Université de Montréal

Exploration bioinformatique des interactions pollen–pistil chez *Solanum chacoense*

par

Valentin Joly

Institut de Recherche en Biologie Végétale
Département de Sciences biologiques
Faculté des Arts et des Sciences

Thèse présentée à la Faculté des Études Supérieures et Postdoctorales
en vue de l'obtention du grade de *Philosophiæ Doctor* (Ph.D.)
en Sciences biologiques

Juillet 2019

Université de Montréal

Institut de Recherche en Biologie Végétale
Département de Sciences biologiques
Faculté des Arts et des Sciences

Cette thèse intitulée

Exploration bioinformatique des interactions pollen–pistil chez *Solanum chacoense*

présentée par

Valentin Joly

a été évaluée par un jury composé des personnes suivantes :

David Morse, Ph.D., président-rapporteur
Daniel P. Matton, Ph.D., directeur de recherche
Daniel Kierzkowski, Ph.D., membre du jury
Jean-Philippe Vielle-Calzada, Ph.D., examinateur externe

Résumé

Rassemblant 107 espèces distribuées dans toute l'Amérique latine, les pommes de terre sauvages (*Solanum* sect. *Petota*) forment un réservoir de germoplasme important pour l'amélioration de leur cousine cultivée (*S. tuberosum*). Pourtant, bien qu'elles vivent en sympatrie et qu'elles aient la capacité physiologique de s'hybrider avec de proches parentes, ces espèces sauvages présentent de fortes barrières d'isolement reproductif. Les résultats préliminaires sur lesquels se fondent cette thèse, présentés au chapitre 1, montrent que les interactions pollen-pistil sont au cœur de cet isolement : dans un pistil hétérospécifique, les tubes polliniques (TP) sont ralentis et leur réceptivité aux signaux chimioattractifs de l'ovule est compromise.

Comme nous le verrons au chapitre 2, l'examen de l'abondante littérature scientifique existante indique qu'un grand nombre de protéines exprimées dans le pistil et le TP ont été caractérisées chez d'autres espèces et que plusieurs contribuent, par leur divergence interspécifique, à l'établissement de points de contrôle précis tout le long du trajet du TP dans le pistil. Parmi elles, les protéines chimioattractives LURE contrôlant le guidage directionnel du TP chez *Torenia* et *Arabidopsis* sont un exemple remarquable de médiateurs moléculaires de l'isolement reproductif. L'objectif de cette thèse est d'isoler des gènes candidats jouant des fonctions analogues chez les Solanacées tubéreuses. C'est dans ce souci que nous avons entrepris d'explorer à grande échelle, dans une approche mêlant bioinformatique, protéomique et transcriptomique, la diversité des gènes exprimés dans les ovules de *S. chacoense*.

Au chapitre 3, nous présentons un nouvel outil de recherche de séquences, KAPPA, spécifiquement dédié à la recherche de peptides riches en cystéines (CRP), une catégorie de protéines particulière dont font partie les chimioattractants ovulaires connus, ainsi que plusieurs autres interactants pollen-pistil. Cet outil a facilité l'identification de nouvelles CRP dans les chapitres suivants. Nous en détaillons l'algorithme, en présentons les principales fonctions et fournissons des résultats démontrant ses performances sur des jeux de données calibrés.

Dans le chapitre 4, nous décrivons une étude du sécrétome ovulaire de *S. chacoense* comparant le contenu protéique des exsudats d'ovules matures (qui sont capables d'attirer le TP) et

d'ovules immatures (qui en sont incapables). Nous présentons la méthode *tissue-free gravity-extraction method* (tf-GEM) utilisée pour collecter les exsudats et détaillons l'analyse bioinformatique effectuée à partir des résultats de spectrométrie de masse. Nous passons en revue les 305 protéines sécrétées identifiées dans les exsudats d'ovules, dont environ la moitié sont régulées entre les deux conditions à l'étude, analysons les catégories fonctionnelles associées, et discutons des applications possibles pour la découverte de gènes candidats.

Le chapitre 5 présente les résultats d'une analyse par biopuce de la réponse transcriptionnelle des ovaires de *S. chacoense* à différents types de pollinisation : conspécifique compatible, conspécifique incompatible, et hétérospécifique compatible. Nous montrons que des gènes ovulaires sont activés ou réprimés à distance dès les premières heures suivant la pollinisation et que cette réponse s'amplifie et devient spécifique à chaque génotype mâle au fur et à mesure de la progression des TP dans le style jusqu'à la fécondation. Nous discutons du rôle de différents médiateurs, notamment l'éthylène, dans la transmission à longue distance du signal de pollinisation.

Au chapitre 6, nous présentons les résultats d'une analyse transcriptomique de plus grande profondeur réalisée par séquençage d'ARN sur des ovules de *S. chacoense* capables (ovules sauvages matures) ou incapables (ovules sauvages immatures, ovules du mutant *frk1* dépourvus de sac embryonnaire) d'attirer les TP. Nous montrons comment notre méthode hybride, recourant à la fois à des assemblages *de novo* et *ab initio* contre le génome récemment publié de *S. chacoense*, nous a conduits à obtenir un jeu de 4353 transcrits enrichis dans l'ovule. Nous détaillons les améliorations portées aux annotations existantes du génome, avec notamment la découverte de nouveaux loci codant pour des CRP reproductives. Nous présentons une analyse du profil d'expression des gènes dans l'ovule, discutons des catégories fonctionnelles associées, et terminons avec l'identification d'une liste de chimioattractants candidats.

Enfin, le chapitre 7 présente en détail ces chimioattractants candidats, les techniques utilisées pour exprimer et purifier ces protéines en bactéries et en levures, ainsi que les dispositifs microfluidiques élaborés pour en vérifier la capacité à attirer le tube pollinique. Nous incluons les résultats préliminaires obtenus avec cette méthode et concluons sur les développements possibles de notre projet.

Mots-clés *Solanum*, pommes de terre sauvages, isolement reproductif, ovule, guidage du tube pollinique, chimioattraction, protéines riches en cystéines, transcriptomique, sécrétomique, microfluidique.

Abstract

With 107 species distributed throughout Latin America, wild potatoes (*Solanum* sect. *Petota*) form an important germplasm reservoir for the breeding of their cultivated cousin (*S. tuberosum*). Yet, although they live in sympatry and have the physiological ability to hybridize with close relatives, those wild species exhibit strong reproductive isolation barriers. The preliminary results on which this thesis is based, presented in chapter 1, show that pollen-pistil interactions are at the center of this isolation: in a heterospecific pistil, pollen tubes (PTs) are slowed down and their receptivity to ovular chemoattractive signals is compromised.

As we will see in chapter 2, the review of the abundant existing scientific literature indicates that a large number of proteins expressed in pistils and PTs have been characterized in other species and that several contribute, through their interspecific divergence, to the establishment of specific checkpoints along the entire path of PTs in pistil. Among them, the LURE chemoattractive proteins controlling the directional guidance of PTs in *Torenia* and *Arabidopsis* are an outstanding example of molecular mediators of reproductive isolation. The objective of this thesis is to isolate candidate genes playing similar functions in tuber-bearing Solanaceous species. It is with this in mind that we have undertaken a large-scale exploration of the diversity of genes expressed in the ovules of *S. chacoense*, in an approach combining bioinformatics, proteomics and transcriptomics.

In chapter 3, we present a new sequence search tool, KAPPA, specifically dedicated to the detection of cystein-rich peptides (CRPs), a particular class of proteins that includes known ovular chemoattractants, as well as several other pollen–pistil interactants. This tool has facilitated the identification of new CRPs in the following chapters. We detail the algorithm, present its main functions and provide results demonstrating its performance on calibrated datasets.

In chapter 4, we describe a study of the ovule secretome of *S. chacoense* comparing the protein content of exudates from mature ovules (which are able to attract PTs) and immature ovules (which are unable to do so). We present the tissue-free gravity-extraction method (tfGEM) used to collect exudates and detail the bioinformatics analysis performed on mass spec-

trometry results. We review the 305 secreted proteins identified in ovule exudates, about half of which are regulated between the two conditions under study, analyze the associated functional categories, and discuss possible applications for the discovery of candidate genes.

Chapter 5 presents the results of a microarray analysis of the transcriptional response of *S. chacoense* ovaries to different pollination types: conspecific compatible, conspecific incompatible, and heterospecific compatible. We show that ovarian genes are activated or repressed at a distance as early as the first few hours after pollination and that this response is amplified and becomes specific to each male genotype as the PTs progress in the style until fertilization. We discuss the role of different mediators, including ethylene, in the long-range transmission of the pollination signal.

In chapter 6, we present the results of a deeper transcriptomic analysis performed by sequencing RNAs from PT guidance-competent (mature wild-type) ovules and PT guidance-incompetent (immature wild-type and embryo sac-devoid *frk1* mutant) ovules in *S. chacoense*. We show how our hybrid method, relying on both *de novo* and *ab initio* assemblies against the *S. chacoense* genome published recently, led us to obtain a set of 4353 ovule-enriched transcripts. We detail the improvements made to existing genome annotations, including the discovery of new loci encoding reproductive CRPs. We present an analysis of the gene expression profile in the ovule, discuss associated functional categories, and conclude with the identification of a list of candidate chemoattractants.

Finally, chapter 7 details these candidate chemoattractants, the techniques used to express and purify them in bacteria and yeasts, and the microfluidic devices developed to verify their ability to attract the pollen tube. We include preliminary results obtained with this method and conclude on possible developments of our project.

Keywords *Solanum*, wild potatoes, reproductive isolation, ovule, pollen tube guidance, chemoattraction, cysteine-rich proteins, transcriptomics, secretomics, microfluidics.

Table des matières

Liste des tableaux

Liste des figures

Liste des abréviations

AGP	*arabinogalactan protein* : protéine arabinogalactane
APF	*apoplastic fluid* : fluide apoplastique
At	*Arabidopsis thaliana*
ARN	acide ribonucléique
ARF	*auxin response factor* : facteur de réponse à l'auxine
AUC	*area under curve* : aire sous la courbe
BBR	barley b recombinant
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
BLAST	*Basic Local Alignment Search Tool*
BPC	BASIC PENTACYSTEINE1
bZIP	*basic region/leucine zipper motif*
CCP	*conspecific compatible pollination* : pollinisation conspécifique compatible
CIP	*conspecific incompatible pollination* : pollinisation conspécifique incompatible
CRP	*cysteine-rich protein* or *peptide* : protéine *ou* peptide riche en cystéines
CS$\alpha\beta$	*Cysteine-Stabilized* $\alpha\beta$
cTPI	*cytosolic TPI* : TPI cytosolique
CV	*coefficient of variation* : coefficient de variation
DBA	*day before anthesis* : jour avant l'anthèse
DEFL	*defensin-like* : [protéine] de type défensine
DGE	*differential gene expression* : expression différentielle des gènes
DNA	*deoxyribonucleic acid* : acide désoxyribonucléique
ECA1	EARLY CULTURE ABUNDANT 1
ECS	*extracellular space* : espace extracellulaire
EGN	*Evolutionary Gene and genome Network*
EGZ	*extended glycine-zipper*
EIL	ETHYLENE-INSENSITIVE 3-like
EIN3	ETHYLENE-INSENSITIVE 3
ERF	*ethylene-responsive element binding factor*
ES	*embryo sac* : sac embryonnaire

EST	*expressed sequence tag* : marqueur de séquence exprimée
FASTA	*FAST-all*
FDR	*false discovery rate* : taux de faux positifs
FG	*female gametophyte* : gamétophyte femelle
FW	*fresh weight* : masse fraîche
GABA	*γ-amino butyric acid* : acide γ-amino-butyrique
GABA-T	GABA-transaminase
GO	*Gene Ontology*
GPI	glycosylphosphatidylinositol
GSP	*generally secreted protein* : protéine sécrétée de manière non-spécifique
HAP	*hour after pollination* : heure après pollinisation
HCP	*heterospecific compatible pollination* : pollinisation hétérospécifique compatible
HMM	*hidden Markov model* : modèle de Markov caché
Hs	*Homo sapiens*
HSP	*heat shock protein* : protéine de choc thermique
IFR	*isoflavone reductase* : isoflavone réductase
KAPPA	*Key Aminoacid Pattern-based Protein Analyzer*
KEGG	*Kyoto Encyclopedia of Genes and Genomes*
KS	Kolmogorov-Smirnov
LC	*liquid chromatography* : chromatographie liquide
LSP	*leaderless secretory protein* : protéine sécrétée dépourvue de peptide signal
LTP	*lipid-transfer protein* : protéine de transfert des lipides
MADS	*MCM1, AGAMOUS, DEFICIENS, and SRF*
MCPI	*metallocarboxypeptidase inhibitor* : inhibiteur de métallocarboxypeptidase
MIKC	*MADS-box, intervening, keratin-like, and C-terminal domains*
Mm	*Mus musclus*
mRNA	*messenger RNA* : ARN messager
MS	*mass spectrometry* : spectrométrie de masse
MUSCLE	*MUltiple Sequence Comparison by Log-Expectation*
MYB	*myeloblastosis virus*
NAC	*NAM, ATAF, and CUC*
NBT	*nitro-blue tetrazolium* : nitrobleu de tétrazolium
NCBI	*National Center for Biotechnology Information*

NGS	*next-generation sequencing* : séquençage de nouvelle génération
nOSP	*non-specific OSP* : OSP non spécifique
ORF	*open reading frame* : cadre de lecture ouvert
Os	*Oryza sativa*
OSP	*ovule-secreted protein* : protéine sécrétée par l'ovule
oOSP	*ovule-specific OSP* : OSP spécifique à l'ovule
PPEB	*post-pollination ethylene burst* : pic d'éthylène post-pollinisation
PAML	*Phylogenetic Analysis by Maximum Likelihood*
PDMS	*polydimethylsiloxane* : polydiméthylsiloxane
PELPIII	Class III PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN
PFAM	*Protein FAMily*
PHI-BLAST	*Pattern Hit Initiated BLAST*
PlantSecKB	*Plant Secretome and subcellular proteome KnowledgeBase*
PR	PATHOGENESIS-RELATED
PSI-BLAST	*Position-Specific Iterated BLAST*
PSSM	*position-specific scoring matrix* : matrice de poids score-position
PT	*pollen tube* : tube pollinique
RLK	*receptor-like kinase* : protéine kinase de type récepteur
RLP	*receptor-like protein* : protéine de type récepteur
RNA	*ribonucleic acid* : acide ribonucléique : **ARN**
RNA-seq	*RNA sequencing* : séquençage d'ARN
RNase	*ribonuclease* : ribonucléase
ROC	*receiver operating characteristic* : fonction d'efficacité du récepteur
ROS	*reactive oxygen species* : espèce(s) réactive(s) de l'oxygène
Sc	*Solanum chacoense*
SEM	*scanning electron microscopy* : microscopie électronique à balayage
SEP	*stigma exudate protein* : protéine de l'exsudat stigmatique
Sg	*Solanum gandarillasii*
SI	*self-incompatibility* : auto-incompatibilité
SIV	semi-*in vivo*
Sm	*Solanum microdontum*
SP	*signal peptide* : peptide signal
SPRR2	Type 2 SMALL PROLINE-RICH
St	*Solanum tarijense*
TF	*transcription factor* : facteur de transcription

tf-GEM	*tissue-free Gravity Extraction Method*
TMH	*transmembrane helix* : hélice transmembranaire
TNR	*true negative rate* : taux de vrais négatifs
TP	tube pollinique
TPI	*triose phosphate isomerase* : triose phosphate isomérase
TPR	*true positive rate* : taux de vrais positifs
TTS	TRANSMITTING TISSUE-SPECIFIC
UPS	*unconventional protein secretion* : sécrétion non conventionnelle des protéines
UO	*unpollinated ovule* : ovule non-pollinisé
UTR	*untranslated region* : région non traduite
VIC	*vacuum-infiltration centrifugation*
XTH	xyloglucan endotransglucosylase/hydrolase

Élévation

Au-dessus des étangs, au-dessus des vallées,
Des montagnes, des bois, des nuages, des mers,
Par delà le soleil, par delà les éthers,
Par delà les confins des sphères étoilées

Mon esprit, tu te meus avec agilité,
Et, comme un bon nageur qui se pâme dans l'onde,
Tu sillonnes gaiement l'immensité profonde
Avec une indicible et mâle volupté

Envole-toi bien loin de ces miasmes morbides,
Va te purifier dans l'air supérieur,
Et bois, comme une pure et divine liqueur,
Le feu clair qui remplit les espaces limpides

Derrière les ennuis et les vastes chagrins
Qui chargent de leur poids l'existence brumeuse,
Heureux celui qui peut d'une aile vigoureuse
S'élancer vers les champs lumineux et sereins ;

Celui dont les pensées, comme des alouettes,
Vers les cieux le matin prennent un libre essor,
— Qui plane sur la vie, et comprend sans effort
Le langage des fleurs et des choses muettes !

Charles Baudelaire, in *Les Fleurs du mal* (1857)

Remerciements

Mes premiers remerciements vont à mes professeurs, à commencer par mon directeur de recherche Daniel P. Matton, qui m'a accueilli et fait confiance pendant toutes ces années, et qui m'a aussi fait beaucoup voyager. Je remercie également ses confrères David Morse et Daniel Kierzkowski (Université de Montréal), et Jean-Philippe Vielle-Calzada (Cinvestav, Institut National Polytechnique du Mexique), d'avoir accepté de réviser et d'évaluer cette thèse. Ma reconnaissance va également à plusieurs autres professeurs et chercheurs montréalais qui ont fait partie de mes comités-conseil ou qui m'ont soutenu et conseillé de multiples manières : Jean Rivoal et Sonia Dorion, Mario Cappadocia, Anja Geitmann, Sébastien Halary, Simon Joly, B. Franz Lang, Daniel Schoen.

Merci, ou plutôt どうもありがとうございました, au professeur Tetsuya Higashiyama, de l'Université de Nagoya (Japon), qui m'a accueilli dans son équipe à l'été 2016, me donnant ainsi la chance de travailler dans *le* laboratoire de référence dans notre domaine. Des remerciements tout particuliers vont à Masahiro Kanaoka, qui m'a beaucoup appuyé au cours de ce séjour, tant scientifiquement qu'humainement. Merci également à Naoki Yanagisawa et Nao Kamiya d'avoir bien voulu partager leur expertise en matière de microfluidique et de purification de protéines, respectivement, et merci plus généralement à l'équipe tout entière pour la courtoisie de son accueil. Merci enfin à Kathy Tamai pour son amitié et sa complicité, qui se sont maintenues depuis ce séjour.

Tack so mycket à Johan Edqvist et Kristina Blomqvist de l'Université de Linköping (Suède), qui m'ont chaleureusement accueilli et gracieusement hébergé à deux reprises, me donnant ainsi l'occasion de collaborer à leurs recherches sur les LTP et de me former à l'expression de protéines. Merci aussi à eux et à leur famille pour les mémorables sorties botaniques et ornithologiques, et pour l'initiation à l'art de vivre suédois.

Many thanks au professeur William J. Swanson ainsi qu'à son collègue Jan E. Aagaard, de l'Université du Washington (États-Unis), qui m'ont reçu à Seattle pour un séjour de deux mois en 2014. Je remercie en particulier Renee George, du Whitehead Institute, qui m'a accompagné de manière très proactive dans les analyses de variants génétiques.

Remerciements

Muchas gracias à Franco E. Chiarini de l'Université Nationale de Córdoba (Argentine), qui a été notre guide dans les Andes pour collecter certains des échantillons nécessaires à ce projet. Son érudition botanique a été un atout précieux pour repérer et identifier nos spécimens, et sa bonne humeur a égayé nos soirées.

Je souhaiterais remercier également mes collègues, en particulier Yang Liu, Claire Viallet et Faïza Tebbji, qui m'ont précédé dans ce projet et qui y ont grandement contribué, ainsi que Caroline Daigle, qui m'a formé lors de mon arrivée au laboratoire. Je pense aussi aux autres membres de l'équipe, Audrey Loubert-Hudon, Fangwen Bai, Samuel B. Drory et Rachid Benhamman, qui m'ont fait une place et aidé plus d'une fois, ainsi qu'aux membres des laboratoires Morse, Cappadocia et Rivoal.

Une mention spéciale doit toutefois être décernée à Benjamin Mazin, qui aura tour à tour été partenaire de labo, psychanalyste (gratuit), expert en humour absurde, critique d'art dramatique, œnologue, herboriste, aromathérapeute, et même coach sportif. En bref : un ami. Accessoirement, c'est aussi à Benjamin que je dois la rencontre de mon conjoint et celle de mon futur directeur de post-doc. Merci Benjamin !

Ce travail n'aurait pas été possible sans l'implication des nombreux stagiaires qui se sont succédé au laboratoire. Merci en particulier aux cinq étudiants du programme des futurs leaders dans les Amériques (PFLA) que j'ai eu la chance de recevoir. *Muchísmias gracias* à Mariana Quiroga et Laura González (Université Nationale de Córdoba, Argentine), Carlos Bravo (Université Nationale Autonome du Mexique) et Federico Ceriotti (Université Nationale de Cuyo, Argentine), et *muito obrigado* à Kelly Rodrigues (Université fédérale du Rio Grande do Sul, Brésil). Merci également aux six stagiaires de premier cycle que j'ai encadrés au laboratoire : Tissicca Hour, Ella Gangbe, Francis Banville et Maude Dorval (Université de Montréal), Andréa Davrinche (Université Paris-VI) et Anna Zaro (Université de Barcelone).

Merci à la microscopiste Louise Pelletier d'avoir bien voulu me faire profiter de son expertise, sa bonne humeur et sa générosité. Merci également au personnel technique et de soutien de l'institut, en particulier Nicolas Boivin, Denis Lauzer et Dave Smith, pour leur aide et leur patience. Merci enfin au personnel des serres du Jardin botanique de Montréal, notamment l'horticultrice spécialisée Nancy Robert et ses collègues de la fin de semaine, la phytopathologiste Marie-Michèle Bouchard et la contre-maître Lise Lacouture.

Remerciements

Ce doctorat a bénéficié du soutien financier de plusieurs organismes. Je remercie notamment le Fonds de recherche du Québec – Nature et technologies (FRQNT ; subvention nº 2011-PR-148467 et bourse de doctorat B2 nº 199137) et Hydro-Québec, qui m'a octroyé à deux reprises une bourse d'excellence d'un an. Mes remerciements vont également au Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG ; subvention nº RGPIN-2014-03883) pour le financement du laboratoire. J'aimerais aussi exprimer ma gratitude aux donateurs du fonds Marie-Victorin de l'Institut de recherche en biologie végétale (IRBV) et du Fonds de bourses en sciences biologiques (FBSB) du Département, ainsi qu'à ceux de la bourse d'excellence Catherine-Frédette et de la bourse au mérite de la Faculté des études supérieures et postdoctorales (FESP). Merci également aux Amis du Jardin botanique de Montréal de m'avoir octroyé une bourse Pehr-Kalm et, au passage, de nous avoir aidés financièrement avec nos installations de culture à l'IRBV.

Je suis très reconnaissant d'avoir pu voyager dans le monde grâce au programme d'été de la Société japonaise pour la promotion de la science (JSPS), aux bourses de stages internationaux du FRQNT octroyées par l'entremise du Centre SÈVE, au fonds de bourses Jacques-Rousseau de l'IRBV, à la bourse de voyage G.-H.-Duff de la Société canadienne de biologie végétale (SCBV), aux subventions de voyage octroyées par la Direction des relations internationales (DRI) de l'Université de Montréal, à la bourse d'appui à la diffusion des résultats de recherche de la FESP, et à la subvention PARSECS de la Fédération des associations étudiantes du campus de l'Université de Montréal (FAÉCUM).

Merci enfin à mes deux familles de chaque côté de l'Atlantique. Merci à mes grand-mères Madeleine et Léontine, qui ne sont pas étrangères à mon goût pour les fleurs et à ma vocation de biologiste, ainsi qu'à Alain et Claudine, Jean et Anne-Marie, Paule et Robert, et tant d'autres. Merci aux Québécois de m'avoir fait une place parmi eux. Merci à Maxime pour son amour sans frontières, de Montréal à Reykjavík.

CHAPITRE 1

Introduction et problématique

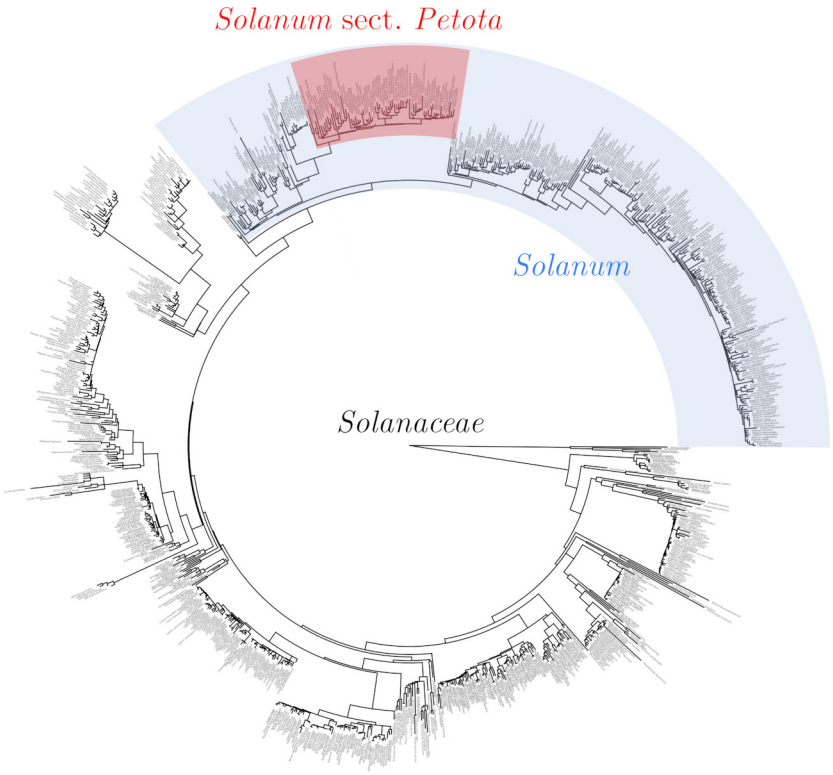
1.1 Présentation des espèces d'intérêt

La famille des Solanacées, taxon âgé d'environ 10 millions d'années, comprend 92 genres et 2300 espèces[1] dont l'organisation phylogénétique est décrite à la figure 1.1a. Plusieurs d'entre elles présentent un intérêt économique majeur : le tabac, le piment, l'aubergine, plusieurs plantes médicinales ainsi que la tomate et la pomme de terre. Ces deux dernières appartiennent au genre *Solanum*, qui regroupe environ 1200 espèces réparties en plusieurs sections, dont *Petota*, celle des pommes de terre.

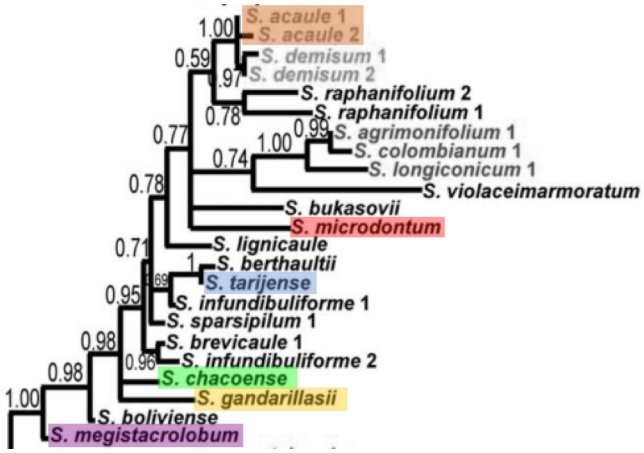
Domestiquée il y a 6000 à 10 000 ans sur les hauteurs andines du Sud du Pérou[2,3], la pomme de terre cultivée (*Solanum tuberosum* L.) est rapidement devenue une denrée alimentaire fondamentale dans le monde entier[4]. Avec une production mondiale de 377 millions de tonnes en 2016, pour une valeur brute totale de 111 milliards de dollars, la pomme de terre se hisse en quatrième position parmi les plantes cultivées dans le monde, derrière le riz, le blé et le maïs[5].

Les défis à relever actuellement pour l'amélioration de la pomme de terre comprennent une augmentation des rendements, une plus grande compatibilité avec les besoins de l'industrie agroalimentaire, une meilleure résistance aux stress abiotiques (p. ex., la sécheresse) et aux pathogènes[6,7]. Pour identifier des traits d'intérêt à introduire chez *S. tuberosum*, les améliorateurs peuvent tirer parti de la grande diversité botanique des pommes de terre, qui comprend 4 espèces cultivées et 107 espèces sauvages divisées en 4 clades[3], ditribuées dans toute l'Amérique latine, du Nouveau-Mexique à la Patagonie[8,9].

Adaptées à un large éventail de conditions environnementales et climatiques, les pommes de terre sauvages possèdent des traits d'intérêt pour l'agriculture, comme une plus grande résistance aux pathogènes et à la sécheresse. Plusieurs études génomiques effectuées récemment dans



(a) Phylogénie des Solanacées. Position du genre *Solanum* et de la section des pommes de terre (*Solanum* sect. *Petota*) dans la famille des Solanacées[10].



(b) Extrait du phylogramme bayésien de *Solanum* sect. *Petota* réalisé par Rodríguez et Spooner en étudiant la nitrate réductase. La longueur des branches représente le nombre estimé de substitutions par site[11].

Fig. 1.1. Phylogénie des espèces d'intérêt.

le clade des pommes de terre (*Solanum* sect. *Petota*) ont en effet mis au jour la grande diversité génétique des espèces sauvages et son utilité pour l'amélioration variétale[12,13].

Parmi elles, *Solanum chacoense*, une espèce du clade 4a (figure 1.1), intéresse tout particulièrement la recherche agronomique, au point que son génome a récemment été séquencé[14]. En effet, *S. chacoense* présente une grande résistance à une vaste gamme de pathogènes, comprenant le virus Y de la pomme de terre[15,16], les bactéries causant le flétrissement et la verticilliose[17,18] ainsi que la gale commune[19,20], ou encore les champignons responsables du mildiou[21] ainsi que de la pourriture du collet et de la maladie dite de la jambe noire[22].

S. chacoense possède également un métabolisme secondaire particulier produisant de grandes quantités de glycoalcaloïdes spéciaux, comme les leptines et les leptinines[23]. Ces composés sont efficaces contre le doryphore de la pomme de terre, un insecte phytoravageur particulièrement nuisible aux cultures[24]. Des travaux sont actuellement menés pour mieux comprendre les gènes impliqués dans ces voies métaboliques chez *S. chacoense*[25,26] et pour les introduire chez la pomme de terre cultivée[27,28,29,30].

1.2 Problématique de l'isolement reproductif

Cet intérêt grandissant pour les parents sauvages d'espèces cultivées (*crop wild relatives*, CWR) dans le contexte agronomique est confronté au problème de l'isolement reproductif, c'est-à-dire des barrières à l'hybridation séparant les espèces qu'il devient nécessaire de surmonter pour effectuer des introgressions[31]. L'existence, chez les pommes de terre sauvages, d'un grand nombre d'espèces entretenant des relations phylogénétiques plus ou moins étroites en font un cas idéal pour étudier les mécanismes sous-jacents à cet isolement.

Dans ce travail, nous avons choisi de nous intéresser plus particulièrement à *S. chacoense* ainsi qu'à trois de ses plus proches parentes présentées au tableau 1.1. L'espèce *S. chacoense*, diploïde, est déjà utilisée comme modèle alternatif à la pomme de terre cultivée tétraploïde pour la recherche sur la reproduction sexuée. Par exemple, *S. chacoense* a été utilisée pour décrypter les mécanismes gouvernant l'auto-incompatibilité gamétophytique à S-RNase[32,33,34,35], pour mieux comprendre la cytomécanique du tube pollinique[36,37,38], et pour explorer plusieurs voies de signalisation reproductives impliquant des peptides RALF[39,40], des MAP kinases[41,42,43,44] et des récepteurs kinases[45,46,47,48].

Les quatre espèces d'intérêt dans ce travail présentent un paradoxe. En effet, d'une part,

notre travail en serre nous a permis de montrer que l'obtention d'hybrides est possible dans toutes les combinaisons formées par nos espèces d'intérêt. De fait, elles présentent le même EBN (*Endosperm Balance Number*, voir tableau 1.1). Ce niveau de ploïdie « effectif », qui caractérise les Solanacées, doit être le même pour que deux espèces puissent se croiser[49,50]. Par ailleurs, nous avons pu constater la présence d'hybrides interspécifiques sur le terrain, lors d'un voyage dans la province argentine de Salta en 2012, dans les herbiers de l'Université Nationale de Córdoba (Argentine) et de l'INTA Balcarce (Argentine) ou encore dans la banque de germoplasme américaine NRSP6, notamment dans les cas suivants : *S. acaule* × *S. chacoense* [a] et *S. microdontum* × *S. chacoense*[b], mais aussi *S. acaule* × *S. microdontum*[c], *S. acaule* × *S. megistacrolobum*[d] et *S. acaule* × *S. tarijense*[e].

Tab. 1.1. Liste des espèces d'intérêt, avec leur code international à trois lettres, leur niveau de ploïdie, leur *Endosperm Balance Number*, le numéro de référence des plantes dans les serres de l'IRBV et le numéro des accessions correspondantes à la banque américaine de germoplasme de pommes de terre NRSP6.

Espèce	Code	Ploïdie	EBN	Génotypes	Accessions NRSP6
S. chacoense	chc	$2n = 2x$	2	V22, G4	PI458314 × PI230582
S. gandarillasii	gnd	$2n = 2x$	2	S14, S15	PI597751, PI545864
S. tarijense	tar	$2n = 2x$	2	S11, S12	PI195206, PI473243
S. microdontum	mcd	$2n = 2x$	2	AS20, S6	s.o.*, PI500041

*Le génotype AS20 a été directement collecté sur le terrain dans la province de Salta (Argentine).

Toutefois, en dépit de l'existence de ces hybrides, l'examen des herbiers précédemment cités ainsi que le travail sur le terrain nous ont permis de constater que les populations des espèces parentales « pures » restent largement majoritaires dans la nature. Dès lors, comment concilier le maintien desdites populations et la survenue marginale d'hybridations interspécifiques ? Comment les espèces parentales ont-elles évité d'être diluées au fil des croisements entre espèces ? Plusieurs mécanismes, que nous détaillons ci-après, sont classiquement évoqués pour expliquer l'isolement reproductif des plantes à fleurs.

a. Herbier de l'INTA Balcarce. Accessions : Bal 69005 Haw 6658 x ; Bal 70002 x Hof 1509 x ; Bal 70013 x Hof 1571 x ; Bal 70072 x Hof 1734 x ; Bal 71129 x Oka 3921 x.

b. Herbier de l'INTA Balcarce. Accessions : Bal 73013 A OL 4819 A ; Bal 75058 Oka 6119 ; Bal 75061 Oka 6127 ; Bal 75063 Oka 6129 ; Bal 75064 Oka 6130 ; Bal 78156 B Oka 6889 B ; Bal 78157 Oka 6890 ; Bal 78159 Oka ; Bal 90056 ; Bal 90067 SCl 4599.

c. Herbier de l'INTA Balcarce. Accession Bal 83144 A Oka 7636 A.

d. Herbier de l'INTA Balcarce. Accessions : Bal 66058 x HHR 3904 x ; Bal 70002 x Hof 1509 x ; Bal 70022 A Hof 1608 ; Bal 71007 A Oka 3686 ; Bal 71108 x Oka 3885 x ; Bal 71123 x Oka 3914 x ; Bal 71127 x Oka 3918 x ; Bal 71129 x Oka 3921 x ; Bal 71185 B Oka 3988 B ; 1033 Bal 71206 Oka 4019 et autres.

e. Herbier de l'INTA Balcarce. Accession Bal 70002 x Hof 1509 x.

a. Ségrégation spatio-temporelle des populations

Le cas le plus simple et le plus commun de barrière à l'hybridation est la ségrégation spatiale des populations[51,52]. Il s'agit d'une barrière prézygotique prépollinisation, c.-à-d. qu'elle vise à prévenir la formation du zygote en empêchant la pollinisation croisée. La cause de cette séparation peut être exogène (ex. : un obstacle physique durable comme une chaîne de montagnes sépare les deux populations) ou endogène (ex. : les deux populations ne sont pas aptes à vivre sur la zone de distribution de l'autre en raison d'une spécialisation à l'habitat[53]). On a d'ailleurs identifié certains gènes associés à ces barrières, comme *HaCDPK3* chez le tournesol[54] ou *AtHKT1* chez *Arabidopsis*[55].

Toutefois, ces barrières géographiques semblent négligeables dans le cadre de notre travail, car les populations de nos six espèces coexistent dans une vaste zone de sympatrie, qui s'étend du sud-est du Pérou au nord-ouest de l'Argentine (figure 1.2), en dépit du fait que des zones de relative allopatrie existent jusqu'en Équateur (*Solanum acaule*) et dans la région de Buenos Aires (*S. chacoense*). Un voyage de terrain dans la province argentine de Salta, située au cœur de la zone de sympatrie, nous a d'ailleurs permis de constater que des individus hétérospécifiques pouvaient cohabiter très étroitement dans plusieurs types d'habitat : prairies humides, alpages, forêts tropicales, etc.

Lorsqu'elle ne peut être spatiale, la séparation des populations peut aussi être temporelle : l'hybridation entre des individus d'espèces différentes peut en effet être limitée par le décalage des périodes de floraison. Des gènes gouvernant ce phénomène ont été identifiés chez *Arabidopsis*[56,57,58], le riz[59], la tomate[60], le blé[61], l'orge[62] ou encore le tournesol[63]. Toutefois, là encore, il semble peu probable que des barrières temporelles assurent l'isolement reproductif de nos espèces d'intérêt, car celles-ci fleurissent toutes au même moment durant l'été austral, de décembre à mars, *S. chacoense* présentant toutefois une période de floraison plus longue, de novembre à juillet.

b. Spécialisation du pollinisateur

Une autre barrière prézygotique typique est la mise en œuvre de stratégies d'attraction différentielle des pollinisateurs. Ainsi, on a décrit à ce jour de nombreux exemples de taxons chez qui la forme, la couleur, le contenu en nectar ou l'odeur des fleurs varie d'une espèce à l'autre[64,65].

Plusieurs gènes gouvernant la couleur des fleurs et divergeant entre les espèces ont été décrits. Ces gènes sont souvent impliqués dans la biosynthèse des pigments en codant pour des

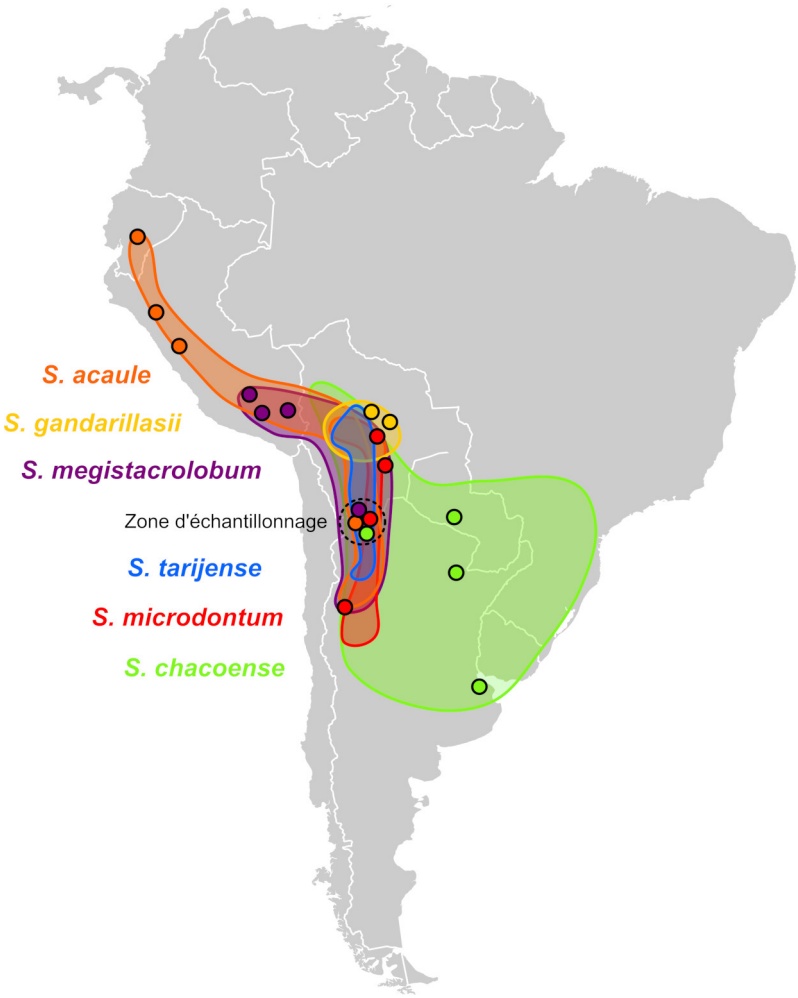


Fig. 1.2. Distribution des populations des espèces d'intérêt en Amérique du Sud. La zone en pointillés signale l'origine de la plupart des génotypes utilisés dans ce projet, à l'exception des individus de *S. gandarillasii* qui proviennent de la zone vosine en jaune. (Carte adatpée de l'*Atlas of wild pototes*[9].)

enzymes ou des régulateurs de leurs voies de synthèse. On peut citer certains facteurs de transcription de la famille *MYB* comme *AN2* chez *Petunia*[66,67] ou *ROSEA1/2* et *VENOSA* chez *Antirrhinum*[68,69], de même que le gène *F3'H* d'*Ipomoea*[70,71,72,73] ou encore certains loci à effet pléiotrope responsables d'un changement de pollinisateur (entre colibris et abeilles) chez *Mimulus*[74,75].

Toutefois, en dépit de légères variations dans la taille, la forme et la pigmentation de la corolle, le parfum ou encore la structure inflorescentielle, l'organisation florale est assez homogène chez ces six espèces (figure 1.3). Il arrive même que des génotypes conspécifiques présentent des morphologies plus éloignées l'une de l'autre que des génotypes d'espèces différentes. Cette plas-

ticité phénotypique des pommes de terre est d'ailleurs à l'origine de remaniements fréquents de leur taxonomie. Par exemple, une de nos espèces d'intérêt, *S. tarijense*, a récemment « absorbé » *S. berthaultii* en raison d'une grande proximité morphologique[76] confirmée par la génétique[77].

(a) *S. chacoense* **(b)** *S. gandarillasii* **(c)** *S. tarijense* **(d)** *S. microdontum*

Fig. 1.3. Morphologie florale des espèces d'intérêt. L'inflorescence cymeuse présente de 2 à 25 fleurs, pour un diamètre total de 2 à 15 cm. Elle est portée par un pédoncule de 2 à 11 cm qui se ramifie en courts pédicelles pubescents (1 à 2 cm) aboutissant chacun à une fleur actinomorphe et pentamère. Le périanthe est hétérochlamydé, gamosépale et gamopétale. La corolle étoilée à pentagonale, de 1 à 4 cm de diamètre, comprend cinq rayons collenchymateux reliés par de légers voiles parenchymateux blancs à violacés. L'androcée est iso-, haplo- et synstémone. Les anthères de 5 à 7 mm, extrorses, lancéolés, connivents, supportés par des filets de 1 à 2 mm, s'ouvrent sur des pores de déhiscence supérieurs. L'ovaire supère, bicarpellé, à placentation axile, de couleur vert glauque, se prolonge par un style droit d'environ 1 mm de largeur par 8 à 15 mm de longueur, qui se termine par un stigmate vert, cireux et légèrement arrondi. (Photographies : D.P. Matton, 2007)

On peut ainsi suggérer que nos espèces d'intérêt sont vraisemblablement pollinisées de manière indifférenciée par les mêmes insectes. Par exemple, les tomates sauvages, taxon proche de nos espèces d'intérêt réparti au Pérou et au Chili, seraient toutes pollinisées dans 99.996 % des cas par un seul insecte hyménoptère de l'espèce *Centris buchholzi* (Boris Igić, données non publiées). Ceci nous amène à dire que la pollinisation croisée entre pommes de terre sympatriques devrait être fréquente en nature.

c. Barrières postzygotiques

Il existe une autre catégorie de barrières contre l'hybridation très étudiées chez les plantes, que l'on qualifie de postzygotiques en ce qu'elles interviennent après la formation du zygote hybride pour diminuer sa viabilité ou, au moins, sa fertilité.

L'étude d'hybrides non viables ou stériles a mené dès 1934 à la définition de ce que l'on dénommera plus tard le modèle de Dobzhansky-Muller[78] (DM). D'après celui-ci, la divergence de deux espèces à partir d'un même ancêtre commun est marquée par des mutations réciproques sur deux loci particuliers que l'on peut appeler *A* et *B*. Alors que l'ancêtre commun

présentait le génotype *AABB*, ses deux espèces-filles présentent respectivement les génotypes *AAbb* et *aaBB*. Si elles se croisent, l'hybride aura un génotype *AaBb*. La mise en contact des allèles mutés *a* et *b* conduit alors à la nécrose de l'hybride.

Chez les plantes, les incompatibilités DM reposent souvent sur des gènes complémentaires de résistance aux maladies. Dans ce cas, la formation de l'hybride met au contact deux partenaires dont l'interaction déclenche le mécanisme de nécrose habituellement utilisé pour détruire un organe touché par une maladie. Nous savons que c'est le cas chez la tomate avec les gènes *Cf-2* et *RCR3*[79,80], chez le riz avec *hbd2*[81] ou encore chez la laitue avec *Rin4*[82]. En dehors des gènes associés aux maladies, on peut citer *HPA1* et *HPA2* chez *Arabidopsis*[83].

Par ailleurs, plusieurs cas d'incompatibités cytonucléaires ont été décrits. Ils résultent d'un conflit entre les génomes cytoplasmiques (notamment le génome mitochondrial) provenant de la femelle et le génome nucléaire, pour moitié hérité du mâle[84]. Ce conflit génère le plus souvent un phénomène appelé stérilité mâle cytoplasmique (*cytoplasmic male sterility*, CMS) marqué par l'absence du pollen voire des anthères entiers chez les hybrides[85].

La CMS est due à l'existence de loci mitochondriaux codant en général pour des ORF chimériques de protéines essentielles à la respiration cellulaire, comme l'ATP-synthase[86]. Lorsqu'il y a croisement intraspécifique entre individus de génotypes proches, la stérilité mâle causée par ces loci est inhibée par l'activité de gènes nucléaires appelés « gènes restaurateurs », qui font notamment partie de la famille PPR (*pentatricopeptide repeat*)[87,88]. Ainsi, des couples formés par un locus CMS mitochondrial et un locus restaurateur nucléaire peuvent évoluer entre deux espèces divergentes selon le modèle DM et ainsi former une barrière postzygotique forte contre l'hybridation interspécifique. Pour l'heure toutefois, la CMS a uniquement été décrite entre des génotypes d'une même espèce cultivée[89].

Qu'en est-il de nos espèces d'intérêt ? Comme précisé plus haut, bien que rares, des individus hybrides existent dans la nature. De plus, en serres, il est possible d'obtenir des individus hybrides viables et capables de fleurir en croisant *S. chacoense* avec n'importe laquelle des autres espèces d'intérêt. (Il faut alors forcer le croisement en effectuant une pollinisation hétérospécifique saturante.) Si l'on ne peut pas formellement exclure l'existence de barrières postzygotiques chez nos espèces de pommes de terre sauvages, il faut malgré tout envisager l'intervention de barrières avant la genèse du zygote.

1.3 Interactions pollen–pistil

Nous venons de voir que les barrières d'isolement reproductif intervenant avant la pollinisation (barrières prézygotiques prépollinisation) ou après la fécondation (barrières postzygotiques) sont insuffisantes pour expliquer le maintien des espèces sauvages de pommes de terre. Ceci nous amène à suggérer l'importance des barrières intervenant entre la pollinisation et la fécondation, qui contrôlent les interactions entre le tube pollinique (partenaire mâle) et le pistil (partenaire femelle). Le lecteur pourra justement se reporter au chapitre 2 pour des explications plus détaillées sur les interactions pollen–pistil, les acteurs moléculaires connus à ce jour et leur lien avec l'isolement reproductif.

La figure 4.1 (page 114) présente une vue schématique des interactions pollen–pistil chez *Solanum*. La pollinisation, assurée par des insectes hyménoptères, marque le début du processus de fertilisation. Elle est immédiatement suivie par l'adhésion du pollen sur les papilles stigmatiques. Les étapes suivantes sont l'hydratation puis la germination du grain de pollen, dont la cellule végétative grandit de façon polaire en formant un tube pollinique qui pénètre entre les papilles stigmatiques pour atteindre le tissu de transmission. Le tube pollinique poursuit sa croissance dans le style et la cellule générative subit une mitose. Chez *S. chacoense*, le tube pollinique entre dans le style 12 heures après la pollinisation (HAP), atteint les deux tiers de sa longueur 24 HAP et son extrémité distale 30 HAP.

Le tube pollinique pénètre alors dans l'ovaire et parvient au voisinage d'un ovule. Des signaux moléculaires émis par l'ovule le guident ensuite vers le funicule puis le micropyle. Il peut alors libérer ses deux noyaux génératifs, dont le premier féconde l'oosphère pour donner le zygote, de ploïdie $2kn$, qui produira l'embryon, alors que second féconde la cellule centrale pour donner la cellule-mère de l'albumen, de ploïdie $3kn$. Cette double fécondation se produit 36 HAP. Le tégument de l'ovule deviendra celui de la graine, qui protégera l'embryon et son albumen. Ces graines seront contenues dans un fruit dérivant de l'ovaire.

Ce projet se doctorat se fonde sur plusieurs résultats expérimentaux préliminaires vérifiant l'implication des interactions pollen–pistil dans l'isolement reproductif de nos espèces d'intérêt, que nous présentons ci-après. Les méthodes expérimentales utilisées sont détaillées dans la légende des figures.

a. Tests de pollinisation mixte

Les tests de pollinisation mixte consistent à mettre en concurrence, dans un même pistil, du pollen de deux espèces différentes. La figure 1.4 en montre le principe, ainsi que le résultat obtenu pour le premier test que nous avons effectué, entre *S. chacoense* et sa plus proche parente dans la phylogénie : *S. gandarillasii*. Lorsque l'on pollinise un pistil de *S. chacoense* avec un mélange équilibré de pollen des deux espèces, on observe que seuls 8 % des descendants sont hybrides. Il existe donc un mécanisme de préférence à l'espèce, impliquant un ou plusieurs points de contrôle entre la pollinisation et la fécondation, y compris chez deux espèces proches.

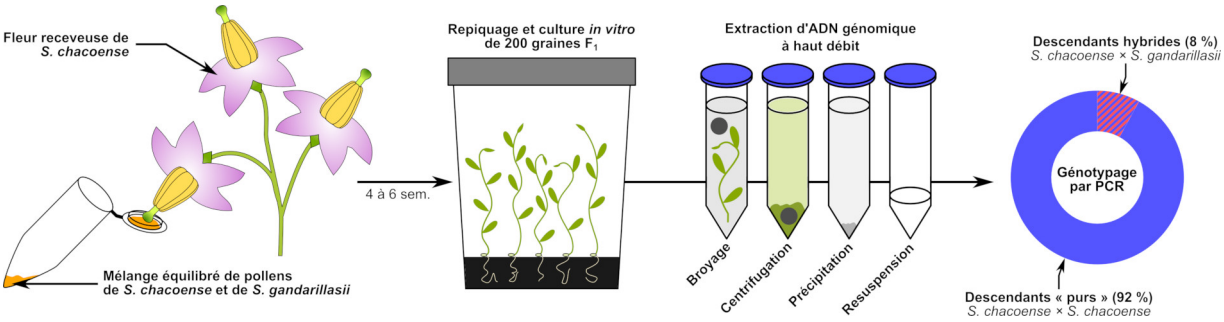


Fig. 1.4. Test de pollinisation mixte impliquant *S. chacoense* et *S. gandarillasii*. Vingt fleurs de *S. chacoense* (G4) ont été pollinisées avec un mélange de 50 mg de pollen de *S. chacoense* (V22) et 50 mg de pollen de *S. gandarillasii* (S15). Un mois plus tard, les graines issues de ce croisement ont été repiquées *in vitro*. Les 200 plantules ainsi obtenues ont subi une extraction d'ADN génomique puis un génotypage par PCR qui a consisté à amplifier un locus discriminant, dont la taille est différente chez les deux espèces. Le génotype de chaque descendant a été identifié après migration électrophorétique des produits de PCR en gel d'agarose.

b. Ralentissement des tubes polliniques hétérospécifiques

Comment expliquer la faible proportion d'hybrides constatée plus haut ? Une hypothèse est que les tubes polliniques sont ralentis lorsqu'ils progressent dans un style hétérospécifique. Nous avons souhaité tester cette hypothèse chez nos espèces d'intérêt en effectuant des cinétiques de croissance du tube pollinique *in vivo*. La figure 1.5 présente la cinétique de croissance comparée du pollen de quatre espèces dans le pistil de *S. chacoense*. Nous pouvons constater que le pollen conspécifique compatible (V22, courbe verte), présente une croissance en deux temps : d'abord une phase de croissance modérée à 170 µm/s de 0 à 12 heures, puis une phase de croissance rapide à 330 µm/s de 12 à 36 heures.

Ce patron biphasique de croissance est caractéristique des espèces à pollen binucléé, c.-à-d. contenant, au moment de la pollinisation, une cellule végétative et une cellule générative[90].

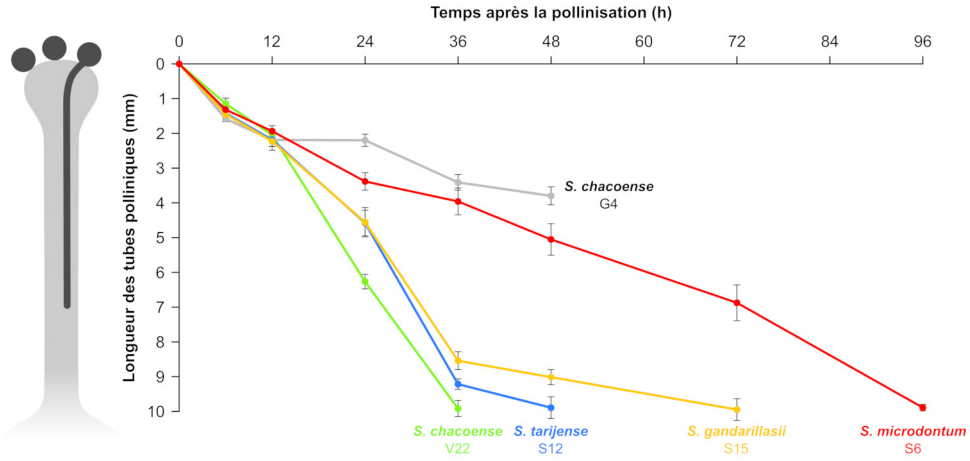
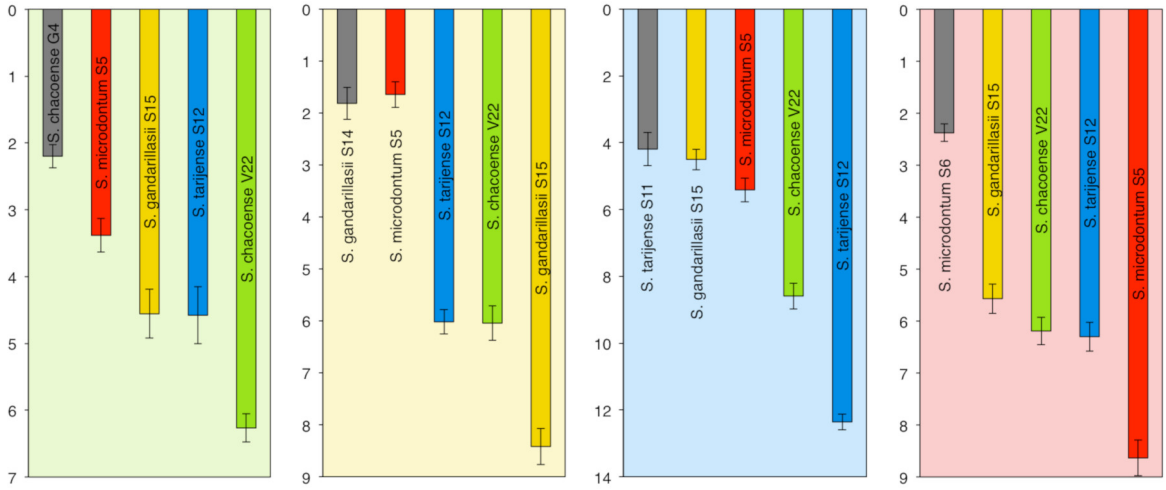


Fig. 1.5. Cinétique de croissance de tubes polliniques de différentes espèces dans le style de *Solanum chacoense*, génotype G4. Les fleurs ont été pollinisées de façon saturante, puis recueillies à intervalles réguliers (0, 6, 12, 24, 36, 48, 72 et 96 HAP). Les styles ont été prélevés et fixés au FAA (éthanol 50 %, H_2O 35 %, acide acétique glacial 10 %, formaline 5 %) pendant une nuit sur agitateur rotatif. Ils ont ensuite été lavés 3 fois au PBS (*phosphate buffer saline*), puis perméabilisés et décolorés au NaOH 2 M pendant 24 h à température ambiante sur un agitateur orbital à 100 tr/min. Après 3 lavages supplémentaires au PBS, les styles ont été incubés dans quelques gouttes de solution décolorée de bleu d'aniline à 0,1 % dans untampon K_3PO_4 à pH 7,5. Les spécimens ont été observés au grossissement 50 × sur un microscope à fluorescence *Axio Observer.Z1* (*Zeiss*) doté d'un filtre de type DAPI. Les valeurs reportées dans le graphique sont les moyennes des longueurs d'au moins 30 tubes polliniques par échantillon.



(a) *S. chacoense* **(b)** *S. gandarillasii* **(c)** *S. tarijense* **(d)** *S. microdontum*

Fig. 1.6. Longueur comparée de tubes polliniques en croissance dans le style des quatre espèces. Les mesures, données en millimètres, ont été faites 24 HAP (*S. chacoense* et *S. microdontum*) ou 36 HAP (*S. tarijense* et *S. gandarillasii*).

La première phase, lente, est dite autotrophique : le pollen tire son énergie de ses propres ressources, bien qu'il dépende à la marge de l'eau et des sucres du pistil. La seconde phase, plus rapide, est dite hétérotrophique : le pollen peut prélever des nutriments du pistil dans le tissu de transmission.

Le basculement entre ces deux phases se produit lorsque le tube pollinique franchit la zone de transition stigmate–style, un « goulet d'étrangelement » physique connu pour être l'endroit (i) où les tubes polliniques conspécifiques incompatibles (G4, dans notre exemple) sont majoritairement arrêtés par la réaction d'autoincompatibilité propre à nos espèces d'intérêt et (ii) où les tubes polliniques compatibles (en l'occurrence V22) poursuivent leur trajet et subissent la division de leur cellule générative[91].

Si les tubes polliniques des deux espèces les plus proches (*S. gandarillasii* en jaune et *S. tarijense* en bleu) est ralenti dans le pistil, le patron biphasique de croissance semble conservé. En revanche, les tubes polliniques de *S. microdontum*, espèce plus éloignée, présentent une croissance lente et monophasique. L'expérience a été répétée avec les mêmes pollens dans les styles d'autres espèces.

La figure 1.6 montre que, quelle que soit l'espèce du pistil, les tubes polliniques hétérospécifiques sont ralentis par rapport aux tubes conspécifiques compatibles. Chez *S. chacoense* et *S. gandarillasii*, le retard semble être davantage marqué pour l'espèce la plus éloignée, *S. microdontum*. Toutefois, dans tous les cas, les tubes polliniques hétérospécifiques parviennent à accomplir leur croissance et à entrer dans l'ovaire.

Ceci laisse présager l'existence d'interactions pollen–pistil spécifiques à l'espèce, dont le bon accomplissement conditionne la vitesse de croissance des tubes polliniques.

c. Spécificité à l'espèce de la chimioattraction par l'ovule

Comme nous le verrons en détail à la section 2.3.1, il a été démontré chez des espèces telles que le maïs[92], *Torenia fournieri*[93] ou encore *Arabidopsis thaliana*[94], la chimioattraction du tube pollinique par l'ovule repose sur de petites protéines sécrétées spécifiques à l'espèce.

Or, la spécificité à l'espèce du guidage directionnel du tube pollinique a justement été mise en évidence par le biais de tests de semi-*in vivo* chez *S. chacoense*. Ces derniers consistent à disposer, sur une matrice gélifiée, des tubes polliniques émergeant de la base d'un pistil et un amas

d'ovules sécrétant dans la gélose un gradient de protéines de guidage, puis a étudier la réorientation des tubes polliniques en direction des ovules.

Les premiers résultats obtenus en 2009 par Édith Lafleur[95] (figure 1.7) montraient déjà que les tubes polliniques conspécifiques étaient nettement plus réceptifs aux signaux de guidage que les tubes hétérospécifiques. Les expériences menées ensuite par Yang Liu, qui ont eu recours à une méthode quantitative plus robuste, ont confirmé ces observations[96]. Comme nous le voyons sur la figure 1.8, qui représente l'attraction de tubes polliniques d'espèces variées par des ovules de *S. chacoense*, seuls les tubes polliniques conspécifiques ($p = 0{,}008$) et ceux de l'espèce la plus proche (*S. gandarillasii*, $p = 0{,}003$) sont significativement attirés. Le pollen des autres espèces, plus éloignées, se répartit de façon aléatoire, donc indépendante de la présence des ovules. Ceci est cohérent avec l'expérience précédente et laisse supposer que la chimioattraction repose sur des protéines qui entrent en jeu dans la spécificité à l'espèce.

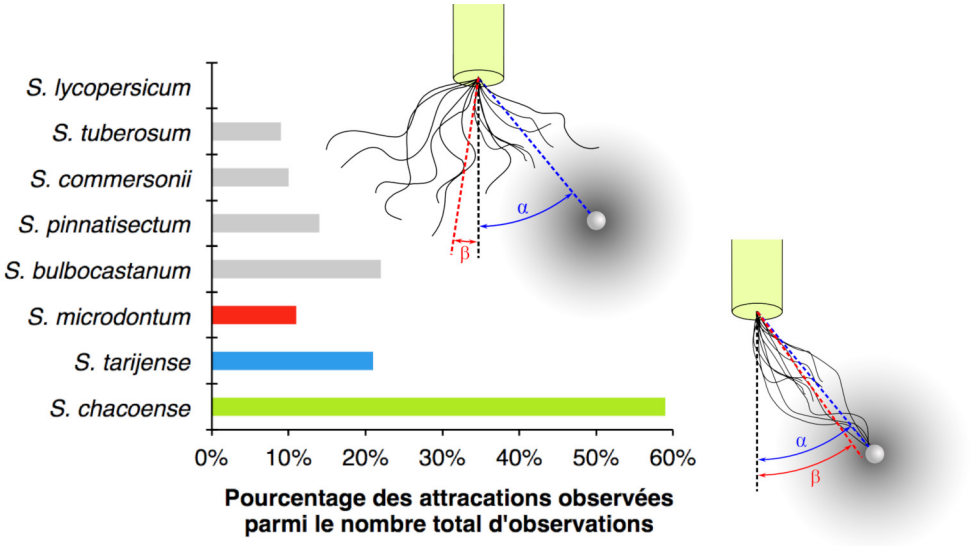


Fig. 1.7. Résultats des tests de guidage d'É. Lafleur. Des ovules de *S. chacoense* (G4) ont été présentés à du pollen de différentes espèces ayant grandi dans un style conspécifique. L'expérience a été reproduite plusieurs fois dans chaque cas. Le graphique représente la proportion d'essais dans lesquels l'attraction des tubes polliniques était clairement observée.

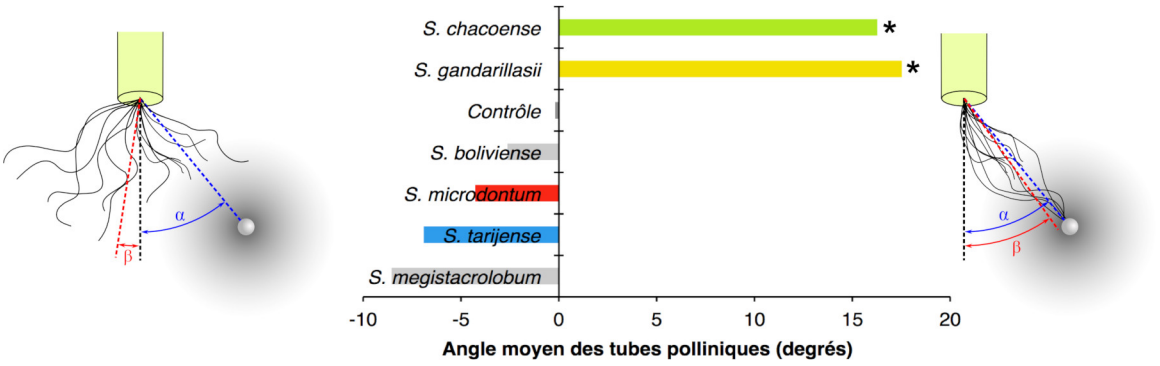


Fig. 1.8. Résultats des tests de guidage de Y. Liu. Des tubes polliniques de *S. chacoense* (V22) en croissance dans un style conspécifique (G4) à des ovules de différentes espèces. Pour au moins 100 tubes polliniques, et dans plusieurs réplicats, on mesure l'angle β de déviation du tube par rapport à sa trajectoire initiale dans le style.Un test de Kolmogrov-Smirnov est utilisé pour comparer leur distribution avec un contrôle négatif (pas d'ovule).

1.4 Objectifs et plan de la thèse

Nous venons de voir que le pistil est un lieu important de l'isolement reproductif chez nos espèces d'intérêt. Le but de ce projet de doctorat était justement d'identifier des acteurs moléculaires impliqués et d'en caractériser les fonctions. Or, comme nous le verrons au chapitre 2, les candidats sont très nombreux : un grand nombre de protéines exprimées dans le pistil et le tube pollinique ont en effet été caractérisées chez d'autres espèces. Plusieurs contribuent, par leur divergence interspécifique, à l'établissement de points de contrôle précis le long du trajet du tube pollinique dans le pistil.

Les résultats préliminaires présentés aux figures 1.7 et 1.8 et les travaux menés précédemment au laboratoire par Édith Lafleur[95], Yang Liu[96] et Claire Viallet[97], de même que l'existence d'une abondante littérature chez d'autres espèces (*cf.* section 2.3.1) nous ont conduits à nous concentrer sur le guidage directionnel du tube pollinique par l'ovule pour la suite de ce travail. Aucune protéine ovulaire chimioattractive n'ayant été identifiée chez les Solanacées, le premier défi à relever était d'isoler une liste de gènes candidats susceptibles de coder pour ces attractants afin de les tester fonctionnellement. C'est dans ce souci que nous avons entrepris d'explorer à grande échelle, dans une approche mêlant bioinformatique, protéomique et transcriptomique, la diversité des gènes exprimés dans les ovules de *S. chacoense*.

Au chapitre 3, nous présenterons le logiciel KAPPA, un logiciel de recherche de séquences spécifiquement adapté à la détection et au regroupement de peptides riches en cystéines (CRP),

une catégorie de protéines particulière dont font partie les chimioattractants ovulaires connus chez d'autres espèces. Nous décrirons ensuite les trois approches utilisées pour explorer l'expression des gènes dans les ovules de *S. chacoense* : sécrétomique comparée de l'ovule mature et immature (chapitre 4), transcriptomique par biopuce de l'ovule avant et après pollinisation (chapitre 5) et transcriptomique par RNA-seq de l'ovule sauvage, de l'ovule immature et de l'ovule mutant dépourvu de sac embryonnaire (chapitre 6). Nous discuterons en conclusion de l'identification de chimioattractants candidats chez *S. chacoense* et proposerons une boîte à outils méthodologique pour en vérifier la fonction (chapitre 7).

Bibliographie

1. Hunziker, A. T. (2001). Genera Solanacearum *: the genera of* Solanaceae *illustrated, arranged according to a new system.* Gantner, Ruggell, Liechtenstein. ISBN : 978-3-904144-77-3. [cit. p. 21]

2. Spooner, D. M. ; McLean, K. ; Ramsay, G. ; Waugh, R. et Bryan, G. J. (2005). A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc. Natl. Acad. Sci. U. S. A.*, 102(41), 14694–9. DOI : 10.1073/pnas.0507400102. [cit. p. 21]

3. Spooner, D. M. ; Ghislain, M. ; Simon, R. ; Jansky, S. H. et Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Bot. Rev.*, 80(4), 283–383. DOI : 10.1007/s12229-014-9146-y. [cit. p. 21]

4. Birch, P. R. J. ; Bryan, G. ; Fenton, B. et coll. (2012). Crops that feed the world 8 : Potato : are the trends of increased global production sustainable ? *Food Sec.*, 4(4), 477–508. DOI : 10.1007/s12571-012-0220-1. [cit. p. 21]

5. Food and Agriculture Organization of the United Nations (2016). FAO statistics division data for year 2016. URL : http ://www.fao.org/faostat/en. [cit. p. 21]

6. Halterman, D. ; Guenthner, J. ; Collinge, S. ; Butler, N. et Douches, D. (2015). Biotech potatoes in the 21st century : 20 years since the first biotech potato. *Am. J. Potato Res.*, 93(1), 1–20. DOI : 10.1007/s12230-015-9485-1. [cit. p. 21]

7. Hameed, A. ; Zaidi, S. S.-E.-A. ; Shakir, S. et Mansoor, S. (2018). Applications of new breeding technologies for potato improvement. *Front. Plant Sci.*, 9, 925. DOI : 10.3389/fpls.2018.00925. [cit. p. 21]

8. Hijmans, R. J. et Spooner, D. M. (2001). Geographic distribution of wild potato species. *Am. J. Bot.*, 88(11), 2101–12. DOI : 10.2307/3558435. [cit. p. 21]

9. Hijmans, R. J. ; Spooner, D. M. ; Salas, A. R. ; Guarino, L. et de la Cruz, J. (2002). *Atlas of wild potatoes.*, volume 10 de *Systematic and ecogeographic studies on crop genepools*. International Plant Genetic Resources Institute, Rome, Italie. ISBN : 978-92-9043-518-1. URL : https ://www.bioversityinternational.org/e-library/publications/detail/atlas-of-wild-potatoes/. [cit. p. 21, 26]

10. Goldberg, E. E. ; Kohn, J. R. ; Lande, R. et coll. (2010). Species selection maintains self-incompatibility. *Science*, 330(6003), 493–5. DOI : 10.1126/science.1194513. [cit. p. 22]

11. Rodríguez, F. et Spooner, D. M. (2009). Nitrate reductase phylogeny of potato (*Solanum* sect. *Petota*) genomes with emphasis on the origins of the polyploid species. *Syst. Bot.*, 34(1), 207–19. DOI : 10.1600/036364409787602195. [cit. p. 22]

12. Hardigan, M. A. ; Laimbeer, F. P. E. ; Newton, L. et coll. (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U. S. A.*, 114(46), E9999–10008. DOI : 10.1073/pnas.1714380114. [cit. p. 23]

13. Li, Y. ; Colleoni, C. ; Zhang, J. et coll. (2018). Genomic analyses yield markers for identifying agronomically important genes in potato. *Mol. Plant*, 11(3), 473–84. DOI : 10.1016/j.molp.2018.01.009. [cit. p. 23]

14. Leisner, C. P. ; Hamilton, J. P. ; Crisovan, E. et coll. (2018). Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. *Plant J.*, 94(3), 562–70. DOI : 10.1111/tpj.13857. [cit. p. 23]

15. Bravo-Almonacid, F. ; Rudoy, V. ; Welin, B. et coll. (2012). Field testing, gene flow assessment and pre-commercial studies on transgenic *Solanum tuberosum* spp. *tuberosum* (cv. *Spunta*) selected for PVY resistance in Argentina. *Transgenic Res.*, 21(5), 967–82. DOI : 10.1007/s11248-011-9584-9. [cit. p. 23]

16. Duan, H. ; Richael, C. et Rommens, C. M. (2012). Overexpression of the wild potato *eIF4E-1* variant *Eva1* elicits *Potato virus Y* resistance in plants silenced for native *eIF4E-1*. *Transgenic Res.*, 21(5), 929–38. DOI : 10.1007/s11248-011-9576-9. [cit. p. 23]

17. Bae, J. ; Halterman, D. et Jansky, S. (2008). Development of a molecular marker associated with *Verticillium* wilt resistance in diploid interspecific potato hybrids. *Mol. Breed.*, 22(1), 61–9. DOI : 10.1007/s11032-008-9156-8. [cit. p. 23]

18. Chen, L. ; Guo, X. ; Xie, C. et coll. (2013). Nuclear and cytoplasmic genome components of *Solanum tuberosum* + *S. chacoense* somatic hybrids and three SSR alleles related to bacterial wilt resistance. *Theor. Appl. Genet.*, 126(7), 1861–72. DOI : 10.1007/s00122-013-2098-5. [cit. p. 23]

19. Braun, S. R. ; Endelman, J. B. ; Haynes, K. G. et Jansky, S. H. (2017). Quantitative trait loci for resistance to common scab and cold-induced sweetening in diploid potato. *Plant Genome*, 10(3). DOI : 10.3835/plantgenome2016.10.0110. [cit. p. 23]

20. Jansky, S. ; Douches, D. et Haynes, K. (2017). Transmission of scab resistance to tetraploid potato via unilateral sexual polyploidization. *Am. J. Potato Res.*, 95(3), 272–7. DOI : 10.1007/s12230-017-9628-7. [cit. p. 23]

21. Faccio, P. ; Vazquez-Rovere, C. ; Hopp, E. et coll. (2011). Increased tolerance to wheat powdery mildew by heterologous constitutive expression of the *Solanum chacoense snakin-1* gene. *Czech J. Genet. Plant Breed.*, 47(Special issue), S135–41. DOI : 10.17221/3268-CJGPB. [cit. p. 23]

22. Almasia, N. I. ; Bazzini, A. A. ; Hopp, H. E. et Vazquez-Rovere, C. (2008). Overexpression of snakin-1 gene enhances resistance to *Rhizoctonia solani* and *Erwinia carotovora* in transgenic potato plants. *Mol. Plant Pathol.*, 9(3), 329–38. DOI : 10.1111/j.1364-3703.2008.00469.x. [cit. p. 23]

23. Mweetwa, A. M. ; Hunter, D. ; Poe, R. et coll. (2012). Steroidal glycoalkaloids in *Solanum chacoense*. *Phytochemistry*, 75, 32–40. DOI : 10.1016/j.phytochem.2011.12.003. [cit. p. 23]

24. Lorenzen, J. H. ; Balbyshev, N. F. ; Lafta, A. M. et coll. (2001). Resistant potato selections contain leptine and inhibit development of the Colorado potato beetle (*Coleoptera* : *Chrysomelidae*). *J. Econ. Entomol.*, 94(5), 1260–7. [cit. p. 23]

25. Manrique-Carpintero, N. C. ; Tokuhisa, J. G. ; Ginzberg, I. ; Holliday, J. A. et Veilleux, R. E. (2013). Sequence diversity in coding regions of candidate genes in the glycoalkaloid biosynthetic pathway of wild potato species. *G3 : Genes, Genomes, Genet.*, 3(9), 1467–79. DOI : 10.1534/g3.113.007146. [cit. p. 23]

26. Manrique-Carpintero, N. C. ; Tokuhisa, J. G. ; Ginzberg, I. et Veilleux, R. E. (2014). Allelic variation in genes contributing to glycoalkaloid biosynthesis in a diploid interspecific population of potato. *Theor. Appl. Genet.*, 127(2), 391–405. DOI : 10.1007/s00122-013-2226-2. [cit. p. 23]

27. Cooper, S. G. ; Douches, D. S. et Grafius, E. J. (2009). Combining engineered resistance, avidin, and natural resistance derived from *Solanum chacoense* Bitter to control Colorado potato beetle (*Coleoptera* : *Chrysomelidae*). *J. Econ. Entomol.*, 102(3), 1270–80. DOI : 10.1603/029.102.0354. [cit. p. 23]

28. Ginzberg, I. ; Thippeswamy, M. ; Fogelman, E. et coll. (2012). Induction of potato steroidal glycoalkaloid biosynthetic pathway by overexpression of cDNA encoding primary metabolism HMG-CoA reductase and squalene synthase. *Planta*, 235(6), 1341–53. DOI : 10.1007/s00425-011-1578-6. [cit. p. 23]

29. Molnár, I. ; Besenyei, E. ; Thieme, R. et coll. (2017). Mismatch repair deficiency increases the transfer of antibiosis and antixenosis properties against Colorado potato beetle in somatic hybrids of *Solanum tuberosum* + *S. chacoense*. *Pest Manag. Sci.*, 73(7), 1428–37. DOI : 10.1002/ps.4473. [cit. p. 23]

30. Crossley, M. S. ; Schoville, S. D. ; Haagenson, D. M. et Jansky, S. H. (2018). Plant resistance to Colorado potato beetle (*Coleoptera* : *Chrysomelidae*) in diploid F2 families derived from crosses between cultivated and wild potato. *J. Econ. Entomol.*, 111(4), 1875–84. DOI : 10.1093/jee/toy120. [cit. p. 23]

31. Prohens, J. ; Gramazio, P. ; Plazas, M. et coll. (2017). Introgressiomics : a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica*, 213(7), 158. DOI : 10.1007/s10681-017-1938-9. [cit. p. 23]

32. Boivin, N. ; Morse, D. et Cappadocia, M. (2014). Degradation of S-RNase in compatible pollen tubes of *Solanum chacoense* inferred by immunogold labeling. *J. Cell Sci.*, 127(Pt 19), 4123–7. DOI : 10.1242/jcs.154823. [cit. p. 23]

33. Soulard, J. ; Boivin, N. ; Morse, D. et Cappadocia, M. (2014). eEF1A is an S-RNase binding factor in self-incompatible *Solanum chacoense*. *PLoS One*, 9(2), e90206. DOI : 10.1371/journal.pone.0090206. [cit. p. 23]

34. Soulard, J. ; Qin, X. ; Boivin, N. ; Morse, D. et Cappadocia, M. (2013). A new dual-specific incompatibility allele revealed by absence of glycosylation in the conserved C2 site of a *Solanum chacoense* S-RNase. *J. Exp. Bot.*, 64(7), 1995–2003. DOI : 10.1093/jxb/ert059. [cit. p. 23]

35. Liu, B. ; Boivin, N. ; Morse, D. et Cappadocia, M. (2012). A time course of GFP expression and mRNA stability in pollen tubes following compatible and incompatible pollinations in *Solanum chacoense*. *Sex. Plant Reprod.*, 25(3), 205–13. DOI : 10.1007/s00497-012-0192-5. [cit. p. 23]

36. Aouar, L. ; Chebli, Y. et Geitmann, A. (2010). Morphogenesis of complex plant cell shapes : the mechanical role of crystalline cellulose in growing pollen tubes. *Sex. Plant Reprod.*, 23(1), 15–27. DOI : 10.1007/s00497-009-0110-7. [cit. p. 23]

37. Parre, E. et Geitmann, A. (2005). Pectin and the role of the physical properties of the cell wall in pollen tube growth of *Solanum chacoense*. *Planta*, 220(4), 582–92. DOI : 10.1007/s00425-004-1368-5. [cit. p. 23]

38. Parre, E. et Geitmann, A. (2005). More than a leak sealant. The mechanical properties of callose in pollen tubes. *Plant Physiol.*, 137(1), 274–86. DOI : 10.1104/pp.104.050773. [cit. p. 23]

39. Chevalier, E. ; Loubert-Hudon, A. et Matton, D. P. (2013). ScRALF3, a secreted RALF-like peptide involved in cell-cell communication between the sporophyte and the female gametophyte in a solanaceous species. *Plant J.*, 73(6), 1019–33. DOI : 10.1111/tpj.12096. [cit. p. 23]

40. Germain, H. ; Chevalier, E. ; Caron, S. et Matton, D. P. (2005). Characterization of five RALF-like genes from *Solanum chacoense* provides support for a developmental role in plants. *Planta*, 220(3), 447–54. DOI : 10.1007/s00425-004-1352-0. [cit. p. 23]

41. Daigle, C. et Matton, D. P. (2015). Genome-wide analysis of MAPKKKs shows expansion and evolution of a new MEKK class involved in solanaceous species sexual reproduction. *BMC Genomics*, 16, 1037. DOI : 10.1186/s12864-015-2228-3. [cit. p. 23]

42. Lafleur, E. ; Kapfer, C. ; Joly, V. et coll. (2015). The FRK1 mitogen-activated protein kinase kinase kinase (MAPKKK) from *Solanum chacoense* is involved in embryo sac and pollen development. *J. Exp. Bot.*, 66(7), 1833–43. DOI : 10.1093/jxb/eru524. [cit. p. 23]

43. O'Brien, M. ; Gray-Mitsumune, M. ; Kapfer, C. ; Bertrand, C. et Matton, D. P. (2007). The ScFRK2 MAP kinase kinase kinase from *Solanum chacoense* affects pollen development and viability. *Planta*, 225(5), 1221–31. DOI : 10.1007/s00425-006-0432-8. [cit. p. 23]

44. Gray-Mitsumune, M. ; O'Brien, M. ; Bertrand, C. et coll. (2006). Loss of ovule identity induced by overexpression of the fertilization-related kinase 2 (ScFRK2), a MAPKKK from *Solanum chacoense*. *J. Exp. Bot.*, 57(15), 4171–87. DOI : 10.1093/jxb/erl194. [cit. p. 23]

45. Germain, H. ; Gray-Mitsumune, M. ; Houde, J. et coll. (2013). The *Solanum chacoense* ovary receptor kinase 11 (ScORK11) undergoes tissue-dependent transcriptional, translational and post-translational regulation. *Plant Physiol. Biochem.*, 70, 261–8. DOI : 10.1016/j.plaphy.2013.05.036. [cit. p. 23]

46. Germain, H. ; Gray-Mitsumune, M. ; Lafleur, E. et Matton, D. P. (2008). ScORK17, a transmembrane receptor-like kinase predominantly expressed in ovules is involved in seed development. *Planta*, 228(5), 851–62. DOI : 10.1007/s00425-008-0787-0. [cit. p. 23]

47. Germain, H. ; Houde, J. ; Gray-Mitsumune, M. et coll. (2007). Characterization of ScORK28, a transmembrane functional protein receptor kinase predominantly expressed in ovaries from the wild potato species *Solanum chacoense*. *FEBS Lett.*, 581(26), 5137–42. DOI : 10.1016/j.febslet.2007.10.001. [cit. p. 23]

48. Germain, H. ; Rudd, S. ; Zotti, C. et coll. (2005). A 6374 unigene set corresponding to low abundance transcripts expressed following fertilization in *Solanum chacoense* Bitt, and characterization of 30 receptor-like kinases. *Plant Mol. Biol.*, 59(3), 515–32. DOI : 10.1007/s11103-005-0536-8. [cit. p. 23]

49. Johnston, S. A. ; den Nijs, T. P. ; Peloquin, S. J. et Hanneman, R. E. (1980). The significance of genic balance to endosperm development in interspecific crosses. *Theor. Appl. Genet.*, 57(1), 5–9. DOI : 10.1007/BF00276002. [cit. p. 24]

50. Carputo, D. ; Monti, L. ; Werner, J. E. et Frusciante, L. (1999). Uses and usefulness of endosperm balance number. *Theor. Appl. Genet.*, 98(3–4), 478–84. DOI : 10.1007/s001220051095. [cit. p. 24]

51. Rieseberg, L. H. et Willis, J. H. (2007). Plant speciation. *Science*, 317(5840), 910–4. DOI : 10.1126/science.1137729. [cit. p. 25]

52. Sobel, J. M. ; Chen, G. F. ; Watt, L. R. et Schemske, D. W. (2010). The biology of speciation. *Evolution*, 64(2), 295–315. DOI : 10.1111/j.1558-5646.2009.00877.x. [cit. p. 25]

53. Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.*, 173 (5), 579–88. DOI : 10.1086/597611. [cit. p. 25]

54. Lexer, C. ; Lai, Z. et Rieseberg, L. H. (2004). Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids : implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *New Phytol.*, 161(1), 225–33. DOI : 10.1046/j.1469-8137.2003.00925.x. [cit. p. 25]

55. Rus, A. ; Baxter, I. ; Muthukumar, B. et coll. (2006). Natural variants of At*HKT1* enhance Na^+ accumulation in two wild populations of *Arabidopsis*. *PLoS Genet.*, 2(12), e210. DOI : 10.1371/journal.pgen.0020210. [cit. p. 25]

56. Balasubramanian, S. ; Sureshkumar, S. ; Agrawal, M. et coll. (2006). The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nat. Genet.*, 38(6), 711–5. DOI : 10.1038/ng1818. [cit. p. 25]

57. Caicedo, A. L. ; Stinchcombe, J. R. ; Olsen, K. M. ; Schmitt, J. et Purugganan, M. D. (2004). Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proc. Natl. Acad. Sci. U. S. A.*, 101(44), 15670–5. DOI : 10.1073/pnas.0406232101. [cit. p. 25]

58. Schwartz, C. ; Balasubramanian, S. ; Warthmann, N. et coll. (2009). Cis-regulatory changes at *FLOWERING LOCUS T* mediate natural variation in flowering responses of *Arabidopsis thaliana*. *Genetics*, 183(2), 723–32. DOI : 10.1534/genetics.109.104984. [cit. p. 25]

59. Takahashi, Y. ; Teshima, K. M. ; Yokoi, S. ; Innan, H. et Shimamoto, K. (2009). Variations in Hd1 proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice. *Proc. Natl. Acad. Sci. U. S. A.*, 106(11), 4555–60. DOI : 10.1073/pnas.0812092106. [cit. p. 25]

60. Chen, K.-Y. ; Cong, B. ; Wing, R. ; Vrebalov, J. et Tanksley, S. D. (2007). Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. *Science*, 318(5850), 643–5. DOI : 10.1126/science.1148428. [cit. p. 25]

61. Yan, L. ; Loukoianov, A. ; Blechl, A. et coll. (2004). The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science*, 303(5664), 1640–4. DOI : 10.1126/science.1094305. [cit. p. 25]

62. Turner, A. (2005). The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. *Science*, 310(5750), 1031–4. DOI : 10.1126/science.1117619. [cit. p. 25]

63. Blackman, B. K. ; Strasburg, J. L. ; Raduski, A. R. ; Michaels, S. D. et Rieseberg, L. H. (2010). The role of recently derived FT paralogs in sunflower domestication. *Curr. Biol.*, 20(7), 629–35. DOI : 10.1016/j.cub.2010.01.059. [cit. p. 25]

64. Grant, V. G. (1971). *Plant speciation*. Columbia University Press, New York (N. Y.), États-Unis. ISBN : 978-0-231-03208-7. [cit. p. 25]

65. Stebbins, G. L. (1970). Adaptive radiation of reproductive characteristics in Angiosperms, I : Pollination mechanisms. *Ann. Rev. Ecol. Syst.*, 1(1), 307–26. DOI : 10.1146/annurev.es.01.110170.001515. [cit. p. 25]

66. Hoballah, M. E. ; Gübitz, T. ; Stuurman, J. et coll. (2007). Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell*, 19(3), 779–90. DOI : 10.1105/tpc.106.048694. [cit. p. 26]

67. Quattrocchio, F. ; Wing, J. ; van der Woude, K. et coll. (1999). Molecular analysis of the anthocyanin2 gene of *Petunia* and its role in the evolution of flower color. *Plant Cell*, 11(8), 1433–44. DOI : 10.1105/tpc.11.8.1433. [cit. p. 26]

68. Schwinn, K. ; Venail, J. ; Shang, Y. et coll. (2006). A small family of *MYB*-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell*, 18(4), 831–51. DOI : 10.1105/tpc.105.039255. [cit. p. 26]

69. Whibley, A. C. ; Langlade, N. B. ; Andalo, C. et coll. (2006). Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science*, 313(5789), 963–6. DOI : 10.1126/science.1129161. [cit. p. 26]

70. Des Marais, D. L. et Rausher, M. D. (2010). Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. *Evolution*, 64(7), 2044–54. DOI : 10.1111/j.1558-5646.2010.00972.x. [cit. p. 26]

71. Hoshino, A. ; Morita, Y. ; Choi, J.-D. et coll. (2003). Spontaneous mutations of the flavonoid 3'-hydroxylase gene conferring reddish flowers in the three morning glory species. *Plant Cell Physiol.*, 44(10), 990–1001. DOI : 10.1093/pcp/pcg143. [cit. p. 26]

72. Streisfeld, M. A. et Rausher, M. D. (2009). Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. *New Phytol.*, 183(3), 751–63. DOI : 10.1111/j.1469-8137.2009.02929.x. [cit. p. 26]

73. Zufall, R. A. (2003). The genetic basis of a flower color polymorphism in the common morning glory (*Ipomoea purpurea*). *J. Hered.*, 94(6), 442–8. DOI : 10.1093/jhered/esg098. [cit. p. 26]

74. Bradshaw, H. D. et Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, 426(6963), 176–8. DOI : 10.1038/nature02106. [cit. p. 26]

75. Streisfeld, M. A. et Rausher, M. D. (2009). Altered trans-regulatory control of gene expression in multiple anthocyanin genes contributes to adaptive flower color evolution in *Mimulus aurantiacus*. *Mol. Biol. Evol.*, 26(2), 433–44. DOI : 10.1093/molbev/msn268. [cit. p. 26]

76. Spooner, D. M. et van den Berg, R. G. (1992). Species limits and hypotheses of hybridization of *Solanum berthaultii* Hawkes and *S. tarijense* Hawkes : Morphological data. *Taxon*, 41(4), 685. DOI : 10.2307/1222394. [cit. p. 27]

77. Spooner, D. M. ; Fajardo, D. et Bryan, G. J. (2007). Species limits of *Solanum berthaultii* Hawkes and S. tarijense Hawkes and the implications for species boundaries in *Solanum* sect. *Petota*. *Taxon*, 56(4), 987. DOI : 10.2307/25065899. [cit. p. 27]

78. Dobzhansky, T. (1934). Studies on hybrid sterility. *Z. Zellforsch. Mikrosk. Anat.*, 21(2), 169–223. DOI : 10.1007/BF00374056. [cit. p. 27]

79. Krüger, J. ; Thomas, C. M. ; Golstein, C. et coll. (2002). A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of autonecrosis. *Science*, 296(5568), 744–7. DOI : 10.1126/science.1069288. [cit. p. 28]

80. Rooney, H. C. E. ; Van't Klooster, J. W. ; van der Hoorn, R. A. L. et coll. (2005). *Cladosporium* Avr2 inhibits tomato Rcr3 protease required for *Cf-2*–dependent disease resistance. *Science*, 308 (5729), 1783–6. DOI : 10.1126/science.1111404. [cit. p. 28]

81. Yamamoto, E. ; Takashi, T. ; Morinaka, Y. et coll. (2010). Gain of deleterious function causes an autoimmune response and Bateson-Dobzhansky-Muller incompatibility in rice. *Mol. Genet. Genomics*, 283(4), 305–15. DOI : 10.1007/s00438-010-0514-y. [cit. p. 28]

82. Jeuken, M. J. W. ; Zhang, N. W. ; McHale, L. K. et coll. (2009). *Rin4* causes hybrid necrosis and race-specific resistance in an interspecific lettuce hybrid. *Plant Cell*, 21(10), 3368–78. DOI : 10.1105/tpc.109.070334. [cit. p. 28]

83. Bikard, D. ; Patel, D. ; Le Mette, C. et coll. (2009). Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science*, 323(5914), 623–6. DOI : 10.1126/science.1165917. [cit. p. 28]

84. Levin, D. A. (2003). The cytoplasmic factor in plant speciation. *Syst. Bot.*, 28(1), 5–11. DOI : 10.1043/0363-6445-28.1.5. [cit. p. 28]

85. Chase, C. D. (2007). Cytoplasmic male sterility : a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.*, 23(2), 81–90. DOI : 10.1016/j.tig.2006.12.004. [cit. p. 28]

86. Hanson, M. R. et Bentolila, S. (2004). Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell*, 16(Suppl. 1), S154–69. DOI : 10.1105/tpc.015966. [cit. p. 28]

87. Lurin, C. ; Andrés, C. ; Aubourg, S. et coll. (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, 16(8), 2089–103. DOI : 10.1105/tpc.104.022236. [cit. p. 28]

88. O'Toole, N. ; Hattori, M. ; Andres, C. et coll. (2008). On the expansion of the pentatricopeptide repeat gene family in plants. *Mol. Biol. Evol.*, 25(6), 1120–8. DOI : 10.1093/molbev/msn057. [cit. p. 28]

89. Rieseberg, L. H. et Blackman, B. K. (2010). Speciation genes in plants. *Ann. Bot.*, 106(3), 439–55. DOI : 10.1093/aob/mcq126. [cit. p. 28]

90. Stephenson, A. G. ; Travers, S. E. ; Mena-Ali, J. I. et Winsor, J. A. (2003). Pollen performance before and during the autotrophic-heterotrophic transition of pollen tube growth. *Philos. Trans. R. Soc., B*, 358(1434), 1009–18. DOI : 10.1098/rstb.2003.1290. [cit. p. 30]

91. Hormaza, J. I. et Herrero, M. (1994). Gametophytic competition and selection. *In* Williams, E. G. ; Clarke, A. E. et Knox, R. B., éditeurs : *Genetic control of self-incompatibility and reproductive development in flowering plants*, volume 2 de *Advances in Cellular and Molecular Biology of Plants*, pages 372–400. Springer, Dordrecht, Pays-Bas. ISBN : 978-90-481-4340-5. DOI : 10.1007/978-94-017-1669-7_18. [cit. p. 32]

92. Márton, M. L. ; Cordts, S. ; Broadhvest, J. et Dresselhaus, T. (2005). Micropylar pollen tube guidance by Egg Apparatus 1 of maize. *Science*, 307(5709), 573–6. DOI : 10.1126/science.1104954. [cit. p. 32]

93. Okuda, S. ; Tsutsui, H. ; Shiina, K. et coll. (2009). Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature*, 458(7236), 357–61. DOI : 10.1038/nature07882. [cit. p. 32]

94. Takeuchi, H. et Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol.*, 10(12), e1001449. DOI : 10.1371/journal.pbio.1001449. [cit. p. 32]

95. Lafleur, É. (2009). *Rôle de la protéine ScFRK1 dans le développement du sac embryonnaire et son impact sur le guidage des tubes polliniques*. Mémoire de maîtrise, Université de Montréal. URL : http ://hdl.handle.net/1866/3617. [cit. p. 33, 34]

96. Liu, Y. (2015). *The plant ovule omics : an integrative approach for pollen-pistil interactions and pollen tube guidance studies in solanaceous species*. Thèse de doctorat, Université de Montréal. URL : http ://hdl.handle.net/1866/13589. [cit. p. 33, 34]

97. Viallet, C. (2016). *Identification de protéines impliquées dans le guidage du tube pollinique par les ovules de* Solanum chacoense. Mémoire de maîtrise, Université de Montréal. [cit. p. 34]

CHAPITRE 2

Interactions pollen–pistil et isolement reproductif

Le présent chapitre vise à effectuer un tour d'horizon des interactions pollen–pistil connues à ce jour, en mettant en avant celles qui pourraient contribuer, par leur spécificité, à l'isolement reproductif. Nous suivrons le trajet du tube pollinique dans le pistil en présentant successivement les événements moléculaires connus dans le stigmate, dans le style et dans l'ovaire. Ici encore, le lecteur pourra se référer à la figure 4.1 (page 114) pour une vue schématique des interactions pollen–pistil chez *Solanum*.

2.1 Interactions pollen–stigmate

Le tout premier contact entre pollen et pistil se produit à la surface du stigmate. Une fois arrimé à ce dernier, le grain de pollen subit une hydratation guidée par les tissus femelles. Ceci conduit à sa germination sous la forme d'un tube pollinique qui va pénétrer dans le stigmate. Finement régulées au niveau moléculaire, les interactions pollen–stigmate sont un premier point de contrôle déterminant l'acceptation ou le rejet du pollen[1].

2.1.1 Adhésion du pollen

Il est courant de distinguer, chez les Angiospermes, les taxons à stigmate « humide » (ex. : *Liliaceae*, *Solanaceae*) présentant une sécrétion constante et abondante accessible de façon non spécifique à tous les grains de pollen[2], de ceux à stigmate « sec » (ex. : *Brassicaceae*) chez qui la reconnaissance du pollen précède et détermine une sécrétion localisée et spécifique[3].

Chez les espèces à stigmate sec, la liaison du pollen sur le stigmate semble déjà présenter une forme de spécificité à l'espèce, ou du moins au taxon proche : le stigmate d'*Arabidopsis thaliana* lie mieux le pollen conspécifique que celui provenant de différentes espèces du genre *Brassica*. Ces dernières, lorsque croisées entre elles, lient indifféremment le pollen con- ou hétérospécifique[4].

Cette relative spécificité impliquerait, du côté pollinique, une variabilité des motifs du manteau d'exine entre les espèces[5]. Les gènes contrôlant la synthèse ou l'arrangement de la sporopollenine pourraient être impliqués dans cette diversité[6,7,8,9,10,11,12,13,14]. Sur le versant femelle, il semble que la cuticule lipidique du stigmate soit spécifiquement impliquée[15,16,17]. On ignore encore la nature précise des acteurs moléculaires sous-jacents, bien que des approches expérimentales précises aient été suggérées pour les identifier[18].

La liaison du pollen au stigmate est également marquée par la mise en contact des sucres, protéines et lipides des deux partenaires. Ce phénomène a été abondamment décrit chez *B. oleracea* (chou commun) chez qui il implique SLG et SLR1, deux glycoprotéines proches des médiateurs de l'autoincompatibilité sporophytique, localisées dans la paroi des papilles stigmatiques[16,19,20,21]. Elles interagissent avec de petites protéines polliniques basiques de la famille PCP, respectivement PCP-A1 et PCP-A2[22]. Bien que ces interactions semblent peu spécifiques à l'espèce[23], il existe un certain polymorphisme des protéines SLG et SLR1 chez les Brassicacées[24]. On a également décrit des protéines proches des PCP chez plusieurs espèces[25]. Un tel mécanisme pourrait donc conférer une spécificité à l'espèce.

2.1.2 Hydratation du pollen

a. Hydration locale et contrôlée chez les espèces à stigmate sec

L'hydratation du grain de pollen à la surface du stigmate pourrait être également marquée, chez les espèces à stigmate sec, par une spécificité à l'espèce. On sait par exemple que le pollen des Brassicacées est capable de s'hydrater à la surface du stigmate d'*Arabidopsis*, mais que cette hydratation est impossible avec des plantes issues d'autres familles[26].

On a abondamment décrit le rôle des lipides du pollen et du stigmate dans l'accomplissement correct de l'hydratation[27,28]. Plusieurs protéines contrôlant l'organisation de ces lipides pourraient affecter la spécificité à l'espèce. Les premières études sur la question ont mis en avant

des protéines riches en glycine (GRP) capables de se lier aux lipides grâce à un domaine oléosine[29]. Par exemple, chez *Arabidopsis*, GRP17 assure une hydratation correcte du pollen en empêchant l'agrégation des lipides[30]. En outre, les protéines GRP évoluent plus rapidement que la moyenne entre les écotypes d'*A. thaliana* ainsi qu'entre les espèces proches chez les Brassicacées, ce qui en fait de bonnes candidates pour la spécificité à l'espèce[31,32]. La lipase extracellulaire EXL4 a également été caractérisée dans le pollen d'*A. thaliana*. Agissant de pair avec GRP17, cette estérase/lipase de type GDSL déstabilise les lipides polliniques pour faciliter l'hydratation. D'autres membres de la famille EXL pourraient être impliqués[33].

La protéine stigmatique Exo70A1, mise en évidence chez *B. napus* (colza) et *A. thaliana*, est impliquée dans la sécrétion polarisée de vésicules à destination du pollen compatible[34]. Exo70A1 est une composante du complexe de l'exocyste, impliqué dans l'attachement à la membrane des vésicules en provenance de l'appareil de Golgi[35]. Une étude microscopique a permis de confirmer cette hypothèse au niveau cellulaire : en cas de pollinisation compatible, des vésicules (*A. thaliana*) ou des corps multivésiculaires (*B. napus*) sont fusionnés à la membrane des papilles stigmatiques, cette fusion n'ayant plus lieu si Exo70A1 n'est pas exprimée ou si une pollinisation incompatible est effectuée[36]. Plus récemment, il a été montré qu'une autre composante du complexe de l'exocyste, SEC3A, participe à la détermination du site de germination du pollen[37].

Si l'on sait que ces vésicules contiennent de l'eau destinée à hydrater le pollen, on a remarqué qu'elles pourraient aussi véhiculer des pompes calciques. En effet, le contact entre le stigmate d'*A. thaliana* et la paroi du pollen compatible entraîne un efflux calcique des papilles stigmatiques vers le pollen, circonscrit à leur site d'interaction. Ceci est corrélé à une hausse de l'expression de la pompe calcique ACA13 et à la fusion à la membrane des vésicules la contenant. Ni le pollen conspécifique incompatible ni le pollen hétérospécifique de *B. rapa* ne peuvent induire ce flux calcique[38].

En outre, les phosphoinositides semblent être requis pour réguler l'arrimage des vésicules à la membrane. La mutation de la phosphatase RHD4 (qui dégrade le phosphatidylinositol-4-phosphate, PI4P) ou celle des kinases PI4Kβ1 et β2 (qui le génèrent) conduit à un ralentissement de l'hydratation du pollen compatible chez *A. thaliana*[39]. De plus, le nombre de graines est réduit, ce qui suggère que ce composé intervient aussi dans les interactions pollen-pistil ultérieures. On sait également qu'un équilibre fin entre les niveaux de PI4P et de $PI4,5P_2$ est requis pour accomplir une endocytose correcte à l'extrémité du tube pollinique[40].

Plusieurs études récentes ont eu recours à la technique de « LM–RNA-seq » pour caractériser le transcriptome des papilles stigmatiques de plusieurs Brassicacées[41,42,43]. La technique consiste à effectuer un séquençage d'ARN à haut débit (RNA-seq) à partir de tissus isolés spécifiquement par micro-dissection au laser (LM) sur des coupes histologiques. Un tiers des transcrits papillaires identifiés chez *A. thaliana*, *A. halleri* et *B. rapa* était spécifique aux papilles stigmatiques, le reste représentant des gènes exprimés également dans d'autres tissus. Parmi les fonctions les plus représentées, on trouve le métabolisme des lipides et la modification des membranes plasmiques et des parois cellulaires, ce qui est cohérent avec les deux points précédents. En outre, bien que de 17 000 à 21 000 transcrits aient été assemblés dans chaque espèce, seuls 12 000 sont communs aux trois. Ceci laisse présager l'existence de transcrits stigmatiques spécifiques à l'espèce[41].

Sur le versant pollinique, plusieurs protéines entrent en jeu lors de l'hydratation pour contrer le stress osmotique causé par l'entrée d'eau massive et rapide en provenance du stigmate. C'est le cas du canal mécanosensible MSL8 qui permet réguler la concentration en osmolytes[44], ainsi que des aquaporines qui participent également au contrôle de la turgescence du pollen[45].

b. Exsudat stigmatique chez les espèces à stigmate humide

Les protéines mentionnées plus haut appartiennent toutes à des espèces à stigmate sec, chez qui une reconnaissance relativement spécifique du pollen déclenche une hydratation localisée. Les espèces à stigmate humide, dont font partie les Solanacées, présentent en revanche une sécrétion riche et visqueuse, constamment présente à la surface du stigmate lorsque la fleur est à maturité. On sait de longue date que cet exsudat contient des sucres, lipides, protéines, acides aminés et phénols[46,47,48]. La présence de formes réactives de l'oxygène (ROS) et de monoxyde d'azote (NO) a été attestée plus récemment[49], de même que celle d'ions calcium[50].

Cette sécrétion abondante est accessible à n'importe quel type de pollen[51]. Faut-il pour autant en conclure qu'il n'existe aucune interaction pollen–stigmate spécifique chez ces espèces ? De récentes études protéomiques mettent en avant la complexité de l'exsudat stigmatique chez le lis (*Lilium longiflorum*) et l'olivier (*Olea europaea*), avec respectivement 51 et 57 protéines décrites[52], ce qui en ferait un « environnement extracellulaire biochimiquement actif », propice à un contrôle fin de la germination du pollen[53]. Seul un quart des protéines est partagé entre les deux espèces testées, ce qui laisse présager que l'exsudat stigmatique, loin d'être conservé d'une espèce à l'autre, pourrait présenter une forme de spécificité au pollen.

Parmi les protéines présentes dans le sécrétome, on compte notamment :

- des enzymes de dégradation de la paroi, par exemple des pectinases, qui pourraient faciliter la pénétration du tube pollinique ;
- de nombreux types d'hydrolases (glycosylases, estérases-lipases, protéases, etc.) suspectées par les auteurs de faciliter l'assimilation de certains gros composés par le pollen ;
- des protéines reproductives connues, telles que la chimiocyanine[54] et la protéine SCA[55] ;
- des protéines de résistance aux chocs thermiques et de défense, qui permettraient de protéger le fragile environnement stigmatique face aux stress biotiques et abiotiques tels que la température[56] ;
- parmi les protéines de défense, des chitinases, déjà connues pour faciliter la pénétration des hyphes fongiques dans les racines et qui pourraient jouer un rôle similaire avec le pollen[57].

Il est intéressant de constater que parmi ces protéines, jusqu'à 65 % ne possèdent pas de peptide signal prédit. Il pourrait s'agir de protéines n'empruntant pas la voie classique de sécrétion médiée par le réticulum endoplasmique et l'appareil de Golgi, mais les voies non conventionnelles récemment décrites chez les plantes[58]. Une autre hypothèse est la présence, dans l'exsudat stigmatique, d'exosomes extracellulaires contenant des protéines normalement cytoplasmiques, comme retrouvé dans les fluides apoplastiques du tournesol[59].

L'activité sécrétoire du pollen, moins connue, semble toutefois impliquée dans les interactions pollen–pistil : une étude a mis au jour l'existence de nanovésicules sécrétées par le pollen, appelées *pollensomes*, d'une taille de moins de 60 nm, qui pourraient être impliquées dans des interactions avec le pistil[60].

2.1.3 Germination du pollen

Après l'hydratation, le pollen germe et forme un tube pollinique (TP) qui pénètre dans le stigmate puis dans le style et croît jusqu'à atteindre un ovule. Chez *Arabidopsis*, l'événement de germination survient au bout de 30 min environ[31] et est associé à des signaux particuliers dans l'interaction pollen–stigmate.

Le calcium cytosolique semble particulièrement impliqué, puisque trois vagues calciques ont été mises en évidence par imagerie FRET dans les papilles stigmatiques lors de l'interaction avec le pollen[61] : une première survient juste après l'hydratation, au site précis de l'interaction,

une deuxième vague se produit lorsque le TP germe, puis une troisième a lieu lorsqu'il pénètre entre les papilles stigmatiques. Les réseaux de signalisation sous-jacents à ces oscillations sont encore inconnus. On peut toutefois suspecter l'intervention de protéines sensibles du calcium, telle que les kinases dépendantes au calcium (CDPK), les protéines de type calmoduline (CML) ou celles de type calcineurine B (CBL). Ces trois types de protéines sont exprimées à l'apex du TP[62].

Par ailleurs, les métabolites secondaires présents dans le pistil semblent jouer un rôle dans le contrôle de la germination. En effet, une analyse par spectrométrie de masse d'extraits de pistil d'*Arabidopsis thaliana* a permis d'identifier les azadécalines sulfinylées, petites molécules cycliques de formule générale $C_{10}H_{19}NOS$, qui stimulent la germination du pollen *in vitro*[63]. On sait aussi que le kaempférol, un flavonoïde de type flavonol, retrouvé dans le pistil de *Petunia*[64,65] est capable d'induire la germination et la croissance du TP. Il en va de même des brassinostéroïdes retrouvés chez *A. thaliana*[66].

2.2 Interactions pollen–style

Une fois qu'il a germé, le TP doit se frayer un chemin dans les espaces intercellulaires du tissu de transmission, qui occupe la zone centrale du style. Avant de nous pencher sur la façon dont il interagit avec le style, attardons-nous sur les mécanismes permettant au TP d'accomplir sa croissance polaire.

2.2.1 Mécanismes d'élongation du tube pollinique

À l'instar des hyphes fongiques ou des poils racinaires, les TP sont des cellules en croissance polaire. Leur vitesse d'élongation est l'une des plus élevées dans le règne végétal[67]. Il a été suggéré que cette grande vitesse soit le fruit de pressions de sélection poussant le TP à être le premier à féconder un ovule[68]. L'élongation rapide du TP repose sur une intense activité sécrétoire localisée à l'apex, dans une région également dénommée « cône de croissance ». Des vésicules sont massivement adressées à la membrane apicale pour apporter le nouveau matériel membranaire et pariétal nécessaire à la croissance, le surplus étant récupéré par endocytose[69].

Les nouvelles parois formées par l'élongation du TP subissent un certain nombre de modifications chimiques[70]. Par exemple, les pectines sont désestérifiées par des pectine méthylestérases (PME) sécrétées comme VANGUARD1[71], ce qui contribue à en modifier la conformation et

à rigidifier les parois latérales, tandis que des inhibiteurs de PME interviennent à l'apex du TP pour en empêcher la rigidification[72].

D'autres protéines sécrétées par le TP jouent un rôle important dans le maintien de sa structure. C'est le cas des extensines à répétitions riches en leucines (LRX) 8 à 11 d'*Arabidopsis*, dont la mutation affecte la composition de la paroi[73,74]. Il a récemment été démontré que les LRX interagissent avec les peptides RALF4 et 19[75] pour activer deux récepteurs kinases polliniques, ANXUR1 (ANX1) et ANX2, dont le rôle dans la croissance du pollen avait déjà été mis en évidence précédemment. Chez le double mutant *anx1/anx2*, le TP s'arrête de façon anticipée et éclate avant d'arriver à la hauteur du sac embryonnaire[76,77], tandis que lorsqu'on surexprime ANX1 et ANX2, l'exocytose à l'apex du tube s'emballe, ce qui conduit à l'accumulation de matériel pariétal qui bloque la croissance du pollen[78]. Les cibles de la signalisation RALF/LRX/ANX restent encore inconnues, de même que son lien éventuel avec d'autres protéines impliquées dans l'élongation du TP telles que ATUNIS1 et 2[79] et MARIS[80].

L'ion calcium (Ca^{2+}) semble être au cœur de la signalisation interne au TP. En effet, des oscillations de la concentration intracellulaire en calcium ($[Ca^{2+}]_i$) sont visibles à l'apex du TP en croissance[81] et synchroniseraient, par le biais d'une signalisation calcique complexe, les multiples activités cellulaires nécessaires à l'élongation du TP[82], notamment la dynamique du cytosquelette[83].

De nombreuses zones d'ombre demeurent toutefois au sujet du calcium. Par exemple, on ignore encore quels canaux, pompes et échangeurs régulent l'influx et l'efflux calciques dans le TP. On sait toutefois que les canaux CNGC18[84] et des GLR (*glutamate-like receptors*)[85,86] identifiés dans le TP seraient partiellement responsables des oscillations de la $[Ca^{2+}]_i$. En outre, les cibles moléculaires précises de la signalisation calcique restent également inconnues. Parmi les 230 protéines polliniques d'*Arabidopsis* connues pour avoir un lien avec le calcium, on retrouve des calmodulines ou encore des protéines kinases liant le calcium (CDPK), qui auraient la capacité de déployer le signal calcique vers une très grande quantité de cibles[87].

Outre le calcium, les GTPases ROP (*Rho Of Plants*) pourraient être un autre régulateur central de la croissance du TP. Lorsqu'elles sont liées à la GTP, les ROP peuvent activer un large éventail de cibles, dont des protéines kinases[88,89,90]. Peu de cibles précises des ROP sont attestées dans le TP, mais on peut citer les protéines RIC3 et 4, qui contrôlent la dynamique des microfilaments d'actine en lien avec le calcium[91], ou encore une NADPH oxydase membranaire de type RBOH qui génère des formes réactives de l'oxygène (ROS), lesquelles sont suspectées d'activer un canal calcique membranaire encore inconnu et de participer ainsi au

contrôle des oscillations calciques dans le TP[92].

2.2.2 Contrôle stylaire de la croissance du pollen

Le tissu de transmission est formé de petites cellules cylindriques sécrétant une matrice extracellulaire riche en sucres, glycolipides et glycoprotéines[93]. Cette matrice contient les composants nécessaires pour nourrir et guider le TP[94]. On sait par exemple que chez le mutant *ntt* (*no transmitting tract*) d'*Arabidopsis*, dont le tissu de transmission est fortement altéré, les TP sont ralentis ou même arrêtés[95]. NTT est un facteur de transcription à doigts de zinc contrôlant l'expression de HAF (HALF FILLED), un autre facteur de transcription, de type bHLH, gouvernant la production de matrice extracellulaire de tissu de transmission[96].

La capacité du TP à pénétrer dans le style repose en particulier sur la sécrétion d'enzymes polliniques dégradant la paroi telles que des β-expansines[97], des extensines[98] ou des polygalacturonases[99]. Une étude récente a en particulier mis en avant l'importance de la *O*-fucosyltransférase pollinique 1 d'*Arabidopsis* (AtOFT1) dans ce contexte[100]. Une régulation correcte de la pression de turgescence dans le TP, assurée par des protéines comme la céramidase TOD1, est également essentielle pour la pénétration dans le style[101].

Sur le versant femelle, l'élongation correcte du TP dans le tissu de transmission requiert un gradient de signaux d'attraction chimiques sécrétés (chimiotaxie) ou liés au substrat stylaire (haptotaxie)[102]. Par exemple, chez le lis, le guidage du TP dans le stigmate et le style est contrôlé par une petite protéine appelée chimiocyanine[54]. Elle agit de pair avec SCA, une protéine riche en cystéines (CRP) de la famille des *lipid transfer proteins* (LTP), qui permet l'adhésion du TP au style[55,103]. (La pectine stylaire est également requise pour cette adhésion[104].) Les homologues de la chimiocyanine et de SCA chez *Arabidopsis*, la plantacyanine[105] et la protéine LTP5[106] respectivement, participent également à l'orientation des TP. Chimio- et plantacyanine appartiennent au groupe des phytocyanines, suspectées de contrôler la concentration des formes réactives de l'oxygène (ROS).

En dehors des protéines mentionnées ci-dessus, l'acide γ-aminobutyrique (GABA), présent dans le pistil d'*Arabidopsis* participerait également à l'élongation directionnelle des TP. Le GABA est présent sous la forme d'un gradient stylaire, le composé étant de plus en plus concentré à mesure que l'on approche de l'ovaire. Chez le mutant *pollen on pistil 2* (*pop2*) d'*Arabidopsis*, déficient pour une GABA-transaminase et présentant de ce fait une quantité excessive de GABA dans le style, les TP ne parviennent pas à progresser correctement[107]. Les gènes associés à

la dynamique de la paroi cellulaire semblent faire partie des cibles de la signalisation GABAergique[108].

Chez le tabac, les glycoprotéines de type arabinogalactane (AGP), riches en hydroxyproline, contribuent également au guidage et à la nutrition des TP[109]. On connaît en particulier les protéines TTS1 et TTS2, qui sont associées à la matrice extracellulaire chez *N. tabacum*[110] et *N. alata*[111]. Ces protéines sont réparties dans le style en un gradient de glycosylation, de plus en plus élevé à mesure que l'on se rapproche de l'ovaire, qui pourrait participer à l'orientation des TP[112].

Rares sont les cas d'interactions pollen–style où les partenaires moléculaires sont connus sur les versants mâle et femelle. Chez la tomate, deux récepteurs kinases LePRK1 et LePRK2, formant un complexe dans la membrane du TP ont été décrits en 2003[113]. Lorsque mises au contact d'extraits stylaires, les deux protéines se dissocient et leurs domaines intracellulaires interagissent avec la protéine pollinique KPP[114] ; ceci initie une cascade de signalisation qui contrôlerait probablement le cytosquelette[115]. Des expériences *in vitro* ont montré que le domaine extracellulaire des LePRK pouvait lier différents interactants du pollen, comme LAT52[116] et SHY[117], et du pistil, comme LeSTIG1[118] ou encore le peptide STIL (*Style Interactor for LePRKs*) qui est capable de déphosphoryler LePRK2[119].

2.2.3 Capacitation du tube pollinique

Il existe des différences notables dans les caractéristiques et le comportement de TP cultivés *in vitro* et *in vivo*, qui mettent en évidence l'existence d'une capacitation du TP par le pistil. Il est connu, par exemple, que le tube pollinique croît plus vite dans le pistil qu'*in vitro*[120]. On sait aussi que son passage dans le style le rend plus réceptif aux signaux chimioattractifs émis par l'ovule[94,121], ce qui a été vérifié chez *Torenia fournieri* en étudiant le guidage de TP ayant crû dans des sections de style de longueurs différentes[122]. Plus récemment, on a identifié chez la même espèce un polysaccharide ovulaire baptisé AMOR qui est requis pour conférer aux TP la compétence à répondre aux signaux de guidage[123,124].

Des études transcriptomiques à large spectre ont permis de mettre en évidence les altérations de l'expression des gènes du TP causées par la croissance dans le pistil. Dès 2009, une étude par biopuce effectuée chez *Arabidopsis* par Qin et coll. a comparé le transcriptome du pollen sec, de TP cultivés *in vitro* et semi-*in vivo* (SIV)[125]. Cette dernière méthode permet de mieux cerner l'effet que la croissance dans le style a sur le pollen. Tout d'abord, on pollinise une fleur de façon

classique, de sorte que les tubes germent et commencent leur élongation dans le pistil. Ensuite, on sectionne le style à sa base pour le placer dans un milieu de culture liquide pour pollen. Les TP finissent par émerger du pistil et l'expérimentateur peut les prélever pour en extraire les ARN[94].

Cette étude a mis en évidence 7044 gènes exprimés dans les TP SIV, dont 383 sont spécifiques au pollen SIV (c.-à-d. absents du pollen sec, du pollen *in vitro* et des autres parties de la plante utilisées comme contrôles). Leur catégorisation fonctionnelle montre un enrichissement net en récepteurs membranaires associés aux réponses de défense lié à la présence d'une famille de récepteurs TIR-NBS-LRR. Ceux-ci sont connus pour être impliqués dans la reconnaissance des effecteurs protéiques dérivés des pathogènes[126,127]. On les suspecte donc d'être des médiateurs importants de l'activation transcriptionnelle du tube pollinique. De plus, les protéines TIR-NBS-LRR sont connues pour évoluer rapidement[126] ; ce trait, typique des protéines reproductives[128], pourrait conférer un haut degré de spécificité à l'interaction pollen–style, et donc être impliqué dans la préférence à l'espèce.

L'étude de Qin et coll. a également permis de mettre au jour les trois facteurs de transcription MYB97, MYB101 et MYB120, dont l'expression est induite par la croissance dans le style et qui contrôleraient l'expression de nombreux gènes en aval[129,130]. Le triple mutant *myb97/101/120* présente une croissance incontrôlée du pollen et une incapacité à décharger les noyaux spermatiques dans le sac embryonnaire[131]. On peut aussi signaler que c'est parmi les transcrits induits par la pollinisation identifiés par Qin et coll. qu'a été plus tard découvert le récepteur PRK6, responsable de la réception des signaux de chimioattraction ovulaire[132].

Dans les années suivantes, d'autres études par biopuce, mais surtout par séquençage d'ARN à haut débit, ont exploré le transcriptome de pistils contenant des TP en étudiant la modulation du transcriptome à différents temps après la pollinisation, chez *Arabidopsis*[133], le peuplier[134] et *Brassica rapa*[135], les mécanismes d'auto-incompatibilité chez les Solanacées[136], les Brassicacées[137] ou encore le thé (*Camellia sinensis*)[138], le rejet du pollen hétérospécifique chez les tomates sauvages[139,140], le maïs[141] et *Arabidopsis*[142], ou encore les micro-ARN impliqués dans les interactions pollen–pistil chez le riz[143].

Toutefois, les travaux mentionnés plus haut impliquent soit d'obtenir des TP SIV, qui ne sont pas dans des conditions biologiquement réalistes en absence d'interaction directe le pistil, soit d'étudier le transcriptome d'un pistil pollinisé, ce qui ne permet pas de discriminer les contributions respectives des deux partenaires. Une étude récente a contribué à résoudre ce problème chez *Arabidopsis* en exploitant le polymorphisme existant entre les écotypes de l'es-

pèce[144]. Les auteurs ont en effet choisi des génotypes mâle et femelle suffisamment éloignés pour présenter entre eux une fréquence élevée de polymorphismes à nucléotide unique (SNP), avec au moins un SNP retrouvé dans ∼60 % des gènes à l'étude. Il leur a ainsi été possible d'effectuer une « dissection *in silico* » du pistil et de montrer que la pollinisation module l'expression de 277 gènes dans le TP et de 6838 gènes sur le versant femelle.

Cette même étude démontre également que la réponse transcriptionnelle à la pollinisation est dépendante du génotype du TP : lorsque le pollen du triple mutant *myb97/101/120* est utilisé pour la pollinisation, 278 transcrits mâles et 34 transcrits femelles présentent un niveau d'expression différent par rapport à une pollinisation avec des TP sauvages. On retrouve notamment un groupe de CRP polliniques de la famille des thionines, dont les auteurs supposent qu'elles sont sécrétées par le TP sous le contrôle des MYB pour alerter les tissus femelles et préparer la réception du TP. Par ailleurs, les auteurs ont remarqué que les 278 transcrits polliniques MYB-dépendants et induits par la pollinisation présentaient un fort degré de divergence et de diversification interspécifique, avec des signes de sélection positive, ce qui en fait des candidats prometteurs pour mieux comprendre la spécificité à l'espèce des interactions pollen–pistil.

Une autre analyse transcriptomique publiée récemment a confirmé l'existence d'une régulation transcriptionnelle différente dans le pistil pollinisé d'*A. thaliana* ou *A. halleri*, selon que le pollen est con- ou hétérospécifique. Cette étude révèle également les similitudes importantes dans la réponse du pistil à l'hétéropollen et à l'invasion par un pathogène (en l'occurrence, *Fusarium graminearum*), marquée notamment par la baisse de l'expression d'un groupe de CRP de type défensine[142].

2.2.4 Autoincompatibilité gamétophytique

a. Principe général

Le terme d'autoincompatibilité regroupe un ensemble de systèmes variés, retrouvés chez environ 50 % des plantes à fleurs, visant tous à empêcher la survenue des autocroisements[145]. Sur le plan génétique, l'autoincompatibilité est gouvernée par un locus *S* polyallélique et hautement polymorphe, comprenant au moins un déterminant mâle et un déterminant femelle suffisamment liés pour être toujours coségrégés. En cas d'autopollinisation, les deux déterminants appartenant au même allèle *S* entrent en conflit, ce qui provoque une cascade d'événements moléculaires et cellulaires conduisant au rejet du pollen.

Chez les Solanacées, les Rosacées et les Scrophulariacées, on retrouve un système particulier dénommé autoincompatibilité gamétophytique (AIG). Le déterminant femelle est dans ce cas une ribonucléase S (S-RNase) sécrétée dans le tissu de transmission ; le déterminant mâle est une protéine SLF (*S-Locus F-box*) exprimée dans le tube pollinique[146]. Le style étant diploïde, ses S-RNases peuvent appartenir à deux allèles *S* différents. Le tube pollinique, haploïde, n'exprime en revanche qu'un seul allèle *S*.

Les mécanismes de l'AIG chez les Solanacées sont encore flous à ce jour. Il semblerait que les S-RNases soient internalisées par le tube pollinique de façon indifférenciée[147] pour ensuite interagir avec SCF, un complexe comprenant SLF et d'autres protéines. Si le croisement est compatible, les S-RNases seraient ubiquitinylées par SCF puis dégradées par le protéasome, laissant ainsi le pollen intact. À l'inverse, si le croisement est incompatible, les S-RNases ne seraient pas dégradées et provoqueraient la mort du tube pollinique en détruisant ses ARN[148]. Les protéines HT[149,150], riches en asparagine, et la protéine 120K[151] sont également requises pour accomplir la réaction d'autoincompatibilité.

b. Incompatibilité unilatérale

Chez les Solanacées, toutes les espèces ne possèdent pas de système d'AIG : on distingue ainsi les espèces autoincompatibles (AI) et autocompatibles (AC). Deux espèces AC peuvent se croiser de façon réciproque. Il se peut également que deux espèces AI puissent se croiser, à condition que les partenaires sexuels n'aient pas les mêmes allèles *S*, comme c'est le cas chez nos espèces d'intérêt. Il existe toutefois une barrière interspécifique particulière, asymétrique, appelée incompatibilité ou incongruité unilatérale (IU). Connue notamment chez les Solanacées, l'IU implique une espèce AC et une espèce AI en respectant que l'on appelle couramment la « loi AI × AC » : le pollen de l'espèce AI peut féconder l'espèce AC, mais la réciproque n'est pas vraie[152].

Les bases génétiques et moléculaires de l'IU sont encore très peu connues. Les premières études sur le sujet suggéraient que l'IU ne reposerait que sur les S-RNases[153,154]. Par exemple, le comportement des TP hétérospécifiques dans le style de *Nicotiana bonariensis* varie selon l'allèle de la S-RNase[155]. Si l'on exprime un gène de S-RNase de *N. alata* (AI) chez une espèce de tabac AC, les TP hétérospécifiques qui pouvaient normalement croître dans le style sont arrêtés[156].

Toutefois, on sait qu'il existe des exceptions à la loi AI × AC : certaines espèces AI présentent en effet des isolats d'invididus redevenus AC qui, malgré la perte des S-RNases, sont toujours capables de rejeter le pollen AC hétérospécifique[157,158]. On sait aussi que selon l'espèce du style, les TP de *S. lycopersicum* subissant l'IU ne sont pas arrêtés au même endroit : l'IU peut être hâtive ou tardive, ce qui suggère l'existence de plusieurs médiateurs la gouvernant[159].

Plusieurs études génétiques mettent justement en évidence que l'IU requiert des protéines distinctes des S-RNases, mais tout de même associées à l'AIG. On sait par exemple que les protéines HT-A et HT-B de *Solanum* sont liées à un locus à caractère quantitatif (*quantitative trait locus*, QTL) de l'IU sur le chromosome 12[159]. On sait aussi que la culline 1, protéine pollinique du complexe SCF, est impliquée dans l'IU chez les tomates[160].

En outre, il a été montré que lorsqu'on inhibe l'expression de l'extensine PELPIII dans le tissu de transmission du style de *N. alata*, le rejet habituellement observé des TP de *N. obtusifolia* et *N. trigonophylla* ne survient plus[161]. PELPIII est ainsi la première protéine sans rapport avec l'AIG dont le rôle dans l'isolement interspécifique stylaire est attesté.

Désormais, les auteurs présentent donc l'IU comme un phénomène complexe faisant intervenir des médiateurs de l'AIG mais aussi des mécanismes qui lui sont spécifiques[162].

2.3 Interactions pollen–ovule

2.3.1 Guidage du tube pollinique par l'ovule

Lorsqu'il est entré dans l'ovaire, le tube pollinique poursuit son chemin dans le tissu de transmission pour arriver à la hauteur d'un ovule, puis longer son funicule jusqu'à bifurquer vers le micropyle, point d'entrée du sac embryonnaire à féconder.

L'étude microscopique du guidage du tube pollinique chez *Arabidopsis* a conduit certains auteurs à distinguer deux types de guidage en fonction de la structure femelle impliquée. Higashiyama et coll. (2003)[163] distinguent ainsi (i) le guidage *funiculaire*, contrôlé à la fois par le sporophyte et le gamétophyte et destiné à faire sortir les TP du tissu de transmission pour les attirer sur le septum puis dans le funicule, et (ii) le guidage *micropylaire* sous le contrôle du seul gamétophyte, qui intervient ensuite pour guider finement les TP vers le micropyle au moyen de signaux d'attraction mais aussi de répulsion[164].

Toutefois, la découverte des premiers médiateurs du guidage a conduit certains auteurs à envisager une autre classificiation, fondée à la fois sur la nature et sur la distance d'action des signaux de guidage. Dresselhaus et Franklin-Tong (2013)[165] définissent ainsi (i) le guidage *micropylaire*, reposant sur de petites protéines sécrétées par le gamétophyte (notamment les cellules synergides), agissant au voisinage immédiat du micropyle, en général à moins de 200 µm et (ii) le guidage *ovulaire* agissant à des distances plus importantes et reposant sur la diffusion de petits composés chimiques non protéiques pouvant provenir de tissus sporopytiques. Nous nous tiendrons ci-après à cette dernière classification.

a. Guidage ovulaire

Les médiateurs du guidage ovulaire comprennent plusieurs petites molécules identifiées chez *Arabidopsis*. Tout d'abord, le GABA, déjà évoqué plus haut, est très concentré au voisinage du micropyle[107]. Il agirait comme agent chimiotropique négatif : de faibles concentrations de GABA autorisent la croissance du pollen alors que de fortes concentrations l'inhibent[108]. Ceci a conduit les auteurs à suggérer que le GABA intervient pour ralentir les TP approchant du micropyle de sorte de faciliter leur contrôle par d'autres médiateurs de guidage. On sait justement que les TP ont une vitesse de croissance moins élevée près du micropyle[166]. On ne sait pas encore si le GABA pourrait participer à la répulsion de TP hétérospécifiques.

Le monoxyde d'azote (NO) produit par les cellules entourant le micropyle, pourrait également exercer un contrôle négatif sur le guidage des TP, comme on l'a montré d'abord *in vitro* chez le lis[167] puis *in vivo* chez *A. thaliana*[168] au moyen du mutant *atnos1*[169]. Le NO semble affecter la signalisation calcique dans le tube pollinique. Bien qu'on ne connaisse pas ses cibles dans le pollen, on sait que d'une manière générale, le NO peut traverser les membranes par simple diffusion, puis en agir (i) en activant des protéines telles que la guanylate cyclase, qui génère de la GMPc et initie des cascades intracellulaires de transduction du signal ou (ii) en provoquant des modifications post-traductionnelles des protéines telles que la nitrosylation ou la nitration[170].

Plusieurs travaux ont également mis en évidence le rôle important du calcium dans les interactions pollen–ovule. On sait qu'il est très concentré dans les synergides de plusieurs espèces[171,172,173,174,175,176] et que des oscillations calciques y sont observées jusqu'à la fécondation[177]. Ces oscillations calciques ont aussi lieu à l'apex du tube pollinique ; elles sont dépendantes de la distance séparant ce dernier des ovules à féconder[178,179]. Les canaux calciques polliniques de type *glutamate receptor-like* (GLR), et en particulier CNGC18 (Cyclic Nucleotide-

Gated Channel 18), participent justement au guidage des TP[85]. Il a été récemment mis en évidence que les chaperones CORNICHON HOMOLOG (CNIH) contrôlent la localisation subcellulaire des canaux GLR dans le TP[86]. Le canal CNGC18, qui permet un influx calcique dans le pollen, est ouvert par un acide aminé atypique présente dans les tissus femelles : la D-sérine. Celle-ci est obtenue par isomérisation de la L-sérine, l'énantiomère lévogyre classique retrouvé dans les protéines. On sait d'ailleurs qu'une des enzymes responsables de cette transformation, la sérine racémase SR1, est fortement exprimée au voisinage du micropyle, probablement pour former un gradient[85,180].

Des canaux ioniques ont également été découverts dans la membrane du réticulum endoplasmique du tube pollinique : les échangeurs de cations/protons CHX21 et CHX23, qui régulent l'assimilation réticulaire d'ions K^+ de façon pH-dépendante, sont ainsi impliqués dans la réponse pollinique aux signaux de guidage. Les TP du double mutant *chx21/chx23* échouent en effet à quitter le tissu de transmission et à se diriger vers les ovules[181].

La façon dont tous ces signaux sont intégrés pour former un réseau de signalisation dans le tube pollinique est encore très obscure. Toutefois, deux protéines MAPK (*mitogen-activated protein kinases*), MPK3 et MPK6, seraient au cœur de cette signalisation[182]. Les TP du double mutant *mpk3/mpk6* sont en effet incapables de croître sur le funicule d'*Arabidopsis*, donc de percevoir les signaux de guidage ovulaire. Ils demeurent toutefois réceptifs aux signaux micropylaires, de diffusion plus restreinte, que nous allons maintenant présenter.

b. Guidage micropylaire

En dehors des petites molécules que nous venons de décrire, les médiateurs du guidage des TP comprennent aussi des protéines, découvertes plus récemment[164,183]. Ces molécules, plus grandes, ne diffusent guère à plus de 200 µm autour du micropyle.

La première protéine de guidage ayant été découverte est ZmEA1 (Egg Apparatus 1), une petite protéine de 94 acides aminés découverte chez le maïs[184]. Sécrétée par l'oosphère et les synergides, ZmEA1 s'accumule dans les 5 à 6 assises cellulaires du nucelle séparant le micropyle du sac embryonnaire et est capable d'attirer les TP à une distance de 100 à 150 µm. ZmEA1 se lie à l'apex du tube pollinique puis y est internalisé et dégradé. Une autre protéine de la même famille, ZmEAL1 (EA1-like 1) a été caractérisée chez le maïs. Synthétisée par l'oosphère, elle est transportée au pôle chalazal du sac embryonnaire dans de petites vésicules et contrôle le destin des cellules antipodales[185].

L'attractant ZmEA1 présente un certain degré de spécificité à l'espèce : il est capable d'attirer les TP du maïs (*Zea mays*) mais pas ceux de l'espèce proche *Tripsacum dactyloides*, ni ceux du tabac[186]. Par ailleurs, ZmEA1 est la première protéine de guidage a avoir été exprimée par transgénèse chez une autre espèce, en l'occurrence dans les synergides d'*A. thaliana* qui sont ainsi devenues capables d'attirer le pollen de maïs *in vitro*[187].

La recherche sur les molécules attractantes a également été accélérée par l'étude de la plante *Torenia fournieri*, une eudicotylédone de la famille des *Linderniaceae*. Contrairement à la plupart des autres espèces modèles, dont le sac embryonnaire est enfoui dans les tissus sporophytiques de l'ovule, celui de *T. fournieri* est protubérant et facilement accessible à l'expérimentateur[121].

Ce modèle a permis de montrer rapidement que les attractants de *Torenia* étaient synthétisés et sécrétés par les cellules synergides[188,189]. Leur identité a été établie en 2009 : il s'agit de petites CRP proche des défensines, appelées LURE1 (TfCRP1) et LURE2 (TfCRP3)[190]. Les LURE sont synthétisées par les synergides et accumulées dans l'appareil filiforme du sac embryonnaire, système de parois épaissies et invaginées occupant le pôle micropylaire du gamétophyte[191].

Dès 2006, des tests de guidage ont prouvé qu'il existait une préférence intraspécifique de la chimioattraction des TP au sein des genres *Torenia* et *Lindernia*, en particulier entre *T. fournieri* et *T. concolor*[192]. En 2011, l'orthologue des LURE de *T. fournieri* a été découvert chez *T. concolor* : il s'agit de la protéine TcLURE1 (TcCRP1) qui attire préférentiellement les TP conspécifiques[193]. Seuls 8 des 62 acides aminés sont différents entre les deux protéines, ce qui suggère qu'il s'agit de résidus clés pour la chimioattraction et l'interaction avec le partenaire pollinique.

C'est justement l'hypothèse d'une divergence rapide des protéines chimioattractives qui a conduit la même équipe de recherche à découvrir, en 2012, les homologues des LURE chez *Arabidopsis thaliana*[194]. La famille des AtLURE comprend 6 gènes, dont 4 codant pour des protéines attractives (AtCRP810_1.1 à 4), un codant pour une protéine proche mais non fonctionnelle (AtCRP810_1.5) et un pseudogène (AtCRP810_1.6). Les AtLURE sont sécrétées par les synergides. Toutefois, des protéines synthétisées par d'autres cellules du sac embryonnaire sont importantes pour le guidage, comme le facteur de transcription CCG de la cellule centrale voisine[195] qui, avec son interactant CBP1[196], contrôle d'une manière encore inconnue l'expression des LURE, ou encore la protéine membranaire GEX3 de l'oosphère[197].

Tout comme les TfLURE, les AtLURE sont des CRP de type *defensin-like*. (Elles ont toutefois un patron de cystéines différent.) Elles ont été identifiées de façon originale par analyse phylogénétique des protéines *defensin-like* connues chez *A. thaliana* [198] et de leurs orthologues chez sa plus proche parente *A. lyrata*. De tous les groupes de *defensin-like* existants, le groupe CRP810_1 est l'unique chez qui les CRP se sont diversifiées après la spéciation.

Bien que les AtLURE aient peu divergé de leurs orthologues AlLURE, elles attirent deux fois moins bien les TP d'*A. lyrata* que ceux conspécifiques. Cette préférence à l'espèce n'est toutefois pas symétrique, les AlLURE attirant aussi bien les TP con- qu'hétérospécifiques [194]. Par ailleurs, si l'on exprime les AtLURE chez *T. fournieri*, non seulement les TP sont attirés mais parviennent même à atteindre le sac embryonnaire, quoique la décharge des noyaux spermatiques ne puisse avoir lieu.

c. Réception des signaux de guidage

En 2016, l'équipe de Tetsuya Higashiyama s'est intéressée aux PRK (*pollen receptor kinases*), une famille de 8 récepteurs kinases à répétitions riches en leucine (LRR) exprimés dans le pollen d'*A. thaliana* et dont les orthologues chez la tomate sont les récepteurs LePRK1 et 2 dont on a parlé plus haut. La mutation systématique de ces récepteurs a permis de mettre en évidence que PRK6 est responsable de la réception des signaux de guidage : il est en effet le seul membre de la famille PRK dont la mutation abolit la capacité du TP à être attiré par la protéine AtLURE1.2 purifiée [132]. Toutefois, PRK3 semble être également impliqué puisque le double mutant *prk3/prk6* présente une incapacité à percevoir les signaux de guidage encore plus forte que le simple mutant *prk6*.

La façon dont le signal perçu par PRK6 est transduit à l'intérieur du TP demeure obscure. Nous avons vu plus haut que les ROP sont suspectées d'être des acteurs importants dans la croissance polaire du tube pollinique. Lorsqu'elle sont liées à la GTP, les ROP exploitent l'hydrolyse de cette dernière en GDP pour activer une série de cibles en aval dont, en l'occurrence, des protéines contrôlant la dynamique des microfilaments d'actine. Après avoir interagi avec leur cible, les ROP restent liées à la GDP et demeurent inactives jusqu'à ce qu'un facteur d'échange de guanosine ROP-GEF remplace la GDP par de la nouvelle GTP. Il a justement été montré que PRK6 interagit avec un facteur ROP-GEF à l'apex du TP [132]. La cascade d'événements reliant cette interaction à la réorientation du TP reste toutefois à découvrir.

Par ailleurs, des résultats récents de cristallographie ont permis de visualiser l'interaction

entre le domaine extracellulaire de PRK6 et son ligand LURE[199]. Il est intéressant de remarquer que, bien que le peptide AtLURE1.2 mature possède 23 acides aminés différents par rapport à son orthologue chez *A. lyrata*, aucun de ces acides aminés divergents n'est directement impliqué dans la liaison avec PRK6. La préférence à l'espèce du guidage du TP pourrait donc reposer sur des altérations plus subtiles dans la conformation de la surface de contact avec le récepteur.

Il est important de signaler qu'en 2016 également, parallèlement à la découverte des PRK, l'équipe de Wei-Cai Yang a mis au jour quatre autres récepteurs polliniques d'*A. thaliana*, MALE DISCOVERER 1 (MDIS1) et MDIS2, et leurs interactants MDIS1-INTERACTING RLK 1 (MIK1) et MIK2[200]. Des expériences *in vitro* ont démontré qu'AtLURE1.2 peut se lier aux domaines extracellulaires de MDIS1, MIK1 et MIK2, et que cette liaison induit la dimérisation et la phosphorylation de ces récepteurs, ce qui initierait une cascade de signalisation dans le TP. Toutefois, les mutants *mdis1* et *mik1/mik2* ne présentent qu'une altération partielle du guidage du TP, moins forte que dans les mutants *prk*. On peut malgré tout suspecter un rôle important de MDIS1 dans la préférence à l'espèce, son expression dans le TP de la proche parente *Capsella rubella* augmentant la réponse aux AtLURE1.2[200].

De futurs travaux permettront de mieux comprendre la façon dont sont intégrés les signaux médiés par PRK, MDIS et MIK, mais aussi par les récepteurs LOST IN POLLEN TUBE GUIDANCE 1 (LIP1) et LIP2 dont la mutation rend le TP incapable de s'orienter vers le micropyle de l'ovule, et moins réceptif aux AtLURE[201].

d. Importance de l'architecture de l'ovule

On a vu que l'existence de cellules synergides fonctionnelles est indispensable pour la production des attractants chez le maïs, *Arabidopsis* et *Torenia*. Ceci vient confirmer plusieurs criblages génétiques qui établissaient dès 1995 une corrélation entre architecture correcte de l'ovule et fonctionnalité du guidage. En effet, les mutants *bel1*[202], *sin1*[203], *ant*[204], *tso1*[205] et *maa1*[206] d'*A. thaliana* présentent à la fois une altération de la structure ou du nombre des ovules et des anomalies de croissance du tube pollinique.

Pour d'autres mutants du guidage des TP découverts chez *A. thaliana* et le maïs, la protéine affectée par la mutation est connue et caractérisée. Il s'agit toujours de protéines jouant un rôle fondamental (*house-keeping function*) dans la cellule, notamment dans l'expression génique :

– MYB98, un facteur de transcription de type R2R3 spécifique aux synergides[191], qui contrôlerait la formation de l'appareil filiforme ainsi que l'expression des CRPs[194,207,208] ;

- CCG (CENTRAL CELL GUIDANCE), un facteur de transcription spécifique à la cellule centrale, proche des protéines gouvernant l'initiation de la transcription [195] qui interagit avec CBP1 [196] ;
- MAA3 (MAGATAMA3), une hélicase importante pour le métabolisme des ARN [209] et dont la mutation produit dans la cellule centrale des noyaux polaires plus petits et incapables de fusionner [206] ;
- PDIL2-1 (protein disulfide isomerase-like 2-1), une protéine du réticulum endoplasmique impliquée dans le repliement des protéines [210] ;
- POD1 (POLLEN DEFECTIVE IN GUIDANCE 1), un interactant de la calréticuline CRT3 [211], une protéine chaperone permettant le repliement des protéines dans le réticulum endoplasmique [212] ;
- ZmDSUL, une protéine du maïs de type diSUMO-like impliquée dans la modification des protéines [213].

Les attractants d'*Arabidopsis* AtLURE sont justement regulés à la baisse chez les mutants *myb98*, *ccg* et *maa3* [194]. Le réseau connectant tous ces gènes demeure encore largement inconnu.

2.3.2 Arrivée du tube pollinique

a. Réception du tube pollinique

Lorsqu'il a atteint le micropyle, le tube pollinique ralentit, traverse l'appareil filiforme et parvient dans le sac embryonnaire [214]. Son extrémité éclate pour délivrer les deux cellules spermatiques. Ces dernières féconderont l'oosphère et la cellule centrale pour donner respectivement le zygote et l'albumen : on parle de double fécondation.

La réception du tube pollinique est finement contrôlée par des interactions impliquant notamment une des deux synergides, qui va dégénérer presque au même moment que la rupture du tube [215]. Chez *T. fournieri*, le délai séparant ces deux événements n'est que d'une seconde [216]. Chez *A. thaliana*, la synergide réceptrice dégénère entre la réception et l'éclatement du pollen [217]. Il semblerait que la signalisation calcique dans la synergide soit associée à ce phénomène : la concentration en ions Ca^{2+} augmente au pôle micropylaire des synergides jusqu'à atteindre un maximum lors de l'éclatement du TP [177,179]. De plus, il a été démontré que la dégénérescence de la synergide réceptive est initiée avant l'éclatement par simple contact avec le TP, sans que l'on

sache si ce contact est direct ou médié par des molécules sécrétées[218]. Sur ce dernier point, on peut rappeler que la mutation des facteurs de transcription polliniques MYB97, 101 et 120, entraîne des défauts de réception du TP en même temps qu'une baisse d'expression d'une famille de thionines suspectées d'être sécrétées à destination de l'ovule[129] (voir section 2.2.3).

Là encore, la réception du tube pollinique pourrait être une barrière à l'hybridation interspécifique. On sait en effet de longue date que le tube pollinique n'est pas arrêté par l'ovule et continue sa croissance de façon aberrante en cas de croisements interspécifiques au sein du genre *Rhododendron*[219]. Plus récemment, on a montré que le même phénotype était retrouvé lorsqu'on croise *A. thaliana* avec ses proches parentes *A. lyrata* et *Cardamine flexuosa*[220].

La réception du tube implique des protéines membranaires. La première à avoir été identifiée est le récepteur FERONIA (FER), aussi appelé SIRÈNE (SRN), exprimé dans la membrane des synergides[220]. Les TP en croissance dans un ovaire mutant *fer* adoptent le phénotype de croissance aberrante évoqué ci-dessus. FER est un récepteur kinase de la famille CrRLK1L présentant un domaine extracellulaire de type malectine[221]. Il a été montré que, dans la racine, ce dernier se liait à une petite sécrétée de la famille RALF (*rapid alkalinization factor*) induisant l'activation d'une pompe à protons ATP-asique médiée par les GTP-ases RAC/ROP, le changement de pH aboutissant à un arrêt de la croissance racinaire[222]. Puisqu'il existe une trentaine de gènes *RALF* chez *Arabidopsis*[223], on peut supposer qu'il existe un ligand RALF pollinique particulier à FER.

Plusieurs protéines coopérant avec FER dans la réception du tube pollinique ont été découvertes en étudiant des mutants ovulaires partageant le même phénotype que *fer*, parmi lesquels le mutant *nortia* (*nta*). NTA est une protéine à sept domaines transmembranaires exprimée dans les synergides. Lors de l'arrivée du TP, FER interagit avec NTA en induisant sa relocalisation à la membrane plasmique, autour du pôle micropylaire des synergides[224]. NTA appartient à la famille MLO (Mildew Resistance Locus O), qui contient d'autres récepteurs membranaires, ainsi qu'une protéine liant la calmoduline et donc potentiellement impliquée dans la signalisation calcique. Il a justement été suggéré que la relocalisation membranaire de NTA dépend d'une concentration suffisante en calcium dans la synergide[225].

Les mutants *evan* (*evn*) et *turan* (*tun*) présentent eux aussi un phénotype similaire à *fer*. EVN, une protéine de type UDP-glycosyltransférase, et TURAN, une dolichol kinase, sont supsectées de contrôler la *N*-glycosylation de protéines exprimées par les synergides pour assurer la réception du TP[226].

LORELEI (LRE), une petite protéine ancrée à la membrane par une liaison GPI (glycosylphosphatidylinositol), est aussi impliquée dans la réception du TP[227]. LRE semble jouer deux rôles distincts. D'une part, avant l'arrivée du TP, LRE interagit avec le domaine extracellulaire de FER et participe, à la manière d'une chaperone, à sa relocalisation dans l'appareil filiforme des synergides[228]. D'autre part, LRE participe ensuite avec FER à la production de ROS dans l'appareil filiforme pour assurer à la bonne réception du TP[229]. Cette seconde fonction de LRE ne semble pas dépendante de sa localisation dans l'ovule, puisque l'expression ectopique de LRE dans le TP permet d'annuler le phénotype des ovules mutants *lre*[230]. Par ailleurs, une autre protéine découverte récemment, EN14, interagit également avec FER dans l'appareil filiforme pour assurer la bonne réception du TP[231].

La protéine peroxysomale AMC contrôlerait aussi la réception du tube pollinique. Le mutant *amc* (*abstinence by mutual consent*) présente en effet un défaut de réception du TP, mais uniquement, comme son nom l'indique, quand le pollen et le pistil sont tous les deux mutants, ce qui suggère que les deux partenaires partagent des éléments de signalisation communs[232]. Les peroxysomes sont de petits organites produisant des molécules de signalisation telles que le NO et les ROS, qui pourraient affecter le tube pollinique.

b. Décharge des noyaux spermatiques

L'éclatement du tube pollinique, permettant la libération des cellules spermatiques, est lui aussi contrôlé, bien que l'on sache encore peu comment. Chez le maïs, on a décrit la protéine ZmES4, une CRP de type *defensin-like* qui s'accumule dans des vésicules au pôle micropylaire des synergides avant d'être sécrétée juste avant l'éclatement du pollen[233]. L'expression de ZmES4 et d'autres protéines spécifiques aux synergides pourrait être gouvernée par le facteur de transcription VERDANDI (VDD). Les mutants *vdd* présentent une altération des synergides et des antipodales corrélée à une incapacité à décharger les cellules spermatiques[234].

ZmES4 interagit avec le canal potassique KZM1, enchâssé dans la membrane pollinique, et provoque la rupture du TP[233]. Cette interaction présente une préférence à l'espèce. On ne sait pas si l'influx d'ions K^+ conduit directement à un changement osmotique faisant éclater le pollen, ou s'il initie une vague de dépolarisation plus ample, médiée par d'autres canaux. On a justement identifié ACA9, une pompe calcique pollinique ATPasique activée par la calmoduline, dont les mutants sont incapables de réaliser la rupture du tube pollinique[235]. Il est intéressant de remarquer que l'homéostat potassique est également impliqué dans le maintien de l'intégrité du TP du riz (*Oryza sativa*), dont le récepteur RUPO interagit avec le canal potassique

HAK1[236].

Enfin, il faut mentionner que les récepteurs ANX1 et 2 évoqués plus haut, ainsi que deux autres RLK dénommés BUDDHA'S PAPER SEAL 1 (BUPS1) and BUPS2 et découverts récemment[237], sont impliqués dans la rupture du TP. Au cours de la croissance du TP dans le pistil, ces quatre récepteurs interagissent normalement avec les peptides polliniques RALF4 et 19 dans le cadre d'une signalisation autocrine nécessaire au maintien de l'intrégité du TP. En revanche, lors de la réception du TP, le peptide ovulaire RALF34, qui lie ces récepteurs avec une plus grande affinité, en déloge RALF4 et 19 et provoque par là même la rupture du TP[238].

2.3.3 Double fécondation

a. Fusion des gamètes

Après l'éclatement du tube pollinique, les cellules génératives mâles parviennent au pôle chalazal de la synergide dégénérée. Après un délai estimé à 8 min chez *Arabidopsis*[239], la première fusionne avec la cellule centrale, la seconde avec l'oosphère ; cette fusion des membranes plasmiques s'appelle la plasmogamie[240]. Ensuite, chaque noyau spermatique fusionne avec le noyau de la cellule femelle dans laquelle il a pénétré : il s'agit de la caryogamie. Chez *Arabidopsis*, la caryogamie a lieu quelques instants plus tôt dans la cellule centrale que dans l'oosphère[241]. Les deux cellules spermatiques semblent être identiques, au sens où la fusion de l'une ou l'autre avec la cellule centrale et l'oosphère semble aléatoire[239,242,243,244].

L'attachement des gamètes fait intervenir la protéine GAMETE EXPRESSED 2 (GEX2), localisée dans la membrane de la cellule générative. Chez le mutant *gex2*, les cellules spermatiques ne parviennent pas à adhérer ni à fusionner aux gamètes femelles[245]. GEX2 contient des domaines domaines extracellulaires de type immunoglobuline semblables à ceux retrouvés dans les protéines de fusion des gamètes des mammifères et des algues, comme la protéine FUS1 de *Chlamydomonas reinhardtii*[246,247]. On ne sait pas encore si cet attachement des gamètes médié par GEX2 participe à l'isolement reproductif chez les Angiospermes, mais il est intéressant de remarquer que son homologue FUS1 présente une divergence interspécifique importante chez les algues, qui contribue aux barrières de spéciation[248].

La plasmogamie a recours à la protéine membranaire HAPLESS 2 (HAP2), aussi appelée GENERATIVE CELL-SPECIFIC 1 (GCS1)[249,250], un agent fusogène connu chez la plupart

des eucaryotes (sauf les champignons)[251,252] et dont l'expression est contrôlée, comme celle de *GEX2*, par le facteur de transcription pollinique de type MYB *DUO1*[253]. Les cellules génératives *hap2* restent attachées à l'oosphère ou à la cellule centrale, mais ne fusionne pas[249]. Il a été récemment démontré que la protéine HAP2 de *Chlamydomonas* entretient des similitudes notables avec les protéines de fusion membranaire de classe II des virus enveloppés[254], et que son orthologue chez *Arabidopsis* possède une hélice apicale lui permettant de s'insérer dans la membrane des cellules femelles à féconder[252].

On ne sait pas encore comment la fonction de HAP2 s'articule aux autres protéines dont on a montré un rôle dans la fusion des gamètes. Par exemple, chez le mutant *glauce* (*glc*), la fusion a lieu avec l'oosphère mais pas avec la cellule centrale[255]. Le mutant *glc* est déficient pour une acyltransférase de type BAHD exprimée dans la cellule centrale. Cette famille d'enzymes, associée au métabolisme secondaire, pourrait être impliquée dans la synthèse par la cellule centrale de médiateurs de la plasmogamie. On sait aussi que la protéine mitochondriale ANK6, exprimée fortement dans le tube pollinique et le sac embryonnaire avant la fécondation, est requise pour la fusion des gamètes[256].

Chez *Arabidopsis*, HAP2 est localisée dans le système endomembranaire de la cellule générative. Lors de l'interaction avec l'oosphère, HAP2 est relocalisée à la membrane. Cet adressage est contrôlé par la protéine EC1, une CRP de type ECA1 (Early Culture Abundant 1) exprimée dans l'oosphère, empaquetée dans des vésicules et sécrétée lors de l'interaction avec le tube pollinique, à l'endroit précis où les deux gamètes fusionnent[241]. EC1 pourrait avoir un récepteur encore inconnu à la surface de la cellule générative.

Ce phénomène de relocalisation à la membrane d'un agent fusogène est déjà attesté chez les animaux, par exemple avec le fusogène IZUMO1 du spermatozoïde des mammifères[247] qui est reconnu par la protéine JUNO à la surface de l'ovule[257]. Chez les mammifères, on a aussi décrit la relocalisation membranaire de tétraspanines de type CD9 dans l'ovule[258,259,260]. On connaît aussi des tétraspanines chez les plantes ; *Arabidopsis* en contient 17, dont deux (TET11 et 12) sont retrouvées dans la membrane des cellules génératives du pollen et une (TET9) a été décrite dans la membrane de l'oosphère et de la cellule centrale[261].

b. Prévention de la polyspermie

L'ovule nouvellement fécondé doit rapidement établir des barrières pour empêcher la polyspermie (ou polytubie), c.-à-d. la fécondation par de nouveaux TP[262]. Il est très rare que

l'oosphère et la cellule centrale soient fécondées par les cellules spermatiques de deux TP différents[263,264]. Cette situation, appelée hétérofécondation, concerne seulement 1 % des fécondations chez *Arabidopsis*[265,266] et chez le maïs[267].

Le mécanisme sous-jacent est encore peu connu. Chez les mutants *maa1* et *maa3* d'*Arabidopsis*, déjà décrits plus haut, plusieurs TP peuvent être attirés simultanément par un même ovule[206]. L'ovule de type sauvage est en revanche capable de repousser les TP surnuméraires[94]. Deux facteurs seraient en jeu : (i) l'émission de signaux de répulsion par l'ovule – de petits peptides, le NO et les ROS ont été proposés comme candidats[268] – et (ii) la dégénérescence de la synergide restée intacte, éliminant du même coup la source des attractants[266,269]. Il a été justement démontré qu'une signalisation médiée par l'éthylène entrait en jeu suite à la fécondation pour induire cette dégnérescence[270].

Une autre observation intéressante a été faite chez *Arabidopsis* : si jamais une première synergide a dégénéré alors que le tube pollinique n'a pas été capable de fusionner, la seconde synergide ne dégénère pas. Elle peut donc continuer à sécréter des molécules attractantes pour attirer un second tube pollinique : on parle de sauvetage de la fécondation (*fertilization recovery*)[269]. La seconde synergide ne dégénérera qu'après la fusion réussie d'un nouveau tube pollinique, ce qui représente un second blocage contre la polyspermie.

Remarques conclusives

Ce tour d'horizon nous aura permis de constater que, de l'adhésion au stigmate à la fusion des gamètes, le pistil est jalonné d'une multitude de points de contrôle dont la participation aux barrières interspécifiques est suspectée, voire avérée. Parmi les protéines que nous avons commentées, une catégorie particulière se détache : celle des protéines riches en cystéines ou CRP. Ces protéines ont en commun la présence d'un peptide signal sécrétoire N-terminal, une petite taille (moins de 150 acides aminés), et surtout un nombre important de cystéines (au moins 6) espacées selon un patron particulier qui permet de diviser les CRP en familles[271].

Les CRP ont ceci de particulier que leur squelette de cystéines est hautement conservé, alors que les autres acides aminés peuvent diverger largement entre paralogues ou orthologues. Les CRP sont ainsi fréquemment impliquées dans des fonctions requérant des interactions à haut degré de spécificité, telles que les interactions plante–pathogène, plante–symbionte[272] ou encore les interactions sexuelles[273]. Ainsi, on a vu que SCA et LTP5 contrôlent l'adhésion et le gui-

dage stylaire du TP, LAT52 et LeSTIG interagissent avec des récepteurs polliniques, SCR/SP11 et PrsS sont impliquées dans l'autoincompatibilité, les peptides LURE dans l'attraction micropylaire du TP, ZmES4 dans la rupture du TP ou encore EC1 dans la fusion des gamètes.

La recherche de nouvelles CRP propres à nos espèces d'intérêt, notamment dans le contexte du guidage du tube pollinique par l'ovule, était donc un objectif important de ce projet de doctorat. C'est dans ce souci que nous avons développé l'outil de recherche bioinformatique KAPPA, présenté au chapitre suivant, et spécifiquement dédié à la recherche et à l'analyse des CRP.

Bibliographie

1. Doucet, J. ; Lee, H. K. et Goring, D. R. (2016). Pollen acceptance or rejection : a tale of two pathways. *Trends Plant Sci.*, 21(12), 1058–67. DOI : 10.1016/j.tplants.2016.09.004. [cit. p. 43]

2. Heslop-Harrison, Y. (1981). Stigma characteristics and angiosperm taxonomy. *Nord. J. Bot.*, 1(3), 401–20. DOI : 10.1111/j.1756-1051.1981.tb00707.x. [cit. p. 43]

3. Heslop-Harrison, Y. et Shivanna, K. R. (1977). The receptive surface of the Angiosperm stigma. *Ann. Bot.*, 41(6), 1233–58. DOI : 10.1093/oxfordjournals.aob.a085414. [cit. p. 43]

4. Zinkl, G. M. ; Zwiebel, B. I. ; Grier, D. G. et Preuss, D. (1999). Pollen-stigma adhesion in *Arabidopsis* : a species-specific interaction mediated by lipophilic molecules in the pollen exine. *Development*, 126(23), 5431–40. [cit. p. 44]

5. Domínguez, E. ; Mercado, J. A. ; Quesada, M. A. et Heredia, A. (1999). Pollen sporopollenin : degradation and structural elucidation. *Sex. Plant Reprod.*, 12(3), 171–8. DOI : 10.1007/s004970050189. [cit. p. 44]

6. Paxson-Sowders, D. M. ; Dodrill, C. H. ; Owen, H. A. et Makaroff, C. A. (2001). DEX1, a novel plant protein, is required for exine pattern formation during pollen development in *Arabidopsis*. *Plant Physiol.*, 127(4), 1739–49. DOI : 10.1104/pp.010517. [cit. p. 44]

7. Ariizumi, T. ; Hatakeyama, K. ; Hinata, K. et coll. (2003). A novel male-sterile mutant of *Arabidopsis thaliana*, *faceless pollen-1*, produces pollen with a smooth surface and an acetolysis-sensitive exine. *Plant Mol. Biol.*, 53(1-2), 107–16. DOI : 10.1023/B :PLAN.0000009269.97773.70. [cit. p. 44]

8. Yang, C. ; Vizcay-Barrena, G. ; Conner, K. et Wilson, Z. A. (2007). MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. *Plant Cell*, 19(11), 3530–48. DOI : 10.1105/tpc.107.054981. [cit. p. 44]

9. Aarts, M. G. ; Hodge, R. ; Kalantidis, K. et coll. (1997). The *Arabidopsis* MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. *Plant J.*, 12(3), 615–23. DOI : 10.1046/j.1365-313X.1997.00615.x. [cit. p. 44]

10. Zhang, Z.-B. ; Zhu, J. ; Gao, J.-F. et coll. (2007). Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. *Plant J.*, 52(3), 528–38. DOI : 10.1111/j.1365-313X.2007.03254.x. [cit. p. 44]

11. Ariizumi, T. ; Hatakeyama, K. ; Hinata, K. et coll. (2004). Disruption of the novel plant protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility in *Arabidopsis thaliana*. *Plant J.*, 39(2), 170–81. DOI : 10.1111/j.1365-313X.2004.02118.x. [cit. p. 44]

12. Chang, H.-S. ; Zhang, C. ; Chang, Y.-H. et coll. (2012). *NO PRIMEXINE AND PLASMA MEMBRANE UNDULATION* is essential for primexine deposition and plasma membrane undulation during microsporogenesis in *Arabidopsis*. *Plant Physiol.*, 158(1), 264–72. DOI : 10.1104/pp.111.184853. [cit. p. 44]

13. Guan, Y.-F. ; Huang, X.-Y. ; Zhu, J. et coll. (2008). *RUPTURED POLLEN GRAIN1*, a member of the MtN3/saliva gene family, is crucial for exine pattern formation and cell integrity of microspores in *Arabidopsis*. *Plant Physiol.*, 147(2), 852–63. DOI : 10.1104/pp.108.118026. [cit. p. 44]

14. Hu, J. ; Wang, Z. ; Zhang, L. et Sun, M.-X. (2014). The *Arabidopsis* exine formation defect (EFD) gene is required for primexine patterning and is critical for pollen fertility. *New Phytol.*, 203(1), 140–54. DOI : 10.1111/nph.12788. [cit. p. 44]

15. Heizmann, P. ; Luu, D.-T. et Dumas, C. (2000). The clues to species specificity of pollination among *Brassicaceae*. *Sex. Plant Reprod.*, 13(3), 157–61. DOI : 10.1007/s004970000052. [cit. p. 44]

16. Luu, D. T. ; Heizmann, P. et Dumas, C. (1997). Pollen-stigma adhesion in kale is not dependent on the self-(in)compatibility genotypes. *Plant Physiol.*, 115(3), 1221–30. DOI : 10.1104/pp.115.3.1221. [cit. p. 44]

17. Stead, A. D. ; Roberts, I. N. et Dickinson, H. G. (1980). Pollen-stigma interaction in *Brassica oleracea* : the role of stigmatic proteins in pollen grain adhesion. *J. Cell Sci.*, 42, 417–23. URL : http ://jcs.biologists.org/content/42/1/417. [cit. p. 44]

18. Zinkl, G. et Preuss, D. (2000). Dissecting *Arabidopsis* pollen–stigma interactions reveals novel mechanisms that confer mating specificity. *Ann. Bot.*, 85(Suppl 1), 15–21. DOI : 10.1006/anbo.1999.1066. [cit. p. 44]

19. Lalonde, B. A. ; Nasrallah, M. E. ; Dwyer, K. G. et coll. (1989). A highly conserved *Brassica* gene with homology to the S-locus-specific glycoprotein structural gene. *Plant Cell*, 1(2), 249–58. DOI : 10.1105/tpc.1.2.249. [cit. p. 44]

20. Luu, D. T. ; Marty-Mazars, D. ; Trick, M. ; Dumas, C. et Heizmann, P. (1999). Pollen-stigma adhesion in *Brassica* spp involves SLG and SLR1 glycoproteins. *Plant Cell*, 11(2), 251–62. DOI : 10.1105/tpc.11.2.251. [cit. p. 44]

21. Umbach, A. L. ; Lalonde, B. A. ; Kandasamy, M. K. ; Nasrallah, J. B. et Nasrallah, M. E. (1990). Immunodetection of protein glycoforms encoded by two independent genes of the self-incompatibility multigene family of *Brassica*. *Plant Physiol.*, 93(2), 739–47. DOI : 10.1104/pp.93.2.739. [cit. p. 44]

22. Doughty, J. ; Hedderson, F. ; McCubbin, A. et Dickinson, H. (1993). Interaction between a coating-borne peptide of the *Brassica* pollen grain and stigmatic *S* (self-incompatibility)-locus-specific glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.*, 90(2), 467–71. DOI : 10.1073/pnas.90.2.467. [cit. p. 44]

23. Luu, D.-T. ; Passelègue, E. ; Dumas, C. et Heizmann, P. (1998). Pollen-stigma capture is not species discriminant within the *Brassicaceae* family. *C. R. Acad. Sci. III, Sci. Vie*, 321(9), 747–55. DOI : 10.1016/S0764-4469(98)80015-2. [cit. p. 44]

24. Inaba, R. et Nishio, T. (2002). Phylogenetic analysis of *Brassiceae* based on the nucleotide sequences of the *S*-locus related gene, SLR1. *Theor. Appl. Genet.*, 105(8), 1159–65. DOI : 10.1007/s00122-002-0968-3. [cit. p. 44]

25. Takayama, S. (2000). Isolation and characterization of pollen coat proteins of *Brassica campestris* that interact with *S* locus-related glycoprotein 1 involved in pollen–stigma adhesion. *Proc. Natl. Acad. Sci. U. S. A.*, 97(7), 3765–70. DOI : 10.1073/pnas.040580797. [cit. p. 44]

26. Hülskamp, M. ; Kopczak, S. D. ; Horejsi, T. F. ; Kihl, B. K. et Pruitt, R. E. (1995). Identification of genes required for pollen-stigma recognition in *Arabidopsis thaliana*. *Plant J.*, 8(5), 703–14. DOI : 10.1046/j.1365-313X.1995.08050703.x. [cit. p. 44]

27. Wolters-Arts, M. ; Lush, W. M. et Mariani, C. (1998). Lipids are required for directional pollen-tube growth. *Nature*, 392(6678), 818–21. DOI : 10.1038/33929. [cit. p. 44]

28. Wolters-Arts, M. ; Van Der Weerd, L. ; Van Aelst, A. C. et coll. (2002). Water-conducting properties of lipids during pollen hydration. *Plant, Cell Environ.*, 25(4), 513–9. DOI : 10.1046/j.1365-3040.2002.00827.x. [cit. p. 44]

29. Mayfield, J. A. ; Fiebig, A. ; Johnstone, S. E. et Preuss, D. (2001). Gene families from the *Arabidopsis thaliana* pollen coat proteome. *Science*, 292(5526), 2482–5. DOI : 10.1126/science.1060972. [cit. p. 45]

30. Mayfield, J. A. et Preuss, D. (2000). Rapid initiation of *Arabidopsis* pollination requires the oleosin-domain protein GRP17. *Nat. Cell Biol.*, 2(2), 128–30. DOI : 10.1038/35000084. [cit. p. 45]

31. Fiebig, A. ; Kimport, R. et Preuss, D. (2004). Comparisons of pollen coat genes across *Brassicaceae* species reveal rapid evolution by repeat expansion and diversification. *Proc. Natl. Acad. Sci. U. S. A.*, 101(9), 3286–91. DOI : 10.1073/pnas.0305448101. [cit. p. 45, 47]

32. Schein, M. ; Yang, Z. ; Mitchell-Olds, T. et Schmid, K. J. (2004). Rapid evolution of a pollen-specific oleosin-like gene family from *Arabidopsis thaliana* and closely related species. *Mol. Biol. Evol.*, 21(4), 659–69. DOI : 10.1093/molbev/msh059. [cit. p. 45]

33. Updegraff, E. P. ; Zhao, F. et Preuss, D. (2009). The extracellular lipase EXL4 is required for efficient hydration of *Arabidopsis* pollen. *Sex. Plant Reprod.*, 22(3), 197–204. DOI : 10.1007/s00497-009-0104-5. [cit. p. 45]

34. Samuel, M. A. ; Chong, Y. T. ; Haasen, K. E. et coll. (2009). Cellular pathways regulating responses to compatible and self-incompatible pollen in *Brassica* and *Arabidopsis* stigmas intersect at Exo70A1, a putative component of the exocyst complex. *Plant Cell*, 21(9), 2655–71. DOI : 10.1105/tpc.109.069740. [cit. p. 45]

35. Synek, L. ; Sekereš, J. et Zárský, V. (2014). The exocyst at the interface between cytoskeleton and membranes in eukaryotic cells. *Front. Plant Sci.*, 4, 543. DOI : 10.3389/fpls.2013.00543. [cit. p. 45]

36. Safavian, D. et Goring, D. R. (2013). Secretory activity is rapidly induced in stigmatic papillae by compatible pollen, but inhibited for self-incompatible pollen in the *Brassicaceae*. *PLoS One*, 8(12), e84286. DOI : 10.1371/journal.pone.0084286. [cit. p. 45]

37. Li, Y. ; Tan, X. ; Wang, M. et coll. (2017). Exocyst subunit SEC3A marks the germination site and is essential for pollen germination in *Arabidopsis thaliana*. *Sci. Rep.*, 7, 40279. DOI : 10.1038/srep40279. [cit. p. 45]

38. Iwano, M. ; Igarashi, M. ; Tarutani, Y. et coll. (2014). A pollen coat-inducible autoinhibited Ca^{2+}-ATPase expressed in stigmatic papilla cells is required for compatible pollination in the *Brassicaceae*. *Plant Cell*, 26(2), 636–49. DOI : 10.1105/tpc.113.121350. [cit. p. 45]

39. Chapman, L. A. et Goring, D. R. (2011). Misregulation of phosphoinositides in *Arabidopsis thaliana* decreases pollen hydration and maternal fertility. *Sex. Plant Reprod.*, 24(4), 319–26. DOI : 10.1007/s00497-011-0172-1. [cit. p. 45]

40. Zhao, L. ; Huang, J. ; Zhao, Z. et coll. (2010). The Skp1-like protein SSK1 is required for cross-pollen compatibility in *S*-RNase-based self-incompatibility. *Plant J.*, 62(1), 52–63. DOI : 10.1111/j.1365-313X.2010.04123.x. [cit. p. 45]

41. Osaka, M. ; Matsuda, T. ; Sakazono, S. et coll. (2013). Cell type-specific transcriptome of *Brassicaceae* stigmatic papilla cells from a combination of laser microdissection and RNA sequencing. *Plant Cell Physiol.*, 54(11), 1894–906. DOI : 10.1093/pcp/pct133. [cit. p. 46]

42. Sankaranarayanan, S. ; Jamshed, M. ; Deb, S. et coll. (2013). Deciphering the stigmatic transcriptional landscape of compatible and self-incompatible pollinations in *Brassica napus* reveals a rapid stigma senescence response following compatible pollination. *Mol. Plant*, 6(6), 1988–91. DOI : 10.1093/mp/sst066. [cit. p. 46]

43. Matsuda, T. ; Matsushima, M. ; Nabemoto, M. et coll. (2015). Transcriptional characteristics and differences in *Arabidopsis* stigmatic papilla cells pre- and post-pollination. *Plant Cell Physiol.*, 56 (4), 663–73. DOI : 10.1093/pcp/pcu209. [cit. p. 46]

44. Hamilton, E. S. ; Jensen, G. S. ; Maksaev, G. et coll. (2015). Mechanosensitive channel MSL8 regulates osmotic forces during pollen hydration and germination. *Science*, 350(6259), 438–41. DOI : 10.1126/science.aac6014. [cit. p. 46]

45. Pérez Di Giorgio, J. A. ; Barberini, M. L. ; Amodeo, G. et Muschietti, J. P. (2016). Pollen aquaporins : what are they there for ? *Plant Signal. Behav.*, 11(9), e1217375. DOI : 10.1080/15592324.2016.1217375. [cit. p. 46]

46. Konar, R. N. et Linskens, H. F. (1966). Physiology and biochemistry of the stigmatic fluid of *Petunia hybrida*. *Planta*, 71(4), 372–87. DOI : 10.1007/BF00396322. [cit. p. 46]

47. Martin, F. W. (1969). Compounds from the stigmas of ten species. *Am. J. Bot.*, 56(9), 1023. DOI : 10.2307/2440924. [cit. p. 46]

48. Labarca, C. et Loewus, F. (1972). The nutritional role of pistil exudate in pollen tube wall formation in *Lilium longiflorum* : I. Utilization of injected stigmatic exudate. *Plant Physiol.*, 50(1), 7–14. DOI : 10.1104/pp.50.1.7. [cit. p. 46]

49. Zafra, A. ; Rodríguez-García, M. I. et Alché, J. D. D. (2010). Cellular localization of ROS and NO in olive reproductive tissues during flower development. *BMC Plant Biol.*, 10, 36. DOI : 10.1186/1471-2229-10-36. [cit. p. 46]

50. Zienkiewicz, K. ; Rejón, J. D. ; Suárez, C. et coll. (2011). Whole-organ analysis of calcium behaviour in the developing pistil of olive (*Olea europaea* L.) as a tool for the determination of key events in sexual plant reproduction. *BMC Plant Biol.*, 11, 150. DOI : 10.1186/1471-2229-11-150. [cit. p. 46]

51. Sanchez, A. M. ; Bosch, M. ; Bots, M. et coll. (2004). Pistil factors controlling pollination. *Plant Cell*, 16 Suppl, S98–106. DOI : 10.1105/tpc.017806. [cit. p. 46]

52. Rejón, J. D. ; Delalande, F. ; Schaeffer-Reiss, C. et coll. (2013). Proteomics profiling reveals novel proteins and functions of the plant stigma exudate. *J. Exp. Bot.*, 64(18), 5695–705. DOI : 10.1093/jxb/ert345. [cit. p. 46]

53. Rejón, J. D. ; Delalande, F. ; Schaeffer-Reiss, C. et coll. (2014). The plant stigma exudate. A biochemically active extracellular environment for pollen germination ? *Plant Signal. Behav.*, 9(4), e28274. DOI : 10.4161/psb.28274. [cit. p. 46]

54. Kim, S. ; Mollet, J.-C. ; Dong, J. et coll. (2003). Chemocyanin, a small basic protein from the lily stigma, induces pollen tube chemotropism. *Proc. Natl. Acad. Sci. U. S. A.*, 100(26), 16125–30. DOI : 10.1073/pnas.2533800100. [cit. p. 47, 50]

55. Park, S. Y. ; Jauh, G. Y. ; Mollet, J. C. et coll. (2000). A lipid transfer-like protein is necessary for lily pollen tube adhesion to an *in vitro* stylar matrix. *Plant Cell*, 12(1), 151–64. DOI : 10.1105/tpc.12.1.151. [cit. p. 47, 50]

56. Hedhly, A. (2011). Sensitivity of flowering plant gametophytes to temperature fluctuations. *Environ. Exp. Bot.*, 74, 9–16. DOI : 10.1016/j.envexpbot.2011.03.016. [cit. p. 47]

57. Salzer, P. ; Hebe, G. et Hager, A. (1997). Cleavage of chitinous elicitors from the ectomycorrhizal fungus *Hebeloma crustuliniforme* by host chitinases prevents induction of K^+ and Cl^- release, extracellular alkalinization and H_2O_2 synthesis of *Picea abies* cells. *Planta*, 203(4), 470–9. DOI : 10.1007/s004250050216. [cit. p. 47]

58. Ding, Y. ; Wang, J. ; Wang, J. et coll. (2012). Unconventional protein secretion. *Trends Plant Sci.*, 17(10), 606–15. DOI : 10.1016/j.tplants.2012.06.004. [cit. p. 47]

59. Regente, M. ; Corti-Monzón, G. ; Maldonado, A. M. et coll. (2009). Vesicular fractions of sunflower apoplastic fluids are associated with potential exosome marker proteins. *FEBS Lett.*, 583(20), 3363–6. DOI : 10.1016/j.febslet.2009.09.041. [cit. p. 47]

60. Prado, N. ; Alché, J. D. D. ; Casado-Vela, J. et coll. (2014). Nanovesicles are secreted during pollen germination and pollen tube growth : a possible role in fertilization. *Mol. Plant*, 7(3), 573–7. DOI : 10.1093/mp/sst153. [cit. p. 47]

61. Iwano, M. ; Shiba, H. ; Miwa, T. et coll. (2004). Ca^{2+} dynamics in a pollen grain and papilla cell during pollination of *Arabidopsis*. *Plant Physiol.*, 136(3), 3562–71. DOI : 10.1104/pp.104.046961. [cit. p. 47]

62. Zhou, L. ; Fu, Y. et Yang, Z. (2009). A genome-wide functional characterization of *Arabidopsis* regulatory calcium sensors in pollen tubes. *J. Integr. Plant Biol.*, 51(8), 751–61. DOI : 10.1111/j.1744-7909.2009.00847.x. [cit. p. 48]

63. Qin, Y. ; Wysocki, R. J. ; Somogyi, A. et coll. (2011). Sulfinylated azadecalins act as functional mimics of a pollen germination stimulant in *Arabidopsis* pistils. *Plant J.*, 68(5), 800–15. DOI : 10.1111/j.1365-313X.2011.04729.x. [cit. p. 48]

64. Mo, Y. ; Nagel, C. et Taylor, L. P. (1992). Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proc. Natl. Acad. Sci. U. S. A.*, 89(15), 7213–7. DOI : 10.1073/pnas.89.15.7213. [cit. p. 48]

65. Guyon, V. N. ; Astwood, J. D. ; Garner, E. C. ; Dunker, A. K. et Taylor, L. P. (2000). Isolation and characterization of cDNAs expressed in the early stages of flavonol-induced pollen germination in petunia. *Plant Physiol.*, 123(2), 699–710. DOI : 10.1104/pp.123.2.699. [cit. p. 48]

66. Vogler, F. ; Schmalzl, C. ; Englhart, M. ; Bircheneder, M. et Sprunck, S. (2014). Brassinosteroids promote *Arabidopsis* pollen germination and growth. *Plant Reprod.*, 27(3), 153–67. DOI : 10.1007/s00497-014-0247-x. [cit. p. 48]

67. Rounds, C. M. et Bezanilla, M. (2013). Growth mechanisms in tip-growing plant cells. *Annu. Rev. Plant Biol.*, 64, 243–65. DOI : 10.1146/annurev-arplant-050312-120150. [cit. p. 48]

68. Williams, J. H. ; Edwards, J. A. et Ramsey, A. J. (2016). Economy, efficiency, and the evolution of pollen tube growth rates. *Am. J. Bot.*, 103(3), 471–83. DOI : 10.3732/ajb.1500264. [cit. p. 48]

69. Grebnev, G. ; Ntefidou, M. et Kost, B. (2017). Secretion and endocytosis in pollen tubes : models of tip growth in the spot light. *Front. Plant Sci.*, 8, 154. DOI : 10.3389/fpls.2017.00154. [cit. p. 48]

70. Mollet, J.-C. ; Leroux, C. ; Dardelle, F. et Lehner, A. (2013). Cell wall composition, biosynthesis and remodeling during pollen tube growth. *Plants*, 2(1), 107–47. DOI : 10.3390/plants2010107. [cit. p. 48]

71. Jiang, L. ; Yang, S.-L. ; Xie, L.-F. et coll. (2005). *VANGUARD1* encodes a pectin methylesterase that enhances pollen tube growth in the *Arabidopsis* style and transmitting tract. *Plant Cell*, 17 (2), 584–96. DOI : 10.1105/tpc.104.027631. [cit. p. 48]

72. Chebli, Y. et Geitmann, A. (2017). Cellular growth in plants requires regulation of cell wall biochemistry. *Curr. Opin. Cell Biol.*, 44, 28–35. DOI : 10.1016/j.ceb.2017.01.002. [cit. p. 49]

73. Ndinyanka Fabrice, T. ; Vogler, H. ; Draeger, C. et coll. (2017). LRX proteins play a crucial role in pollen grain and pollen tube cell wall development. *Plant Physiol.*, 176(3), 1981–92. DOI : 10.1104/pp.17.01374. [cit. p. 49]

74. Sede, A. R. ; Borassi, C. ; Wengier, D. L. et coll. (2017). *Arabidopsis* pollen extensins LRX are required for cell wall integrity during pollen tube growth. *FEBS Lett.*, 592(2), 233–43. DOI : 10.1002/1873-3468.12947. [cit. p. 49]

75. Mecchia, M. A. ; Santos-Fernandez, G. ; Duss, N. N. et coll. (2017). RALF4/19 peptides interact with LRX proteins to control pollen tube growth in *Arabidopsis*. *Science*, 358(6370), 1600–3. DOI : 10.1126/science.aao5467. [cit. p. 49]

76. Boisson-Dernier, A. ; Roy, S. ; Kritsas, K. et coll. (2009). Disruption of the pollen-expressed FERONIA homologs ANXUR1 and ANXUR2 triggers pollen tube discharge. *Development*, 136 (19), 3279–88. DOI : 10.1242/dev.040071. [cit. p. 49]

77. Miyazaki, S. ; Murata, T. ; Sakurai-Ozato, N. et coll. (2009). *ANXUR1* and *2*, sister genes to *FERONIA/SIRENE*, are male factors for coordinated fertilization. *Curr. Biol.*, 19(15), 1327–31. DOI : 10.1016/j.cub.2009.06.064. [cit. p. 49]

78. Boisson-Dernier, A. ; Lituiev, D. S. ; Nestorova, A. et coll. (2013). ANXUR receptor-like kinases coordinate cell wall integrity with growth at the pollen tube tip via NADPH oxidases. *PLoS Biol.*, 11(11), e1001719. DOI : 10.1371/journal.pbio.1001719. [cit. p. 49]

79. Franck, C. ; Westermann, J. ; Bürssner, S. et coll. (2018). The protein phosphatases ATUNIS1 and ATUNIS2 regulate cell wall integrity in tip-growing cells. *Plant Cell*, 30(8), 1906–23. DOI : 10.1105/tpc.18.00284. [cit. p. 49]

80. Boisson-Dernier, A. ; Franck, C. M. ; Lituiev, D. S. et Grossniklaus, U. (2015). Receptor-like cytoplasmic kinase MARIS functions downstream of CrRLK1L-dependent signaling during tip growth. *Proc. Natl. Acad. Sci. U. S. A.*, 112(39), 12211–6. DOI : 10.1073/pnas.1512375112. [cit. p. 49]

81. Steinhorst, L. et Kudla, J. (2013). Calcium – a central regulator of pollen germination and tube growth. *Biochim. Biophys. Acta*, 1833(7), 1573–81. DOI : 10.1016/j.bbamcr.2012.10.009. [cit. p. 49]

82. Damineli, D. S. C. ; Portes, M. T. et Feijó, J. A. (2017). Oscillatory signatures underlie growth regimes in *Arabidopsis* pollen tubes : computational methods to estimate tip location, periodicity, and synchronization in growing cells. *J. Exp. Bot.*, 68(12), 3267–81. DOI : 10.1093/jxb/erx032. [cit. p. 49]

83. Hepler, P. K. (2016). The cytoskeleton and its regulation by calcium and protons. *Plant Physiol.*, 170(1), 3–22. DOI : 10.1104/pp.15.01506. [cit. p. 49]

84. Gao, Q.-F. ; Gu, L.-L. ; Wang, H.-Q. et coll. (2016). Cyclic nucleotide-gated channel 18 is an essential Ca^{2+} channel in pollen tube tips for pollen tube guidance to ovules in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.*, 113(11), 3096–101. DOI : 10.1073/pnas.1524629113. [cit. p. 49]

85. Michard, E. ; Lima, P. T. ; Borges, F. et coll. (2011). Glutamate receptor-like genes form Ca^{2+} channels in pollen tubes and are regulated by pistil D-serine. *Science*, 332(6028), 434–7. DOI : 10.1126/science.1201101. [cit. p. 49, 57]

86. Wudick, M. M. ; Portes, M. T. ; Michard, E. et coll. (2018). CORNICHON sorting and regulation of GLR channels underlie pollen tube Ca^{2+} homeostasis. *Science*, 360(6388), 533–6. DOI : 10.1126/science.aar6464. [cit. p. 49, 57]

87. Rahmati Ishka, M. ; Brown, E. ; Weigand, C. et coll. (2018). A comparison of heat-stress transcriptome changes between wild-type *Arabidopsis* pollen and a heat-sensitive mutant harboring a knockout of *cyclic nucleotide-gated cation channel 16* (*cngc16*). *BMC Genomics*, 19(1), 549. DOI : 10.1186/s12864-018-4930-4. [cit. p. 49]

88. Chen, W. ; Gong, P. ; Guo, J. et coll. (2018). Glycolysis regulates pollen tube polarity via Rho GTPase signaling. *PLoS Genet.*, 14(4), e1007373. DOI : 10.1371/journal.pgen.1007373. [cit. p. 49]

89. Fehér, A. et Lajkó, D. B. (2015). Signals fly when kinases meet Rho-of-plants (ROP) small G-proteins. *Plant Sci.*, 237, 93–107. DOI : 10.1016/j.plantsci.2015.05.007. [cit. p. 49]

90. Feng, Q.-N. ; Kang, H. ; Song, S.-J. et coll. (2016). *Arabidopsis* RhoGDIs are critical for cellular homeostasis of pollen tubes. *Plant Physiol.*, 170(2), 841–56. DOI : 10.1104/pp.15.01600. [cit. p. 49]

91. Lee, Y. J. et Yang, Z. (2008). Tip growth : signaling in the apical dome. *Curr. Opin. Plant Biol.*, 11(6), 662–71. DOI : 10.1016/j.pbi.2008.10.002. [cit. p. 49]

92. Mangano, S. ; Juárez, S. P. D. et Estevez, J. M. (2016). ROS regulation of polar growth in plant cells. *Plant Physiol.*, 171(3), 1593–605. DOI : 10.1104/pp.16.00191. [cit. p. 50]

93. Lennon, K. A. ; Roy, S. ; Hepler, P. K. et Lord, E. M. (1998). The structure of the transmitting tissue of *Arabidopsis thaliana* (L.) and the path of pollen tube growth. *Sex. Plant Reprod.*, 11(1), 49–59. DOI : 10.1007/s004970050120. [cit. p. 50]

94. Palanivelu, R. et Preuss, D. (2006). Distinct short-range ovule signals attract or repel *Arabidopsis thaliana* pollen tubes *in vitro*. *BMC Plant Biol.*, 6, 7. DOI : 10.1186/1471-2229-6-7. [cit. p. 50, 51, 52, 66]

95. Crawford, B. C. W. ; Ditta, G. et Yanofsky, M. F. (2007). The NTT gene is required for transmitting-tract development in carpels of *Arabidopsis thaliana*. *Curr. Biol.*, 17(13), 1101–8. DOI : 10.1016/j.cub.2007.05.079. [cit. p. 50]

96. Crawford, B. C. W. et Yanofsky, M. F. (2011). HALF FILLED promotes reproductive tract development and fertilization efficiency in *Arabidopsis thaliana*. *Development*, 138(14), 2999–3009. DOI : 10.1242/dev.067793. [cit. p. 50]

97. Grobe, K. ; Becker, W. M. ; Schlaak, M. et Petersen, A. (1999). Grass group I allergens (beta-expansins) are novel, papain-related proteinases. *Eur. J. Biochem.*, 263(1), 33–40. DOI : 10.1046/j.1432-1327.1999.00462.x. [cit. p. 50]

98. Stratford, S. ; Barne, W. ; Hohorst, D. L. et coll. (2001). A leucine-rich repeat region is conserved in pollen extensin-like (Pex) proteins in monocots and dicots. *Plant Mol. Biol.*, 46(1), 43–56. DOI : 10.1023/A :1010659425399. [cit. p. 50]

99. Ogawa, M. ; Kay, P. ; Wilson, S. et Swain, S. M. (2009). ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE1 (ADPG1), ADPG2, and QUARTET2 are polygalacturonases required for cell separation during reproductive development in emphArabidopsis. *Plant Cell*, 21 (1), 216–33. DOI : 10.1105/tpc.108.063768. [cit. p. 50]

100. Smith, D. K. ; Harper, J. F. et Wallace, I. S. (2018). A potential role for protein O-fucosylation during pollen-pistil interactions. *Plant Signal. Behav.*, 13(5), e1467687. DOI : 10.1080/15592324.2018.1467687. [cit. p. 50]

101. Chen, L.-Y. ; Shi, D.-Q. ; Zhang, W.-J. et coll. (2015). The *Arabidopsis* alkaline ceramidase TOD1 is a key turgor pressure regulator in plant cells. *Nat. Commun.*, 6, 6030. DOI : 10.1038/ncomms7030. [cit. p. 50]

102. Edlund, A. F. ; Swanson, R. et Preuss, D. (2004). Pollen and stigma structure and function : the role of diversity in pollination. *Plant Cell*, 16 Suppl, S84–97. DOI : 10.1105/tpc.015800. [cit. p. 50]

103. Lord, E. M. (2001). Adhesion molecules in lily pollination. *Sex. Plant Reprod.*, 14(1-2), 57–62. DOI : 10.1007/s004970100076. [cit. p. 50]

104. Mollet, J.-C. (2000). A lily stylar pectin is necessary for pollen tube adhesion to an *in vitro* stylar matrix. *Plant Cell*, 12(9), 1737–50. DOI : 10.1105/tpc.12.9.1737. [cit. p. 50]

105. Dong, J. ; Kim, S. T. et Lord, E. M. (2005). Plantacyanin plays a role in reproduction in *Arabidopsis*. *Plant Physiol.*, 138(2), 778–89. DOI : 10.1104/pp.105.063388. [cit. p. 50]

106. Chae, K. ; Kieslich, C. A. ; Morikis, D. ; Kim, S.-C. et Lord, E. M. (2009). A gain-of-function mutation of *Arabidopsis* lipid transfer protein 5 disturbs pollen tube tip growth and fertilization. *Plant Cell*, 21(12), 3902–14. DOI : 10.1105/tpc.109.070854. [cit. p. 50]

107. Palanivelu, R. ; Brass, L. ; Edlund, A. F. et Preuss, D. (2003). Pollen tube growth and guidance is regulated by *POP2*, an *Arabidopsis* gene that controls GABA levels. *Cell*, 114(1), 47–59. DOI : 10.1016/S0092-8674(03)00479-3. [cit. p. 50, 56]

108. Renault, H. ; El Amrani, A. ; Palanivelu, R. et coll. (2011). GABA accumulation causes cell elongation defects and a decrease in expression of genes encoding secreted and cell wall-related proteins in *Arabidopsis thaliana*. *Plant Cell Physiol.*, 52(5), 894–908. DOI : 10.1093/pcp/pcr041. [cit. p. 51, 56]

109. Wu, G. ; Gu, Y. ; Li, S. et Yang, Z. (2001). A genome-wide analysis of *Arabidopsis* Rop-interactive CRIB motif-containing proteins that act as Rop GTPase targets. *Plant Cell*, 13(12), 2841–56. DOI : 10.1105/tpc.13.12.2841. [cit. p. 51]

110. Cheung, A. Y. ; Wang, H. et Wu, H. M. (1995). A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. *Cell*, 82(3), 383–93. DOI : 10.1016/0092-8674(95)90427-1. [cit. p. 51]

111. Wu, H.-M. ; Wong, E. ; Ogdahl, J. et Cheung, A. Y. (2000). A pollen tube growth-promoting arabinogalactan protein from *Nicotiana alata* is similar to the tobacco TTS protein. *Plant J.*, 22 (2), 165–76. DOI : 10.1046/j.1365-313x.2000.00731.x. [cit. p. 51]

112. Wu, H.-M. ; Wang, H. et Cheung, A. Y. (1995). A pollen tube growth stimulatory glycoprotein is deglycosylated by pollen tubes and displays a glycosylation gradient in the flower. *Cell*, 82(3), 395–403. DOI : 10.1016/0092-8674(95)90428-X. [cit. p. 51]

113. Wengier, D. ; Valsecchi, I. ; Cabanas, M. L. et coll. (2003). The receptor kinases LePRK1 and LePRK2 associate in pollen and when expressed in yeast, but dissociate in the presence of style extract. *Proc. Natl. Acad. Sci. U. S. A.*, 100(11), 6860–5. DOI : 10.1073/pnas.0631728100. [cit. p. 51]

114. Kaothien, P. ; Ok, S. H. ; Shuai, B. et coll. (2005). Kinase partner protein interacts with the LePRK1 and LePRK2 receptor kinases and plays a role in polarized pollen tube growth. *Plant J.*, 42 (4), 492–503. DOI : 10.1111/j.1365-313X.2005.02388.x. [cit. p. 51]

115. Gui, C.-P. ; Dong, X. ; Liu, H.-K. et coll. (2014). Overexpression of the tomato pollen receptor kinase LePRK1 rewires pollen tube growth to a blebbing mode. *Plant Cell*, 26(9), 3538–55. DOI : 10.1105/tpc.114.127381. [cit. p. 51]

116. Tang, W. ; Ezcurra, I. ; Muschietti, J. et McCormick, S. (2002). A cysteine-rich extracellular protein, LAT52, interacts with the extracellular domain of the pollen receptor kinase LePRK2. *Plant Cell*, 14(9), 2277–87. DOI : 10.1105/tpc.003103. [cit. p. 51]

117. Guyon, V. ; Tang, W.-H. ; Monti, M. M. et coll. (2004). Antisense phenotypes reveal a role for SHY, a pollen-specific leucine-rich repeat protein, in pollen tube growth. *Plant J.*, 39(4), 643–54. DOI : 10.1111/j.1365-313X.2004.02162.x. [cit. p. 51]

118. Tang, W. ; Kelley, D. ; Ezcurra, I. ; Cotter, R. et McCormick, S. (2004). LeSTIG1, an extracellular binding partner for the pollen receptor kinases LePRK1 and LePRK2, promotes pollen tube growth *in vitro*. *Plant J.*, 39(3), 343–53. DOI : 10.1111/j.1365-313X.2004.02139.x. [cit. p. 51]

119. Wengier, D. L. ; Mazzella, M. A. ; Salem, T. M. ; McCormick, S. et Muschietti, J. P. (2010). STIL, a peculiar molecule from styles, specifically dephosphorylates the pollen receptor kinase LePRK2 and stimulates pollen tube growth *in vitro*. *BMC Plant Biol.*, 10, 33. DOI : 10.1186/1471-2229-10-33. [cit. p. 51]

120. Taylor, L. P. et Hepler, P. K. (1997). Pollen germination and tube growth. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 48, 461–91. DOI : 10.1146/annurev.arplant.48.1.461. [cit. p. 51]

121. Higashiyama, T. (1998). Guidance *in vitro* of the pollen tube to the naked embryo sac of *Torenia fournieri*. *Plant Cell*, 10(12), 2019–32. DOI : 10.1105/tpc.10.12.2019. [cit. p. 51, 58]

122. Okuda, S. ; Suzuki, T. ; Kanaoka, M. M. et coll. (2013). Acquisition of LURE-binding activity at the pollen tube tip of *Torenia fournieri*. *Mol. Plant*, 6(4), 1074–90. DOI : 10.1093/mp/sst050. [cit. p. 51]

123. Mizukami, A. G. ; Inatsugi, R. ; Jiao, J. et coll. (2016). The AMOR arabinogalactan sugar chain induces pollen-tube competency to respond to ovular guidance. *Curr. Biol.*, 26(8), 1091–7. DOI : 10.1016/j.cub.2016.02.040. [cit. p. 51]

124. Jiao, J. ; Mizukami, A. G. ; Sankaranarayanan, S. et coll. (2017). Structure-activity relation of AMOR sugar molecule that activates pollen-tubes for ovular guidance. *Plant Physiol.*, 173(1), 354–63. DOI : 10.1104/pp.16.01655. [cit. p. 51]

125. Qin, Y. ; Leydon, A. R. ; Manziello, A. et coll. (2009). Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genet.*, 5(8), e1000621. DOI : 10.1371/journal.pgen.1000621. [cit. p. 51]

126. Meyers, B. C. (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell*, 15(4), 809–34. DOI : 10.1105/tpc.009308. [cit. p. 52]

127. Burch-Smith, T. M. et Dinesh-Kumar, S. P. (2007). The functions of plant TIR domains. *Sci. Signaling*, 2007(401), pe46. DOI : 10.1126/stke.4012007pe46. [cit. p. 52]

128. Swanson, W. J. et Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. *Nat. Rev. Genet.*, 3(2), 137–44. DOI : 10.1038/nrg733. [cit. p. 52]

129. Leydon, A. R. ; Beale, K. M. ; Woroniecka, K. et coll. (2013). Three MYB transcription factors control pollen tube differentiation required for sperm release. *Curr. Biol.*, 23(13), 1209–14. DOI : 10.1016/j.cub.2013.05.021. [cit. p. 52, 62]

130. Leydon, A. R. ; Chaibang, A. et Johnson, M. A. (2014). Interactions between pollen tube and pistil control pollen tube identity and sperm release in the *Arabidopsis* female gametophyte. *Biochem. Soc. Trans.*, 42(2), 340–5. DOI : 10.1042/BST20130223. [cit. p. 52]

131. Liang, Y. ; Tan, Z.-M. ; Zhu, L. et coll. (2013). MYB97, MYB101 and MYB120 function as male factors that control pollen tube-synergid interaction in *Arabidopsis thaliana* fertilization. *PLoS Genet.*, 9(11), e1003933. DOI : 10.1371/journal.pgen.1003933. [cit. p. 52]

132. Takeuchi, H. et Higashiyama, T. (2016). Tip-localized receptors control pollen tube growth and LURE sensing in *Arabidopsis*. *Nature*, 531(7593), 245–8. DOI : 10.1038/nature17413. [cit. p. 52, 59]

133. Boavida, L. C. ; Borges, F. ; Becker, J. D. et Feijó, J. A. (2011). Whole genome analysis of gene expression reveals coordinated activation of signaling and metabolic pathways during pollen-pistil interactions in *Arabidopsis*. *Plant Physiol.*, 155(4), 2066–80. DOI : 10.1104/pp.110.169813. [cit. p. 52]

134. Rao, P. ; Chen, Z. ; Yang, X. et coll. (2017). Dynamic transcriptomic analysis of the early response of female flowers of *Populus alba* × *P. glandulosa* to pollination. *Sci. Rep.*, 7(1), 6048. DOI : 10.1038/s41598-017-06255-3. [cit. p. 52]

135. Tan, H. ; Zhang, J. ; Qi, X. et coll. (2018). Integrated metabolite profiling and transcriptome analysis reveals a dynamic metabolic exchange between pollen tubes and the style during fertilization of *Brassica napus*. *Plant Mol. Biol.* DOI : 10.1007/s11103-018-0740-y. [cit. p. 52]

136. Zhao, P. ; Zhang, L. et Zhao, L. (2015). Dissection of the style's response to pollination using transcriptome profiling in self-compatible (*Solanum pimpinellifolium*) and self-incompatible (*Solanum chilense*) tomato species. *BMC Plant Biol.*, 15, 119. DOI : 10.1186/s12870-015-0492-7. [cit. p. 52]

137. Zhang, T. ; Gao, C. ; Yue, Y. et coll. (2017). Time-course transcriptome analysis of compatible and incompatible pollen-stigma interactions in *Brassica napus* L. *Front. Plant Sci.*, 8, 682. DOI : 10.3389/fpls.2017.00682. [cit. p. 52]

138. Ma, Q. ; Chen, C. ; Zeng, Z. et coll. (2018). Transcriptomic analysis between self- and cross-pollinated pistils of tea plants (*Camellia sinensis*). *BMC Genomics*, 19(1), 289. DOI : 10.1186/s12864-018-4674-1. [cit. p. 52]

139. Pease, J. B. ; Guerrero, R. F. ; Sherman, N. A. ; Hahn, M. W. et Moyle, L. C. (2016). Molecular mechanisms of postmating prezygotic reproductive isolation uncovered by transcriptome analysis. *Mol. Ecol.*, 25(11), 2592–608. DOI : 10.1111/mec.13679. [cit. p. 52]

140. Broz, A. K. ; Guerrero, R. F. ; Randle, A. M. et coll. (2017). Transcriptomic analysis links gene expression to unilateral pollen-pistil reproductive barriers. *BMC Plant Biol.*, 17(1), 81. DOI : 10.1186/s12870-017-1032-4. [cit. p. 52]

141. Wang, M. ; Chen, Z. ; Zhang, H. ; Chen, H. et Gao, X. (2018). Transcriptome analysis provides insight into the molecular mechanisms underlying *gametophyte factor 2*-mediated cross-incompatibility in maize. *Int. J. Mol. Sci.*, 19(6), 1757. DOI : 10.3390/ijms19061757. [cit. p. 52]

142. Mondragón-Palomino, M. ; John-Arputharaj, A. ; Pallmann, M. et Dresselhaus, T. (2017). Similarities between reproductive and immune pistil transcriptomes of *Arabidopsis* species. *Plant Physiol.*, 174(3), 1559–1575. DOI : 10.1104/pp.17.00390. [cit. p. 52, 53]

143. Wang, K. ; Wang, X. ; Li, M. ; Shi, T. et Yang, P. (2017). Low genetic diversity and functional constraint of miRNA genes participating pollen-pistil interaction in rice. *Plant Mol. Biol.*, 95 (1–2), 89–98. DOI : 10.1007/s11103-017-0638-0. [cit. p. 52]

144. Leydon, A. R. ; Weinreb, C. ; Venable, E. et coll. (2017). The molecular dialog between flowering plant reproductive partners defined by SNP-informed RNA-sequencing. *Plant Cell*. DOI : 10.1105/tpc.16.00816. [cit. p. 53]

145. Takayama, S. et Isogai, A. (2005). Self-incompatibility in plants. *Annu. Rev. Plant Biol.*, 56, 467–89. DOI : 10.1146/annurev.arplant.56.032604.144249. [cit. p. 53]

146. McClure, B. (2006). New views of S-RNase-based self-incompatibility. *Curr. Opin. Plant Biol.*, 9 (6), 639–46. DOI : 10.1016/j.pbi.2006.09.004. [cit. p. 54]

147. Luu, D. T. ; Qin, X. ; Morse, D. et Cappadocia, M. (2000). S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. *Nature*, 407(6804), 649–51. DOI : 10.1038/35036623. [cit. p. 54]

148. McClure, B. A. et Franklin-Tong, V. (2006). Gametophytic self-incompatibility : understanding the cellular mechanisms involved in "self" pollen tube inhibition. *Planta*, 224(2), 233–45. DOI : 10.1007/s00425-006-0284-2. [cit. p. 54]

149. McClure, B. ; Mou, B. ; Canevascini, S. et Bernatzky, R. (1999). A small asparagine-rich protein required for S-allele-specific pollen rejection in *Nicotiana*. *Proc. Natl. Acad. Sci. U. S. A.*, 96(23), 13548–53. DOI : 10.1073/pnas.96.23.13548. [cit. p. 54]

150. O'Brien, M. ; Kapfer, C. ; Major, G. et coll. (2002). Molecular analysis of the stylar-expressed *Solanum chacoense* small asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility in *Nicotiana*. *Plant J.*, 32(6), 985–96. DOI : 10.1046/j.1365-313X.2002.01486.x. [cit. p. 54]

151. Hancock, C. N. ; Kent, L. et McClure, B. A. (2005). The stylar 120 kDa glycoprotein is required for *S*-specific pollen rejection in *Nicotiana*. *Plant J.*, 43(5), 716–23. DOI : 10.1111/j.1365-313X.2005.02490.x. [cit. p. 54]

152. Lewis, D. et Crowe, L. K. (1958). Unilateral interspecific incompatibility in flowering plants. *Heredity*, 12(2), 233–56. DOI : 10.1038/hdy.1958.26. [cit. p. 54]

153. Pandey, K. K. (1981). Evolution of unilateral incompatibility in flowering plants : further evidence in favour of twin specificities controlling intra- and interspecific incompatibility. *New Phytol.*, 89 (4), 705–28. DOI : 10.1111/j.1469-8137.1981.tb02349.x. [cit. p. 54]

154. Chetelat, R. T. et Deverna, J. W. (1991). Expression of unilateral incompatibility in pollen of *Lycopersicon pennellii* is determined by major loci on chromosomes 1, 6 and 10. *Theor. Appl. Genet.*, 82(6), 704–12. DOI : 10.1007/BF00227314. [cit. p. 54]

155. Pandey, K. K. (1973). Phases in the *S*-gene expression, and *S*-allele interaction in the control of interspecific incompatibility. *Heredity*, 31(3), 381–400. DOI : 10.1038/hdy.1973.93. [cit. p. 54]

156. Murfett, J. ; Strabala, T. J. ; Zurek, D. M. et coll. (1996). S RNase and interspecific pollen rejection in the genus *Nicotiana* : multiple pollen-rejection pathways contribute to unilateral incompatibility between self-incompatible and self-compatible species. *Plant Cell*, 8(6), 943–58. DOI : 10.1105/tpc.8.6.943. [cit. p. 54]

157. Camadro, E. L. et Peloquin, S. J. (1981). Cross-incompatibility between two sympatric polyploid *Solanum* species. *Theor. Appl. Genet.*, 60(2), 65–70. DOI : 10.1007/BF00282417. [cit. p. 55]

158. Liedl, B. E. ; McCormick, S. et Mutschler, M. A. (1996). Unilateral incongruity in crosses involving *Lycopersicon pennellii* and *L. esculentum* is distinct from self-incompatibility in expression, timing and location. *Sex. Plant Reprod.*, 9(5), 299–308. DOI : 10.1007/BF02152705. [cit. p. 55]

159. Covey, P. A. ; Kondo, K. ; Welch, L. et coll. (2010). Multiple features that distinguish unilateral incongruity and self-incompatibility in the tomato clade. *Plant J.*, 64(3), 367–78. DOI : 10.1111/j.1365-313X.2010.04340.x. [cit. p. 55]

160. Li, W. et Chetelat, R. T. (2010). A pollen factor linking inter- and intraspecific pollen rejection in tomato. *Science*, 330(6012), 1827–30. DOI : 10.1126/science.1197908. [cit. p. 55]

161. Eberle, C. A. ; Anderson, N. O. ; Clasen, B. M. ; Hegeman, A. D. et Smith, A. G. (2013). PELPIII : the class III pistil-specific extensin-like *Nicotiana tabacum* proteins are essential for interspecific incompatibility. *Plant J.*, 74(5), 805–14. DOI : 10.1111/tpj.12163. [cit. p. 55]

162. Bedinger, P. A. ; Chetelat, R. T. ; McClure, B. et coll. (2011). Interspecific reproductive barriers in the tomato clade : opportunities to decipher mechanisms of reproductive isolation. *Sex. Plant Reprod.*, 24(3), 171–87. DOI : 10.1007/s00497-010-0155-7. [cit. p. 55]

163. Higashiyama, T. ; Kuroiwa, H. et Kuroiwa, T. (2003). Pollen-tube guidance : beacons from the female gametophyte. *Curr. Opin. Plant Biol.*, 6(1), 36–41. DOI : 10.1016/S1369-5266(02)00010-9. [cit. p. 55]

164. Higashiyama, T. et Hamamura, Y. (2008). Gametophytic pollen tube guidance. *Sex. Plant Reprod.*, 21(1), 17–26. DOI : 10.1007/s00497-007-0064-6. [cit. p. 55, 57]

165. Dresselhaus, T. et Franklin-Tong, N. (2013). Male-female crosstalk during pollen germination, tube growth and guidance, and double fertilization. *Mol. Plant*, 6(4), 1018–36. DOI : 10.1093/mp/sst061. [cit. p. 56]

166. Stewman, S. F. ; Jones-Rhoades, M. ; Bhimalapuram, P. et coll. (2010). Mechanistic insights from a quantitative analysis of pollen tube guidance. *BMC Plant Biol.*, 10, 32. DOI : 10.1186/1471-2229-10-32. [cit. p. 56]

167. Prado, A. M. ; Porterfield, D. M. et Feijó, J. A. (2004). Nitric oxide is involved in growth regulation and re-orientation of pollen tubes. *Development*, 131(11), 2707–14. DOI : 10.1242/dev.01153. [cit. p. 56]

168. Prado, A. M. ; Colaço, R. ; Moreno, N. ; Silva, A. C. et Feijó, J. A. (2008). Targeting of pollen tubes to ovules is dependent on nitric oxide (NO) signaling. *Mol. Plant*, 1(4), 703–14. DOI : 10.1093/mp/ssn034. [cit. p. 56]

169. Guo, F.-Q. ; Okamoto, M. et Crawford, N. M. (2003). Identification of a plant nitric oxide synthase gene involved in hormonal signaling. *Science*, 302(5642), 100–3. DOI : 10.1126/science.1086770. [cit. p. 56]

170. Besson-Bard, A. ; Pugin, A. et Wendehenne, D. (2008). New insights into nitric oxide signaling in plants. *Annu. Rev. Plant Biol.*, 59, 21–39. DOI : 10.1146/annurev.arplant.59.032607.092830. [cit. p. 56]

171. Chaubal, R. et Reger, B. (1990). Relatively high calcium is localized in synergid cells of wheat ovaries. *Sex. Plant Reprod.*, 3(2), 98–102. DOI : 10.1007/BF00198852. [cit. p. 56]

172. Chaubal, R. et Reger, B. (1993). Prepollination degeneration in mature synergids of pearl millet : an examination using antimonate fixation to localize calcium. *Sex. Plant Reprod.*, 6(4), 225–38. DOI : 10.1007/BF00231899. [cit. p. 56]

173. Huang, B. et Russell, S. (1992). Synergid degeneration in *Nicotiana* : a quantitative, fluorochromatic and chlorotetracycline study. *Sex. Plant Reprod.*, 5(2), 151–155. DOI : 10.1007/BF00194875. [cit. p. 56]

174. Jensen, W. A. (1965). The ultrastructure and histochemistry of the synergids of cotton. *Am. J. Bot.*, 52(3), 238–56. DOI : 10.1002/j.1537-2197.1965.tb06781.x. [cit. p. 56]

175. Tian, H.-Q. et Russell, S. D. (1997). Calcium distribution in fertilized and unfertilized ovules and embryo sacs of *Nicotiana tabacum* L. *Planta*, 202(1), 93–105. DOI : 10.1007/s004250050107. [cit. p. 56]

176. Tirlapur, U. (1993). Visualization of membrane calcium and calmodulin in embryo sacs *in situ* and isolated from *Petunia hybrida* L. and *Nicotiana tabacum* L. *Ann. Bot.*, 71(2), 161–7. DOI : 10.1006/anbo.1993.1020. [cit. p. 56]

177. Denninger, P. ; Bleckmann, A. ; Lausser, A. et coll. (2014). Male-female communication triggers calcium signatures during fertilization in *Arabidopsis*. *Nat. Commun.*, 5, 4645. DOI : 10.1038/ncomms5645. [cit. p. 56, 61]

178. Shi, Y.-Y. ; Tao, W.-J. ; Liang, S.-P. ; Lü, Y. et Zhang, L. (2009). Analysis of the tip-to-base gradient of CaM in pollen tube pulsant growth using *in vivo* CaM-GFP system. *Plant Cell Rep.*, 28(8), 1253–64. DOI : 10.1007/s00299-009-0725-z. [cit. p. 56]

179. Iwano, M. ; Ngo, Q. A. ; Entani, T. et coll. (2012). Cytoplasmic Ca^{2+} changes dynamically during the interaction of the pollen tube with synergid cells. *Development*, 139(22), 4202–9. DOI : 10.1242/dev.081208. [cit. p. 56, 61]

180. Fujitani, Y. ; Nakajima, N. ; Ishihara, K. et coll. (2006). Molecular and biochemical characterization of a serine racemase from *Arabidopsis thaliana*. *Phytochemistry*, 67(7), 668–74. DOI : 10.1016/j.phytochem.2006.01.003. [cit. p. 57]

181. Lu, Y. ; Chanroj, S. ; Zulkifli, L. et coll. (2011). Pollen tubes lacking a pair of K^+ transporters fail to target ovules in *Arabidopsis*. *Plant Cell*, 23(1), 81–93. DOI : 10.1105/tpc.110.080499. [cit. p. 57]

182. Guan, Y. ; Lu, J. ; Xu, J. ; McClure, B. et Zhang, S. (2014). Two mitogen-activated protein kinases, MPK3 and MPK6, are required for funicular guidance of pollen tubes in *Arabidopsis*. *Plant Physiol.*, 165(2), 528–33. DOI : 10.1104/pp.113.231274. [cit. p. 57]

183. Takeuchi, H. et Higashiyama, T. (2011). Attraction of tip-growing pollen tubes by the female gametophyte. *Curr. Opin. Plant Biol.*, 14(5), 614–21. DOI : 10.1016/j.pbi.2011.07.010. [cit. p. 57]

184. Márton, M. L. ; Cordts, S. ; Broadhvest, J. et Dresselhaus, T. (2005). Micropylar pollen tube guidance by Egg Apparatus 1 of maize. *Science*, 307(5709), 573–6. DOI : 10.1126/science.1104954. [cit. p. 57]

185. Krohn, N. G. ; Lausser, A. ; Juranić, M. et Dresselhaus, T. (2012). Egg cell signaling by the secreted peptide ZmEAL1 controls antipodal cell fate. *Dev. Cell*, 23(1), 219–25. DOI : 10.1016/j.devcel.2012.05.018. [cit. p. 57]

186. Uebler, S. ; Dresselhaus, T. et Márton, M. (2013). Species-specific interaction of EA1 with the maize pollen tube apex. *Plant Signal. Behav.*, 8(10), e25682. DOI : 10.4161/psb.25682. [cit. p. 58]

187. Márton, M. L. ; Fastner, A. ; Uebler, S. et Dresselhaus, T. (2012). Overcoming hybridization barriers by the secretion of the maize pollen tube attractant ZmEA1 from *Arabidopsis* ovules. *Curr. Biol.*, 22(13), 1194–8. DOI : 10.1016/j.cub.2012.04.061. [cit. p. 58]

188. Higashiyama, T. ; Yabe, S. ; Sasaki, N. et coll. (2001). Pollen tube attraction by the synergid cell. *Science*, 293(5534), 1480–3. DOI : 10.1126/science.1062429. [cit. p. 58]

189. Higashiyama, T. (2002). The synergid cell : attractor and acceptor of the pollen tube for double fertilization. *J. Plant Res.*, 115(1118), 149–60. DOI : 10.1007/s102650200020. [cit. p. 58]

190. Okuda, S. ; Tsutsui, H. ; Shiina, K. et coll. (2009). Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature*, 458(7236), 357–61. DOI : 10.1038/nature07882. [cit. p. 58]

191. Kasahara, R. D. ; Portereiko, M. F. ; Sandaklie-Nikolova, L. ; Rabiger, D. S. et Drews, G. N. (2005). *MYB98* is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell*, 17(11), 2981–92. DOI : 10.1105/tpc.105.034603. [cit. p. 58, 60]

192. Higashiyama, T. ; Inatsugi, R. ; Sakamoto, S. et coll. (2006). Species preferentiality of the pollen tube attractant derived from the synergid cell of *Torenia fournieri*. *Plant Physiol.*, 142(2), 481–91. DOI : 10.1104/pp.106.083832. [cit. p. 58]

193. Kanaoka, M. M. ; Kawano, N. ; Matsubara, Y. et coll. (2011). Identification and characterization of TcCRP1, a pollen tube attractant from *Torenia concolor*. *Ann. Bot.*, 108(4), 739–47. DOI : 10.1093/aob/mcr111. [cit. p. 58]

194. Takeuchi, H. et Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol.*, 10(12), e1001449. DOI : 10.1371/journal.pbio.1001449. [cit. p. 58, 59, 60, 61]

195. Chen, Y.-H. ; Li, H.-J. ; Shi, D.-Q. et coll. (2007). The central cell plays a critical role in pollen tube guidance in *Arabidopsis*. *Plant Cell*, 19(11), 3563–77. DOI : 10.1105/tpc.107.053967. [cit. p. 58, 61]

196. Li, H.-J. ; Zhu, S.-S. ; Zhang, M.-X. et coll. (2015). *Arabidopsis* CBP1 is a novel regulator of transcription initiation in central cell-mediated pollen tube guidance. *Plant Cell*, 27(10), 2880–93. DOI : 10.1105/tpc.15.00370. [cit. p. 58, 61]

197. Alandete-Saez, M. ; Ron, M. et McCormick, S. (2008). GEX3, expressed in the male gametophyte and in the egg cell of *Arabidopsis thaliana*, is essential for micropylar pollen tube guidance and plays a role during early embryogenesis. *Mol. Plant*, 1(4), 586–98. DOI : 10.1093/mp/ssn015. [cit. p. 58]

198. Silverstein, K. A. T. ; Graham, M. A. ; Paape, T. D. et VandenBosch, K. A. (2005). Genome organization of more than 300 defensin-like genes in *Arabidopsis*. *Plant Physiol.*, 138(2), 600–10. DOI : 10.1104/pp.105.060079. [cit. p. 59]

199. Zhang, X. ; Liu, W. ; Nagae, T. T. et coll. (2017). Structural basis for receptor recognition of pollen tube attraction peptides. *Nat. Commun.*, 8(1), 1331. DOI : 10.1038/s41467-017-01323-8. [cit. p. 60]

200. Wang, T. ; Liang, L. ; Xue, Y. et coll. (2016). A receptor heteromer mediates the male perception of female attractants in plants. *Nature*, 531(7593), 241–4. DOI : 10.1038/nature16975. [cit. p. 60]

201. Liu, J. ; Zhong, S. ; Guo, X. et coll. (2013). Membrane-bound RLCKs LIP1 and LIP2 are essential male factors controlling male-female attraction in *Arabidopsis*. *Curr. Biol.*, 23(11), 993–8. DOI : 10.1016/j.cub.2013.04.043. [cit. p. 60]

202. Hulskamp, M. ; Schneitz, K. et Pruitt, R. E. (1995). Genetic evidence for a long-range activity that directs pollen tube guidance in *Arabidopsis*. *Plant Cell*, 7(1), 57–64. DOI : 10.1105/tpc.7.1.57. [cit. p. 60]

203. Schneitz, K. ; Hülskamp, M. ; Kopczak, S. D. et Pruitt, R. E. (1997). Dissection of sexual organ ontogenesis : a genetic analysis of ovule development in *Arabidopsis thaliana*. *Development*, 124 (7), 1367–76. URL : http ://dev.biologists.org/content/124/7/1367. [cit. p. 60]

204. Elliott, R. C. ; Betzner, A. S. ; Huttner, E. et coll. (1996). *AINTEGUMENTA*, an *APETALA2*-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. *Plant Cell*, 8(2), 155–68. DOI : 10.1105/tpc.8.2.155. [cit. p. 60]

205. Hauser, B. A. ; Villanueva, J. M. et Gasser, C. S. (1998). *Arabidopsis* TSO1 regulates directional processes in cells during floral organogenesis. *Genetics*, 150(1), 411–23. URL : https ://www.genetics.org/content/150/1/411. [cit. p. 60]

206. Shimizu, K. K. et Okada, K. (2000). Attractive and repulsive interactions between female and male gametophytes in *Arabidopsis* pollen tube guidance. *Development*, 127(20), 4511–8. URL : http ://dev.biologists.org/content/127/20/4511. [cit. p. 60, 61, 66]

207. Punwani, J. A. ; Rabiger, D. S. et Drews, G. N. (2007). MYB98 positively regulates a battery of synergid-expressed genes encoding filiform apparatus localized proteins. *Plant Cell*, 19(8), 2557–68. DOI : 10.1105/tpc.107.052076. [cit. p. 60]

208. Punwani, J. A. ; Rabiger, D. S. ; Lloyd, A. et Drews, G. N. (2008). The *MYB98* subcircuit of the synergid gene regulatory network includes genes directly and indirectly regulated by MYB98. *Plant J.*, 55(3), 406–14. DOI : 10.1111/j.1365-313X.2008.03514.x. [cit. p. 60]

209. Shimizu, K. K. ; Ito, T. ; Ishiguro, S. et Okada, K. (2008). *MAA3* (*MAGATAMA3*) helicase gene is required for female gametophyte development and pollen tube guidance in *Arabidopsis thaliana*. *Plant Cell Physiol.*, 49(10), 1478–83. DOI : 10.1093/pcp/pcn130. [cit. p. 61]

210. Wang, H. ; Boavida, L. C. ; Ron, M. et McCormick, S. (2008). Truncation of a protein disulfide isomerase, PDIL2-1, delays embryo sac maturation and disrupts pollen tube guidance in *Arabidopsis thaliana*. *Plant Cell*, 20(12), 3300–11. DOI : 10.1105/tpc.108.062919. [cit. p. 61]

211. Li, H.-J. ; Xue, Y. ; Jia, D.-J. et coll. (2011). POD1 regulates pollen tube guidance in response to micropylar female signaling and acts in early embryo patterning in *Arabidopsis*. *Plant Cell*, 23(9), 3288–302. DOI : 10.1105/tpc.111.088914. [cit. p. 61]

212. Christensen, A. ; Svensson, K. ; Thelin, L. et coll. (2010). Higher plant calreticulins have acquired specialized functions in *Arabidopsis*. *PLoS One*, 5(6), e11342. DOI : 10.1371/journal.pone.0011342. [cit. p. 61]

213. Srilunchang, K.-O. ; Krohn, N. G. et Dresselhaus, T. (2010). DiSUMO-like DSUL is required for nuclei positioning, cell specification and viability during female gametophyte maturation in maize. *Development*, 137(2), 333–45. DOI : 10.1242/dev.035964. [cit. p. 61]

214. Leshem, Y. ; Johnson, C. et Sundaresan, V. (2013). Pollen tube entry into the synergid cell of *Arabidopsis* is observed at a site distinct from the filiform apparatus. *Plant Reprod.*, 26(2), 93–9. DOI : 10.1007/s00497-013-0211-1. [cit. p. 61]

215. Kessler, S. A. et Grossniklaus, U. (2011). She's the boss : signaling in pollen tube reception. *Curr. Opin. Plant Biol.*, 14(5), 622–7. DOI : 10.1016/j.pbi.2011.07.012. [cit. p. 61]

216. Higashiyama, T. ; Kuroiwa, H. ; Kawano, S. et Kuroiwa, T. (2000). Explosive discharge of pollen tube contents in *Torenia fournieri*. *Plant Physiol.*, 122(1), 11–4. DOI : 10.1104/pp.122.1.11. [cit. p. 61]

217. Sandaklie-Nikolova, L. ; Palanivelu, R. ; King, E. J. ; Copenhaver, G. P. et Drews, G. N. (2007). Synergid cell death in *Arabidopsis* is triggered following direct interaction with the pollen tube. *Plant Physiol.*, 144(4), 1753–62. DOI : 10.1104/pp.107.098236. [cit. p. 61]

218. Leydon, A. R. ; Tsukamoto, T. ; Dunatunga, D. et coll. (2015). Pollen tube discharge completes the process of synergid degeneration that is initiated by pollen tube-synergid interaction in *Arabidopsis*. *Plant Physiol.*, 169(1), 485–96. DOI : 10.1104/pp.15.00528. [cit. p. 62]

219. Williams, E. ; Kaul, V. ; Rouse, J. et Palser, B. (1986). Overgrowth of pollen tubes in embryo sacs of *Rhododendron* following interspecific pollinations. *Aust. J. Bot.*, 34(4), 413. DOI : 10.1071/BT9860413. [cit. p. 62]

220. Escobar-Restrepo, J.-M. ; Huck, N. ; Kessler, S. et coll. (2007). The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science*, 317(5838), 656–60. DOI : 10.1126/science.1143562. [cit. p. 62]

221. Lindner, H. ; Müller, L. M. ; Boisson-Dernier, A. et Grossniklaus, U. (2012). CrRLK1L receptor-like kinases : not just another brick in the wall. *Curr. Opin. Plant Biol.*, 15(6), 659–69. DOI : 10.1016/j.pbi.2012.07.003. [cit. p. 62]

222. Haruta, M. ; Sabat, G. ; Stecker, K. ; Minkoff, B. B. et Sussman, M. R. (2014). A peptide hormone and its receptor protein kinase regulate plant cell expansion. *Science*, 343(6169), 408–11. DOI : 10.1126/science.1244454. [cit. p. 62]

223. Olsen, A. N. ; Mundy, J. et Skriver, K. (2002). Peptomics, identification of novel cationic *Arabidopsis* peptides with conserved sequence motifs. *In Silico Biol.*, 2, 441–51. DOI : 10.1385/ABAB :120 :3 :169. [cit. p. 62]

224. Kessler, S. A. ; Shimosato-Asano, H. ; Keinath, N. F. et coll. (2010). Conserved molecular components for pollen tube reception and fungal invasion. *Science*, 330(6006), 968–71. DOI : 10.1126/science.1195211. [cit. p. 62]

225. Ngo, Q. A. ; Vogler, H. ; Lituiev, D. S. ; Nestorova, A. et Grossniklaus, U. (2014). A calcium dialog mediated by the *FERONIA* signal transduction pathway controls plant sperm delivery. *Dev. Cell*, 29(4), 491–500. DOI : 10.1016/j.devcel.2014.04.008. [cit. p. 62]

226. Lindner, H. ; Kessler, S. A. ; Müller, L. M. et coll. (2015). *TURAN* and *EVAN* mediate pollen tube reception in *Arabidopsis* synergids through protein glycosylation. *PLoS Biol.*, 13(4), e1002139. DOI : 10.1371/journal.pbio.1002139. [cit. p. 62]

227. Capron, A. ; Gourgues, M. ; Neiva, L. S. et coll. (2008). Maternal control of male-gamete delivery in *Arabidopsis* involves a putative GPI-anchored protein encoded by the *LORELEI* gene. *Plant Cell*, 20(11), 3038–49. DOI : 10.1105/tpc.108.061713. [cit. p. 63]

228. Li, C. ; Yeh, F.-L. ; Cheung, A. Y. et coll. (2015). Glycosylphosphatidylinositol-anchored proteins as chaperones and co-receptors for FERONIA receptor kinase signaling in *Arabidopsis*. *eLife*, 4, e06587. DOI : 10.7554/eLife.06587.001. [cit. p. 63]

229. Duan, Q. ; Kita, D. ; Johnson, E. A. et coll. (2014). Reactive oxygen species mediate pollen tube rupture to release sperm for fertilization in *Arabidopsis*. *Nat. Commun.*, 5, 3129. DOI : 10.1038/ncomms4129. [cit. p. 63]

230. Liu, X. ; Castro, C. ; Wang, Y. et coll. (2016). The role of LORELEI in pollen tube reception at the interface of the synergid cell and pollen tube requires the modified eight-cysteine motif and the receptor-like kinase FERONIA. *Plant Cell*, 28(5), 1035–52. DOI : 10.1105/tpc.15.00703. [cit. p. 63]

231. Hou, Y. ; Guo, X. ; Cyprys, P. et coll. (2016). Maternal ENODLs are required for pollen tube reception in *Arabidopsis*. *Curr. Biol.*, 26(17), 2343–50. DOI : 10.1016/j.cub.2016.06.053. [cit. p. 63]

232. Boisson-Dernier, A. ; Frietsch, S. ; Kim, T.-H. ; Dizon, M. B. et Schroeder, J. I. (2008). The peroxin loss-of-function mutation *abstinence by mutual consent* disrupts male-female gametophyte recognition. *Curr. Biol.*, 18(1), 63–8. DOI : 10.1016/j.cub.2007.11.067. [cit. p. 63]

233. Amien, S. ; Kliwer, I. ; Márton, M. L. et coll. (2010). Defensin-like ZmES4 mediates pollen tube burst in maize via opening of the potassium channel KZM1. *PLoS Biol.*, 8(6), e1000388. DOI : 10.1371/journal.pbio.1000388. [cit. p. 63]

234. Matias-Hernandez, L. ; Battaglia, R. ; Galbiati, F. et coll. (2010). *VERDANDI* is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in *Arabidopsis*. *Plant Cell*, 22(6), 1702–15. DOI : 10.1105/tpc.109.068627. [cit. p. 63]

235. Schiøtt, M. ; Romanowsky, S. M. ; Baekgaard, L. et coll. (2004). A plant plasma membrane Ca^{2+} pump is required for normal pollen tube growth and fertilization. *Proc. Natl. Acad. Sci. U. S. A.*, 101(25), 9502–7. DOI : 10.1073/pnas.0401542101. [cit. p. 63]

236. Liu, L. ; Zheng, C. ; Kuang, B. et coll. (2016). Receptor-like kinase RUPO interacts with potassium transporters to regulate pollen tube growth and integrity in rice. *PLoS Genet.*, 12(7), e1006085. DOI : 10.1371/journal.pgen.1006085. [cit. p. 64]

237. Zhu, L. ; Chu, L.-C. ; Liang, Y. et coll. (2018). The *Arabidopsis* CrRLK1L protein kinases BUPS1 and BUPS2 are required for normal growth of pollen tubes in the pistil. *Plant J.* DOI : 10.1111/tpj.13963. [cit. p. 64]

238. Ge, Z. ; Bergonci, T. ; Zhao, Y. et coll. (2017). *Arabidopsis* pollen tube integrity and sperm release are regulated by RALF-mediated signaling. *Science*, 358, 1596–1600. DOI : 10.1126/science.aao3642. [cit. p. 64]

239. Hamamura, Y. ; Saito, C. ; Awai, C. et coll. (2011). Live-cell imaging reveals the dynamics of two sperm cells during double fertilization in *Arabidopsis thaliana*. *Curr. Biol.*, 21(6), 497–502. DOI : 10.1016/j.cub.2011.02.013. [cit. p. 64]

240. Igawa, T. ; Yanagawa, Y. ; Miyagishima, S.-Y. et Mori, T. (2013). Analysis of gamete membrane dynamics during double fertilization of *Arabidopsis*. *J. Plant Res.*, 126(3), 387–94. DOI : 10.1007/s10265-012-0528-0. [cit. p. 64]

241. Sprunck, S. ; Rademacher, S. ; Vogler, F. et coll. (2012). Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science*, 338(6110), 1093–7. DOI : 10.1126/science.1223944. [cit. p. 64, 65]

242. Berger, F. ; Hamamura, Y. ; Ingouff, M. et Higashiyama, T. (2008). Double fertilization – caught in the act. *Trends Plant Sci.*, 13(8), 437–43. DOI : 10.1016/j.tplants.2008.05.011. [cit. p. 64]

243. Ingouff, M. ; Sakata, T. ; Li, J. et coll. (2009). The two male gametes share equal ability to fertilize the egg cell in *Arabidopsis thaliana*. *Curr. Biol.*, 19(1), R19–20. DOI : 10.1016/j.cub.2008.11.025. [cit. p. 64]

244. Liu, Y. ; Misamore, M. J. et Snell, W. J. (2010). Membrane fusion triggers rapid degradation of two gamete-specific, fusion-essential proteins in a membrane block to polygamy in *Chlamydomonas*. *Development*, 137(9), 1473–81. DOI : 10.1242/dev.044743. [cit. p. 64]

245. Mori, T. ; Igawa, T. ; Tamiya, G. ; Miyagishima, S.-Y. et Berger, F. (2014). Gamete attachment requires GEX2 for successful fertilization in *Arabidopsis*. *Curr. Biol.*, 24(2), 170–5. DOI : 10.1016/j.cub.2013.11.030. [cit. p. 64]

246. Misamore, M. J. ; Gupta, S. et Snell, W. J. (2003). The *Chlamydomonas* Fus1 protein is present on the mating type plus fusion organelle and required for a critical membrane adhesion event during fusion with *minus* gametes. *Mol. Biol. Cell*, 14(6), 2530–42. DOI : 10.1091/mbc.E02-12-0790. [cit. p. 64]

247. Inoue, N. ; Ikawa, M. ; Isotani, A. et Okabe, M. (2005). The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature*, 434(7030), 234–8. DOI : 10.1038/nature03362. [cit. p. 64, 65]

248. Ferris, P. J. ; Pavlovic, C. ; Fabry, S. et Goodenough, U. W. (1997). Rapid evolution of sex-related genes in *Chlamydomonas*. *Proc. Natl. Acad. Sci. U. S. A.*, 94(16), 8634–9. DOI : 10.1073/pnas.94.16.8634. [cit. p. 64]

249. Mori, T. ; Kuroiwa, H. ; Higashiyama, T. et Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat. Cell Biol.*, 8(1), 64–71. DOI : 10.1038/ncb1345. [cit. p. 64, 65]

250. von Besser, K. ; Frank, A. C. ; Johnson, M. A. et Preuss, D. (2006). *Arabidopsis HAP2* (*GCS1*) is a sperm-specific gene required for pollen tube guidance and fertilization. *Development*, 133(23), 4761–9. DOI : 10.1242/dev.02683. [cit. p. 64]

251. Wong, J. L. et Johnson, M. A. (2010). Is HAP2-GCS1 an ancestral gamete fusogen ? *Trends Cell Biol.*, 20(3), 134–41. DOI : 10.1016/j.tcb.2009.12.007. [cit. p. 65]

252. Fedry, J. ; Forcina, J. ; Legrand, P. et coll. (2018). Evolutionary diversification of the HAP2 membrane insertion motifs to drive gamete fusion across eukaryotes. *PLoS Biol.*, 16(8), e2006357. DOI : 10.1371/journal.pbio.2006357. [cit. p. 65]

253. Brownfield, L. ; Hafidh, S. ; Borg, M. et coll. (2009). A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genet.*, 5(3), e1000430. DOI : 10.1371/journal.pgen.1000430. [cit. p. 65]

254. Fédry, J. ; Liu, Y. ; Péhau-Arnaudet, G. et coll. (2017). The ancient gamete fusogen HAP2 is a eukaryotic class II fusion protein. *Cell*, 168(5), 904–915.e10. DOI : 10.1016/j.cell.2017.01.024. [cit. p. 65]

255. Leshem, Y. ; Johnson, C. ; Wuest, S. E. et coll. (2012). Molecular characterization of the *glauce* mutant : a central cell-specific function is required for double fertilization in *Arabidopsis*. *Plant Cell*, 24(8), 3264–77. DOI : 10.1105/tpc.112.096420. [cit. p. 65]

256. Yu, F. ; Shi, J. ; Zhou, J. et coll. (2010). ANK6, a mitochondrial ankyrin repeat protein, is required for male-female gamete recognition in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.*, 107 (51), 22332–7. DOI : 10.1073/pnas.1015911107. [cit. p. 65]

257. Bianchi, E. ; Doe, B. ; Goulding, D. et Wright, G. J. (2014). Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature*, 508(7497), 483–7. DOI : 10.1038/nature13203. [cit. p. 65]

258. Kaji, K. ; Oda, S. ; Shikano, T. et coll. (2000). The gamete fusion process is defective in eggs of Cd9-deficient mice. *Nat. Genet.*, 24(3), 279–82. DOI : 10.1038/73502. [cit. p. 65]

259. Le Naour, F. ; Rubinstein, E. ; Jasmin, C. ; Prenant, M. et Boucheix, C. (2000). Severely reduced female fertility in CD9-deficient mice. *Science*, 287(5451), 319–21. DOI : 10.1126/science.287.5451.319. [cit. p. 65]

260. Miyado, K. ; Yamada, G. ; Yamada, S. et coll. (2000). Requirement of CD9 on the egg plasma membrane for fertilization. *Science*, 287(5451), 321–4. DOI : 10.1126/science.287.5451.321. [cit. p. 65]

261. Boavida, L. C. ; Qin, P. ; Broz, M. ; Becker, J. D. et McCormick, S. (2013). *Arabidopsis* tetraspanins are confined to discrete expression domains and cell types in reproductive tissues and form homo- and heterodimers when expressed in yeast. *Plant Physiol.*, 163(2), 696–712. DOI : 10.1104/pp.113.216598. [cit. p. 65]

262. Spielman, M. et Scott, R. J. (2008). Polyspermy barriers in plants : from preventing to promoting fertilization. *Sex. Plant Reprod.*, 21(1), 53–65. DOI : 10.1007/s00497-007-0063-7. [cit. p. 65]

263. Nakel, T. ; Tekleyohans, D. G. ; Mao, Y. et coll. (2017). Triparental plants provide direct evidence for polyspermy induced polyploidy. *Nat. Commun.*, 8(1), 1033. DOI : 10.1038/s41467-017-01044-y. [cit. p. 66]

264. Grossniklaus, U. (2017). Polyspermy produces tri-parental seeds in maize. *Curr. Biol.*, 27(24), R1300–2. DOI : 10.1016/j.cub.2017.10.059. [cit. p. 66]

265. Huck, N. (2003). The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception. *Development*, 130(10), 2149–59. DOI : 10.1242/dev.00458. [cit. p. 66]

266. Beale, K. M. ; Leydon, A. R. et Johnson, M. A. (2012). Gamete fusion is required to block multiple pollen tubes from entering an *Arabidopsis* ovule. *Curr. Biol.*, 22(12), 1090–4. DOI : 10.1016/j.cub.2012.04.041. [cit. p. 66]

267. Wu, C.-C. ; Diggle, P. K. et Friedman, W. E. (2013). Kin recognition within a seed and the effect of genetic relatedness of an endosperm to its compatriot embryo on maize seed development. *Proc. Natl. Acad. Sci. U. S. A.*, 110(6), 2217–22. DOI : 10.1073/pnas.1220885110. [cit. p. 66]

268. Palanivelu, R. et Tsukamoto, T. (2012). Pathfinding in angiosperm reproduction : pollen tube guidance by pistils ensures successful double fertilization. *Wiley Interdiscip. Rev. : Dev. Biol.*, 1 (1), 96–113. DOI : 10.1002/wdev.6. [cit. p. 66]

269. Kasahara, R. D. ; Maruyama, D. ; Hamamura, Y. et coll. (2012). Fertilization recovery after defective sperm cell release in *Arabidopsis*. *Curr. Biol.*, 22(12), 1084–9. DOI : 10.1016/j.cub.2012.03.069. [cit. p. 66]

270. Völz, R. ; Heydlauff, J. ; Ripper, D. ; von Lyncker, L. et Groß-Hardt, R. (2013). Ethylene signaling is required for synergid degeneration and the establishment of a pollen tube block. *Dev. Cell*, 25 (3), 310–6. DOI : 10.1016/j.devcel.2013.04.001. [cit. p. 66]

271. Silverstein, K. A. T. ; Moskal, W. A. ; Wu, H. C. et coll. (2007). Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J.*, 51(2), 262–80. DOI : 10.1111/j.1365-313X.2007.03136.x. [cit. p. 66]

272. Marshall, E. ; Costa, L. M. et Gutierrez-Marcos, J. (2011). Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J. Exp. Bot.*, 62 (5), 1677–86. DOI : 10.1093/jxb/err002. [cit. p. 66]

273. Bircheneder, S. et Dresselhaus, T. (2016). Why cellular communication during plant reproduction is particularly mediated by CRP signalling. *J. Exp. Bot.*, 67(16), 4849–61. DOI : 10.1093/jxb/erw271. [cit. p. 66]

CHAPITRE 3

KAPPA : détection et *clustering* des protéines riches en cystéines

Ce chapitre a été publié le 31 janvier 2015 sous le titre *KAPPA, a simple algorithm for discovery and clustering of proteins defined by a key amino acid pattern : a case study of the cysteine-rich proteins* dans le périodique *Bioinformatics*, 31(11), 1716–23. Il est accessible en ligne à https://doi.org/10.1093/bioinformatics/btv047.

Auteurs Valentin Joly et Daniel P. Matton

Contributions Valentin Joly a réalisé l'ensemble de la recherche publiée dans cet article, ainsi que la rédaction du manuscrit. Daniel P. Matton a contribué à la conception du travail, et s'est chargé de sa supervision et de la révision critique du manuscrit.

Mention légale Le texte ci-après est reproduit et adapté avec l'aimable autorisation de Oxford University Press. Le texte original fait l'objet de la mention légale suivante :

> **This is a pre-copyedited, author-produced version of an article accepted for publication in *Bioinformatics* following peer review. The version of record : *Joly V and Matton DP (2015). KAPPA, a simple algorithm for discovery and clustering of proteins defined by a key amino acid pattern : a case study of the cysteine-rich proteins. Bioinformatics, 31(11), 1716–23. DOI : 10.1093/bioinformatics/btv047* is available online at : https://academic.oup.com/bioinformatics/article/31/11/1716/2365319.**

3.1 Résumé

Motivation Les protéines présentant un patron basé sur un acide aminé-clé jouent un rôle central dans l'échange de signaux entre les bactéries, les animaux et les plantes. Ce sont également des médiateurs importants de la communication intercellulaire au sein d'un même organisme. Leur description et leur caractérisation ouvrent la voie à une meilleure compréhension de la signalisation moléculaire chez un vaste éventail d'organismes, et à des applications possibles pour la recherche médicale et agronomique. Le patron évolutif constrasté de ces protéines rend difficile leur détection et leur regroupement par les méthodes classiques de recherche de séquences. Dans cet article, nous présentons KAPPA (*Key Aminoacid Pattern-based Protein Analyzer*), un nouveau programme multi-plateformes capable de détecter ces protéines dans un jeu de séquences donné, d'analyser leurs patrons et de les regrouper en les comparant à des patrons de référence (recherche *ab initio*) ou en les comparant entre eux (recherche *de novo*).

Résultats Dans cette étude, nous employons l'exemple concret des protéines riches en cystéines (*cysteine-rich proteins*, CRPs) pour démontrer que la similarité de deux patrons de cystéines peut être évaluée précisément et efficacement au moyen d'un nouvel outil quantitatif créé pour KAPPA : le κ-score. Nous montrons également le clair avantage offert par KAPPA par rapport aux autres outils de recherche de séquence pour la recherche *ab initio* de nouvelles CRP. Pour terminer, nous présentons les fonctionnalités de regroupement (*clustering*) et de sous-regroupement (*subclustering*) de KAPPA, qui permettent de générer rapidement des groupes de CRP cohérents sans avoir besoin de séquences de référence.

Disponibilité and implémentation Les exécutables KAPPA sont disponibles pour Linux, Windows et Mac OS à `http://kappa-sequence-search.sourceforge.net`.

3.2 Abstract

Motivation Proteins defined by a key amino acid pattern are key players in the exchange of signals between bacteria, animals and plants, as well as important mediators for cell–cell communication within a single organism. Their description and characterization open the way to a better knowledge of molecular signalling in a broad range of organisms, and to possible application in medical and agricultural research. The contrasted pattern of evolution in these proteins

makes it difficult to detect and cluster them with classical sequence-based search tools. Here, we introduce *Key Aminoacid Pattern-based Protein Analyzer* (KAPPA), a new multi-platform program to detect them in a given set of proteins, analyze their pattern and cluster them by comparison to reference patterns (*ab initio* search) or internal pairwise comparison (*de novo* search).

Results In this study, we use the concrete example of cysteine-rich proteins (CRPs) to show that the similarity of two cysteine patterns can be precisely and efficiently assessed by a quantitative tool created for KAPPA: the κ-score. We also demonstrate the clear advantage of KAPPA over other classical sequence search tools for *ab initio* search of new CRPs. Eventually, we present *de novo* clustering and subclustering functionalities that allow to rapidly generate consistent groups of CRPs without a seed reference.

Availability and implementation KAPPA executables are available for Linux, Windows and Mac OS at `http://kappa-sequence-search.sourceforge.net`.

3.3 Introduction

In recent years, novel types of proteins defined by a key amino acid pattern—often referred to as 'X-rich proteins'—have emerged as important and diversified actors of molecular signalling in animals and plants. Although glycine-rich proteins,[1] proline-rich peptides[2] and leucine-rich repeats- containing proteins[3] are nowadays the subject of increasing research, cysteine-rich proteins (CRPs) still remain the most extensively studied ones.[4,5,6,7]

Although traditional sequence search tools were successfully used to characterize the diversity of CRPs in various organisms, implementation of numerous manual data curation steps was always necessary, with the drawback of being both tedious and time-consuming. Moreover, manual intervention always introduces the risk of a subjective, hence, potentially skewed or biased analysis. In this article, we introduce *Key Aminoacid Pattern-based Protein Analyzer* (KAPPA), a new automated sequence search program dedicated to the discovery and clustering of 'X-rich proteins', and we assess its performance on plant CRPs.

Although quite heterogeneous, CRPs share three common features: (i) a small size (50–200 amino acids); (ii) the presence of a signal peptide at their N-termini to allow secretion; (iii) a mature protein comprising six or more cysteines (usually from 6 to 14). They are involved in a

wide range of functions in living organisms. For instance, antimicrobial peptides (AMPs) such as defensins are broadly studied for their role in human innate immunity against viruses, bacteria and fungi,[8] and in plant responses to pathogens.[9] Plant–bacteria symbiotic equilibrium relies on CRPs as well,[10] while venoms from snakes, spiders and scorpions have also adopted the CRPs as neurotoxins and myonecrotic agents.[11] Thus, characterization of these communication mediators opens the way to future applications for human health and agronomy.[12]

In addition to their inter-organism communication functions, CRPs appear to be important messengers for cell–cell signalling within a single individual. They are particularly involved in the control of developmental processes, such as seed, root and stomata development,[13] but also in sexual reproduction in plants[14,15,16] and in animals.[17]

Numerous families of CRPs like defensins, thionins, albumins, snakins, lipid-transfer proteins, rapid alkalinization factors (RALFs), etc. have already been described, each of them being characterized by a precise cysteine spacing pattern. In several plant species, these CRPs can represent up to 2–3 % of the total genome.[18] Describing their diversity and evolution now becomes a major stake in plant, animal and microbe biology.

The structural particularities of CRPs are a challenging issue for bioinformaticians: on the one hand, the cysteine backbone—which governs maintenance of disulfide bonds, hence, the 3D structure of proteins—is highly conserved in a given CRP family, even between distant species. For instance, all defensins share a γ-core and a cysteine-stabilized (CS) $\alpha\beta$ motif.[19] On the other hand, the remaining residues in the sequence—involved in fine and specific recognition functions—can exhibit a fast evolutive speed, often underlined by a positive selection. Hence, global sequence identity can be extremely low in a given CRP family (Figure S3.1, page 225). As shown in our study, this dual evolutive pattern makes it difficult to discover paralogues and orthologues of already known CRPs with classical *ab initio* sequence similarity search tools.

In addition, the cysteine backbone itself can evolve—though at a lesser speed—and give rise to new families of CRPs that may be specific to a given taxon. Literature suggests that CRPs have been largely under-predicted;[18] it is therefore essential to be able to detect and cluster CRPs *de novo*, without necessarily relying on a set of reference proteins.

The KAPPA workflow, presented in Figure S3.2 (page 227), meets this sequence search challenge by extracting and comparing cysteine patterns by means of a quantitative similarity index called κ-score. This mapping step allows to detect CRPs that are similar to reference patterns

(*ab initio* search) or to cluster CRPs having similar patterns without relying on a reference (*de novo* search). A BLAST-based subclustering step then allows to refine clustering, analyzing the remaining amino acids of the sequence and to visualize output groups of CRPs graphically.

KAPPA is a free program coded in Python 3 and is executable on most of existing operating systems (Linux, Mac OS, Windows). It supports sequence search parallelization by multithreading. UNIX manual pages and a complete user's guide are also available.

3.4 Methods

3.4.1 Cysteine pattern extraction

The first step consists in providing KAPPA with one or several FASTA files containing protein sequences. Several scripts bundled with KAPPA can be used to detect open reading frames (ORFs) in a set of nucleotide sequences (`kappa_findorfs`), translate them into proteins (`kappa_translate`) and predict secretion (`kappa_secretion`). The latter program relies on SignalP[20] and SecretomeP.[21] After sequence import and FASTA identifier parsing, KAPPA prefilters proteins to retain only those susceptible to be true CRPs, according to amino acid length (options `-l` and `-L`) and number of cysteines (options `-m` and `-M`). The user then obtain a list of *target proteins*.

For each of them, KAPPA analyzes cysteine spacing and creates a *pattern*, a 1D vector in which each value corresponds to the number of amino acids in a sequence block between two cysteines. The first block corresponds to amino acids before the first cysteine, the last block to those located after the last cysteine. Two consecutive cysteines define an empty block, represented by a zero in the pattern. *Pattern length* describes the number of blocks in a pattern. For example, sequence `XCXXXCCXCXX` gives pattern $\{1, 3, 0, 1, 2\}$, whose length is 5.

The mapping step then consists in comparing each target pattern to reference patterns called *query patterns*. If an *ab initio* search is performed, the user must provide query patterns in a separate text file (option `-q`). The user can write this file directly or build it from a set of reference proteins with the `kappa_extract_patterns` script. Unlike target patterns, each position in a query pattern comprises two values: a minimum and a maximum number of amino acids. These values can be equal or replaced by letter *n* to define an unknown number of amino acids.

Alternatively, if a *de novo* search is chosen, no reference is required, since an all-versus-all pairwise comparison of patterns will be performed: query patterns are simply target patterns themselves.

3.4.2 Mapping

To determine the level of similarity between a query and a target pattern, KAPPA aligns them in all possible configurations, i.e. considering all possible shifts between them. For each alignment, a κ-score reflecting pattern homology is computed. The maximum κ-score obtained among all possible alignments between the two patterns is retained as the final score.

The κ-score incorporates three indicators varying between 0 and 1 called *block identity*, *key persistance* and *alignment coverage*, which describe different aspects of pattern similarity. Figure 3.1 presents an example of κ-score calculation for two given patterns.

Block identity I aims at describing conservation of the number of aminoacids between query and target aligned blocks. To begin, KAPPA computes local query-target variation δ_k at each aligned position k, as shown in Equation 3.1 where Q_k^- and Q_k^+ are the minimum and maximum values in the query pattern and T_k the value in the target pattern at position k.

$$\delta_k = \begin{cases} 0 & \text{if} \quad Q_k^- \leq T_k \leq Q_k^+ \\ \frac{Q_k^- - T_k}{Q_k^- + T_k} & \text{if} \quad T_k < Q_k^- \\ \frac{T_k - Q_k^+}{T_k + Q_k^+} & \text{if} \quad T_k > Q_k^+ \end{cases} \tag{3.1}$$

Local identity at position k is the difference $1 - \delta_k$. Before being incorporated in the calculation of global identity I, local identity values are adjusted with the normalized logisitic function $\hat{f}$ described in Equation 3.2.

$$\hat{f}_\alpha(x) = \frac{f_\alpha(x) - f_\alpha(0)}{f_\alpha(1) - f_\alpha(0)} \quad \text{with} \quad f_\alpha(x) = \frac{1}{1 + e^{2(\alpha - 10x)}} \tag{3.2}$$

This step allows penalizing low values and favouring high values in a more or less restrictive manner, depending on the value attributed to the user-defined stringency parameter α. The user can thus choose to be stringent (i.e. considering only perfect identities) or more permissive. Stringency values are comprised between 0 (low) and 10 (high). The default is 7.0

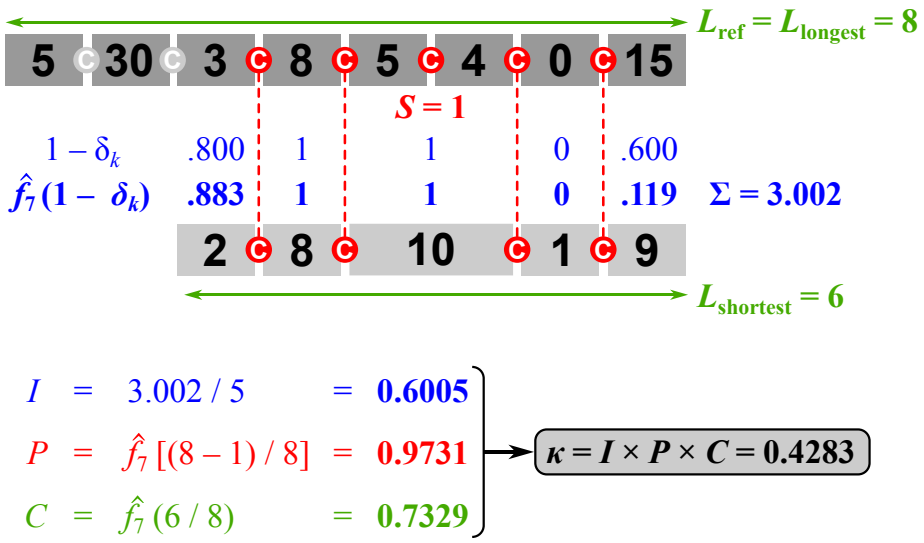


Figure 3.1. Example of κ-score calculation between two cysteine patterns, with default strigency parameters ($\alpha_I = \alpha_P = \alpha_C = 7$) and one loss or gain of cysteine allowed.

Identity I is simply computed as shown in Equation 3.3, where L corresponds to the number of positions in the alignment. Stringency on identity α_I can be set with option `-I`.

$$I = \frac{1}{L} \sum_{k=1}^{L} \hat{f}_{\alpha_I}(1 - \delta_k) \tag{3.3}$$

Because residues located before the first cysteine or after the last one are generally excluded from cysteine pattern analyses, KAPPA provides options `-n` and `-c` to define a number of N- and C-term blocks to be ignored in the calculation of identity. The default value for both of these options is 1.

Key persistence P accounts for maintenance of the cysteine content itself. Indeed, despite high conservation of cysteine backbones within a given CRP family, gain or loss of cysteine can occur marginally in the course of evolution. Options `-kg` and `-kl` can then be used to specify a number of X-to-C or C-to-X substitutions allowed in the query pattern, respectively. If one or more substitutions need to be simulated by KAPPA to achieve better identity, persistence P will decrease, as shown in Equation 3.4, where S is the number of substitutions simulated, and L is the length of the pattern where substitutions occurred. As well as identity, persistence is finally adjusted with the $\hat{f}$ function. The stringency α_P applied to persistence adjustment through $\hat{f}$ can be set with option `-P`.

$$P = \hat{f}_{\alpha_P}\left(\frac{L-S}{L}\right) \tag{3.4}$$

Finally, alignment coverage C describes how wide the alignment is compared to a reference number of blocks L_{ref}. By default, L_{ref} is the size of the longest pattern, so that coverage is maximized only when two patterns with similar size are fully aligned. However, one may want to allow the query pattern to be fully included in the targeted one without making coverage decrease. In this case, option `-i` can be enforced. L_{ref} will then be the size of the query pattern (*ab initio* search) or the one of the shortest pattern (*de novo* search). Raw coverage is adjusted with the $\hat{f}$ function, taking in account a specific stringency level α_C set with option `-C`.

$$C = \hat{f}_{\alpha_C}\left(\frac{L}{L_{\text{ref}}}\right) \tag{3.5}$$

The κ-score is eventually computed as the product of I, P and C. In case of an *ab initio* search, all sequences matching to a given reference cysteine pattern with a κ-score greater than a threshold defined with option `-S` will be assigned to it and exported in a FASTA file.

3.4.3 Clustering

If a *de novo* search is performed, a two-step clustering immediately follows mapping. Figure S3.3 (page 230) gives an overview of the clustering process. First, preclusters are formed recursively by connecting all sequences pairs having a κ-score above a threshold defined with option `-S1`. Second, similarity is assessed for all possible pairs of sequences belonging to two different preclusters. For two given preclusters, if more than a certain percentage of these inter-precluster sequence pairs (defined with option `-F`) have a κ-score above a certain threshold (specified with option `-S2`), the two preclusters are fused into one final cluster. Final clusters can result from an unfused precluster as well as from the fusion of two or more preclusters. They are exported in separate FASTA files.

Clusters can be seen as similarity networks. It is then interesting to assess their density (i.e. the number of links between sequences) and compactness (i.e. the level of similarity between sequences). Statistical indicators are computed taking in account all sequence–sequence similarities represented in the cluster, but also more specifically focusing on intra- and inter-sample similarities only. In addition, option `-G` can be enforced to generate network files that

can be imported into Cytoscape[22] to graphically visualize clusters.

3.4.4 Subclustering

Mapping and clustering steps allow assembling families of proteins sharing similar cysteine backbones. However, the remaining elements of the sequence are also of interest to refine this clustering. Indeed, CRPs also contain motifs or domains conserved on a more or less large scale that can be used to define subgroups within clusters. Users may want to split large families of CRPs into smaller, motif-defined subclusters; or to analyse paralogy and orthology relationships within a cluster.

The subclustering step consists in using BLASTp to make pairwise comparisons of all sequences within each group, relying on a network approach that was first developed in clustering tools such as EGN.[23] Two sequences are grouped into the same subcluster, if BLAST characteristic values pass a user-defined threshold: e-value, percentage of identity, percentage of positive matches, hit length, alignment coverage (options `-sE`, `-sI/-sJ`, `-sP/-sQ`, `-sL` and `-sC`, respectively). Moreover, the user can use option `-sR` to enforce a reciprocity condition: two sequences will be considered similar if they are reciprocal best or near-best hits. Here again, several density and compactness indicators are computed for each subcluster and graphical visualization with Cytoscape is possible through option `-G`.

3.4.5 Optimization

Several options affecting mapping and clustering granularity are to be set by the user. Therefore, finding the optimal combination of parameters to fit biological reality is a crucial issue. Two kinds of pitfalls especially need attention: (i) the 'snowball effect' taking place when most of the proteins are clustered into the same big group because of too permissive settings and (ii) the accumulation of singletons due to too stringent parameters.

To overcome these problems, the `kappa` script allows the user to provide several values for each option instead of one. The script will then execute KAPPA on the input target proteins, considering all possible combinations of parameters among those provided by the user. The output is a large table presenting, for each combination tested, a broad range of statistical indicators the user can explore to determine the most suitable settings.

3.4.6 Performance tests

Reference proteins In experiments described in Sections 3.5.1 and 3.5.2, we used reference small families of CRPs from *Arabidopsis* and rice as query proteins. LTP1s (subfamily 1 of non-specific lipid-transfer proteins) were retrieved from Edstam *et al.* [4]. Three OsLTP1s that were obviously not LTP1s in terms of cysteine pattern were removed from the dataset (Os11g02330.1, Os11g02379.1 and Os12g02290.1), because they would have skewed the analysis. True defensins and snakins come from groups CRP0000 and CRP2700 published by Silverstein *et al.* [18], respectively.

Expected outputs In Section 3.5.2, we assessed the quality of outputs from several sequence-based search tools by comparing them to 'expected outputs', i.e. known lipid-transfer proteins (LTPs) and defensin-like proteins (DEFLs). Reference LTPs correspond to those described in *Arabidopsis* and rice by Edstam *et al.* [4] (all subfamilies described in the study) and Silverstein *et al.* [18] (groups CRP3800–4958). Reference DEFLs are those described in Silverstein *et al.* [18] only (groups CRP0000–1520).

Target proteomes We used the *Arabidopsis thaliana* [24] and *Oryza sativa* subsp. *japonica* [25] proteomes available on line from Phytozome v9.1 (`http://www.phytozome.net`) as target sets of proteins for assessment of KAPPA performance in Section 3.5.

Dataset calibration Studies from which reference proteins were retrieved used older releases of the *Arabidopsis* and rice proteomes. To ensure consistency between our datasets, we did not use reference proteins directly, but the corresponding sequences from the Phytozome proteomes mentioned earlier. In a small minority of cases, no homologue—or only a distant one—was found, since some proteins turned to be obsolete or because they corresponded to pseudogenes rather than true proteins. We simply discarded them. This only affected 'expected outputs', not query proteins.

Software In Section 3.5.1, MUSCLE [26] was used for pairwise and multiple protein alignments; identities were computed on the hit with a homemade script. In Section 3.5.2, we used HMMER 3.0, [27] BLASTp, [28] PSI-BLAST and PHI-BLAST (position-specific iterated and pattern hit-initiated BLAST, respectively) [29] from the BLAST+ 2.2.29 suite. All programs were used with default parameters, unless otherwise specified.

3.5 Results

3.5.1 Relevance of the κ-score

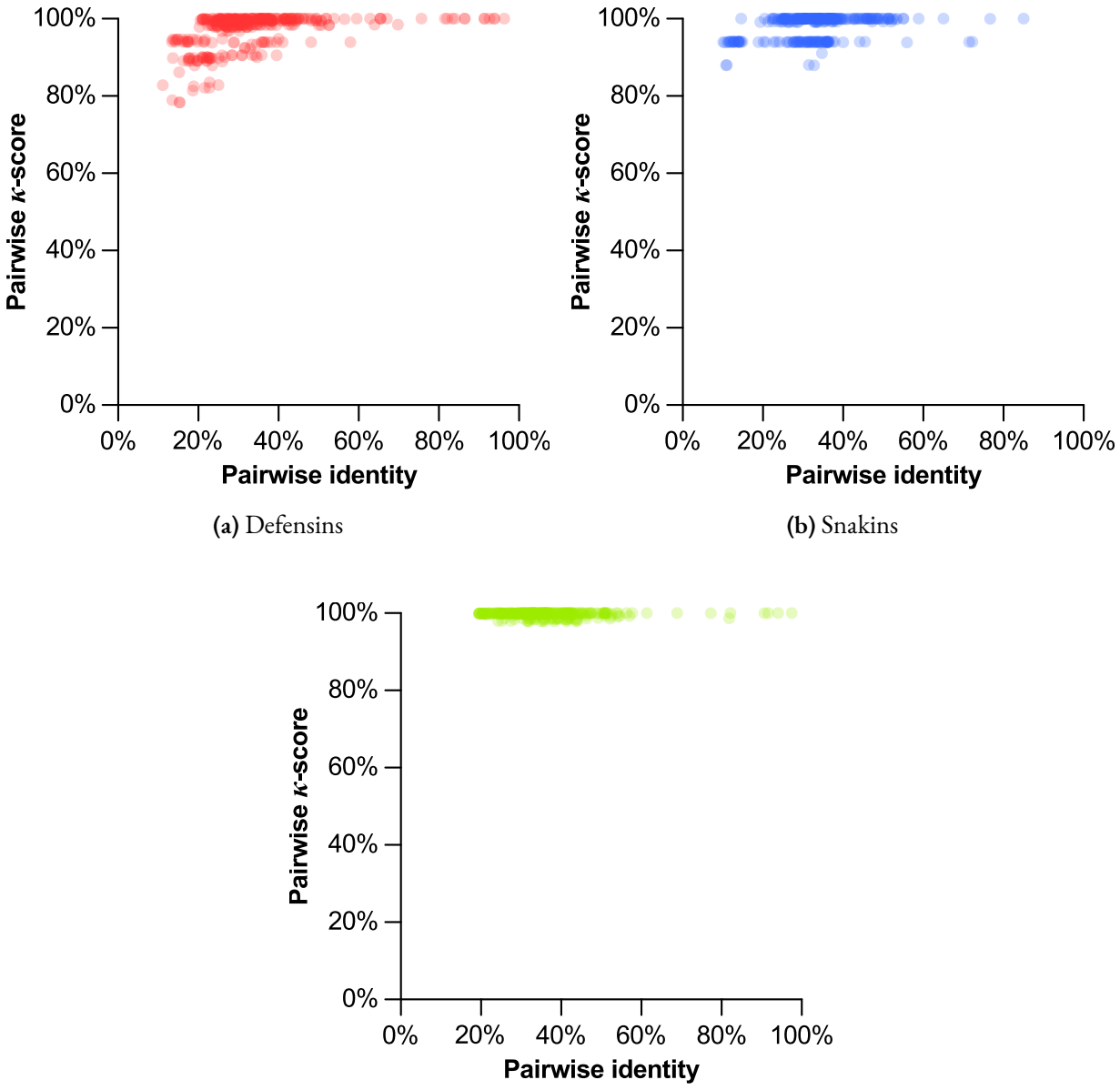
We first tried to check if the κ-score is relevant to describe conservation of the cysteine backbone within two well-known CRP families: true defensins and snakins described by Silverstein *et al.* [18] and the LTP1 subfamily described by Edstam *et al.* [4]. In all cases, we analyzed together proteins from *A. thaliana* and *O. sativa*. Multiple alignments of these proteins can be found in Figure S3.1 (page 225). Within each family, proteins were pairwise aligned with MUSCLE and identity on the alignment was computed. Besides, KAPPA was used to determine the κ-score for each pair of proteins.

As can be seen in Figure 3.2, pairwise identity can vary to a large extent while the κ-score always remains high within a given CRP family. This holds especially true for LTP1s, which have conserved cysteine spacing. However, defensins include proteins with a more distant pattern, which led to a slight decrease of the κ-score for a few data points. The general conclusion of this test is that the κ-score is a simple but reliable quantitative tool to assess conservation of the cysteine spacing.

3.5.2 Ab initio discovery of CRPs

KAPPA provides an *ab initio* sequence search function, consisting in detecting proteins matching to a given reference cysteine pattern. We compared performance of KAPPA, HMMER, BLASTp, PSI-BLAST and PHI-BLAST in detecting known LTPs and LTP-like proteins in the *Arabidopsis* proteome (Figures 3.3 and S3.4, page 231). KAPPA was provided with a consensus cysteine pattern made with `kappa_extract_patterns` after alignment of AtLTP1s with MUSCLE (Table S3.1, page 256). HMMER was provided with a position-specific scoring matrix (PSSM) made with hmmbuild using the same alignment. Programs from the BLAST+ suite were given AtLTP1s as query proteins. In addition, a perfect consensus cysteine pattern of AtLTP1s was given to PHI-BLAST.

As we can see on Figure 3.3, KAPPA was sensitive because the vast majority ($\sim$98 %) of the 128 AtLTPs described by Silverstein *et al.* [18] and Edstam *et al.* [4] were detected. Only three known AtLTPs were not detected by KAPPA: AT1G05450.1, AT3G63095.1 and AT5G38197.1. The first one was not found simply because it does not contain any cysteine



(a) Defensins

(b) Snakins

(c) LTP1s

Figure 3.2. Comparison of all-versus-all pairwise identities and κ-scores within three reference CRP families. KAPPA was used with default parameters, except for defensins for which one gain or loss of cysteine was allowed.

residue. The two latter ones do have cysteines, but their spacing is different from the canonical LTP pattern. Therefore, their κ-score was low and they were discarded by KAPPA. One may conclude that these three proteins were inappropriately described as LTPs in previous studies.

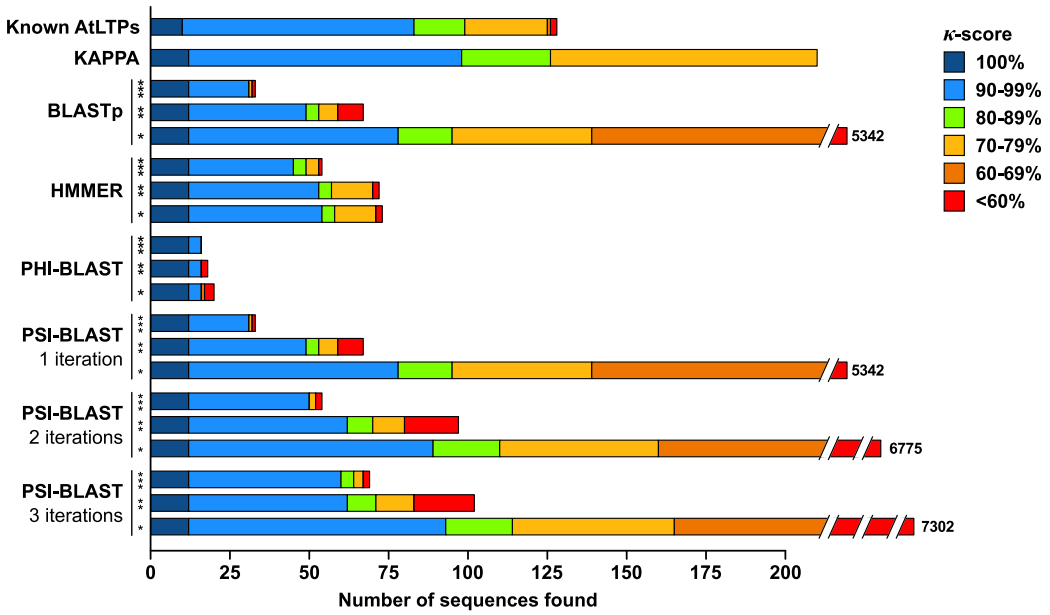
Interestingly, KAPPA also detected 85 new LTP-like proteins. This is mainly due to a better performance with respect to sequence search strategies used in previous studies, but also to the use of a more recent release of the *Arabidopsis* proteome (14 out of the 85 new LTPs were not present in The Arabidopsis Information Resource (TAIR6) database used by Silverstein *et al.*). Sequence identifiers of proteins detected by KAPPA are listed in Table S3.2 (page 257).

Contrary to other methods, KAPPA also achieved high specificity because the κ-score made it possible to reject non-related sequences efficiently. Here, we have used a 70 % threshold, which appeared to be an adequate value to work with divergent CRPs such as LTPs (Figure S3.5, page 233). Not only KAPPA provided the best sensitivity-specificity trade-off, but its execution was also $\sim$9 times faster than the second best method, PSI-BLAST with 3 iterations (Figure S3.6, page 234).

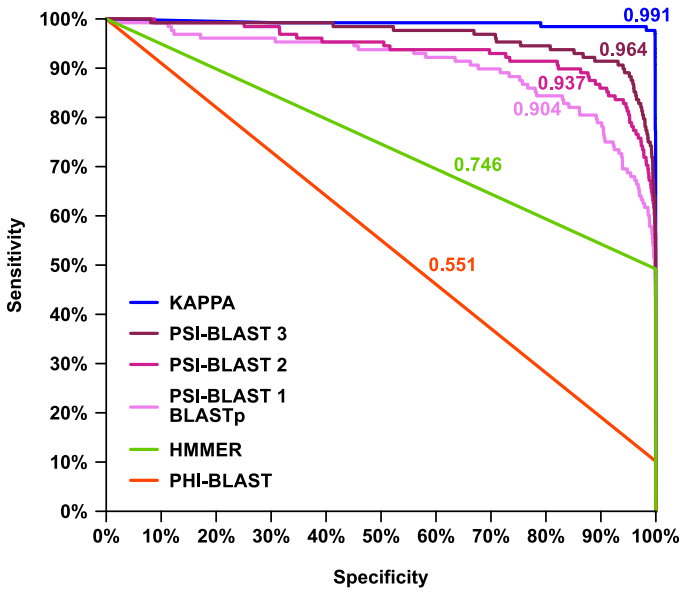
BLASTp and PSI-BLAST could easily find the LTP-like proteins presenting a high sequence identity to input AtLTP1s. However, more dissimilar LTP-like proteins could not be retrieved unless the stringency level was decreased. In this situation, a high number of non-related proteins accumulated in the output since they have the same low sequence similarity to AtLTP1s. Figure S3.4c (page 231) shows that the κ-score correctly discriminated LTP-like proteins from other sequences, including LTP-like with a low sequence identity, whereas the BLASTp e-value did not make a difference. Figures S3.4d–g (pages 232 to 232) show this was the same for outputs from other programs.

PHI-BLAST and HMMER appeared to be way more efficient in detecting target LTP-like CRPs without accumulating non-related proteins. However, they were also much more restrictive. Although they were able to give more weight to cysteine residues in the similarity search, both tools seemed unable to deal with variation in cysteine number and spacing. In contrast, KAPPA took advantage of a more or less stringent calculation of block identity I and key residue persistence P to address this issue.

Another problem may be due to the fact we used proteins stemming from the same LTP subfamily and from one single species. In this situation, residues that are common to all input proteins can not only be cysteine characteristic of all LTPs, but also other residues that are specific to AtLTP1s only. Although the first ones are expected to be highly conserved in other



(a) Comparison of outputs from different programs. AtLTP1s were used as input sequences to parse the *A. thaliana* proteome available in Phytozome 9.1. KAPPA was used with parameters specified in Table S3.1 (page 256). Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of AtLTP1s. Known AtLTPs described by Silverstein *et al.* [18] and Edstam *et al.* [4] were used to define a minimal expected output.



(b) Sensitivity-specificity ROC plot comparing performances of KAPPA and other programs, using known AtLTPs as reference true sequences. Sensitivity refers to the true positive rate while specificity corresponds to the true negative rate. Values indicated on the graph correspond to the area under the curve.

Figure 3.3. Assessment of *ab initio* sequence search performance for KAPPA and other programs.

LTP subfamilies and in other species, the latter may have changed more rapidly in the course of evolution. Yet, the algorithms to which we compared KAPPA gave the same weight to all of these residues, which can limit their performance in detecting more distant LTP-like proteins.

KAPPA's efficiency in *ab initio* sequence search was also demonstrated taking other examples. We first looked for AtLTP1s in rice (Figure S3.7, page 235), and then turned to study other families of plant CRPs described by Silverstein *et al.* [18]: defensins (Figures S3.8 and S3.9, pages 237 to 239) and snakins (Figures S3.10 and S3.11, pages 241 to 243). Finally, KAPPA was tested with other key residues: 2008 human and 1917 mouse proteins containing the extended glycine zipper (EGZ) motif described by Kim *et al.* [30] were discovered (Figures S3.12 and S3.13, pages 245 to 247). Moreover, KAPPA could also efficiently find all members of the small proline-rich protein 2 (SPRR2) family[31] in human and mouse proteomes (Figures S3.14 and S3.15, pages 249 to 251).

As shown in Figure S3.16 (page 253), BLASTp, PSI-BLAST and HMMER could perform as well as KAPPA when studying protein families displaying high sequence identity (e.g. snakins, SPRR2s). PHI-BLAST also efficiently detected target proteins when the key residue spacing is highly conserved (e.g. snakins, EGZs). However, KAPPA's performance was equal or superior to other programs in all cases, combining high sensitivity and specificity. KAPPA thus appears to be a reliable, all-purpose tool to detect any type of protein displaying a key amino acid pattern, with a clear advantage over existing methods when dealing with families displaying extensive sequence divergence or small variations in the key residue spacing.

3.5.3 De novo clustering of CRPs

Once potential CRPs are detected by the pre-filtering function or by a reference-guided *ab initio* search, it may be of interest to split them into clusters defined by precise cysteine spacing. Indeed, Edstam *et al.* [4] described 79 LTPs in *A. thaliana* and divided them into nine clusters and three singletons. Likewise, Silverstein *et al.* [18] found 131 new LTPs fragmented into 23 clusters and 8 singletons. In both of these studies, clustering is performed in a more or less arbitrary way, taking into account not only pattern but also sequence similarity. Thus, this experimental procedure can be long, tedious and partially subjective.

KAPPA introduces the possibility of clustering proteins (i) *de novo*, i.e. without relying on a set of reference patterns; (ii) in an automated fashion and (iii) using clear quantitative criteria. As a first step, the κ-score can be used to perform a pattern-based clustering of proteins. Second,

the remaining sequences can be used to launch a subclustering step based on classical sequence similarity values (e.g. percentage of identities and positives).

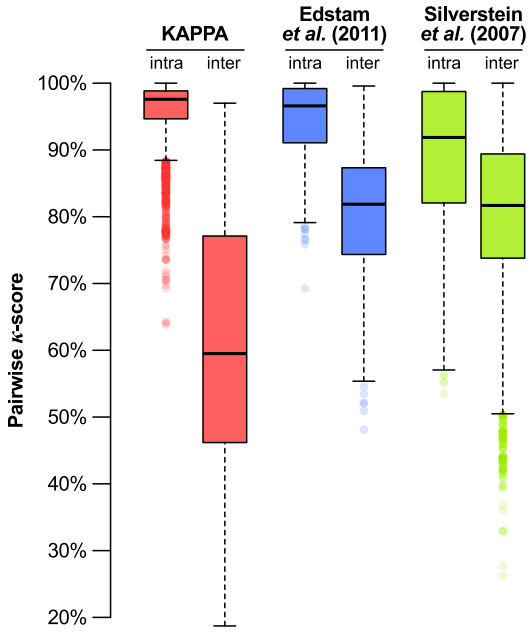
Because KAPPA detected 210 LTPs in the previous section, we tested its ability to divide them into clusters that would be consistent (high intra-group similarity) and distinct from each other (lower inter-group similarity). The *de novo* clustering function was used as follows: first, all pairs of proteins connected with a κ-score $\sim$97 % were recursively bound to form preclusters, which were fused to form final clusters if $\geq$90 % of inter-group sequence pairs had a κ-score above 95 %. These settings, chosen with the KAPPA optimization function, yielded 21 LTP clusters and 46 singletons.

Figure 3.4a compares the structure of AtLTPs clusters formed by KAPPA, Edstam *et al.* [4] and Silverstein *et al.* [18]. The stringent criteria we used to build KAPPA clusters led to a high intra-cluster pairwise κ-score with a narrow distribution that is clearly distinct from a low, wide distribution of inter-cluster κ-scores. Cluster homogeneity is still clear for Edstam's clusters, but starts to be questionable for Silverstein's clusters, because a notable proportion of intra-cluster pattern similarities are lower than inter-cluster similarities.

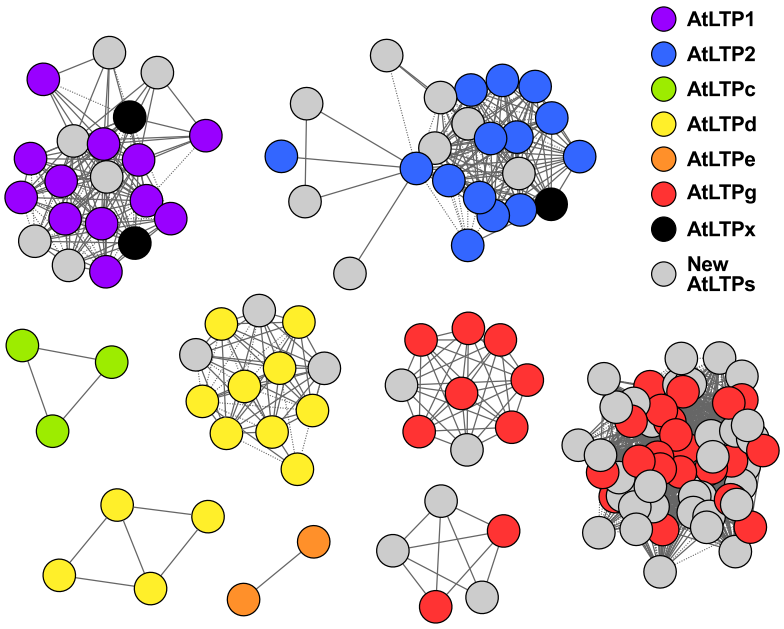
Figure 3.4b and Table S3.3 (page 267) shed light on the relative homogeneity between clusters generated automatically by KAPPA and manually by Edstam *et al.*: KAPPA perfectly reformed clusters AtLTP1, AtLTP2, AtLTPc and AtLTPe and extended the two first with new proteins (in grey). Clusters AtLTPd and AtLTPg were fractioned into 2 and 3 KAPPA clusters, respectively. Attention paid to protein alignments (Figure S3.17, page 254) clearly shows this separation is supported by differences in the cysteine spacing. Moreover, 3 of the 4 Edstam's singletons (AtLTPx, in black) could now be assigned to clusters.

The nine KAPPA clusters represented in Figure 3.4b gathered 136 proteins (65 % of the total) and contained on average 15.1 proteins per cluster. These proteins are quite close to the original AtLTP1s that were used as queries for the *ab initio* search, with a mean κ-score of 91.0 % with respect to the AtLTP1 cysteine pattern.

The remaining 74 proteins were, however, distributed into 12 smaller clusters (2.2 proteins per cluster) and 48 singletons. Their isolated status probably reflects their more distant nature as LTP-like proteins. Indeed, they all correspond to LTPs that were not described by Edstam *et al.* (except one AtLTPg) and their mean κ-score with respect to the AtLTP1 pattern was much lower (75.7 %).



(a) Distributions of κ-scores for all possible pairs of sequences belonging to the same cluster (intra) or to two different clusters (inter) in AtLTPs clusters generated by KAPPA (210 LTPs, 21 clusters and 46 singletons), Edstam *et al.* (79 LTPs, 6 clusters and 3 singletons) and Silverstein *et al.* (131 LTPs, 23 clusters and 8 singletons).



(b) Graphical view of the 9 KAPPA clusters containing at least one of the 79 LTPs from Edstam *et al.* [4] generated with Cytoscape. Coloured nodes correspond to LTPs described by KAPPA and Edstam *et al.*; gray nodes represent LTPs newly described by KAPPA. Edge length reflects pattern similarity: the longer the edge, the smaller the pairwise κ-score.

Figure 3.4. Compared clustering performance on AtLTPs.

Together, these results suggest that the KAPPA clustering function is able to reproduce and improve previous, manually formed groups.

3.6 Discussion

By providing an automated pipeline specifically dedicated to the evolutive specificities of proteins defined by a key amino acid pattern, like CRPs, KAPPA fills a gap in the landscape of sequence search computational tools. Table S3.4 (page 268) presents salient advantages of KAPPA over more conventional approaches.

The κ-score appears to be a robust, quantitative and objective tool to describe cysteine pattern similarity and explore all its aspects: block identity I accounts for subtle modifications of cysteine spacing within a given family, while pattern persistence P allows to investigate emergence of new groups of CRPs within a given taxon.

Numerous families of CRPs have been detected and characterized so far in a small number of model organisms. Moreover, the number and availability of new sequenced proteomes is quickly growing, especially for non-model organisms. Considering the functional importance of CRPs in plants and animals, there is an increasing need to characterize their orthologues in these new proteomes.

Focusing on the cysteine spacing with the κ-score, the KAPPA *ab initio* search function made it possible not only to detect new members of CRP families within reference species themselves, but also orthologues belonging to more distant species. Furthermore, the *ab initio* search is also flexible, since mapping parameters, especially stringency options, enable users to choose an optimal research framework. One can indeed decide to be strict in the similarity search, using a high κ-score threshold or high stringency parameters, or more permissive and detect new '-like' proteins.

Traditional sequence search approaches such as BLASTp do not make a difference between key residues and the rest of the sequence; hence, true homologues of input CRPs are intermingled with non-related, low similarity sequences. Although HMMER and PSI/ PHI-BLAST can give more weight to cysteines and other conserved residues, they are less performant in dealing automatically with extensive divergence of blocks between cysteines, and with fine modifications of the cysteine spacing itself. This is why previous reports dealing with CRPs always implemented curation steps relying on manual review of data or home-made scripts to obtain

a consistent final protein dataset.

Though often leading to correct results, this time-consuming approach is suboptimal when dealing with large-scale, proteome-wide studies. KAPPA addresses this issue by providing a fast and accurate analysis pipeline based on quantitative criteria. Moreover, the optimization functionality makes it easy to determine the best parameters for a given sequence search experiment. Besides providing an automated pipeline, KAPPA also innovates by relieving the user of the need for reference cysteine patterns. Indeed, the *de novo* search and clustering function applied to a whole proteome can uncover totally new families of CRPs without a reference.

In the challenging case of proteins defined by a key amino acid pattern, KAPPA provides a new and accurate detection method over HMMER-based sequence search strategies implemented in previous CRP studies[4,18] and in gene-finding pipelines such as SPADA.[32]

Implementation of KAPPA on available proteomes opens the way to a better and quicker understanding of the diversity, evolution and functions these peculiar proteins, as shown for the CRPs, EGZ proteins and SPRR2s.

Bibliography

1. Mousavi, A. et Hotta, Y. (2005). Glycine-rich proteins: a class of novel proteins. *Appl. Biochem. Biotechnol.*, 120(3), 169–74. DOI: 10.1385/ABAB:120:3:169. [cit. p. 91]

2. Scocchi, M. ; Tossi, A. et Gennaro, R. (2011). Proline-rich antimicrobial peptides: converging to a non-lytic mechanism of action. *Cell. Mol. Life Sci.*, 68(13), 2317–30. DOI: 10.1007/s00018-011-0721-7. [cit. p. 91]

3. Bella, J. ; Hindle, K. L. ; McEwan, P. A. et Lovell, S. C. (2008). The leucine-rich repeat structure. *Cell. Mol. Life Sci.*, 65(15), 2307–33. DOI: 10.1007/s00018-008-8019-0. [cit. p. 91]

4. Edstam, M. M. ; Viitanen, L. ; Salminen, T. A. et Edqvist, J. (2011). Evolutionary history of the non-specific lipid transfer proteins. *Mol. Plant*, 4(6), 947–64. DOI: 10.1093/mp/ssr019. [cit. p. 91, 98, 99, 102, 103, 104, 105, 107]

5. Giacomelli, L. ; Nanni, V. ; Lenzi, L. et coll. (2012). Identification and characterization of the defensin-like gene family of grapevine. *Mol. Plant-Microbe Interact.*, 25(8), 1118–31. DOI: 10.1094/MPMI-12-11-0323. [cit. p. 91]

6. Hanks, J. N. ; Snyder, A. K. ; Graham, M. A. et coll. (2005). Defensin gene family in *Medicago truncatula*: structure, expression and induction by signal molecules. *Plant Mol. Biol.*, 58(3), 385–99. DOI: 10.1007/s11103-005-5567-7. [cit. p. 91]

7. Silverstein, K. A. T. ; Graham, M. A. ; Paape, T. D. et VandenBosch, K. A. (2005). Genome organization of more than 300 defensin-like genes in *Arabidopsis*. *Plant Physiol.*, 138(2), 600–10. DOI: 10.1104/pp.105.060079. [cit. p. 91]

8. Pasupuleti, M. ; Schmidtchen, A. et Malmsten, M. (2012). Antimicrobial peptides: key components of the innate immune system. *Crit. Rev. Biotechnol.*, 32(2), 143–71. DOI: 10.3109/07388551.2011.594423. [cit. p. 92]

9. Odintsova, T. et Egorov, T. (2012). Plant antimicrobial peptides. *In Plant Signaling Peptides*, volume 16 de *Signaling and Communication in Plants*, pages 107–33. Springer, Berlin, Heidelberg, Allemagne. ISBN: 978-3-642-27602-6. DOI: 10.1007/978-3-642-27603-3_7. [cit. p. 92]

10. van de Velde, W. ; Zehirov, G. ; Szatmari, A. et coll. (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*, 327(5969), 1122–6. DOI: 10.1126/science.1184057. [cit. p. 92]

11. Wong, E. S. W. et Belov, K. (2012). Venom evolution through gene duplications. *Gene*, 496(1), 1–7. DOI: 10.1016/j.gene.2012.01.009. [cit. p. 92]

12. de Souza Cândido, E. ; e Silva Cardoso, M. H. ; Sousa, D. A. et coll. (2014). The use of versatile plant antimicrobial peptides in agribusiness and human health. *Peptides*, 55, 65–78. DOI: 10.1016/j.peptides.2014.02.003. [cit. p. 92]

13. Marshall, E. ; Costa, L. M. et Gutierrez-Marcos, J. (2011). Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J. Exp. Bot.*, 62 (5), 1677–86. DOI: 10.1093/jxb/err002. [cit. p. 92]

14. Chevalier, É. ; Loubert-Hudon, A. ; Zimmerman, E. L. et Matton, D. P. (2011). Cell–cell communication and signalling pathways within the ovule: from its inception to fertilization. *New Phytol.*, 192(1), 13–28. DOI: 10.1111/j.1469-8137.2011.03836.x. [cit. p. 92]

15. Chevalier, E. ; Loubert-Hudon, A. et Matton, D. P. (2013). ScRALF3, a secreted RALF-like peptide involved in cell-cell communication between the sporophyte and the female gametophyte in a solanaceous species. *Plant J.*, 73(6), 1019–33. DOI: 10.1111/tpj.12096. [cit. p. 92]

16. Higashiyama, T. (2010). Peptide signaling in pollen-pistil interactions. *Plant Cell Physiol.*, 51(2), 177–89. DOI: 10.1093/pcp/pcq008. [cit. p. 92]

17. Koppers, A. J. ; Reddy, T. et O'Bryan, M. K. (2011). The role of cysteine-rich secretory proteins in male fertility. *Asian J. Androl.*, 13(1), 111–7. DOI: 10.1038/aja.2010.77. [cit. p. 92]

18. Silverstein, K. A. T. ; Moskal, W. A. ; Wu, H. C. et coll. (2007). Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J.*, 51(2), 262–80. DOI: 10.1111/j.1365-313X.2007.03136.x. [cit. p. 92, 98, 99, 101, 102, 103, 104, 105, 107]

19. Zhu, S. ; Gao, B. et Tytgat, J. (2005). Phylogenetic distribution, functional epitopes and evolution of the CS$\alpha\beta$ superfamily. *Cell. Mol. Life Sci.*, 62(19-20), 2257–69. DOI: 10.1007/s00018-005-5200-6. [cit. p. 92]

20. Petersen, T. N. ; Brunak, S. ; von Heijne, G. et Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8(10), 785–6. DOI: 10.1038/nmeth.1701. [cit. p. 93]

21. Bendtsen, J. D. ; Jensen, L. J. ; Blom, N. ; Von Heijne, G. et Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng., Des. Sel.*, 17(4), 349–56. DOI: 10.1093/protein/gzh037. [cit. p. 93]

22. Shannon, P. ; Markiel, A. ; Ozier, O. et coll. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11), 2498–504. DOI: 10.1101/gr.1239303. [cit. p. 97]

23. Halary, S. ; McInerney, J. O. ; Lopez, P. et Bapteste, E. (2013). EGN: a wizard for construction of gene and genome similarity networks. *BMC Evol. Biol.*, 13, 146. DOI: 10.1186/1471-2148-13-146. [cit. p. 97]

24. Swarbreck, D. ; Wilks, C. ; Lamesch, P. et coll. (2008). The *Arabidopsis* information resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.*, 36(Database issue), D1009–14. DOI: 10.1093/nar/gkm965. [cit. p. 98]

25. Ouyang, S. ; Zhu, W. ; Hamilton, J. et coll. (2007). The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.*, 35(Database issue), D883–7. DOI: 10.1093/nar/gkl976. [cit. p. 98]

26. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32(5), 1792–7. DOI: 10.1093/nar/gkh340. [cit. p. 98]

27. Finn, R. D. ; Clements, J. et Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.*, 39(Web Server issue), W29–37. DOI: 10.1093/nar/gkr367. [cit. p. 98]

28. Altschul, S. F. ; Gish, W. ; Miller, W. ; Myers, E. W. et Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215(3), 403–10. DOI: 10.1016/S0022-2836(05)80360-2. [cit. p. 98]

29. Altschul, S. F. ; Madden, T. L. ; Schäffer, A. A. et coll. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25(17), 3389–402. DOI: 10.1093/nar/25.17.3389. [cit. p. 98]

30. Kim, S. ; Jeon, T.-J. ; Oberai, A. et coll. (2005). Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.*, 102(40), 14278–83. DOI: 10.1073/pnas.0501234102. [cit. p. 103]

31. Cabral, A. ; Voskamp, P. ; Cleton-Jansen, A. M. et coll. (2001). Structural organization and regulation of the small proline-rich family of cornified envelope precursors suggest a role in adaptive barrier function. *J. Biol. Chem.*, 276(22), 19231–7. DOI: 10.1074/jbc.M100336200. [cit. p. 103]

32. Zhou, P. ; Silverstein, K. A. ; Gao, L. et coll. (2013). Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). *BMC Bioinf.*, 14(1), 335. DOI: 10.1186/1471-2105-14-335. [cit. p. 107]

CHAPITRE 4

Exploration du sécrétome ovulaire

Ce chapitre a été publié le 21 septembre 2015 sous le titre *The plant ovule secretome : a different view toward pollen–pistil interactions* dans *Journal of Proteome Research*, 14(11), 4763–75. Il est accessible en ligne à https ://doi.org/10.1021/acs.jproteome.5b00618.

Auteurs Yang Liu*, Valentin Joly*, Sonia Dorion, Jean Rivoal et Daniel P. Matton (*co-premiers auteurs)

Contributions Valentin Joly est co-premier auteur avec Yang Liu ; leurs contributions au travail publié sont équivalentes. Yang Liu a effectué les travaux expérimentaux au laboratoire, notamment la collecte des exsudats protéiques. Les tests de pureté des exsudats ont été effectués par Sonia Dorion et Jean Rivoal. Valentin Joly s'est consacré aux analyses bioinformatiques des données de spectrométrie de masse sur lesquelles s'appuient le manuscrit. Yang Liu et Valentin Joly ont travaillé ensemble à l'interprétation de ces résultats, à la rédaction du manuscrit, à la préparation des figures et du matériel supplémentaire. Daniel P. Matton a contribué à la conception du travail, et s'est chargé de sa supervision et de la révision critique du manuscrit.

4.1 Résumé

Lors de la reproduction sexuée des plantes, l'échange continu de signaux entre le pollen et le pistil (stigmate, style et ovaire) joue un rôle important pour la reconnaissance et la sélection du pollen. Il contribue à établir des barrières à l'hybridation et, *in fine*, à produire un jeu de graines optimal. Après avoir navigué dans le stigmate et dans le style, les tubes polliniques atteignent leur destination finale : l'ovule. Cette étape terminale est elle aussi régulée par de nombreux signaux émanant du sac embryonnaire de l'ovule et qui englobent une grande variété de molécules. Toutefois, la spécificité à l'espèce de l'interaction pollen–ovule repose en majorité sur des protéines sécrétées et leurs récepteurs. L'isolement de gènes candidats impliqués dans les interactions pollen–pistil s'est, jusqu'ici, principalement appuyé sur des approches transcriptomiques, sans tenir compte de potentielles régulations post-transcriptionnelles. Pour résoudre ce problème, des exsudats d'ovules ont été extraits à partir de l'espèce de pomme de terre sauvage *Solanum chacoense* en utilisant une méthode d'extraction gravitationnelle sans tissu (*tissue-free gravity-extraction method*, tf-GEM). La combinaison du séquençage d'ARN et des analyses protéomiques basées sur la spectrométrie de masse ont conduit à l'identification de 305 protéines sécrétées, dont 58 % étaient spécifiques à l'ovule. Des analyses comparatives effectuées entre des ovules matures (attirant les tubes polliniques) et des ovules immatures (ne les attirant pas) ont révélé que la dernière étape du développement du sac embryonnaire affectait presque la moitié du sécrétome ovulaire. Parmi les 128 protéines enrichies à l'anthèse, 106 n'étaient pas régulés au niveau de l'ARNm, ce qui souligne l'importance des régulations post-transcriptionnelles dans le développement reproductif.

Mots-clés Interactions pollen–pistil, exsudats ovulaires, sécrétome, chimioattraction, guidage du tube pollinique, nano-LC–MS/MS sans gel, quantification *label-free*

4.2 Abstract

During plant sexual reproduction, continuous exchange of signals between the pollen and the pistil (stigma, style, and ovary) plays important roles in pollen recognition and selection, establishing breeding barriers and, ultimately, leading to optimal seed set. After navigating through the stigma and the style, pollen tubes (PTs) reach their final destination, the ovule. This ultimate step is also regulated by numerous signals emanating from the embryo sac (ES) of the ovule. These signals encompass a wide variety of molecules, but species-specificity of the pollen–ovule interaction relies mainly on secreted proteins and their receptors. Isolation

of candidate genes involved in pollen–pistil interactions has mainly relied on transcriptomic approaches, overlooking potential post-transcriptional regulation. To address this issue, ovule exudates were collected from the wild potato species Solanum chacoense using a tissue-free gravity-extraction method (tf-GEM). Combined RNA-seq and mass spectrometry-based proteomics led to the identification of 305 secreted proteins, of which 58 % were ovule-specific. Comparative analyses using mature ovules (attracting PTs) and immature ovules (not attracting PTs) revealed that the last maturation step of ES development affected almost half of the ovule secretome. Of 128 upregulated proteins in anthesis stage, 106 were not regulated at the mRNA level, emphasizing the importance of post-transcriptional regulation in reproductive development.

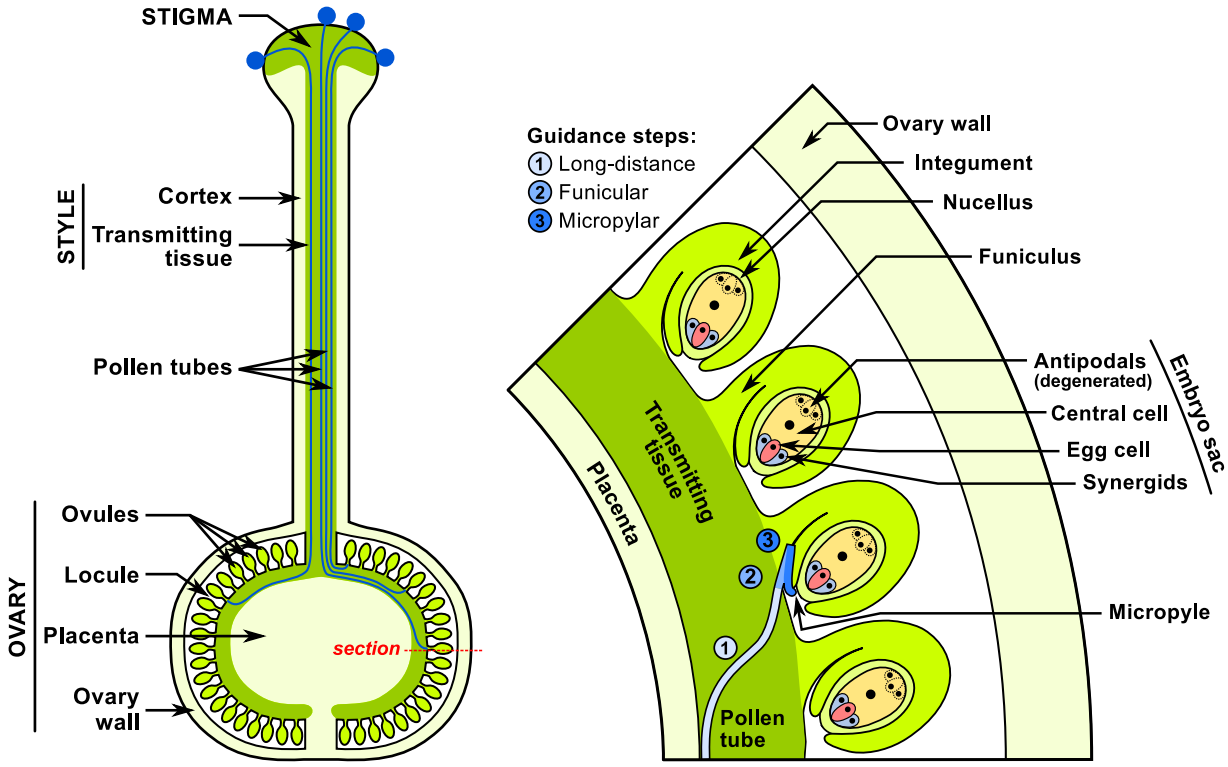
Keywords Pollen–pistil interactions, ovule exudates, secretome, chemoattraction, pollen tube guidance, gel-free nano-LC–MS/MS, label-free quantification

4.3 Introduction

Communication through extracellular signals starts with the synthesis and release of signaling molecules from a signaling cell(s). Some molecules may remain bound to the signaling cell and only affect cells in its vicinity. Such a case can be seen in the attachment of sperm to egg cells during fertilization in mammals via the interaction between sperm-tethered membrane protein Izumo[1] and its sole surface receptor Juno[2] on the egg cell to achieve proper gamete binding. In most cases, however, signal molecules are secreted, followed by their active transport or diffusion. Detection of the signals by the target cell then relies on specific receptor proteins that relay this information inside the cell, where a second signaling phase starts, leading to appropriate downstream responses. One way to address the compendium of secreted proteins that could play roles in extracellular signaling is to analyze the secretome of a specific tissue or organ under various developmental conditions or following biotic or abiotic stresses. The secretome refers to the proteins and peptides that are secreted out of the plasma membrane and, more specifically in plants, to the cell wall and apoplastic fluid (APF). So far, secretomic studies have been reported for some tissues in angiosperms including root cap, seeds, seedlings, fruit pericarps, leaves, and stems *in planta*, as well as in cell suspension cultures.[3,4,5,6,7,8,9,10,11] Secretomes of reproductive tissues have also been conducted in tobacco, maize, lily, and olive tree stigmas[12,13,14] as well as in PTs.[15] In gymnosperm species, the secretome of pollination drops was also reported.[16] These originate from ovular secretion that fills the micropyle of the ovule to enhance pollen capture and facilitate pollen germination.

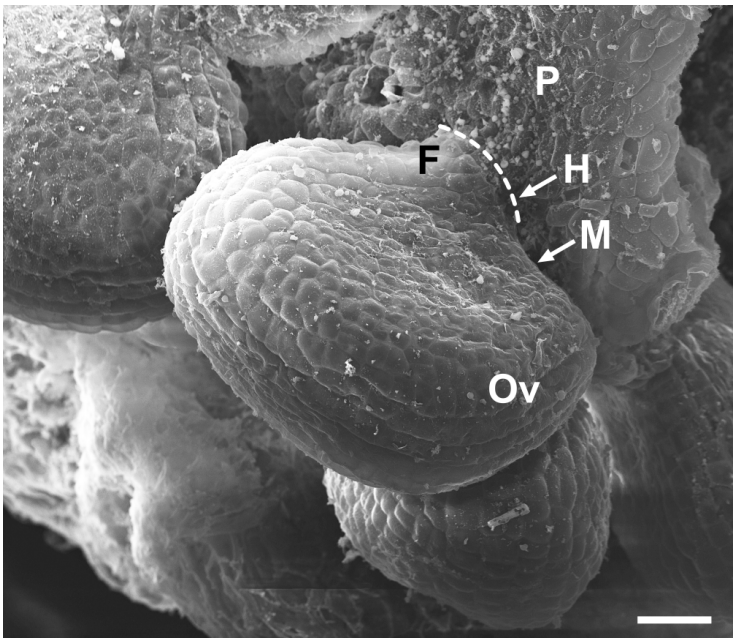
Depending on the biological context, secreted proteins participate in a wide range of processes, directly or indirectly affecting their neighboring cellular environment. For instance, analysis of the root tip secretome revealed numerous proteins released to the extracellular space (ECS) that protect root tips from fungal infection.[10] To resist low-temperature stress, seedlings were reported to secrete antifreeze molecules in response to low-temperature stress.[17] Besides protecting against microbial infections and abiotic stresses, plants also use exudates to distinguish dissimilar individuals from their own siblings in the context of plant competition.[18]

Here, we report the first secretome study of a plant ovule, an organ deeply enclosed in the carpel, the female reproductive organ consisting of the stigma, style, and ovary. From pollen landing on the stigma until fertilization, PTs are believed to be precisely guided to target each available ovule.[19] In solanaceous species, the PT path is a continuous tract of specialized cells running from the stigma surface through the stylar transmitting tract that ultimately splits along the ovary septa and merges with the placental epithelium (Figure 4.1a), as showed in *Solanum lycopersicum*[20] and *Nicotiana alata*.[21] In the preovular guidance stage, pollen grains germinate and elongate extracellularly in the transmitting tract of the style. After entering the ovary, ovular guidance cues take over to eventually steer the PT toward the micropyle (Figures 4.1b–d), a small opening at the ovule surface through which the PT penetrates to effect double fertilization.[22,23] Ovular guidance can be divided into three distinct steps, as depicted in Figure 4.1b. A long-distance activity was suggested to describe ovular signals that change PT trajectory from the intercellular space of the transmitting cells to the septum, associated with a tendency for PTs to emerge from the ovary proximal position to the stylar end.[24] This phase is illustrated by the phenotype observed in the *Arabidopsis* mutant for the pollen-expressed cation/proton exchangers *CHX21* and *CHX23*. In this mutant, PTs do not respond to ovular cues and fail to exit the transmitting tract.[25] The second step involves funicular guidance, where PTs adhere to and migrate up the funiculus, a stalk-like structure that attaches the ovule to the placenta, and is followed by a third step, described as micropylar guidance, that targets the PT at close range, within ~100–150 µm to the micropyle.[26,27,28] The latter two steps can be further distinguished on the basis of the phenotype observed in the *Arabidopsis mpk3/mpk6* mutant.[29] Here, the majority of mutant PTs exit the transmitting tract, but they either get lost in the septum or take a much longer time to eventually target the ovule, exhibiting a funicular guidance defect *in vivo*. Interestingly, in the *mpk3/mpk6* mutant, micropylar guidance is not affected, as demonstrated by a semi *in vivo* (SIV) guidance assay, emphasizing the distinct nature of different signals controlling funicular and micropylar guidance, at least in *Arabidopsis*.

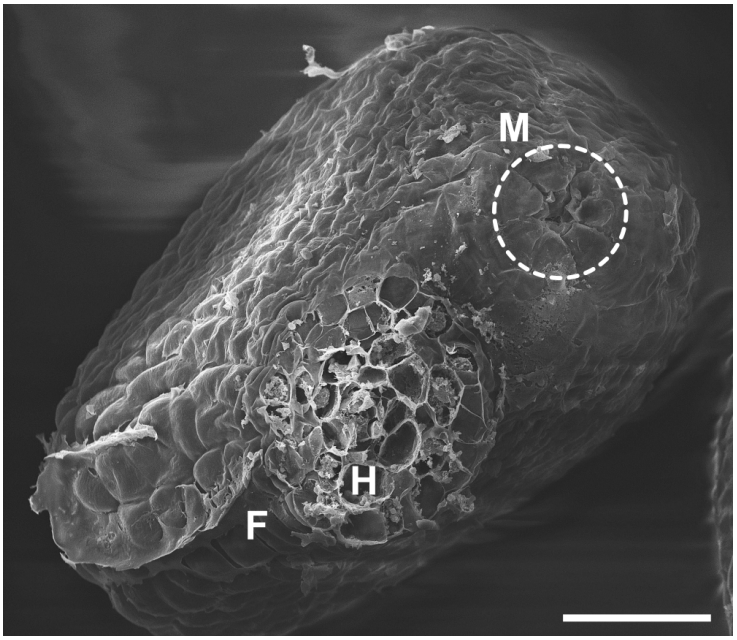


(a) Longitudinal section of the pistil **(b)** Close-up of an ovary transverse section

Figure 4.1. Schematic depiction of PT growth in the *S. chacoense* pistil. (a) Dark green indicates the PT pathway, a continuous cell tract originating from the stigma surface, extending through the style down to the placental epithelium, where it connects to the ovules. At the mature stage of an ovule, the ES consists of an egg cell, two synergid cells, and a central cell, embedded in one layer of integument in *S. chacoense*. **(b)** PT typically approaches the micropyle of an ovule through three guidance steps. Step 1 (long- distance guidance), PT emerges from the intercellular space of the transmitting tract to the placental surface; step 2 (funicular guidance), PT climbs up the funiculus; and step 3 (micropylar guidance), PT navigates toward and into the micropyle. **(c–d)** Scanning electron microscopy (SEM) image of *S. chacoense* wild-type ovules.



(c) Scanning electron microscopy (SEM) image of *S. chacoense* wild-type ovules connected to the placental tissue. The micropylar opening is hidden, facing the placental side. Dotted lines trace the position of the hilum. Noticeably, the particular morphology of *S. chacoense* ovules brings the micropyle in the immediate vicinity of the placental surface, rendering the micropyle directly accessible to emerging PTs.



(d) SEM image of a *S. chacoense* ovule detached from the placenta. The dashed circle marks the micropylar opening. F, funiculus; H, hilum; M, micropyle; Ov, ovule; P, placenta. Scale bar = 30 μm.

Figure 4.1. Schematic depiction of PT growth in the *S. chacoense* pistil. (suite)

Among the multistep control points of PT guidance, long-distance guidance activity, although shown to be governed by the ovule, is largely unexplored.[24] Funicular guidance appears to be associated with the ovule sporophytic tissues,[30] as demonstrated by the *pop2*,[31] *siz1-2*,[32] and *pdil2-1*[33] mutants, where PTs migrate up the funiculus but become lost near the micropyle. Micropylar guidance is controlled by the female gametophyte.[34] Such short-distance attractants have been identified in *Torenia*[35,36] and maize[37] through transcriptomic analysis and in *Arabidopsis*[38] through comparative phylogenetic studies. Collectively, these results suggest that the ovule is an active player in male–female gametophyte communication, guiding PTs to navigate through the ovarian mucilage to achieve double fertilization.[22,39]

Here, we present the ovule secretome as a different approach to study pollen–pistil interactions. By examining the protein composition of this bioactive environment, we anticipate gaining more insight into pollen–ovule interaction processes, involving PT growth stimulation, competency acquisition, guidance, and PT repulsion. To this aim, comparative proteomic analyses were conducted on mature attracting ovules, as well as immature ovules, unable to attract PTs, as shown in SIV assays.

4.4 Experimental section

4.4.1 Plant materials and growth conditions

Solanum chacoense Bitt. individuals were greenhouse-grown under long-day conditions (16 h light/8 h dark). Two genotypes, G4 ($S_{12}S_{14}$ self-incompatibility alleles) and V22 ($S_{11}S_{13}$), were used to perform compatible pollinations: G4 () × V22 ().

4.4.2 Semi *in vivo* PT guidance assay for *S. chacoense*

SIV guidance assays were performed as described previously.[40] In brief, 24 h after pollination (HAP), styles were excised and placed on BK solid medium,[41] allowing PTs to grow properly when exiting from the style around 30 HAP. Clusters of 5 to 10 ovules connected by placental tissue were positioned ~700 µm away from the end of the style (equivalent to the radius of the ovary), with a 45° angle. Two developmental stages were tested: mature ovules from flowers at anthesis (ES stage FG7) and immature ovules from 6 to 7 mm flower buds taken two days before anthesis (2DBA, ES stage FG6). Guidance response of emerging PTs

to ovule clusters was observed in bright field with an Axio Observer.Z1 microscope equipped with an AxioCam HRm camera (Zeiss). The turning angle of each PT was measured with ImageJ (http://imagej.nih.gov/ij). Angle distribution was compared between conditions using a two-sample Kolmogorov–Smirnov (KS) statistical test ($n > 80$).

4.4.3 Scanning electron microscopy (SEM)

S. chacoense G4 ovaries were hand-dissected to remove the pericarp (ovary wall). A group of ovules connected by placental tissues was fixed, dehydrated, and critical-point-dried as described previously.[42] Since *S. chacoense* ovules align tightly around the placental tissues, some ovules from the clusters were removed under a stereomicroscope in order to leave enough space for single-ovule observation. The tissues were coated with gold–palladium and viewed in a JEOL JSM-35 SEM.

4.4.4 Collection of ovule exudates

The pericarp was removed from the ovary. Ovule clusters connected by placental tissues were aligned on a 0.5 % agarose strip (0.5 cm × 1 cm) and incubated in a humidified chamber for 24 h in darkness at room temperature. Ovules were then taken out with a needle, and the gel strip containing ovule exudates was placed in a 0.5 mL centrifuge tube perforated at the bottom. A small amount of glass wool was placed under the gel piece to allow the flow-through of ovule exudates to the collection tube and to retain the gel matrix in place during centrifugation. The device was then inserted in a 1.5 mL tube and centrifuged 5 min at 5000 rpm. Exudates collected in the flow-through were centrifuged for 10 min at 14 400 rpm and 4 °C to remove residual impurities. The supernatant was flash frozen in liquid nitrogen and stored at −80 °C.

4.4.5 Total protein extraction

For western blot analyses and enzymatic assays, three biological replicates of whole ovules (50–70 mg fresh weight, FW) were collected from flowers at anthesis, placed on dry ice, and stored at −80 °C until use. Samples were then ground on ice with a pestle in a buffer containing 30 mM Tris·HCl pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.1 % (v/v) Triton X–100, 10 % (v/v) glycerol, 5 mM DTT, 5 % (w/v) insoluble PVPP, 5 mM ε-amino caproic acid, 1 mM benzamidine, 1 µg mL^{-1} leupeptin, and 2 mM PMSF, with a 3:1

ratio (milliliters of extraction buffer per gram of FW). Homogenates were centrifuged 15 min at 12000 × *g* at 4 °C. Supernatants were further centrifuged for 5 min. Clarified supernatants were used immediately for enzyme activity measurement. For immunoblot analysis, an aliquot of the supernatant was immediately heat-denatured in SDS sample buffer and kept frozen at −20 °C until use. Proteins were quantified with the Bradford method using bovine serum albumin as a standard.[43]

4.4.6 Enzymatic Assay

Three biological replicates of ovule exudates were obtained for enzymatic assays. Lyophilized samples were solubilized in a buffer containing 30 mM Tris·HCl pH 7.5, 5 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA, 5 mM DTT, 5 mM ε-amino caproic acid, 1 mM benzamidine, 1 µg mL^{-1} leupeptin, and 2 mM PMSF. The samples were centrifuged for 5 min at 12000 × *g* and 4 °C to eliminate possible insoluble materials. Supernatants from the three samples were pooled in a final volume of 20 µL and used immediately for enzymatic activity assays. Triose-phosphate isomerase (TPI) activity assays were performed according to a protocol described previously.[44] One unit (U) of enzyme activity corresponds to the appearance of the reaction product at a rate of 1 µmol min^{-1}. Activity was tested in two independent experiments. Intracellular contamination of exudates was assessed as the percentage of TPI activity present in the exudates over that of the total ovule extract.

4.4.7 SDS-PAGE and immunoblot analysis

Three independent ovule samples were obtained to perform western blots. SDS-PAGE analysis was performed on 15 % acrylamide gels with 9.0, 3.0, 1.0, 0.3 and 0.1 µg of proteins from ovules at anthesis and 9.0 µg of proteins from ovule exudates. Proteins were then transferred onto a nitrocellulose membrane at 70 V for 60 min. Membranes were incubated with an anti-cytosolic TPI (cTPI) antibody (1/500 dilution) for 1 h at room temperature.[44] Polypeptides were detected using goat anti-rabbit IgG secondary antibodies (1/10 000 dilution) conjugated to alkaline phosphatase (Promega). The reaction was visualized using BCIP (5-bromo-4-chloro3-indolyl-phosphate) and NBT (nitro-blue tetrazolium) and was allowed to develop for 20 min at room temperature.

4.4.8 RNA sequencing and *de novo* assembly

Total RNA was extracted using the TRIzol reagent (Invitrogen) as recommended by the manufacturer. Two next-generation sequencing platforms were used to generate the ovule transcriptome in parallel. First, the 454 GS-FLX Titanium platform was used to perform RNA-seq on ovules at anthesis and at 2DBA. cDNA libraries were constructed for each condition with the rapid library preparation kit (Roche) after an mRNA enrichment step performed with Dynabeads Oligo-$(dT)_{25}$ (Invitrogen). Reads were *de novo* assembled with Newbler software (Roche).[45] A second RNA-seq was performed with the Illumina HiSeq 2000 platform, on the anthesis sample only. The cDNA library was prepared using the TruSeq cDNA preparation kit (Illumina). A *de novo* assembly was then performed using Trinity software.[46] Both assemblies were used to generate a reference ovule transcriptome for protein identification.

4.4.9 Quantification of RNA Expression

In order to assess differential gene expression (DGE) for ovule-secreted proteins, three biological replicates of mRNA samples from anthesis and 2DBA ovules were sequenced on the Illumina platform as described above. Bowtie 2 software was used to align reads to all contigs encoding an ovule-secreted protein.[47] RSEM[48] and edgeR[49] were used to compute expression fold-changes between conditions. Transcripts with a fold-change below -2.0 or above 2.0 ($p \leq 0.05$) were considered to be differentially regulated.

4.4.10 Mass spectrometry (MS)

One microgram (1 µg) of lyophilized exudates from three biological replicates was trypsin-digested for 8 h at 37 °C. Samples were desalted with Ziptips (Millipore) and separated on a reversed-phase column (150 µm i.d. × 150 mm) equipped with a precolumn (0.3 mm × 5 mm) using a 56 min gradient from 10 to 60 % (v/v) acetonitrile in 0.2 % (v/v) formic acid and a 600 nL min^{-1} flow rate on a 2D-nanoLC system (Eksigent). Liquid chromatography (LC) was connected to LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). Full-scan MS spectra were acquired in the Orbitrap with a mass resolution of 60 000. The mass window for precursor ion selection was set to 2 m/z. Each full MS spectrum was followed by 12 MS/MS scans, where the 12 most abundant ions above a threshold of 10 000 counts with charge ≥ 2 were subjected to collision-induced dissociation in the linear ion trap at the Institute for Research in Immunology and Cancer (IRIC, Université de Montréal). The dynamic exclusion

was set to 45 s. The collision energy was set to 35, with an activation Q of 0.25.

4.4.11 Protein identification and quantification

Spectral processing and peak list generation was performed using Mascot Distiller v2.5.1 (Matrix Science). The data were searched with the Mascot search engine, v2.3.01, against concatenated forward and reversed six-frame translations of all 454 and Illumina contigs containing 100 662 forward sequences. The reversed sequences served as decoys, as opposed to the forward sequences in the target database. Tolerance was set at 15ppm for precursor and 0.5Da for fragment ions. Variable modifications were specified, including carbamidomethylation of cysteine, oxidation of methionine, deamidation, and phosphorylation of serine, threonine, and tyrosine residues. The false discovery rate (FDR) was calculated as the ratio of decoy matches versus target matches. Initially, proteins at 5 % FDR with at least 2 peptides detected were reported. We applied a curation step to these contigs, as explained in Figure S4.1 (page 269). After curation, proteins at 3 % with at least 6 assigned peptides were reported in the ovule secretome.

Label-free protein quantification was performed using the ProteoProfile program (`http://www.thibault.iric.ca/ProteoProfile/`). Mascot peptide identifications were matched to MS peak intensity extracted from the aligned MS raw data files (tolerances set to m/z: 15 ppm and retention times: 1 min). Only peptides with a minimum intensity value of 10 000 counts were analyzed. Normalization was performed across different replicates and conditions for peptide intensities during each LC–MS run to adjust the median of their logarithms to zero. Protein intensities were summed using the median of all associated peptides. Only proteins defined by at least two quantified peptides were reported in the ovule secretome. The mean coefficient of variation (CV) of protein intensity for ovule secretome at anthesis was 0.372, and 56.3 % of proteins were below this mean (Figure S4.2, page 270). Proteins showing a fold-change below -2.0 or above 2.0 between conditions with a p-value below 0.05 following a Student's t-test were reported as being differentially secreted.

4.4.12 *In silico* predictions

Protein sequences were analyzed *in silico* for secretion predictions with the SignalP 4.0[50] and SecretomeP 1.0[51] (score $\geq$0.5) programs. Sequences were also blasted against the `nr` database using BLASTp with default parameters.[52] The best informative BLAST hits were retained. The Blast2GO program was used to annotate the ovule secretome.[53] Protein family and

domain information were retrieved from PFAM (http://pfam.sanger.ac.uk/).

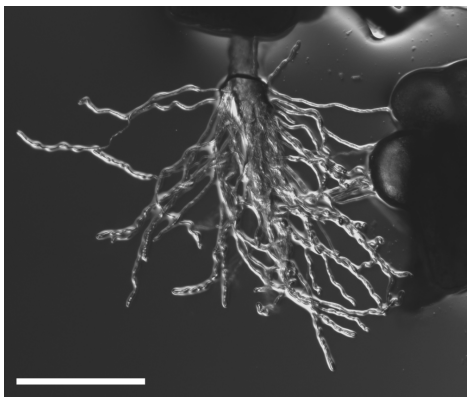
4.4.13 Estimates of diversifying positive selection

Positive selection estimates were conducted on cysteine-rich peptides (CRPs) in the ovule secretome based on the ratio of divergence at nonsynonymous and synonymous sites (d_N/d_S). For each CRP investigated, closest orthologs in 10 solanaceous species were retrieved using the reciprocal best BLAST hits method and aligned codon-by-codon. The `codeml` program from the PAML suite [54] was run to compute likelihoods of the M7 (neutralist) and M8 (selectionist) models using the phylogeny described by Goldberg *et al.* [55]. A likelihood ratio test was finally performed to determine if positive selection was acting on the aligned sequences.

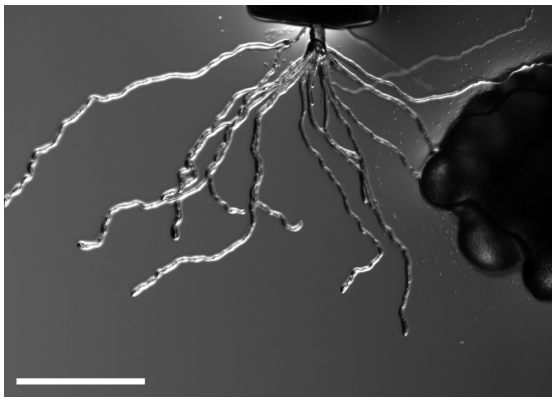
4.5 Results

4.5.1 PTs are attracted by mature ovules

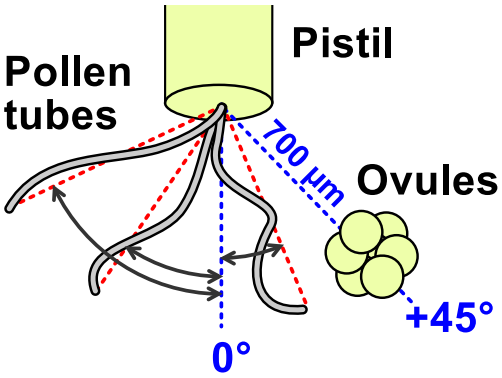
The chemotropic effect from ovules at two developmental stages was examined. Figures 4.2a and 4.2b shows typical PT behavior toward each ovule condition. Most PTs grew directly toward mature ovule clusters, whereas 2DBA ovule clusters showed lesser or no attraction. On the basis of the angles defined in Figure 4.2c for this assay, a Kolmogorov–Smirnov (KS) test was employed to statistically distinguish random growth from PT attraction by ovules (Figure 4.2d, $n > 80$). KS test results indicated that immature ovules exhibited significantly different PT behavior ($p = 0.02$) compared to that of ovules at anthesis stage.



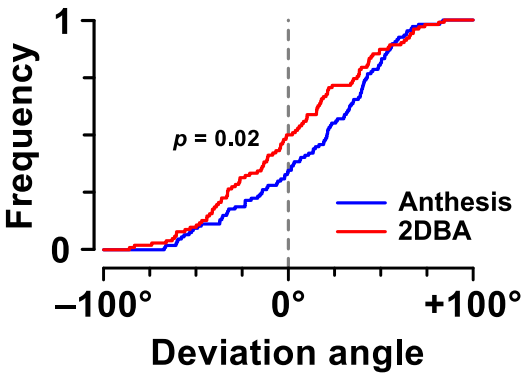
(a) PT guidance assay with anthesis ovule clusters. Scale bar: 400 µm.



(b) PT guidance assay with 2DBA ovule clusters. Scale bar: 400 µm.



(c) Schematic depiction of SIV PT guidance assay in *S. chacoense*. Double-headed arrows define the angle of each PT in this assay.



(d) The angles of PT curvature between two developmental stages were compared by the KS test. Results obtained from KS test indicated immature ovules exhibited significantly different PT behavior ($p = 0.02$) as compared to that of ovules at anthesis stage.

Figure 4.2. Chemotropism of PTs toward anthesis ovule clusters and immature ovule clusters (2DBA) in *S. chacoense*.

4.5.2 Secretome protein isolation through a modified tissue-free Gravity-Extraction Method (tf-GEM)

In order to isolate proteins that could be involved in PT–ovule interactions, ovule exudates were collected. Two methods are commonly used for *in planta* protein collection, namely, the vacuum-infiltration centrifugation (VIC) method[56] and the gravity-extraction method (GEM).[57] In the VIC method, a vacuum is applied to infiltrate buffer into the ECS and then the APF is harvested by centrifugation. In an effort to minimize cell damage during vacuum infil-

tration, the GEM was developed, where plant tissues are centrifuged directly at low speed to extract APF without bathing ECS in extraction buffer/H_2O in advance. Due to the fragility of the ovule, a GEM-based system was specifically tailored for ovule secretome studies and named tf-GEM for tissue free-GEM. A similar approach was also recently used to study PT secretome.[14] The workflow is schematically described in Figure 4.3. Briefly, a gel-based medium is used as a support for the tissue considered. The sample is placed on a gel, allowing the exudates to soak in the solid medium. Following incubation, ovules are removed, and the supporting gel is spun to recover the exudates. It should also be noted that both VIC and GEM methods aim to collect all apoplastic proteins including cell wall proteins, whereas the tf-GEM herein, by removal of ovule clusters before spinning, is intended to collect secreted proteins that diffuse from the apoplastic space.

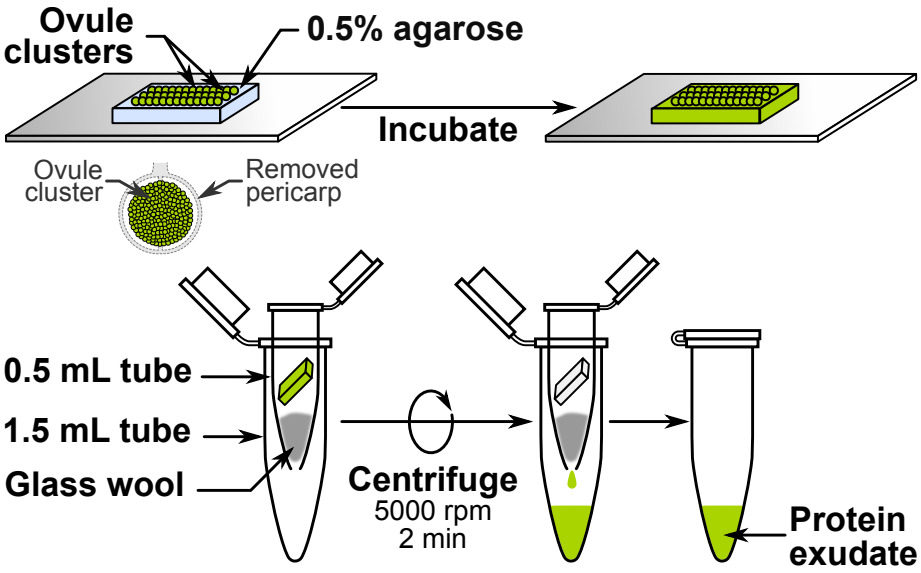


Figure 4.3. Tissue-free GEM system workflow for ovule secretome isolation. After removal of the pericarp, ovule clusters are incubated on an agarose matrix. Following an overnight incubation, ovules are removed from the gel matrix under a stereomicroscope. The matrix is laid on top of glass wool in a small perforated tube, and exudates are collected in a larger tube by centrifugation. Collected protein exudates are flash frozen in liquid nitrogen and stored at −80 °C until use.

4.5.3 Purity assessment

To monitor if dissection caused contamination of the ovule exudates due to intracellular protein leakage, presence and activity of a control protein, triose phosphate isomerase (TPI), was tested. First, a western blot was performed using an antibody specific for cytosolic TPI

(cTPI), the main isoform of TPI present in plant tissues.[44] This experiment revealed no trace of cTPI in ovule exudates, whereas a clear band was observed when using total ovule proteins as control (Figure 4.4). Second, a highly sensitive enzymatic assay assessing total TPI activity in the exudates[44] showed that intracellular contamination was limited to 0.79 % (Table 4.1). These results confirm that tf-GEM is a noninvasive technique that minimizes cytosolic content leakage.

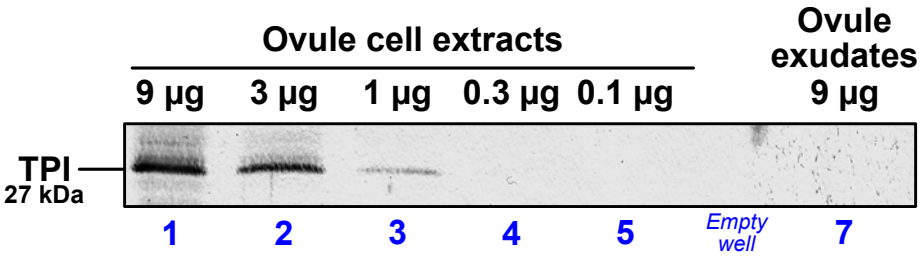


Figure 4.4. Ovule exudates' purity assessment. Western blot analyses of the cytosolic TPI isoform were performed from whole ovule extracts and ovule exudates to detect cytosolic leakage stemming from tissue wounding. Lanes 1–5, dilution series (9, 3, 1, 0.3 and 0.1 µg) of total ovule proteins. Lane 7, ovule exudates (9 µg) obtained from the tf-GEM protein collection.

Table 4.1. Assessment of TPI enzymatic activity in whole ovules extracts and ovule exudates from *S. chacoense*. Data (±) indicate standard deviation of the means. Contamination was calculated from TPI activity in ovule exudates divided by the one from whole ovule extracts.

TPI enzymatic activity ($U \cdot mg^{-1}$ prot)		Cytosolic contamination
Whole ovules	Exudates	
12.928 ± 1.675	0.102 ± 0.013	0.79 %

4.5.4 Curation of transcriptomic data drastically increased coverage for protein identification

Maximization of protein identification by MS requires a high-quality reference proteome, which can be a major hurdle when using a nonmodel species like *S. chacoense*. Although annotated genomes are available for other solanaceous species such as potato (*S. tuberosum*)[58] and tomato (*S. lycopersicum*),[59] reproductive proteins are often found to be highly divergent

(e.g., LURE-type PT attractants or self-incompatibility S-RNases). Thus, the usage of an annotated genome derived from a related species in MS may increase the risk of missing species-specific proteins. Therefore, we used transcriptomic data from *S. chacoense* to build our own reference proteome. First, RNA-seq assemblies from both 454 and Illumina platforms were translated into six frames and concatenated into an initial reference proteome. On the basis of this database, Mascot assigned 363 proteins (FDR = 5 %) to the ovule secretome. However, the large size of the RNA-seq database, the presence of untranslated regions (UTR), and the possibility of undesired peptide assignment to the junctions of concatenated amino acid sequences are factors that can increase the risk of false peptide discovery. In this study, a data curation protocol was designed to maximize and improve protein identification, as illustrated in Figure S4.1 (page 269). We applied the curation protocol to the 363 initially identified proteins. These contigs were queried against `refseq_rna` and `refseq_protein` databases from NCBI with the BLASTn and BLASTx programs, respectively, in order to delimitate open reading frames (ORFs) and to resolve frameshifted and chimeric contigs. Following this protocol, 362 curated proteins were obtained. On the basis of this refined reference proteome, Mascot assigned 305 proteins (FDR = 3 %) to the ovule secretome. On average, 17 peptides (ranging from 6 to 106 peptides) were assigned to each protein, compared to 4 peptides for the former search. Accordingly, by using all identified peptides, the average protein coverage was increased from 13.6 to 67.3 %, as calculated by Protein Coverage Summarizer (`http://omics.pnl.gov/`). This 5-fold increase in protein coverage increased the reliability of protein identification and demonstrated the feasibility and necessity of our curation protocol to MS sequencing, using transcriptomic data as reference.

4.5.5 Ovule secretome annotation

In all, 305 proteins were reported for the ovule secretome, with at least 6 peptides identified and a protein FDR of 3 %. These are hereafter designated as ovule-secreted proteins (OSPs). Table S4.1 (page 272) provides a catalog of these 305 proteins, with descriptions and annotations as predicted from bioinformatic analyses. Although a noninvasive method was used that did not require centrifugation or bathing the sample for exudates collection, it is noteworthy that the number of proteins recovered is consistent with other plant secretomic studies.[15,60,61] Biological process GO (gene ontology) categories were used to globally describe the ovule secretome (Figure S4.3, page 271). The most represented GO categories are associated with metabolism (GO:0008152; 26 %), cell growth and/or maintenance (GO:0009987; 19 %), single-organism process (GO:0044699; 21 %), and physiological response to stimulus

(GO:0050896; 8 %). To a lesser extent, proteins that are associated with developmental process, growth, signaling, and reproduction account for 8 %. In parallel, PFAM domain and family information were queried. Proteins with the same domain or family accession were pooled and displayed in Table S4.2 (page 272).

Following Blast2GO annotation of biological processes, we examined protein secretion patterns of OSPs to study their predicted cellular localization. Protein secretion can proceed through different pathways. The best known is the classical ER–Golgi secretory pathway, where proteins require an N-terminal signal peptide (SP) for protein sorting. Lesser known are proteins secreted independently of the ER–Golgi pathway. A growing body of evidence demonstrates the importance of this unconventional protein secretion (UPS) pathway, i.e., the production of leaderless secretory proteins (LSPs), which could account for more than 50 % of plant secretomes.[62] In the ovule secretome, 64 (21 %) proteins were predicted to possess a SP. Some 86 proteins (28 %) were predicted as LSPs by SecretomeP while not being predicted by SignalP. Similarly, SP-containing proteins also accounted for 20 % of root cap secretome[10] and of the SIV-PT secretome.[15]

4.5.6 Specificity of the ovule secretome

In order to sort out proteins that are unique to the ovule secretome, OSPs were separated into ovule-specific OSPs (oOSPs) and nonspecific OSPs (nOSP), by comparing them to the generally secreted proteins (GSPs) described in the PlantSecKB curated databases,[63] and stigma exudate proteins (SEPs) from lily (*Lilium longiflorum*), olive (*Olea europaea*),[12,13] and tobacco (*Nicotiana tabacum*).[14] SEPs were treated separately from GSPs since stigma exudates are active in in vitro PT chemotropism[64] and might share a certain level of similarity with ovule exudates. The ovule secretome was then queried against the two aforementioned subproteomes with BLASTp. Putative orthologs based on best hits with an e-value below 1×10^{-10}, a minimum alignment coverage of 80 %, and a minimum amino acid identity of 50 % are reported in Table S4.3 (page 272).

As shown in Figure 4.5, 116 (38 %) and 52 (17 %) OSPs have a putative ortholog in the GSP and SEP databases, respectively, with an overlap of 41 proteins (13 %) in all three profiles. OSPs shown to have a putative ortholog detected in either GSPs or SEPs were named nonspecific OSPs (nOSPs). The existence of nOSPs indicates that common factors are secreted into the extracellular matrix, regardless of the tissues examined. Such factors include pathogenesis-related (PR) proteins, like chitinase, β-1,3-glucanase, and peroxidases. Besides PR proteins, peptidases,

heat shock proteins HSP70 and HSP90, and chaperonin 60 are also released to the extracellular environment. These proteins are mainly involved in plant defense against pathogens and stress responses.[57] A self-incompatibility ribonuclease (S-RNase, OSP286) was also found to be common among all three secretome profiles, although its putative ortholog in SEP database is only 32 % identical to OSP286 due to the highly divergent nature of S-RNases.[65]

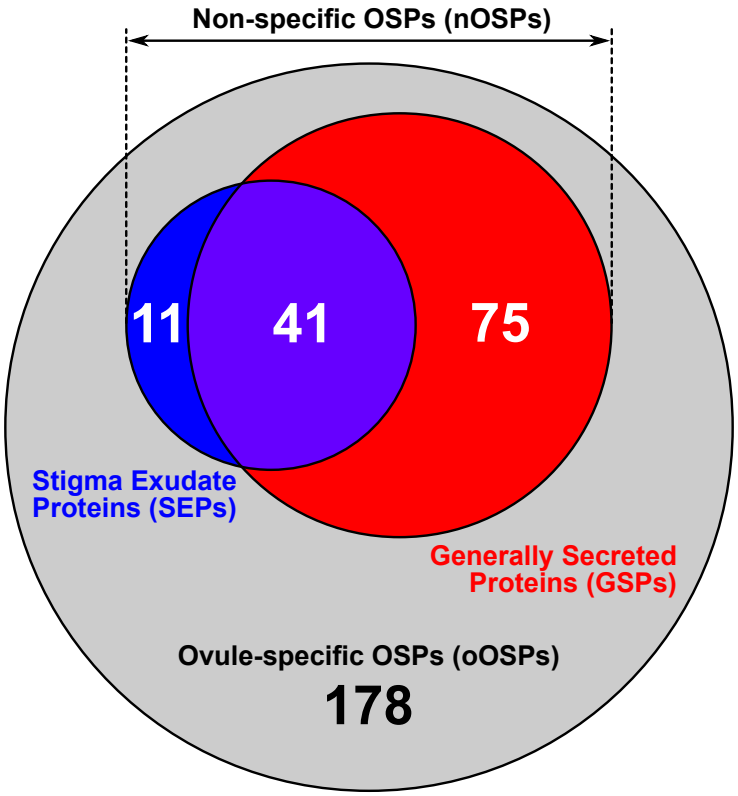


Figure 4.5. Diagram illustrating the number of oOSPs and nOSPs in the ovule secretome as compared to GSPs obtained from PlantSecKB database and SEPs collected from lily, olive, and tobacco stigma exudates.

In contrast to nOSPs, 178 OSPs did not have any ortholog in other known secretomes. These oOSPs accounted for 58 % of the ovule secretome. Notably, these include a γ-amino butyric acid-transaminase (GABA-T) OSP131, which was previously shown to be involved in sporophytic PT guidance in *Arabidopsis*.[31] Four members of the CRP family were also detected in the ovule secretome (OSP21, 123, 227, 305), belonging to lipid-transfer proteins (LTP), Early Culture Abundant 1 (ECA1) gametogenesis-related family, and thionin-like CRP subgroups.

4.5.7 Label-free quantification revealed differentially secreted proteins between mature and immature ovules

As expected from the PT guidance assay, ovules at anthesis exhibited a distinct protein profile from that of nonattracting ovules (2DBA) in terms of secretion abundance. Compared to the 2DBA secretome, 128 and 7 proteins were, respectively, up- and down-regulated at anthesis, whereas the abundance of the majority of proteins (56 %) remained unchanged (Figure 4.6). Morphologically, the overall size of the ovule is not significantly different between 2DBA ovules and mature ones, but ES development is not yet completed in 2DBA ovules.[66] At this stage, polar nuclei fusion in the central cell has not yet occurred while antipodal cells degeneration has not yet completed. Although synergid cells and egg cell are fully cellularized by 2DBA, vacuolar development inside the egg apparatus has just initiated.[66] Since cell wall ingrowth at the micropylar apex of the synergid cells starts after the formation of a fully vacuolated, pear-shaped egg apparatus, the development of the filiform apparatus, although initiated, is most probably incomplete. Accordingly, immature ovules at 2DBA did not attract PTs (Figure 4.2). Indeed, synergids and the filiform apparatus had been previously shown to be involved in PTs attractant secretion.[34,67]

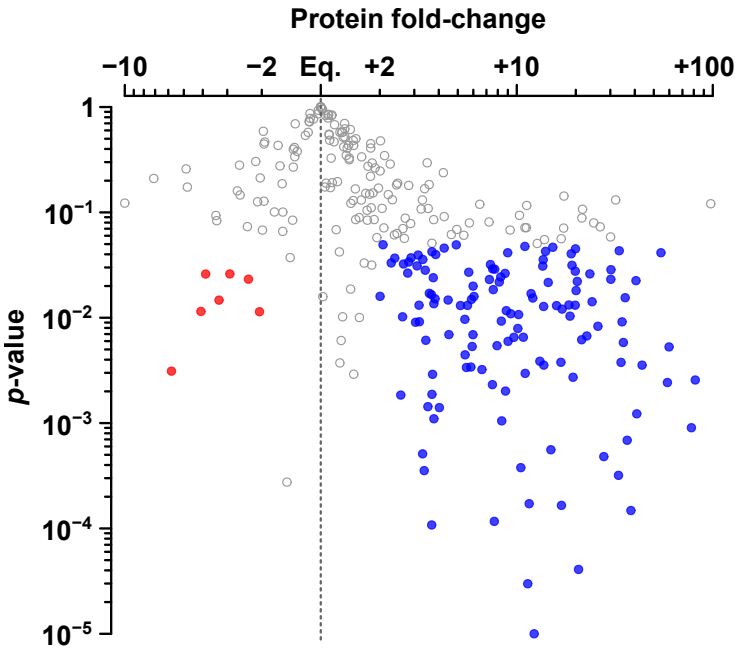


Figure 4.6. Protein secretion ratio in anthesis vs 2DBA ovules. Each dot represents a secreted protein in ovule secretome. Protein secretion ratios with a value < 1.0 were inverted to avoid fractional number. A minus symbol ($-$) was added to these values to indicate a negative regulation. Blue dots highlight OSPs that have a secretion fold-change of anthesis/2DBA > 2.0 and a p-value < 0.05. Red dots highlight OSPs that have a secretion fold-change of anthesis/2DBA < -2.0 and a p-value < 0.05.

On the basis of this result, we observe that, although deeply embedded in the ovule nucellus, the maturation step arising from stage FG6 to FG7 (mature ES) affected almost half (44 %) of the ovule secretome. The secretion status of ovule-expressed proteins is thus closely associated with late ES development. To extend our secretomic study, we conducted DGE analysis for the 305 OSPs. As shown in Figure 4.7, there is no correlation between gene expression and secretion abundance. Of the 128 upregulated proteins at anthesis (14-fold on average), 83 % (106) were not regulated at the mRNA level (1-fold on average), emphasizing the importance of post-transcriptional regulation in reproductive development and the necessity of conducting proteomic analysis in parallel to transcriptomic studies.

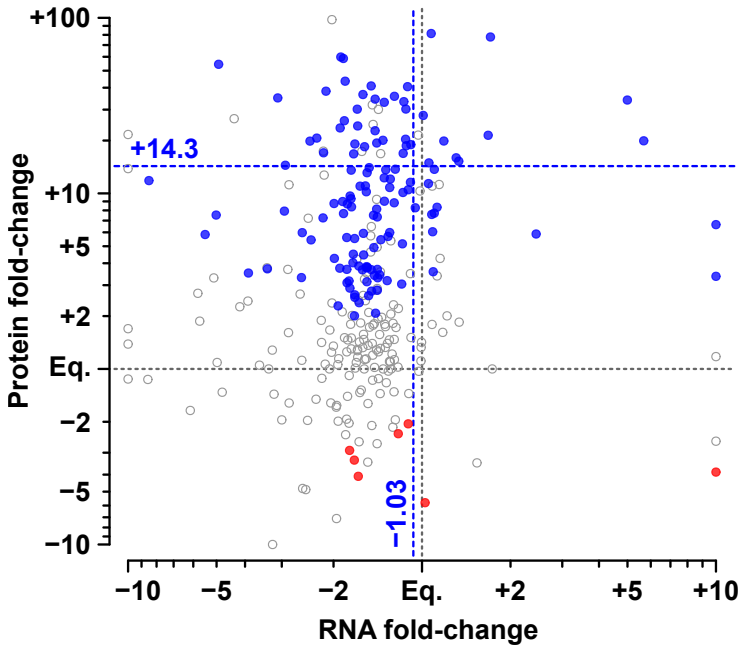


Figure 4.7. Fold-change correlations between protein secretion abundance and gene expression during the 2DBA to anthesis transition. Each dot represents a secreted protein in the ovule secretome. Blue dots highlight OSPs that are upregulated at anthesis at the protein level (fold-change > 2.0, p-value < 0.05). Red dots highlight OSPs that are downregulated at anthesis. Mean RNA (-1.03) and protein ($+14.3$) fold-changes from upregulated OSPs are represented as dotted lines.

4.6 Discussion

4.6.1 tf-GEM: A new approach for secretomic studies

In this study, we established tf-GEM as a simple and efficient method to isolate tissue exudates with minimal cytosolic contamination. First, as shown in Figure 4.3, instead of severing individual ovules from the placenta, ovule clusters were placed on the agarose strip, preventing placental cells from directly touching the gel, thus minimizing cell leakage from wounded plant tissues. Second, unlike the VIC and GEM methods, tf-GEM does not require centrifugation of plant tissues that can lead to cell breakage. The noninvasive nature of the tf-GEM is also substantiated by its very low cytosolic contamination levels (Figure 4.4 and Table 4.1).

4.6.2 A long-distance guidance SIV assay for PT attraction

The guidance mechanism operating behind our SIV system differs from that in previous models designed for *Torenia*[26] and *Arabidopsis*.[27] While these aimed at observing micropylar guidance of a single ovule within a short distance between a PT and an ovule (∼100–150 µm), the current assay investigates guidance capacity of ovule clusters from a longer distance at ∼700 µm. These two types of SIV assays can be distinguished based on different PT behaviors in each system. In the single-ovule SIV assay, PTs make an abrupt turn toward the micropyle only when they are near an ovule. In our assay, PTs gradually change their trajectory toward the ovule clusters, and this attraction can be seen as soon as PTs exit the bottom-end of the style. Although it is quite possible that short-range diffusible signals are also secreted onto the medium when ovule exudates are collected, the underlying mechanism behind our SIV assay may point to a long-distance guidance mechanism.

4.6.3 The ovule secretome defines a microenvironment for pollen–pistil interactions before fertilization

In the ovular guidance stage, as soon as PTs exit the transmitting tract, they are in direct contact with the locular fluid produced by the placental tissues in the ovary. The carbohydrate composition of this fluid was studied previously in the angiosperm *Gasteria verrucosa* and includes fructose, sucrose, and glucose as the main components.[68] These carbohydrates are important for creating proper osmolarity for PT growth. In contrast, the protein composition had not been addressed. Profiling of the ovule secretome in *S. chacoense* provides the first description of the protein composition of the locular fluid, more specifically of ovular origin, based on the large difference observed between the mature vs. immature ovule secretomes. Our data suggests that proteins secreted in the locular fluid create a dynamic microenvironment that functions not only in PT growth and guidance but also in controlling late self-incompatibility, stress tolerance, and regulation of unconventional secretion from the ovules, to name a few. Here, we highlighted the importance of a few OSPs based on the different functional categories with which they are associated in pollen–pistil interactions.

a. Intra- and interspecific incompatibility

S-RNases are involved in gametophytic self-incompatibility in *Solanaceae*, *Scrophulariaceae*, and *Rosaceae*, preventing self-fertilization and inbreeding.[69] BLASTp search assigned

two putative orthologous S-RNases to OSP286: one from tobacco stigma exudates (S_{15}-RNase)[14] in the SEP database and one from potato style (S_2-RNase)[70] in the GSP database. OSP286 corresponds to the S_{14}-RNase of *S. chacoense*.[71] Although S-RNases are mostly expressed in the stylar transmitting tract where they exert their major function, i.e., self-PT growth cessation, in situ hybridization revealed that they are also expressed in the ovule integument and in the epidermal cells of the placenta that are continuous with the transmitting tract tissue,[72] a localization consistent with the ovule secretome. Furthermore, accumulation of OSP286 is developmentally regulated, being 11-fold more abundant at anthesis compared to 2DBA (Table S4.1, page 272), consistent with the accumulation of the S-RNases and HT-B modifier genes in the style where peak accumulation occurs at the anthesis stage.[71] We thus postulate that ovular-secreted S-RNases may act as a late checkpoint for self-PTs that would have survived their journey through the style. This is not uncommon since ribonuclease activity can be quite variable between different S-alleles and even from style to style for the same allele.[73]

OSP30 is similar to PELPIII, a class III pistil-specific extensin-like protein with arabinogalactan protein (AGP) properties initially isolated from tobacco pistil,[74] where it was found to be expressed in the transmitting tract of the style and further translocated in the growing PT cell wall.[75,76] More recently, PELPIII was also shown to be required for proper rejection of interspecific pollen.[77] Compared to PELP III, OSP30 is proline-rich (16 %) at the N-terminus, shares more than 50 % identity with PELPIII at the C-terminus, and is predicted to be heavily *O*-glycosylated.[78] AGPs have been detected in the ES of *A. thaliana* ovules[79] as well as in other reproductive tissues.[80] A member of the AGP family, the transmitting tissue-specific (TTS) protein is a PT growth simulant that also induces chemotactic behavior of SIV grown PTs in tobacco[81] and *Nicotiana alata*.[82] OSP30 is developmentally regulated, with a 7-fold increase during ES maturation between the 2DBA and anthesis stages. Apart from rejecting interspecific PTs, OSP30 might be involved in PT directional growth through its interaction with PTs.

b. PT guidance

OSP131 is a γ-amino butyric acid-transaminase (GABA-T), closely related to the *A. thaliana* GABA-T POP2, with 73.2 % amino acid sequence identity. GABA is a glutamate derivative whose accumulation leads to defective PT elongation in the transmitting tract via putative Ca^{2+}-permeable channels on the PT plasma membrane.[83,84] POP2 degrades GABA in the pistil and contributes to the formation of a GABA gradient from the stigma (lowest) to the inner integument (highest) and was shown to be important in leading the PTs to the micropyle.[31] However, compared to POP2, which bears a N-terminal mitochondrial target-

ing sequence, OSP131 lacks such a sequence, as predicted by both TargetP[85] and PSORT.[86] In contrast, OSP131 is predicted as a LSP by SecretomeP. Since POP2 is a single-copy gene in *Arabidopsis*, the *S. chacoense* ovule transcriptome was queried for other isoforms of OSP131 that might be targeted to the mitochondria. A mitochondrial isoform of GABA-T sharing 68.5 % amino acid identity with OSP131 was found, but it is not detected in the ovule secretome. Two other GABA-T isoforms sharing 79 and 93 % amino acid sequence identity with OSP131 were also detected in the transcriptome and predicted as LSPs. However, they were not detected in the ovule exudates. This suggests that in *S. chacoense* ovules OSP131 is delivered through the UPS pathway and might function in the extracellular matrix of the ovary to mediate a GABA gradient. Interestingly, solanaceous species like potato and tomato have both mitochondria-targeted and LSP GABA-T, whereas *Arabidopsis* has only the mitochondrial isoform (data not shown).

Besides GABA-T, four CRPs are also present in the ovule secretome. CRPs are of particular interest since short-distance guidance signaling involves several small secreted CRPs. The significance of diverse CRPs in pollen–pistil interaction has been recently reviewed in detail.[87,88,89] According to cysteine motifs classified previously from plants,[90] OSP21 (8-cys) belongs to the LTP subgroup of CRPs. OSP227 (6-cys) falls into the ECA1 gametogenesis-related family, whereas OSP305 (12-cys) is predicted to be a thionin-like protein. These three CRPs harbor a SP followed by a mature peptide of roughly 90 amino acids. Another thionin-like CRP, OSP123, is larger in size, with a predicted mature peptide of 133 amino acids and a peculiar organization of 18 cysteines comprising three blocks of six cysteines, of which the first two have an identical cysteine pattern, sharing 84 % sequence identity. Each block is preceded by a basic amino acid residue (lysine or arginine), suggesting the possibility of proteolytic processing through the presence of monobasic cleavage sites,[91] similar to what is observed in the BTH6 barley thionin.[92]

An orthology survey was conducted on these secreted CRPs in 10 related solanaceous species (Table S4.4, page 272). Data indicated that OSP21 (LTP) and OSP227 (ECA1) were both under positive selection ($d_N/d_S > 1$). No reciprocal best-BLAST hit was assigned to thionin-like OSP123 and OSP305, likely due to their high divergence. Such divergent CRPs could be involved in reproductive processes in a species-specific manner. In particular, the secretion of OSP227 is regulated in the mature ES and marked as an upregulated OSP in the ovule secretome (Table S4.1, page 272). In *Arabidopsis*, 119 ECA1 genes remain functionally unknown,[93] although subcellular localization of several ECA1 genes was associated with the synergid cells.[94,95] The putative ortholog of OSP227 was found only in *Solanum tarijense*, a

very closely related species to *S. chacoense*, among 10 other solanaceous species, suggesting that OSP227 is highly divergent and could function in a species-specific manner, as expected for short-distance guidance signals.[27,96]

c. Stress tolerance

OSP130 is predicted to be an isoflavone reductase-like (IFR-like) protein, one of the LSPs. OSP130 is 66 % identical (81 % similar) to the pistil-expressed and pollination-enhanced CP100, an IFR-like protein from *S. tuberosum*.[97] Such IFR-like proteins in maize and *Arabidopsis* were shown to be involved in defense against oxidative stress.[98,99] Transgenic plants overexpressing rice IFR-like gene exhibited growing tolerance to reactive oxygen species (ROS) in leaves and suspension-cultured cells.[100] In the context of ES development, the central cell was shown to be the main source of ROS before pollination. ROS were also shown to affect the central cell fate and were detected in synergid cells after pollination.[101] Furthermore, high levels of ROS under the control of the receptor-kinase FERONIA were observed at the female gametophyte entrance, where ROS were shown to mediate PT rupture in order to release its sperm cells.[102] Considering OSP130 secretion is developmentally upregulated by 6-fold in anthesis ovules compared to that in 2DBA (Table S4.1, page 272), it will be interesting to examine the localization of OSP130 in the ovule and explore its potential role in ROS regulation, ES development, and PT guidance.

d. Unconventional secretion

Four nOSPs were annotated as 14-3-3 proteins. The extracellular localization of 14-3-3's was confirmed by in situ immunolocalization in root border cells of pea and maize.[10] 14-3-3's are encoded by a gene family containing multi-isoforms.[103,104] One characteristic of 14-3-3's is their highly conserved core and divergent N- and C-terminus sequences, which possibly explains their organelle-specific functions.[105] 14-3-3 proteins are generally involved in signal transduction. They can modulate protein function through interaction with their phosphorylated binding partners and are therefore considered to be phospho-sensors. 14-3-3 proteins were also reported to stimulate unconventional secretion.[106] The presence of four distinct 14-3-3 proteins in our data set may relate to the high percentage of LSPs found in the ovule secretome. On the other hand, phosphorylation is one of the fastest responses to relay a stimulus to the cellular machinery. This post-translational modification could be employed by PTs in response to chemoattractant cues in order to promptly change their growth direction. It is pos-

sible that ovule-emitted 14-3-3's are used to modulate signal transduction of phosphorylated proteins in PTs once they are internalized. Of all of the 14-3-3 family OSPs, only the secretion of OSP23 is upregulated at anthesis, suggesting that it may be involved in reproductive-related processes.

4.6.4 ES developmental stage influences secretion status of 44 % of the ovule secretome

The developmental stage of the ES between FG6 and FG7 stages influenced the secretion of 44 % of total OSPs. Thus, the late transition from an ES that has unfused polar nuclei, immature filiform apparatus, and not fully degenerated antipodal cells to a fully mature ES leads to a drastic change in protein secretion. Since ovules at 2DBA are defective in PT attraction (Figure 4.2), these 128 OSPs could be involved in PT directional growth and guidance and thus provided a candidate pool for the study of isolating PT attractants. These OSPs may be directly secreted from the ES or secreted from the sporophytic tissue of the ovule, mainly the integument. Of major interest, this comparative analysis between mature and immature ovules revealed that, during the last maturation step of ES development, 106 of the 128 upregulated proteins are not regulated at the mRNA level (Figures 4.6 and 4.7), suggesting a late translation of the mRNAs that had accumulated at the 2DBA stage and/or the selective secretion of already-made proteins at anthesis stage, emphasizing the importance of post-transcriptional regulation in reproductive development.

4.7 Conclusions

The study of ovule secretome in *S. chacoense* reveals that various ovular signals are secreted to the ovarian locules, which form an active microenvironment to interact with approaching PTs. Comparative proteomic analysis allowed us to identify 128 secreted proteins potentially involved in pollen–ovule interactions, including PT guidance processes that are closely associated with the ES transition from the penultimate stage onward. Of these 128 upregulated OSPs from anthesis stage ovules, the majority was found to be unregulated at the mRNA level. This lack of correlation in gene expression and protein secretion suggests a strict regulation exerted from the ES over ovular signal secretion, vindicating this novel approach in the study of pollen–pistil interactions as a robust alternative to transcriptomic studies.

Bibliography

1. Inoue, N. ; Ikawa, M. ; Isotani, A. et Okabe, M. (2005). The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature*, 434(7030), 234–8. DOI: 10.1038/nature03362. [cit. p. 112]

2. Bianchi, E. ; Doe, B. ; Goulding, D. et Wright, G. J. (2014). Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature*, 508(7497), 483–7. DOI: 10.1038/nature13203. [cit. p. 112]

3. Floerl, S. ; Majcherczyk, A. ; Possienke, M. et coll. (2012). *Verticillium longisporum* infection affects the leaf apoplastic proteome, metabolome, and cell wall properties in *Arabidopsis thaliana*. *PLoS One*, 7(2), e31435. DOI: 10.1371/journal.pone.0031435. [cit. p. 112]

4. Goulet, C. ; Goulet, C. ; Goulet, M.-C. et Michaud, D. (2010). 2-DE proteome maps for the leaf apoplast of *Nicotiana benthamiana*. *Proteomics*, 10(13), 2536–44. DOI: 10.1002/pmic.200900382. [cit. p. 112]

5. Gupta, S. ; Wardhan, V. ; Verma, S. et coll. (2011). Characterization of the secretome of chickpea suspension culture reveals pathway abundance and the expected and unexpected secreted proteins. *J. Proteome Res.*, 10(11), 5006–15. DOI: 10.1021/pr200493d. [cit. p. 112]

6. Konozy, E. H. E. ; Rogniaux, H. ; Causse, M. et Faurobert, M. (2013). Proteomic analysis of tomato (*Solanum lycopersicum*) secretome. *J. Plant Res.*, 126(2), 251–66. DOI: 10.1007/s10265-012-0516-4. [cit. p. 112]

7. Pandey, A. ; Rajamani, U. ; Verma, J. et coll. (2010). Identification of extracellular matrix proteins of rice (*Oryza sativa* L.) involved in dehydration-responsive network: a proteomic approach. *J. Proteome Res.*, 9(7), 3443–64. DOI: 10.1021/pr901098p. [cit. p. 112]

8. Regente, M. ; Pinedo, M. ; Elizalde, M. et de la Canal, L. (2012). Apoplastic exosome-like vesicles: a new way of protein secretion in plants? *Plant Signal. Behav.*, 7(5), 544–6. DOI: 10.4161/psb.19675. [cit. p. 112]

9. Song, Y. ; Zhang, C. ; Ge, W. et coll. (2011). Identification of NaCl stress-responsive apoplastic proteins in rice shoot stems by 2D-DIGE. *J. Proteomics*, 74(7), 1045–67. DOI: 10.1016/j.jprot.2011.03.009. [cit. p. 112]

10. Wen, F. ; VanEtten, H. D. ; Tsaprailis, G. et Hawes, M. C. (2007). Extracellular proteins in pea root tip and border cell exudates. *Plant Physiol.*, 143(2), 773–83. DOI: 10.1104/pp.106.091637. [cit. p. 112, 113, 126, 134]

11. Zhou, L. ; Bokhari, S. A. ; Dong, C.-J. et Liu, J.-Y. (2011). Comparative proteomics analysis of the root apoplasts of rice seedlings in response to hydrogen peroxide. *PLoS One*, 6(2), e16723. DOI: 10.1371/journal.pone.0016723. [cit. p. 112]

12. Rejón, J. D. ; Delalande, F. ; Schaeffer-Reiss, C. et coll. (2014). The plant stigma exudate. A biochemically active extracellular environment for pollen germination? *Plant Signal. Behav.*, 9(4), e28274. DOI: 10.4161/psb.28274. [cit. p. 112, 126]

13. Rejón, J. D. ; Delalande, F. ; Schaeffer-Reiss, C. et coll. (2013). Proteomics profiling reveals novel proteins and functions of the plant stigma exudate. *J. Exp. Bot.*, 64(18), 5695–705. DOI: 10.1093/jxb/ert345. [cit. p. 112, 126]

14. Sang, Y. L. ; Xu, M. ; Ma, F. F. et coll. (2012). Comparative proteomic analysis reveals similar and distinct features of proteins in dry and wet stigmas. *Proteomics*, 12(12), 1983–98. DOI: 10.1002/pmic.201100407. [cit. p. 112, 123, 126, 132]

15. Hafidh, S. ; Potěšil, D. ; Fíla, J. et coll. (2014). In search of ligands and receptors of the pollen tube: the missing link in pollen tube perception. *Biochem. Soc. Trans.*, 42(2), 388–94. DOI: 10.1042/BST20130204. [cit. p. 112, 125, 126]

16. Prior, N. ; Little, S. A. ; Pirone, C. et coll. (2013). Application of proteomics to the study of pollination drops. *Appl. Plant Sci.*, 1(4), 1300008. DOI: 10.3732/apps.1300008. [cit. p. 112]

17. Gupta, R. et Deswal, R. (2012). Low temperature stress modulated secretome analysis and purification of antifreeze protein from *Hippophae rhamnoides*, a Himalayan wonder plant. *J. Proteome Res.*, 11(5), 2684–96. DOI: 10.1021/pr200944z. [cit. p. 113]

18. Badri, D. V. ; De-la Peña, C. ; Lei, Z. et coll. (2012). Root secreted metabolites and proteins are involved in the early events of plant-plant recognition prior to competition. *PLoS One*, 7(10), e46640. DOI: 10.1371/journal.pone.0046640. [cit. p. 113]

19. Higashiyama, T. ; Kuroiwa, H. et Kuroiwa, T. (2003). Pollen-tube guidance: beacons from the female gametophyte. *Curr. Opin. Plant Biol.*, 6(1), 36–41. DOI: 10.1016/S1369-5266(02)00010-9. [cit. p. 113]

20. Webb, M. C. et Williams, E. G. (1988). The pollen tube pathway in the pistil of *Lycopersicon peruvianum*. *Ann. Bot.*, 61(4), 415–23. DOI: 10.1093/oxfordjournals.aob.a087573. [cit. p. 113]

21. Cornish, E. C. ; Pettitt, J. M. ; Bonig, I. et Clarke, A. E. (1987). Developmentally controlled expression of a gene associated with self-incompatibility in *Nicotiana alata*. *Nature*, 326(6108), 99–102. DOI: 10.1038/326099a0. [cit. p. 113]

22. Dresselhaus, T. et Franklin-Tong, N. (2013). Male-female crosstalk during pollen germination, tube growth and guidance, and double fertilization. *Mol. Plant*, 6(4), 1018–36. DOI: 10.1093/mp/sst061. [cit. p. 113, 116]

23. Higashiyama, T. et Takeuchi, H. (2015). The mechanism and key molecules involved in pollen tube guidance. *Annu. Rev. Plant Biol.*, 66, 393–413. DOI: 10.1146/annurev-arplant-043014-115635. [cit. p. 113]

24. Hulskamp, M. ; Schneitz, K. et Pruitt, R. E. (1995). Genetic evidence for a long-range activity that directs pollen tube guidance in *Arabidopsis*. *Plant Cell*, 7(1), 57–64. DOI: 10.1105/tpc.7.1.57. [cit. p. 113, 116]

25. Lu, Y. ; Chanroj, S. ; Zulkifli, L. et coll. (2011). Pollen tubes lacking a pair of K^+ transporters fail to target ovules in *Arabidopsis*. *Plant Cell*, 23(1), 81–93. DOI: 10.1105/tpc.110.080499. [cit. p. 113]

26. Higashiyama, T. (1998). Guidance *in vitro* of the pollen tube to the naked embryo sac of *Torenia fournieri*. *Plant Cell*, 10(12), 2019–32. DOI: 10.1105/tpc.10.12.2019. [cit. p. 113, 131]

27. Palanivelu, R. et Preuss, D. (2006). Distinct short-range ovule signals attract or repel *Arabidopsis thaliana* pollen tubes *in vitro*. *BMC Plant Biol.*, 6, 7. DOI: 10.1186/1471-2229-6-7. [cit. p. 113, 131, 134]

28. Stewman, S. F. ; Jones-Rhoades, M. ; Bhimalapuram, P. et coll. (2010). Mechanistic insights from a quantitative analysis of pollen tube guidance. *BMC Plant Biol.*, 10, 32. DOI: 10.1186/1471-2229-10-32. [cit. p. 113]

29. Guan, Y. ; Lu, J. ; Xu, J. ; McClure, B. et Zhang, S. (2014). Two mitogen-activated protein kinases, MPK3 and MPK6, are required for funicular guidance of pollen tubes in *Arabidopsis*. *Plant Physiol.*, 165(2), 528–33. DOI: 10.1104/pp.113.231274. [cit. p. 113]

30. Lausser, A. ; Kliwer, I. ; Srilunchang, K.-O. et Dresselhaus, T. (2009). Sporophytic control of pollen tube growth and guidance in maize. *J. Exp. Bot.*, 61(3), 673–82. DOI: 10.1093/jxb/erp330. [cit. p. 116]

31. Palanivelu, R. ; Brass, L. ; Edlund, A. F. et Preuss, D. (2003). Pollen tube growth and guidance is regulated by *POP2*, an *Arabidopsis* gene that controls GABA levels. *Cell*, 114(1), 47–59. DOI: 10.1016/S0092-8674(03)00479-3. [cit. p. 116, 127, 132]

32. Ling, Y. ; Zhang, C. ; Chen, T. et coll. (2012). Mutation in SUMO E3 ligase, SIZ1, disrupts the mature female gametophyte in *Arabidopsis*. *PLoS One*, 7(1), e29470. DOI: 10.1371/journal.pone.0029470. [cit. p. 116]

33. Wang, H. ; Boavida, L. C. ; Ron, M. et McCormick, S. (2008). Truncation of a protein disulfide isomerase, PDIL2-1, delays embryo sac maturation and disrupts pollen tube guidance in *Arabidopsis thaliana*. *Plant Cell*, 20(12), 3300–11. DOI: 10.1105/tpc.108.062919. [cit. p. 116]

34. Higashiyama, T. ; Yabe, S. ; Sasaki, N. et coll. (2001). Pollen tube attraction by the synergid cell. *Science*, 293(5534), 1480–3. DOI: 10.1126/science.1062429. [cit. p. 116, 128]

35. Okuda, S. ; Tsutsui, H. ; Shiina, K. et coll. (2009). Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature*, 458(7236), 357–61. DOI: 10.1038/nature07882. [cit. p. 116]

36. Kanaoka, M. M. ; Kawano, N. ; Matsubara, Y. et coll. (2011). Identification and characterization of TcCRP1, a pollen tube attractant from *Torenia concolor*. *Ann. Bot.*, 108(4), 739–47. DOI: 10.1093/aob/mcr111. [cit. p. 116]

37. Márton, M. L. ; Cordts, S. ; Broadhvest, J. et Dresselhaus, T. (2005). Micropylar pollen tube guidance by Egg Apparatus 1 of maize. *Science*, 307(5709), 573–6. DOI: 10.1126/science.1104954. [cit. p. 116]

38. Takeuchi, H. et Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol.*, 10(12), e1001449. DOI: 10.1371/journal.pbio.1001449. [cit. p. 116]

39. Chevalier, É. ; Loubert-Hudon, A. ; Zimmerman, E. L. et Matton, D. P. (2011). Cell–cell communication and signalling pathways within the ovule: from its inception to fertilization. *New Phytol.*, 192(1), 13–28. DOI: 10.1111/j.1469-8137.2011.03836.x. [cit. p. 116]

40. Lafleur, E. ; Kapfer, C. ; Joly, V. et coll. (2015). The FRK1 mitogen-activated protein kinase kinase kinase (MAPKKK) from *Solanum chacoense* is involved in embryo sac and pollen development. *J. Exp. Bot.*, 66(7), 1833–43. DOI: 10.1093/jxb/eru524. [cit. p. 116]

41. Brewbacker, J. L. et Kwack, B. H. (1963). The essential role of calcium ion in pollen germination and pollen tube growth. *Am. J. Bot.*, 50(9), 859–965. DOI: 10.1002/j.1537-2197.1963.tb06564.x. [cit. p. 116]

42. Gray-Mitsumune, M. ; O'Brien, M. ; Bertrand, C. et coll. (2006). Loss of ovule identity induced by overexpression of the fertilization-related kinase 2 (ScFRK2), a MAPKKK from *Solanum chacoense*. *J. Exp. Bot.*, 57(15), 4171–87. DOI: 10.1093/jxb/erl194. [cit. p. 117]

43. Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, 72, 248–54. DOI: 10.1016/0003-2697(76)90527-3. [cit. p. 118]

44. Dorion, S. ; Parveen ; Jeukens, J. ; Matton, D. P. et Rivoal, J. (2005). Cloning and characterization of a cytosolic isoform of triosephosphate isomerase developmentally regulated in potato leaves. *Plant Sci.*, 168(1), 183–94. DOI: 10.1016/j.plantsci.2004.07.029. [cit. p. 118, 124]

45. Margulies, M. ; Egholm, M. ; Altman, W. E. et coll. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376–80. DOI: 10.1038/nature03959. [cit. p. 119]

46. Grabherr, M. G. ; Haas, B. J. ; Yassour, M. et coll. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.*, 29(7), 644–52. DOI: 10.1038/nbt.1883. [cit. p. 119]

47. Langmead, B. ; Trapnell, C. ; Pop, M. et Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, 10(3), R25. DOI: 10.1186/gb-2009-10-3-r25. [cit. p. 119]

48. Li, B. et Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinf.*, 12, 323. DOI: 10.1186/1471-2105-12-323. [cit. p. 119]

49. Robinson, M. D. ; McCarthy, D. J. et Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–40. DOI: 10.1093/bioinformatics/btp616. [cit. p. 119]

50. Petersen, T. N. ; Brunak, S. ; von Heijne, G. et Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8(10), 785–6. DOI: 10.1038/nmeth.1701. [cit. p. 120]

51. Bendtsen, J. D. ; Jensen, L. J. ; Blom, N. ; Von Heijne, G. et Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng., Des. Sel.*, 17(4), 349–56. DOI: 10.1093/protein/gzh037. [cit. p. 120]

52. Altschul, S. F. ; Gish, W. ; Miller, W. ; Myers, E. W. et Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215(3), 403–10. DOI: 10.1016/S0022-2836(05)80360-2. [cit. p. 120]

53. Conesa, A. ; Götz, S. ; García-Gómez, J. M. et coll. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674–6. DOI: 10.1093/bioinformatics/bti610. [cit. p. 120]

54. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, 24(8), 1586–91. DOI: 10.1093/molbev/msm088. [cit. p. 121]

55. Goldberg, E. E. ; Kohn, J. R. ; Lande, R. et coll. (2010). Species selection maintains self-incompatibility. *Science*, 330(6003), 493–5. DOI: 10.1126/science.1194513. [cit. p. 121]

56. Terry, M. E. et Bonner, B. A. (1980). An examination of centrifugation as a method of extracting an extracellular solution from peas, and its use for the study of indoleacetic acid-induced growth. *Plant Physiol.*, 66(2), 321–5. DOI: 10.1104/pp.66.2.321. [cit. p. 122]

57. Jung, Y.-H. ; Jeong, S.-H. ; Kim, S. H. et coll. (2008). Systematic secretome analyses of rice leaf and seed callus suspension-cultured cells: workflow development and establishment of high-density two-dimensional gel reference maps. *J. Proteome Res.*, 7(12), 5187–210. DOI: 10.1021/pr8005149. [cit. p. 122, 127]

58. The Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. *Nature*, 475(7355), 189–95. DOI: 10.1038/nature10158. [cit. p. 124]

59. The Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485(7400), 635–41. DOI: 10.1038/nature11119. [cit. p. 124]

60. Cheng, F.-Y. ; Blackburn, K. ; Lin, Y.-M. ; Goshe, M. B. et Williamson, J. D. (2009). Absolute protein quantification by LC/MS(E) for global analysis of salicylic acid-induced plant protein secretion responses. *J. Proteome Res.*, 8(1), 82–93. DOI: 10.1021/pr800649s. [cit. p. 125]

61. Cho, W. K. ; Chen, X. Y. ; Chu, H. et coll. (2009). Proteomic analysis of the secretome of rice calli. *Physiol. Plant.*, 135(4), 331–41. DOI: 10.1111/j.1399-3054.2008.01198.x. [cit. p. 125]

62. Agrawal, G. K. ; Jwa, N.-S. ; Lebrun, M.-H. ; Job, D. et Rakwal, R. (2010). Plant secretome: unlocking secrets of the secreted proteins. *Proteomics*, 10(4), 799–827. DOI: 10.1002/pmic.200900514. [cit. p. 126]

63. Lum, G. ; Meinken, J. ; Orr, J. ; Frazier, S. et Min, X. J. (2014). PlantSecKB: the plant secretome and subcellular proteome KnowledgeBase. *Comput. Mol. Biol.*, 4(1), 1–17. DOI: 10.5376/cmb.2014.04.0001. [cit. p. 126]

64. Kim, S. G. ; Kim, S. T. ; Wang, Y. et coll. (2010). Overexpression of rice isoflavone reductase-like gene (OsIRL) confers tolerance to reactive oxygen species. *Physiol. Plant.*, 138(1), 1–9. DOI: 10.1111/j.1399-3054.2009.01290.x. [cit. p. 126]

65. Matton, D. P. ; Maes, O. ; Laublin, G. et coll. (1997). Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. *Plant Cell*, 9(10), 1757–66. DOI: 10.1105/tpc.9.10.1757. [cit. p. 127]

66. Chevalier, E. ; Loubert-Hudon, A. et Matton, D. P. (2013). ScRALF3, a secreted RALF-like peptide involved in cell-cell communication between the sporophyte and the female gametophyte in a solanaceous species. *Plant J.*, 73(6), 1019–33. DOI: 10.1111/tpj.12096. [cit. p. 128]

67. Kasahara, R. D. ; Portereiko, M. F. ; Sandaklie-Nikolova, L. ; Rabiger, D. S. et Drews, G. N. (2005). *MYB98* is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell*, 17(11), 2981–92. DOI: 10.1105/tpc.105.034603. [cit. p. 128]

68. Willemse, M. T. M. et Franssen-Verheijen, M. A. W. (1988). Ovular development and pollen tube growth in the ovary of *Gasteria verrucosa* (Mill.) H. Duval as condition for fertilization. *In* Cresti, M. ; Gori, P. et Pacini, E., éditeurs : *Sexual Reproduction in Higher Plants*, pages 357–62. Springer, Berlin, Allemagne. ISBN: 978-3-642-73271-3. DOI: 10.1007/978-3-642-73271-3_57. [cit. p. 131]

69. McClure, B. ; Cruz-García, F. et Romero, C. (2011). Compatibility and incompatibility in S-RNase-based systems. *Ann. Bot.*, 108(4), 647–58. DOI: 10.1093/aob/mcr179. [cit. p. 131]

70. Kaufmann, H. ; Salamini, F. et Thompson, R. D. (1991). Sequence variability and gene structure at the self-incompatibility locus of *Solanum tuberosum*. *Mol. Gen. Genetics*, 226(3), 457–66. DOI: 10.1007/BF00260659. [cit. p. 132]

71. O'Brien, M. ; Kapfer, C. ; Major, G. et coll. (2002). Molecular analysis of the stylar-expressed *Solanum chacoense* small asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility in *Nicotiana*. *Plant J.*, 32(6), 985–96. DOI: 10.1046/j.1365-313X.2002.01486.x. [cit. p. 132]

72. Matton, D. P. ; Bertrand, C. ; Laublin, G. et Cappadocia, M. (1998). Molecular aspects of self-incompatibility in tuber-bearing *Solanum* species. *In* Khurana, S. M. P. ; Chandra, R. et Upadhya, M. D., éditeurs : *Comprehensive potato biotechnology*, pages 97–113. Malhotra Publishing House, New Delhi, Inde. ISBN: 978-81-85048-39-0. [cit. p. 132]

73. Qin, X. ; Liu, B. ; Soulard, J. ; Morse, D. et Cappadocia, M. (2006). Style-by-style analysis of two sporadic self-compatible *Solanum chacoense* lines supports a primary role for S-RNases in determining pollen rejection thresholds. *J. Exp. Bot.*, 57(9), 2001–13. DOI: 10.1093/jxb/erj147. [cit. p. 132]

74. Goldman, M. H. ; Pezzotti, M. ; Seurinck, J. et Mariani, C. (1992). Developmental expression of tobacco pistil-specific genes encoding novel extensin-like proteins. *Plant Cell*, 4(9), 1041–51. DOI: 10.1105/tpc.4.9.1041. [cit. p. 132]

75. Bosch, M. ; Knudsen, J. S. ; Derksen, J. et Mariani, C. (2001). Class III pistil-specific extensin-like proteins from tobacco have characteristics of arabinogalactan proteins. *Plant Physiol.*, 125(4), 2180–8. DOI: 10.1104/pp.125.4.2180. [cit. p. 132]

76. de Graaf, B. H. J. ; Knuiman, B. A. ; Derksen, J. et Mariani, C. (2003). Characterization and localization of the transmitting tissue-specific PELPIII proteins of *Nicotiana tabacum*. *J. Exp. Bot.*, 54(380), 55–63. DOI: 10.1093/jxb/erg002. [cit. p. 132]

77. Eberle, C. A. ; Anderson, N. O. ; Clasen, B. M. ; Hegeman, A. D. et Smith, A. G. (2013). PELPIII: the class III pistil-specific extensin-like *Nicotiana tabacum* proteins are essential for interspecific incompatibility. *Plant J.*, 74(5), 805–14. DOI: 10.1111/tpj.12163. [cit. p. 132]

78. Steentoft, C. ; Vakhrushev, S. Y. ; Joshi, H. J. et coll. (2013). Precision mapping of the human *O*-GalNAc glycoproteome through SimpleCell technology. *EMBO J.*, 32(10), 1478–88. DOI: 10.1038/emboj.2013.79. [cit. p. 132]

79. Coimbra, S. ; Almeida, J. ; Junqueira, V. ; Costa, M. L. et Pereira, L. G. (2007). Arabinogalactan proteins as molecular markers in *Arabidopsis thaliana* sexual reproduction. *J. Exp. Bot.*, 58(15-16), 4027–35. DOI: 10.1093/jxb/erm259. [cit. p. 132]

80. Pereira, A. M. ; Masiero, S. ; Nobre, M. S. et coll. (2014). Differential expression patterns of arabinogalactan proteins in *Arabidopsis thaliana* reproductive tissues. *J. Exp. Bot.*, 65(18), 5459–71. DOI: 10.1093/jxb/eru300. [cit. p. 132]

81. Cheung, A. Y. ; Wang, H. et Wu, H. M. (1995). A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. *Cell*, 82(3), 383–93. DOI: 10.1016/0092-8674(95)90427-1. [cit. p. 132]

82. Wu, H.-M. ; Wong, E. ; Ogdahl, J. et Cheung, A. Y. (2000). A pollen tube growth-promoting arabinogalactan protein from *Nicotiana alata* is similar to the tobacco TTS protein. *Plant J.*, 22 (2), 165–76. DOI: 10.1046/j.1365-313x.2000.00731.x. [cit. p. 132]

83. Renault, H. ; El Amrani, A. ; Palanivelu, R. et coll. (2011). GABA accumulation causes cell elongation defects and a decrease in expression of genes encoding secreted and cell wall-related proteins in *Arabidopsis thaliana*. *Plant Cell Physiol.*, 52(5), 894–908. DOI: 10.1093/pcp/pcr041. [cit. p. 132]

84. Yu, G.-H. ; Zou, J. ; Feng, J. et coll. (2014). Exogenous γ-aminobutyric acid (GABA) affects pollen tube growth via modulating putative Ca^{2+}-permeable membrane channels and is coupled to negative regulation on glutamate decarboxylase. *J. Exp. Bot.*, 65(12), 3235–48. DOI: 10.1093/jxb/eru171. [cit. p. 132]

85. Emanuelsson, O. ; Nielsen, H. ; Brunak, S. et von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.*, 300(4), 1005–16. DOI: 10.1006/jmbi.2000.3903. [cit. p. 133]

86. Nakai, K. et Horton, P. (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.*, 24(1), 34–6. DOI: 10.1016/S0968-0004(98)01336-X. [cit. p. 133]

87. Chae, K. et Lord, E. M. (2011). Pollen tube growth and guidance: roles of small, secreted proteins. *Ann. Bot.*, 108(4), 627–36. DOI: 10.1093/aob/mcr015. [cit. p. 133]

88. Higashiyama, T. (2010). Peptide signaling in pollen-pistil interactions. *Plant Cell Physiol.*, 51(2), 177–89. DOI: 10.1093/pcp/pcq008. [cit. p. 133]

89. Marshall, E. ; Costa, L. M. et Gutierrez-Marcos, J. (2011). Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J. Exp. Bot.*, 62 (5), 1677–86. DOI: 10.1093/jxb/err002. [cit. p. 133]

90. Silverstein, K. A. T. ; Moskal, W. A. ; Wu, H. C. et coll. (2007). Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J.*, 51(2), 262–80. DOI: 10.1111/j.1365-313X.2007.03136.x. [cit. p. 133]

91. Rholam, M. et Fahy, C. (2009). Processing of peptide and hormone precursors at the dibasic cleavage sites. *Cell. Mol. Life Sci.*, 66(13), 2075–91. DOI: 10.1007/s00018-009-0007-5. [cit. p. 133]

92. Plattner, S. ; Gruber, C. ; Stadlmann, J. et coll. (2015). Isolation and characterization of a thionin proprotein-processing enzyme from barley. *J. Biol. Chem.*, 290(29), 18056–67. DOI: 10.1074/jbc.M115.647859. [cit. p. 133]

93. Sprunck, S. ; Hackenberg, T. ; Englhart, M. et Vogler, F. (2014). Same same but different: sperm-activating EC1 and ECA1 gametogenesis-related family proteins. *Biochem. Soc. Trans.*, 42(2), 401–7. DOI: 10.1042/BST20140039. [cit. p. 133]

94. Jones-Rhoades, M. W. ; Borevitz, J. O. et Preuss, D. (2007). Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins. *PLoS Genet.*, 3 (10), 1848–61. DOI: 10.1371/journal.pgen.0030171. [cit. p. 133]

95. Steffen, J. G. ; Kang, I.-H. ; Macfarlane, J. et Drews, G. N. (2007). Identification of genes expressed in the *Arabidopsis* female gametophyte. *Plant J.*, 51(2), 281–92. DOI: 10.1111/j.1365-313X.2007.03137.x. [cit. p. 133]

96. Higashiyama, T. ; Inatsugi, R. ; Sakamoto, S. et coll. (2006). Species preferentiality of the pollen tube attractant derived from the synergid cell of *Torenia fournieri*. *Plant Physiol.*, 142(2), 481–91. DOI: 10.1104/pp.106.083832. [cit. p. 134]

97. van Eldik, G. J. ; Ruiter, R. K. ; Colla, P. H. et coll. (1997). Expression of an isoflavone reductase-like gene enhanced by pollen tube growth in pistils of *Solanum tuberosum*. *Plant Mol. Biol.*, 33 (5), 923–9. DOI: 10.1023/A:1005749913203. [cit. p. 134]

98. Babiychuk, E. ; Kushnir, S. ; Belles-Boix, E. ; Van Montagu, M. et Inzé, D. (1995). *Arabidopsis thaliana* NADPH oxidoreductase homologs confer tolerance of yeasts toward the thiol-oxidizing drug diamide. *J. Biol. Chem.*, 270(44), 26224–31. DOI: 10.1074/jbc.270.44.26224. [cit. p. 134]

99. Petrucco, S. (1996). A maize gene encoding an NADPH binding enzyme highly homologous to isoflavone reductases is activated in response to sulfur starvation. *Plant Cell*, 8(1), 69–80. DOI: 10.1105/tpc.8.1.69. [cit. p. 134]

100. Kim, S. ; Mollet, J.-C. ; Dong, J. et coll. (2003). Chemocyanin, a small basic protein from the lily stigma, induces pollen tube chemotropism. *Proc. Natl. Acad. Sci. U. S. A.*, 100(26), 16125–30. DOI: 10.1073/pnas.2533800100. [cit. p. 134]

101. Martin, M. V. ; Fiol, D. F. ; Sundaresan, V. ; Zabaleta, E. J. et Pagnussat, G. C. (2013). *oiwa*, a female gametophytic mutant impaired in a mitochondrial manganese-superoxide dismutase, reveals crucial roles for reactive oxygen species during embryo sac development and fertilization in *Arabidopsis*. *Plant Cell*, 25(5), 1573–91. DOI: 10.1105/tpc.113.109306. [cit. p. 134]

102. Duan, Q. ; Kita, D. ; Johnson, E. A. et coll. (2014). Reactive oxygen species mediate pollen tube rupture to release sperm for fertilization in *Arabidopsis*. *Nat. Commun.*, 5, 3129. DOI: 10.1038/ncomms4129. [cit. p. 134]

103. Wu, K. ; Rooney, M. F. et Ferl, R. J. (1997). The *Arabidopsis* 14-3-3 multigene family. *Plant Physiol.*, 114(4), 1421–31. DOI: 10.1104/pp.114.4.1421. [cit. p. 134]

104. DeLille, J. M. ; Sehnke, P. C. et Ferl, R. J. (2001). The *Arabidopsis* 14-3-3 family of signaling regulators. *Plant Physiol.*, 126(1), 35–8. DOI: 10.1104/pp.126.1.35. [cit. p. 134]

105. Chung, H.-J. ; Sehnke, P. C. et Ferl, R. J. (1999). The 14-3-3 proteins: cellular regulators of plant metabolism. *Trends Plant Sci.*, 4(9), 367–71. DOI: 10.1016/S1360-1385(99)01462-4. [cit. p. 134]

106. Carreño, F. R. ; Goñi, C. N. ; Castro, L. M. et Ferro, E. S. (2005). 14-3-3 epsilon modulates the stimulated secretion of endopeptidase 24.15. *J. Neurochem.*, 93(1), 10–25. DOI: 10.1111/j.1471-4159.2004.02967.x. [cit. p. 134]

CHAPITRE 5

Réponse transcriptionnelle à distance de l'ovaire à la pollinisation

Ce chapitre a été publié le 24 juin 2019 sous le titre *Pollination type recognition from a distance by the ovary is revealed through a global transcriptomic analysis* dans le périodique *Plants*, 8(6), 185. Il est accessible en ligne à https://doi.org/10.3390/plants8060185.

Auteurs Valentin Joly*, Faïza Tebbji*, André Nantel et Daniel P. Matton (*co-premiers auteurs)

Contributions Valentin Joly est co-premier auteur avec Faïza Tebbji ; leurs contributions au travail présenté sont équivalentes. FaïzaTebbji s'est chargée des travaux expérimentaux au laboratoire, notamment de la conception de la bio-puce sur laquelle repose cet article, de la génération des banques d'ADNc à partir des échantillons d'intérêt et des expériences d'hybridation, ainsi que d'analyses préliminaires des données d'expression différentielle. Valentin Joly a repris les analyses quantitatives d'expression différentielle, effectué les analyses d'orthologie, d'annotation structurale et fonctionnelle, de *clustering*, et s'est chargé du commentaire critique de ces résultats. Valentin Joly et Faïza Tebbji ont contribué à parts égales à la rédaction du manuscrit. André Nantel a fourni les ressources techniques et méthodologiques nécessaires aux expériences d'hybridation. Daniel P. Matton a contribué à la conception du travail, et s'est chargé de sa supervision et de la révision critique du manuscrit.

5.1 Résumé

La reproduction sexuée des plantes à fleurs implique des contacts étroits et des interactions continues entre le tube pollinique en croissance et les structures reproductrices femelles. Ces interactions peuvent déclencher des réponses dans les régions distales de la fleur bien avant la fécondation. Alors que la sénescence des pétales induite par la pollinisation a été largement étudiée, on en sait beaucoup moins sur la façon dont la pollinisation est perçue à distance dans l'ovaire, et sur le degré de spécificité de cette réponse à différents génotypes polliniques. Pour répondre à cette question, nous avons effectué une analyse transcriptomique globale de l'ovaire d'une espèce sauvage de pomme de terre, *Solanum chacoense*, à différents temps après des pollinisations compatible, incompatible et hétérospécifique. Dans tous les cas, la pénétration des tubes polliniques dans le stigmate a été initialement perçue comme une blessure. Par la suite, alors que la croissance des tubes polliniques continuaient leur croissance dans le style, un nombre croissant de gènes sont devenus spécifiques à chaque génotype pollinique. Des analyses de classification fonctionnelle ont révélé des différences nettes entre la réponse aux pollinisations compatible et hétérospécifique. Par example, la première a induit des gènes reliés aux espèces réactives de l'oxygène (ROS) alors que la seconde a affecté des gènes associés à la signalisation à l'éthylène. À l'opposé, la réponse à la pollinisation incompatible est restée plus proche d'un stress de blessure. Notre analyse montre que chaque type de pollinisation produit une signature moléculaire spécifique générant à distance des réponses diverses et spécifiques dans l'ovaire en préparation pour la fécondation.

Mots-clés Signalisation à longue distance, interactions pollen–pistil, signatures moléculaires associées au pollen, barrières d'isolement reproductif postpollinisation.

5.2 Abstract

Sexual reproduction in flowering plants involves intimate contact and continuous interactions between the growing pollen tube and the female reproductive structures. These interactions can trigger responses in distal regions of the flower well ahead of fertilization. While pollination-induced petal senescence has been studied extensively, less is known about how pollination is perceived at a distance in the ovary, and how specific this response is to various pollen genotypes. To address this question, we performed a global transcriptomic analysis in the ovary of a wild potato species, *Solanum chacoense*, at various time points following compatible, incompatible, and heterospecific pollinations. In all cases, pollen tube penetration in

the stigma was initially perceived as a wounding aggression. Then, as the pollen tubes grew in the style, a growing number of genes became specific to each pollen genotype. Functional classification analyses revealed sharp differences in the response to compatible and heterospecific pollinations. For instance, the former induced reactive oxygen species (ROS)-related genes while the latter affected genes associated to ethylene signaling. In contrast, incompatible pollination remained more akin to a wound response. Our analysis reveals that every pollination type produces a specific molecular signature generating diversified and specific responses at a distance in the ovary in preparation for fertilization.

Keywords Long distance signalling, pollen–pistil interactions, pollen-associated molecular signatures, postmating isolation barriers.

5.3 Introduction

In Angiosperms, sexual reproduction is initiated by pollen landing on the stigma papillae. After hydration, pollen grains produce a pollen tube (PT) that grows through the internal tissue of the carpel, guided by physical as well as chemotropic cues originating from both the style and the ovary to finally deliver its two sperm cells to the female gametophyte.[1] One sperm cell fuses with the egg cell forming the zygote while the second sperm cell fuses with the central cell to form the endosperm that surrounds and provides nutrients to the developing embryo. From the onset of pollen grains landing on a receptive stigma surface until effective fertilization, multiple interactions are initiated and a complex and intricate cross talk between the pollen and the pistil is established.[2] The decision to accept or reject the pollen starts with pollen capture and adhesion, followed by pollen hydration and germination. At this stage, pollen grains might already be rejected, as found in species expressing sporophytic self-compatibility (SI), like in the *Brassicaceae* family.[3] In species expressing gametophytic SI systems, like in the *Papaveraceae* and in the *Solanaceae*, PT recognition and rejection occurs either soon after pollen germination[4] or later on during PT growth in the transmitting tissue of the style,[5] respectively. Being highly specialized structures, pollen,[6,7,8,9,10] stigma/style,[11,12,13,14] ovary[15,16,17,18,19] as well as individual cells within the ovule[20,21] express a specific transcriptome. During pollen–pistil interactions, continuous intimate contact and concomitant signal exchanges are bound to further modulate these transcriptomes. Since the first large-scale report of pistil-induced gene expression in the PT by Qin *et al.* in 2009,[22] several studies have investigated transcriptional changes taking place in pollinated pistils in the context of compatible,[23,24,25] incompatible,[26,27,28] and interspecific crosses.[29,30,31,32,33]

In the abovementioned studies, the modulated genes were isolated from the tissues in direct contact, however there is still the question of what goes on in distal structures before PTs reach the ovary. Long-distance signaling during plant reproduction was described almost 150 years ago with the discovery of pollination-induced ovule maturation in orchid species,[34,35,36] which was later found to be associated with interorgan ethylene signaling.[37,38] Pollination was also shown to be required to complete female gametophyte development in other species such as almond,[39] maize,[40] and tobacco.[41] Moreover, pollination is known to trigger several other physiological responses in the flower,[42] including petal senescence in orchids[43] and *Petunia*,[44] or changes in floral scent, for example in thistles.[45] Again, ethylene was identified or suggested as the mediator of this long-distance signaling.[42]

Such responses in distal organs require the modulation of genes at a distance following pollination, well before PTs reach the ovules. Indeed, several studies revealed that pollination induces the expression of ethylene biosynthesis genes in the flowers of orchids,[38,46] tomato,[47] and tobacco.[41] Moreover, Lantin *et al.* showed that the *SPP2* gene from the wild potato species *Solanum chacoense*, which encodes a dioxygenase, is activated at a distance in the ovary by both compatible pollination and stigma wounding.[48] This first observation on a single gene prompted us to expand the analysis and explore the global transcriptional response of *S. chacoense* ovaries to pollination. Although comprehensive transcriptomic studies were performed recently on pollination-induced responses in corollas,[49,50] no large-scale study has yet addressed the specific issue of long-distance communication between growing PTs and ovules.

In this work, we set out to understand how precisely the ovary can interpret pollination from a distance in preparation for fertilization, and how specific this response is to various pollination types. To address these questions, we have used a global transcriptomic approach to monitor gene expression profiles in *S. chacoense* ovaries at different times following conspecific compatible (CCP), conspecific incompatible (CIP), and heterospecific compatible (HCP) pollinations as well as from stigma wounding and touch treatments.

5.4 Results & Discussion

5.4.1 Experimental Design

We used an ovule cDNA microarray consisting of 7741 sequences representing 6374 unigenes[19] to globally analyze the ovule transcriptomic response following pollination. The microarray included cDNAs from various developmental stages, from unpollinated ovules (UOs) to fertilized ovules until late torpedo stage embryos, sequenced in the form of expressed sequence tags (ESTs).[51]

Since gametophytic SI is an important PT rejection mechanism in our model species, involving pollination-induced regulation of pistil transcripts,[52,53,54] we chose to compare the effect of CCP and CIP on gene expression in the ovary. Moreover, the existence of cross-incompatibility barriers affecting pollen–pistil interactions in wild potatoes[55] led us to include a HCP condition in our study. To minimize incongruity problems, we chose to perform HCP with pollen from a closely related self-incompatible species, *S. microdontum*, which was previously shown to make fertile hybrids with *S. chacoense*.[56] Finally, to investigate the possible relationship between PT perception and wound or mechanical stress, we also included a stigma wounding (W) and a touch (T) condition in our design, the latter involving mock pollinations made with inert zirconia/silica microbeads.

To choose the best time points for the analysis, PT growth kinetics were monitored in vivo (Section 5.4.1). The three PT types germinated equally and had undistinguishable growth until 12 h after pollination (HAP), where they all reached ∼2 mm in length. To determine if the ovary could accurately discriminate between pollination types before any visible difference in PT growth, 6 HAP was chosen as the first time point. Then, after growing slowly (∼170 $\mu m\, h^{-1}$) until they emerged from the stigma 12 HAP, CCP PTs dramatically sped up (∼330 $\mu m\, h^{-1}$) to finally exit the style around 30 HAP. This biphasic growth pattern is characteristic for species that shed bicellular pollen (containing a vegetative and a generative cell) like solanaceous species.[57] The first phase, termed the autotrophic phase, is characterized by a period of slow growth where PTs rely on their stored reserves. Next, the heterotrophic phase is characterized by a faster growth rate, the pollen being fed by nutrient made available from the stylar transmitting tissue. In *S. chacoense*, PTs normally reach the first available ovules a few hours later to effect fertilization.[58]

Interestingly, CIP and HCP PTs displayed a steady but slower monophasic growth pattern.

CIP PTs were all stopped by the SI reaction before they reached mid-style, whereas HCP PTs faced suboptimal growth in the heterospecific style, which is often described as incongruity.[59] A control pollination in *S. microdontum* (Section 5.4.1, light gray line) confirmed that HCP PTs grow faster in their conspecific pistils. Since most CIP PTs were already arrested 24 HAP while CCP PTs had not yet reached the ovary, this was chosen as the second reference time point. Finally, in case interorgan signaling needed more time to be detected, a late time point, 48 HAP, was also selected to determine late pollination effects, especially for CIP and HCP. The same time points (6, 24, and 48 h) were used to examine transcript regulation after the wounding and touch treatments.

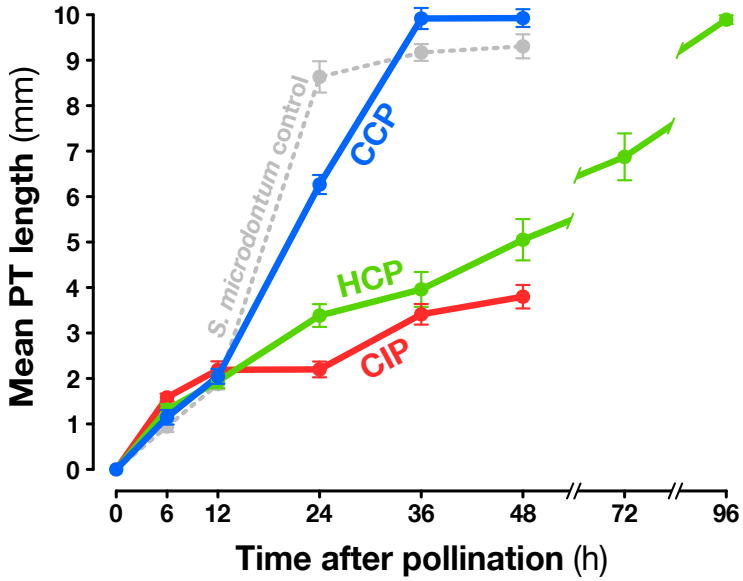


Figure 5.1. Pollen tube (PT) growth kinetics after conspecific compatible (CCP), conspecific incompatible (CIP), and heterospecific compatible (HCP) pollinations. PT length was measured after aniline blue staining of *S. chacoense* G4 pistils pollinated with *S. chacoense* G4 (CIP, red) and V22 (CCP, blue), as well as *S. microdontum* (HCP, green). The light gray line represents PT kinetics after intraspecific *S. microdontum* pollination.

5.4.2 Expression Profiling of Pollination-Responsive Genes

For each time point in each pollination condition, four ovule samples were collected from a large number of plants grown in the same greenhouse. After RNA extraction and cDNA library construction, half of these biological replicates were labeled with Cy3 and the other half with Cy5 to account for the possibility of dye bias. Following the procedure used by

Tebbji *et al.*,[19] competitive hybridizations were made against the same pooled control obtained from different UO replicates. To confirm the reliability of this approach, six additional control hybridizations were made with individual UO replicates against the pooled control (Dataset S5.1a).

The exploratory nature of our study led us to opt for relaxed criteria to select regulated genes: Transcripts showing a greater than ± 1.5-fold expression difference between test and control hybridizations with $P \leq 0.05$ were retained for further analysis (Dataset S5.1b–h). In the end, 1441 ovary transcripts showed a significant change in abundance in at least one pollination condition, with 163, 598, and 1184 of them regulated 6, 24, and 48 HAP, respectively (Figure 5.2). To investigate how regulated transcripts behaved across the different conditions under study, we employed a dual approach involving *k*-means hierarchical clustering (Section 5.4.2, Dataset S5.1i) and Venn diagrams comparing pollinations at each time-point (Section 5.4.2) and vice versa (Section 5.4.2). Statistics about correlations and coregulations between conditions are shown in Figure S5.1 and Table S5.1, respectively.

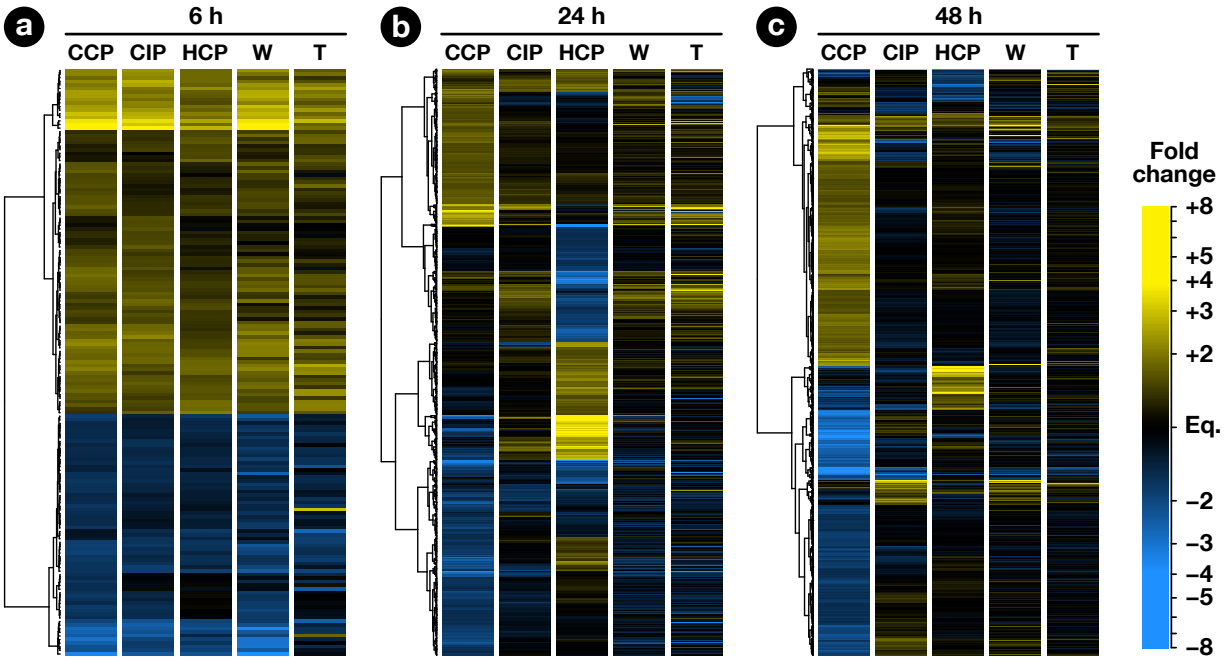


Figure 5.2. Hierarchical clustering analysis of regulated genes. Each row represents a gene and each column represents a condition. At each time point, the analysis was performed using genes regulated in at least one pollination condition. Euclidean distances between expression ratios in CCP, CIP and HCP vs. unpollinated ovule (UO) were used in a hierarchial clustering analysis based on Ward's method. Expression ratios in stigma wounding (W) and touch (T) conditions vs. UO were then added to the heatmap.

Several analyses were performed to better understand the potential functions of regulated

transcripts (Dataset S5.2). First, we proceeded with BLASTn and BLASTx searches against the National Center for Biotechnology Information (NCBI) RefSeq database to find potential homologs in other species and give descriptions to our transcripts (Dataset S5.3), and then performed a functional classification into GO (*Gene Ontology*) categories and subsequent enrichment analyses (Datasets S4 and S5). Finally, we used the closest BLASTx hit of each EST to perform a variety of *in silico* predictions, in particular putative transcription factors (Section 5.4.3 and Figure S5.2, Table S5.2) and secreted proteins (Section 5.4.3 and Figure S5.3, Table S5.3), as well as metabolic enzymes (Table S5.4).

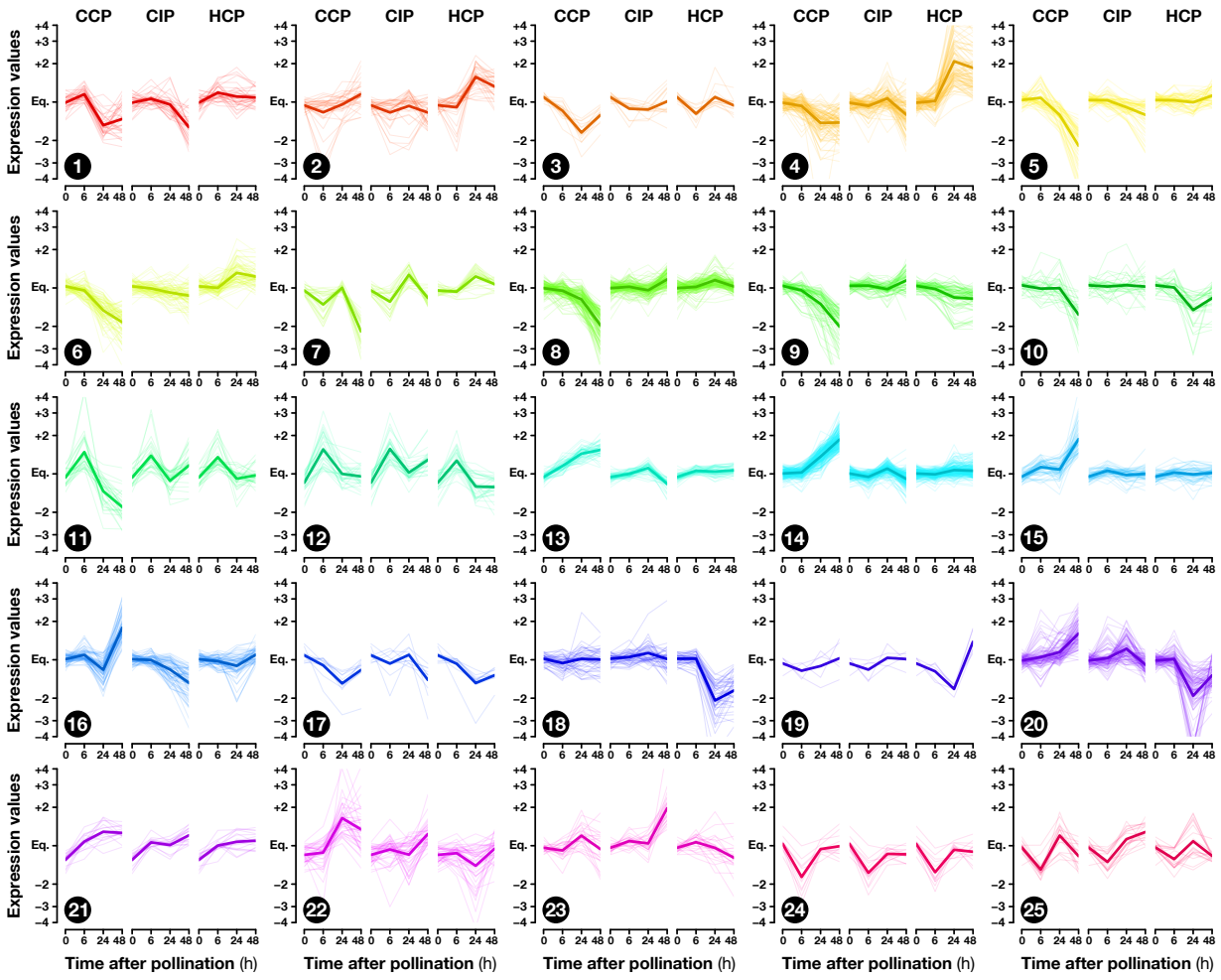


Figure 5.3. Transcription expression profiles in the 25 clusters obtained by *k*-means clustering.

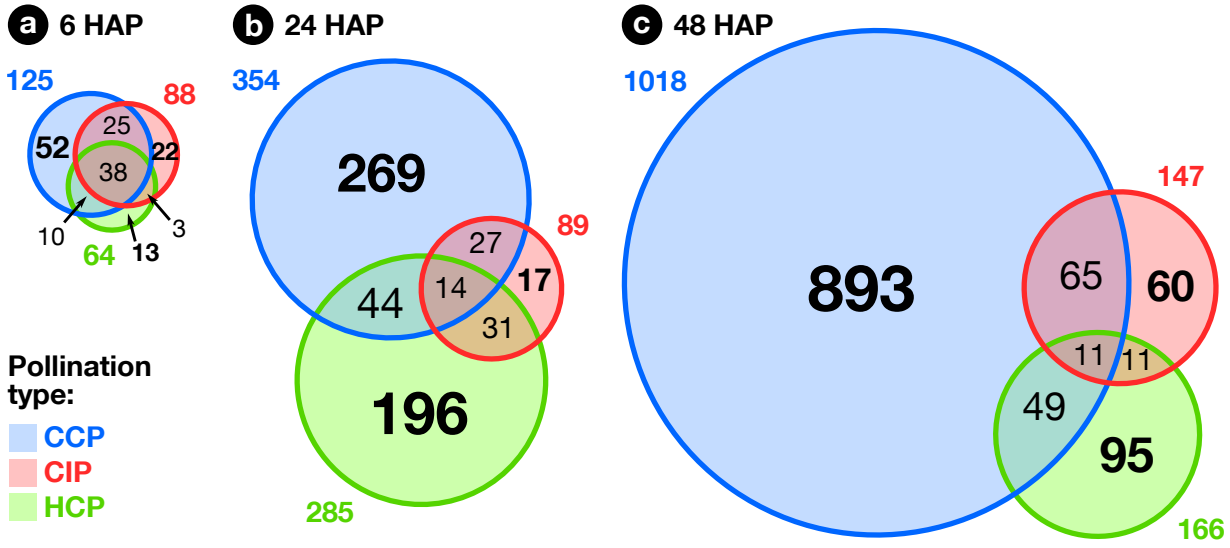


Figure 5.4. Overlap between pollination responses. Venn diagrams showing the overlap between CCP-, CIP-, and HCP-regulated genes at 6, 24 and 48 h after pollination (HAP).

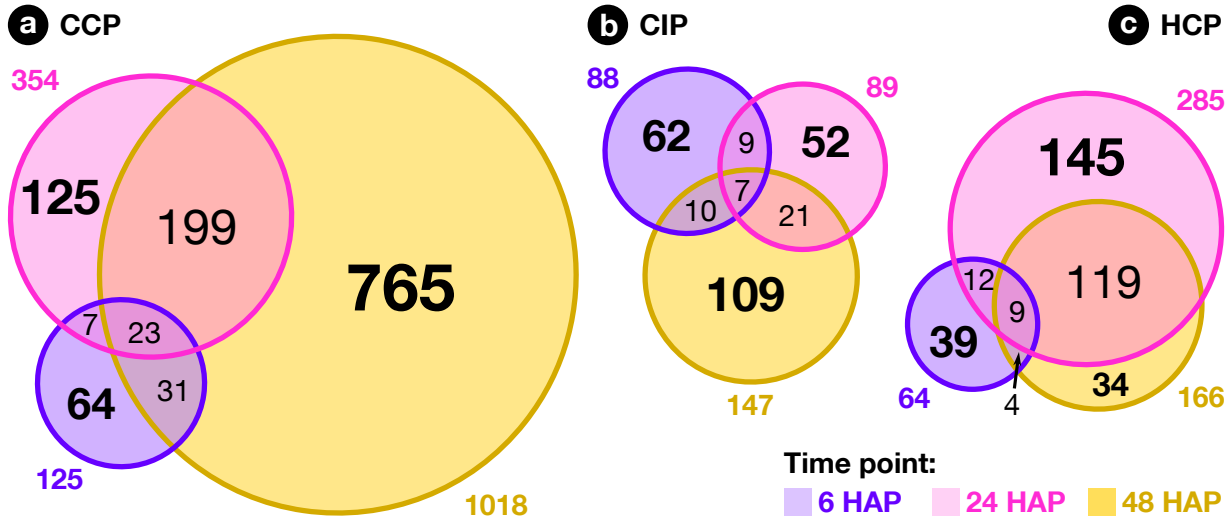


Figure 5.5. Overlap between time points. Venn diagrams showing the overlap between lists of genes modulated 6, 24 and 48 HAP after CCP, CIP, and HCP.

5.4.3 Early Response to Pollination

At 6 HAP, all pollination types had germinated equally and PTs had reached ~1.5 mm. In all, 163 ovary transcripts showed a statistically significant change in abundance 6 h after CCP, CIP, or HCP (Section 5.4.2a). As can be seen in Figure 5.2a, the three pollination types induced a globally similar response in the ovary, with significant overlaps between regulated genes (Table S5.1a–c). Interestingly, pollination responses were also highly correlated to the stigma wounding condition with R^2 coefficents ranging from 0.80 to 0.93 (Figure S5.1), but remained clearly distinct from a simple touch response ($R^2 \leq 0.05$), suggesting that the early response following pollination corresponds to the perception of a wounding aggression, due to PT penetration and growth in the stigma.

Functional categories significantly enriched in common and coregulated transcripts included defense-related GO-terms such as "defense response to fungus" (Dataset S5.4a–g). Interestingly, at 24 and 48 HAP, this category remained enriched only in transcripts regulated by CIP or HCP, but not CCP (clusters 4 and 18 on Section 5.4.2; Dataset S5.5d and r), suggesting that the response to CIP and HCP remains more akin to a defense response than CCP at later time points. The GO enrichment analysis also revealed that the non-specific response to pollination 6 HAP was correlated to the modulation of signaling-related categories such as "auxin transport" and "cellular response to reactive oxygen species". Moreover, transcription factors predicted to belong to ERF and ARF families were also regulated 6 HAP (Section 5.4.3, Table S5.2a), pointing to a possible involvement of phytohormones in the mediation of early ovary responses to pollination.

In *S. chacoense*, the ovule secretome was shown to consitute a dynamic microenvironnement in preparation for terminal pollen–pistil interactions.[60] Therefore, we investigated the presence of transcripts predicted to encode secreted proteins (SPs) in our dataset (Section 5.4.3, Table S5.3). Interestingly, they represented 31 to 40% of the transcripts regulated 6 HAP, while they accounted for only 7.5% of non-regulated transcripts, which represents a significant enrichment (Table S5.3a). An example of SPs induced 6 HAP were xyloglucan endotransglucosylase/hydrolases (XTHs), a group of cell wall-loosening enzymes previously reported to play a role during host invasion by parasitic plant haustoria.[61] Besides XTHs, 15, 11, and 3 cysteine-rich proteins (CRPs) were modulated by CCP, CIP, and HCP, respectively (Section 5.4.3). This peculiar category of small, secreted, rapidly evolving proteins, with ≥6 cysteines and a mature size ≤150 aa, was shown to be involved in several species-specific pollen–pistil interactions.[62] Here, CRPs exhibited a statistically significant enrichment 6 HAP, representing up

to 13% of transcripts induced by pollination, and only 0.8% of not regulated transcripts (Table S5.3a). Interestingly, different CRP families exhibited distinct regulation patterns: lipid-transfer proteins (LTPs) were induced by pollination, while other families such as γ-thionins and metallocarboxypeptidase inhibitors (MCPIs) were repressed. LTPs were previously shown to control PT adhesion and pre-ovular guidance in the pistil,[63] while thionin-like proteins were reported to be embryo sac-dependent CRPs with potential roles in PT-ovule interactions.[64] MCPIs, on the other hand, are known to be ovary and fruit development regulators in tomato plants.[65] All this underlines that the early pollination signal participates in the dynamic remodelling of the ovule secretome, affecting proteins susceptible to play key roles for ovary development and functionality.

Finally, even though the three pollination types produced a globally similar response in the ovary 6 HAP, specific profiles already started to be visible, with 52, 22, and 13 transcripts specifically regulated in CCP, CIP, and HCP, respectively (Section 5.4.2a). Moreover, transcripts up-regulated in both conspecific pollinations (CCP and CIP) were specifically enriched in several proteins similar to known regulators of ovule specification and development: ARGONAUTE 4 proteins,[66] as well as the AGAMOUS-LIKE 11[67] and AUXIN RESPONSE FACTOR 5[68] transcription factors (Dataset S5.4d). This points to a possible role of conspecific pollination as a signal triggering ovule and female gametophyte development, possibly mediated by ethylene and auxin, as demonstrated previously in other species.[38,39,40,41]

5.4.4 Pollination Response after Completion of the SI Reaction

At the second time point, 24 HAP, the majority of CIP PTs had ceased growth, while HCP and CCP PTs had reached around one and two thirds of the style's length, respectively (Section 5.4.1). Compared to 6 HAP, an amplification of the ovary response to both compatible pollinations was visible, with 354 and 285 transcripts modulated in CCP and HCP, respectively (Section 5.4.2b) and a very limited overlap with early responses (Section 5.4.2a–c and Figure S5.1). Moreover, a larger proportion of those transcripts became specific to CCP (76% or 269/354) and HCP (69% or 196/285). Even though 58 transcripts appeared in the overlap between CCP and HCP on Section 5.4.2b, the majority of them (57%) had in fact opposite regulations (Table S5.1d). All this suggests that the two pollination types are now perceived as distinct events by the ovary, as confirmed by the low correlation coefficient on Figure S5.1 ($R^2 = 0.03$).

This is further supported by the analysis of enriched functional categories. Among them,

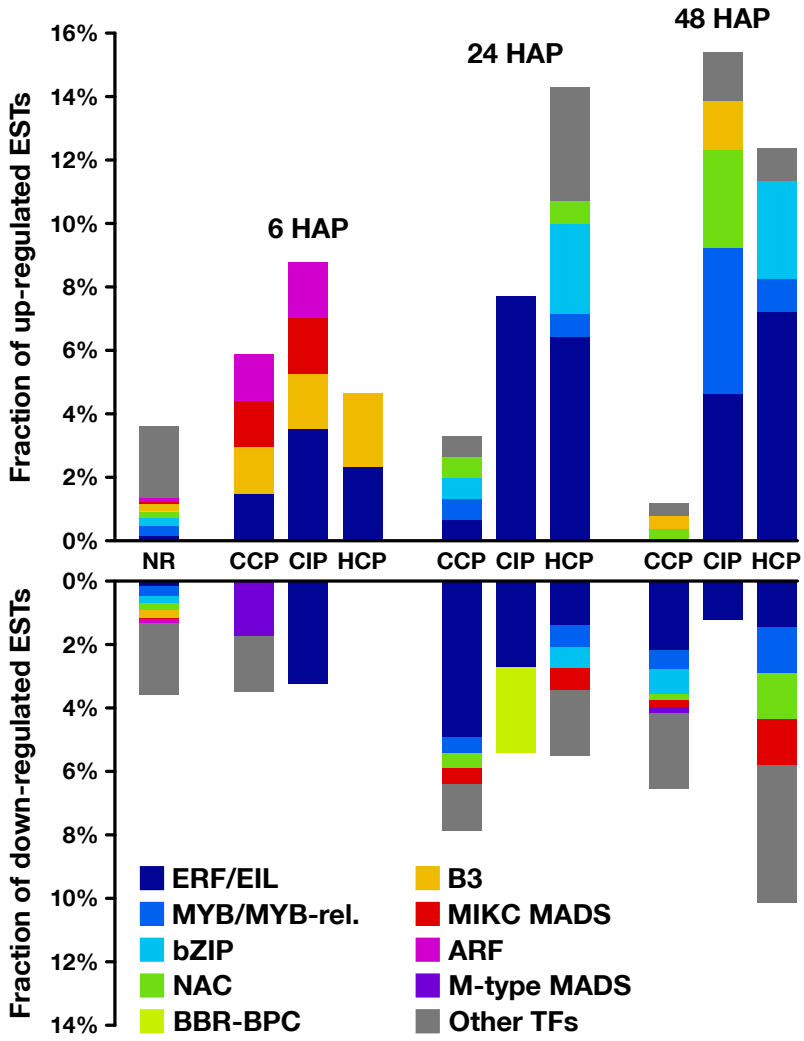


Figure 5.6. Transcription factor (TF) predictions on regulated genes. For each time point and each pollination condition, proportion of up- (**top**) and down-regulated (**bottom**) genes predicted to encode transcription factors belonging to different families, as identified by the PlantTFDB prediction tool. Corresponding data on non-regulated (*NR*) transcripts is shown as a reference. ARF: auxin response factor; BBR: barley b recombinant; BPC: BASIC PENTACYSTEINE1; bZIP: basic region/leucine zipper motif; EIL: ETHYLENE-INSENSITIVE 3-like; ERF: ethylene-responsive element binding factor; MADS: MCM1, AGAMOUS, DEFICIENS, and SRF; MIKC: MADS-box, intervening, keratin-like, and C-terminal domains; MYB: myeloblastosis virus; NAC: NAM, ATAF, and CUC.

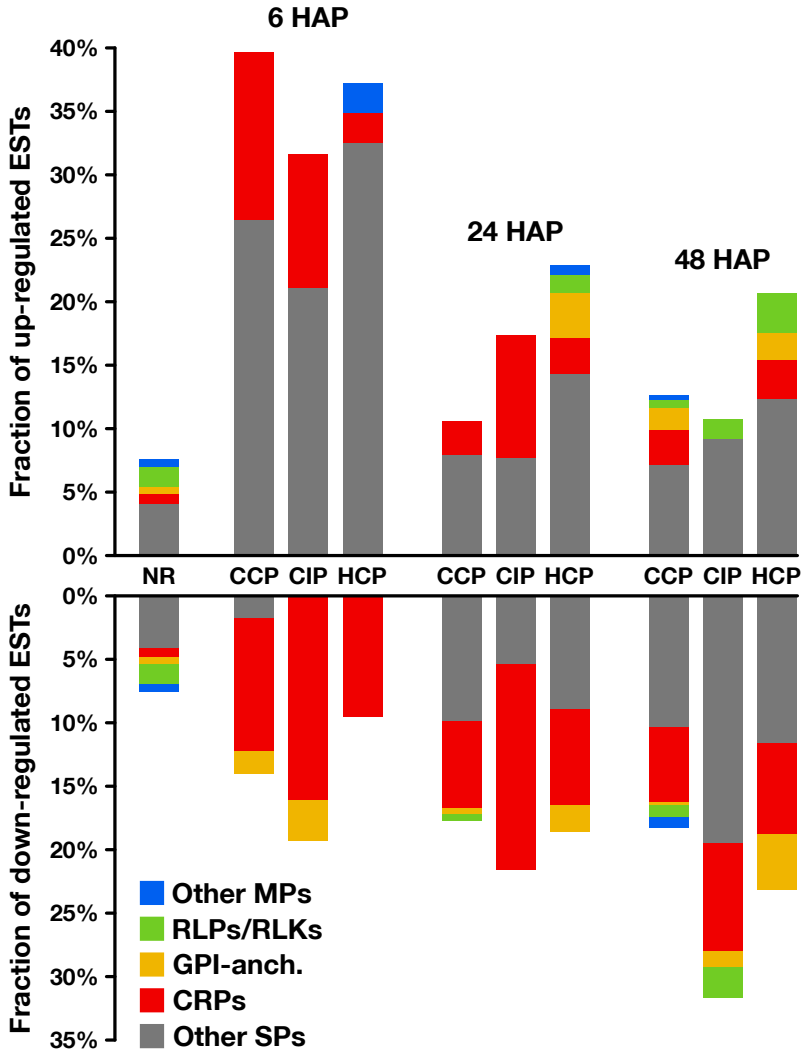


Figure 5.7. Secretion predictions on regulated genes. For each time point and each pollination condition, proportion of up- (**top**) and down-regulated (**bottom**) genes predicted to encode proteins possessing a secretory signal peptide. Among them, cysteine-rich proteins (*CRPs*, red) and other non-membrane secreted proteins (*Other SPs*, gray), predicted GPI-anchored proteins (*GPI-anch.*, yellow), putative receptor-like proteins and receptor-like kinases (*RLPs/RLKs*, green), and other membrane-bound secreted proteins (*Other MPs*, blue). Corresponding data on non-regulated (*NR*) transcripts is shown as a reference.

phytohormone-related GO terms exhibited contrasted responses after CCP and HCP. In particular, categories related to signaling mediated by the diffusible hormone ethylene were enriched in transcripts up-regulated by HCP and down-regulated by CCP (Dataset S5.4l). GO-terms "response to ethylene" and "ethylene metabolic process" were also significantly over-represented in cluster 4 on Section 5.4.2 (Dataset S5.5d), while putative ERF/EIL transcription factors shared a consistent enrichment profile (Section 5.4.3, Table S5.2b), which suggests that ethylene is a key mediator of pollination-specific, long-distance signaling in the pistil. In line with this, previous studies revealed the existence of post-pollination ethylene bursts (PPEBs) occuring in the stigma/style of other solanaceous species such as *Petunia*[69] and tobacco,[70,71] whose timing and/or intensity could vary according to the pollination type.

More recently, additional roles were discovered for ethylene signaling in the context of ovule and PT function. While pollination-induced ethylene accumulation in immature tobacco flowers was shown to be correlated to female gametophyte maturation,[41] ethylene was also demonstrated to control PT elongation in *Arabidopsis* by affecting the organization of actin microfilaments.[72] Furthermore, the over-accumulation of the ethylene signal-transducing protein ETHYLENE INSENSITIVE 3 (EIN3) in synergid cells was shown to lead to PT attraction defects in *Arabidopsis*.[73] Here, cluster 8, which gathers transcripts specifically down 24 and 48 h after CCP (Section 5.4.2), included two EIN3-like proteins that remained stable after CIP and HCP (Dataset S5.3d). This suggests that the differential ethylene response in CCP vs. HCP might allow the ovary to get specifically prepared for compatible PT guidance.

Besides ethylene, genes up-regulated in HCP and down-regulated in CCP were also enriched in the GO-term "response to abscisic acid" (Dataset S5.4l), while "response to gibberellin" was over-represented in cluster 9, consisting of transcripts specifically down in CCP (Dataset S5.5i). On the other hand, transcripts up-regulated by CCP and/or down-regulated by HCP were enriched in categories related to auxin, brassinosteroids, and jasmonic acid (Dataset S5.4j and m, Dataset S5.5r and t). This denotes the existence of a complex cross-talk between phytohormone signaling pathways coordinating the ovary response to pollination.

In contrast, GO-terms associated with the response to reactive oxygen species (ROS) were over- represented in genes specifically up-regulated by CCP (Dataset S5.4h), as confirmed by the clustering analysis (Section 5.4.2, clusters 14 and 22; Dataset S5.5n and v). Interestingly, besides being key players of rapid long-distance signaling,[74] ROS are known to control pollen germination[75,76] and PT growth.[77,78,79] We could therefore hypothesize that ROS may convey the CCP signal at a distance, or be part of the ovule response to a different CCP signal.

Interestingly, ovule-emitted ROS were shown to control PT rupture, with no influence however, on the pollination status.[80] Further work is therefore required to better understand how CCP-induced modulation of ROS-related genes affects ovule function in preparation for interactions restricted to conspecific PTs.

As demonstrated by metabolic pathway (Table S5.4b) and GO (Dataset S5.5t) enrichment analyses, enzymes of the secondary metabolism, especially those involved in anthocyanin, flavone, and favonol biosynthesis, were over-represented in transcripts from cluster 20, which were up-regulated by CCP and down-regulated by HCP (Section 5.4.2). Flavonoids, have been extensively studied as messenger molecules during pollination, especially for the control of pollen germination.[81] Interestingly, flavonoids were also shown to play a key role for the maintenance of ROS homeostasis in the context of PT growth.[82] All this suggests that CCP-induced flavonoid production by the ovary could be a mechanism favoring conspecific PT growth in the pistil, in preparation for species-preferential pollen–ovule interactions.

Moreover, cluster 18 (Section 5.4.2) contained a γ-aminobutyric acid (GABA) transaminase, an enzyme responsible for the control of γ-aminobutyric acid (GABA), which was specifically down-regulated after HCP while it remained stable after CCP (Dataset S5.3d). GABA is a key signaling compound controlling PT elongation[83] and known to form a gradient in the pistil, with increasing concentrations from the stigma to the ovule, whose disruption impairs proper PT directional growth.[84] Therefore, HCP-induced disruption of pistil GABA levels could constitute another mechanism facilitating the rejection of heterospecific pollen.

In contrast to the ample, antagonistic ovary response to CCP and HCP, the number of transcripts regulated by CIP 24 HAP remained stable compared to 6 HAP (89 vs. 88, Section 5.4.2b). Interestingly, transcipts specifically regulated by CIP were enriched in GO-terms such as "gene silencing", "DNA packaging", and "chromating remodeling" (Dataset S5.4i), pointing to a possible epigenetic modulation of gene expression in the ovary as a consequence of self-pollination.

5.4.5 Fertilization and Late Pollination Responses

In *S. chacoense*, as in many *Solanum* species, conspecific fertilization takes place from 36 HAP until 48 HAP as determined by aniline blue staining of the PTs that had reached the ovules (data not shown) and by the fertilization-induced activation of ribosomal proteins.[85] As expected, the highest number of ovule-modulated genes, 1018, were isolated 48 HAP from a

fully compatible pollination (CCP) that lead to fertilization, thus including a large number of genes regulated immediately following fertilization (Section 5.4.2c). Among them, 253 (25%) were already regulated before fertilization 6 or 24 HAP, suggesting a dual role for these genes before and after fertilization, during pollination and early embryogenesis (Section 5.4.2a and Table S5.1g). Remaining genes were particularly present in clusters 5–10 (down-regulation) and 14–16 (up-regulation) and were, as expected, enriched in functional categories related to cell proliferation and gene expression (Dataset S5.4o, Dataset S5.5e–j and n–p) underlining that CCP induced the reprogramming of ovule transcriptome toward embryo development.

In contrast, only 147 genes were modulated by CIP. Although only a limited increase was observed in the total number of genes modulated between 24 and 48 HAP in CIP (going up from 89 to 147), the nature of the modulated genes was strikingly different with 74% (109/147) genes specifically expressed 48 HAP (Section 5.4.2b and Table S5.1h). In fact, as can be seen on Figure 5.2c, this modulation of the ovule response to CIP 48 HAP still closely resembles a wound response, as confirmed by the high statistical correlation between the two conditions ($R^2 = 0.70$, Figure S5.1).

Compared to CCP, only 166 genes were modulated in HCP at 48 HAP, an important reduction from 24 HAP (Section 5.4.2c). Among them, no ribosomal protein genes were up-regulated (out of the 65 available on the microarray), indicating that fertilization had yet taken place (Dataset S5.3c). This was confirmed by aniline blue staining of *S. microdontum* PTs 48 HAP, showing that most of the tubes had only travelled 60% of the style's length (Section 5.4.1). Interestingly, 77% of the HCP genes regulated at 48 HAP (128 out of 166) were common with the ones expressed 24 HAP, a situation not observed in other pollination types where little overlap was observed between successive time points (Section 5.4.2a–c, Table S5.1g–i). This is further confirmed by the high statistical correlation between HCP responses 24 and 48 HAP ($R^2 = 0.83$, Figure S5.1). Importantly, responses to mid-style CCP PTs (24 HAP) and HCP PTs (48 HAP) exhibit a very low correlation coefficient ($R^2 = 0.01$), confirming that the late response to HCP is not simply a non-specific response to PTs located in the central region of the style. In terms of functional annotations, the categories enriched 48 h after HCP mostly overlap those enriched 24 HAP and discussed in the previous section (Datasets S4 and S5). All this shows that HCP is perceived by the ovary as a single, continuous signal, that is clearly distinct from CCP.

5.5 Conclusions

The present study shows that, after being all initially perceived as a wounding aggression, each pollination type produced its own transcriptomic signature at a distance in the ovary, in a way that may prepare subsequent species–preferential pollen–ovule interactions (Section 5.5). We have shown that ROS and ethylene are potential messengers acting at a distance to convey the presence of CCP and HCP PTs, respectively. But how could distinct pollination types elicit such antagonistic long-distance responses in the pistil? A recent study revealed that pollination triggers the expression of three pollinic MYB transcription factors, whose mutation causes significant changes in the post-pollination pistil transcriptome. Interestingly, those MYBs were shown to control the expression of rapidly evolving thionin-like CRPs that are secreted by the growing PTs in the pistil, and suspected to control proper PT reception by the ovule.[24] Such divergent PT-secreted proteins could serve as initial signals specific to each pollination type, allowing the pistil to discriminate between CCP, CIP, and HCP PTs.

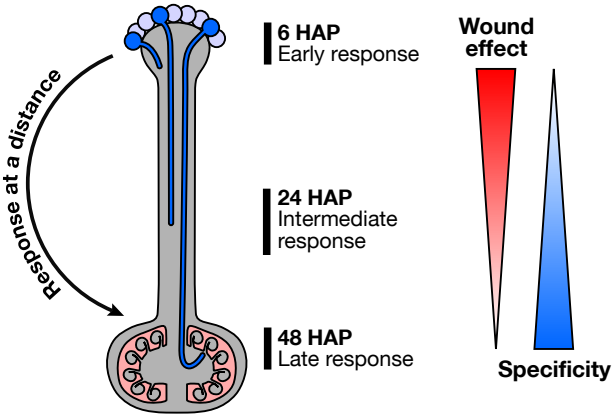


Figure 5.8. Outline of the at a distance ovary response following pollination. Early pollination response, irrespective of the pollination type, is akin to a wound response. As the PT grow, each pollination type is then recognized as distinct and produce a highly specific transcriptomic signature in the ovary, before PT arrival.

5.6 Materials & Methods

5.6.1 Plant Material and Pollination Conditions

Self-incompatible *Solanum chacoense* and *S. microdontum* accessions ($2n = 2x = 24$) obtained from the NRSP-6 US Potato Genebank (Sturgeon Bay, WI, USA) were glasshouse-grown with a 16 h light/8 h dark cycle. The *S. chacoense* G4 genotype ($S_{12}S_{14}$ SI alleles) was used as the female progenitor. *S. chacoense* pollen from G4 and V22 ($S_{11}S_{13}$) genotypes was used for CIP and CCP, respectively. HCP was performed with pollen from *S. microdontum* PI500041 accession.

Wounding treatments were performed by slightly crushing the upper region of the style with small forceps, as described previously.[48] Touch treatments consisted in mock pollinations accomplished by gently touching the stigmas with sterile 100 µm zirconia/silica beads (BioSpec Products Inc., Bartlesville, OK, USA).

5.6.2 Pollen Tube Growth Assay and Aniline Blue Staining

Flowers were collected from 6 to 96 h after pollination. Pistils were dissected and fixed overnight in FAA (ethanol 50%, water 35% glacial acetic acid 10%, formalin 5%, washed twice with water and then softened in 2 M NaOH for 24 h at room temperature. After rinsing, pistils were stained with 0.1% aniline blue in K_3PO_4 buffer (pH 7.5) and slightly squashed between a slide and coverslip. Pictures were taken on a Zeiss Axio Observer.Z1 fluorescence microscope equipped with an AxioCam HRm camera (Carl Zeiss Canada, Canada) and analyzed with ImageJ (https://imagej.nih.gov/ij).

5.6.3 RNA Isolation and Microarray Experimental Design

Ovules were collected 6, 24, and 48 h after each treatment and used for RNA extraction and probe preparation. RNA from UOs served as controls. Four independent biological replicates were produced for each time point. To estimate reproducibility and to produce control data for statistical analyses, a large number of UOs were isolated and separated in seven independent control groups. RNA from randomly selected pairs of controls was hybridized on six microarrays. Microarray experiments and data analysis were performed as described previ-

ously.[19] The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov) and are accessible through GEO Series accession number GSE21957.[86]

5.6.4 Differential Expression Analysis

Transcripts with a significant expression fold-change in CCP, CIP, HCP, W or T vs. UO ($\geq$1.5 or ≤ -1.5; $P \leq 0.05$, Welch's t-test) were considered to be regulated. To draw the heatmaps (Figure 5.2), a hierarchical clustering analysis was performed at each time point using genes regulated in at least one pollination condition. Euclidean distances between CCP, CIP, and HCP vs. UO expression ratios were used to connect transcripts based on Ward's method.[87] Expression ratios in W and T vs. UO were then added to the heatmaps. Clusters presented on Section 5.4.2 were obtained by a similar hierarchical clustering analysis applied to Pearson's correlation coefficients of UO, CCP, CIP, and HCP vs. UO expression ratios at all time points, based on Ward's method. Dendrograms were then split into clusters using k-means clustering with $k = 25$. Figure S5.1 presents pairwise squared correlation coefficients (R^2) of CCP, CIP, HCP, W, and T vs. UO expression values obtained by linear least-squares regression analysis.

5.6.5 Sequence Annotation

ESTs were compared to the NCBI `refseq_rna` and `refseq_protein` databases v. 87[88] using BLASTn and BLASTx v. 2.2.29+, respectively.[89] Descriptions were then manually assigned to each EST based on the most similar hits. Automated functional classification into Gene Ontology (GO) categories was performed with Blast2GO v. 5.2.5.[90] The best BLASTx hits for each EST were used for transcription factor predictions with the PlantTFDB v. 4.0 prediction tool,[91] enzyme code retrieval and metabolic pathway assignment based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database v. 90.0,[92] and signal peptide predictions with SignalP v. 4.1.[93] Proteins predicted to contain a signal peptide were further checked for the presence of transmembrane helices (TMH) with TMHMM v. 2.0.[94]

Proteins with one predicted TMH were submitted to HMMER v. 3.1b2 (http://hmmer.org/) to check for the presence of kinase domains (KDs; motifs *Pkinase* and *Pkinase_Tyr*) and leucine-rich repeats (LRRs; motifs *LRR_1*, *LRR_2*, *LRR_4*, *LRR_5*, *LRR_6*, *LRR_8*, *LRR_9*, *LRRNT*, and *LRRNT_2*) defined in the Pfam database v. 32.0.[95] Among them, those with at least one KD and one LRR were classified as potential LRR receptor-like kinases (LRR-

RLKs); those with at least one KD but no LRR were tagged as potential RLKs; remaining proteins with one TMH were considered as potential receptor-like proteins (RLPs). Proteins with two or more predicted TMHs were tagged as other membrane proteins.

Proteins that were not predicted to have a TMH were inspected for the presence of predicted of glycosylphosphatidylinositol (GPI) anchors with the PredGPI program.[96] Remaining proteins were split into cysteine-rich proteins (6 cysteines or more, mature part smaller than 150 aa) and other secreted proteins with KAPPA v. 1.0.[97]

Enrichment analyses based on all those predictions were made using Fisher's exact tests. A prediction was considered enriched in a given condition when $P \leq 0.05$.

Bibliography

1. Mizuta, Y. et Higashiyama, T. (2018). Chemical signaling for pollen tube guidance at a glance. *J. Cell Sci.*, 131(2), jcs208447. DOI: 10.1242/jcs.208447. [cit. p. 147]

2. Johnson, M. A. ; Harper, J. F. et Palanivelu, R. (2019). A fruitful journey: pollen tube navigation from germination to fertilization. *Annu. Rev. Plant Biol.*, 70(1). DOI: 10.1146/annurev-arplant-050718-100133. [cit. p. 147]

3. Doucet, J. ; Lee, H. K. et Goring, D. R. (2016). Pollen acceptance or rejection: a tale of two pathways. *Trends Plant Sci.*, 21(12), 1058–67. DOI: 10.1016/j.tplants.2016.09.004. [cit. p. 147]

4. Wang, L. ; Lin, Z. ; Triviño, M. et coll. (2018). Self-incompatibility in *Papaver* pollen: programmed cell death in an acidic environment. *J. Exp. Bot.* DOI: 10.1093/jxb/ery406. [cit. p. 147]

5. Williams, J. S. ; Wu, L. ; Li, S. ; Sun, P. et Kao, T.-H. (2015). Insight into S-RNase-based self-incompatibility in *Petunia*: recent findings and future directions. *Front. Plant Sci.*, 6, 41. DOI: 10.3389/fpls.2015.00041. [cit. p. 147]

6. Becker, J. D. ; Boavida, L. C. ; Carneiro, J. ; Haury, M. et Feijó, J. A. (2003). Transcriptional profiling of *Arabidopsis* tissues reveals the unique characteristics of the pollen transcriptome. *Plant Physiol.*, 133(2), 713–25. DOI: 10.1104/pp.103.028241. [cit. p. 147]

7. Honys, D. et Twell, D. (2003). Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.*, 132(2), 640–652. DOI: 10.1104/pp.103.020925. [cit. p. 147]

8. Honys, D. et Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biol.*, 5(11), R85. DOI: 10.1186/gb-2004-5-11-r85. [cit. p. 147]

9. Pina, C. ; Pinto, F. ; Feijó, J. A. et Becker, J. D. (2005). Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiol.*, 138(2), 744–756. DOI: 10.1104/pp.104.057935. [cit. p. 147]

10. Wang, Y. ; Zhang, W.-Z. ; Song, L.-F. et coll. (2008). Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiol.*, 148 (3), 1201–1211. DOI: 10.1104/pp.108.126375. [cit. p. 147]

11. Swanson, R. ; Clark, T. et Preuss, D. (2005). Expression profiling of *Arabidopsis* stigma tissue identifies stigma-specific genes. *Sex. Plant Reprod.*, 18(4), 163–171. DOI: 10.1007/s00497-005-0009-x. [cit. p. 147]

12. Tung, C.-W. ; Dwyer, K. G. ; Nasrallah, M. E. et Nasrallah, J. B. (2005). Genome-wide identification of genes expressed in *Arabidopsis* pistils specifically along the path of pollen tube growth. *Plant Physiol.*, 138(2), 977–989. DOI: 10.1104/pp.105.060558. [cit. p. 147]

13. Li, M. ; Xu, W. ; Yang, W. ; Kong, Z. et Xue, Y. (2007). Genome-wide gene expression profiling reveals conserved and novel molecular functions of the stigma in rice. *Plant Physiol.*, 144(4), 1797–1812. DOI: 10.1104/pp.107.101600. [cit. p. 147]

14. Quiapim, A. C. ; Brito, M. S. ; Bernardes, L. A. S. et coll. (2009). Analysis of the *Nicotiana tabacum* stigma/style transcriptome reveals gene expression differences between wet and dry stigma species. *Plant Physiol.*, 149(3), 1211–1230. DOI: 10.1104/pp.108.131573. [cit. p. 147]

15. Ma, L. ; Sun, N. ; Liu, X. et coll. (2005). Organ-specific expression of *Arabidopsis* genome during development. *Plant Physiol.*, 138(1), 80–91. DOI: 10.1104/pp.104.054783. [cit. p. 147]

16. Yu, H.-J. ; Hogan, P. et Sundaresan, V. (2005). Analysis of the female gametophyte transcriptome of *Arabidopsis* by comparative expression profiling. *Plant Physiol.*, 139(4), 1853–1869. DOI: 10.1104/pp.105.067314. [cit. p. 147]

17. Peiffer, J. A. ; Kaushik, S. ; Sakai, H. et coll. (2008). A spatial dissection of the *Arabidopsis* floral transcriptome by MPSS. *BMC Plant Biol.*, 8, 43. DOI: 10.1186/1471-2229-8-43. [cit. p. 147]

18. Vriezen, W. H. ; Feron, R. ; Maretto, F. ; Keijman, J. et Mariani, C. (2008). Changes in tomato ovary transcriptome demonstrate complex hormonal regulation of fruit set. *New Phytol.*, 177(1), 60–76. DOI: 10.1111/j.1469-8137.2007.02254.x. [cit. p. 147]

19. Tebbji, F. ; Nantel, A. et Matton, D. P. (2010). Transcription profiling of fertilization and early seed development events in a solanaceous species using a 7.7 K cDNA microarray from *Solanum chacoense* ovules. *BMC Plant Biol.*, 10, 174. DOI: 10.1186/1471-2229-10-174. [cit. p. 147, 149, 151, 163]

20. Dresselhaus, T. ; Lörz, H. et Kranz, E. (1994). Representative cDNA libraries from few plant cells. *Plant J.*, 5(4), 605–610. DOI: 10.1046/j.1365-313X.1994.05040605.x. [cit. p. 147]

21. Wuest, S. E. ; Vijverberg, K. ; Schmidt, A. et coll. (2010). *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr. Biol.*, 20(6), 506–512. DOI: 10.1016/j.cub.2010.01.051. [cit. p. 147]

22. Qin, Y. ; Leydon, A. R. ; Manziello, A. et coll. (2009). Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genet.*, 5 (8), e1000621. DOI: 10.1371/journal.pgen.1000621. [cit. p. 147]

23. Boavida, L. C. ; Borges, F. ; Becker, J. D. et Feijó, J. A. (2011). Whole genome analysis of gene expression reveals coordinated activation of signaling and metabolic pathways during pollen-pistil interactions in *Arabidopsis*. *Plant Physiol.*, 155(4), 2066–80. DOI: 10.1104/pp.110.169813. [cit. p. 147]

24. Leydon, A. R. ; Weinreb, C. ; Venable, E. et coll. (2017). The molecular dialog between flowering plant reproductive partners defined by SNP-informed RNA-sequencing. *Plant Cell*. DOI: 10.1105/tpc.16.00816. [cit. p. 147, 161]

25. Tan, H. ; Zhang, J. ; Qi, X. et coll. (2018). Integrated metabolite profiling and transcriptome analysis reveals a dynamic metabolic exchange between pollen tubes and the style during fertilization of *Brassica napus*. *Plant Mol. Biol.* DOI: 10.1007/s11103-018-0740-y. [cit. p. 147]

26. Zhao, P. ; Zhang, L. et Zhao, L. (2015). Dissection of the style's response to pollination using transcriptome profiling in self-compatible (*Solanum pimpinellifolium*) and self-incompatible (*Solanum chilense*) tomato species. *BMC Plant Biol.*, 15, 119. DOI: 10.1186/s12870-015-0492-7. [cit. p. 147]

27. Zhang, T. ; Gao, C. ; Yue, Y. et coll. (2017). Time-course transcriptome analysis of compatible and incompatible pollen-stigma interactions in *Brassica napus* L. *Front. Plant Sci.*, 8, 682. DOI: 10.3389/fpls.2017.00682. [cit. p. 147]

28. Ma, Q. ; Chen, C. ; Zeng, Z. et coll. (2018). Transcriptomic analysis between self- and cross-pollinated pistils of tea plants (*Camellia sinensis*). *BMC Genomics*, 19(1), 289. DOI: 10.1186/s12864-018-4674-1. [cit. p. 147]

29. Pease, J. B. ; Guerrero, R. F. ; Sherman, N. A. ; Hahn, M. W. et Moyle, L. C. (2016). Molecular mechanisms of postmating prezygotic reproductive isolation uncovered by transcriptome analysis. *Mol. Ecol.*, 25(11), 2592–608. DOI: 10.1111/mec.13679. [cit. p. 147]

30. Broz, A. K. ; Guerrero, R. F. ; Randle, A. M. et coll. (2017). Transcriptomic analysis links gene expression to unilateral pollen-pistil reproductive barriers. *BMC Plant Biol.*, 17(1), 81. DOI: 10.1186/s12870-017-1032-4. [cit. p. 147]

31. Wang, M. ; Chen, Z. ; Zhang, H. ; Chen, H. et Gao, X. (2018). Transcriptome analysis provides insight into the molecular mechanisms underlying *gametophyte factor 2*-mediated cross-incompatibility in maize. *Int. J. Mol. Sci.*, 19(6), 1757. DOI: 10.3390/ijms19061757. [cit. p. 147]

32. Mondragón-Palomino, M. ; John-Arputharaj, A. ; Pallmann, M. et Dresselhaus, T. (2017). Similarities between reproductive and immune pistil transcriptomes of *Arabidopsis* species. *Plant Physiol.*, 174(3), 1559–1575. DOI: 10.1104/pp.17.00390. [cit. p. 147]

33. Rao, P. ; Chen, Z. ; Yang, X. et coll. (2017). Dynamic transcriptomic analysis of the early response of female flowers of *Populus alba* × *P. glandulosa* to pollination. *Sci. Rep.*, 7(1), 6048. DOI: 10.1038/s41598-017-06255-3. [cit. p. 147]

34. Hildebrand, F. (1863). Die Fruchtbildung der Orchideen, ein Beweis für die doppelte Wirkung des Pollen. *Bot. Zeitung*, 21(44), 329–333. URL: https://www.biodiversitylibrary.org/page/33866669#page/345. [cit. p. 148]

35. Treub, M. M. (1883). Notes sur l'embryon, le sac embryonnaire et l'ovule. 4. L'action des tubes polliniques sur le développement des ovules chez les orchidées. *Ann. Jard. Bot. Buitenzorg*, 3, 122–128. URL: https://play.google.com/books/reader?id=6xdYAAAAMAAJ&hl=fr_CA&pg=GBS.PA120. [cit. p. 148]

36. Guignard, L. (1886). Sur la pollinisation et ses effets chez les orchidées. *Ann. Sci. Nat. Sér. 7 : Bot.*, 4, 202–240. URL: https://gallica.bnf.fr/ark:/12148/bpt6k5518127z/f209. [cit. p. 148]

37. Zhang, X. S. et O'Neill, S. D. (1993). Ovary and gametophyte development are coordinately regulated by auxin and ethylene following pollination. *Plant Cell*, 5(4), 403–418. DOI: 10.1105/tpc.5.4.403. [cit. p. 148]

38. O'Neill, S. D. ; Bui, A. Q. ; Potter, D. et Zhang, X. S. (2017). Pollination of orchid flowers: quantitative and domain-specific analysis of ethylene biosynthetic and hormone-induced gene expression. *Int. J. Plant Sci.*, 178(3), 188–210. DOI: 10.1086/690107. [cit. p. 148, 155]

39. Pimienta, E. et Polito, V. S. (1983). Embryo sac development in almond [*Prunus dulcis* (Mill.) D. A. Webb] as affected by cross-, self- and non-pollination. *Ann. Bot.*, 51(4), 469–479. DOI: 10.1093/oxfordjournals.aob.a086492. [cit. p. 148, 155]

40. Mól, R. ; Filek, M. ; Machackova, I. et Matthys-Rochon, E. (2004). Ethylene synthesis and auxin augmentation in pistil tissues are important for egg cell differentiation after pollination in maize. *Plant Cell Physiol.*, 45(10), 1396–1405. DOI: 10.1093/pcp/pch167. [cit. p. 148, 155]

41. Brito, M. S. ; Bertolino, L. T. ; Cossalter, V. et coll. (2015). Pollination triggers female gametophyte development in immature *Nicotiana tabacum* flowers. *Front. Plant Sci.*, 6, 561. DOI: 10.3389/fpls.2015.00561. [cit. p. 148, 155, 158]

42. O'Neill, S. D. (1997). Pollination regulation of flower development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 48, 547–574. DOI: 10.1146/annurev.arplant.48.1.547. [cit. p. 148]

43. Huda, M. K. et Wilcock, C. C. (2012). Rapid floral senescence following male function and breeding systems of some tropical orchids. *Plant Biol.*, 14(2), 278–284. DOI: 10.1111/j.1438-8677.2011.00507.x. [cit. p. 148]

44. Shibuya, K. ; Niki, T. et Ichimura, K. (2013). Pollination induces autophagy in petunia petals via ethylene. *J. Exp. Bot.*, 64(4), 1111–1120. DOI: 10.1093/jxb/ers395. [cit. p. 148]

45. Theis, N. et Raguso, R. A. (2005). The effect of pollination on floral fragrance in thistles. *J. Chem. Ecol.*, 31(11), 2581–2600. DOI: 10.1007/s10886-005-7615-9. [cit. p. 148]

46. O'Neill, S. D. ; Nadeau, J. A. ; Zhang, X. S. ; Bui, A. Q. et Halevy, A. H. (1993). Interorgan regulation of ethylene biosynthetic genes by pollination. *Plant Cell*, 5(4), 419–432. DOI: 10.1105/tpc.5.4.419. [cit. p. 148]

47. Llop-Tous, I. ; Barry, C. S. et Grierson, D. (2000). Regulation of ethylene biosynthesis in response to pollination in tomato flowers. *Plant Physiol.*, 123(3), 971–978. DOI: 10.1104/pp.123.3.971. [cit. p. 148]

48. Lantin, S. ; O'Brien, M. et Matton, D. P. (1999). Pollination, wounding and jasmonate treatments induce the expression of a developmentally regulated pistil dioxygenase at a distance, in the ovary, in the wild potato *Solanum chacoense* Bitt. *Plant Mol. Biol.*, 41(3), 371–386. DOI: 10.1023/A:1006375522626. [cit. p. 148, 162]

49. Broderick, S. R. ; Wijeratne, S. ; Wijeratn, A. J. et coll. (2014). RNA-sequencing reveals early, dynamic transcriptome changes in the corollas of pollinated petunias. *BMC Plant Biol.*, 14(1), 307. DOI: 10.1186/s12870-014-0307-2. [cit. p. 148]

50. Chen, C. ; Zeng, L. et Ye, Q. (2018). Proteomic and biochemical changes during senescence of *Phalaenopsis* 'red dragon' petals. *Int. J. Mol. Sci.*, 19(5), 1317. DOI: 10.3390/ijms19051317. [cit. p. 148]

51. Germain, H. ; Rudd, S. ; Zotti, C. et coll. (2005). A 6374 unigene set corresponding to low abundance transcripts expressed following fertilization in *Solanum chacoense* Bitt, and characterization of 30 receptor-like kinases. *Plant Mol. Biol.*, 59(3), 515–32. DOI: 10.1007/s11103-005-0536-8. [cit. p. 149]

52. O'Brien, M. ; Kapfer, C. ; Major, G. et coll. (2002). Molecular analysis of the stylar-expressed *Solanum chacoense* small asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility in *Nicotiana*. *Plant J.*, 32(6), 985–96. DOI: 10.1046/j.1365-313X.2002.01486.x. [cit. p. 149]

53. Feng, J. ; Chen, X. ; Wu, Y. et coll. (2006). Detection and transcript expression of S-RNase gene associated with self-incompatibility in apricot (*Prunus armeniaca* L.). *Mol. Biol. Rep.*, 33(3), 215–221. DOI: 10.1007/s11033-006-0011-x. [cit. p. 149]

54. Liu, B. ; Morse, D. et Cappadocia, M. (2009). Compatible pollinations in *Solanum chacoense* decrease both S-RNase and S-RNase mRNA. *PLoS One*, 4(6), e5774. DOI: 10.1371/journal.pone.0005774. [cit. p. 149]

55. Maune, J. F. ; Camadro, E. L. et Erazzú, L. E. (2018). Cross-incompatibility and self-incompatibility: unrelated phenomena in wild and cultivated potatoes? *Botany*, 96(1), 33–45. DOI: 10.1139/cjb-2017-0070. [cit. p. 149]

56. Brücher, H. (1953). Über das natürliche Vorkommen von Hybriden zwischen *Solanum simplicifolium* und *Solanum subtilius* im Aconquija Gebirge. *Z. Indukt. Abstamm. Vererbungsl.*, 85, 12–19. DOI: 10.1007/BF00311572. [cit. p. 149]

57. Stephenson, A. G. ; Travers, S. E. ; Mena-Ali, J. I. et Winsor, J. A. (2003). Pollen performance before and during the autotrophic-heterotrophic transition of pollen tube growth. *Philos. Trans. R. Soc., B*, 358(1434), 1009–18. DOI: 10.1098/rstb.2003.1290. [cit. p. 149]

58. Chantha, S.-C. ; Emerald, B. S. et Matton, D. P. (2006). Characterization of the plant Notchless homolog, a WD repeat protein involved in seed development. *Plant Mol. Biol.*, 62(6), 897–912. DOI: 10.1007/s11103-006-9064-4. [cit. p. 149]

59. Hogenboom, N. G. et Mather, K. (1975). Incompatibility and incongruity: two different mechanisms for the non-functioning of intimate partner relationships [and comment]. *Proc. R. Soc. London, Ser. B*, 188(1092), 361–375. DOI: 10.1098/rspb.1975.0025. [cit. p. 150]

60. Liu, Y. ; Joly, V. ; Dorion, S. ; Rivoal, J. et Matton, D. P. (2015). The plant ovule secretome: a different view toward pollen–pistil interactions. *J. Proteome Res.*, 14(11), 4763–4775. DOI: 10.1021/acs.jproteome.5b00618. [cit. p. 154]

61. Olsen, S. et Krause, K. (2017). Activity of xyloglucan endotransglucosylases/hydrolases suggests a role during host invasion by the parasitic plant *Cuscuta reflexa*. *PLoS One*, 12(4), e0176754. DOI: 10.1371/journal.pone.0176754. [cit. p. 154]

62. Bircheneder, S. et Dresselhaus, T. (2016). Why cellular communication during plant reproduction is particularly mediated by CRP signalling. *J. Exp. Bot.*, 67(16), 4849–61. DOI: 10.1093/jxb/erw271. [cit. p. 154]

63. Chae, K. ; Gonong, B. J. ; Kim, S.-C. et coll. (2010). A multifaceted study of stigma/style cysteine-rich adhesin (SCA)-like *Arabidopsis* lipid transfer proteins (LTPs) suggests diversified roles for these LTPs in plant growth and reproduction. *J. Exp. Bot.*, 61(15), 4277–4290. DOI: 10.1093/jxb/erq228. [cit. p. 155]

64. Jones-Rhoades, M. W. ; Borevitz, J. O. et Preuss, D. (2007). Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins. *PLoS Genet.*, 3 (10), 1848–61. DOI: 10.1371/journal.pgen.0030171. [cit. p. 155]

65. Molesini, B. ; Rotino, G. L. ; Dusi, V. et coll. (2018). Two metallocarboxypeptidase inhibitors are implicated in tomato fruit development and regulated by the inner no outer transcription factor. *Plant Sci.*, 266, 19–26. DOI: 10.1016/j.plantsci.2017.10.011. [cit. p. 155]

66. Hernández-Lagana, E. ; Rodríguez-Leal, D. ; Lúa, J. et Vielle-Calzada, J.-P. (2016). A multigenic network of ARGONAUTE4 clade members controls early megaspore formation in *Arabidopsis*. *Genetics*, 204(3), 1045–1056. DOI: 10.1534/genetics.116.188151. [cit. p. 155]

67. Ocarez, N. et Mejía, N. (2016). Suppression of the D-class MADS-box AGL11 gene triggers seedlessness in fleshy fruits. *Plant Cell Rep.*, 35(1), 239–254. DOI: 10.1007/s00299-015-1882-x. [cit. p. 155]

68. Liu, S. ; Zhang, Y. ; Feng, Q. et coll. (2018). Tomato AUXIN RESPONSE FACTOR 5 regulates fruit set and development via the mediation of auxin and gibberellin signaling. *Sci. Rep.*, 8(1), 2971. DOI: 10.1038/s41598-018-21315-y. [cit. p. 155]

69. Singh, A. ; Evensen, K. B. et Kao, T. H. (1992). Ethylene synthesis and floral senescence following compatible and incompatible pollinations in *Petunia inflata*. *Plant Physiol.*, 99(1), 38–45. [cit. p. 158]

70. De Martinis, D. ; Cotti, G. ; te Lintel Hekker, S. ; Harren, F. J. M. et Mariani, C. (2002). Ethylene response to pollen tube growth in *Nicotiana tabacum* flowers. *Planta*, 214(5), 806–812. DOI: 10.1007/s00425-001-0684-2. [cit. p. 158]

71. Bhattacharya, S. et Baldwin, I. T. (2012). The post-pollination ethylene burst and the continuation of floral advertisement are harbingers of non-random mate selection in *Nicotiana attenuata*. *Plant J.*, 71(4), 587–601. DOI: 10.1111/j.1365-313X.2012.05011.x. [cit. p. 158]

72. Jia, H. ; Yang, J. ; Liesche, J. et coll. (2017). Ethylene promotes pollen tube growth by affecting actin filament organization via the cGMP-dependent pathway in *Arabidopsis thaliana*. *Protoplasma*, 255 (1), 273–284. DOI: 10.1007/s00709-017-1158-0. [cit. p. 158]

73. Zhang, C. ; Teng, X.-D. ; Zheng, Q.-Q. et coll. (2018). Ethylene signaling is critical for synergid cell functional specification and pollen tube attraction. *Plant J.*, 96(1), 176–187. DOI: 10.1111/tpj.14027. [cit. p. 158]

74. Waszczak, C. ; Carmody, M. et Kangasjärvi, J. (2018). Reactive oxygen species in plant signaling. *Annu. Rev. Plant Biol.*, 69(1), 209–236. DOI: 10.1146/annurev-arplant-042817-040322. [cit. p. 158]

75. Speranza, A. ; Crinelli, R. ; Scoccianti, V. et Geitmann, A. (2012). Reactive oxygen species are involved in pollen tube initiation in kiwifruit. *Plant Biol.*, 14(1), 64–76. DOI: 10.1111/j.1438-8677.2011.00479.x. [cit. p. 158]

76. Smirnova, A. V. ; Matveyeva, N. P. et Yermakov, I. P. (2014). Reactive oxygen species are involved in regulation of pollen wall cytomechanics. *Plant Biol.*, 16(1), 252–257. DOI: 10.1111/plb.12004. [cit. p. 158]

77. Potocký, M. ; Jones, M. A. ; Bezvoda, R. ; Smirnoff, N. et Zársky, V. (2007). Reactive oxygen species produced by NADPH oxidase are involved in pollen tube growth. *New Phytol.*, 174(4), 742–751. DOI: 10.1111/j.1469-8137.2007.02042.x. [cit. p. 158]

78. Kaya, H. ; Nakajima, R. ; Iwano, M. et coll. (2014). Ca^{2+}-activated reactive oxygen species production by *Arabidopsis* RbohH and RbohJ is essential for proper pollen tube tip growth. *Plant Cell*, 26(3), 1069–1080. DOI: 10.1105/tpc.113.120642. [cit. p. 158]

79. Lassig, R. ; Gutermuth, T. ; Bey, T. D. ; Konrad, K. R. et Romeis, T. (2014). Pollen tube NAD(P)H oxidases act as a speed control to dampen growth rate oscillations during polarized cell growth. *Plant J.*, 78(1), 94–106. DOI: 10.1111/tpj.12452. [cit. p. 158]

80. Duan, Q. ; Kita, D. ; Johnson, E. A. et coll. (2014). Reactive oxygen species mediate pollen tube rupture to release sperm for fertilization in *Arabidopsis*. *Nat. Commun.*, 5, 3129. DOI: 10.1038/ncomms4129. [cit. p. 159]

81. Forbes, A. M. ; Meier, G. P. ; Haendiges, S. et Taylor, L. P. (2014). Structure-activity relationship studies of flavonol analogues on pollen germination. *J. Agric. Food Chem.*, 62(10), 2175–2181. DOI: 10.1021/jf405688d. [cit. p. 159]

82. Muhlemann, J. K. ; Younts, T. L. B. et Muday, G. K. (2018). Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. *Proc. Natl. Acad. Sci. U. S. A.*, 115(47), E11188–E11197. DOI: 10.1073/pnas.1811492115. [cit. p. 159]

83. Yu, G.-H. ; Zou, J. ; Feng, J. et coll. (2014). Exogenous γ-aminobutyric acid (GABA) affects pollen tube growth via modulating putative Ca^{2+}-permeable membrane channels and is coupled to negative regulation on glutamate decarboxylase. *J. Exp. Bot.*, 65(12), 3235–48. DOI: 10.1093/jxb/eru171. [cit. p. 159]

84. Palanivelu, R. ; Brass, L. ; Edlund, A. F. et Preuss, D. (2003). Pollen tube growth and guidance is regulated by *POP2*, an *Arabidopsis* gene that controls GABA levels. *Cell*, 114(1), 47–59. DOI: 10.1016/S0092-8674(03)00479-3. [cit. p. 159]

85. Chantha, S. ; Tebbji, F. et Matton, D. P. (2007). From the Notch signaling pathway to ribosome biogenesis. *Plant Signal. Behav.*, 2(3), 168–170. DOI: 10.4161/psb.2.3.3724. [cit. p. 159]

86. Edgar, R. ; Domrachev, M. et Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, 30(1), 207–210. DOI: 10.1093/nar/30.1.207. [cit. p. 163]

87. Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.*, 58 (301), 236–244. DOI: 10.1080/01621459.1963.10500845. [cit. p. 163]

88. O'Leary, N. A. ; Wright, M. W. ; Brister, J. R. et coll. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, 44 (D1), D733–D745. DOI: 10.1093/nar/gkv1189. [cit. p. 163]

89. Altschul, S. F. ; Gish, W. ; Miller, W. ; Myers, E. W. et Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215(3), 403–10. DOI: 10.1016/S0022-2836(05)80360-2. [cit. p. 163]

90. Götz, S. ; García-Gómez, J. M. ; Terol, J. et coll. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.*, 36(10), 3420–3435. DOI: 10.1093/nar/gkn176. [cit. p. 163]

91. Jin, J. ; Tian, F. ; Yang, D.-C. et coll. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.*, 45(D1), D1040–D1045. DOI: 10.1093/nar/gkw982. [cit. p. 163]

92. Kanehisa, M. ; Sato, Y. ; Kawashima, M. ; Furumichi, M. et Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, 44(D1), D457–D462. DOI: 10.1093/nar/gkv1070. [cit. p. 163]

93. Nielsen, H. (2017). Predicting secretory proteins with SignalP. *In* Kihara, D., éditeur : *Protein Function Prediction*, volume 1611 de *Methods in Molecular Biology*, pages 59–73. Humana Press, New York, NY, United States. ISBN: 978-1-4939-7013-1. DOI: 10.1007/978-1-4939-7015-5_6. [cit. p. 163]

94. Krogh, A. ; Larsson, B. ; von Heijne, G. et Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, 305 (3), 567–580. DOI: 10.1006/jmbi.2000.4315. [cit. p. 163]

95. El-Gebali, S. ; Mistry, J. ; Bateman, A. et coll. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.*, 47(D1), D427–D432. DOI: 10.1093/nar/gky995. [cit. p. 163]

96. Pierleoni, A. ; Martelli, P. L. et Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC Bioinf.*, 9, 392. DOI: 10.1186/1471-2105-9-392. [cit. p. 164]

97. Joly, V. et Matton, D. P. (2015). KAPPA, a simple algorithm for discovery and clustering of proteins defined by a key amino acid pattern: a case study of the cysteine-rich proteins. *Bioinformatics*, 31 (11), 1716–1723. DOI: 10.1093/bioinformatics/btv047. [cit. p. 164]

CHAPITRE 6

Identification de gènes candidats dans le transcriptome ovulaire

Ce chapitre est tiré d'un manuscrit en cours de préparation, visant à décrire les particularités du transcriptome des ovules de *Solanum chacoense*, ainsi qu'à analyser l'expression différentielle de gènes dans des ovules sauvages matures, des ovules sauvages immatures et des ovules dépourvus de sac embryonnaire issus du mutant *Scfrk1*. Nous en présentons ici le texte provisoire, en nous concentrant sur l'identification de plusieurs gènes candidats pour les interactions pollen-ovule dans notre espèce d'intérêt.

Auteurs Valentin Joly*, Yang Liu* et Daniel P. Matton (*co-premiers auteurs)

Contributions Valentin Joly est co-premier auteur avec Yang Liu ; leurs contributions au travail présenté sont équivalentes. Yang Liu a effectué la collecte des tissus, les extractions d'ARN, les RT-PCR, les pollinisations et le travail de microscopie. Valentin Joly s'est chargé du travail bioinformatique : nettoyage et assemblage des lectures de séquençage contre le génome et *de novo*, analyses d'expression différentielle, annotations structurales et fonctionnelles et calculs d'enrichissement associés, production des figures, tableaux et jeux de données. Valentin Joly et Yang Liu ont contribué à parts égales commentaire critique de ces résultats et à la rédaction du manuscrit. Daniel P. Matton a contribué à la conception du travail, et s'est chargé de sa supervision et de la révision du manuscrit.

6.1 Résumé

Les parents sauvages de la pomme de terre (*Solanum* sect. *Petota*) forment un réservoir important de germoplasme pour l'amélioration de la pomme de terre, et ont été le point focal de récentes études génomiques visant à identifier des traits d'intérêt agronomique à introduire chez *S. tuberosum*. Pourtant, ces espèces présentent des barrières d'isolement reproductif qu'il faut surmonter pour produire des hybrides interspécifiques. Parmi elles, les barrières prézygotiques postpollinisation reposant sur les interactions pollen-pistil revêtent une grande importance, en particulier le guidance du tube pollinique (TP) par les ovules. Chez *Arabidopsis*, cette interaction est médiée par des récepteurs kinases (RLK) et leurs ligands protéiques riches en cystéines (CRP) de type défensine (DEFL). Dans ce contexte, il devient nécessaire d'augmenter notre compréhension des gènes gouvernant la fonction de l'ovule chez les pommes de terre sauvages. Nous présentons ici une méthode d'assemblage hybride originale qui nous a permis de produire un transcriptome ovulaire de *S. chacoense*. Ceci nous a conduits à améliorer les annotations existantes du génome de cette espèce avec la découverte de 7044 nouveaux loci géniques, et d'identifier une liste complémentaire de 20 503 gènes *de novo*. Une recherche *in silico* des principaux composants de signalisation a établi l'existence de 2124 facteurs de transcription ainsi que de 4750 protéines empruntant la voie de sécrétion, dont 478 CRPs et 392 RLKs. Des analyses d'expression différentielle entre les ovules et les feuilles de *S. chacoense* nous ont permis de définir une liste de 4353 gènes fortement enrichis dans les ovules, ainsi qu'un sous-ensemble de 323 gènes régulés à la baisse à la fois dans (i) les ovules légèrement immatures et (ii) les ovules du mutant *frk1* dépourvus de sac embryonnaire, qui sont les uns comme les autres incapables d'attirer les TP. Il est important de constater que cette dernière liste contenait 4 protéines DEFL, qui deviennent donc de bonnes candidates pour contrôler la chimioattraction du TP. Notre étude fournit ainsi un réservoir de gènes reproductifs d'intérêt qu'il faut maintenant caractériser pour en comprendre les fonctions particulières dans l'ovule de *S. chacoense*.

Mots-clés Pommes de terre sauvages, *Solanum chacoense*, transcriptome, ovule, gamétophyte femelle, maturation du sac embryonnaire, guidage du tube pollinique, récepteurs kinases, protéines riches en cystéines.

6.2 Abstract

Wild potato relatives (*Solanum* sect. *Petota*) form a valuable germaplsm reservoir for potato breeding, and have been the focus of recent genome-wide studies aiming to identify agronomically desirable traits to be introgressed into *S. tuberosum*. Yet, those species exhibit reproductive isolation barriers that need to be overcome to produce interspecific hybrids. Among them, prezygotic postpollination barriers relying on pollen-pistil interactions are of critical importance, in particular pollen tube (PT) guidance by ovules. In *Arabidopsis*, this interaction is mediated by receptor-like kinases (RLKs) and their defensin-like (DEFL) cysteine-rich peptide (CRP) ligands. In this context, it becomes necessary to improve our understanding of genes governing ovule function in wild potatoes. Here, we present an original hybrid assembly method that allowed to produce an ovular transcriptome of *S. chacoense*. This led us to improve existing genome annotations with the discovery of 7044 new genic loci, and to identify a list of 20 503 additional *de novo* genes. *In silico* search for major signaling components identified 2124 transcription factors as well as 4750 secretory proteins including 478 CRPs and 392 RLKs. Differential gene expression analyses between *S. chacoense* ovules and leaves allowed us to define a set of 4353 genes highly enriched in ovules, as well as a list of 323 genes downregulated in (i) slightly immature and (ii) ES-less mutant *frk1* ovules, which are both unable to attract PTs. Importantly, this list contained 4 DEFL proteins, which are thus good candidate PT chemoattractants. Our study thus provides a reservoir of genes of interest that now need to be characterized in order to understand their specific functions in *S. chacoense* ovules.

Keywords Wild potatoes, *Solanum chacoense*, transcriptome, ovule, female gametophyte, embryo sac maturation, PT guidance, receptor-like kinases, cysteine-rich proteins.

6.3 Introduction

In Angiosperms, the ovule is a female reproductive structure consisting of the female gametophyte (FG), also known as the embryo sac (ES), integument(s) and nucellus. So far, plant ovules have been shown to control a wide range of pollen–pistil interactions through various key signaling molecules, including pollen tube (PT) elongation, competency control, guidance, growth arrest and discharge, as well as gamete fusion and embryonic development.[1] Interestingly, though immotile, ovules can also communicate with the male gametophyte at a distance, as observed from specific genes being induced or enhanced in the ovule by pollination events even before PTs are in their vicinity.[2,3,4] Moreover, the ES is a major ovular signaling hub where

extensive cell-cell communication events take place to effect fertilization, leading to seed set.[5]

Ovule and ES-expressed genes have been identified through expression profile comparisons between wild-type and mutant ovules in the dicot species *Arabidopsis thaliana*[6,7,8,9,10] and in monocot species like maize[11,12] and rice[13,14] using enzymatic maceration and micro-dissection techniques. Interestingly, ES-dependent and ovular sporophyte-expressed genes were also uncovered, revealing a complex gametophyte-sporophyte crosstalk.[7,8,15] Genes expressed in individual cell types of the FG have also been identified in *Arabidopsis*.[16]

In this study, we chose to focus on *Solanum chacoense*, a wild diploid potato species native to South America,[17] whose genome was recently sequenced.[18] With desirable agronomical traits, including adaptation to highly diverse environments and better disease resistance, wild potato relatives constitute a valuable germplasm reservoir for potato breeding.[19] Therefore, they were the subject of recent genome-wide sequencing analyses aiming to identify genes of interest to be introgressed into the cultivated potato[20]. The necessity to overcome interspecific hybridization barriers then becomes a challenge posed to potato breeders. In this context, a better understanding of the molecular mechanisms governing ovule development and function is of critical importance, as ovules were shown to be involved in species-specific reproductive interactions, such as PT guidance.[21,22,23]

Here, we set out to investigate gene expression in *S. chacoense* ovules by means of a hybrid transcriptomic approach. Genome-guided assembly allowed us to expand existing *S. chacoense* genome annotations with 7044 novel genes, while *de novo* assembly yielded a catalog of 20 503 additional transcript clusters originating from reads that could not map to the genome. Differential gene expression (DGE) analyses performed between ovules and leaves led us to define a set of 4353 ovule-enriched transcripts, and to discuss their enrichment in specific structural features and functional categories predicted *in silico*.

Furthermore, we explored the dynamics of the *S. chacoense* ovule transcriptome by comparing DGE in wild-type mature ovules to (i) ES-devoid ovules from the *fertilization-related kinase 1* (*Scfrk1*) mutant,[24] and (ii) slightly immature ovules obtained two days before anthesis (DBA). Comparison between these three ovule conditions allowed us to pinpoint a series of ES-specific genes, including cysteine-rich peptides (CRPs), that are good candidates to control successful, species-preferential pollen-ovule interactions.

6.4 Results & Discussion

6.4.1 Hybrid transcriptome assembly

RNA sequencing (RNA-seq) analyses were performed on *S. chacoense* genotype G4, in three ovule conditions, with three replicates per condition: ovules from flowers at anthesis (hereafter referred to as condition "Anth"), ovules from flowers collected 2 DBA (condition "2 DBA"), and ovules from *Scfrk1* flowers at anthesis (condition "*frk1*"). In parallel, RNAs from young leaves (condition "Leaf") were also sequenced in order to identify ovule-enriched transcripts in *S. chacoense*. A total of 1.3 billion 100-bp paired-end sequencing reads were obtained on a Illumina HiSeq 2000 platform, with an average of 108 million reads generated from each run (Table S6.1).

The hybrid pipeline depicted on Figure S6.1 was then developed to perform both genome-guided and *de novo* assemblies. Briefly, Illumina reads were trimmed, filtered, and aligned onto the *S. chacoense* genome sequence recently obtained by Leisner *et al.*[18] from the diploid inbred clone M6.[25] In all, 89.7% reads were mapped and used by Cufflinks[26] for genome-guided transcript assembly (Table S6.1). An additional 10.3% unmapped reads were obtained and assembled *de novo* with Trinity[27] in an effort to exploit the full potential of our data. *De novo* contigs were then clustered into groups based on sequence similarity, referred to as "*de novo* genes" in the rest of this study. To ensure data consistency and eliminate low-quality sequences, only transcripts predicted to contain a coding sequence (CDS) of 150 aa or more were reported for both genome-guided and *de novo* assemblies.

As shown on Table 6.1, our hybrid assembly of leaf and ovule tissues led us to expand existing *S. chacoense* genome annotations with 41 123 novel transcript isoforms found for known genes. Initial genome annotations made by Leisner *et al.* relied on RNA-seq data from leaves, tubers, stolons, flowers, flower buds, and fruits.[18] Here, a significant fraction of reads aligned to unannotated regions and allowed us to describe 7044 novel genes encoding 10 710 CDS-containing transcript variants. This highlights the specificity of the ovule transcriptome and its utility to improve genome annotation. The distribution of novel genes and transcripts along the reference genome can be visualized on Figure 6.1 (assembled pseudomolecules) and Figure S6.2 (unanchored scaffolds), while detailed information about gene and transcript coordinates is available in Dataset S6.1.

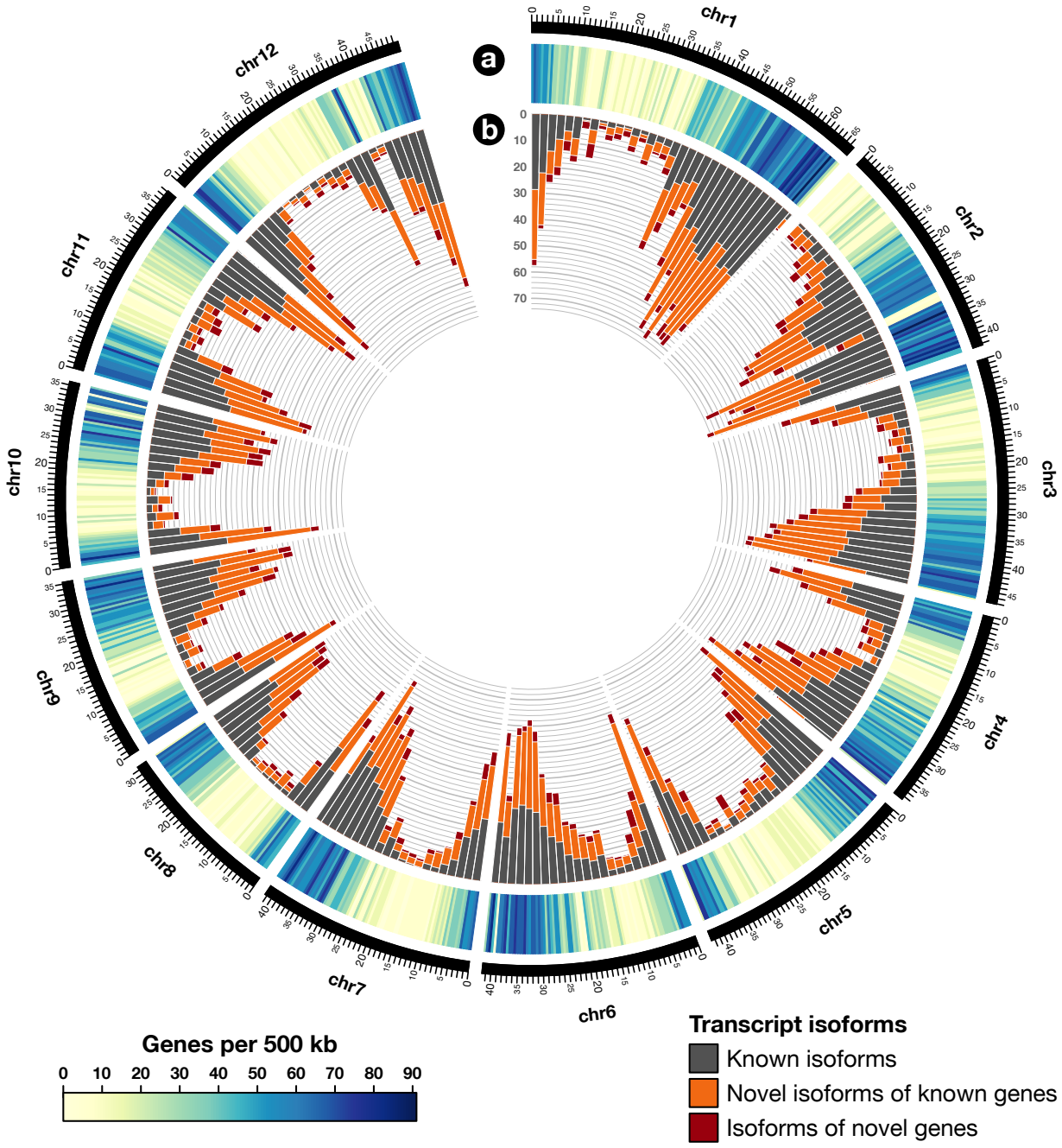


Figure 6.1. Distribution of gene and transcript annotations along *S. chacoense* chromosomes 1 to 12. (a) Heatmap depicting gene density, expressed in number of genes per 500 kb. **(b)** Histogram view of transcript annotations, in terms of number of transcripts per 1.5 Mb: in gray, transcript isoforms described by Leisner *et al.*;[18] in orange, new isoforms found for known genes; in red, isoforms of novel genes detected original to our assembly.

Table 6.1. Details about gene and transcript sequences assembled in this study.

Sequence types	Genes	Transcripts
Known genes, genome-mapped*	37 740	90 247
*Known transcript isoforms**		*49 124*
New isoforms found for known genes		*41 123*
New genes, genome-mapped	7 044	10 710
New genes, unmapped	20 503	40 478
Total sequences	**65 287**	**141 435**

*Genes and transcripts described by Leisner *et al.* (2018)

In addition, *de novo* assembly of unmapped reads allowed to obtain a list of 40 478 additional CDS-containing contigs, with no match to the reference genome, forming a set of 20 503 *de novo* genes. Those transcripts possibly result from (i) sequence divergence between the M6 and G4 genotypes, (ii) incomplete assembly of the M6 genome sequence, as the reference genome used is this study (826 Mb) is slightly shorter than the size estimated by flow cytometry (882 Mb),[18] (iii) non-nuclear RNAs (chloroplastic, mitochondrial, viral), (iv) masked regions in the reference genome, (v) sequencing and/or assembly errors stemming from our own data. Nevertheless, this supplementary gene list is of great importance to ensure a better sensitivity in our search for divergent ovule genes governing specific interactions with PTs.

6.4.2 Differential gene expression analayses

DGE analyses were performed on all genome-guided and *de novo* transcripts with the same trimmed reads used for the assembly, using the Salmon[28] and DESeq2[29] programs. Consistency of gene expression data obtained from the three biological replicates in each condition was first evaluated by means of correlation and principal component analyses. As shown in Figure S6.3, both approches revealed that replicas were highly consistent within each condition to the exception of Anth replica #3. As a consequence, gene expression values were computed without considering this replica (Dataset S6.2a). Significance of pairwise gene expression fold-changes (FCs) between conditions was assessed with the Wald test, after which p-values were adjusted using a Benjamini-Hochberg multiple testing correction. Genes with $p_{adj} \leq 0.05$ were considered to be significantly regulated. Figures 6.2 and S6.7 provide a cartography of DGE data along the reference genome sequence. The reliability of gene expression data was confirmed by semi-quantitative RT-PCR using selected genes that showed different expression patterns across the four conditions under study (Figure S6.8). Primers used for RT-PCR are listed in Table S6.2.

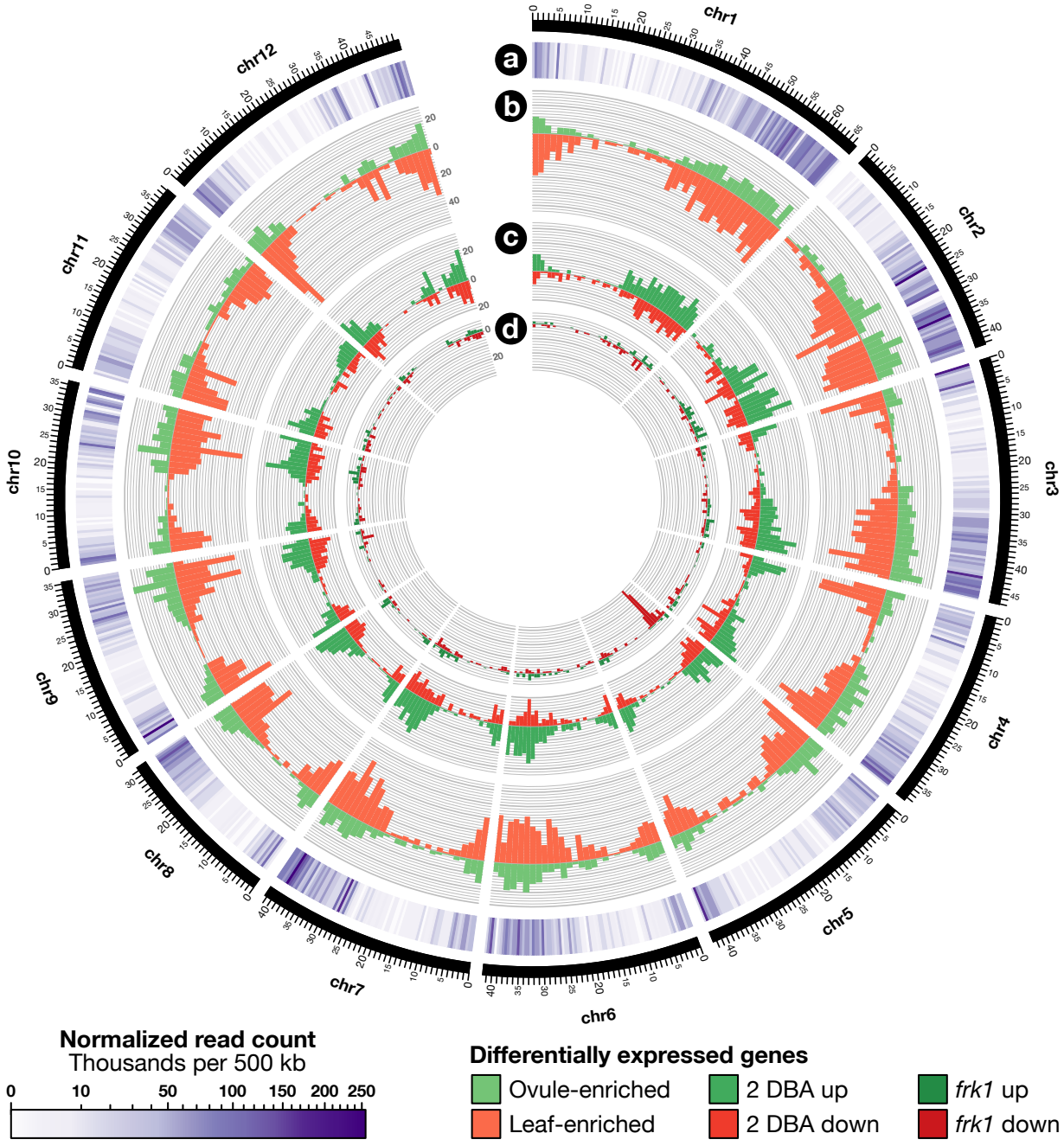


Figure 6.2. Distribution of differentially expressed genes along *S. chacoense* chromosomes 1 to 12. (a) Heatmap representing the expression level, in terms of normalized read count per 500 kb. **(b–d)** Histograms depicting the number of genes up- and down-regulated in Anth vs. Leaf, 2 DBA vs. Anth, and *frk1* vs. Anth, respectively, per 1.5 Mb.

As an initial step to dissect our transcriptome, we first compared gene expression levels between mature WT ovules and leaves, with the aim of delineating a list of transcripts that may play specific reproductive roles in the ovule. By applying an arbitrary cut-off ($FC_{Anth/Leaf} \geq 5$, $p_{adj} \leq 0.05$), we identified a list of 4353 ovule-enriched genes (Dataset S6.2b and Figure S6.4). Interestingly, half (52%) of those transcripts were encoded by novel genes identified in this study (827 genome-mapped genes and 1424 *de novo* genes).

Next, expression FCs were examined between the Anth and 2 DBA conditions. In immature 2 DBA ovules, the overall size is not significantly different from that of mature ones, but the majority of ovules have not yet completed their developmental program. Their ES is indeed still at the FG6 stage, with unfused polar nuclei, immature filiform apparatus, and not fully degenerated antipodal cells. Therefore, PTs are not attracted to 2 DBA ovules.[30] In constrat, ovules from the Anth condition exhibit a guidance-competent mature ES that has reached its final FG7 developmental stage. Previous proteomic investigation revealed that the FG6-to-FG7 maturation step affected the secretion levels of 44% of total ovule proteins emitted in the ovular locule.[31] Here, 4117 and 2815 genes with $FC_{2DBA/Anth} \geq 2$ and ≤ -2 were defined as up- and down-regulated in 2 DBA ovules, respectively (Dataset S6.2c and Figure S6.5). These two gene lists provide interesting candidates possibly mediating female gametophyte development, ovular sporophytic tissue development as well as gametophyte-sporophyte crosstalk.

The third comparison we made was between the Anth and *frk1* conditions. In the *frk1* mutant, 94% ovules do not possess an ES while the remaining 6% contain an altered embryo sac with fewer cells.[24] Consistently, *frk1* mutant ovules were shown to be unable to attract semi-*in vivo*-grown PTs.[24] This result was confirmed here on Figure 6.3, which presents *in vivo* guidance defects of WT PTs in a *frk1* ovary. DGE comparison allowed to identify 708 and 676 genes with $FC_{frk1/Anth} \geq 2$ and ≤ -2 as up- and down-regulated in the *frk1* condition, respectively (Dataset S6.2d and Figure S6.6). We may hypothesize that the latter are enriched in ES-specific genes and, to a lesser extent, in genes that depend on the regulatory activity of the FRK1 kinase. This list is of specific interest to identify genes involved in ES function, and PT-ES interactions such as PT guidance and reception, and double fertilization. In contrast, genes up-regulated in *frk1* ovules may be specific to ovular sporophytic tissues (integument and/or nucellus) and their expression may depend on signals emitted by a functional ES. Again, this is of particular interest for the identification of genes mediating female gametophyte-sporophyte crosstalk.

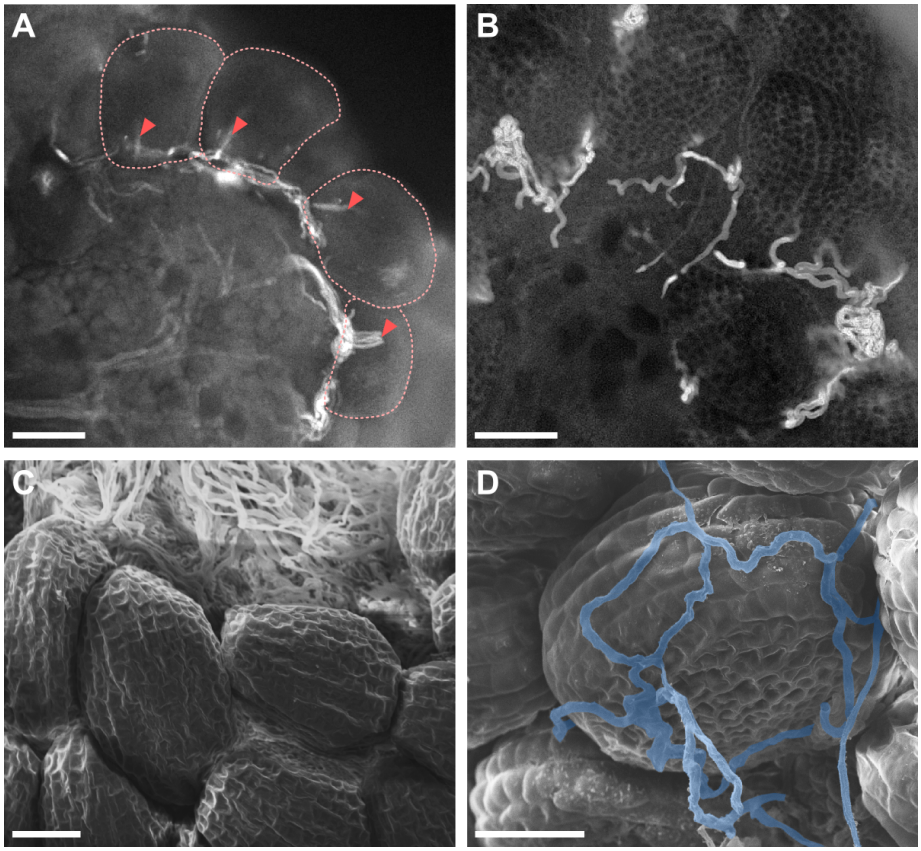


Figure 6.3. PT guidance in wild-type and *frk1* plants. Upper panels are squashes of pollinated pistils (pericarp removed). PTs were stained with aniline blue. **(a)** PT behavior near wild-type ovules. Dotted red lines indicate the contour of each single ovule. Red arrowhead indicates PTs that turned sharply to target the ovule for double fertilization. **(b)** PT behavior near *frk1* ovules. PTs migrated randomly on the surface of ovules. Lower panels are SEM images of in vivo pollinated pistils. **(c)** PT behavior near wild-type ovules showing PTs elongated along the placental epithelium, with the micropyle facing the placental side of the ovule. **(d)** PT behavior near *frk1* ovules. PTs that elongated randomly on the surface of the ovule were colored in blue. Scale bar: 100 μm.

In order to identify genes specifically required for mature ovule and mature ES functionality, especially in the context of PT guidance, we will draw special attention to the overlap between putative ES-dependant genes (down-regulated in *frk1*) and genes involved in terminal stages of ovule maturation (down-regulated in 2 DBA). As shown on Figure 6.4, 323 genes exhibited this particular expression profile, of which 218 (67%) were ovule-enriched.

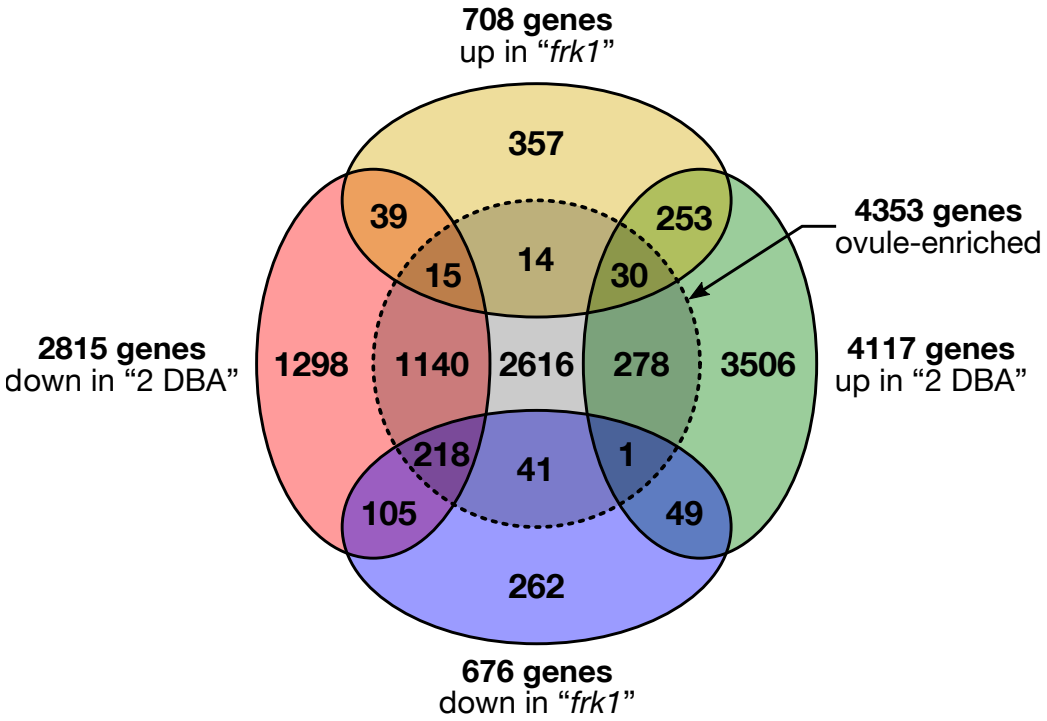


Figure 6.4. Euler diagram presenting the overlaps between regulated gene sets.

6.4.3 Functional exploration of the ovule transcriptome

In order to provide a more informative insight into our ovule transcriptome, several *in silico* analyses were accomplished on all *S. chacoense* representative protein sequences. BLAST comparison against the RefSeq protein database allowed us to find homologs in other solanaceous species and to assign descriptions to our transcripts (Dataset S6.3a), while InterProScan was used to detect specific motifs in our proteins (Dataset S6.3b). All this allowed us to associate Gene Ontology (GO) functional categories to our genes (Dataset S6.3c and d).

A Pfam enrichment analysis allowed us to identify 149 protein domains that were over-represented among ovule-enriched transcripts (Table S6.7) As shown on Figure 6.5, top enriched domains include cytochrome P450 (PF00067), SRF-type DNA-binding and dimerisa-

tion domain (PF00319), oxygenase superfamily (PF03171), plant self-incompatibility protein S1 (PF05938), and plant invertase/pectin methylesterase inhibitor (PF04043).

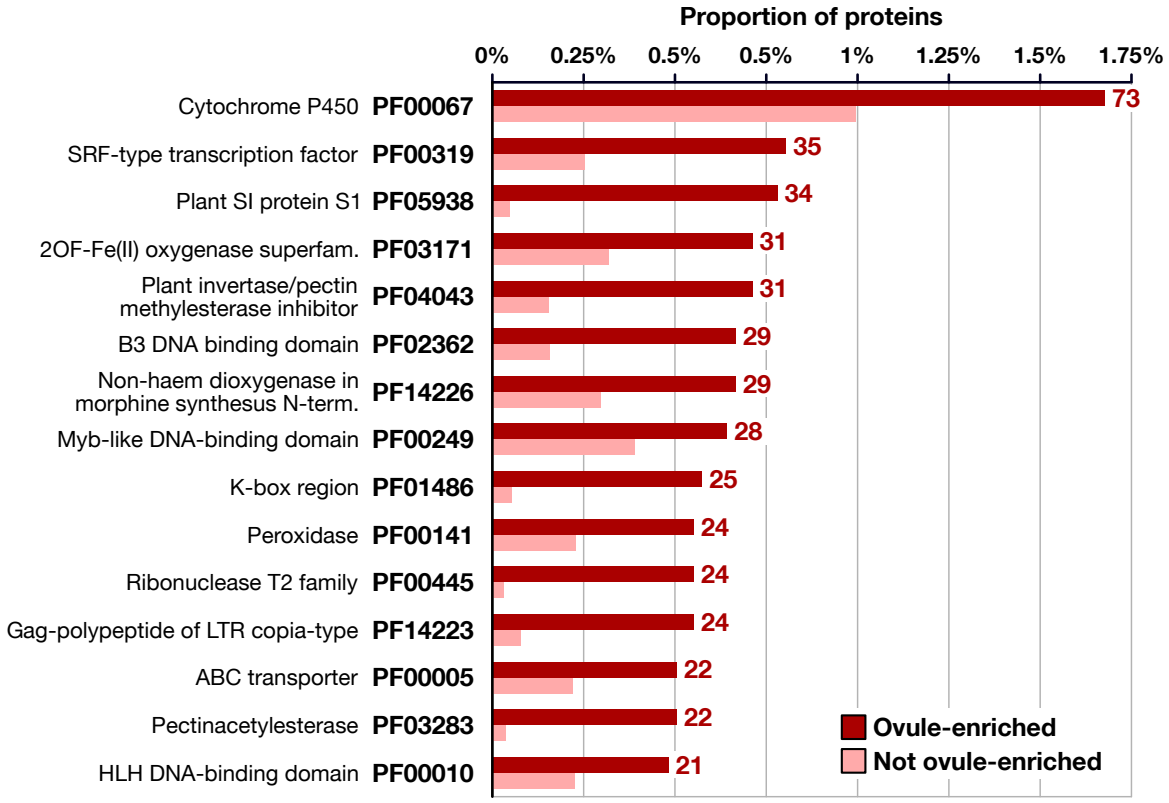


Figure 6.5. Pfam domains enriched in ovule-enriched transcripts. Number of sequences bearing each Pfam domain was counted in ovule-enriched dataset versus non-enriched. Fisher's exact test was performed to select significantly highly represented Pfam domains ($p \leq 0.05$) in ovule-enriched dataset. Only the first 15 Pfam domains that are associated with the largest number of sequences were included.

Besides, a Pfam domain characteristic for phytosulfokine precursor proteins (PSK, PF06404) also appeared to be enriched in ovules (Table S6.7). PSKs are disulfated pentapeptides that were shown to promote PT growth *in vitro*.[32,33] Among seven paralogous PSK preproproteins found in *S. chacoense*, three of them (Sc02g22130, Sc04g18405 and Sc10g15160) were highly expressed in ovules versus leaf. This finding is in line with the expression profile of PSKs in *Arabidopsis*, where PSK2 promoter activity was found in a defined region within the ovule while PSK4 was found all over mature ovules.[33]

In addition, 18 Pfam domains of unknown function (DUF) were enriched in ovule versus leaf, namely DUF563, DUF588, DUF702, DUF966, DUF1338, DUF1985, DUF2431, DUF3336, DUF3403, DUF3444, DUF4094, DUF4216, DUF4228, DUF4281, DUF4283, DUF4371, DUF4378 and DUF4666. Here we extracted 45 transcripts associated with these

DUFs. Table S6.8) provides their identity along with DGE data in order to facilitate future investigation on these unknown domains.

Moreover, a series of structural predictions were made on peptide sequences led us to identify 2261 potential metabolic enzymes (Dataset S6.3d and Table S6.6), 2124 putative transcription factors (TFs; Dataset S6.3e and Table S6.5) and 4750 proteins predicted to contain a secretory signal peptide (Dataset S6.3f). Among the latter, we detected 478 cysteine-rich proteins, 2469 *N*-glycosylated proteins, 128 glycosylphosphatidylinositol (GPI)-anchored proteins, and 1044 proteins bearing transmembrane helices (Dataset S6.3g–j and Table S6.3). Finally, examination of intracellular kinase motifs and extracellular binding domains in predicted membrane-bound proteins allowed us to identify a subset of 392 and 403 putative receptor-like kinases (RLKs) and proteins (RLPs), respectively (Dataset S6.3k and Table S6.4). Figure 6.6 presents the overlap between ovule-enriched, 2DBA- and *frk1*-down genes associated to some of those predictions.

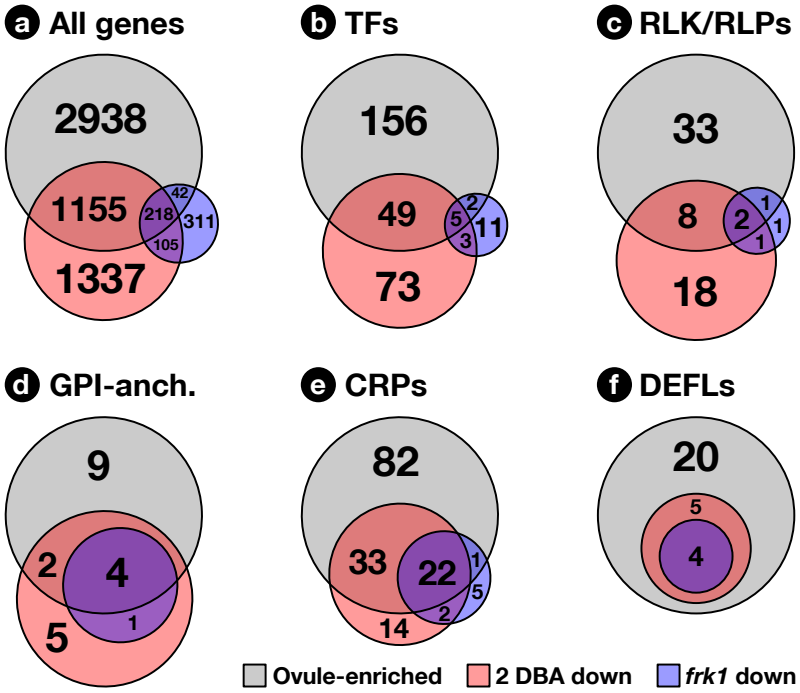


Figure 6.6. Venn diagrams depicting overlaps between ovule-enriched, 2DBA- and *frk1*-down genes associated to various predictions. (a) Whole set of genes under study. **(b)** Transcription factors. **(c)** Receptor-like kinases and proteins. **(d)** GPI-anchored proteins. **(e)** Cysteine-rich peptides. **(f)** Defensin-like peptides.

Transcription factors. In *Arabidopsis*, several TFs were shown to control gene expression during ovule development and at various stages of pollen-pistil interactions. Examples include

BELL1 (BEL1), AGAMOUS (AG) and SEEDSTICK (STK), controlling correct determination of integument identity,[34,35] SPOROCYTELESS/NOZZLE (SPL/NZZ) involved in megasporocyte development,[36,37] as well as other TFs required for micropylar PT guidance, such as MYB98[38] and CENTRAL CELL GUIDANCE (CCG).[39]

To better understand gene regulation during ovule development in *S. chacoense*, a genome-wide search for TFs was performed using the 674 non-redundant, high quality DNA-binding motifs provided the Plant TF database (PlantTFDB) v. 4.0,[40]) which led us to generate a list of 2124 TFs from *S. chacoense*, whose size is comparable to the 2451 TFs predicted in *Arabidopsis*[41] (Dataset S6.3e). Of these, 212 (10%) were ovule-enriched (Figure 6.6b and Table S6.5). Consistently, Pfam domain enrichment analyses revealed that five DNA-binding domains typical for TFs were significantly over-represented among ovule-enriched genes (Table S6.7), including SRF-type DNA-binding and dimerisation domain (PF00319), B3 DNA binding domain (PF02362), Myb-like DNA-binding domain (PF00249), K-box region (PF01486), and helix-loop-helix DNA-binding domain (PF00010). This leads us to speculate their involvement in ovule development or reproduction. In particular, 5 predicted ovule-enriched TFs were also down-regulated in both *frk1* and 2 DBA conditions. Among them, Sc12g13610 was predicted to be a MYB domain-containing transcription factor resembling MYB108, which was shown to control stamen and pollen development in *Arabidopsis*.[42] Besides, Sc01g33850 encodes a predicted MADS transcription factor with similarity to AGAMOUS-LIKE 62 (AGL62), which is involved in central cell proliferation[43] and underlying interspecific postzygotic barriers[44] in *Arabidopsis*.

Receptor-like kinases. In *Arabidopsis*, several RLKs were found to play an important role in ovule development, including STRUBBELIG (SUB),[45] ARABIDOPSIS CRINKLY4 (ACR4),[46] and ERECTA-family genes (ER and ERLs).[47] During pollen-pistil interactions, RLKs and their secreted ligands were demonstrated to play key roles in the ability of both sexual partners to perceive signaling molecules emitted by the other.[48] For example, on the male side, several *Arabidopsis* leucine-rich repeat (LRR)-RLKs were shown to be involved in PT response to ovular LURE chemoattratants: POLLEN RECEPTOR KINASE 6 (PRK6), which functions in cooperation with other members of the PRK family,[49] and the closely related MALE DISCOVERER 1 (MDIS1) and MDIS2.[50] In *Arabidopsis*, 4 pollen RLKs of the *Catharanthus roseus* RLK1-LIKE (CrRLK1L) family, ANXUR1 (ANX1) and ANX2 and BUDDHA'S PAPER SEAL 1 (BUPS1) and BUPS2, were found to regulate PT integrity under the control of RALF peptides.[51] Fewer RLKs were described on the female side. FERONIA (FER), another CrRLK1L, is expressed by synergid cells, localized at the

filiform apparatus, and required for correct PT reception probably under the control of RALF peptide(s) yet to be identified.[52] Two other ovule CrRLK1Ls, HERCULES RECEPTOR KINASE 1 (HERK1) and ANJEA (ANJ), were recently found to act cooperatively with FER for PT reception.[53]

The *Arabidopsis* RLK family consists of over 600 members[54] while the rice genome encodes more than 1000 RLKs[55]. However, genome-wide identification of RLKs has never been performed in *S. chacoense*. We therefore surveyed our transcriptome for putative RLKs, i.e. proteins predicted to contain a signal peptide, a single transmembrane helix, and an intracellular kinase domain. This led us to generate a list of 392 putative RLKs in *S. chacoense*. Domain predictions on the extracellular parts of those proteins allowed to define different subsets of RLKs, the two largest being LRR-RLKs and lectin-RLKs, with 153 and 117 members, respectively. Importantly, two previously described *S. chacoense* leucine-rich repeat receptor-like kinases (LRR-RLKs) predominantly expressed in the ovary, OVULE RECEPTOR KINASE 17[56] and 28[57] (ScORK17 and 28), were correctly assembled as Sc01g23160 and Sc03g28590, respectively, and detected among predicted RLKs. An additional list of 403 receptor-like proteins (RLPs), which possessed all expected features except the intracellular kinase domain, was also made (Dataset S6.3k and Table S6.4). As shown on Figure 6.6c, 44 putative RLKs/RLPs were found to be ovule-enriched, while 22 exhibited enhanced expression during the late ES maturation step (i.e., down-regulated in 2 DBA vs. Anth) and 5 were potentially ES-expressed (i.e., down-regulated in *frk1* vs. Anth). Only two RLKs (Sc04g18140 and ScUng65060) and one RLP (ScUng40850) were down-regulated in both 2 DBA and *frk1* conditions. Interestingly, ScUng65060 exhibits a malectin domain in its extracellular region. This feature is also found in CrRLK1L-type receptors like FER.[58]

GPI-anchored proteins. In parallel to RLKs, another important surface receptors were found to play an important role in reproductive development. Maternal GPI-anchored proteins such as the synergid-expressed LORELEI (LRE)[59] and ENODLs[60] were found to associate to the juxtamembrane portion of the FER ectodomain, and be reequired for correct PT reception. In the *S. chacoense* genome, 128 proteins carrying both a signal peptide and a predicted GPI anchor in the mature sequence were detected Dataset S6.3i). Of these, 13 putative GPI-anchored proteins were ovule-enriched, 12 were down-regulated in 2 DBA and 5 were down-regulated in *frk1* (Figure 6.6d and Table S6.3). Of the latter, 4 were also down-regulated in 2 DBA and ovule-enriched (Sc07g18620, Sc10g02930, Sc12g18655, Sc12g18660). Interestingly, BLASTp detected a similarity between Sc10g02930 and LORELEI-LIKE GPI-ANCHORED PROTEIN 2 (LLG2; Dataset S6.3a).

Cysteine-rich peptides. CRPs form a large protein category sharing three common features: (i) a small size, (ii) the presence of a secretory signal peptide and (iii) a high number of cysteine-stabilized disulfide bridges whose organization is highly conserved within each CRP family. In the mature peptide, amino acids in between cysteine residues can be largely divergent, which enables CRPs to exert highly specific recognition functions for plant reproduction, development, and immunity.[61] Indeed, CRPs from different families were shown to be secreted ligands of critical importance in the context of pollen-pistil interactions.[62] For instance, the lily SCA and *Arabidopsis* LTP5 lipid-transfer proteins control stylar PT guidance and adhesion.[63,64] In tomatoes, pollen LAT52 and stylar LeSTIG are able to bind to PT-expressed LePRK receptors.[65,66]. SP11/SCR and PrsS mediate self-incompatibility in the *Brassicaceae* and *Papaver*, respectively.[67,68]

CRPs are also known to mediate several key pollen-ovule interactions, and arge-scale analyses performed in *Arabidopsis* revealed that 53% of the FG-specific genes encode CRPs.[69] In *Arabidopsis*, egg cell-expressed EC1 proteins are involved in the control of gamete fusion.[70] Interestingly, apart from EC1, other CRPs identified in the context of pollen-ovule interactions all belong the defensin-like (DEFL) family. In *Torenia fournieri* and *T. concolor*, synergid cell-expressed TfLURE and TcLURE peptides are responsible for micropylar species-specific PT guidance.[21,71] In *Arabidopsis thaliana*, the AtLUREs, a distinct class of DEFLs secreted by synergids, were found to be involved in PT guidance, again in a species-prefential manner.[22] More recently, the discovery of non-species-specific *Arabidopsis* PT attractants XIUQIU1–4, which belong to the DEFL family as well, allowed to redefine AtLUREs as ovular mediators of reproductive isolation that accelerate the emergence of conspecific PTs onto the septum surface.[23] Besides PT guidance, ZmES4, a DEFL protein expressed in the FG, is known to control PT burst in maize.[72]

Here *in silico* predictions allowed to detect *478* CRPs in the *S. chacoense* genome, defined as proteins with a predicted signal peptide, a mature part of $\leq$150 aa, and $\geq$6 cysteines. This number is in line with the 825 and 598 CRPs detected in the *Arabidopsis* and rice genomes, respectively.[73] As shown on Figure 6.6e, 138 were ovule-enriched, 71 were down-regulated in 2 DBA ovules, and 30 were down-regulated in *frk1* ovules, hence potentially ES-expressed CRPs. Twenty-two (22) CRPs combined the three criteria and are of particular interest to us. Among them, ScUng36640, Sc04g13560 and ScUng36560 encode proteins similar to EC1, while Sc07g13940 belongs to the LTP subgroup. Importantly, this small set also contained 4 DEFL proteins with unknown functions, Sc04g14918, Sc09g21242, Sc10g13555, and ScDng02289, which contain 8 to 12 cysteines in their mature part (Figure 6.6f). Those CRPs

are good candidates to be PT chemoattractants, since they combine the structural features of DEFLs and a DGE profile consistent with the inability of 2 DBA and *frk1* ovules to attract PTs. Besides, a total number of 58 DEFLs were identified in the *S. chacoense* genome. Interestingly, they exhibited extensive sequence divergence despite a conserved cysteine backbone (Figure 6.7), and 29 of them were enriched in ovules compared to leaves, including the 4 PT guidance candidates mentioned above, as well as 25 DEFLs that were not affected in the *frk1* mutant. The latter could be expressed in sporophytic tissues and play key roles in ovule function and in its interactions with PTs (see also Figure 7.3 in the next chapter).

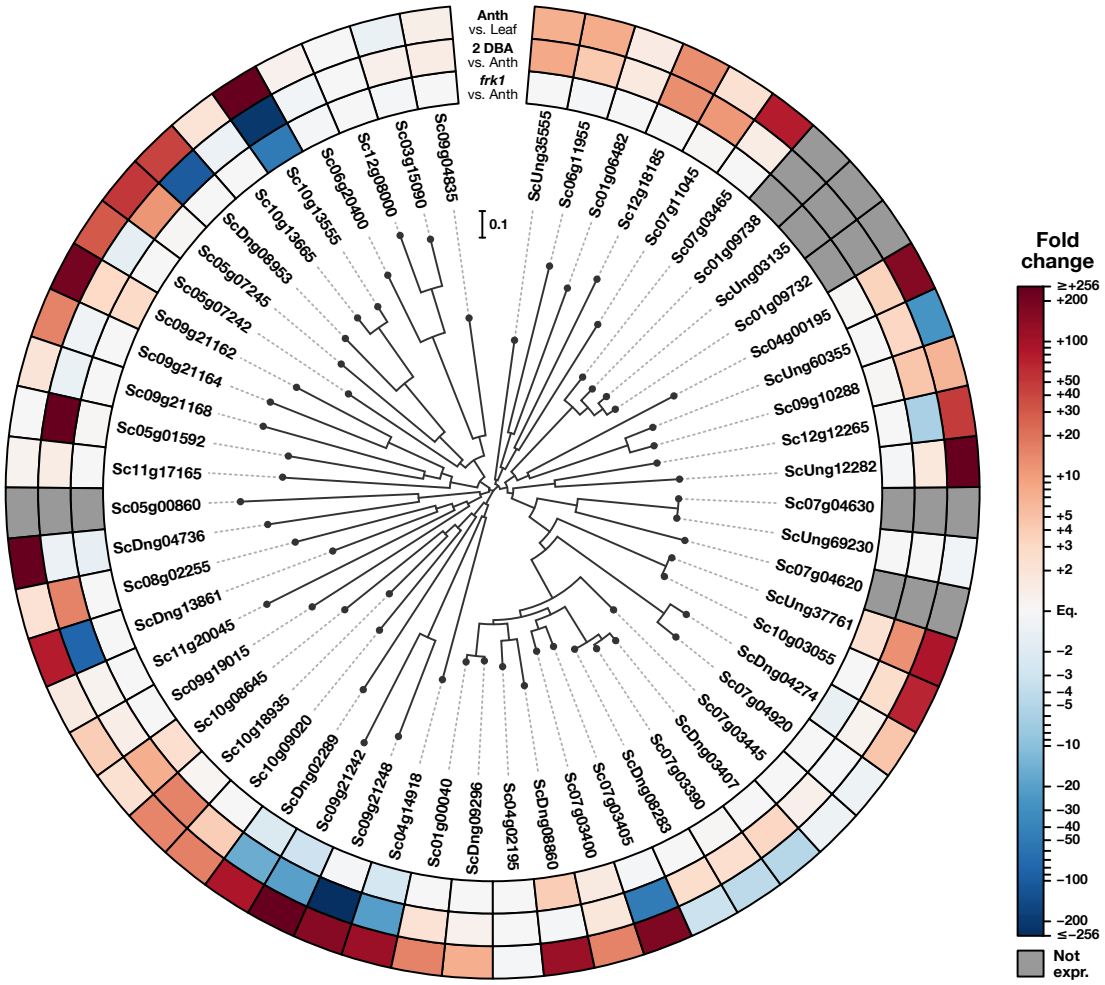


Figure 6.7. Phylogenetic relationships and DGE data of *S. chacoense* DEFLs. Mature aminoacid sequences were aligned with KAPPA and MUSCLE and used to build a neighbor-joining tree. DGE data were then added to the graph.

Further functional assays are now required to assess the exact functions of those candidates genes in ovule development and function. In this effort, we can rely on the advancements made in genome-editing of potatoes using the CRISPR-Cas9 system.[74,75,76]

6.5 Materials & Methods

6.5.1 Plant material and tissue collection

S. chacoense Bitt. ($2n = 2x = 24$) individuals were greenhouse-grown at ~25 °C under long-day conditions (16 h light/8 h dark). The genotypes used were G4, a wild-type genotype derived from PI458314 and PI230582 accessions of the NRSP-6 United States Potato Genebank (Sturgeon Bay, WI, USA), and the *Scfrk1*-S1 knock-down mutant line, which was obtained previously in a G4 background.[24] Ovule samples from the "Anth", "2 DBA", and "*frk1*" were collected by dissecting ovaries to remove the pericarp and flash-frozen in liquid nitrogen upon harvest. For the "Leaf" condition, young G4 leaves were collected and ground in liquid nitrogen. Three biological replicated were made for each condition. Total RNAs were extracted using the TRIzol reagent (Fisher Scientific, Ottawa, ON, Canada) following the manufacturer's instructions, followed by a quick cleanup step with the RNeasy MinElute Cleanup Kit (QIAgen Inc., Toronto, ON, Canada).

6.5.2 RNA sequencing

Total RNAs were sent to the Génome Québec Innovation Center (Montréal, QC, Canada), where cDNA library construction was performed with the TruSeq Stranded mRNA Preparation Kit (Illumina, San Diego, CA, USA). A total of 12 indexed libraries were prepared, one for each replicate, and underwent quality control. Libraries were sequenced on a Illumina HiSeq 2000 platform in a multiplexed setup with four libraries per lane.

6.5.3 Transcriptome assembly

Raw Illumina reads were trimmed and filtered with SHEAR v. 2015-09-13[77] and Scythe v. 0.994 BETA,[78] and then aligned onto the *S. chacoense* M6 genome sequence (chromosomes 1 to 12 and unanchored scaffolds) with the STAR aligner v. 2.5.3a.[79] Genome-guided transcript assembly was performed with Cufflinks v. 2.2.1,[26] using a maximum intron size of 10 000 bp. Genome annotations produced by Leisner *et al.* (2018)[18] were used by Cufflinks for Reference Annotation Based Transcript (RABT) assembly,[80] and sequence repeats described by the same authors were used as a mask. Unmapped reads were assembled *de novo* with Trinity v. 2.8.4.[27] Both genome-guided and *de novo* assemblies were then filtered to eliminate short transcripts

that did not contain a coding sequence of at least 150 bp. *De novo* transcripts were realigned onto the reference genome with GMAP v. 2018-07-04.[81] Those that had a significant alignment to a genomic locus were considered to be genome-mapped, and were merged together with Cufflinks assemblies into one final mapped assembly using Cuffmerge v. 2.2.1.[26] Remaining unmapped *de novo* transcripts were clustered with CD-HIT v. 4.6[82,83]. Two transcripts were placed in the same group if their shared ≥99% aminoacid similarity on their longest predicted coding sequences.

6.5.4 Sequence nomenclature

Gene identifiers were defined systematically in the form *ScAAgBBBBC*, where *AA* specifies the chromosome (a number from 01 to 12, or *Un* for unanchored scaffolds), *BBBB* is a number between 0000 and 9999 corresponding to the gene rank along the chromosome sequence, and *C* is a digit equal to 0 for known genes (i.e., described by Leisner *et al.*) or varying between 1 and 9 for novel genome-mapped genes described in this study. *De novo* gene identifiers, on the other hand, were written in the form *ScDngNNNNN*, where *NNNNN* is the cluster rank in the CD-HIT output.

Transcript identifiers consist of the gene identifier followed by a suffix in the form *.D*, where *D* is the transcript isoform number. For each gene, known transcript isoforms were numbered first, according to Leisner *et al.*'s data. Then, if any, novel transcript isoforms were sorted by total length and numbered consecutively. For each gene, the transcript ending in *.1* was considered to be the representative one. For each transcript, CDSs and corresponding peptides received identifiers made of the transcript identifier followed by a suffix in the form *.cdsE* and *.pepE*, respectively, where *E* is the CDS rank.

6.5.5 Transcriptome annotation

In silico annotations were assigned to the longest peptide sequence predicted in each representative transcript. Protein sequences were first compared to the RefSeq database rel. 93 of the National Center for Biotechnological Information (NCBI)[84] with BLASTp v. 2.8.1+.[85] Domains from different reference databases, including Pfam v. 32.0,[86] were predicted with InterProScan (IPS) v. 5.34-73.0.[87] BLASTp and IPS results were used for Gene Ontology (GO) term and Enzyme Code (EC) assignment with Blast2GO CLI v. 1.4.4.[88] Enzyme codes were mapped to descriptions available in the BRENDA database rel. 2018.2[89] and metabolic path-

ways defined in the KEGG database rel. 90.1.[90] Putative TFs were identified with the PlantTFDB v. 4.0 prediction tool.[40] The presence of N-terminal secretory signal peptides (SPs) in *S. chacoense* protein sequences was predicted with SignalP v 5.0.[91] Proteins predicted to contain SPs were further checked for the presence of putative *N*-glycosylation sites with NetNGlyc v. 1.0[92] and transmembrane helices (TMHs) with TMHMM v. 2.0[93] in the mature peptide. Proteins with one predicted TMH were further classified as putative receptor-like kinases (RLKs) or receptor-like proteins (RLPs) based on the presence or absence of a predicted intracellular kinase domain (Pfam motifs PF00069 and PF07714). Putative RLKs/RLPs were also checked for the presence of extracellular LRR (PF00560, PF01462, PF01463, PF01816, PF07723, PF07725, PF08263, PF12799, PF13306, PF13516, PF13855, PF14580, PF18805, PF18831, PF18837), malectin (PF11721, PF12819), lectin (PF00059, PF00139, PF01453) and LysM (PF01476) domains. Proteins with no predicted TMH were checked for the presence of glycosylphosphatidylinositol (GPI) anchoring sites with PredGPI.[94] Those with no GPI were analyzed with KAPPA v. 2.0[95] and classified as cysteine-rich proteins (CRPs) if they had a mature length of 150 aa at most, with at least 6 cysteines. KAPPA was also used to detect DEFLs, defined as CRPs containing the following pattern: `X{n}CX{n}CX{3}CX{1-}CX{1-}CXCX{n}`.

6.5.6 Differential expression analysis

Salmon v. 0.13.1[28] was used to perform "quasi-mapping" of reads onto both genome-mapped and *de novo* transcript sequences. Read quantification data was then analyzed with DESeq2 v. 1.22.2[29]. Gene-wise normalized read counts for each sample were used to compute pairwise correlation coefficients and to perform a principal component analysis (Figure S6.3), leading to the rejection of Anth replica #3, as discussed earlier. Final fold-change (FC) values were therefore computed after redoing read count normalization without this sample. Binary logarithms of FC values were then shrunken using the APEGLM method[96]. Significance of FCs was assessed with the Wald test, after which *p*-values were adjusted using the Benjamini-Hochberg multiple testing correction method.

Semi-quantitative RT-PCR validation of differential gene expression data was performed on selected genes using 70 ng of total RNA from anthesis, 2DBA, *frk1* ovules and leaf in biological triplicates using the M-MLV reverse transcriptase kit (Invitrogen). The cDNA templates were amplified for 26 cycles for all genes tested. Ubiquitin was used as the internal control. Primer sequences used for RT-PCR analyses are listed in Table S6.2.

6.5.7 Imaging of pollen–ovule interactions *in vivo*

Wild-type flowers (*S. chacoense* genotype G4, S_{12} and S_{14} self-incompatibility alleles) were pollinated with compatible pollen (*S. chacoense* genotype V22, S_{11} and S_{13} alleles). For the optical imaging, pistils were fixed, squashed and stained 40 h after pollination as described previously.[97] The pericarps were carefully removed before squashing the pistils and staining. PT behavior near the ovule was observed with an Axio Imager.M1 microscope equipped with an Axiocam HRc camera (Zeiss). For SEM imaging, pistils were fixed, dehydrated and critical-point-dried as described previously,[98] and mounted on double-taped SEM stubs. Pericarps were dissected to reveal fertilization events prior to coating. Pictures were acquired with a JEOL JSM-35 SEM apparatus.

6.6 Data Availability

Raw sequence data used in this study have been submitted to the NCBI sequence read archive (`http://www.ncbi.nlm.nih.gov/sra`) under the following accession numbers: SRR2782415–9, SRR2782438, SRR2782462, SRR2782513, SRR2782581, and SRR2782690–2.

Bibliography

1. Johnson, M. A. ; Harper, J. F. et Palanivelu, R. (2019). A fruitful journey: pollen tube navigation from germination to fertilization. *Annu. Rev. Plant Biol.*, 70(1). DOI: 10.1146/annurev-arplant-050718-100133. [cit. p. 174]

2. Dong, Y.-H. ; Kvarnheden, A. ; Yao, J.-L. et coll. (1998). Identification of pollination-induced genes from the ovary of apple (*Malus domestica*). *Sex. Plant Reprod.*, 11(5), 277–83. DOI: 10.1007/s004970050154. [cit. p. 174]

3. Lantin, S. ; O'Brien, M. et Matton, D. P. (1999). Pollination, wounding and jasmonate treatments induce the expression of a developmentally regulated pistil dioxygenase at a distance, in the ovary, in the wild potato *Solanum chacoense* Bitt. *Plant Mol. Biol.*, 41(3), 371–386. DOI: 10.1023/A:1006375522626. [cit. p. 174]

4. Joly, V. ; Tebbji, F. ; Nantel, A. et Matton, D. P. (2019). Pollination type recognition from a distance by the ovary is revealed through a global transcriptomic analysis. *Plants*, 8(6), 185. DOI: 10.3390/plants8060185. [cit. p. 174]

5. Sundaresan, V. et Alandete-Saez, M. (2010). Pattern formation in miniature: the female gametophyte of flowering plants. *Development*, 137(2), 179–89. DOI: 10.1242/dev.030346. [cit. p. 175]

6. Yu, H.-J. ; Hogan, P. et Sundaresan, V. (2005). Analysis of the female gametophyte transcriptome of *Arabidopsis* by comparative expression profiling. *Plant Physiol.*, 139(4), 1853–1869. DOI: 10.1104/pp.105.067314. [cit. p. 175]

7. Johnston, A. J. ; Meier, P. ; Gheyselinck, J. et coll. (2007). Genetic subtraction profiling identifies genes essential for *Arabidopsis* reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. *Genome Biol.*, 8(10), R204. DOI: 10.1186/gb-2007-8-10-r204. [cit. p. 175]

8. Jones-Rhoades, M. W. ; Borevitz, J. O. et Preuss, D. (2007). Genome-wide expression profiling of the *Arabidopsis* female gametophyte identifies families of small, secreted proteins. *PLoS Genet.*, 3 (10), 1848–61. DOI: 10.1371/journal.pgen.0030171. [cit. p. 175]

9. Steffen, J. G. ; Kang, I.-H. ; Macfarlane, J. et Drews, G. N. (2007). Identification of genes expressed in the *Arabidopsis* female gametophyte. *Plant J.*, 51(2), 281–92. DOI: 10.1111/j.1365-313X.2007.03137.x. [cit. p. 175]

10. Sánchez-León, N. ; Arteaga-Vázquez, M. ; Alvarez-Mejía, C. et coll. (2012). Transcriptional analysis of the *Arabidopsis* ovule by massively parallel signature sequencing. *J. Exp. Bot.*, 63(10), 3829–42. DOI: 10.1093/jxb/ers075. [cit. p. 175]

11. Yang, H. ; Kaur, N. ; Kiriakopolos, S. et McCormick, S. (2006). EST generation and analyses towards identifying female gametophyte-specific genes in *Zea mays* L. *Planta*, 224(5), 1004–14. DOI: 10.1007/s00425-006-0283-3. [cit. p. 175]

12. Wang, S. S. ; Wang, F. ; Tan, S. J. et coll. (2014). Transcript profiles of maize embryo sacs and preliminary identification of genes involved in the embryo sac-pollen tube interaction. *Front. Plant Sci.*, 5, 702. DOI: 10.3389/fpls.2014.00702. [cit. p. 175]

13. Ohnishi, T. ; Takanashi, H. ; Mogi, M. et coll. (2011). Distinct gene expression profiles in egg and synergid cells of rice as revealed by cell type-specific microarrays. *Plant Physiol.*, 155(2), 881–91. DOI: 10.1104/pp.110.167502. [cit. p. 175]

14. Kubo, T. ; Fujita, M. ; Takahashi, H. et coll. (2013). Transcriptome analysis of developing ovules in rice isolated by laser microdissection. *Plant Cell Physiol.*, 54(5), 750–65. DOI: 10.1093/pcp/pct029. [cit. p. 175]

15. Armenta-Medina, A. ; Huanca-Mamani, W. ; Sanchez-León, N. ; Rodríguez-Arévalo, I. et Vielle-Calzada, J.-P. (2013). Functional analysis of sporophytic transcripts repressed by the female gametophyte in the ovule of *Arabidopsis thaliana*. *PLoS One*, 8(10), e76977. DOI: 10.1371/journal.pone.0076977. [cit. p. 175]

16. Wuest, S. E. ; Vijverberg, K. ; Schmidt, A. et coll. (2010). *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr. Biol.*, 20(6), 506–512. DOI: 10.1016/j.cub.2010.01.051. [cit. p. 175]

17. Hijmans, R. J. ; Spooner, D. M. ; Salas, A. R. ; Guarino, L. et de la Cruz, J. (2002). *Atlas of wild potatoes.*, volume 10 de *Systematic and ecogeographic studies on crop genepools*. International Plant Genetic Resources Institute, Rome, Italie. ISBN: 978-92-9043-518-1. URL: https://www.bioversityinternational.org/e-library/publications/detail/atlas-of-wild-potatoes/. [cit. p. 175]

18. Leisner, C. P. ; Hamilton, J. P. ; Crisovan, E. et coll. (2018). Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. *Plant J.*, 94(3), 562–70. DOI: 10.1111/tpj.13857. [cit. p. 175, 176, 177, 178, 189]

19. Hardigan, M. A. ; Laimbeer, F. P. E. ; Newton, L. et coll. (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U. S. A.*, 114(46), E9999–10008. DOI: 10.1073/pnas.1714380114. [cit. p. 175]

20. Li, Y. ; Colleoni, C. ; Zhang, J. et coll. (2018). Genomic analyses yield markers for identifying agronomically important genes in potato. *Mol. Plant*, 11(3), 473–84. DOI: 10.1016/j.molp.2018.01.009. [cit. p. 175]

21. Okuda, S. ; Tsutsui, H. ; Shiina, K. et coll. (2009). Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature*, 458(7236), 357–61. DOI: 10.1038/nature07882. [cit. p. 175, 187]

22. Takeuchi, H. et Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol.*, 10(12), e1001449. DOI: 10.1371/journal.pbio.1001449. [cit. p. 175, 187]

23. Zhong, S. ; Liu, M. ; Wang, Z. et coll. (2019). Cysteine-rich peptides promote interspecific genetic isolation in *Arabidopsis*. *Science*, 364(6443). DOI: 10.1126/science.aau9564. [cit. p. 175, 187]

24. Lafleur, E. ; Kapfer, C. ; Joly, V. et coll. (2015). The FRK1 mitogen-activated protein kinase kinase kinase (MAPKKK) from *Solanum chacoense* is involved in embryo sac and pollen development. *J. Exp. Bot.*, 66(7), 1833–43. DOI: 10.1093/jxb/eru524. [cit. p. 175, 180, 189]

25. Jansky, S. H. ; Chung, Y. S. et Kittipadukal, P. (2014). M6: a diploid potato inbred line for use in breeding and genetics research. *J. Plant Registr.*, 8(2), 195. DOI: 10.3198/jpr2013.05.0024crg. [cit. p. 176]

26. Trapnell, C. ; Williams, B. A. ; Pertea, G. et coll. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, 28(5), 511–5. DOI: 10.1038/nbt.1621. [cit. p. 176, 189, 190]

27. Grabherr, M. G. ; Haas, B. J. ; Yassour, M. et coll. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.*, 29(7), 644–52. DOI: 10.1038/nbt.1883. [cit. p. 176, 189]

28. Patro, R. ; Duggal, G. ; Love, M. I. ; Irizarry, R. A. et Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, 14(4), 417–9. DOI: 10.1038/nmeth.4197. [cit. p. 178, 191]

29. Love, M. I. ; Huber, W. et Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15(12), 550. DOI: 10.1186/s13059-014-0550-8. [cit. p. 178, 191]

30. Liu, Y. (2015). *The plant ovule omics: an integrative approach for pollen-pistil interactions and pollen tube guidance studies in solanaceous species*. Thèse de doctorat, Université de Montréal. URL: http://hdl.handle.net/1866/13589. [cit. p. 180]

31. Liu, Y. ; Joly, V. ; Dorion, S. ; Rivoal, J. et Matton, D. P. (2015). The plant ovule secretome: a different view toward pollen–pistil interactions. *J. Proteome Res.*, 14(11), 4763–4775. DOI: 10.1021/acs.jproteome.5b00618. [cit. p. 180]

32. Matsubayashi, Y. (2014). Posttranslationally modified small-peptide signals in plants. *Annu. Rev. Plant Biol.*, 65, 385–413. DOI: 10.1146/annurev-arplant-050312-120122. [cit. p. 183]

33. Stührwohldt, N. ; Dahlke, R. I. ; Kutschmar, A. et coll. (2015). Phytosulfokine peptide signaling controls pollen tube growth and funicular pollen tube guidance in *Arabidopsis thaliana*. *Physiol. Plant.*, 153(4), 643–53. DOI: 10.1111/ppl.12270. [cit. p. 183]

34. Reiser, L. ; Modrusan, Z. ; Margossian, L. et coll. (1995). The *BELL1* gene encodes a homeodomain protein involved in pattern formation in the *Arabidopsis* ovule primordium. *Cell*, 83(5), 735–42. DOI: 10.1016/0092-8674(95)90186-8. [cit. p. 185]

35. Brambilla, V. ; Battaglia, R. ; Colombo, M. et coll. (2007). Genetic and molecular interactions between BELL1 and MADS box factors support ovule development in *Arabidopsis*. *Plant Cell*, 19 (8), 2544–56. DOI: 10.1105/tpc.107.051797. [cit. p. 185]

36. Schiefthaler, U. ; Balasubramanian, S. ; Sieber, P. et coll. (1999). Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.*, 96(20), 11664–9. DOI: 10.1073/pnas.96.20.11664. [cit. p. 185]

37. Yang, W.-C. ; Ye, D. ; Xu, J. et Sundaresan, V. (1999). The *SPOROCYTELESS* gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev.*, 13(16), 2108–17. DOI: 10.1101/gad.13.16.2108. [cit. p. 185]

38. Kasahara, R. D. ; Portereiko, M. F. ; Sandaklie-Nikolova, L. ; Rabiger, D. S. et Drews, G. N. (2005). *MYB98* is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell*, 17(11), 2981–92. DOI: 10.1105/tpc.105.034603. [cit. p. 185]

39. Chen, Y.-H. ; Li, H.-J. ; Shi, D.-Q. et coll. (2007). The central cell plays a critical role in pollen tube guidance in *Arabidopsis*. *Plant Cell*, 19(11), 3563–77. DOI: 10.1105/tpc.107.053967. [cit. p. 185]

40. Jin, J. ; Tian, F. ; Yang, D.-C. et coll. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.*, 45(D1), D1040–D1045. DOI: 10.1093/nar/gkw982. [cit. p. 185, 191]

41. Pérez-Rodríguez, P. ; Riaño-Pachón, D. M. ; Corrêa, L. G. G. et coll. (2010). PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.*, 38(Database issue), D822–7. DOI: 10.1093/nar/gkp805. [cit. p. 185]

42. Mandaokar, A. et Browse, J. (2009). MYB108 acts together with MYB24 to regulate jasmonate-mediated stamen maturation in *Arabidopsis*. *Plant Physiol.*, 149(2), 851–62. DOI: 10.1104/pp.108.132597. [cit. p. 185]

43. Kang, I.-H. ; Steffen, J. G. ; Portereiko, M. F. ; Lloyd, A. et Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell*, 20(3), 635–47. DOI: 10.1105/tpc.107.055137. [cit. p. 185]

44. Walia, H. ; Josefsson, C. ; Dilkes, B. et coll. (2009). Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr. Biol.*, 19(13), 1128–32. DOI: 10.1016/j.cub.2009.05.068. [cit. p. 185]

45. Chevalier, D. ; Batoux, M. ; Fulton, L. et coll. (2005). STRUBBELIG defines a receptor kinase-mediated signaling pathway regulating organ development in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.*, 102(25), 9074–9. DOI: 10.1073/pnas.0503526102. [cit. p. 185]

46. Gifford, M. L. ; Dean, S. et Ingram, G. C. (2003). The *Arabidopsis* ACR4 gene plays a role in cell layer organisation during ovule integument and sepal margin development. *Development*, 130(18), 4249–58. DOI: 10.1242/dev.00634. [cit. p. 185]

47. Pillitteri, L. J. ; Bemis, S. M. ; Shpak, E. D. et Torii, K. U. (2007). Haploinsufficiency after successive loss of signaling reveals a role for ERECTA-family genes in *Arabidopsis* ovule development. *Development*, 134(17), 3099–109. DOI: 10.1242/dev.004788. [cit. p. 185]

48. Zhong, S. et Qu, L.-J. (2019). Peptide/receptor-like kinase-mediated signaling involved in male-female interactions. *Curr. Opin. Plant Biol.*, 51, 7–14. DOI: 10.1016/j.pbi.2019.03.004. [cit. p. 185]

49. Takeuchi, H. et Higashiyama, T. (2016). Tip-localized receptors control pollen tube growth and LURE sensing in *Arabidopsis*. *Nature*, 531(7593), 245–8. DOI: 10.1038/nature17413. [cit. p. 185]

50. Wang, T. ; Liang, L. ; Xue, Y. et coll. (2016). A receptor heteromer mediates the male perception of female attractants in plants. *Nature*, 531(7593), 241–4. DOI: 10.1038/nature16975. [cit. p. 185]

51. Ge, Z. ; Bergonci, T. ; Zhao, Y. et coll. (2017). *Arabidopsis* pollen tube integrity and sperm release are regulated by RALF-mediated signaling. *Science*, 358, 1596–1600. DOI: 10.1126/science.aao3642. [cit. p. 185]

52. Escobar-Restrepo, J.-M. ; Huck, N. ; Kessler, S. et coll. (2007). The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science*, 317(5838), 656–60. DOI: 10.1126/science.1143562. [cit. p. 186]

53. Galindo-Trigo, S. ; Blanco-Tourinan, N. ; DeFalco, T. A. et coll. (2018). CrRLK1L receptor-like kinases HERK1 and ANJEA are female determinants of pollen tube reception. *bioRxiv*, page 428854. DOI: 10.1101/428854. [cit. p. 186]

54. Shiu, S.-H. et Bleecker, A. B. (2001). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. U. S. A.*, 98(19), 10763–8. DOI: 10.1073/pnas.181141598. [cit. p. 186]

55. Shiu, S.-H. ; Karlowski, W. M. ; Pan, R. et coll. (2004). Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell*, 16(5), 1220–34. DOI: 10.1105/tpc.020834. [cit. p. 186]

56. Germain, H. ; Gray-Mitsumune, M. ; Lafleur, E. et Matton, D. P. (2008). ScORK17, a transmembrane receptor-like kinase predominantly expressed in ovules is involved in seed development. *Planta*, 228(5), 851–62. DOI: 10.1007/s00425-008-0787-0. [cit. p. 186]

57. Germain, H. ; Houde, J. ; Gray-Mitsumune, M. et coll. (2007). Characterization of ScORK28, a transmembrane functional protein receptor kinase predominantly expressed in ovaries from the wild potato species *Solanum chacoense*. *FEBS Lett.*, 581(26), 5137–42. DOI: 10.1016/j.febslet.2007.10.001. [cit. p. 186]

58. Lindner, H. ; Müller, L. M. ; Boisson-Dernier, A. et Grossniklaus, U. (2012). CrRLK1L receptor-like kinases: not just another brick in the wall. *Curr. Opin. Plant Biol.*, 15(6), 659–69. DOI: 10.1016/j.pbi.2012.07.003. [cit. p. 186]

59. Li, C. ; Yeh, F.-L. ; Cheung, A. Y. et coll. (2015). Glycosylphosphatidylinositol-anchored proteins as chaperones and co-receptors for FERONIA receptor kinase signaling in *Arabidopsis*. *eLife*, 4, e06587. DOI: 10.7554/eLife.06587.001. [cit. p. 186]

60. Hou, Y. ; Guo, X. ; Cyprys, P. et coll. (2016). Maternal ENODLs are required for pollen tube reception in *Arabidopsis*. *Curr. Biol.*, 26(17), 2343–50. DOI: 10.1016/j.cub.2016.06.053. [cit. p. 186]

61. Marshall, E. ; Costa, L. M. et Gutierrez-Marcos, J. (2011). Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J. Exp. Bot.*, 62 (5), 1677–86. DOI: 10.1093/jxb/err002. [cit. p. 187]

62. Bircheneder, S. et Dresselhaus, T. (2016). Why cellular communication during plant reproduction is particularly mediated by CRP signalling. *J. Exp. Bot.*, 67(16), 4849–61. DOI: 10.1093/jxb/erw271. [cit. p. 187]

63. Park, S. Y. ; Jauh, G. Y. ; Mollet, J. C. et coll. (2000). A lipid transfer-like protein is necessary for lily pollen tube adhesion to an *in vitro* stylar matrix. *Plant Cell*, 12(1), 151–64. DOI: 10.1105/tpc.12.1.151. [cit. p. 187]

64. Chae, K. ; Kieslich, C. A. ; Morikis, D. ; Kim, S.-C. et Lord, E. M. (2009). A gain-of-function mutation of *Arabidopsis* lipid transfer protein 5 disturbs pollen tube tip growth and fertilization. *Plant Cell*, 21(12), 3902–14. DOI: 10.1105/tpc.109.070854. [cit. p. 187]

65. Tang, W. ; Ezcurra, I. ; Muschietti, J. et McCormick, S. (2002). A cysteine-rich extracellular protein, LAT52, interacts with the extracellular domain of the pollen receptor kinase LePRK2. *Plant Cell*, 14(9), 2277–87. DOI: 10.1105/tpc.003103. [cit. p. 187]

66. Tang, W. ; Kelley, D. ; Ezcurra, I. ; Cotter, R. et McCormick, S. (2004). LeSTIG1, an extracellular binding partner for the pollen receptor kinases LePRK1 and LePRK2, promotes pollen tube growth *in vitro*. *Plant J.*, 39(3), 343–53. DOI: 10.1111/j.1365-313X.2004.02139.x. [cit. p. 187]

67. Jany, E. ; Nelles, H. et Goring, D. R. (2019). The molecular and cellular regulation of *Brassicaceae* self-incompatibility and self-pollen rejection. *Int. Rev. Cell Mol. Biol.*, 343, 1–35. DOI: 10.1016/bs.ircmb.2018.05.011. [cit. p. 187]

68. Wang, L. ; Lin, Z. ; Triviño, M. et coll. (2019). Self-incompatibility in *Papaver* pollen: programmed cell death in an acidic environment. *J. Exp. Bot.*, 70(7), 2113–23. DOI: 10.1093/jxb/ery406. [cit. p. 187]

69. Huang, Q. ; Dresselhaus, T. ; Gu, H. et Qu, L.-J. (2015). Active role of small peptides in *Arabidopsis* reproduction: expression evidence. *J. Integr. Plant Biol.*, 57(6), 518–21. DOI: 10.1111/jipb.12356. [cit. p. 187]

70. Sprunck, S. ; Rademacher, S. ; Vogler, F. et coll. (2012). Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science*, 338(6110), 1093–7. DOI: 10.1126/science.1223944. [cit. p. 187]

71. Kanaoka, M. M. ; Kawano, N. ; Matsubara, Y. et coll. (2011). Identification and characterization of TcCRP1, a pollen tube attractant from *Torenia concolor*. *Ann. Bot.*, 108(4), 739–47. DOI: 10.1093/aob/mcr111. [cit. p. 187]

72. Amien, S. ; Kliwer, I. ; Márton, M. L. et coll. (2010). Defensin-like ZmES4 mediates pollen tube burst in maize via opening of the potassium channel KZM1. *PLoS Biol.*, 8(6), e1000388. DOI: 10.1371/journal.pbio.1000388. [cit. p. 187]

73. Silverstein, K. A. T. ; Moskal, W. A. ; Wu, H. C. et coll. (2007). Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J.*, 51(2), 262–80. DOI: 10.1111/j.1365-313X.2007.03136.x. [cit. p. 187]

74. Andersson, M. ; Turesson, H. ; Olsson, N. et coll. (2018). Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. *Physiol. Plant.*, 164(4), 378–84. DOI: 10.1111/ppl.12731. [cit. p. 188]

75. Kusano, H. ; Ohnuma, M. ; Mutsuro-Aoki, H. et coll. (2018). Establishment of a modified CRISPR/Cas9 system with increased mutagenesis frequency using the translational enhancer dMac3 and multiple guide RNAs in potato. *Sci. Rep.*, 8(1), 13753. DOI: 10.1038/s41598-018-32049-2. [cit. p. 188]

76. Veillet, F. ; Perrot, L. ; Chauvin, L. et coll. (2019). Transgene-free genome editing in tomato and potato plants using -mediated delivery of a CRISPR/Cas9 cytidine base editor. *Int. J. Mol. Sci.*, 20 (2), E402. DOI: 10.3390/ijms20020402. [cit. p. 188]

77. Pease, J. B. (2017). SHEAR: simple handler for error and adapter removal. URL: https://github.com/jbpease/shear/. [cit. p. 189]

78. Buffalo, V. (2014). Scythe - a Bayesian adapter trimmer (version 0.994 BETA). URL: https://github.com/vsbuffalo/scythe. [cit. p. 189]

79. Dobin, A. ; Davis, C. A. ; Schlesinger, F. et coll. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. DOI: 10.1093/bioinformatics/bts635. [cit. p. 189]

80. Roberts, A. ; Pimentel, H. ; Trapnell, C. et Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-seq. *Bioinformatics*, 27(17), 2325–9. DOI: 10.1093/bioinformatics/btr355. [cit. p. 189]

81. Wu, T. D. et Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9), 1859–75. DOI: 10.1093/bioinformatics/bti310. [cit. p. 190]

82. Li, W. et Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–9. DOI: 10.1093/bioinformatics/btl158. [cit. p. 190]

83. Fu, L. ; Niu, B. ; Zhu, Z. ; Wu, S. et Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150–2. DOI: 10.1093/bioinformatics/bts565. [cit. p. 190]

84. O'Leary, N. A. ; Wright, M. W. ; Brister, J. R. et coll. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, 44 (D1), D733–D745. DOI: 10.1093/nar/gkv1189. [cit. p. 190]

85. Altschul, S. F. ; Gish, W. ; Miller, W. ; Myers, E. W. et Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215(3), 403–10. DOI: 10.1016/S0022-2836(05)80360-2. [cit. p. 190]

86. El-Gebali, S. ; Mistry, J. ; Bateman, A. et coll. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.*, 47(D1), D427–D432. DOI: 10.1093/nar/gky995. [cit. p. 190]

87. Jones, P. ; Binns, D. ; Chang, H.-Y. et coll. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–40. DOI: 10.1093/bioinformatics/btu031. [cit. p. 190]

88. Götz, S. ; García-Gómez, J. M. ; Terol, J. et coll. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.*, 36(10), 3420–3435. DOI: 10.1093/nar/gkn176. [cit. p. 190]

89. Jeske, L. ; Placzek, S. ; Schomburg, I. ; Chang, A. et Schomburg, D. (2019). BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.*, 47(D1), D542–9. DOI: 10.1093/nar/gky1048. [cit. p. 190]

90. Kanehisa, M. ; Furumichi, M. ; Tanabe, M. ; Sato, Y. et Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, 45(D1), D353–61. DOI: 10.1093/nar/gkw1092. [cit. p. 191]

91. Almagro Armenteros, J. J. ; Tsirigos, K. D. ; Sønderby, C. K. et coll. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.*, 37(4), 420–3. DOI: 10.1038/s41587-019-0036-z. [cit. p. 191]

92. Gupta, R. ; Jung, E. et Brunak, S. (2014). Prediction of *N*-glycosylation sites in human proteins. URL: http://www.cbs.dtu.dk/services/NetNGlyc. [cit. p. 191]

93. Krogh, A. ; Larsson, B. ; von Heijne, G. et Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, 305 (3), 567–580. DOI: 10.1006/jmbi.2000.4315. [cit. p. 191]

94. Pierleoni, A. ; Martelli, P. L. et Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC Bioinf.*, 9, 392. DOI: 10.1186/1471-2105-9-392. [cit. p. 191]

95. Joly, V. et Matton, D. P. (2015). KAPPA, a simple algorithm for discovery and clustering of proteins defined by a key amino acid pattern: a case study of the cysteine-rich proteins. *Bioinformatics*, 31 (11), 1716–1723. DOI: 10.1093/bioinformatics/btv047. [cit. p. 191]

96. Zhu, A. ; Ibrahim, J. G. et Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*, 35(12), 2084–92. DOI: 10.1093/bioinformatics/bty895. [cit. p. 191]

97. Matton, D. P. ; Maes, O. ; Laublin, G. et coll. (1997). Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. *Plant Cell*, 9(10), 1757–66. DOI: 10.1105/tpc.9.10.1757. [cit. p. 192]

98. Gray-Mitsumune, M. ; O'Brien, M. ; Bertrand, C. et coll. (2006). Loss of ovule identity induced by overexpression of the fertilization-related kinase 2 (ScFRK2), a MAPKKK from *Solanum chacoense*. *J. Exp. Bot.*, 57(15), 4171–87. DOI: 10.1093/jxb/erl194. [cit. p. 192]

CHAPITRE 7

Perspectives sur la chimioattraction du tube pollinique chez *Solanum*

7.1 Protéines candidates de chimioattraction

Comme discuté précédemment à la section 2.3.1, les chimioattractants connus à ce jour chez *Torenia* et *Arabidopsis* sont des petites protéines riches en cystéines (CRP) de la famille des défensines, dont les principales caractéristiques sont résumées au tableau 7.1. On compte parmi eux les attractants TfLURE1 et 2 identifiés en 2009 chez *Torenia fournieri* par l'analyse du transcriptome de cellules synergides isolées à partir du sac embryonnaire[1], ainsi que l'orthologue TcLURE1 de la proche parente *T. concolor*.[2] Il s'agit de protéines de type défensine (*defensin-like*, DEFL) comprenant les éléments CS$\alpha\beta$ et γ-core caractéristiques de cette famille. La figure figure 7.1a fait ressortir les motifs `CXXXC` et `CXC` correspondants. Les Tf- et TcLURE sont fortement divergentes et permettent un guidage spécifique à l'espèce du tube pollinique (TP).

En 2012, l'analyse phylogénétique des DEFL de la plante-modèle *Arabidopsis thaliana* et de l'espèce voisine *A. lyrata* a permis de mettre au jour la famille CRP810, contenant 6 gènes renommés depuis *AtLURE1.1* à *1.6*[3]. *AtLURE1.6* est en réalité un pseudogène et seuls les cinq premiers membres de la famille sont exprimés dans les cellules synergides et codent pour des protéines. Des tests de guidage ont permis de mettre en évidence que les peptides AtLURE1.1 à 1.4 agissent de façon redondante pour attirer le TP avec une nette préférence à l'espèce. En revanche, le peptide AtLURE1.5, dépourvu de la sixième cystéine caractéristique, est inactif dans le guidage (figure 7.1b).

Tout récemment, Zhong et coll. ont identifié deux nouveaux membres de cette famille, AtLURE1.7 et 1.8, codant pour des peptides chimioattractifs[4]. Leur étude a surtout permis de

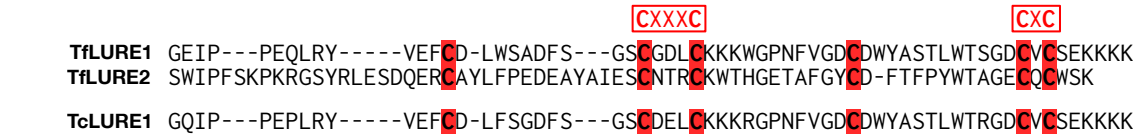
mettre en évidence que le septuple mutant *lure*, obtenu en inactivant les sept gènes *AtLURE1.1* à *1.8* (sauf *1.6*) grâce à la technologie CRISPR/Cas9, ne présente pas de défauts de fertilité. Le phénotype de ce mutant consiste en revanche en une perte de la préséance des TP conspécifiques. Ces derniers ont normalement un temps d'émergence sur le septum nettement plus court que leurs compétiteurs hétérospécifiques, et cet avantage est aboli dans le mutant *lure*. Ce constat a conduit les auteurs à explorer la fonction d'un clade de DEFL voisin des AtLURE, régulé à la baisse dans le mutant *myb98* et comprenant quatre gènes renommés *AtXIUQIU1* à *4*, exprimés eux aussi dans les synergides (figure 7.1c). Des tests fonctionnels ont permis de mettre en évidence que les peptides XIUQIU, qui sont davantage conservés chez les Brassicacées, sont suffisants pour attirer les TP de manière non spécifique à l'espèce.

Tab. 7.1. Propriétés des chimioattractants riches en cystéines identifiés chez *Torenia* et *Arabidopsis*. Les valeurs de longueur (*L*), poids moléculaire (*M*), point isoélectrique (*pI*) et nombre de cystéines (*Cys.*) sont calculées sur la partie mature de la protéine.

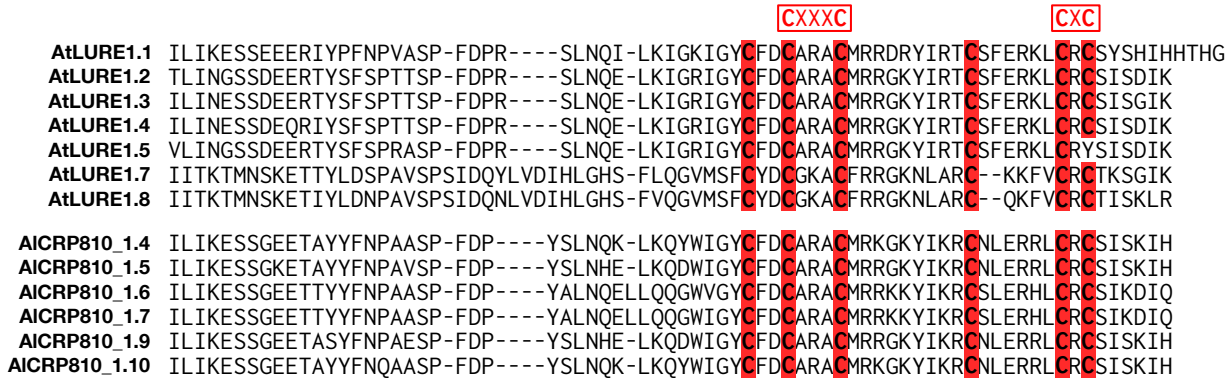
Espèce	Protéine	*L* (aa)	*M* (kDa)	pI	Cys.
Torenia fournieri	TfLURE1	62	7,02	4,736	6
	TfLURE2	70	8,28	5,143	6
Torenia concolor	TcLURE1	62	7,04	5,970	6
Arabidopsis thaliana	AtLURE1.1	75	8,87	8,053	6
	AtLURE1.2	71	8,18	8,053	6
	AtLURE1.3	71	8,21	8,053	6
	AtLURE1.4	71	8,28	8,053	6
	AtLURE1.5*	71	8,26	8,639	5
	AtLURE1.7	74	8,31	8,419	6
	AtLURE1.8	74	8,37	8,273	6
Arabidopsis thaliana	AtXIUQIU1	76	8,63	5,499	6
	AtXIUQIU2	68	7,69	7,980	6
	AtXIUQIU3	55	6,24	7,805	6
	AtXIUQIU4	72	8,08	5,537	6
Arabidopsis lyrata	AlCRP810_1.4	71	8,38	8,463	6
	AlCRP810_1.5	71	8,39	8,258	6
	AlCRP810_1.6	72	8,46	7,512	6
	AlCRP810_1.7	72	8,47	7,512	6
	AlCRP810_1.9	71	8,35	7,512	6
	AlCRP810_1.10	71	8,41	8,463	6

*AtLURE1.5 a perdu sa dernière cystéine et n'a pas la capacité d'attirer les TP.

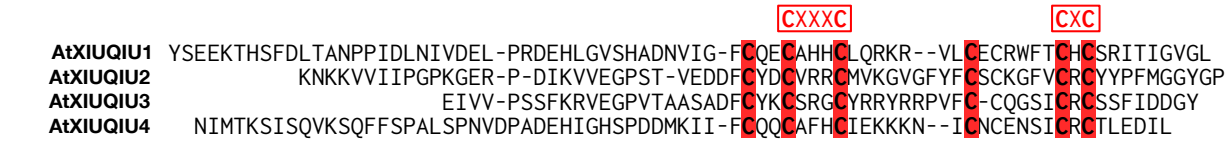
Comme expliqué au chapitre précédent, l'analyse par KAPPA des annotations augmentées que nous avons générées pour le génome de *S. chacoense* ainsi que la liste complémentaire de 20 503 gènes ovulaires assemblés *de novo* a permis de dégager un ensemble de 478 CRPs dont 58 contiennent le motif DEFL `X{n}CX{n}CX{3}CX{1-}CX{1-}CXCX{n}` (Dataset S6.3g). Les relations phylogénétiques de ces DEFL sont présentées à la figure 6.7 et leurs coordonnées génomiques sont données dans le tableau S7.1.



(a) Attractants LURE spécifiques à l'espèce de *Torenia fournieri*[1] et *T. concolor*[2].



(b) Attractants LURE spécifiques à l'espèce d'*Arabidopsis thaliana* et *A. lyrata*[3,4].



(c) Attractants XIUQIU non-spécifiques d'*Arabidopsis thaliana*[4].

Fig. 7.1. Alignement des chimioattractants riches en cystéines identifiés chez *Torenia* et *Arabidopsis*.

La figure 7.2 présente les valeurs d'expression différentielle pour les DEFL exprimées. La comparaison entre les ovules matures sauvages (condition « Anth ») et les feuilles (condition « Leaf ») permet de dégager un ensemble de 29 DEFLs enrichies dans l'ovule ($FC_{Anth/Leaf} \geq 5, p_{adj} \leq 0{,}05$; points jaunes, rouges et violets sur le graphique). Les ovules immatures issus de fleurs cueillies deux jours avant l'anthèse (condition « 2 DBA ») sont incapables d'attirer le TP (voir figure 4.2, page 122). Neuf (9) DEFL ont effectivement une expression significativement moins forte dans l'ovule immature (condition 2 DBA) que dans l'ovule mature ($FC_{2\,DBA/Anth} \leq -2, p_{adj} \leq 0{,}05$; points rouges et violets sur le graphique). Parmi elles, 4 sont également régulées à la baisse dans les ovules issus du mutant *Scfrk1* (condition « *frk1* ») par rapport aux ovules sauvages ($FC_{frk1/Anth} \leq -2, p_{adj} \leq 0{,}05$; points violets sur le graphique), lesquels sont également incompétents pour le guidage du TP car ils ne possèdent pas de sac embryonnaire (voir figure 6.3, page 181).

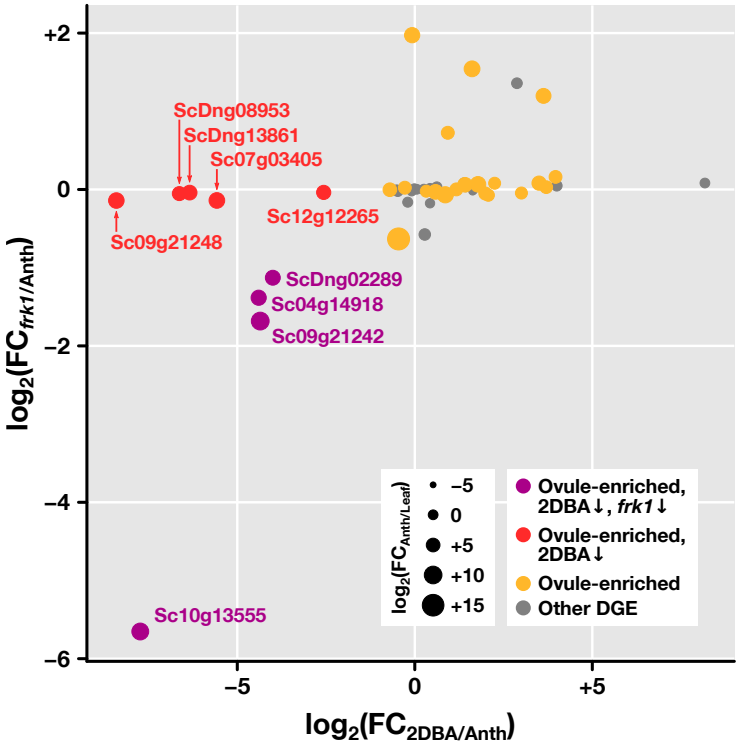


Fig. 7.2. Valeurs d'expression différentielle des DEFL identifiées chez *S. chacoense*.

Les propriétés des 29 DEFL enrichies dans l'ovule sont rappelées, avec leurs valeurs précises d'expression différentielle, dans le tableau 7.2 et leur alignement est présenté à la figure 7.3. Il est intéressant de remarquer qu'exception faite de Sc09g21162 et Sc09g21164, qui n'ont de toute façon par le profil d'expression attendu pour les conditions 2 DBA et *frk1*, toutes les DEFL identifiées ici comportent plus de 6 cystéines contrairement aux attractants connus (figure 7.1).

Ceci suggère l'existence de sous-familles de DEFL particulières aux Solanacées qui possèdent des ponts disulfure supplémentaires, ou subissent une étape de maturation impliquant des clivages protéolytiques. On peut aussi noter que la quasi-totalité de ces DEFL ont été repérées dans les nouveaux transcrits découverts dans le chapitre 6, à savoir ceux codés par de nouveaux gènes identifiés dans le génome grâce aux lectures de séquençage ovulaires, et ceux assemblés *de novo*.

Tab. 7.2. Propriétés et valeurs d'expression différentielle des DEFL enrichies dans l'ovule de *S. chacoense*. Les valeurs de longueur (L), poids moléculaire (M), point isoélectrique (pI) et nombre de cystéines (*Cys.*) sont calculées sur la partie mature de la protéine. Les valeurs d'expression différentielle en gris sont associées à des variations non-significatives ($p_{adj} > 0{,}05$). Les trois sections désignent, respectivement, les gènes enrichis dans l'ovule qui sont également à la baisse dans 2 DBA et *frk1*, ceux qui sont également à la baisse dans 2 DBA seulement, et ceux qui ne sont à la baisse ni dans 2 DBA, ni dans *frk1*.

Gène	L (aa)	M (kDa)	pI	Cys.	$FC_{Anth/Leaf}$	$FC_{2DBA/Anth}$	$FC_{frk1/Anth}$
Sc04g14918	58	6,54	6,707	8	+113,632	−21,075	−2,610
Sc09g21242	81	9,13	6,883	12	+1156,941	−20,445	−3,210
Sc10g13555	51	5,69	4,787	8	+ 495,750	−212,670	−50,296
ScDng02289*	112	11,83	6,751	8	+90,299	−15,994	−2,186
Sc07g03405	47	5,09	7,088	8	+173,491	−47,671	−1,102
Sc09g21248	81	9,16	5,397	13	+155,418	−339,343	−1,103
Sc12g12265	67	7,12	6,942	10	+49,248	−5,923	−1,026
ScDng08953	50	5,65	4,991	8	+42,000	−99,181	−1,037
ScDng13861	43	4,25	5,512	11	+77,718	−81,167	−1,028
Sc01g00040	47	5,38	8,565	8	+15,697	+2,252	+1,001
Sc04g00195	71	8,18	7,073	8	+161,160	+3,435	+1,047
Sc05g07242	78	8,49	6,532	14	+29,825	−1,627	−1,002
Sc05g07245	73	8,41	6,287	12	+49,821	+11,317	+1,058
Sc06g11955	67	7,68	6,737	8	+7,794	+4,203	−1,052
Sc07g03400	47	5,25	7,541	8	+15,993	+1,905	+1,653
Sc07g03465	51	5,52	6,722	10	+77,706	+1,498	−1,021
Sc09g10288	54	5,98	6,160	8	+6,726	+4,753	+1,057
Sc09g21162	48	5,45	5,550	6	+202,367	+3,058	+2,914
Sc09g21164	54	6,13	7,249	6	+15,500	−1,214	+1,016
Sc10g03055	50	5,41	5,830	8	+69,490	+2,666	+1,043
Sc10g09020	58	6,27	4,304	8	+16,302	+3,962	−1,036
Sc10g18935	48	5,04	5,321	8	+14,772	+15,607	+1,118
Sc12g18185	65	7,47	4,240	8	+13,213	+13,018	+1,021
ScDng04736	78	9,26	4,329	10	+44459,374	−1,373	−1,550
ScDng08860	48	5,41	8,609	7	+120,228	−1,053	+3,927
ScDng09296	47	5,34	8,697	8	+7,446	+1,252	−1,015
ScUng12282	64	6,84	6,097	10	+277,555	+1,814	−1,048
ScUng35555	56	6,05	5,207	10	+7,159	+8,019	−1,032
ScUng37761	50	5,47	5,830	8	+90,381	+12,346	+2,293

*Séquence *de novo* ignorée dans les analyses subséquentes.

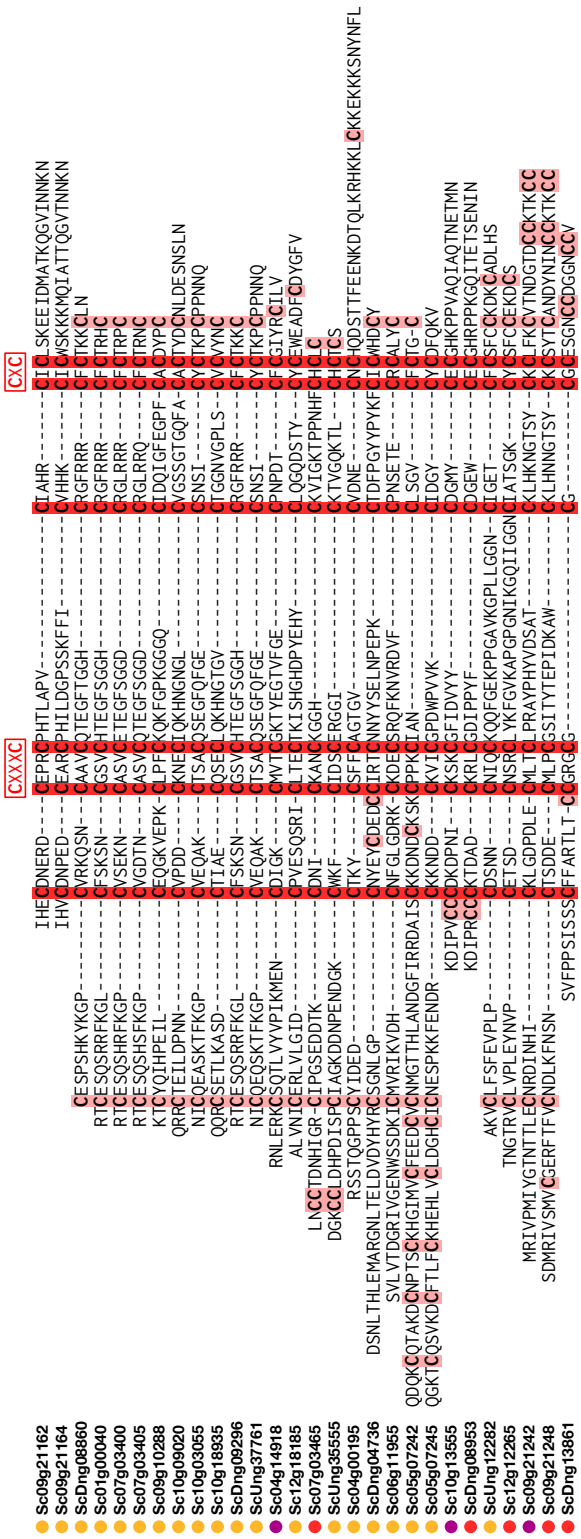


Fig. 7.3. Alignement des DEFL enrichies dans l'ovule de *S. chacoense*. L'alignement a été effectué sur la partie mature des protéines. Les cystéines caractéristiques du motif DEFL apparaissent en rouge vif, les autres en rouge pâle. Les points de couleur renvoient aux catégories définies dans la figure 7.2.

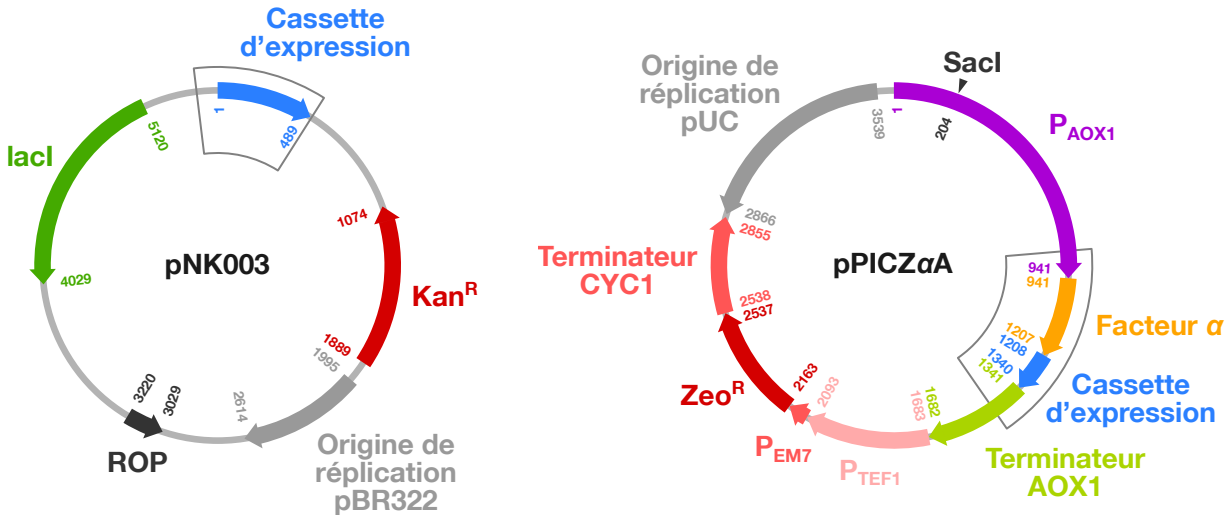
Nous pouvons ainsi délimiter trois groupes de chimioattractants candidats :

1. les 4 DEFL combinant les trois critères d'expression attendus (points violets) : Sc04g14918, Sc09g21242, Sc10g13555 et ScDng02289 ;
2. les 5 DEFL combinant deux des trois critères, à savoir l'enrichissement dans l'ovule et l'expression moins forte dans l'ovule immature (points rouges) : Sc07g03405, Sc09g21248, Sc12g12265, ScDng08953 et ScDng13861 ;
3. les 15 DEFL restantes, qui sont enrichies dans l'ovule mais qui ne sont à la baisse ni dans l'ovule 2 DBA ni dans l'ovule *frk1* (points jaunes).

Parmi les 4 meilleurs candidats, le seul ayant été assemblé *de novo*, ScDng02289, sera écart. des analyses subséquentes. La taille anormalement grande de son domaine C_4-C_5 et le fait qu'il s'apparie à cheval sur deux DEFL successives du génome de *S. tuberosum* indiquent qu'il s'agit probablement d'un contig chimérique. Une analyse directe par PCR sur les ADNc ovulaires de *S. chacoense* permettrait de le confirmer.

Afin de vérifier expérimentalement la fonction des candidats identifiés plus haut, il devient nécessaire de tester *in vitro* leur capacité à attirer le TP. Pour cela, il faut d'abord cloner les séquences codantes des protéines candidates dans des vecteurs d'expression qui seront transformés dans des organismes-hôtes chargés de produire de grandes quantités de protéines. Dans les travaux publiés précédemment sur *Torenia* [1,2] et *Arabidopsis* [3,4,5], les protéines ont été exprimées dans des systèmes bactériens (*E. coli*) ou eucaryotes (cellules d'insecte).

Ainsi, les séquences codantes matures des candidats de *S. chacoense* pourront à leur tour être exprimées dans des systèmes analogues. Le lecteur pourra se reporter aux travaux de deux collègues ayant travaillé sur le guidage chez *S. chacoense*, Yang Liu et Claire Viallet, qui ont détaillé dans leurs thèse et mémoire respectifs des protocoles optimisés pour l'expression et la purification de CRP [6,7]. À la banque de plasmides testés par ces autrices, nous pouvons ajouter ici deux nouveaux vecteurs qui nous ont été conseillés lors de stages dans les laboratoires des professeurs Tetsuya Higashiyama (Université de Nagoya, Japon) et Johan Edqvist (Université de Linköping, Suède). Il s'agit respectivement du vecteur d'expression bactérien pNK003, qui permet l'expression de protéines munies d'une étiquette polyhistidinique aminoterminale (figures 7.4a et S7.1), et du vecteur pPICZαA, qui permet l'expression et la sécrétion par la levure *Pichia pastoris* de protéines portant une étiquette polyhistidinique et un épitope c-myc dans la partie carboxyterminale (figures 7.4b et S7.2). Nous avons commencé les clonages et les tests d'expression pour les candidats Sc09g21242 et Sc09g21248.



(a) Expression bactérienne **(b)** Expression en levure

Fig. 7.4. Cartes des plasmides utilisés pour l'expression des chimioattractants candidats.

7.2 Caractérisation fonctionnelle par microfluidique

7.2.1 Dispositifs microfluidiques

Dans les études publiées jusqu'ici, la réceptivité des tubes polliniques aux protéines ovulaires purifiées a été testée avec la méthode dite du *bead assay*[1,2,3,4]. On place sur un milieu de culture gélosé un style pollinisé suffisamment longtemps à l'avance pour que les tubes polliniques aient parcouru environ les deux tiers de sa longueur, et découpé à sa base, juste au dessus de l'ovaire. On laisse les tubes polliniques émerger de la base du style sur le milieu. Lorsqu'ils sont suffisamment éloignés les uns des autres, on sélectionne un TP en croissance et on lui présente, avec un angle et une distance caractéristiques que l'expérimentateur peut moduler selon ses besoins, une bille de gélatine contenant la protéine à tester à une concentration choisie. Si le TP est réceptif, il changera rapidement de trajectoire pour rejoindre la bille. Ce protocole permet d'obtenir des résultats robustes, et son applicabilité à *S. chacoense* a été démontrée[6,7].

Cependant, il s'agit d'une expérience extrêmement délicate et fastidieuse, dont la mise en œuvre devient difficilement envisageable lorsqu'il s'agit de tester de nombreux candidats. En outre, plusieurs facteurs viennent, dans notre cas, démultiplier le nombre de tests à effectuer. Tout d'abord, on ignore quelle est la distance d'action de nos attractants : il sera donc nécessaire de les tester à courte distance (quelques dizaines de micromètres) et à longue distance (quelques centaines voire milliers de micromètres). On ne sait pas non plus à quelle concentration l'on

devra travailler : les attractants de *Torenia* donnent des résultats significatifs à l'échelle nanomolaire[1], alors qu'il faut porter la concentration à l'échelle micromolaire chez *Arabidopsis*[3]. Enfin, puisque nous nous intéressons également à l'isolement reproductif, nous devrons tester de nombreuses combinaisons d'espèces.

C'est justement dans le souci de faciliter et de systématiser ces tests de guidage que le laboratoire du professeur Higashiyama a développé des dispositifs microfluidiques. Il s'agit de micropuces moulées dans du polydiméthylsiloxane (PDMS) à partir d'une matrice personnalisée gravée dans une plaque de silice, dans lesquelles ont peut contraindre la croissance des TP selon un circuit choisi par l'expérimentateur. Ces dispositifs, déjà largement utilisés pour comprendre la cytomécanique du TP[8,9,10,11,12,13], ont aussi été appliqués au problème spécifique du guidage directionnel[14,15,16].

C'est à l'occasion d'un séjour dans ce laboratoire que nous avons développé, avec l'aide et d'après les travaux de Naoki Yanagisawa, deux dispositifs microfluidiques permettant de tester le guidage des tubes polliniques de *Solanum*. Tout d'abord, le dispositif présenté à la figure 7.5 a été dessiné pour effectuer des tests à longue distance, sur le modèle de ceux existant pour *Torenia*[14]. Le dispositif comprend une cavité pour le style pollinisé (point S) et deux cavités dans lesquelles déposer des ovules ou bien des protéines purifiées (point O et son symétrique par rapport à la droite $x = 0$). Suite au moulage, le dispositif de PDMS est percé en ces trois points au moyen de poinçons chirurgicaux de 1 mm (orifice stylaire) ou 1,25 mm (orifices ovulaires). Après nettoyage, la micropuce est déposée, relief vers le bas, sur une lamelle de microscope ou sur la fenêtre d'une boîte de Pétri à fond de verre (*glass-bottom Petri dish*). Du milieu de culture peut être injecté dans les trois cavités et infiltré dans les microcanaux grâce à un passage dans une cloche à vide.

Après dépôt du style pollinisé (prélevé 24 HAP dans le cas de *S. chacoense*) au point S, le dispositif est maintenu à température constante (∼22 °C) dans un milieu fortement humide jusqu'à la sortie des TP, lesquels se rendent jusqu'au point T. À cet endroit, ils peuvent être mis au contact des chimioattractants se répartissant en gradient le long de l'axe RL. La longueur de ce dernier peut être ajustée pour mettre en place des gradients plus ou moins abrupts (figure S7.3). Il est alors possible de surveiller, pendant plusieurs heures, le comportement des TP.

La figure figure 7.6 présente des résultats préliminaires obtenus avec un tel dispositif chez *S. chacoense*. On peut constater l'existence d'un guidage à longue distance présentant une préférence à l'espèce. Après purification des chimioattractants candidats, ce dispositif pourra être

utilisé directement avec une solution liquide de protéines, comme testé précédemment chez *Torenia*[14,15], ou bien encore avec un fragment de gélatine contenant la protéine d'intérêt, permettant une diffusion plus lente et contrôlée.

Dans un deuxième temps, nous nous sommes fondés sur les travaux effectués par Naoki Yanagisawa sur le guidage à courte distance chez *Arabidopsis*[16] pour développer une micropuce alternative présentée aux figures 7.7–9, adaptée pour visualiser le guidage à courte distance. Ici, le dispositif est constitué d'un « sandwich » de deux couches de PDMS dont les faces moulées sont placées l'une contre l'autre. La couche inférieure est consistuée des trois cavités représentés en gris sur la figure 7.7. Elle est posée sur la face de verre, le relief vers le haut cette fois. La couche supérieure ne comprend que de minces canaux horizontaux transverses d'une largeur et d'une profondeur de 1 µm, représentés par les lignes bleues sur la figure 7.7, qui viennent connecter les trois cavités de la couche inférieure. C'est également dans la couche supérieure que sont percés au poinçon les quatre orifices entourant les points S, P, B (diamètre : 1 mm) et O (diamètre : 2 mm) du schéma.

Après injection du milieu de culture dans le dispositif, un style pollinisé prélevé 24 HAP peut être déposé au point S. Les TP commencent à envahir la cavité autour du point S environ 6 h après. Les fentes situées entre $y = -750$ µm et $y = -1000$ µm (figure 7.8) sont conçues pour limiter le nombre de TP atteignant la partie centrale du dispositif. Une fois que les TP ont dépassé la position $y = -1000$ µm, une goutelette de protéine candidate à une concentration donnée peut être déposée au point P. Les protéines vont alors diffuser des points G_i aux points D_i, en formant un gradient de concentration horizontal entre les points L_i et les points R_i, perpendiculairement à la direction de croissance des tubes polliniques (figure 7.9). Chaque point L_i agit alors comme un petit micropyle artificiel sécrétant un gradient d'attractants, et il devient possible de tester et quantifier la réceptivité des TP. Une modification de la forme des cavités latérales permet même de tester simultanément des gradients d'intensités différentes (figure S7.4). La cavité entourant le point O est utile pour injecter le milieu de culture mais aussi, éventuellement, pour introduire des ovules, des protéines, ou tout autre matériel

Ce type de dispositif, déjà validé pour tester la réceptivité des TP d'*Arabidopsis* aux signaux LURE[16], est donc un outil prometteur pour le criblage fonctionnel à haut débit de nos chimioattractants candidats. Dans le but de mieux comprendre et quantifier le comportement des TP, nous présentons à la section suivante une modélisation théorique de la diffusion des protéines dans le dispositif à courte distance.

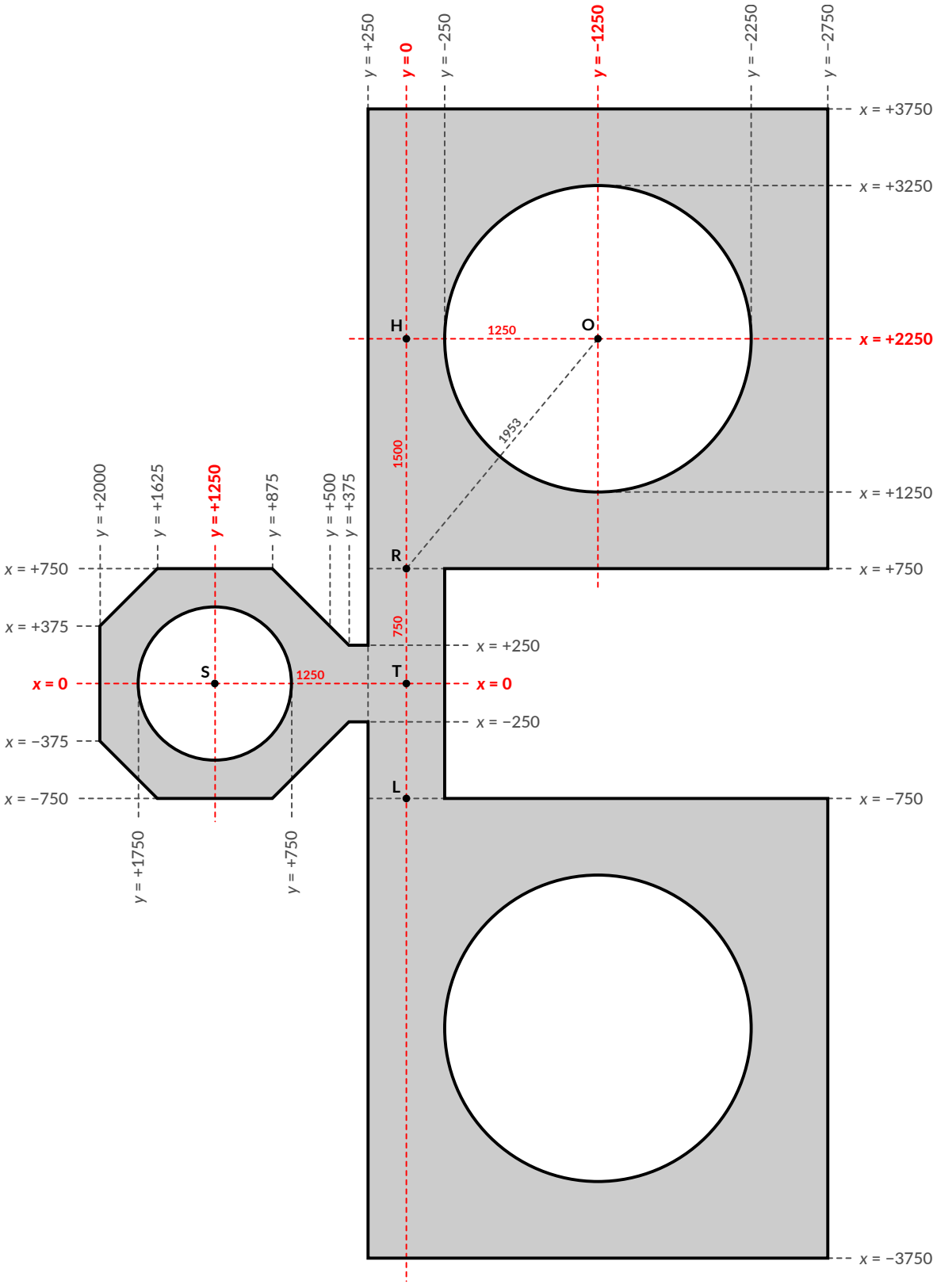
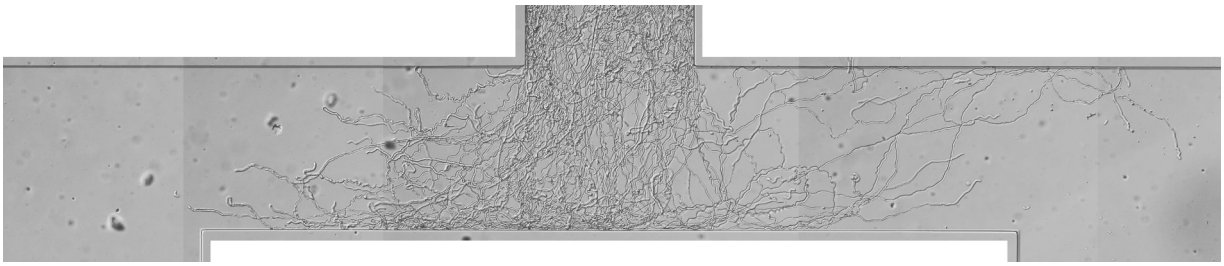
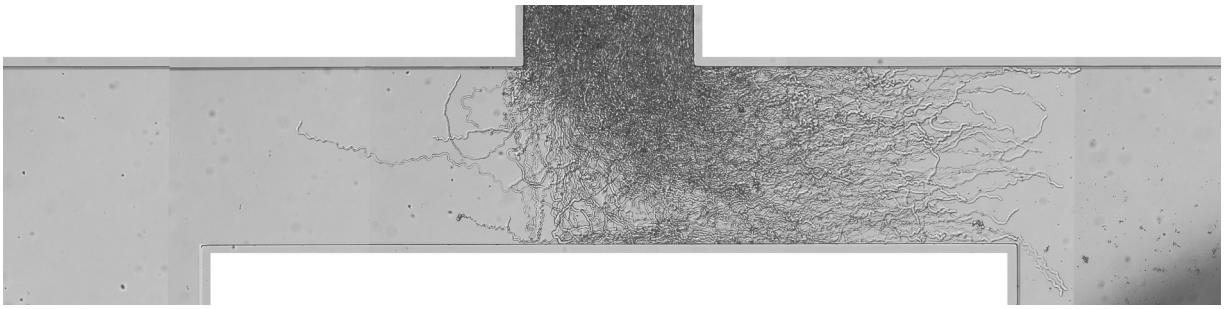


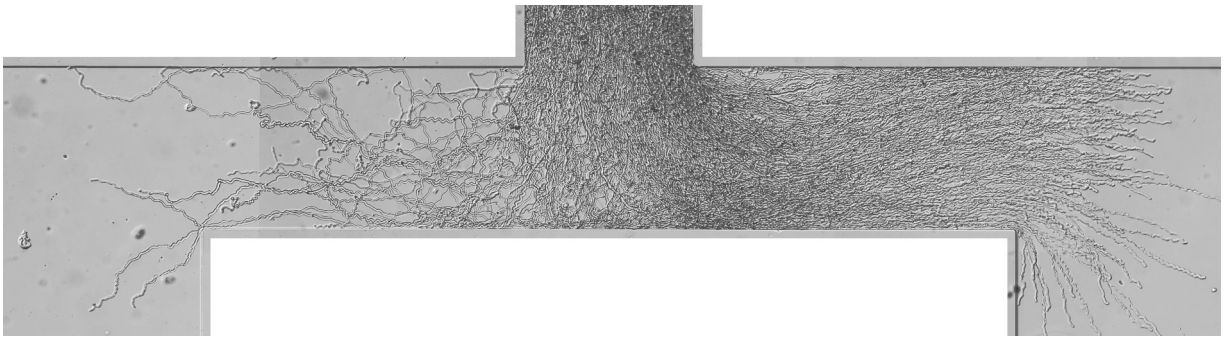
Fig. 7.5. Schéma du dispositif microfluidique pour tests de guidage à longue distance semi-*in vivo* chez *Solanum*. Les coordonnées sont données en micromètres. Profondeur : 100 µm.



(a) Témoin négatif : pas d'ovules à gauche ni à droite



(b) Test unilatéral : pas d'ovules à gauche, ovules de *S. chacoense* à droite



(c) Test bilatéral : ovules de *S. microdontum* à gauche, ovules de *S. chacoense* à droite

Fig. 7.6. Tests de guidage préliminaires les dispositifs microfluidiques à longue distance. Dans les trois cas, l'orifice supérieur (non montré) est muni d'un pistil de *S. chacoense*, génotype V22, pollinisé conspécifiquement par du pollen de génotype G4. Les photographies ont été prises *48*HAP avec un microscope optique en champ clair.

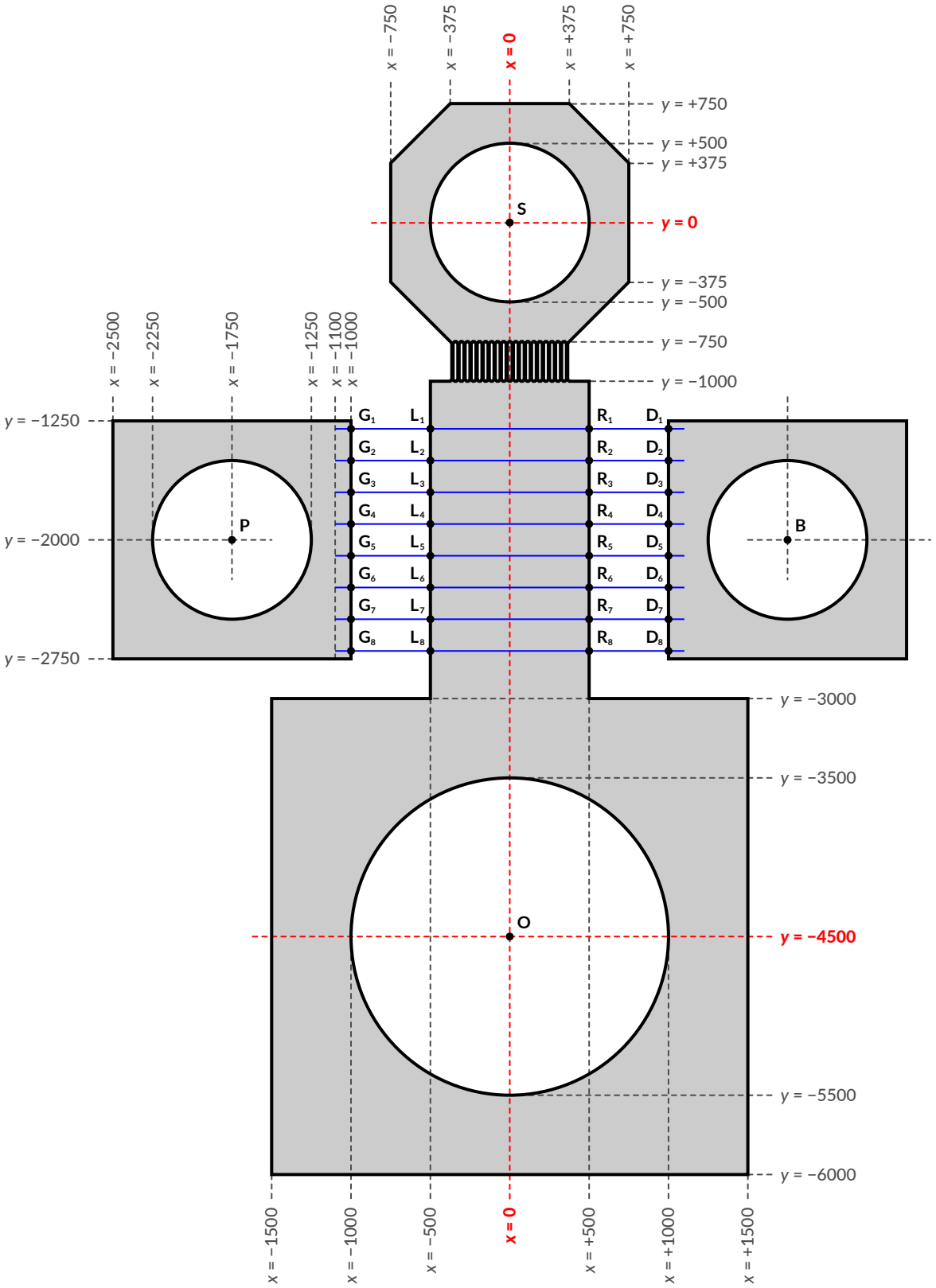


Fig. 7.7. Schéma du dispositif microfluidique pour tests de guidage à courte distance semi-*in vivo* chez *Solanum*. Les coordonnées sont données en micromètres. Profondeur : 30 µm pour la couche inférieure, 1 µm pour les canaux de la couche supérieure (traits bleus).

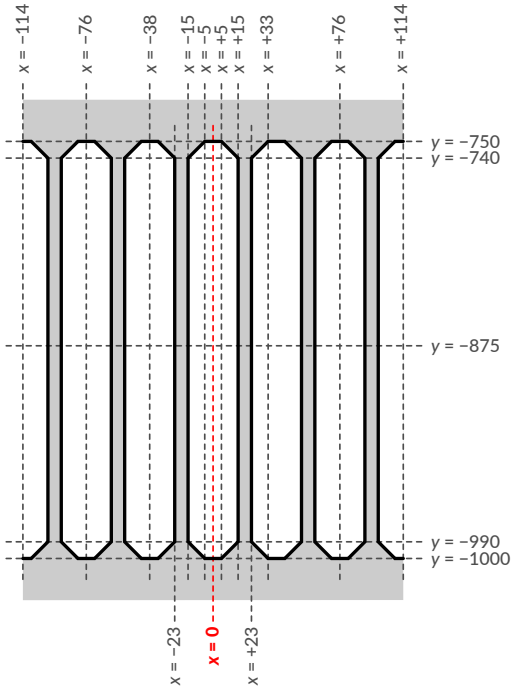


Fig. 7.8. Détail des fentes du dispositif à courte distance. Les fentes sont localisées à la sorte de la cavité stylaire, entre $y = -750$ µm and $y = -1000$ µm.

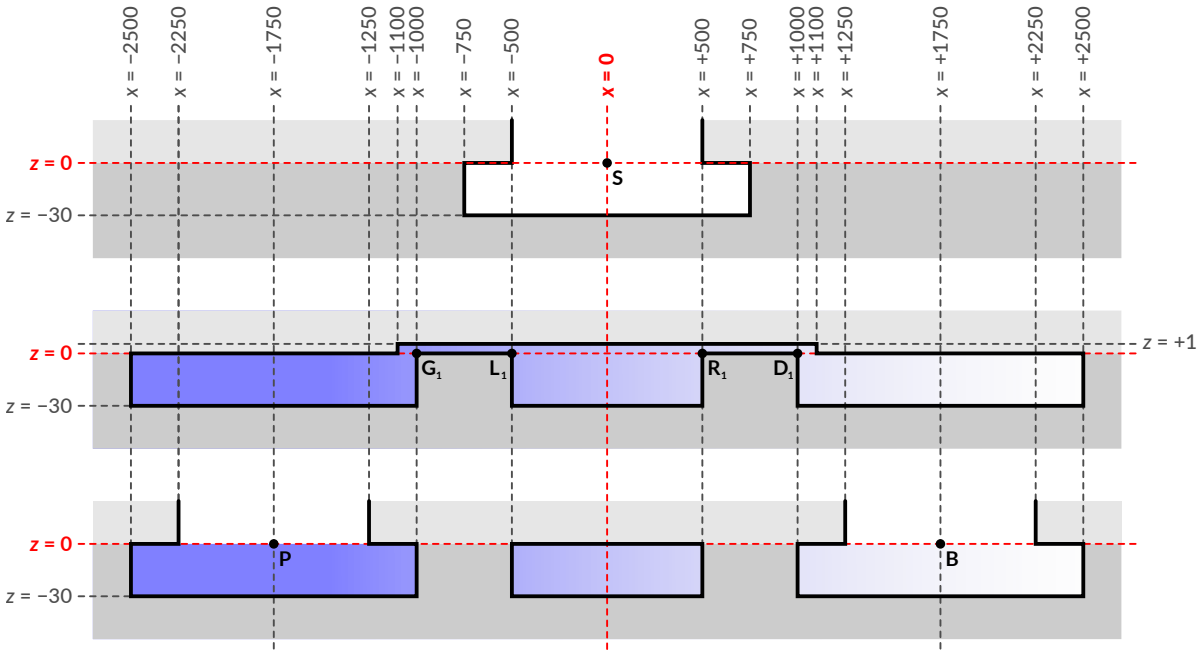


Fig. 7.9. Vue du dispositif à courte distance en coupe transversale dans le plan (x, z). Les sections sont pratiquées à $y = 0$ µm (haut), $y = -1350$ µm (milieu) et $y = -2000$ µm (bas).

7.2.2 Modélisation du flux de protéines

Introduisons tout d'abord quelques rappels théoriques sur la diffusion. Les protéines ont toutes une densité approximativement égale à $\rho = 1{,}37\,\mathrm{g\,cm^{-3}} = 1370\,\mathrm{kg\,m^{-3}}$. Le volume occupé par une protéine peut ainsi être déterminé comme suit :

$$V = \frac{M}{\rho N_A} \tag{7.1}$$

où :
- V est le volume de la protéine (en $\mathrm{m^3}$) ;
- M est la masse molaire de la protéine (en $\mathrm{kg\,mol^{-1}}$) ;
- ρ est la densité de la protéine (en $\mathrm{kg\,m^{-3}}$) ;
- N_A est la constante d'Avogadro (en $\mathrm{mol^{-1}}$).

Considérons la plus petite sphère imaginable pouvant contenir la totalité de la masse de la protéine. Son volume V_{min} et son rayon R_{min} sont reliés par :

$$V_{\mathrm{min}} = \frac{4}{3}\pi R_{\mathrm{min}}^3 \quad \Leftrightarrow \quad R_{\mathrm{min}} = \left(\frac{3V}{4\pi}\right)^{1/3} \tag{7.2}$$

En combinant les équations 7.1 et 7.2, on obtient :

$$R_{\mathrm{min}} = \left(\frac{3M}{4\pi\rho N_A}\right)^{1/3} \tag{7.3}$$

En théorie de la diffusion, le rayon de Stokes-Einstein R_{S} d'une particule est défini comme le rayon d'une sphère hypothétique qui diffuserait à la même vitesse que cette particule. Dans le cas d'une petite protéine globulaire, on peut postuler que :

$$R_{\mathrm{S}} = R_{\mathrm{min}} \tag{7.4}$$

La loi de Stokes décrit la force de frottement F_d excercée sur de très petits objets sphériques (tels que des protéines) se trouvant dans un fluide (tel que de l'eau) :

$$F_d = 6\pi\mu r v \tag{7.5}$$

où :
- F_d est la force de frottement (en N);
- μ est la viscosité dynamique du fluide (en Pa s);
- r est le rayon de l'objet sphérique considéré (en m);
- v est la vitesse de flux relative à l'objet (en $\mathrm{m\,s^{-1}}$).

La viscosité dynamique μ de l'eau (en Pa s) peut être approximée, à n'importe quelle température T (en K), par :

$$\mu = 2.414 \times 10^{\frac{247.8}{T-140}-5} \tag{7.6}$$

Par ailleurs, la relation d'Einstein décrit le coefficient de diffusion D d'une particule en mouvement brownien comme :

$$D = \eta k_B T \tag{7.7}$$

où :

D	est le coefficient de diffusion	(en $m^2\,s^{-1}$);	
η	est la mobilité de la particule	(en $m^2\,s^{-1}\,J^{-1}$),	définie comme $\eta = \frac{v}{F_d}$;
k_B	est la constante de Boltzmann	(en $J\,K^{-1}$);	
T	est la température	(en K).	

En incorporant l'équation 7.7 dans l'équation 7.5, on obtient l'équation de Stokes-Einstein :

$$D = \frac{k_B T}{6\pi r \mu} \tag{7.8}$$

Pour une protéine avec un rayon de Stokes R_S, on obtient, d'après l'équation 7.4 :

$$D = \frac{k_B T}{6\pi R_S \mu} = \frac{k_B T}{6\pi\mu}\left(\frac{4\pi\rho N_A}{3M}\right)^{1/3} \tag{7.9}$$

Le flux de diffusion **J** décrit la quantité d'une substance diffusant au travers d'une unité d'aire par unité de temps. La première loi de Fick indique que la magnitude de ce flux est inversement proportionnelle au gradient de concentration $\nabla\varphi$:

$$\mathbf{J} = -D \underbrace{\left(\frac{\partial\varphi}{\partial x} + \frac{\partial\varphi}{\partial y} + \frac{\partial\varphi}{\partial z}\right)}_{\nabla\varphi} \tag{7.10}$$

où :

J	est le flux de diffusion	(en $mol\,m^{-2}\,s^{-1}$);
D	est le coefficient de diffusion	(en $m^2\,s^{-1}$);
φ	est la concentration	(en $mol\,m^{-3}$);
x, y et z	sont les coordonnées cartésiennes dans l'espace	(en m).

L'équation 7.10 est valide aux conditions suivantes : (i) le coefficient de diffusion D est constant, (ii) il n'y a pas d'autres mécanismes que la diffusion affectant la concentration (p. ex., des réactions chimiques) et (iii) il n'y a pas d'advection, autrement dit, le fluide lui-même n'est pas en mouvement.

Le principe de conservation de la masse permet de dériver la seconde loi de Fick. Cette équation différentielle partielle prédit de quelle manière la concentration change au cours du temps sous l'effet de la diffusion :

$$\frac{\partial \varphi}{\partial t} = D \underbrace{\left(\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} + \frac{\partial^2 \varphi}{\partial z^2} \right)}_{\Delta\varphi = \nabla^2 \varphi} \tag{7.11}$$

Source ponctuelle instantanée. Pour résoudre l'équation 7.11, considérons le système simple unidimensionnel représenté à la figure 7.10. La diffusion se produit selon l'axe des x dans un tube long et étroit, dont la section présente une aire A de taille suffisamment faible pour être négligeable. Posons N la quantité de molécules relâchées en une source ponctuelle $S(x = 0)$ à l'instant initial $t = 0$. La concentration initiale en tout point de l'axe des x est donnée par :

$$\varphi(x, t = 0) = N\delta(x) \tag{7.12}$$

où δ est la fonction delta de Dirac.

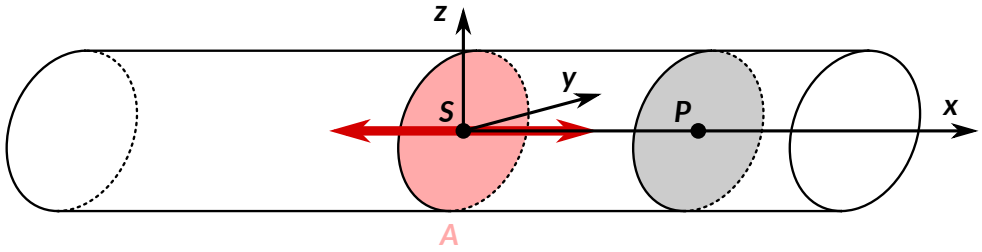


Fig. 7.10. Diffusion depuis une source ponctuelle instantanée dans un système à une dimension.

La solution à l'équation 7.11 prend alors la forme suivante :

$$\varphi(x, t) = \frac{N}{\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt} \right) \tag{7.13}$$

Dans l'équation 7.13, la concentration φ est exprimée en termes de quantité de matière par unité de longueur [NL^{-1}]. Pour transposer cette solution unidimensionnelle à la réalité spatiale, et par là exprimer la concentration en termes de quantité de matière par unité de volume [NL^{-3}], il faut la diviser par l'aire A [L^2] de la section du tube qui avait été négligée au départ :

$$\varphi(x, t) = \frac{N}{A\sqrt{4\pi D t}} \exp\left(-\frac{x^2}{4Dt} \right) \tag{7.14}$$

En supposant que la diffusion se fait de manière isotrope ($D_x = D_y = D_z = D$), l'équation 7.13 peut être généralisée à un système tridimensionnel comme suit :

$$\varphi(x,y,z,t) = \frac{N}{(4\pi Dt)^{\frac{3}{2}}} \exp\left(-\frac{x^2+y^2+z^2}{4Dt}\right) \tag{7.15}$$

Source ponctuelle continue. Si l'on suppose maintenant que S est une source relâchant continûment des molécules dans le système à une concentration constante φ_0 et dans un sens unique, nous nous trouvons dans la situation décrite à la figure 7.11. La solution à l'équation 7.11 devient :

$$\varphi(x,t) = \varphi_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \tag{7.16}$$

où erfc est la fonction d'erreur complémentaire, définie comme :

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = 1 - \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\mathrm{d}t = 1 - \frac{2}{\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n+1}}{n!(2n+1)}$$

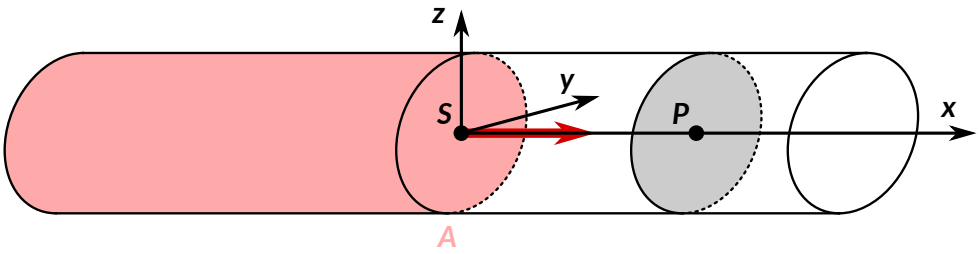


Fig. 7.11. Diffusion depuis une source ponctuelle instantanée dans un système semi-fini à une dimension.

Appliquons ce modèle à notre dispositif microfluidique, en nous plaçant à une température de 25 °C. Dans la mesure où nos chimioattractants candidats sont de petites protéines d'environ 100 aa, considérons une valeur de poids moléculaire $M = 10\,\mathrm{kDa}$. Les valeurs de départ sont donc les suivantes :

Constante de Boltzmann	k_B	=	$1{,}380\,648\,52 \times 10^{-23}\,\mathrm{J\,K^{-1}}$
Constante d'Avogadro	N_A	=	$6{,}022\,140\,857 \times 10^{23}\,\mathrm{mol^{-1}}$
Température	T	=	$293{,}15\,\mathrm{K}$
Densité protéique	ρ	=	$1370\,\mathrm{kg\,m^{-3}}$
Masse molaire protéique	M	=	$10\,\mathrm{kg\,mol^{-1}}$

Avec l'équation 7.4, on peut retrouver le rayon de Stokes R_S de notre protéine attractante :

$$R_S = \left(\frac{3 \times 10}{4\pi \times 1370 \times 6{,}022\,140\,857 \times 10^{23}}\right)^{1/3} = 1{,}425 \times 10^{-9}\,\mathrm{m}$$

L'équation 7.6 permet de calculer la viscosité dynamique du milieu de culture (assimilé à de l'eau) :

$$\mu = 2.414 \times 10^{\frac{247.8}{293.15-140}-5} = 1{,}002 \times 10^{-3}\,\mathrm{Pa\,s}$$

Enfin, on peut utilser ces deux valeurs dans l'équation 7.9 pour calculer le coefficient de diffusion de nos protéines dans le dispositif :

$$D = \frac{1{,}380\,648\,52 \times 10^{-23} \times 293{,}15}{6\pi \times 1{,}425 \times 10^{-9} \times 1{,}002 \times 10^{-3}} = 1{,}504 \times 10^{-10}\,\mathrm{m^2\,s^{-1}}$$

Considérons qu'à $t = 0$, une goutelette de solution protéique de concentration C_0 est déposée dans la cavité entourant le point P. Pour simplifier, on supposera que l'action de déposer la goutelette dans l'orifice homogénéise instantanément le contenu de la cavité, et que donc la concentration de protéine à $t = 0$ est la même dans toute la cavité. On admettra également que la cavité ayant un volume nettement plus grand que les canaux transversaux, la quantité de protéines entrant par diffusion dans ces canaux est négligeable, ce qui implique que la concentration en protéines dans la cavité est réputée constante au cours du temps.

Ainsi, les points G_i décrits à la figure 7.7 sont assimilables à des sources ponctuelles continues, fournissant à chaque instant des protéines à une concentration constante $\varphi_0 = C_0$. La concentration en protéines selon l'axe des x est alors donnée par l'équation 7.16, et il devient alors possible de décrire la diffusion d'une protéine de poids moléculaire $M = 10\,\mathrm{kDa}$ avec une concentration initiale $\varphi_0 = 1\,\mu\mathrm{M}$ entre les points L_i et R_i, en fonction du temps (figure 7.12) et de la position (figure 7.13). La figure 7.14 décrit comment le rapport des concentrations entre le début (L_i) et la fin (R_i) des canaux varie au cours du temps.

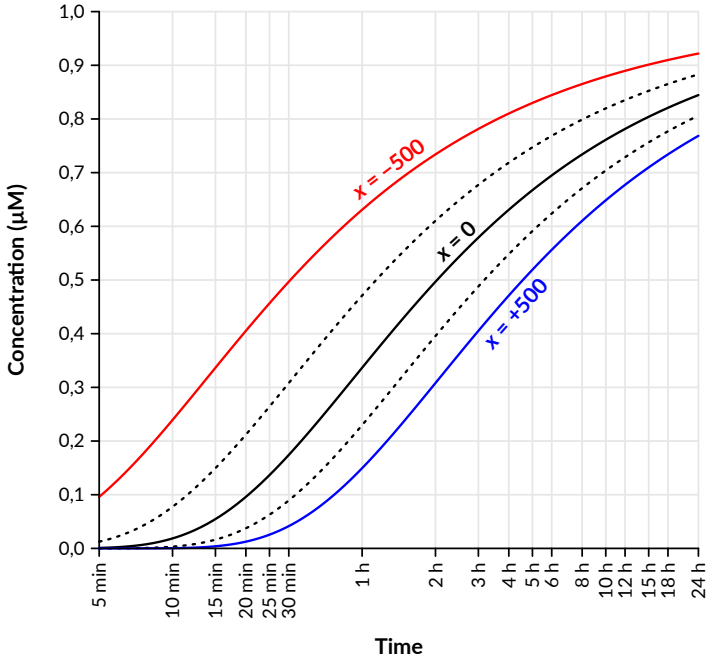


Fig. 7.12. Évolution théorique de la concentration protéique en fonction du temps, à différentes positions dans le microcanal. Les calculs ont été faits aux positions $x = -500$, -250, 0, 250 et 500 µm du microcanal, avec une concentration initiale $\varphi_0 = 1$ µM.

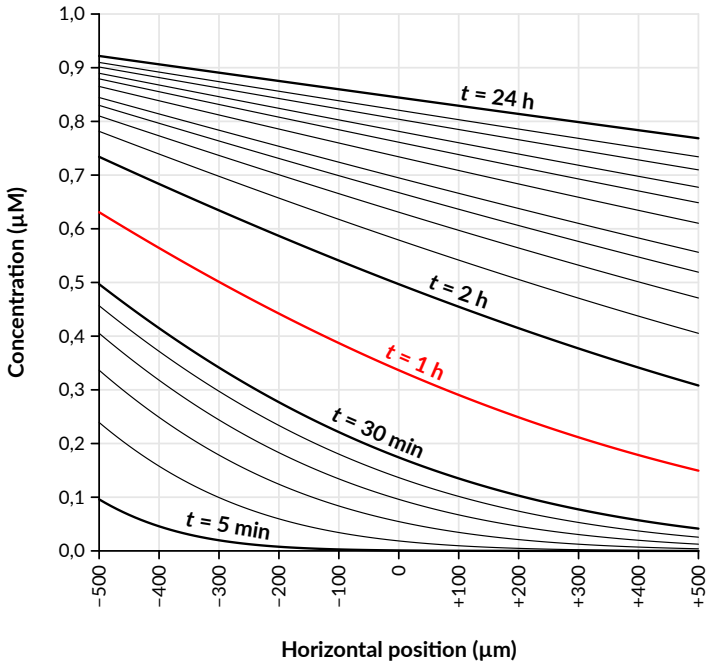


Fig. 7.13. Évolution théorique de la concentration protéique en fonction de la position dans le microcanal, à différents temps. Les calculs ont été faits aux temps donnés à la figure 7.12, avec une concentration initiale $\varphi_0 = 1$ µM.

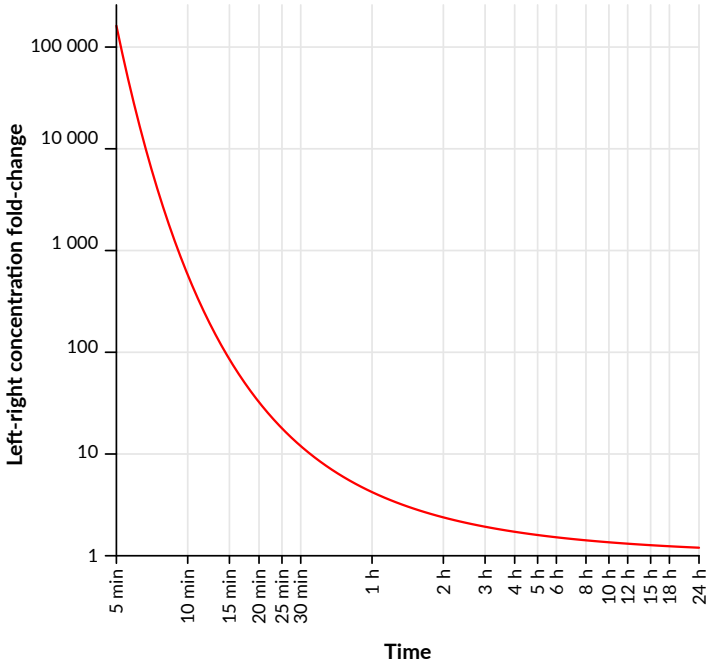


Fig. 7.14. Évolution théorique au cours du temps du rapport des concentrations protéiques entre les extrémités du microcanal. Les calculs ont été faits entre les abscisses $x = -500\,\mu\text{m}$ et $x = 500\,\mu\text{m}$, avec $\varphi_0 = 1\,\mu\text{M}$.

7.3 Remarques conclusives

Cette thèse nous aura permis de mettre en évidence que l'ovule de *S. chacoense* présente un paysage transcriptomique et protéomique particulier, faisant intervenir des gènes susceptibles de jouer des rôles clés dans les interactions pollen-ovule et l'isolement reproductif, comme ceux codant pour les CRP. Nous avons également compris que l'expression génique dans les ovules est finement régulée par plusieurs facteurs : (i) l'absence ou la présence de sac embryonnaire, comme on l'a analysé avec le mutant *frk1*, (ii) le stade de développement de ce dernier, y compris dans les étapes tardives, comme on l'a vu avec les ovules immature 2 DBA, (iii) la présence de TP en croissance dans le pistil et, le cas échéant, leur génotype et leur distance par rapport à l'ovaire. Nous avons aussi constaté que ces facteurs affectent de manière indépendante la transcription des ARN et la sécrétion des protéines.

Pour répondre aux hypothèses soulevées au premier chapitre, nous avons en particulier identifié grâce à l'outil KAPPA une liste de 29 CRP de type défensine enrichies dans l'ovule de *S. chacoense*, dont quatre présentent un profil d'expression spécifique au sac embryonnaire

de l'ovule mature, faisant d'elles de bons chimioattractants candidats. Les dispositifs microfluidiques présentés ci-dessus seront un atout précieux dans la boîte à outils méthodologique des futurs doctorants qui chercheront à en vérifier la fonction.

Une fois les bons attractants identifiés, il sera nécessaire de multiplier les tests de guidage pour déterminer la concentration et la distance optimales d'action de ces protéines et pour évaluer, avec des tests interspécifiques, leur contribution éventuelle à l'isolement reproductif. Il sera possible, par des analyses de mutagenèse dirigée, investiguer la contribution de chaque résidu des attractants dans leur capacité à attirer les TP et à assurer, le cas échéant, une préférence à l'espèce.

La localisation précise de l'expression des attractants devra être évaluée, par exemple grâce à des plantes mutantes exprimant des protéines de fusion GFP ou GUS. D'autre part, comme effectué tout récemment avec les LURE d'*Arabidopsis*[4], il sera intéressant de générer, grâce à la technologie CRISPR/Cas9, des mutants nuls pour les attractants candidats, afin de comprendre à quelle étape du guidage du TP ils contribuent. L'imagerie confocale à deux photons, permettant d'effectuer une dissection optique en profondeur des spécimens observés, serait un outil précieux pour caractériser le phénotype de tels mutants[17].

Par ailleurs, des mutants de transfert latéral d'attractants entre espèces, tels que ceux obtenus en exprimant les attractants ZmEA1 du maïs chez *Arabidopsis*[18], fourniraient une preuve robuste de leur contribution à l'isolement reproductif. Pour toutes ces expériences d'édition du génome, il serait souhaitable de changer de génotype d'étude et choisir, plutôt que le génotype auto-incompatible G4 utilisé à l'IRBV, le clone M6[19]. Ce dernier a en effet été manipulé génétiquement pour l'exprimer le gène inhibiteur de l'auto-incompatibilité *Sli*, rendant possible l'obtention par croisement de lignées mutantes homozygotes. Pour cette raison, M6 est devenu un génotype-modèle utilisé dans plusieurs projets de recherche agronomique, ainsi que pour la génération du premier génome de *S. chacoense*[20].

Enfin, les nouvelles annotations augmentées que nous avons fournies du génome de *S. chacoense* ont permis d'identifier plus de 300 récepteurs kinases (RLK), parmi lesquels se trouve peut-être le récepteur pollinique au chimioattractant. Une analyse du transcriptome du pollen, préférentiellement du TP obtenu semi-*in vivo*, permettrait de constituer là encore une liste de RLK candidats à tester fonctionnellement, ouvrant la voie à de nouveaux projets de recherche sur le versant mâle des interactions pollen-pistil.

Bibliographie

1. Okuda, S. ; Tsutsui, H. ; Shiina, K. et coll. (2009). Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature*, 458(7236), 357–61. DOI : 10.1038/nature07882. [cit. p. 201, 203, 207, 208, 209]

2. Kanaoka, M. M. ; Kawano, N. ; Matsubara, Y. et coll. (2011). Identification and characterization of TcCRP1, a pollen tube attractant from *Torenia concolor*. *Ann. Bot.*, 108(4), 739–47. DOI : 10.1093/aob/mcr111. [cit. p. 201, 203, 207, 208]

3. Takeuchi, H. et Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol.*, 10(12), e1001449. DOI : 10.1371/journal.pbio.1001449. [cit. p. 201, 203, 207, 208, 209]

4. Zhong, S. ; Liu, M. ; Wang, Z. et coll. (2019). Cysteine-rich peptides promote interspecific genetic isolation in *Arabidopsis*. *Science*, 364(6443). DOI : 10.1126/science.aau9564. [cit. p. 201, 203, 207, 208, 222]

5. Takeuchi, H. et Higashiyama, T. (2016). Tip-localized receptors control pollen tube growth and LURE sensing in *Arabidopsis*. *Nature*, 531(7593), 245–8. DOI : 10.1038/nature17413. [cit. p. 207]

6. Liu, Y. (2015). *The plant ovule omics : an integrative approach for pollen-pistil interactions and pollen tube guidance studies in solanaceous species*. Thèse de doctorat, Université de Montréal. URL : http ://hdl.handle.net/1866/13589. [cit. p. 207, 208]

7. Viallet, C. (2016). *Identification de protéines impliquées dans le guidage du tube pollinique par les ovules de* Solanum chacoense. Mémoire de maîtrise, Université de Montréal. [cit. p. 207, 208]

8. Sanati Nezhad, A. ; Naghavi, M. ; Packirisamy, M. ; Bhat, R. et Geitmann, A. (2013). Quantification of cellular penetrative forces using lab-on-a-chip technology and finite element modeling. *Proc. Natl. Acad. Sci. U. S. A.*, 110(20), 8093–8. DOI : 10.1073/pnas.1221677110. [cit. p. 209]

9. Horade, M. ; Yanagisawa, N. ; Mizuta, Y. ; Higashiyama, T. et Arata, H. (2014). Growth assay of individual pollen tubes arrayed by microchannel device. *Microelectron. Eng.*, 118, 25–8. DOI : 10.1016/j.mee.2014.01.017. [cit. p. 209]

10. Burri, J. T. ; Vogler, H. ; Läubli, N. F. et coll. (2018). Feeling the force : how pollen tubes deal with obstacles. *New Phytol.*, 220(1), 187–95. DOI : 10.1111/nph.15260. [cit. p. 209]

11. Geitmann, A. (2017). Microfluidic- and microelectromechanical system (MEMS)-based platforms for experimental analysis of pollen tube growth behavior and quantification of cell mechanical properties. *In* Obermeyer, G. et Feijó, J., éditeurs : *Pollen tip growth : from biophysical aspects to systems biology.*, pages 87–103. Springer, Berlin, Allemagne. ISBN : 978-3-319-56644-3. DOI : 10.1007/978-3-319-56645-0_5. [cit. p. 209]

12. Yanagisawa, N. ; Sugimoto, N. ; Arata, H. ; Higashiyama, T. et Sato, Y. (2017). Capability of tip-growing plant cells to penetrate into extremely narrow gaps. *Sci. Rep.*, 7(1), 1403. DOI : 10.1038/s41598-017-01610-w. [cit. p. 209]

13. Yanagisawa, N. ; Sugimoto, N. ; Higashiyama, T. et Sato, Y. (2018). Development of microfluidic devices to study the elongation capability of tip-growing plant cells in extremely small spaces. *J. Vis. Exp.*, (135). DOI : 10.3791/57262. [cit. p. 209]

14. Horade, M. ; Kanaoka, M. M. ; Kuzuya, M. ; Higashiyama, T. et Kaji, N. (2013). A microfluidic device for quantitative analysis of chemoattraction in plants. *RSC Adv.*, 3(44), 22301. DOI : 10.1039/c3ra42804d. [cit. p. 209, 210]

15. Sato, Y. ; Sugimoto, N. ; Higashiyama, T. et Arata, H. (2015). Quantification of pollen tube attraction in response to guidance by female gametophyte tissue using artificial microscale pathway. *J. Biosci. Bioeng.*, 120(6), 697–700. DOI : 10.1016/j.jbiosc.2015.03.021. [cit. p. 209, 210]

16. Yanagisawa, N. et Higashiyama, T. (2018). Quantitative assessment of chemotropism in pollen tubes using microslit channel filters. *Biomicrofluidics*, 12(2), 024113. DOI : 10.1063/1.5023718. [cit. p. 209, 210]

17. Cheung, A. Y. ; Boavida, L. C. ; Aggarwal, M. ; Wu, H.-M. et Feijó, J. A. (2010). The pollen tube journey in the pistil and imaging the *in vivo* process by two-photon microscopy. *J. Exp. Bot.*, 61(7), 1907–15. DOI : 10.1093/jxb/erq062. [cit. p. 222]

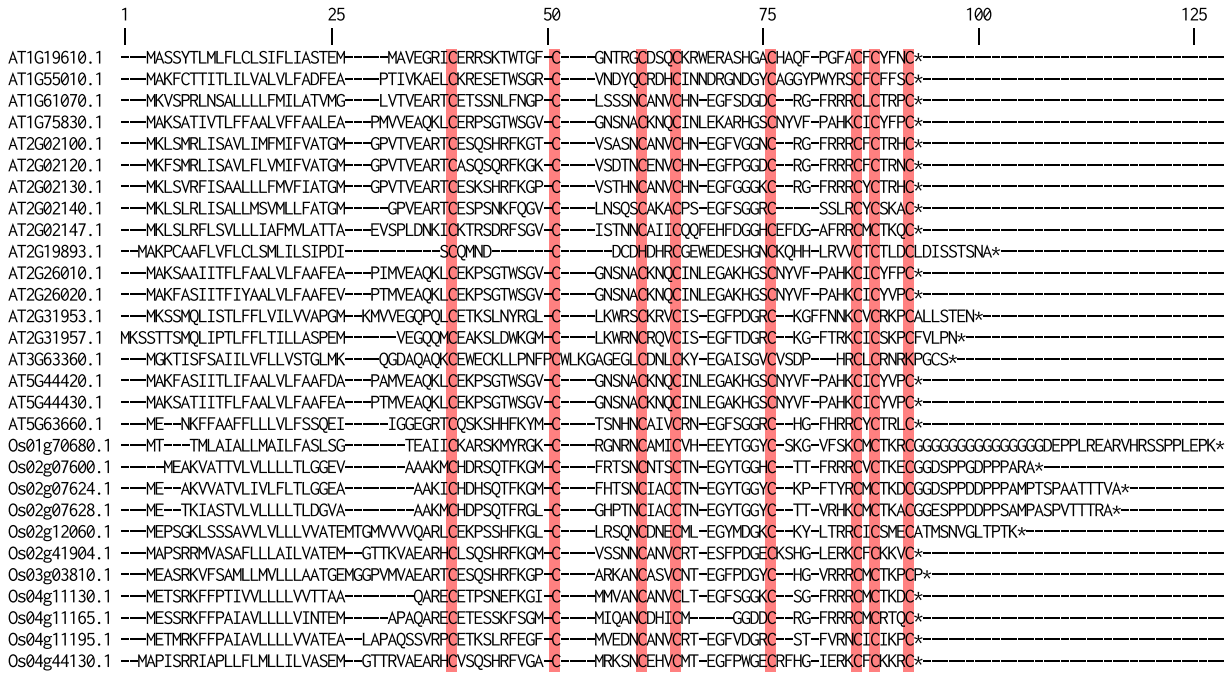
18. Márton, M. L. ; Fastner, A. ; Uebler, S. et Dresselhaus, T. (2012). Overcoming hybridization barriers by the secretion of the maize pollen tube attractant ZmEA1 from *Arabidopsis* ovules. *Curr. Biol.*, 22(13), 1194–8. DOI : 10.1016/j.cub.2012.04.061. [cit. p. 222]

19. Jansky, S. H. ; Chung, Y. S. et Kittipadukal, P. (2014). M6 : a diploid potato inbred line for use in breeding and genetics research. *J. Plant Registr.*, 8(2), 195. DOI : 10.3198/jpr2013.05.0024crg. [cit. p. 222]

20. Leisner, C. P. ; Hamilton, J. P. ; Crisovan, E. et coll. (2018). Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. *Plant J.*, 94(3), 562–70. DOI : 10.1111/tpj.13857. [cit. p. 222]

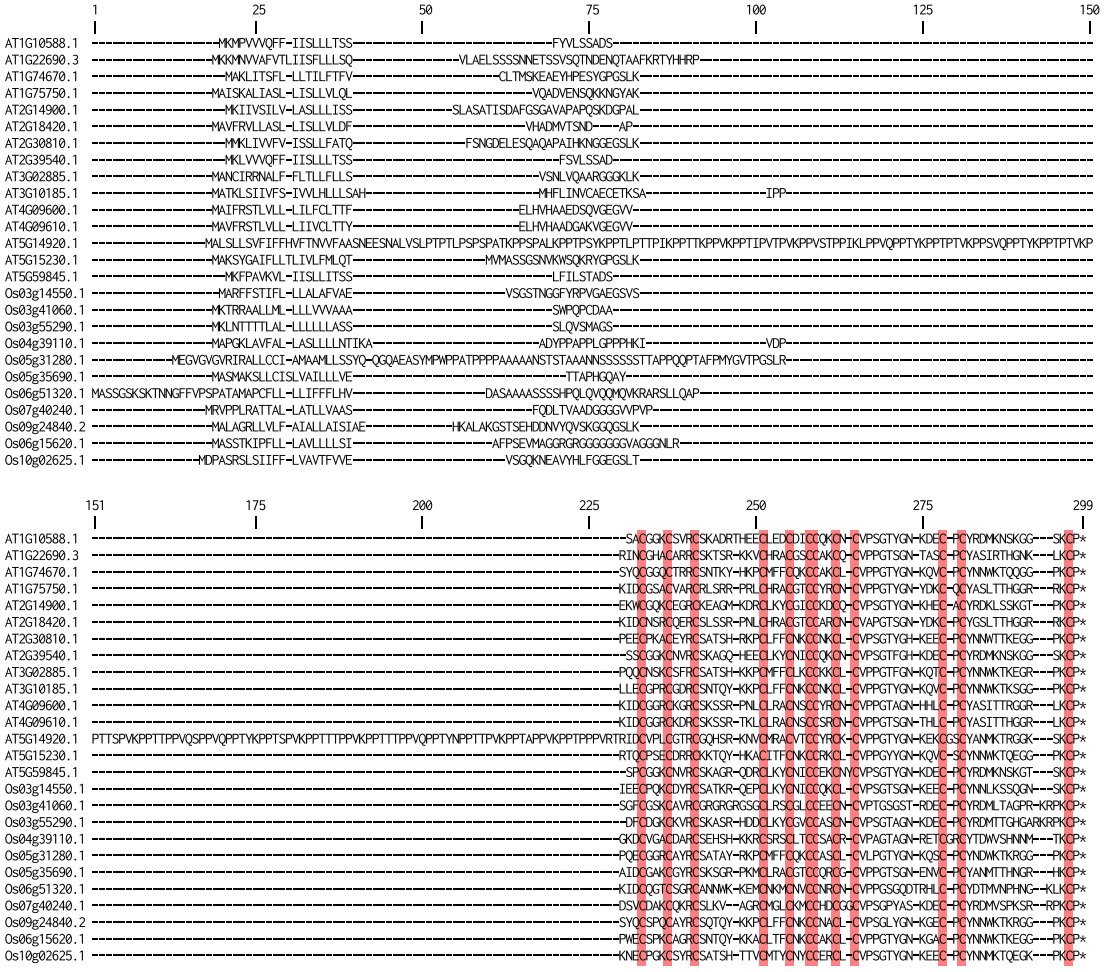
ANNEXE A

Matériel supplémentaire du chapitre 3

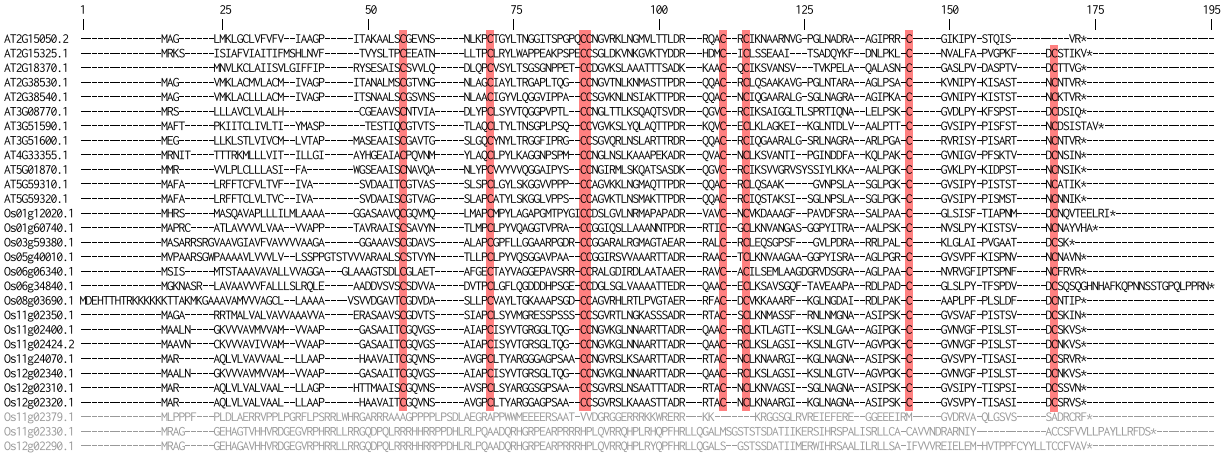


(a) Defensins

Figure S3.1. Multiple alignments of reference input proteins used in this study. *Arabidopsis* thaliana and Oryza sativa protein sequences were retrieved from Silverstein *et al.* (2007) for defensins and snakins and from Edstam *et al.* (2011) for LTP1s. To ensure experimental homogeneity, we used BLASTp to align each of these proteins against the *Arabidopsis* and rice proteomes from the Phytozome 9.1 release. In a small minority of cases, no hit—or only a low quality hit—was found, since some proteins turned to be obsolete or because they corresponded to pseudogenes rather than true proteins. These were discarded. In addition, rice LTP1 sequences coloured in gray in the alignment were not taken into account since they did not display a cysteine pattern. Multiple alignments were generated with MUSCLE (default parameters).

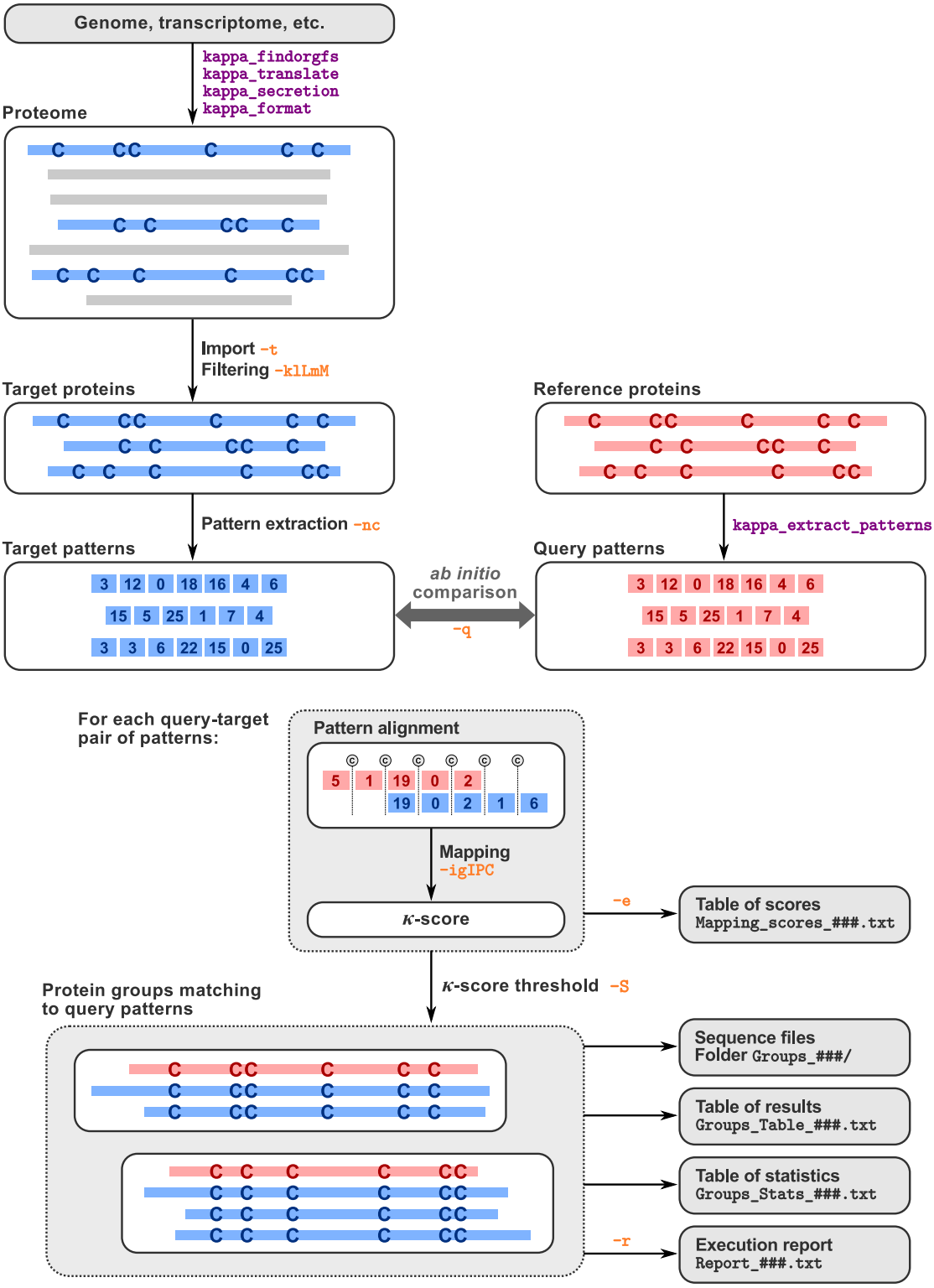


(b) Snakins



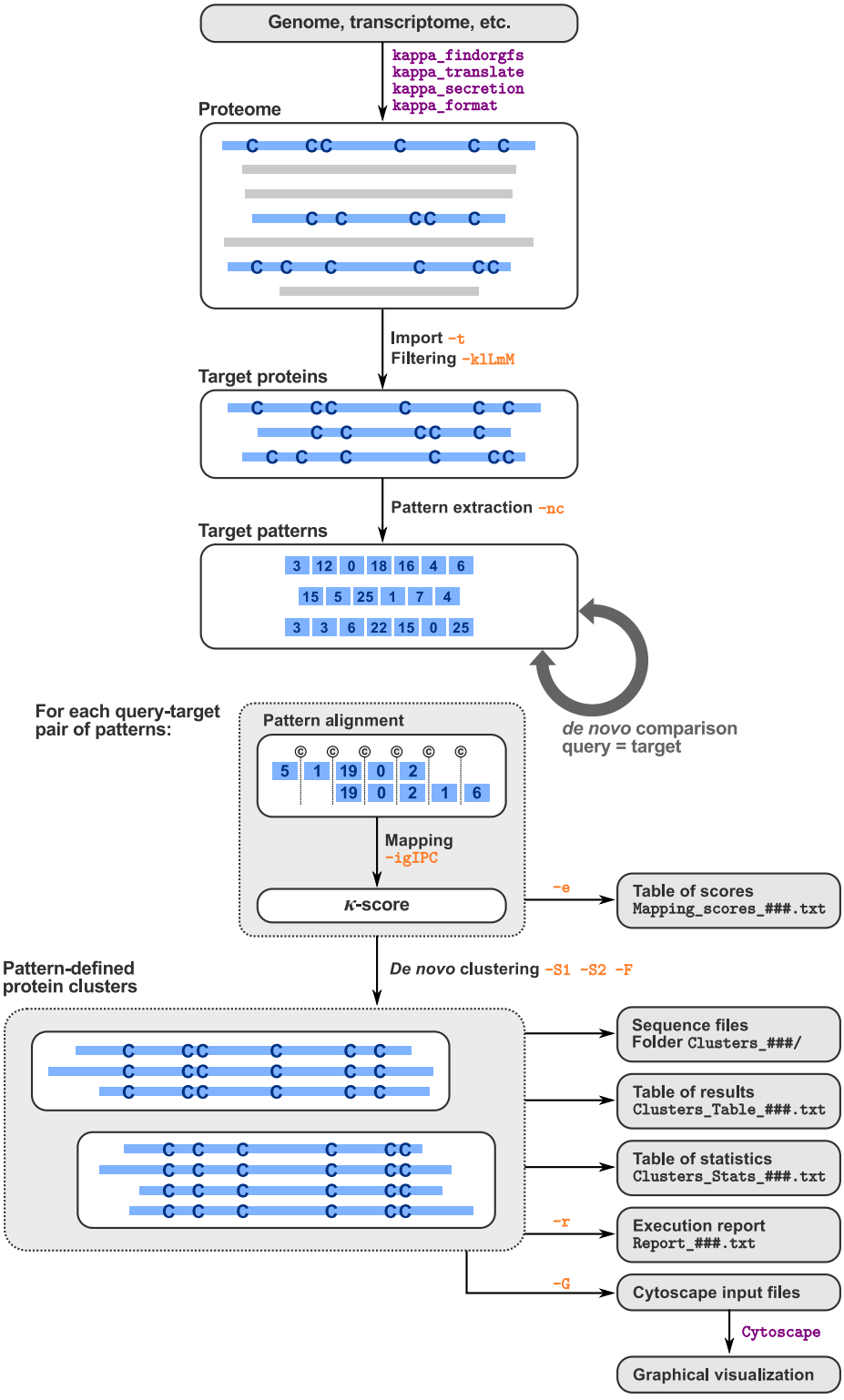
(c) Lipid-transfer proteins 1 (LTP1)

Figure S3.1. (suite)



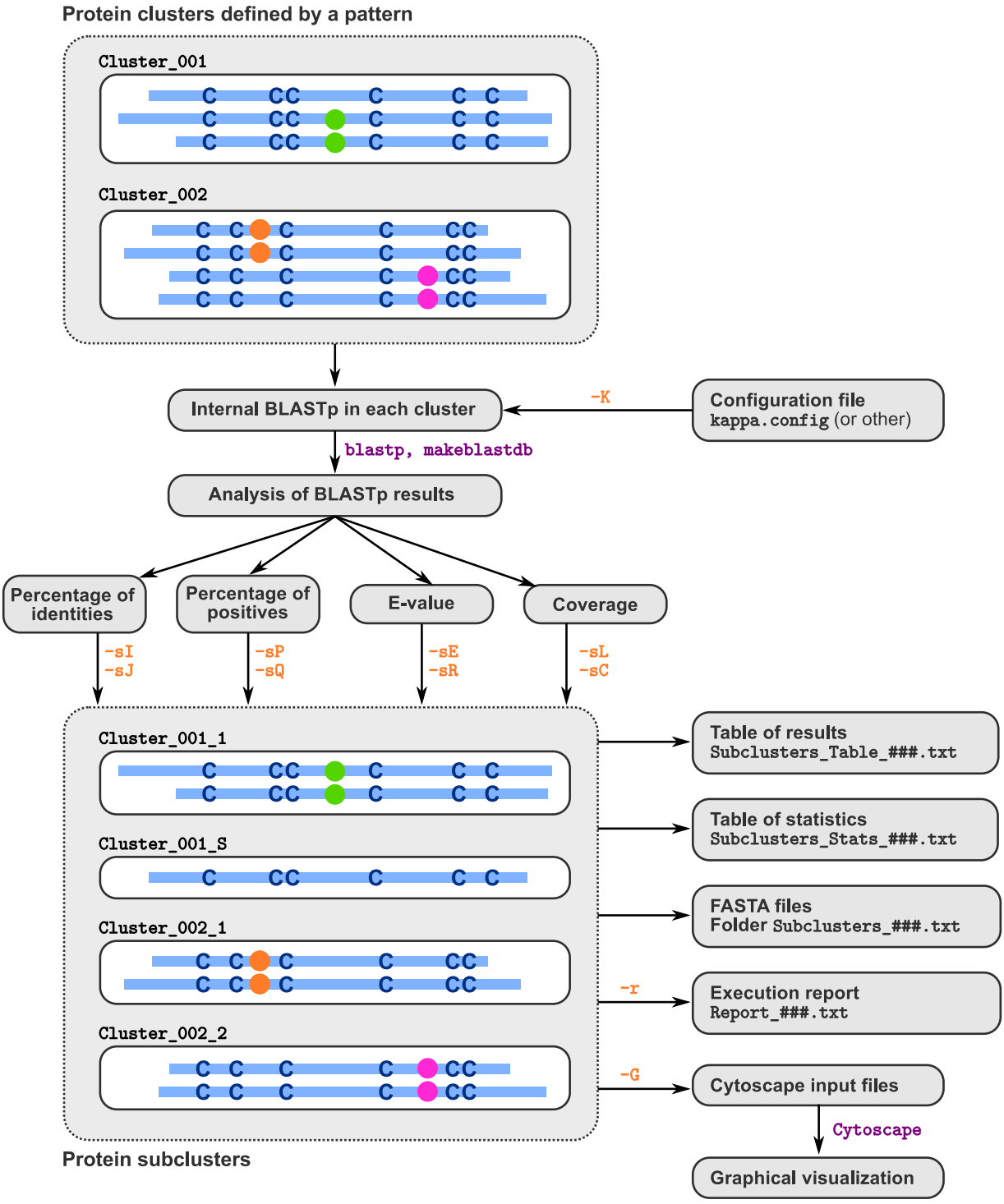
(a) *Ab initio* search

Figure S3.2. Graphical view of the KAPPA workflow.



(b) *De novo* search and clustering

Figure S3.2. (suite)



(c) Subclustering

Figure S3.2. (suite)

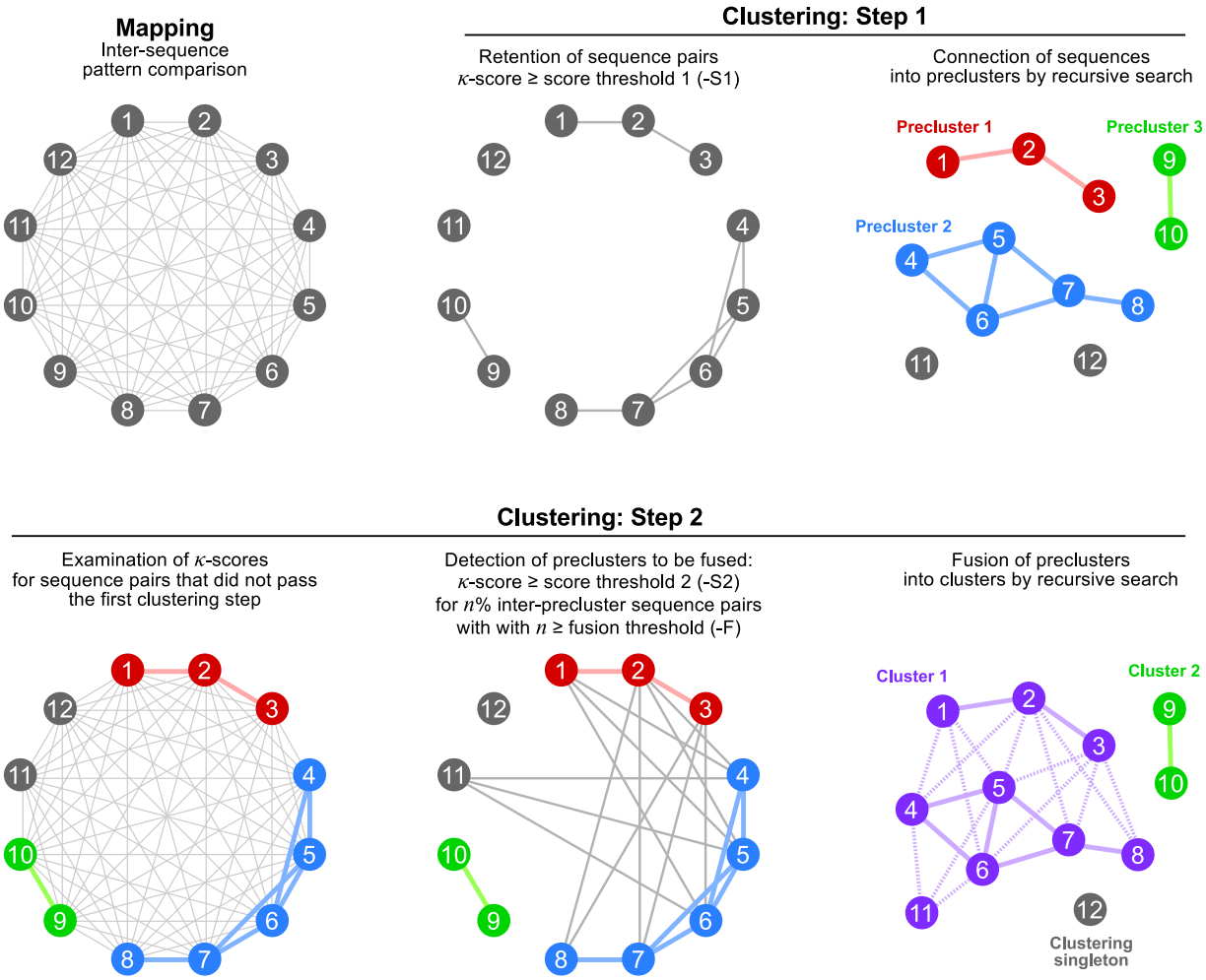
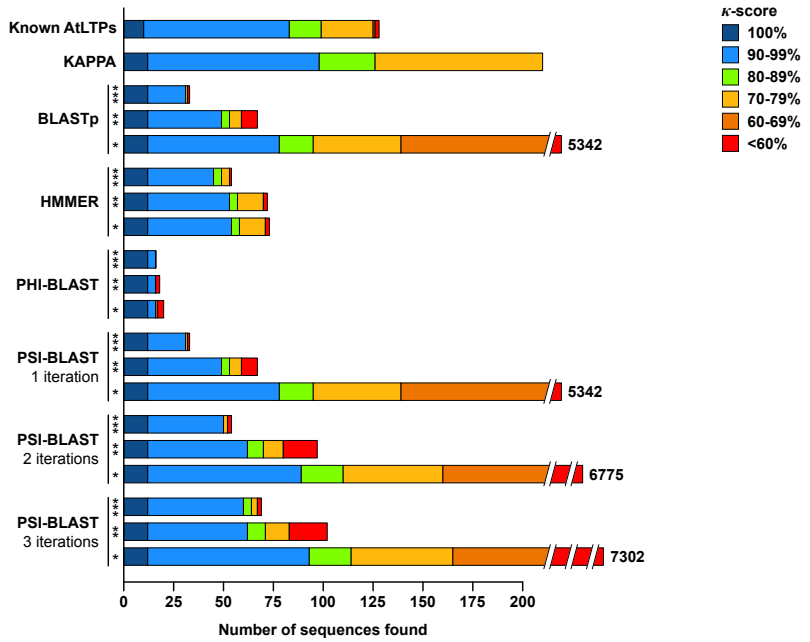
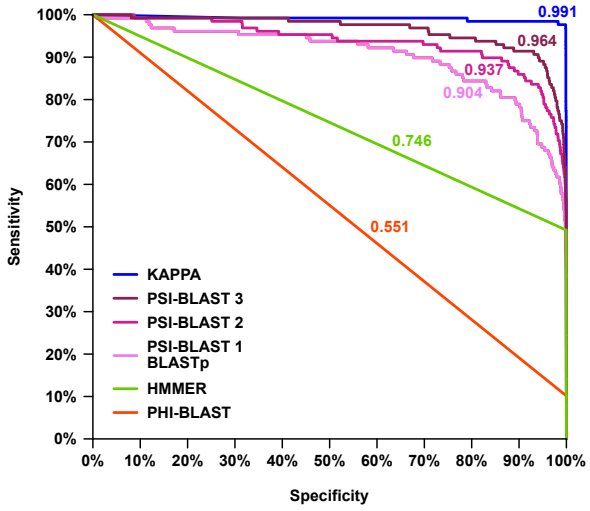


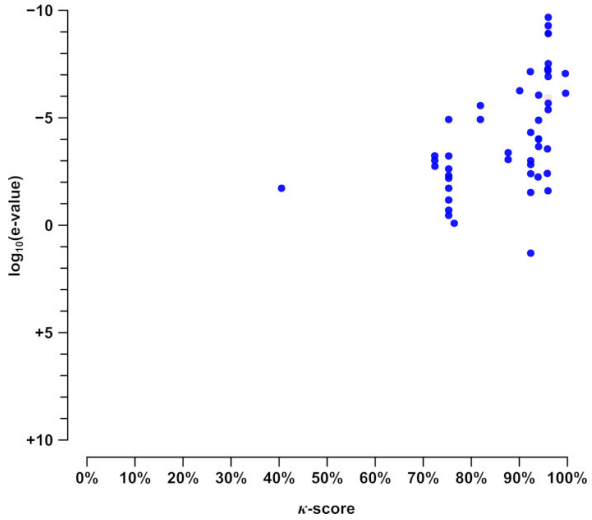
Figure S3.3. Schematic view of KAPPA de novo clustering method.



(a) Comparison of outputs from different programs. AtLTP1s were used as input sequences to parse the *A. thaliana* proteome available in Phytozome 9.1. KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of AtLTP1s. Known AtLTPs were used to define a minimal expected output; they correspond to those described by Silverstein *et al.* (2007) and Edstam *et al.* (2011).

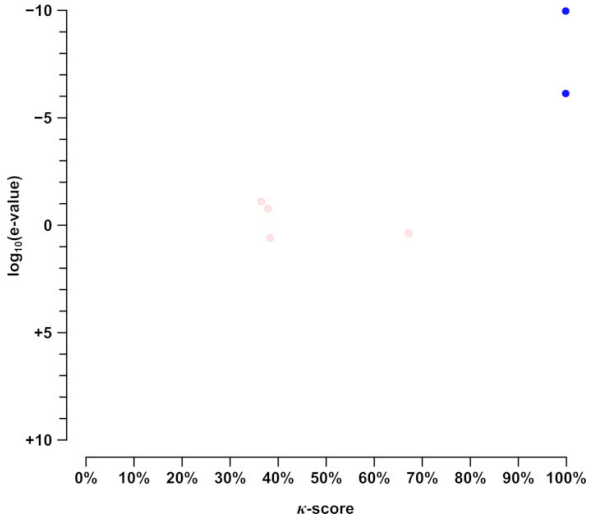


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known AtLTPs as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

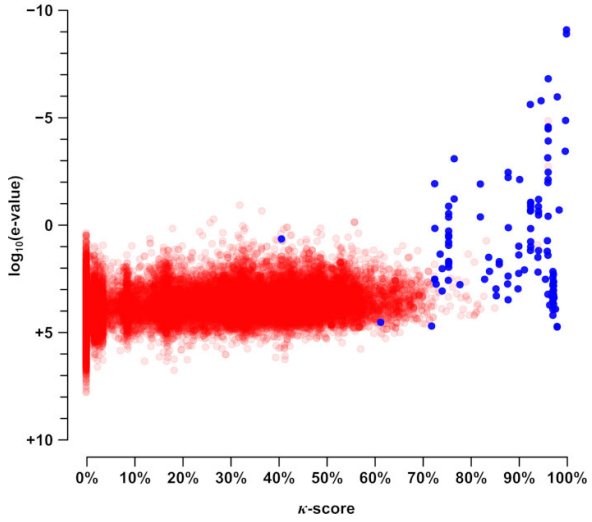


(c) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **HMMER**. Blue dots correspond to known AtLTPs; red dots represent other proteins.

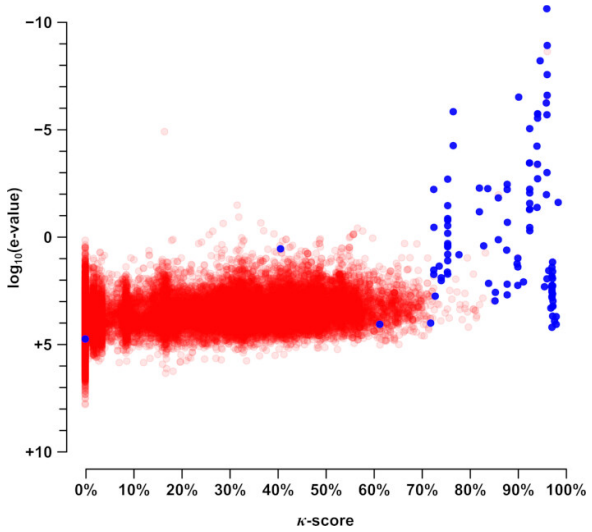
Figure S3.4. Assessment of *ab initio* sequence search performance for KAPPA and other programs: *Arabidopsis* LTP1s in the *Arabidopsis* proteome.



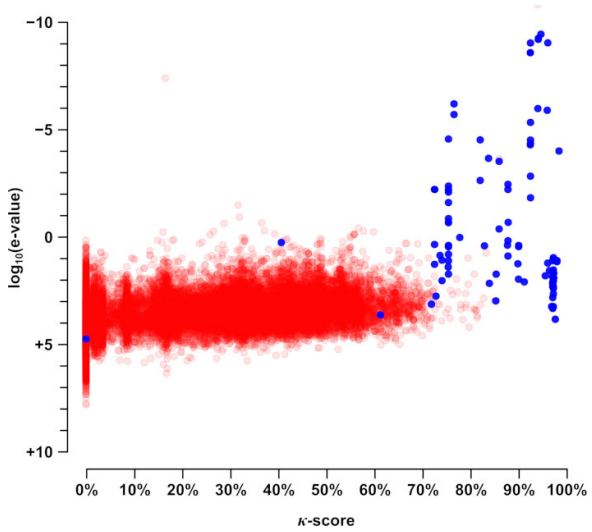
(d) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **PHI-BLAST**. Blue dots correspond to known AtLTPs; red dots represent other proteins.



(e) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to known AtLTPs; red dots represent other proteins.

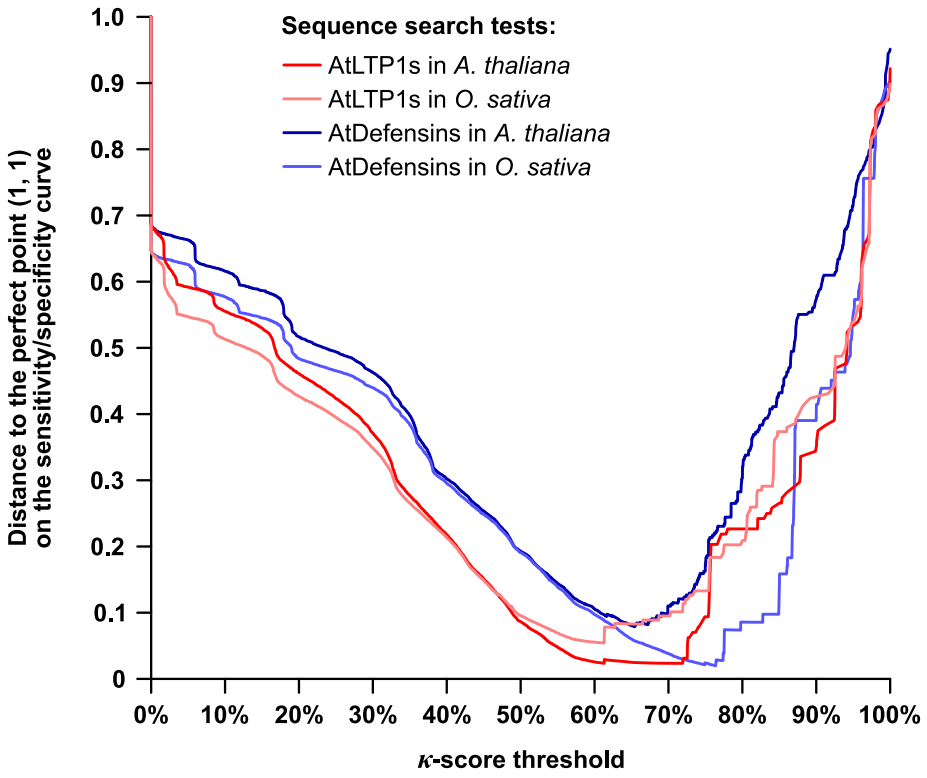


(f) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to known AtLTPs; red dots represent other proteins.

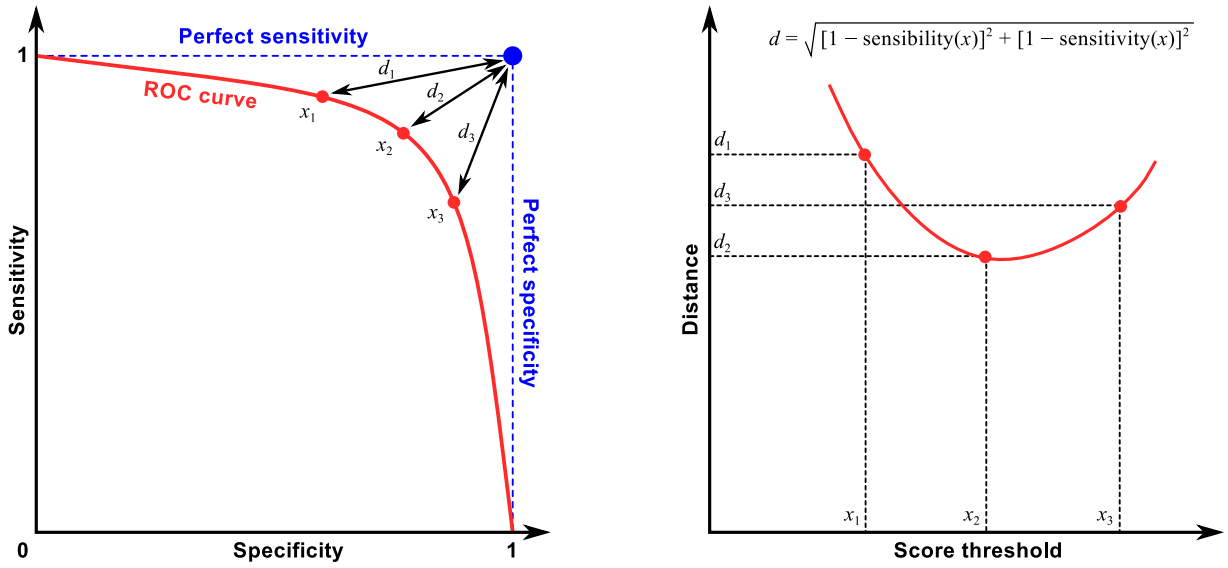


(g) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to known AtLTPs; red dots represent other proteins.

Figure S3.4. (suite)



(a) Optimization results



(b) Distance calculation method

Figure S3.5. Determination of the most suitable κ-score threshold for KAPPA *ab initio* sequence search tests on LTPs and defensins studying distances between the specificity/sensitivity ROC curve and the perfect prediction point (1, 1). These distance were computed for each point along the ROC curve as explained in the box below. This figure shows that a κ-score threshold of about 70 % is generally a good compromise between sensitivity and specificity when studying highly divergent cysteine-rich proteins displaying small modification in their cysteine spacing such as LTPs and defensins. However, the κ-score threshold can be raised to much higher values when exact matches to a given motif are required.

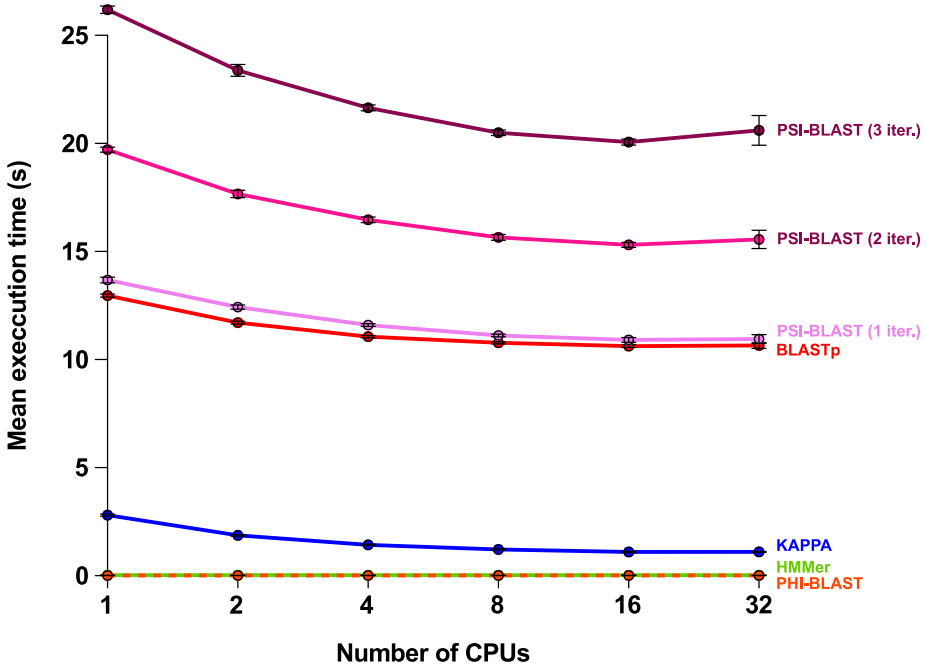
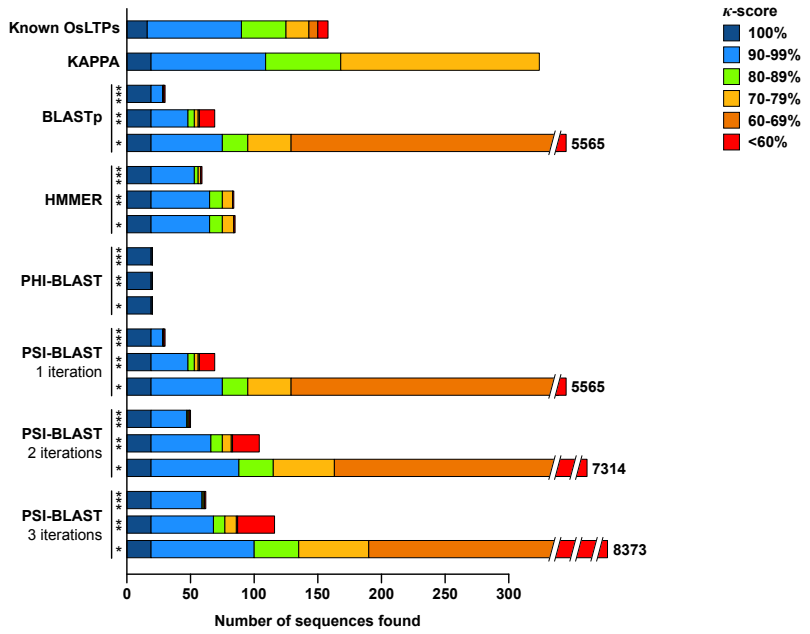
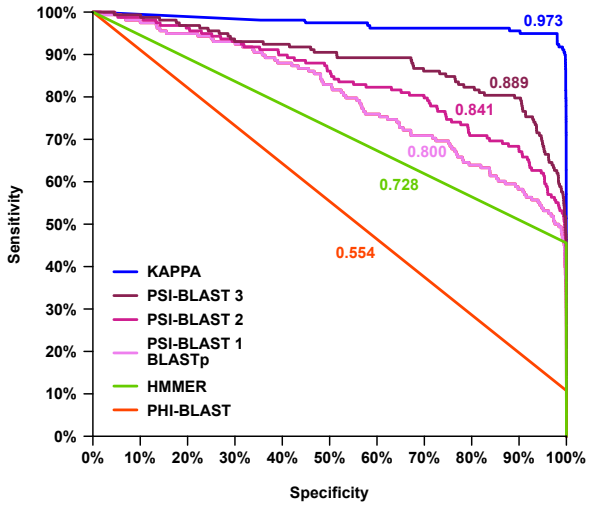


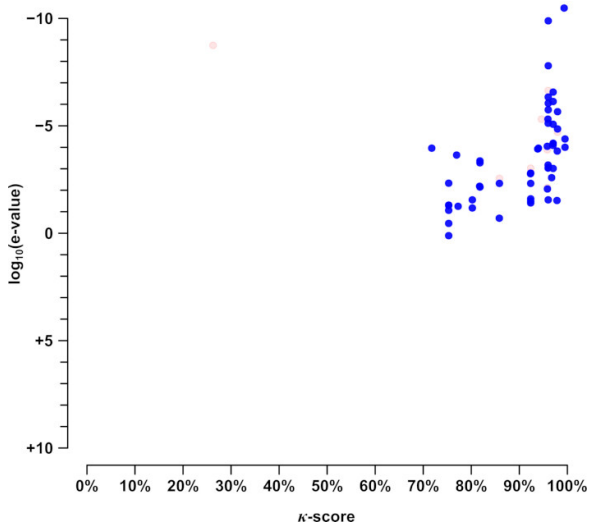
Figure S3.6. Comparison of sequence search speed for KAPPA and other programs. Execution times were measured taking the example of *ab initio* search of LTPs in the *Arabidopsis* proteome shown in Figure S3.4, allocating between 1 and 32 processors to sequence search. All programs were run on a Linux server with a double Intel® Xeon® Octo Core E5-2650 processor and 128 GB DDR3 ECC SDRAM (1600 MHz), executing Ubuntu Server 12.04 LTS. Here are reported "real execution times" measured by the `time` UNIX command.



(a) Comparison of outputs from different programs. AtLTP1s were used as input sequences to parse the rice proteome available in Phytozome 9.1. KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of AtLTP1s. Known OsLTPs were used to define a minimal expected output; they correspond to those described by Silverstein *et al.* (2007) and Edstam *et al.* (2011).

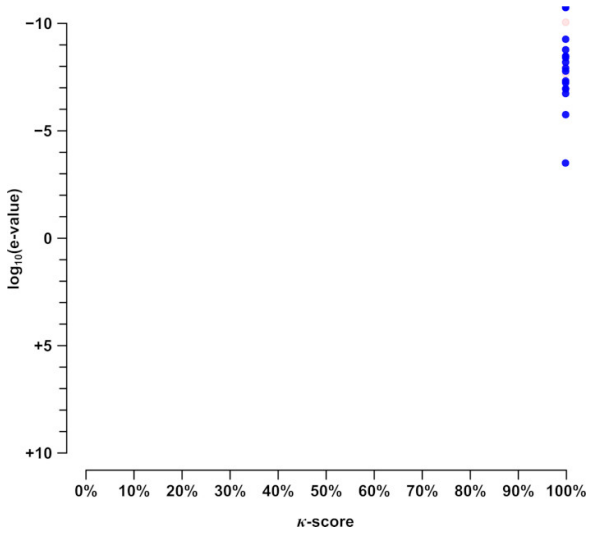


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known OsLTPs as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

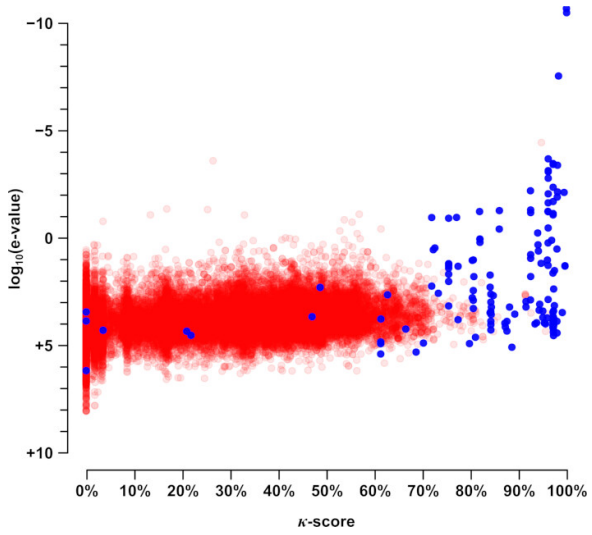


(c) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **HMMER**. Blue dots correspond to known AtLTPs; red dots represent other proteins.

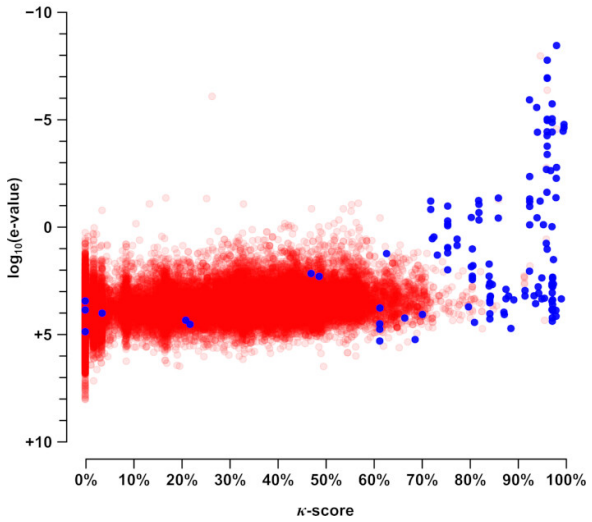
Figure S3.7. Assessment of *ab initio* sequence search performance for KAPPA and other programs: *Arabidopsis* LTP1s in the rice proteome.



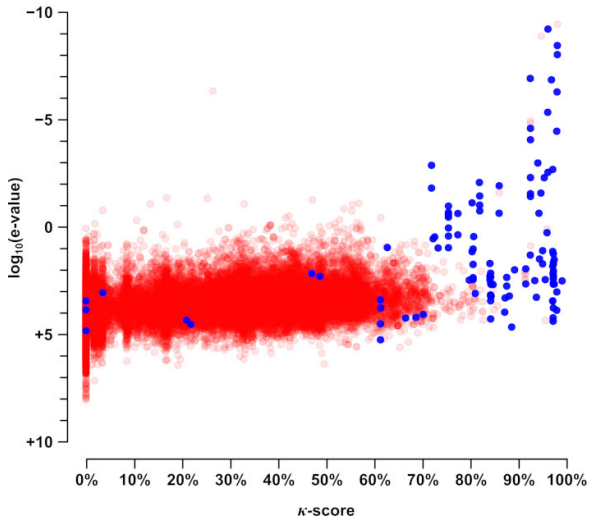
(d) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **PHI-BLAST**. Blue dots correspond to known AtLTPs; red dots represent other proteins.



(e) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to known AtLTPs; red dots represent other proteins.

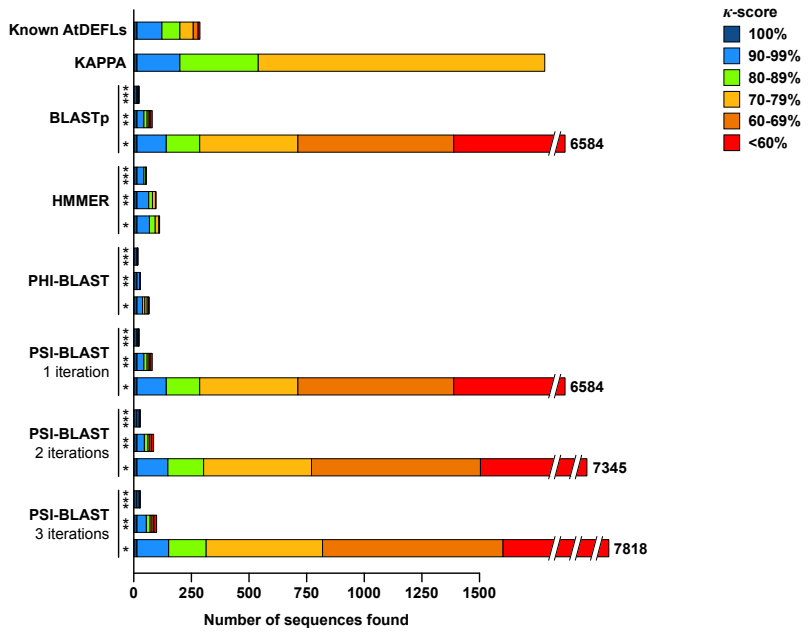


(f) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to known AtLTPs; red dots represent other proteins.

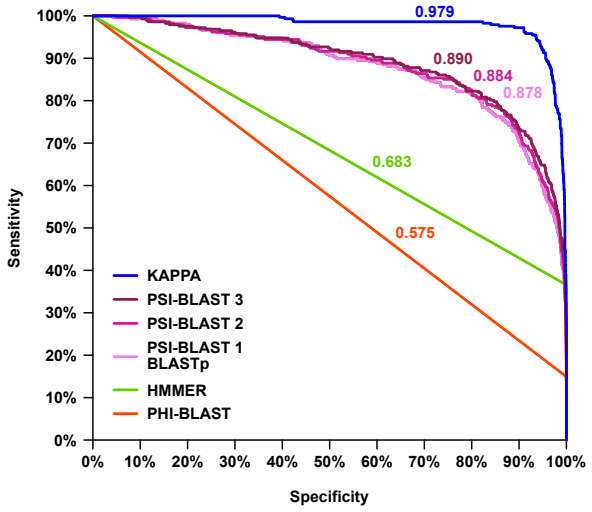


(g) E-value for the best match to a query AtLTP1 vs. κ-score according to the AtLTP1 consensus pattern for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to known AtLTPs; red dots represent other proteins.

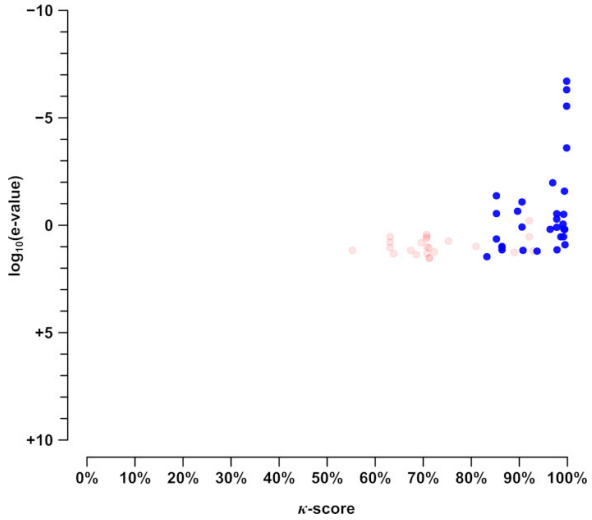
Figure S3.7. (suite)



(a) Comparison of outputs from different programs. *Arabidopsis* defensins were used as input sequences to parse the *A. thaliana* proteome available in Phytozome 9.1. KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of *Arabidopsis* defensins. Known *Arabidopsis* defensin-like proteins (AtDEFLs) described by Silverstein *et al.* (2007) were used to define a minimal expected output.

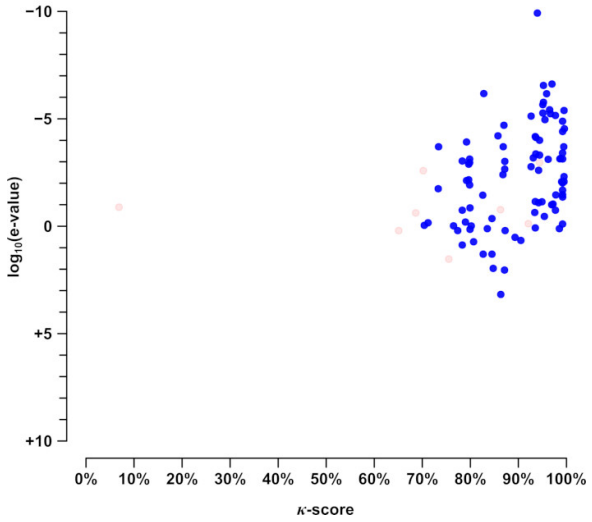


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known AtDEFLs as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

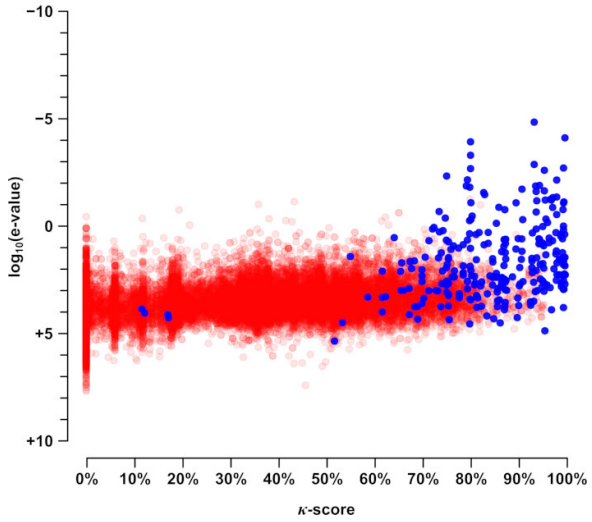


(c) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **HMMER**. Blue dots correspond to known AtDEFLs; red dots represent other proteins.

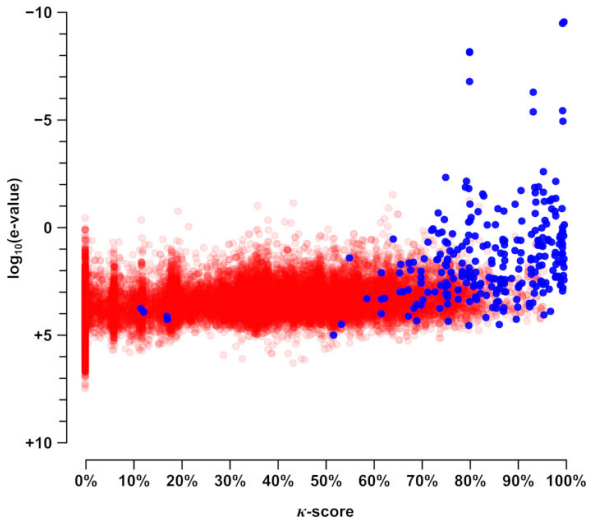
Figure S3.8. Assessment of *ab initio* sequence search performance for KAPPA and other programs: *Arabidopsis* defensins in the *Arabidopsis* proteome.



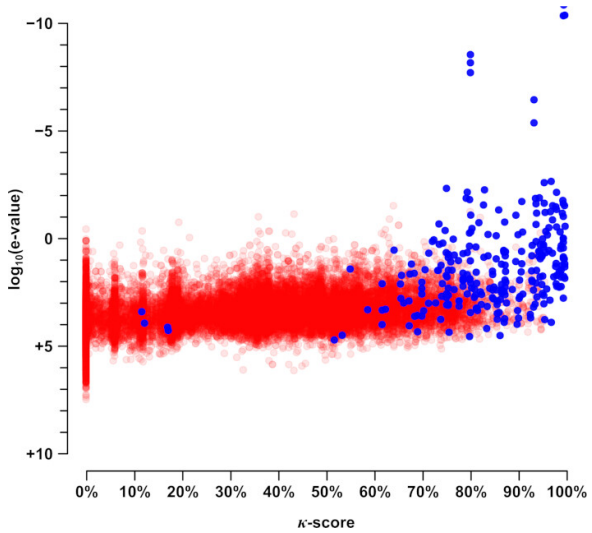
(d) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **PHI-BLAST**. Blue dots correspond to known AtDEFLs; red dots represent other proteins.



(e) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to known AtDEFLs; red dots represent other proteins.

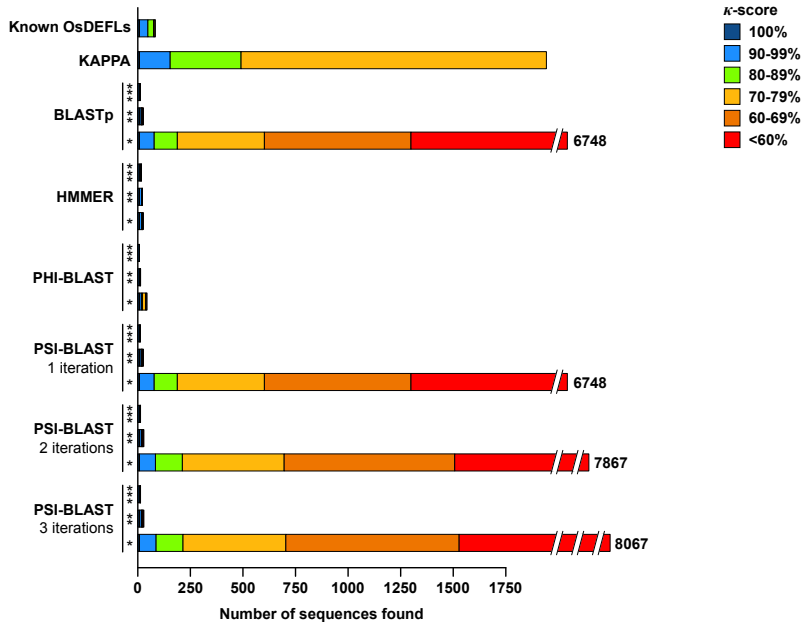


(f) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to known AtDEFLs; red dots represent other proteins.

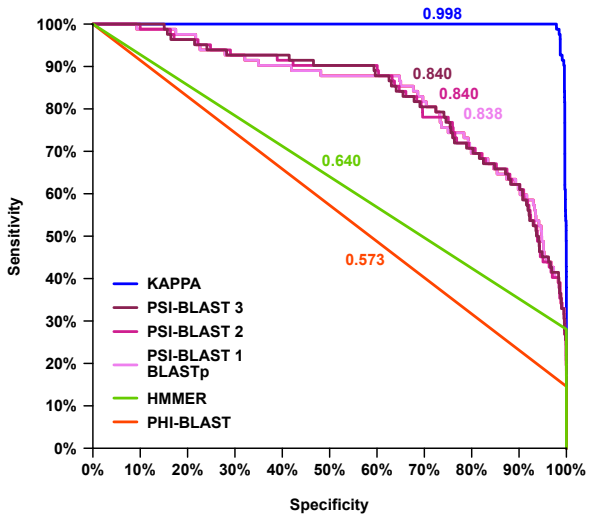


(g) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to known AtDEFLs; red dots represent other proteins.

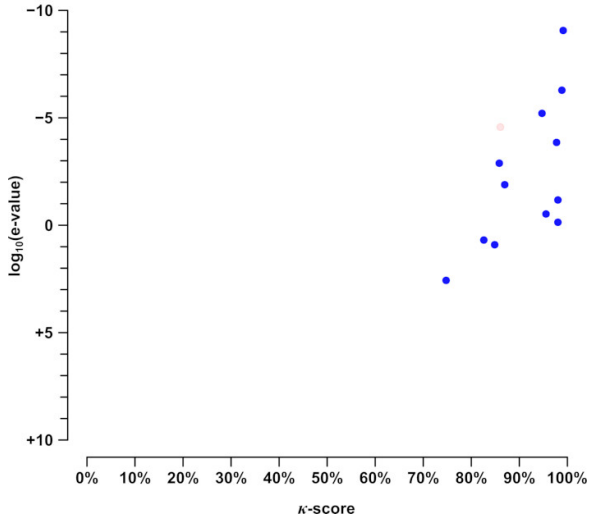
Figure S3.8. (suite)



(a) Comparison of outputs from different programs. *Arabidopsis* defensins were used as input sequences to parse the rice proteome available in Phytozome 9.1. KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of pattern of *Arabidopsis* defensins. Known rice defensin-like proteins (OsDEFLs) described by Silverstein *et al.* (2007) were used to define a minimal expected output.

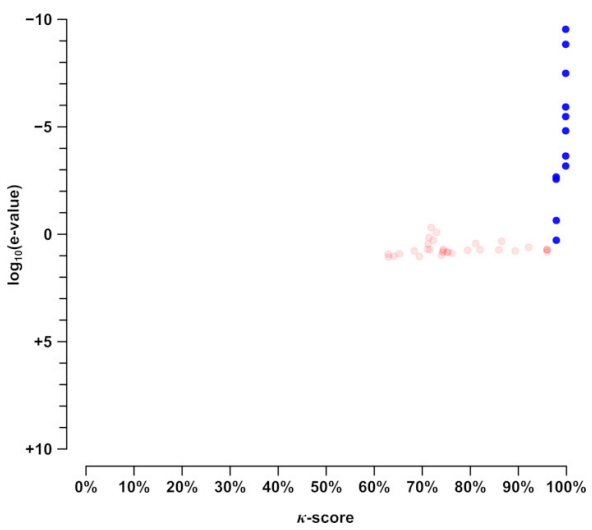


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known OsDEFLs as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

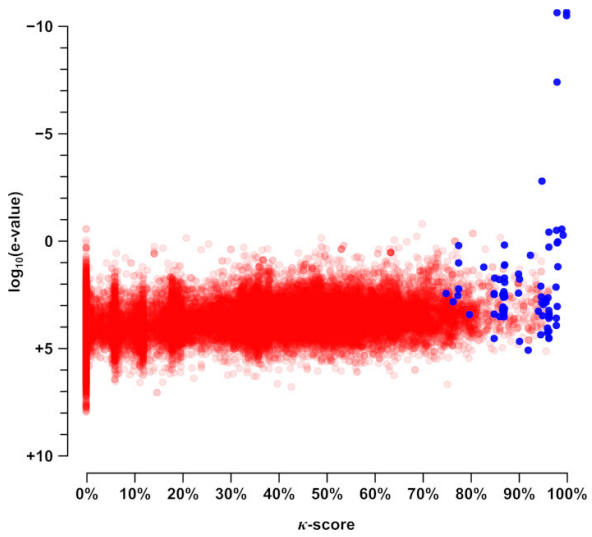


(c) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **HMMER**. Blue dots correspond to known OsDEFLs; red dots represent other proteins.

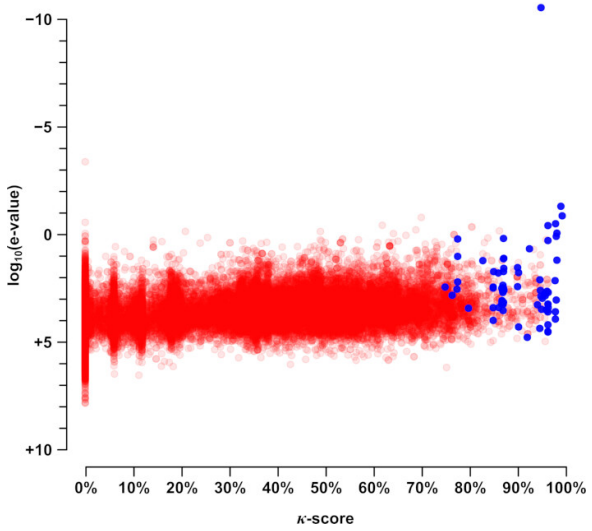
Figure S3.9. Assessment of *ab initio* sequence search performance for KAPPA and other programs: *Arabidopsis* defensins in the rice proteome.



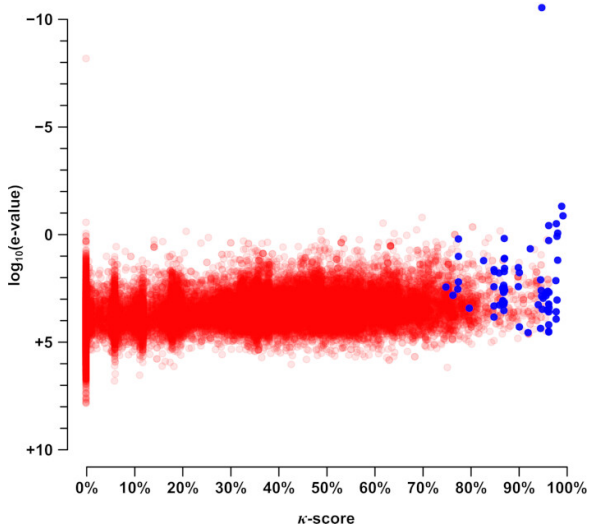
(d) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **PHI-BLAST**. Blue dots correspond to known OsDEFLs; red dots represent other proteins.



(e) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to known OsDEFLs; red dots represent other proteins.

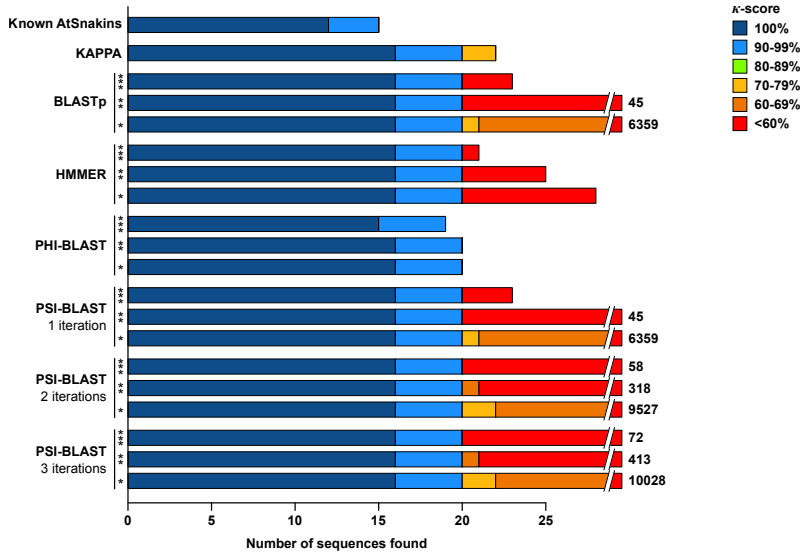


(f) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to known OsDEFLs; red dots represent other proteins.

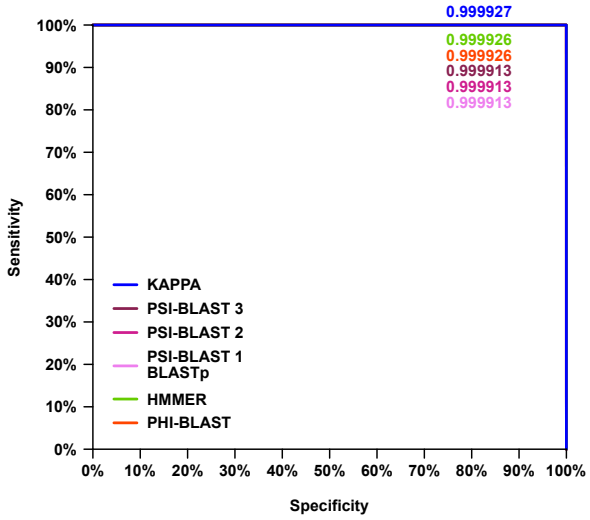


(g) E-value for the best match to a query *Arabidopsis* defensin vs. κ-score according to the consensus pattern of *Arabidopsis* defensins for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to known OsDEFLs; red dots represent other proteins.

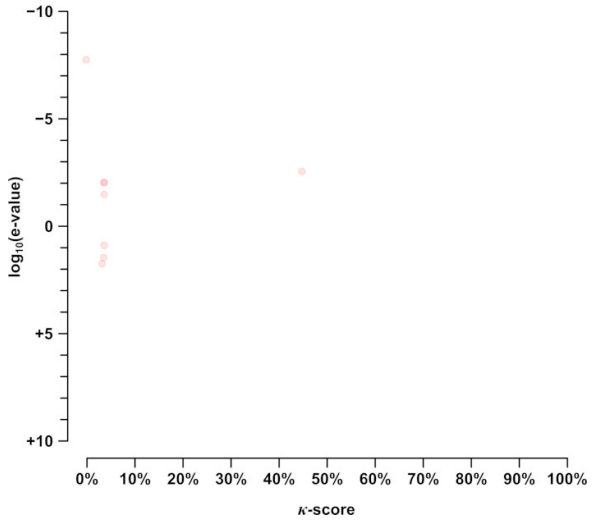
Figure S3.9. (suite)



(a) Comparison of outputs from different programs. *Arabidopsis* snakins were used as input sequences to parse the *A. thaliana* proteome available in Phytozome 9.1. KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of *Arabidopsis* snakins. Known *Arabidopsis* snakins described by Silverstein *et al.* (2007) were used to define a minimal expected output.

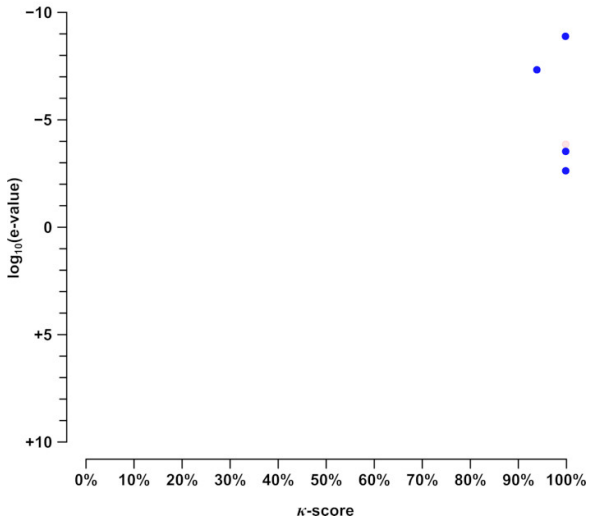


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known snakins as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

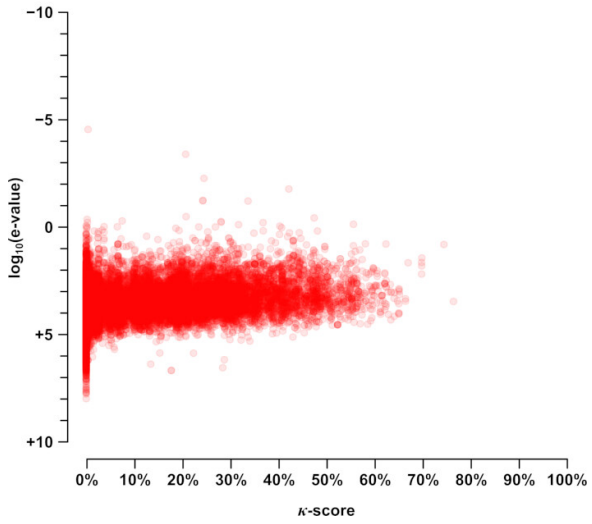


(c) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **HMMER**. Blue dots correspond to known *Arabidopsis* snakins; red dots represent other proteins.

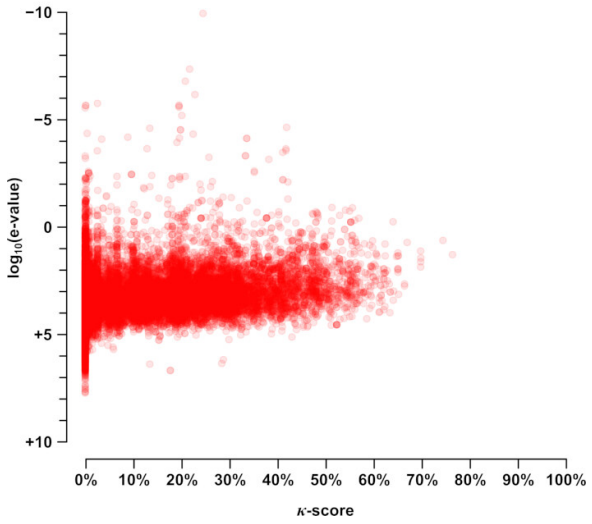
Figure S3.10. Assessment of *ab initio* sequence search performance for KAPPA and other programs: *Arabidopsis* snakins in the *Arabidopsis* proteome.



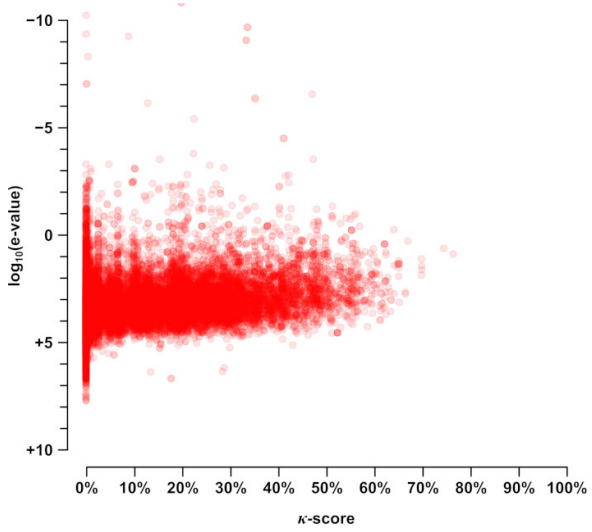
(d) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **PHI-BLAST**. Blue dots correspond to known *Arabidopsis* snakins; red dots represent other proteins.



(e) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to known *Arabidopsis* snakins; red dots represent other proteins.

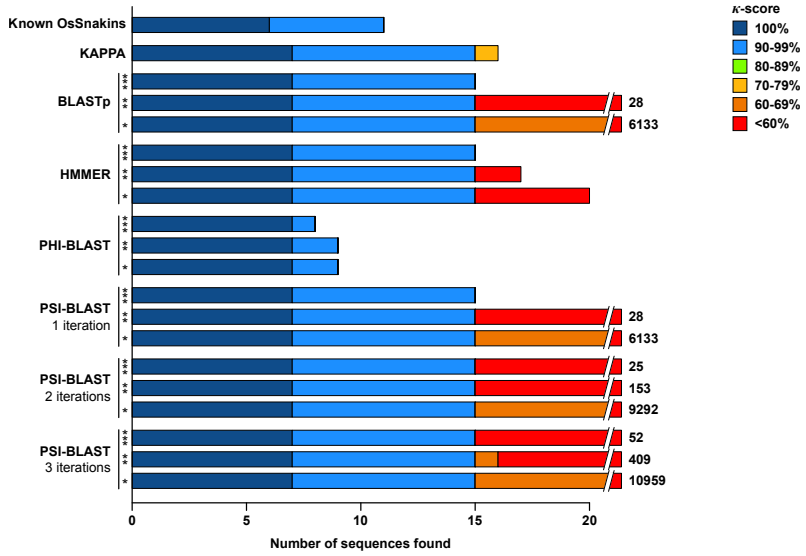


(f) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to known *Arabidopsis* snakins; red dots represent other proteins.

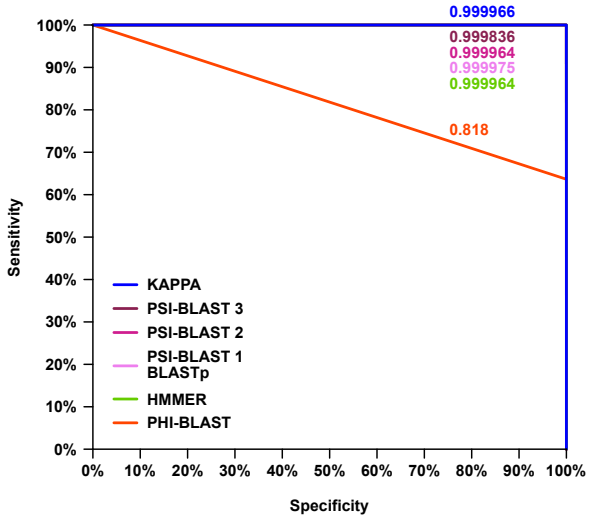


(g) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to known *Arabidopsis* snakins; red dots represent other proteins.

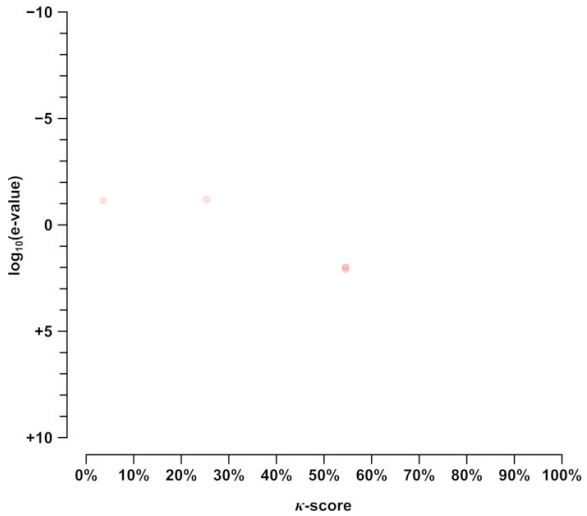
Figure S3.10. (suite)



(a) Comparison of outputs from different programs. *Arabidopsis* snakins were used as input sequences to parse the rice proteome available in Phytozome 9.1. KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of pattern of *Arabidopsis* snakins. Known rice snakins described by Silverstein *et al.* (2007) were used to define a minimal expected output.

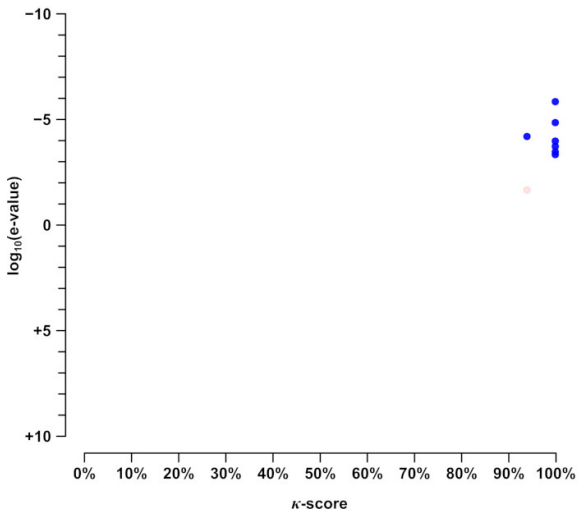


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known snakins as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

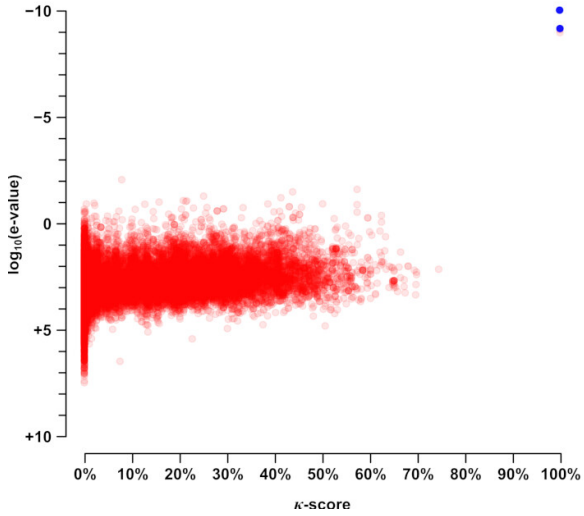


(c) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **HMMER**. Blue dots correspond to known rice snakins; red dots represent other proteins.

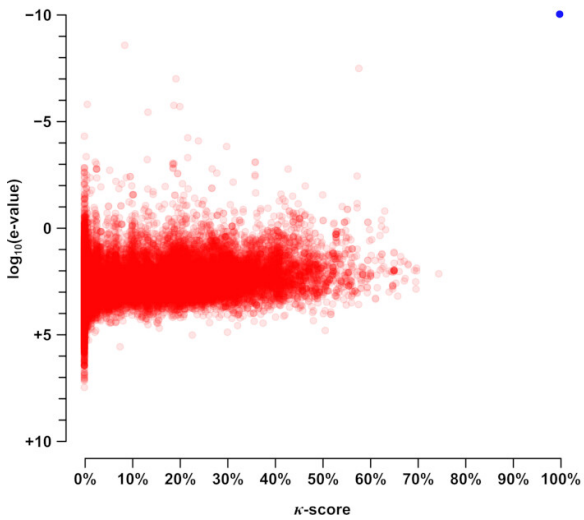
Figure S3.11. Assessment of *ab initio* sequence search performance for KAPPA and other programs: *Arabidopsis* snakins in the rice proteome.



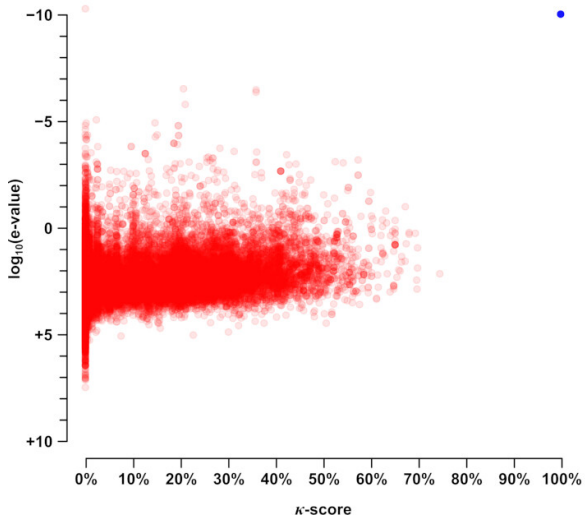
(d) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **PHI-BLAST**. Blue dots correspond to known rice snakins; red dots represent other proteins.



(e) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to known rice snakins; red dots represent other proteins.

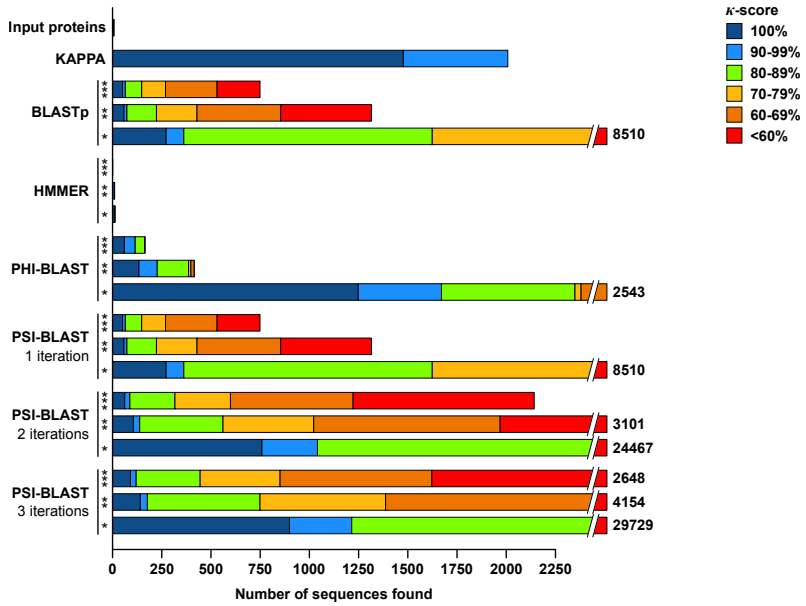


(f) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to known rice snakins; red dots represent other proteins.

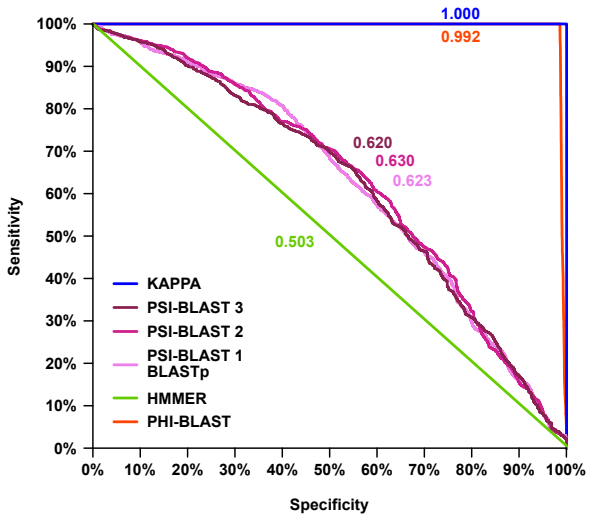


(g) E-value for the best match to a query *Arabidopsis* snakin vs. κ-score according to the consensus pattern of *Arabidopsis* snakins for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to known rice snakins; red dots represent other proteins.

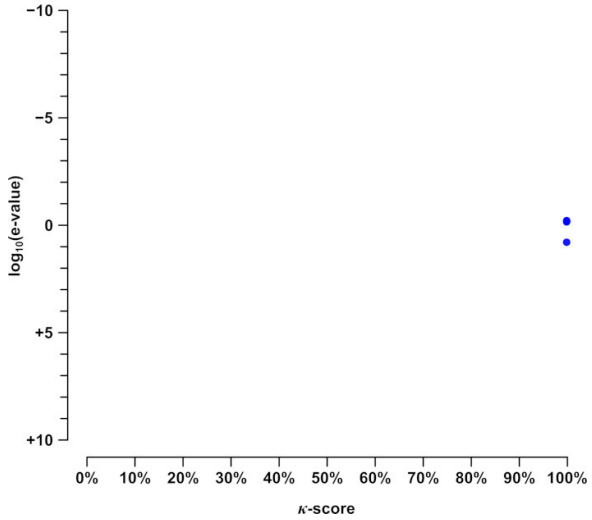
Figure S3.11. (suite)



(a) Comparison of outputs from different programs. Human EGZs described by Kim *et al.* (2005) were used as input sequences to parse the human proteome available from the NCBI website (downloaded on Oct. 29th 2014). KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of HsEGZs. To define a minimal expected output ("predicted HsEGZs"), we built a list of all human proteins containing the EGZ motif by parsing the proteome FASTA file with a regular expression ("regex") approach.

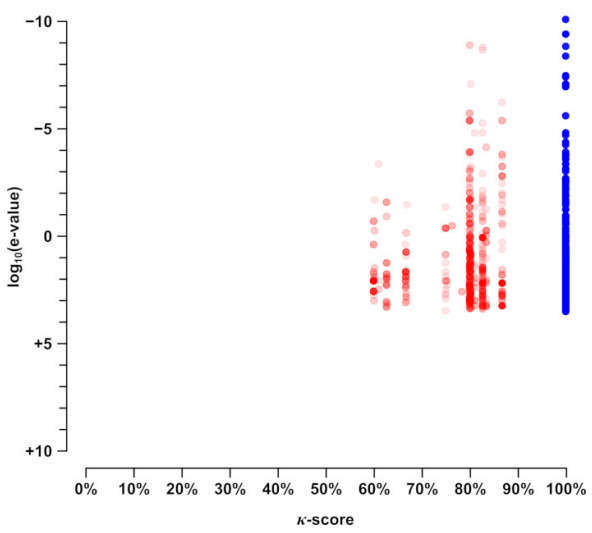


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known HsEGZs as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

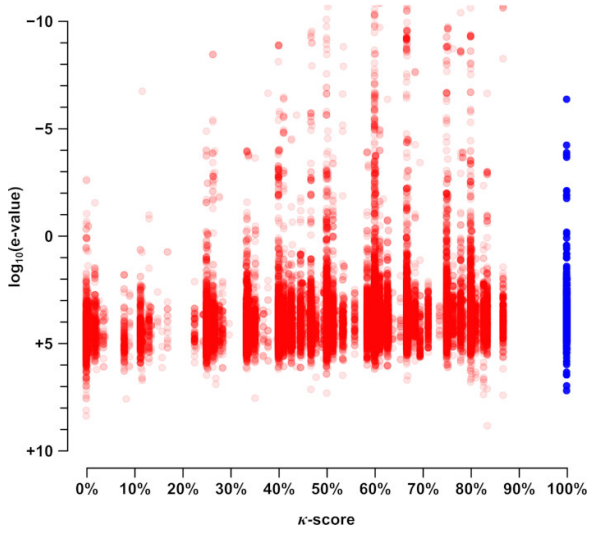


(c) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **HMMER**. Blue dots correspond to predicted HsEGZs; red dots represent other proteins.

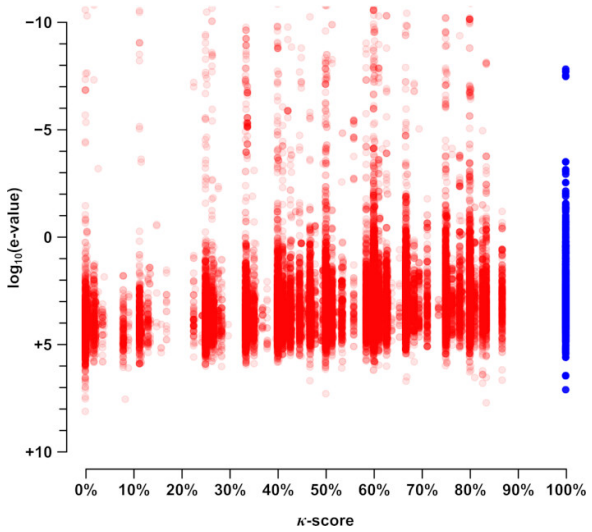
Figure S3.12. Assessment of *ab initio* sequence search performance for KAPPA and other programs: human extended glycine zipper-containing proteins (EGZs) in the human proteome.



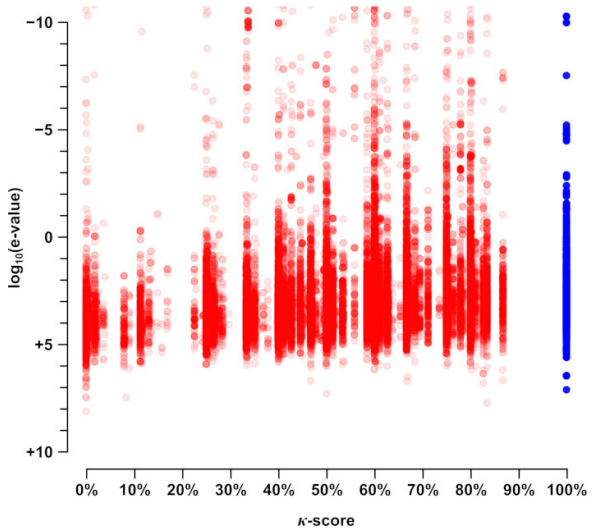
(d) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **PHI-BLAST**. Blue dots correspond to predicted HsEGZs; red dots represent other proteins.



(e) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to predicted HsEGZs; red dots represent other proteins.

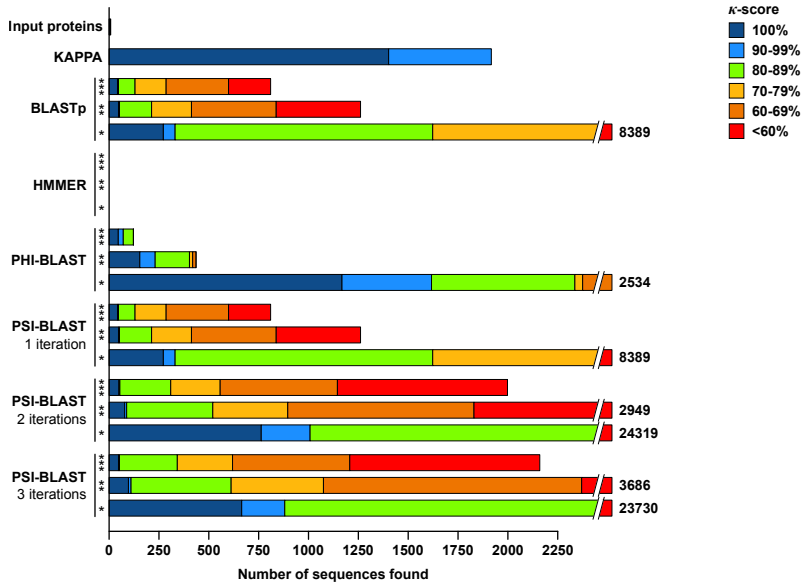


(f) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to predicted HsEGZs; red dots represent other proteins.

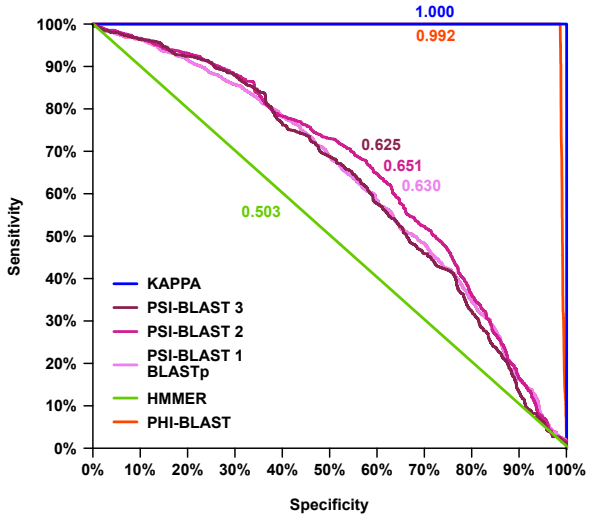


(g) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to predicted HsEGZs; red dots represent other proteins.

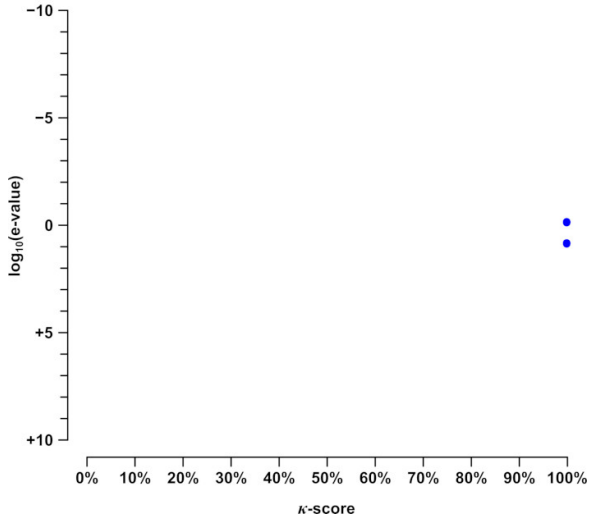
Figure S3.12. (suite)



(a) Comparison of outputs from different programs. Human EGZs described by Kim *et al.* (2005) were used as input sequences to parse the mouse proteome available from the NCBI website (downloaded on Oct. 29th 2014). KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of HsEGZs. To define a minimal expected output ("predicted MmEGZs"), we built a list of all mouse proteins containing the EGZ motif by parsing the proteome FASTA file with a regular expression ("regex") approach.

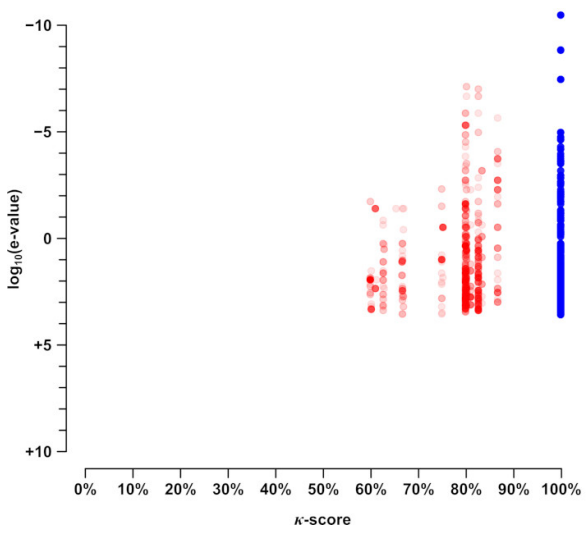


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known MmEGZs as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

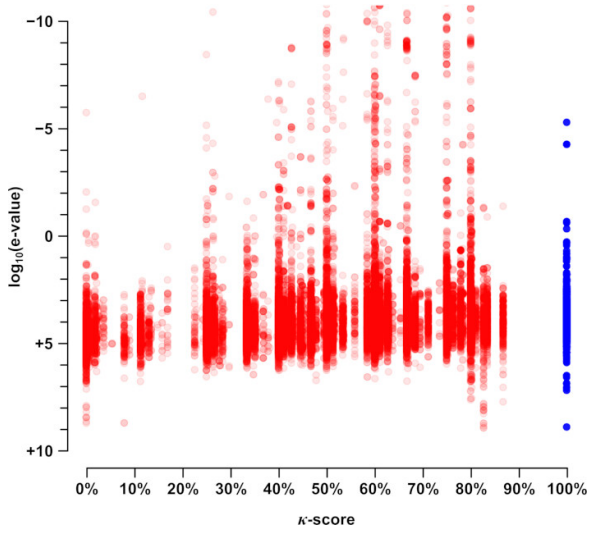


(c) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **HMMER**. Blue dots correspond to predicted MmEGZs; red dots represent other proteins.

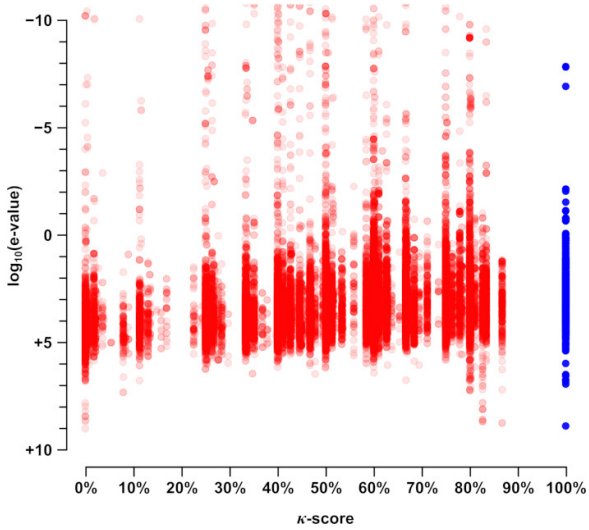
Figure S3.13. Assessment of *ab initio* sequence search performance for KAPPA and other programs: human extended glycine zipper-containing proteins (EGZs) in the mouse proteome.



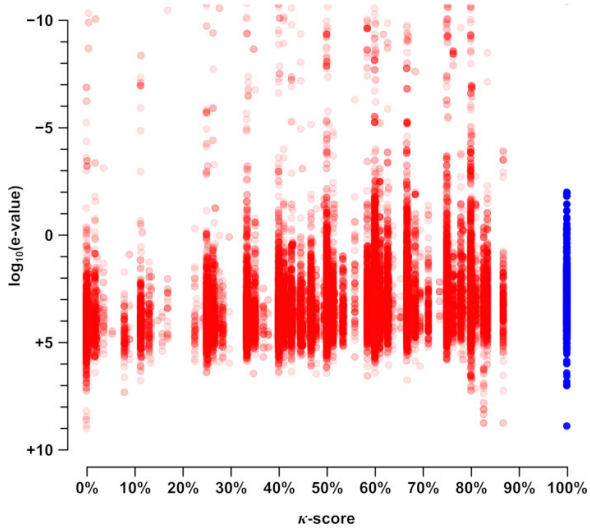
(d) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **PHI-BLAST**. Blue dots correspond to predicted MmEGZs; red dots represent other proteins.



(e) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to predicted MmEGZs; red dots represent other proteins.

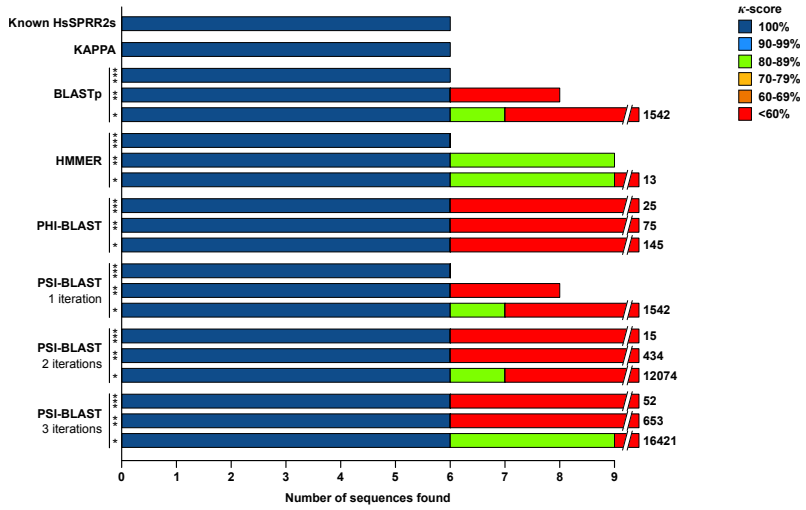


(f) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to predicted MmEGZs; red dots represent other proteins.

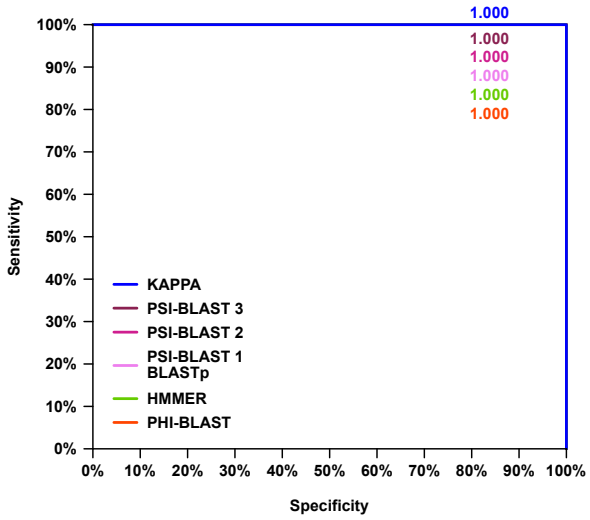


(g) E-value for the best match to a query HsEGZ vs. κ-score according to the consensus pattern of HsEGZs for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to predicted MmEGZs; red dots represent other proteins.

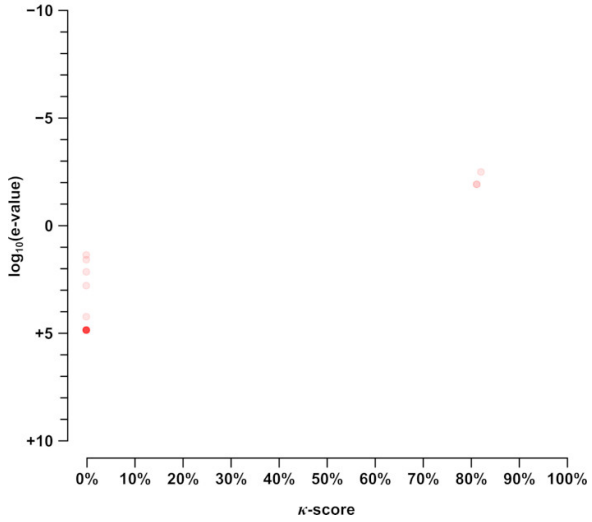
Figure S3.13. (suite)



(a) Comparison of outputs from different programs. Human SPRR2s described by Cabral *et al.* (2001) were used as input sequences to parse the human proteome available from the NCBI website (downloaded on Oct. 29th 2014). KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e-value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of HsSPRR2s.

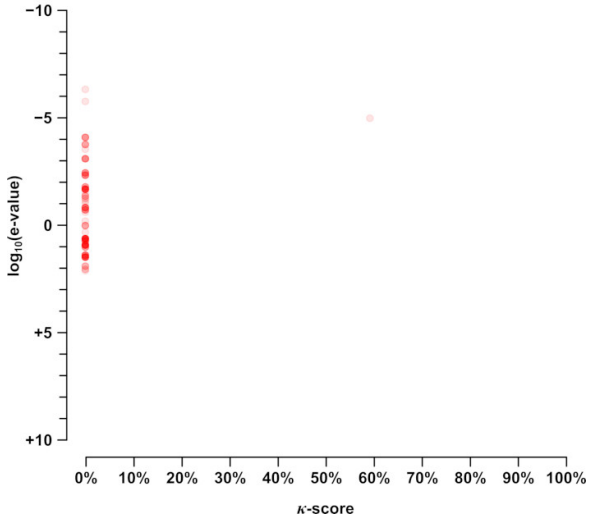


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known HsSPRR2s as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

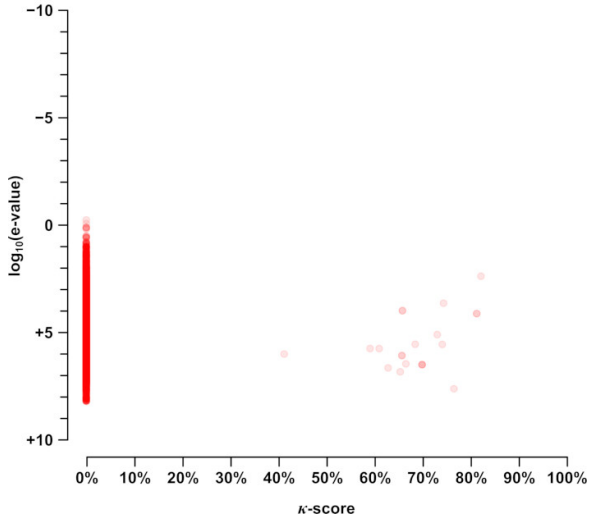


(c) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **HMMER**. Blue dots correspond to known HsSPRR2s; red dots represent other proteins.

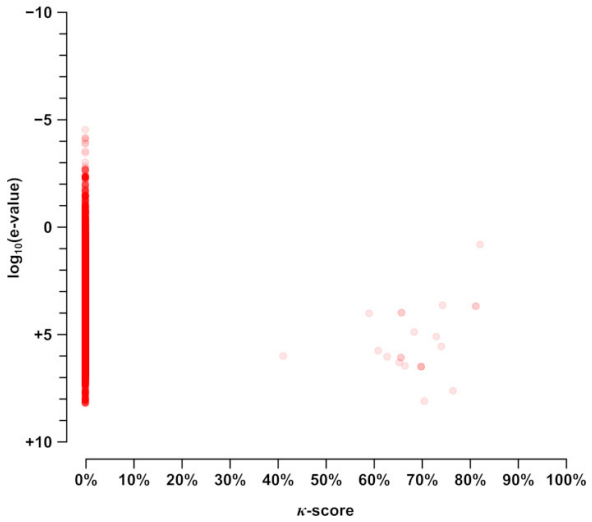
Figure S3.14. Assessment of *ab initio* sequence search performance for KAPPA and other programs: human type 2 small proline-rich proteins (SPRR2s) in the human proteome.



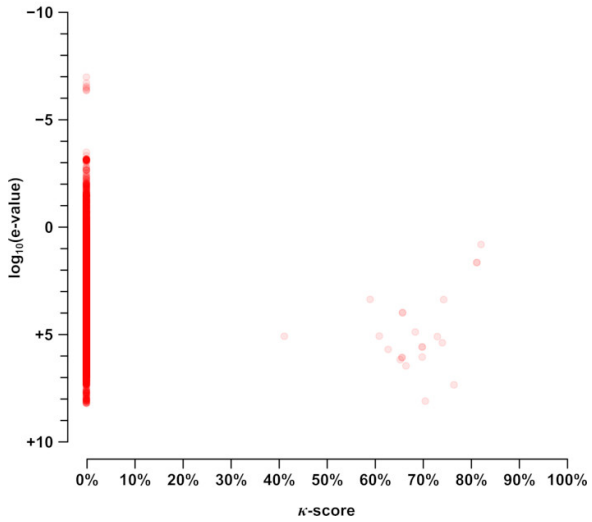
(d) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **PHI-BLAST**. Blue dots correspond to known HsSPRR2s; red dots represent other proteins.



(e) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to known HsSPRR2s; red dots represent other proteins.

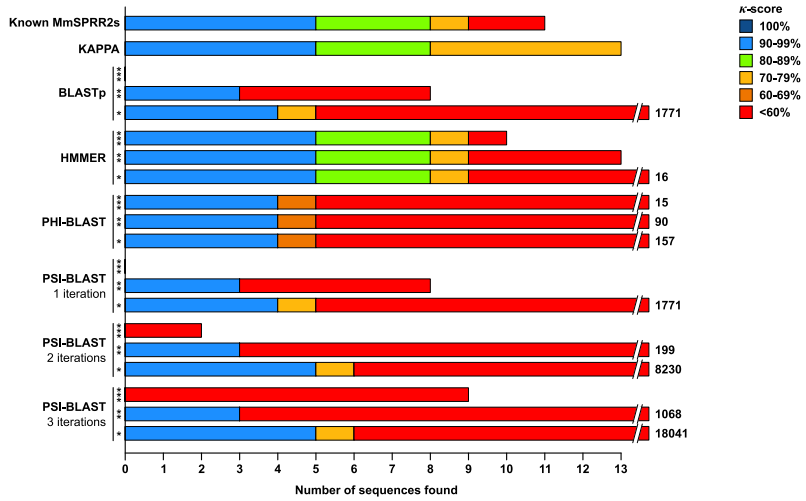


(f) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to known HsSPRR2s; red dots represent other proteins.

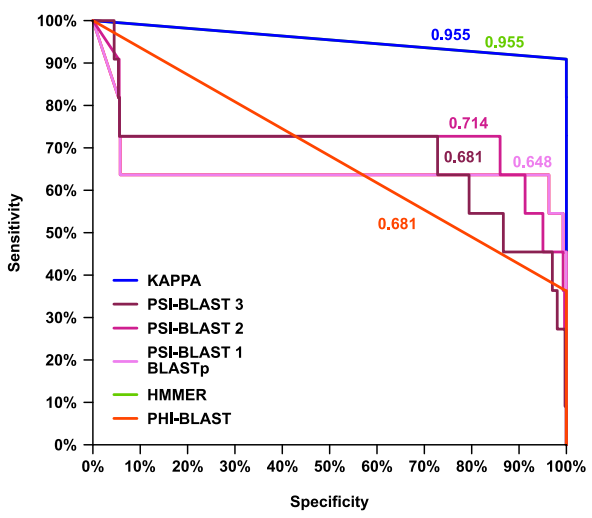


(g) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to known HsSPRR2s; red dots represent other proteins.

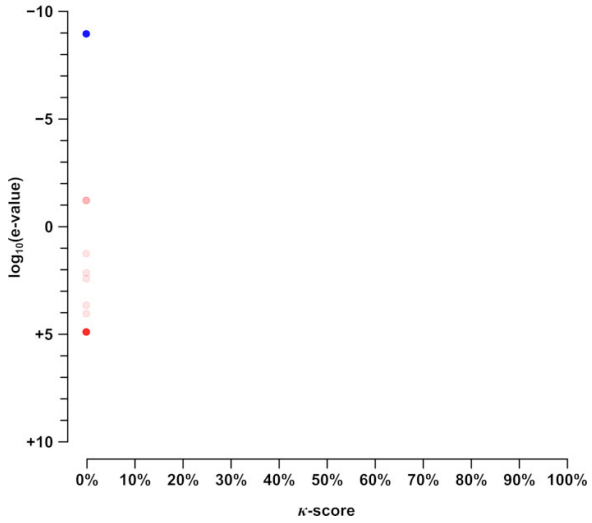
Figure S3.14. (suite)



(a) Comparison of outputs from different programs. Human SPRR2s described by Cabral *et al.* (2001) were used as input sequences to parse the mouse proteome available from the NCBI website (downloaded on Oct. 29th 2014). KAPPA was used with the parameters specified in Table S3.1. Other programs were used with default parameters and three levels of stringency based on e- value: 10^{-3} (***), 1 (**) or 1000 (*). PSI-BLAST was run with 1, 2 and 3 iterations. Colours refer to the κ-score of the output sequences with respect to the consensus pattern of HsSPRR2s. Known MmSPRR2s found in GenBank were used to define a minimal expected output.

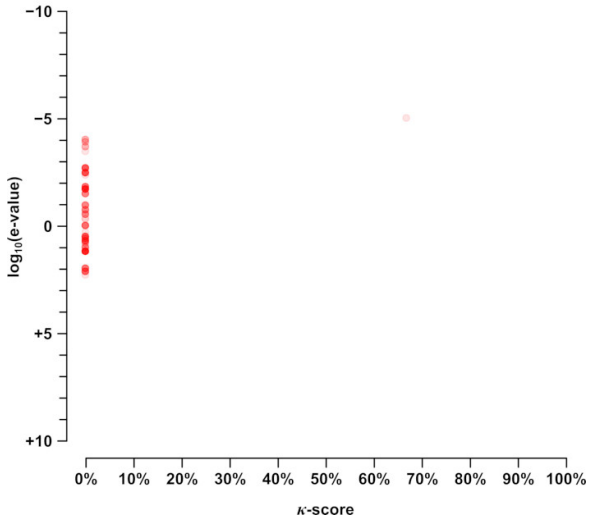


(b) Sensitivity-specificity plot comparing performances of KAPPA and other programs, using known MmSPRR2s as reference true sequences. Sensitivity refers to the true positive rate (TPR) while specificity corresponds to the true negative rate (TNR).

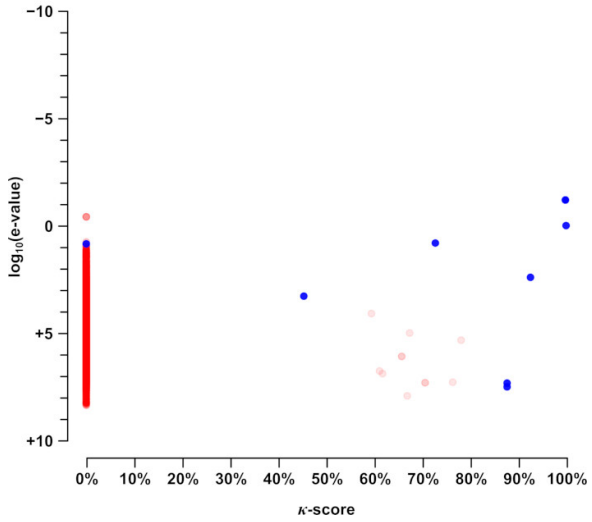


(c) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **HMMER**. Blue dots correspond to known MmSPRR2s; red dots represent other proteins.

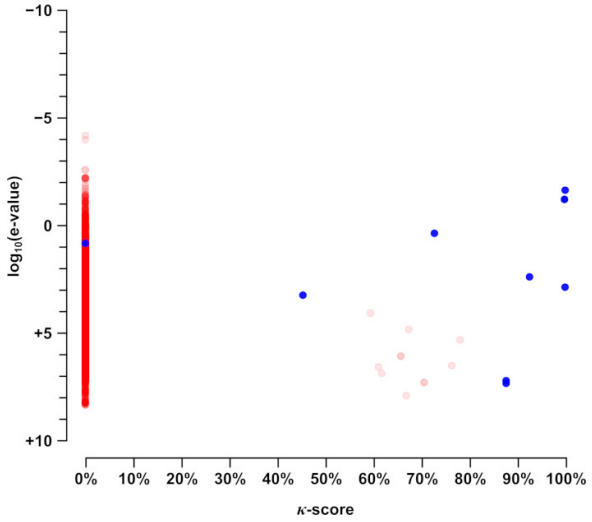
Figure S3.15. Assessment of *ab initio* sequence search performance for KAPPA and other programs: human type 2 small proline-rich proteins (SPRR2s) in the mouse proteome.



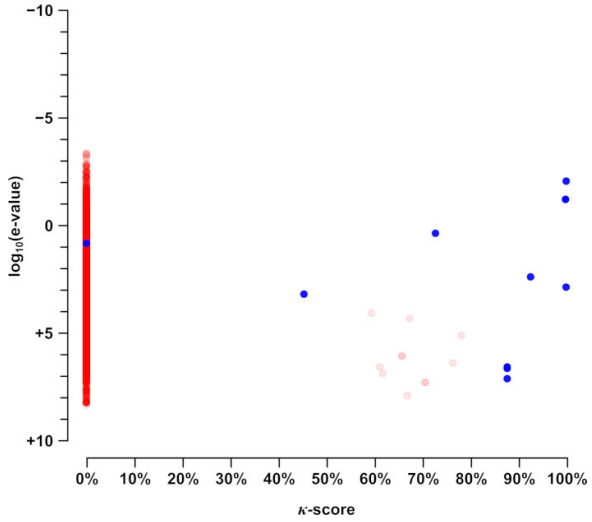
(d) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **PHI-BLAST**. Blue dots correspond to known MmSPRR2s; red dots represent other proteins.



(e) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **BLASTp** (or PSI-BLAST with 1 iteration). Blue dots correspond to known MmSPRR2s; red dots represent other proteins.



(f) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **PSI-BLAST (2 iterations)**. Blue dots correspond to known MmSPRR2s; red dots represent other proteins.



(g) E-value for the best match to a query HsSPRR2 vs. κ-score according to the consensus pattern of HsSPRR2s for sequences found by **PSI-BLAST (3 iterations)**. Blue dots correspond to known MmSPRR2s; red dots represent other proteins.

Figure S3.15. (suite)

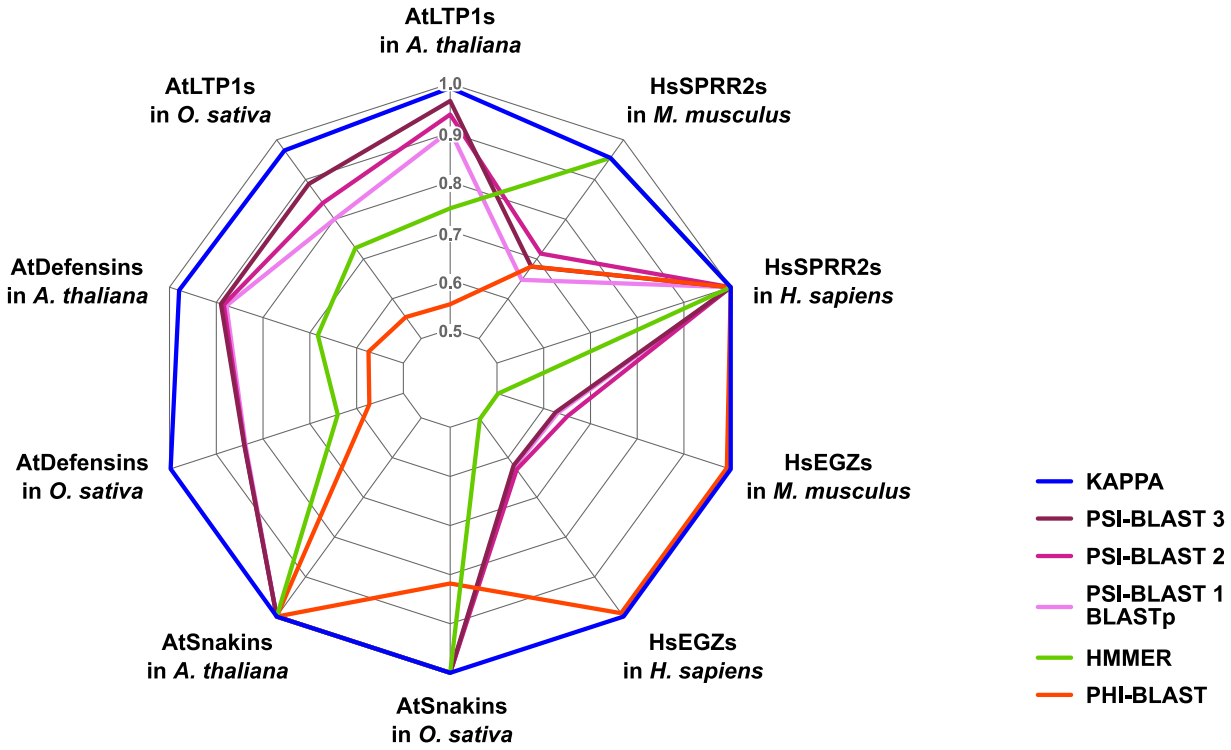
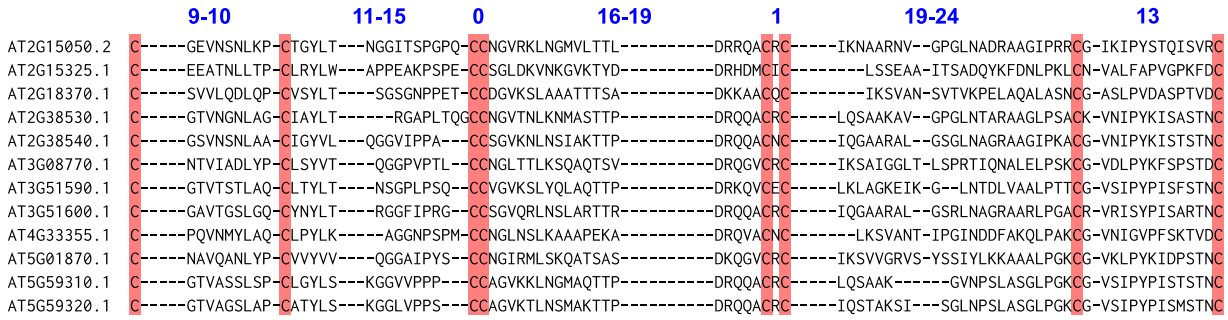
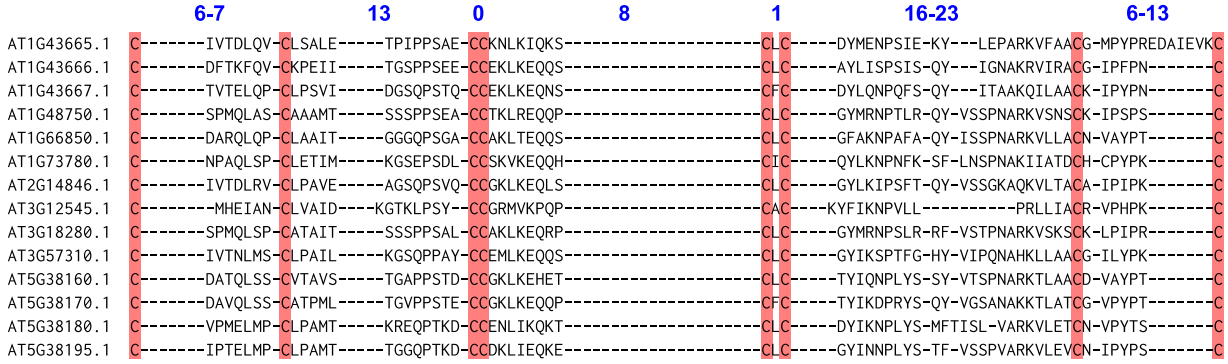


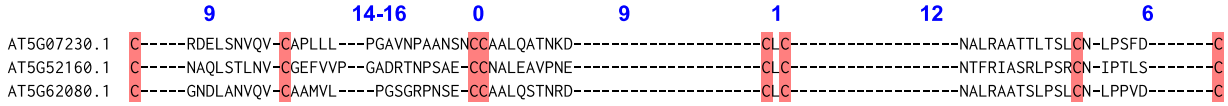
Figure S3.16. Overall sequence search performance of KAPPA and other programs assessed by comparing the area under the specificity/sensitivity ROC curves (AUC) in all *ab initio* sequence search tests performed in this study. Sensitivity corresponds to the true positive rate (TPR) while specificity is equal to the true negative rate (TNR). The AUC integrates both of these indicators to assess detection performance. KAPPA's AUCs were always greater than 0.95 while other programs can give lower AUCs, especially when proteins families with high sequence divergence or small modifications of the key aminoacid spacing are studied.



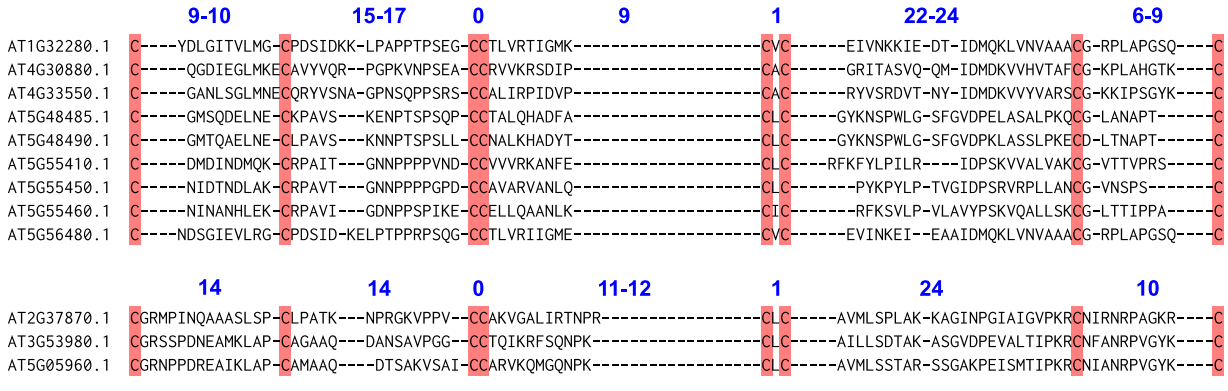
(a) Edstam's AtLTP1 → KAPPA Cluster 03



(b) Edstam's AtLTP2 → KAPPA Cluster 02

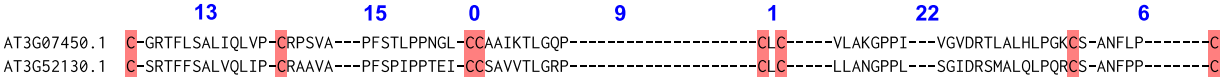


(c) Edstam's AtLTPc → KAPPA Cluster 09

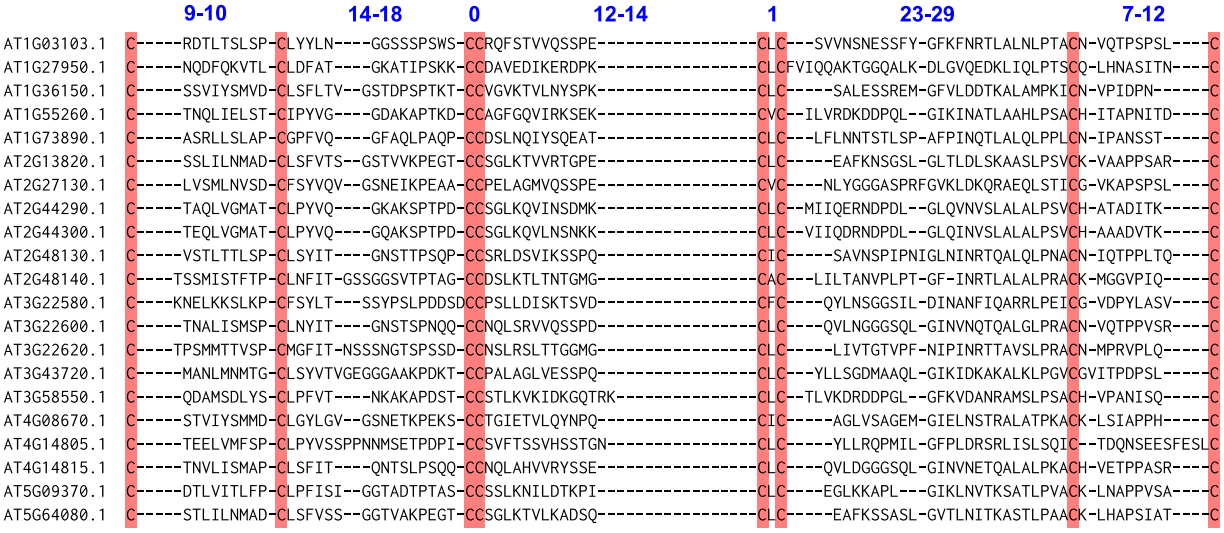


(d) Edstam's AtLTPd → KAPPA Clusters 04 and 08

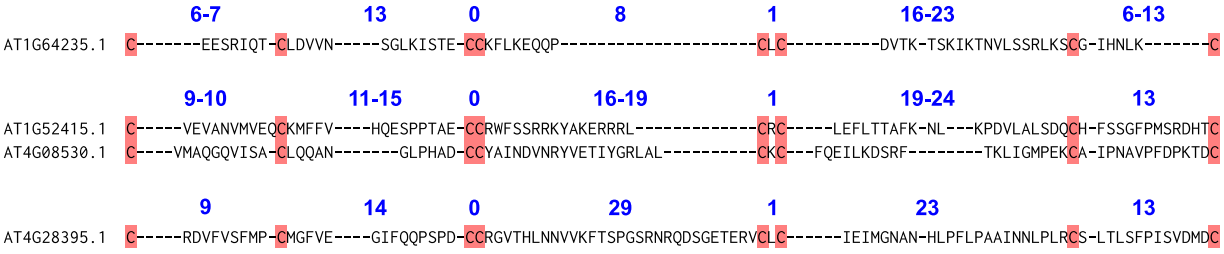
Figure S3.17. Correspondance between clusters formed by Edstam *et al.* (2011) and KAPPA using the 79 LTPs found by Edstam *et al.* (2011). Alignments were generated with MUSCLE. Residues before the first characteristic cysteine and after the last one were discarded. Cysteine spacing is indicated in blue. Edstam's AtLTPg AT1G05450.1 was excluded from KAPPA clustering because it did not contain any cysteine residue.



(e) Edstam's AtLTPe → KAPPA Cluster 16



(f) Edstam's AtLTPg → KAPPA Clusters 01, 05, 06 and 1 singleton.



(g) Edstam's AtLTPx → KAPPA Clusters 02, 03 and 1 singleton

Figure S3.17. (suite)

Table S3.1. Parameters used for KAPPA *ab initio* sequence search tests performed in this study.

(a) KAPPA command-line parameters

Figure	Query proteins	Key aa	Target proteome	KAPPA parameters													
				k	m	M	l	L	n	c	kg	kl	i	I	C	P	S
3.3, S3.4	*A. thaliana* LTP1s	Cys	*A. thaliana*	C	6	None	0	350	1	1	0	0	True	7	7	7	70
S3.5			*O. sativa*	C	6	None	0	350	1	1	0	0	True	7	7	7	70
S3.6	*A. thaliana* defensins	Cys	*A. thaliana*	C	8	None	0	250	1	1	1	1	True	7	7	7	70
S3.7			*O. sativa*	C	8	None	0	250	1	1	1	1	True	7	7	7	70
S3.8	*A. thaliana* snakins	Cys	*A. thaliana*	C	8	None	0	250	1	1	0	0	True	7	7	7	70
S3.9			*O. sativa*	C	8	None	0	250	1	1	0	0	True	7	7	7	70
S3.10	*H. sapiens* EGZs	Gly	*H. sapiens*	G	4	None	0	None	1	1	2	0	True	9	9	1	90
S3.11			*M. musculus*	G	4	None	0	None	1	1	2	0	True	9	9	1	90
S3.12	*H. sapiens* SPRR2s	Pro	*H. sapiens*	P	25	None	0	120	1	1	0	0	True	7	7	7	90
S3.13			*M. musculus*	P	25	None	0	120	1	1	0	0	True	7	7	7	90

(b) KAPPA query patterns

Figure	Query proteins	KAPPA query pattern
3.3, S3.4, S3.5	*A. thaliana* LTP1s	`X{25-30}CX{9}CX{13-15}CCX{19}CXCX{21-24}CX{13}CX{4-5}`
S3.6, S3.7	*A. thaliana* defensins	`X{30-32}CX{10}CX{5}CX{3}CX{9-10}CX{6-8}CXCXXXCX{0-7}`
S3.8, S3.9	*A. thaliana* snakins	`X{29-60}CX{3}CX{3}CX{8}CX{3}CXXCCXXCXCX{11}CXCX{12}CX`
S3.10, S3.11	*H. sapiens* EGZs[a]	`X{n}GXXXGXXXGXXXGX{n}`
S3.12, S3.13	*H. sapiens* SPRR2s[b]	`X{10}PXXPPPXXPXPXXPXPXPX{n}`

[a]The glycine pattern used for these tests corresponds to the EGZ motif only.
[b]The proline pattern used for these tests corresponds to the N-term region of SPRR2s only (up to the 9th proline) which is specific to SPRR2s.

Table S3.2. Sequence identifiers of proteins detected by KAPPA in all *ab initio* sequence search tests performed in this study. For each test, input proteins were used to build a key amino acid pattern used to parse a target proteome. *A. thaliana* and *O. sativa* proteomes were downloaded from Phytozome 9.1. *H. sapiens* and *M. musculus* were downloaded from NCBI.

(a) *A. thaliana* LTPs against the *A. thaliana* proteome → Figure S3.4

Input proteins: 12 known AtLTP1s

AT2G15050.2, AT2G15325.1, AT2G18370.1, AT2G38530.1, AT2G38540.1, AT3G08770.1, AT3G51590.1, AT3G51600.1, AT4G33355.1, AT5G01870.1, AT5G59310.1, AT5G59320.1

Output proteins: 125 known AtLTPs

AT1G03103.1, AT1G05450.2, AT1G07747.1, AT1G12090.1, AT1G12100.1, AT1G18280.1, AT1G27950.1, AT1G32280.1, AT1G36150.1, AT1G43665.1, AT1G43666.1, AT1G43667.1, AT1G48750.1, AT1G52415.1, AT1G55260.1, AT1G62500.1, AT1G62510.1, AT1G62790.1, AT1G64235.1, AT1G66850.1, AT1G70250.1, AT1G73550.1, AT1G73550.2, AT1G73560.1, AT1G73780.1, AT1G73890.1, AT1G76965.1, AT2G10940.2, AT2G13295.1, AT2G13820.1, AT2G14846.1, AT2G15050.1, AT2G15050.2, AT2G15325.1, AT2G16592.1, AT2G16594.1, AT2G18370.1, AT2G27130.1, AT2G37870.1, AT2G38530.1, AT2G38540.1, AT2G44290.1, AT2G44300.1, AT2G45180.1, AT2G48130.1, AT2G48140.1, AT3G07450.1, AT3G08770.1, AT3G11825.1, AT3G12545.1, AT3G18280.1, AT3G22120.1, AT3G22142.1, AT3G22570.1, AT3G22580.1, AT3G22600.1, AT3G22620.1, AT3G29152.1, AT3G43720.1, AT3G51590.1, AT3G51600.1, AT3G52130.1, AT3G53980.1, AT3G53980.2, AT3G57310.1, AT3G58550.1, AT3G59455.1, AT4G00165.2, AT4G08530.1, AT4G08670.1, AT4G12360.1, AT4G12470.1, AT4G12480.1, AT4G12490.1, AT4G12500.1, AT4G12510.1, AT4G12530.1, AT4G12545.1, AT4G12550.1, AT4G12825.1, AT4G14805.1, AT4G14815.1, AT4G15160.1, AT4G22460.1, AT4G22470.1, AT4G22485.1, AT4G22490.1, AT4G22505.1, AT4G22513.1, AT4G22517.1, AT4G22520.1, AT4G22610.1, AT4G22630.1, AT4G22640.1, AT4G22650.1, AT4G22666.1, AT4G28395.1, AT4G30880.1, AT4G33355.1, AT4G33550.1, AT4G33550.2, AT5G01870.1, AT5G05960.1, AT5G07230.1, AT5G09370.1, AT5G13900.1, AT5G38160.1, AT5G38170.1, AT5G38180.1, AT5G38195.1, AT5G44265.1, AT5G46890.1, AT5G46900.1, AT5G48485.1, AT5G48490.1, AT5G52160.1, AT5G55410.1, AT5G55450.1, AT5G55460.1, AT5G56480.1, AT5G59310.1, AT5G59320.1, AT5G62065.1, AT5G62080.1, AT5G64080.1

Output proteins: 85 new AtLTPs

AT1G05410.1, AT1G05410.2, AT1G08610.1, AT1G10610.1, AT1G16650.1, AT1G18480.1, AT1G22060.1, AT1G24580.1, AT1G32600.1, AT1G55260.2, AT1G60700.1, AT1G61900.1, AT1G61900.2, AT1G61900.3, AT1G62530.1, AT1G62530.2, AT1G62790.2, AT1G64255.1, AT1G64580.1, AT1G67856.1, AT1G77250.1, AT1G79080.1, AT2G10940.1, AT2G13820.2, AT2G14080.1, AT2G15050.3, AT2G30700.1, AT2G39620.1, AT2G42620.1, AT2G45700.1, AT2G46850.1, AT2G48140.2, AT3G03580.1, AT3G08770.2, AT3G10790.1, AT3G18090.1, AT3G23780.1, AT3G23780.2, AT3G43720.2, AT3G46820.1, AT3G47860.1, AT3G51570.1, AT3G51730.1, AT3G61170.1, AT3G63370.1, AT4G00165.1, AT4G02100.1, AT4G09680.1, AT4G09680.2, AT4G10600.1, AT4G11270.1, AT4G12520.1, AT4G15160.2, AT4G16860.1, AT4G16890.1, AT4G18593.1, AT4G19890.1, AT4G22666.2, AT4G22970.1, AT4G22970.2, AT4G26350.1, AT4G33210.1, AT4G33355.2, AT4G36650.1, AT4G36650.2, AT4G39940.1, AT5G01720.1, AT5G01800.1, AT5G09370.2, AT5G12210.1, AT5G12210.2, AT5G20885.1, AT5G23340.1, AT5G25350.1, AT5G27920.1, AT5G46470.1, AT5G51370.1, AT5G51370.2, AT5G51380.1, AT5G54062.1, AT5G54320.1, AT5G55410.2, AT5G55840.1, AT5G56380.1, AT5G64080.2

(b) *A. thaliana* LTPs in the *O. sativa* proteome → Figure S3.7

Input proteins: 12 known AtLTP1s

AT2G15050.2, AT2G15325.1, AT2G18370.1, AT2G38530.1, AT2G38540.1, AT3G08770.1, AT3G51590.1, AT3G51600.1, AT4G33355.1, AT5G01870.1, AT5G59310.1, AT5G59320.1

Output proteins: 129 known OsLTPs

Os01g12020.1, Os01g42210.1, Os01g49640.1, Os01g49650.1, Os01g58650.1, Os01g58660.1, Os01g59870.1, Os01g60740.1, Os01g62980.1, Os01g68580.1, Os01g68589.1, Os02g44310.1, Os02g44320.1, Os03g01300.1, Os03g02050.1, Os03g07100.1, Os03g09230.1, Os03g14630.1, Os03g14642.1, Os03g14654.1, Os03g20760.1, Os03g25350.1, Os03g26800.1, Os03g26820.1, Os03g43050.1, Os03g44000.1, Os03g45150.1, Os03g46110.1, Os03g46150.1, Os03g46180.1, Os03g50960.1, Os03g57970.2, Os03g57980.1, Os03g57990.1, Os03g58670.1, Os03g58940.1, Os03g59380.1, Os04g33920.1, Os04g33930.2, Os04g38840.1, Os04g46810.1, Os04g46820.1, Os04g46830.1, Os04g52250.1, Os04g52260.1, Os04g55170.1, Os05g06780.1, Os05g40010.1, Os05g41030.1, Os05g47700.1, Os05g47730.1, Os06g01580.1, Os06g06340.1, Os06g07220.1, Os06g12440.1, Os06g34840.1, Os06g47200.1, Os06g49190.1, Os06g49770.1, Os06g49770.2, Os07g07790.1, Os07g07860.1, Os07g07870.1, Os07g07920.1, Os07g07930.1, Os07g09970.1, Os07g11310.1, Os07g11630.1, Os07g11650.1, Os07g12080.1, Os07g14140.1, Os07g18750.1, Os07g18990.1, Os07g27940.1, Os07g30590.1, Os07g37045.1, Os07g39640.1, Os07g43290.1, Os08g03690.1, Os08g35665.1, Os08g42040.1, Os08g43240.1, Os08g43290.1, Os09g35700.1, Os10g05720.1, Os10g05720.2, Os10g09920.1, Os10g11730.1, Os10g11750.1, Os10g20830.1, Os10g20840.1, Os10g20860.1, Os10g20880.1, Os10g20890.1, Os10g36070.1, Os10g36090.1, Os10g36100.1, Os10g36110.1, Os10g36160.1, Os10g36170.1, Os10g40420.1, Os10g40430.1, Os10g40440.1, Os10g40460.1, Os10g40470.1, Os10g40480.1, Os10g40510.1, Os10g40520.1, Os10g40530.1, Os10g40614.1, Os11g02350.1, Os11g02369.1, Os11g02389.1, Os11g02400.1, Os11g02424.2, Os11g03870.1, Os11g24070.1, Os11g29420.1, Os11g34660.1, Os11g37280.1, Os11g37320.1, Os11g40530.1, Os12g02105.1, Os12g02310.1, Os12g02320.1, Os12g02330.2, Os12g02340.1, Os12g28880.1, Os12g29040.1

Output proteins: 43 new OsLTPs

Os01g07250.1, Os01g16210.1, Os01g62720.1, Os01g62980.2, Os01g70920.3, Os02g02480.2, Os02g56340.1, Os03g05225.1, Os03g09230.2, Os03g18779.1, Os03g46150.2, Os03g57970.1, Os04g33930.1, Os04g40350.1, Os04g40650.1, Os04g40650.2, Os04g53020.1, Os04g55580.1, Os05g32010.1, Os06g47200.2, Os06g47200.3, Os07g11320.1, Os07g11330.1, Os07g11360.1, Os07g11380.1, Os07g11380.2, Os07g11410.1, Os07g11510.1, Os07g43290.2, Os08g42040.2, Os10g32860.2, Os11g02150.1, Os11g02150.2, Os11g02400.2, Os11g18990.1, Os12g02094.1, Os12g02300.1, Os12g02310.2, Os12g04790.1, Os12g04920.1, Os12g14450.1, Os12g19090.1, Os12g38510.1

Table S3.2. (suite)

(c) *A. thaliana* defensins against the *A. thaliana* proteome → Figure S3.8

Input proteins: 18 known *Arabidopsis* defensins

AT1G19610.1, AT1G55010.1, AT1G61070.1, AT1G75830.1, AT2G02100.1, AT2G02120.1, AT2G02130.1, AT2G02140.1, AT2G02147.1, AT2G19893.1, AT2G26010.1, AT2G26020.1, AT2G31953.1, AT2G31957.1, AT3G63360.1, AT5G44420.1, AT5G44430.1, AT5G63660.1

Output proteins: 227 known AtDEFLs

AT1G13605.1, AT1G13607.1, AT1G13755.1, AT1G14755.1, AT1G15757.1, AT1G19610.1, AT1G28335.1, AT1G31772.1, AT1G32763.1, AT1G33607.1, AT1G34047.1, AT1G35435.1, AT1G35537.1, AT1G47317.1, AT1G47540.1, AT1G49435.1, AT1G49715.1, AT1G54445.1, AT1G55010.1, AT1G56233.1, AT1G56553.1, AT1G58055.1, AT1G59833.1, AT1G60983.1, AT1G61070.1, AT1G61688.1, AT1G63522.1, AT1G63535.1, AT1G64107.1, AT1G64195.1, AT1G65352.1, AT1G68905.1, AT1G68907.1, AT1G69818.1, AT1G69828.1, AT1G73603.1, AT1G73607.1, AT1G75830.1, AT2G02100.1, AT2G02120.1, AT2G02130.1, AT2G02140.1, AT2G02147.1, AT2G03913.1, AT2G03931.1, AT2G03933.1, AT2G03936.1, AT2G03937.2, AT2G03955.1, AT2G04034.1, AT2G04045.1, AT2G04046.1, AT2G04425.1, AT2G04925.1, AT2G05117.1, AT2G05335.1, AT2G06166.1, AT2G10535.1, AT2G12465.1, AT2G12475.1, AT2G13542.1, AT2G14365.1, AT2G14935.1, AT2G15535.1, AT2G19893.1, AT2G20070.1, AT2G20208.1, AT2G20463.1, AT2G20465.1, AT2G21465.1, AT2G21725.1, AT2G22121.1, AT2G22345.1, AT2G22805.1, AT2G22807.1, AT2G22941.1, AT2G24615.1, AT2G24625.1, AT2G25185.1, AT2G25295.1, AT2G25305.1, AT2G25344.1, AT2G26010.1, AT2G26020.1, AT2G27145.1, AT2G28355.1, AT2G28405.1, AT2G29045.1, AT2G31953.1, AT2G31957.1, AT2G33233.1, AT2G36255.1, AT2G40995.1, AT2G41997.1, AT2G42885.1, AT2G43510.1, AT2G43520.1, AT2G43530.1, AT2G43535.1, AT2G43550.1, AT3G04540.1, AT3G04545.1, AT3G04903.1, AT3G04943.1, AT3G04945.1, AT3G05727.1, AT3G05730.1, AT3G06985.1, AT3G07005.1, AT3G10195.1, AT3G16895.1, AT3G17155.1, AT3G20993.1, AT3G20997.1, AT3G23167.1, AT3G23715.1, AT3G23727.1, AT3G25265.1, AT3G27831.1, AT3G27835.1, AT3G42473.1, AT3G43083.1, AT3G43505.1, AT3G48231.1, AT3G50925.1, AT3G59930.1, AT3G61172.1, AT3G61175.1, AT3G61177.1, AT3G61182.1, AT3G63360.1, AT4G08028.1, AT4G09153.1, AT4G09647.1, AT4G09795.1, AT4G09984.1, AT4G10595.1, AT4G10603.1, AT4G11485.1, AT4G11760.1, AT4G13095.1, AT4G13968.1, AT4G14276.1, AT4G14785.1, AT4G15733.1, AT4G15735.1, AT4G17713.1, AT4G17718.1, AT4G18823.1, AT4G19035.1, AT4G19038.1, AT4G19905.1, AT4G22115.1, AT4G22210.1, AT4G22212.1, AT4G22214.1, AT4G22217.1, AT4G22230.1, AT4G22235.1, AT4G29033.1, AT4G29273.1, AT4G29280.1, AT4G29283.1, AT4G29285.1, AT4G29290.1, AT4G29300.1, AT4G29305.1, AT4G30064.1, AT4G30067.1, AT4G30070.1, AT4G30074.1, AT4G32714.1, AT4G32717.1, AT4G33465.1, AT4G39917.1, AT5G04045.1, AT5G08055.1, AT5G08315.1, AT5G08505.1, AT5G16453.1, AT5G18403.1, AT5G19172.1, AT5G19175.1, AT5G19315.1, AT5G23035.1, AT5G23212.1, AT5G27495.1, AT5G28288.1, AT5G32619.1, AT5G33355.1, AT5G37473.1, AT5G37474.1, AT5G38317.1, AT5G38330.1, AT5G39365.1, AT5G40155.1, AT5G42223.1, AT5G42232.1, AT5G42235.1, AT5G42242.1, AT5G42797.1, AT5G43518.1, AT5G44420.1, AT5G44430.1, AT5G46873.1, AT5G46877.1, AT5G47075.1, AT5G47077.1, AT5G47175.1, AT5G48543.1, AT5G48905.1, AT5G48945.1, AT5G48953.1, AT5G51845.1, AT5G52605.1, AT5G54215.1, AT5G54225.1, AT5G55132.1, AT5G55565.1, AT5G56368.1, AT5G56369.1, AT5G59105.1, AT5G60553.1, AT5G60615.1, AT5G62623.1, AT5G62627.1, AT5G63660.1

Output proteins: 252 new AtDEFLs

AT1G01780.1, AT1G02180.1, AT1G07170.1, AT1G07170.2, AT1G07170.3, AT1G09794.1, AT1G09920.1, AT1G10588.1, AT1G10588.2, AT1G12064.1, AT1G12660.1, AT1G12663.1, AT1G15100.1, AT1G15120.2, AT1G15600.1, AT1G18250.1, AT1G18250.2, AT1G19310.1, AT1G19320.1, AT1G21835.1, AT1G21864.1, AT1G21866.1, AT1G21925.1, AT1G21928.1, AT1G22510.1, AT1G22690.1, AT1G22690.2, AT1G22690.3, AT1G24147.1, AT1G31175.1, AT1G34047.2, AT1G34792.1, AT1G34795.1, AT1G34800.1, AT1G34805.1, AT1G34810.1, AT1G34815.1, AT1G34820.1, AT1G34825.1, AT1G34830.1, AT1G34840.1, AT1G34850.1, AT1G34860.1, AT1G34930.1, AT1G35035.1, AT1G47540.2, AT1G49030.1, AT1G52855.1, AT1G53010.1, AT1G54390.5, AT1G54390.6, AT1G56415.1, AT1G58242.1, AT1G58245.1, AT1G58248.1, AT1G61105.1, AT1G61255.1, AT1G62370.1, AT1G63830.1, AT1G63830.2, AT1G63830.3, AT1G65295.1, AT1G66100.1, AT1G66610.2, AT1G66660.1, AT1G71866.1, AT1G72060.1, AT1G72175.1, AT1G72260.1, AT1G74620.1, AT1G75030.1, AT1G75040.1, AT1G75050.1, AT1G75750.1, AT1G75750.2, AT1G77525.1, AT1G77682.1, AT1G78922.1, AT1G78995.1, AT2G02470.2, AT2G04240.1, AT2G04240.2, AT2G04675.1, AT2G06925.1, AT2G12880.1, AT2G13895.1, AT2G14288.1, AT2G14900.1, AT2G15010.1, AT2G16050.1, AT2G18420.1, AT2G19000.1, AT2G19690.1, AT2G19690.2, AT2G20825.1, AT2G20875.1, AT2G21320.1, AT2G23240.2, AT2G24480.1, AT2G24617.1, AT2G24790.2, AT2G24860.1, AT2G28790.1, AT2G28920.1, AT2G30000.1, AT2G30810.1, AT2G35200.1, AT2G36320.1, AT2G37030.1, AT2G39540.1, AT2G39900.1, AT2G40935.1, AT2G40935.2, AT2G40935.3, AT2G44390.1, AT2G45010.1, AT2G45010.2, AT2G45135.1, AT2G45135.2, AT2G46493.1, AT2G47890.2, AT3G01170.1, AT3G04720.1, AT3G05345.1, AT3G05870.1, AT3G05870.2, AT3G06310.1, AT3G06310.3, AT3G07200.1, AT3G07200.2, AT3G09922.1, AT3G10185.1, AT3G11200.1, AT3G11200.2, AT3G15534.1, AT3G17670.1, AT3G17670.2, AT3G18470.1, AT3G19220.1, AT3G21890.1, AT3G24010.1, AT3G24465.1, AT3G28210.1, AT3G28620.1, AT3G42830.1, AT3G43180.1, AT3G44716.1, AT3G44716.2, AT3G45470.1, AT3G45555.1, AT3G47180.1, AT3G48205.1, AT3G49550.1, AT3G51960.1, AT3G51960.2, AT3G52800.1, AT3G53075.1, AT3G53080.1, AT3G54040.1, AT3G57480.1, AT3G61980.1, AT3G63390.1, AT4G01340.1, AT4G01575.1, AT4G03965.1, AT4G08910.1, AT4G10240.1, AT4G11360.1, AT4G11650.1, AT4G14746.1, AT4G15248.1, AT4G17680.2, AT4G18110.1, AT4G21620.1, AT4G21620.2, AT4G21720.1, AT4G22230.2, AT4G22235.2, AT4G22250.1, AT4G24204.1, AT4G24204.2, AT4G24204.3, AT4G24973.1, AT4G26370.2, AT4G27310.1, AT4G27660.1, AT4G28190.1, AT4G28820.1, AT4G28820.2, AT4G29460.1, AT4G29470.1, AT4G30370.1, AT4G35480.1, AT4G36000.1, AT4G38960.3, AT5G01015.1, AT5G01070.1, AT5G01070.2, AT5G05610.1, AT5G05610.2, AT5G05910.1, AT5G06130.1, AT5G07040.1, AT5G07225.1, AT5G07500.1, AT5G15230.1, AT5G15230.2, AT5G15260.1, AT5G16430.1, AT5G17610.1, AT5G18800.1, AT5G18800.2, AT5G19430.2, AT5G20570.1, AT5G20570.2, AT5G20570.3, AT5G24610.1, AT5G26673.1, AT5G26692.1, AT5G26717.1, AT5G34828.1, AT5G36720.1, AT5G36805.1, AT5G36910.1, AT5G37055.1, AT5G37900.1, AT5G38378.1, AT5G40590.1, AT5G41390.2, AT5G41430.1, AT5G41450.1, AT5G42200.1, AT5G43200.1, AT5G48545.1, AT5G48655.1, AT5G48655.2, AT5G48655.3, AT5G51900.1, AT5G54148.1, AT5G54150.1, AT5G54470.1, AT5G54990.1, AT5G57123.1, AT5G57820.1, AT5G58412.1, AT5G58787.1, AT5G58787.2, AT5G59845.1, AT5G63740.1, AT5G64816.1, AT5G64816.2, ATCG01060.1

Table S3.2. (suite)

(d) *A. thaliana* defensins against the *O. sativa* proteome → Figure S3.9

Input proteins: 18 *Arabidopsis* defensins

AT1G19610.1, AT1G55010.1, AT1G61070.1, AT1G75830.1, AT2G02100.1, AT2G02120.1, AT2G02130.1, AT2G02140.1, AT2G02147.1, AT2G19893.1, AT2G26010.1, AT2G26020.1, AT2G31953.1, AT2G31957.1, AT3G63360.1, AT5G44420.1, AT5G44430.1, AT5G63660.1

Output proteins: 81 known OsDEFLs

Os01g10550.1, Os01g40220.1, Os01g61360.1, Os01g70680.1, Os02g07440.1, Os02g07495.1, Os02g07550.1, Os02g07575.1, Os02g07600.1, Os02g07624.1, Os02g07628.1, Os02g12060.1, Os02g20130.1, Os02g39625.1, Os02g41904.1, Os02g49540.1, Os02g53570.1, Os02g53590.1, Os02g53600.1, Os02g56870.1, Os03g02255.1, Os03g03810.1, Os03g56682.1, Os04g11130.1, Os04g11165.1, Os04g11195.1, Os04g15740.1, Os04g22235.1, Os04g31250.1, Os04g44130.1, Os05g16395.1, Os06g11308.1, Os06g22880.1, Os06g22925.1, Os06g23060.1, Os06g45320.1, Os06g48665.1, Os06g48690.1, Os07g01700.1, Os07g41290.1, Os08g04520.1, Os08g15505.1, Os08g15545.1, Os08g15550.1, Os09g02160.1, Os09g11790.1, Os10g19892.1, Os10g19894.1, Os10g19898.1, Os10g19902.1, Os10g19904.1, Os10g19912.1, Os10g19914.1, Os10g19925.1, Os10g20540.1, Os10g20550.1, Os10g20560.1, Os10g37290.1, Os11g08170.1, Os11g08220.1, Os11g08235.1, Os11g08240.1, Os11g08250.1, Os11g08260.1, Os11g08265.1, Os11g08270.1, Os11g08280.1, Os11g08285.1, Os11g34990.1, Os11g39910.1, Os11g42520.1, Os11g42525.1, Os11g42530.1, Os11g45360.1, Os11g47229.1, Os11g47278.1, Os12g06750.1, Os12g06760.1, Os12g12220.1, Os12g12230.1, Os12g41790.1

Output proteins: 354 new OsDEFLs

Os01g01700.1, Os01g02490.1, Os01g03310.1, Os01g03320.1, Os01g03380.1, Os01g03390.1, Os01g03400.1, Os01g03464.1, Os01g03680.1, Os01g04040.1, Os01g04050.1, Os01g15800.1, Os01g16120.1, Os01g16210.1, Os01g19340.1, Os01g19800.1, Os01g25110.1, Os01g33350.1, Os01g35330.1, Os01g35330.2, Os01g38700.1, Os01g38760.1, Os01g41140.1, Os01g41180.1, Os01g42720.1, Os01g42720.2, Os01g47280.1, Os01g52110.9, Os01g52890.1, Os01g53650.1, Os01g55710.1, Os01g57240.1, Os01g59610.1, Os01g62260.1, Os01g64230.1, Os01g64620.1, Os01g66070.1, Os01g68589.1, Os01g68840.1, Os01g69040.1, Os01g70920.3, Os02g02630.1, Os02g02650.1, Os02g02870.1, Os02g03800.1, Os02g07930.1, Os02g08014.1, Os02g08200.1, Os02g12490.1, Os02g12490.2, Os02g12490.3, Os02g15010.1, Os02g15020.1, Os02g15080.1, Os02g20934.1, Os02g20934.4, Os02g27300.1, Os02g31050.1, Os02g32080.1, Os02g32840.1, Os02g32990.1, Os02g33705.1, Os02g35144.1, Os02g36950.1, Os02g36950.2, Os02g37570.1, Os02g39800.1, Os02g42412.1, Os02g43519.1, Os02g44260.1, Os02g44700.1, Os02g45054.2, Os02g45054.3, Os02g45054.4, Os02g45710.1, Os02g45780.1, Os02g46720.3, Os02g47870.1, Os02g49710.1, Os02g52550.1, Os02g52790.1, Os02g55480.1, Os02g55480.2, Os02g58500.1, Os03g01270.2, Os03g05270.3, Os03g06580.1, Os03g06580.2, Os03g06580.3, Os03g06580.4, Os03g06835.1, Os03g08710.1, Os03g08710.2, Os03g08840.1, Os03g11230.2, Os03g11600.1, Os03g11600.2, Os03g11874.1, Os03g13360.1, Os03g14300.1, Os03g14550.1, Os03g15820.1, Os03g18420.1, Os03g19059.1, Os03g19180.1, Os03g22680.2, Os03g24184.1, Os03g32230.1, Os03g35920.1, Os03g36080.1, Os03g36439.1, Os03g41060.1, Os03g43100.2, Os03g44100.1, Os03g45960.1, Os03g46060.1, Os03g46070.1, Os03g47270.1, Os03g48020.1, Os03g48020.2, Os03g48104.1, Os03g49270.1, Os03g49280.1, Os03g49300.1, Os03g49310.1, Os03g50030.1, Os03g52360.1, Os03g52370.1, Os03g52370.2, Os03g52380.1, Os03g52390.1, Os03g55290.1, Os03g56410.1, Os03g58280.1, Os03g60600.2, Os03g60840.1, Os03g61440.1, Os03g61470.1, Os03g62530.1, Os03g62570.1, Os04g09390.1, Os04g09560.1, Os04g16970.1, Os04g22220.1, Os04g24410.2, Os04g26400.1, Os04g30250.1, Os04g30250.2, Os04g32700.1, Os04g32800.1, Os04g32810.1, Os04g32820.1, Os04g32830.1, Os04g32850.3, Os04g34540.1, Os04g38470.1, Os04g39110.1, Os04g39110.2, Os04g39360.1, Os04g39380.1, Os04g40350.1, Os04g44600.1, Os04g45010.1, Os04g47840.1, Os04g48300.1, Os04g48310.1, Os04g49000.1, Os04g49160.1, Os04g49170.1, Os04g49670.1, Os04g50176.1, Os04g54490.1, Os04g56760.1, Os04g57320.2, Os04g57780.2, Os04g58120.1, Os04g58380.1, Os05g01730.1, Os05g01730.2, Os05g02250.1, Os05g06014.1, Os05g06014.2, Os05g07830.1, Os05g08854.1, Os05g13420.1, Os05g25930.1, Os05g28730.1, Os05g28780.1, Os05g29740.1, Os05g30410.1, Os05g31470.1, Os05g33010.2, Os05g33830.1, Os05g35850.1, Os05g38880.1, Os05g39940.1, Os05g40020.1, Os05g43380.2, Os05g51040.1, Os05g51660.1, Os06g01340.1, Os06g03682.2, Os06g05120.1, Os06g05970.1, Os06g09520.1, Os06g14070.1, Os06g14200.1, Os06g14640.1, Os06g15620.1, Os06g16060.1, Os06g21960.1, Os06g31800.1, Os06g31890.1, Os06g31930.1, Os06g31960.1, Os06g32020.1, Os06g32160.1, Os06g32240.1, Os06g32350.1, Os06g32550.1, Os06g32600.1, Os06g34650.1, Os06g38460.1, Os06g44610.1, Os06g44820.1, Os06g45580.1, Os06g46436.1, Os06g47440.1, Os07g06834.1, Os07g09120.1, Os07g14150.4, Os07g14470.1, Os07g14740.1, Os07g22498.1, Os07g22840.1, Os07g23470.1, Os07g23770.1, Os07g34964.1, Os07g38010.1, Os07g38410.1, Os07g40240.1, Os07g40370.1, Os07g41890.1, Os07g41920.1, Os07g42040.1, Os07g43130.1, Os07g43830.1, Os07g43840.1, Os07g46390.1, Os07g46390.2, Os07g46390.3, Os07g46700.3, Os08g03579.1, Os08g06090.1, Os08g13380.1, Os08g27260.1, Os08g35070.1, Os08g35980.1, Os09g11210.1, Os09g21710.1, Os09g24840.1, Os09g24840.2, Os09g25590.1, Os09g25784.2, Os09g27860.1, Os09g28480.2, Os09g32660.1, Os09g32690.1, Os09g32690.2, Os09g32690.3, Os09g32730.1, Os09g34310.1, Os09g34980.1, Os09g35690.1, Os09g35880.1, Os09g36720.1, Os09g37500.1, Os10g02300.1, Os10g02300.2, Os10g02300.3, Os10g02625.1, Os10g05900.1, Os10g07606.1, Os10g08780.1, Os10g09870.3, Os10g21406.1, Os10g28380.1, Os10g30230.1, Os10g30310.1, Os10g30850.1, Os10g30850.2, Os10g31850.2, Os10g31850.5, Os10g35670.1, Os10g35930.1, Os10g39100.1, Os10g39770.1, Os10g39820.1, Os10g39936.1, Os10g42270.1, Os11g02250.1, Os11g02820.1, Os11g04280.1, Os11g04281.1, Os11g09460.1, Os11g09680.1, Os11g09680.2, Os11g10160.1, Os11g12600.1, Os11g14010.2, Os11g19500.1, Os11g26580.1, Os11g26710.1, Os11g27540.1, Os11g27540.2, Os11g34440.1, Os11g35000.1, Os11g36430.5, Os11g36480.1, Os11g38710.1, Os11g40200.1, Os11g47690.1, Os12g02210.1, Os12g02800.1, Os12g02800.2, Os12g02800.3, Os12g04090.1, Os12g05280.1, Os12g05660.1, Os12g05660.2, Os12g05660.3, Os12g05660.4, Os12g08210.1, Os12g09465.1, Os12g18410.1, Os12g18410.2, Os12g25130.1, Os12g26960.1, Os12g29390.1, Os12g30170.1, Os12g31670.1, Os12g32374.1, Os12g34796.1, Os12g40470.1, Os12g41310.1, Os12g43380.1, Os12g43390.1, Os12g43430.1, Os12g43440.1, Os12g43450.1, Os12g43810.1, Os12g43820.1, ChrUn.fgenesh.mRNA.35, ChrUn.fgenesh.mRNA.41

(e) *A. thaliana* snakins in *A. thaliana* → Figure S3.10

Input proteins: 15 known *Arabidopsis* snakins

AT1G10588.1, AT1G22690.3, AT1G74670.1, AT1G75750.1, AT2G14900.1, AT2G18420.1, AT2G30810.1, AT2G39540.1, AT3G02885.1, AT3G10185.1, AT4G09600.1, AT4G09610.1, AT5G14920.1, AT5G15230.1, AT5G59845.1

Output proteins: 15 known *Arabidopsis* snakins

AT1G10588.1, AT1G22690.3, AT1G74670.1, AT1G75750.1, AT2G14900.1, AT2G18420.1, AT2G30810.1, AT2G39540.1, AT3G02885.1, AT3G10185.1, AT4G09600.1, AT4G09610.1, AT5G14920.1, AT5G15230.1, AT5G59845.1

Output proteins: 7 new *Arabidopsis* snakins

AT1G10588.2, AT1G22690.1, AT1G22690.2, AT1G75750.2, AT3G27330.1, AT5G15230.2, AT5G38330.1

Table S3.2. (suite)

(f) *A. thaliana* snakins in *O. sativa* → Figure S3.11

Input proteins: 15 *Arabidopsis* snakins

AT1G10588.1, AT1G22690.3, AT1G74670.1, AT1G75750.1, AT2G14900.1, AT2G18420.1, AT2G30810.1, AT2G39540.1, AT3G02885.1, AT3G10185.1, AT4G09600.1, AT4G09610.1, AT5G14920.1, AT5G15230.1, AT5G59845.1

Output proteins: 11 known rice snakins

Os03g14550.1, Os03g41060.1, Os03g55290.1, Os04g39110.1, Os05g31280.1, Os05g35690.1, Os06g15620.1, Os06g51320.1, Os07g40240.1, Os09g24840.2, Os10g02625.1

Output proteins: 5 new rice snakins

Os04g39110.2, Os06g51320.2, Os06g51320.3, Os09g24840.1, Os12g16690.1

(g) *H. sapiens* EGZs against the *H. sapiens* proteome → Figure S3.12

Input proteins: 8 human EGZs

NP_000521.2, NP_004354.2, NP_005579.2, NP_060389.2, NP_079045.1, NP_714928.1, NP_001073592.1, NP_001129602.1

Output proteins: 8 known human EGZs

NP_000521.2, NP_004354.2, NP_005579.2, NP_060389.2, NP_079045.1, NP_714928.1, NP_001073592.1, NP_001129602.1

Output proteins: 2000 new human EGZs

NP_000059.3, NP_000112.1, NP_000214.1, NP_000235.2, NP_000278.3, NP_000302.1, NP_000334.1, NP_000342.2, NP_000412.3, NP_000413.1, NP_000432.1, NP_000436.2, NP_000466.2, NP_000475.1, NP_000492.2, NP_000529.1, NP_000531.2, NP_000578.2, NP_000594.2, NP_000604.1, NP_000788.2, NP_000910.2, NP_000912.3, NP_000997.1, NP_001040.1, NP_001072.2, NP_001127.1, NP_001147.1, NP_001149.2, NP_001161.1, NP_001261.2, NP_001318.1, NP_001348.2, NP_001379.1, NP_001387.2, NP_001399.1, NP_001401.2, NP_001409.3, NP_001420.2, NP_001427.2, NP_001442.2, NP_001444.2, NP_001498.1, NP_001597.2, NP_001611.1, NP_001633.1, NP_001640.1, NP_001703.2, NP_001767.3, NP_001806.2, NP_001808.2, NP_001824.1, NP_001846.3, NP_001888.2, NP_001931.2, NP_001933.2, NP_001935.2, NP_001938.2, NP_002006.2, NP_002027.2, NP_002043.2, NP_002127.1, NP_002169.1, NP_002242.2, NP_002263.3, NP_002272.2, NP_002273.3, NP_002274.1, NP_002275.1, NP_002326.2, NP_002596.1, NP_002607.2, NP_002705.2, NP_002740.2, NP_002766.1, NP_002770.3, NP_002834.3, NP_002837.1, NP_002838.2, NP_002966.1, NP_002967.2, NP_003003.3, NP_003031.3, NP_003032.1, NP_003044.1, NP_003045.2, NP_003061.3, NP_003063.2, NP_003082.1, NP_003088.1, NP_003122.1, NP_003165.2, NP_003176.2, NP_003232.2, NP_003236.3, NP_003308.1, NP_003362.2, NP_003390.4, NP_003449.2, NP_003473.3, NP_003478.1, NP_003657.1, NP_003733.2, NP_003740.2, NP_003793.2, NP_003820.2, NP_003881.2, NP_003923.2, NP_004025.1, NP_004027.2, NP_004047.3, NP_004073.2, NP_004138.1, NP_004226.3, NP_004230.2, NP_004317.2, NP_004387.1, NP_004406.2, NP_004487.2, NP_004492.2, NP_004507.2, NP_004617.2, NP_004629.3, NP_004665.2, NP_004682.2, NP_004705.1, NP_004825.3, NP_004851.2, NP_004865.1, NP_004934.1, NP_005057.1, NP_005076.3, NP_005131.1, NP_005174.2, NP_005258.2, NP_005378.4, NP_005380.2, NP_005387.2, NP_005447.1, NP_005506.3, NP_005547.3, NP_005572.2, NP_005589.1, NP_005641.1, NP_005648.1, NP_005735.2, NP_005817.1, NP_005849.1, NP_005872.2, NP_005912.1, NP_005947.3, NP_005959.2, NP_005985.3, NP_006050.3, NP_006261.1, NP_006275.1, NP_006318.1, NP_006421.2, NP_006431.2, NP_006442.2, NP_006469.2, NP_006517.1, NP_006522.2, NP_006551.1, NP_006565.2, NP_006673.1, NP_006737.2, NP_006740.1, NP_006862.2, NP_008835.5, NP_008864.3, NP_008938.2, NP_008948.1, NP_009027.1, NP_009153.3, NP_009155.1, NP_009176.2, NP_031370.2, NP_031401.1, NP_033720.2, NP_036350.2, NP_036356.1, NP_036425.1, NP_036427.1, NP_036429.2, NP_036524.1, NP_036542.1, NP_037397.2, NP_037404.2, NP_037457.3, NP_037469.2, NP_037473.3, NP_037495.1, NP_054729.3, NP_054798.1, NP_054808.1, NP_055020.1, NP_055028.3, NP_055042.1, NP_055108.2, NP_055127.2, NP_055143.2, NP_055144.3, NP_055150.1, NP_055156.1, NP_055160.2, NP_055172.1, NP_055184.2, NP_055204.3, NP_055402.2, NP_055416.2, NP_055427.2, NP_055445.2, NP_055489.1, NP_055491.1, NP_055514.3, NP_055542.1, NP_055554.1, NP_055582.1, NP_055601.2, NP_055604.3, NP_055673.2, NP_055674.4, NP_055718.4, NP_055740.3, NP_055746.3, NP_055778.2, NP_055795.1, NP_055806.2, NP_055833.2, NP_055850.1, NP_055852.2, NP_055903.2, NP_055940.3, NP_055987.2, NP_055993.2, NP_056001.1, NP_056050.1, NP_056054.2, NP_056062.1, NP_056097.1, NP_056112.1, NP_056125.3, NP_056132.1, NP_056134.2, NP_056193.2, NP_056203.2, NP_056255.2, NP_056258.1, NP_056290.3, NP_056471.1, NP_056508.2, NP_056526.3, NP_056982.1, NP_056984.1, NP_057009.3, NP_057060.2, NP_057128.2, NP_057161.1, NP_057184.1, NP_057211.4, NP_057223.1, NP_057232.2, NP_057323.3, NP_057354.1, NP_057361.3, NP_057368.3, NP_057396.1, NP_057404.2, NP_057541.2, NP_057652.1, NP_057673.2, NP_057722.3, NP_059122.2, NP_059867.3, NP_060022.1, NP_060090.2, NP_060123.3, NP_060125.4, NP_060154.3, NP_060158.2, NP_060164.3, NP_060295.1, NP_060318.3, NP_060425.2, NP_060428.2, NP_060482.2, NP_060516.2, NP_060548.2, NP_060550.2, NP_060553.4, NP_060626.2, NP_060653.3, NP_060681.2, NP_060700.2, NP_060853.3, NP_060859.4, NP_061092.3, NP_061160.3, NP_061164.3, NP_061856.1, NP_061869.2, NP_061889.2, NP_061985.2, NP_064612.2, NP_064714.2, NP_065176.3, NP_065681.1, NP_065706.2, NP_065777.1, NP_065798.2, NP_065804.2, NP_065815.1, NP_065842.1, NP_065872.1, NP_065874.1, NP_065916.2, NP_065934.1, NP_066124.1, NP_066274.2, NP_066921.2, NP_067023.1, NP_067064.2, NP_067070.1, NP_067633.2, NP_068506.2, NP_068755.2, NP_068770.2, NP_068779.1, NP_071401.3, NP_071737.1, NP_071920.1, NP_073150.2, NP_073716.1, NP_073717.1, NP_073718.1, NP_073719.1, NP_075011.1, NP_075049.1, NP_075392.2, NP_075408.1, NP_075461.2, NP_076415.3, NP_076418.3, NP_077269.3, NP_077313.3, NP_077726.1, NP_078781.3, NP_078813.1, NP_078838.1, NP_079152.3, NP_079160.1, NP_079345.1, NP_079418.1, NP_079464.2, NP_079470.1, NP_079524.2, NP_085148.1, NP_085915.2, NP_110398.1, NP_110410.1, NP_112420.1, NP_112480.2, NP_113636.2, NP_113654.4, NP_114032.2, NP_115285.1, NP_115615.2, NP_115667.2, NP_115668.4, NP_115743.1, NP_115755.2, NP_115785.1, NP_115805.1, NP_115871.1, NP_115890.2, NP_115954.1, NP_116114.1, NP_116161.2, NP_116162.1, NP_116241.2, NP_116274.3, NP_116561.1, NP_127460.1, NP_149016.2, NP_149034.2, NP_149052.1, NP_149072.2, NP_149118.2, NP_150648.2, NP_201574.3, NP_207646.2, NP_258260.1, NP_291035.1, NP_443117.3, NP_443129.3, NP_443130.1, NP_443168.2, NP_443176.2, NP_476429.2, NP_478102.2, NP_542417.2, NP_542775.2, NP_542785.1, NP_542786.1, NP_542937.2, NP_567824.1, NP_569120.1, NP_569122.1, NP_570711.1, NP_570712.1, NP_570713.1, NP_570714.1, NP_570715.1, NP_570716.1, NP_570857.2, NP_570858.2, NP_612409.1, NP_612439.2, NP_612444.2, NP_612466.1, NP_612475.1, NP_612808.1, NP_620121.1, NP_620176.1, NP_620601.1, NP_620602.2, NP_620603.2, NP_620614.2, NP_624357.1, NP_624358.2, NP_631961.1, NP_640343.1, NP_653170.3, NP_653232.3, NP_653307.1, NP_659493.2, NP_660159.1, NP_660292.1, NP_663320.2, NP_663625.2, NP_663719.2, NP_663720.1, NP_663747.1, NP_663760.1, NP_680780.1, NP_689414.1, NP_689422.1, NP_689424.2, NP_689550.2, NP_689564.3, NP_689588.2, NP_689662.2, NP_699200.2, NP_703152.1, NP_703194.1, NP_705834.2, NP_722520.2, NP_736610.2, NP_758516.1, NP_758965.2, NP_775487.2, NP_775762.1, NP_775782.2, NP_775835.2, NP_775871.2, NP_777591.3, NP_777596.2, NP_783195.2, NP_787028.1, NP_787050.6, NP_787053.1, NP_788850.1, NP_795352.2, NP_796374.1, NP_835144.1, NP_835739.1, NP_840054.1, NP_847896.1, NP_848126.1, NP_848128.1, NP_848129.1, NP_848130.1, NP_848511.1, NP_848515.1, NP_848516.1, NP_848517.1, NP_848563.1, NP_849188.4, NP_851422.1, NP_852469.1, NP_852668.1, NP_852669.1, NP_853517.2, NP_853636.3, NP_853648.1, NP_853650.1, NP_859059.1, NP_859069.2, NP_862825.1, NP_872363.1, NP_872369.2, NP_872413.1, NP_877495.1, NP_878905.2, NP_892021.1, NP_892120.2, NP_898902.1 → *cf.* page suivante

Table S3.2. (suite)

Output proteins: 2000 new human EGZs (suite)

NP_899050.1, NP_899230.2, NP_919223.1, NP_919267.2, NP_932079.1, NP_937859.1, NP_938011.1, NP_938072.1, NP_940860.2, NP_940881.2, NP_940885.2, NP_941989.1, NP_945187.1, NP_945190.1, NP_945192.1, NP_945345.2, NP_954631.1, NP_955368.1, NP_955373.3, NP_958780.1, NP_958781.1, NP_958782.1, NP_958783.1, NP_958784.1, NP_958785.1, NP_958786.1, NP_958816.1, NP_958817.1, NP_963840.2, NP_982249.1, NP_987100.1, NP_987101.1, NP_991403.1, NP_991404.1, NP_996810.1, NP_996811.1, NP_996812.1, NP_996813.1, NP_996832.1, NP_996882.1, NP_997213.1, NP_997215.1, NP_997460.1, NP_997527.2, NP_997698.1, NP_998885.1, NP_001001671.3, NP_001002251.2, NP_001002252.2, NP_001003704.1, NP_001004342.3, NP_001005407.1, NP_001007027.1, NP_001007564.1, NP_001008271.2, NP_001008564.1, NP_001008783.1, NP_001008844.1, NP_001011547.2, NP_001011724.1, NP_001011725.1, NP_001012426.1, NP_001012427.1, NP_001012526.2, NP_001014364.1, NP_001014830.1, NP_001017926.1, NP_001018047.1, NP_001018048.1, NP_001018082.1, NP_001018098.1, NP_001018494.1, NP_001018861.3, NP_001019554.1, NP_001020014.1, NP_001020083.1, NP_001020767.1, NP_001026896.2, NP_001026904.1, NP_001030014.2, NP_001030025.1, NP_001032624.1, NP_001032625.1, NP_001032629.1, NP_001032728.1, NP_001032895.2, NP_001034485.1, NP_001034681.1, NP_001034760.1, NP_001034761.1, NP_001034762.1, NP_001034794.1, NP_001035047.1, NP_001035814.1, NP_001035877.1, NP_001035879.1, NP_001035915.1, NP_001036024.3, NP_001036068.1, NP_001036188.1, NP_001070145.1, NP_001070730.1, NP_001070909.1, NP_001072982.1, NP_001073136.1, NP_001073349.1, NP_001073350.1, NP_001073351.1, NP_001073590.1, NP_001073591.1, NP_001073882.3, NP_001073903.1, NP_001073926.1, NP_001073977.1, NP_001075109.1, NP_001075221.1, NP_001075222.1, NP_001075223.1, NP_001075224.1, NP_001075949.1, NP_001091645.1, NP_001091896.1, NP_001092879.1, NP_001093592.1, NP_001094093.1, NP_001094343.1, NP_001095867.1, NP_001095868.1, NP_001095869.1, NP_001095896.1, NP_001098673.1, NP_001098674.1, NP_001098675.1, NP_001099009.1, NP_001099039.1, NP_001107571.1, NP_001107869.1, NP_001116423.1, NP_001116513.2, NP_001119528.2, NP_001119529.2, NP_001119530.2, NP_001119531.1, NP_001120693.1, NP_001120694.1, NP_001122316.1, NP_001122317.1, NP_001122318.1, NP_001122319.1, NP_001122320.1, NP_001122321.1, NP_001123368.1, NP_001123505.1, NP_001123506.1, NP_001123507.1, NP_001123508.1, NP_001123577.1, NP_001127870.1, NP_001127906.1, NP_001128350.1, NP_001128512.1, NP_001128513.1, NP_001128563.1, NP_001128653.1, NP_001129163.1, NP_001129488.1, NP_001129496.1, NP_001129601.1, NP_001129603.1, NP_001129737.1, NP_001129743.2, NP_001131022.1, NP_001131145.1, NP_001135451.1, NP_001135452.1, NP_001135745.1, NP_001135746.1, NP_001135748.1, NP_001135749.1, NP_001135750.1, NP_001135796.1, NP_001135797.1, NP_001135842.1, NP_001136027.1, NP_001136112.1, NP_001136115.1, NP_001136272.1, NP_001136402.1, NP_001138408.1, NP_001138516.1, NP_001138737.1, NP_001138738.1, NP_001138933.1, NP_001138934.1, NP_001138935.1, NP_001139014.1, NP_001139015.1, NP_001139016.1, NP_001139107.1, NP_001139434.1, NP_001139697.1, NP_001139698.1, NP_001153208.1, NP_001153508.1, NP_001153509.1, NP_001153581.1, NP_001153582.1, NP_001153583.1, NP_001155018.1, NP_001155102.1, NP_001155973.1, NP_001156423.1, NP_001156424.1, NP_001156425.1, NP_001157394.1, NP_001157650.1, NP_001157651.1, NP_001157653.1, NP_001157654.1, NP_001157655.1, NP_001158058.1, NP_001160005.1, NP_001161210.1, NP_001161328.1, NP_001164024.1, NP_001164672.1, NP_001164673.1, NP_001165422.1, NP_001166110.1, NP_001166111.1, NP_001166176.1, NP_001167539.1, NP_001167551.1, NP_001170777.1, NP_001171590.1, NP_001171591.1, NP_001171592.1, NP_001171689.1, NP_001171690.1, NP_001171691.1, NP_001171742.1, NP_001171744.1, NP_001171745.1, NP_001177276.1, NP_001177277.1, NP_001177278.1, NP_001177719.1, NP_001177765.1, NP_001177766.1, NP_001180298.1, NP_001180394.1, NP_001180395.1, NP_001185744.1, NP_001185832.1, NP_001185833.1, NP_001186210.1, NP_001186326.1, NP_001186621.1, NP_001186622.1, NP_001186623.1, NP_001186692.1, NP_001186693.1, NP_001186871.1, NP_001188358.1, NP_001188379.1, NP_001188380.1, NP_001188388.1, NP_001191230.1, NP_001191231.1, NP_001191232.1, NP_001191807.1, NP_001192273.1, NP_001193858.1, NP_001193861.1, NP_001193871.1, NP_001193872.1, NP_001193873.1, NP_001193874.1, NP_001193889.1, NP_001193890.1, NP_001193938.1, NP_001193980.1, NP_001229267.1, NP_001229416.1, NP_001229417.1, NP_001229488.1, NP_001229489.1, NP_001229696.1, NP_001229697.1, NP_001229854.1, NP_001229861.1, NP_001230066.1, NP_001230228.1, NP_001231879.1, NP_001238978.1, NP_001238979.1, NP_001238980.1, NP_001238981.1, NP_001238982.1, NP_001239263.1, NP_001239604.1, NP_001239606.1, NP_001241663.1, NP_001242975.1, NP_001242976.1, NP_001243084.1, NP_001243243.1, NP_001243452.1, NP_001243650.1, NP_001243718.1, NP_001244109.1, NP_001244110.1, NP_001244953.1, NP_001244960.1, NP_001245340.1, NP_001245341.1, NP_001245342.1, NP_001245343.1, NP_001247418.1, NP_001247419.1, NP_001248335.1, NP_001248336.1, NP_001248772.1, NP_001252541.1, NP_001253991.1, NP_001253992.1, NP_001253993.1, NP_001253994.1, NP_001257577.1, NP_001257578.1, NP_001257894.1, NP_001257895.1, NP_001257972.1, NP_001258112.1, NP_001258113.1, NP_001258570.1, NP_001258804.1, NP_001258867.1, NP_001263260.1, NP_001264092.1, NP_001264992.1, NP_001265156.1, NP_001265157.1, NP_001265158.1, NP_001265159.1, NP_001265170.1, NP_001265307.1, NP_001265518.1, NP_001265533.1, NP_001265565.1, NP_001265595.1, NP_001265596.1, NP_001265597.1, NP_001265659.1, NP_001265841.1, NP_001265842.1, NP_001265843.1, NP_001265844.1, NP_001265845.1, NP_001265846.1, NP_001265847.1, NP_001265868.1, NP_001268433.1, NP_001268916.1, NP_001269166.1, NP_001269167.1, NP_001269346.1, NP_001269441.1, NP_001269681.1, NP_001269719.1, NP_001269720.1, NP_001269785.1, NP_001269787.1, NP_001269789.1, NP_001269790.1, NP_001271168.1, NP_001271346.1, NP_001271347.1, NP_001271348.1, NP_001271349.1, NP_001271350.1, NP_001271351.1, NP_001273328.1, NP_001273560.1, NP_001273561.1, NP_001273644.1, NP_001273646.1, NP_001273651.1, NP_001275585.1, NP_001276325.1, NP_001276326.1, XP_002342143.5, XP_003403803.1, XP_003846382.1, XP_003846613.1, XP_003846671.2, XP_003846716.2, XP_003960511.2, XP_005245072.1, XP_005245073.1, XP_005245074.1, XP_005245075.1, XP_005245076.1, XP_005245077.1, XP_005245079.1, XP_005245080.1, XP_005245081.1, XP_005245082.1, XP_005245083.1, XP_005245084.1, XP_005245085.1, XP_005245086.1, XP_005245087.1, XP_005245340.1, XP_005245341.1, XP_005245343.1, XP_005245479.1, XP_005245618.1, XP_005245629.1, XP_005245637.1, XP_005245741.1, XP_005245742.1, XP_005245743.1, XP_005245746.1, XP_005245747.1, XP_005245768.1, XP_005245769.1, XP_005245770.1, XP_005245771.1, XP_005246169.1, XP_005246170.1, XP_005246171.1, XP_005246172.1, XP_005246173.1, XP_005246174.1, XP_005246175.1, XP_005246177.1, XP_005246178.1, XP_005246180.1, XP_005246181.1, XP_005246184.1, XP_005246194.1, XP_005246195.1, XP_005246196.1, XP_005246197.1, XP_005246437.1, XP_005246438.1, XP_005246439.1, XP_005246711.1, XP_005246712.1, XP_005246713.1, XP_005246905.1, XP_005247049.1, XP_005247050.1, XP_005247052.1, XP_005247053.1, XP_005247073.1, XP_005247457.1, XP_005247639.1, XP_005247640.1, XP_005247641.1, XP_005247642.1, XP_005247964.1, XP_005248009.1, XP_005248576.1, XP_005248577.1, XP_005248602.1, XP_005248616.1, XP_005248836.1, XP_005249035.1, XP_005249036.1, XP_005249046.1, XP_005249047.1, XP_005249237.1, XP_005249617.1, XP_005250003.1, XP_005250004.1, XP_005250005.1, XP_005250050.1, XP_005250244.1, XP_005250245.1, XP_005250285.1, XP_005250482.1, XP_005250483.1, XP_005250606.1, XP_005251033.1, XP_005251034.1, XP_005251035.1, XP_005251036.1, XP_005251037.1, XP_005251038.1, XP_005251039.1, XP_005251040.1, XP_005251041.1, XP_005251108.1, XP_005251344.1, XP_005251421.1, XP_005251477.1, XP_005251493.1, XP_005251494.1, XP_005251495.1, XP_005251510.1, XP_005251511.1, XP_005251679.1, XP_005251680.1, XP_005251694.1, XP_005251899.1, XP_005251904.1, XP_005251905.1, XP_005251906.1, XP_005252102.1, XP_005252103.1, XP_005252119.1, XP_005252120.1, XP_005252359.1, XP_005252362.1, XP_005252390.1, XP_005252506.1, XP_005252621.1, XP_005252622.1, XP_005252623.1, XP_005252624.1, XP_005252625.1, XP_005252626.1, XP_005252627.1, XP_005252628.1, XP_005252629.1, XP_005252630.1, XP_005252741.1, XP_005252811.1, XP_005252812.1, XP_005252813.1, XP_005252940.1, XP_005253007.1, XP_005253101.1, XP_005253102.1, XP_005253103.1, XP_005253104.1, XP_005253105.1, XP_005253210.1, XP_005253283.1, XP_005253629.1, XP_005253630.1, XP_005253632.1, XP_005253633.1, XP_005254281.1, XP_005254446.1, XP_005254448.1, XP_005254538.1, XP_005254540.1, XP_005254541.1, XP_005254542.1, XP_005254691.1, XP_005254692.1, XP_005254924.1, XP_005254946.1, XP_005254990.1, XP_005255053.1, XP_005255118.1, XP_005255119.1, XP_005255120.1, XP_005255121.1, XP_005255122.1, XP_005255123.1, XP_005255126.1, XP_005255127.1, XP_005255128.1, XP_005255129.1, XP_005255134.1, XP_005255137.1, XP_005255138.1, XP_005255139.1, XP_005255320.1, XP_005255334.1, XP_005255335.1, XP_005255336.1, XP_005255337.1, XP_005255515.1, XP_005255690.1, XP_005255709.1,XP_005255767.1, XP_005255768.1, XP_005255769.1, XP_005255848.1, XP_005255849.1, XP_005255869.1, XP_005256332.2, XP_005256514.1, XP_005256556.1, XP_005256557.1, XP_005256732.1, XP_005256746.1, XP_005256747.1, XP_005256776.1, XP_005256945.1, XP_005256946.1, XP_005256947.1, XP_005256948.1, XP_005257168.1, XP_005257289.2, XP_005257290.2, XP_005257291.2, XP_005257292.2, XP_005257364.1, XP_005257400.1, XP_005257406.1, XP_005257504.2, XP_005257604.1, XP_005257605.1, XP_005257661.2, XP_005258470.2, XP_005258472.2, XP_005258491.1, XP_005258708.1, XP_005258726.1, XP_005258727.1, XP_005258730.1, XP_005258731.1, XP_005258732.1, XP_005258773.2, XP_005258845.1, XP_005258846.1, XP_005258890.1, XP_005259020.1, XP_005259037.1, XP_005259039.1, XP_005259231.1, XP_005259327.1, XP_005259328.1, XP_005259329.1, XP_005259330.1, XP_005259331.1, XP_005259441.1, XP_005259442.1, XP_005259452.1, XP_005259455.1, XP_005259486.1, XP_005259506.1, XP_005259576.1, XP_005259577.1, XP_005259627.2, XP_005259628.1, XP_005259629.1, XP_005259697.1, XP_005259716.1, XP_005259717.1, XP_005259763.1, XP_005259952.1, XP_005259974.1, XP_005259975.1, XP_005259977.1, XP_005260061.1, XP_005260062.1, XP_005260063.1, XP_005260064.1 → *cf.* page suivante

Table S3.2. (suite)

Output proteins: 2000 new human EGZs (suite)

XP_005260065.1, XP_005260066.1, XP_005260067.1, XP_005260068.1, XP_005260069.1, XP_005260085.1, XP_005260087.1, XP_005260088.1, XP_005260089.1, XP_005260090.1, XP_005260091.1, XP_005260092.1, XP_005260398.1, XP_005260400.1, XP_005260466.1, XP_005260602.1, XP_005260988.1, XP_005260990.1, XP_005261472.1, XP_005261473.1, XP_005261474.1, XP_005261476.1, XP_005261477.1, XP_005261522.1, XP_005261657.1, XP_005261779.1, XP_005261780.1, XP_005261781.1, XP_005262065.1, XP_005262068.1, XP_005262073.2, XP_005262075.2, XP_005262121.1, XP_005262184.1, XP_005262283.1, XP_005262286.1, XP_005262664.1, XP_005262666.1, XP_005262667.1, XP_005262915.1, XP_005263126.1, XP_005263143.1, XP_005263144.1, XP_005263145.1, XP_005263147.1, XP_005263149.1, XP_005263150.1, XP_005263151.1, XP_005263152.1, XP_005263154.1, XP_005263155.1, XP_005263158.1, XP_005263235.1, XP_005263359.2, XP_005263439.1, XP_005263440.1, XP_005263441.1, XP_005263442.1, XP_005263443.1, XP_005263920.1, XP_005263949.1, XP_005263950.1, XP_005264101.1, XP_005264102.1, XP_005264103.1, XP_005264104.1, XP_005264105.1, XP_005264106.1, XP_005264107.1, XP_005264109.1, XP_005264110.1, XP_005264111.1, XP_005264112.1, XP_005264113.1, XP_005264114.1, XP_005264115.1, XP_005264116.1, XP_005264117.1, XP_005264118.1, XP_005264119.1, XP_005264120.1, XP_005264121.1, XP_005264122.1, XP_005264123.1, XP_005264125.1, XP_005264126.1, XP_005264128.1, XP_005264159.1, XP_005264161.1, XP_005264162.1, XP_005264199.1, XP_005264429.2, XP_005264764.1, XP_005264835.1, XP_005265445.1, XP_005265446.1, XP_005265460.1, XP_005265476.1, XP_005265858.1, XP_005266077.1, XP_005266389.1, XP_005266390.1, XP_005266548.1, XP_005266549.1, XP_005266550.1, XP_005266551.1, XP_005266553.1, XP_005266554.1, XP_005266555.1, XP_005266556.1, XP_005266558.1, XP_005266603.1, XP_005266604.1, XP_005266605.1, XP_005266606.1, XP_005266607.1, XP_005266608.1, XP_005266610.1, XP_005266612.1, XP_005266672.1, XP_005266964.1, XP_005266967.1, XP_005266968.1, XP_005267293.1, XP_005267296.1, XP_005267371.1, XP_005267530.1, XP_005267533.1, XP_005267695.1, XP_005268132.1, XP_005268191.1, XP_005268192.1, XP_005268270.1, XP_005268272.1, XP_005268581.1, XP_005268645.1, XP_005268646.1, XP_005268647.1, XP_005268673.1, XP_005268674.1, XP_005268675.1, XP_005268677.1, XP_005268768.1, XP_005268772.1, XP_005268779.1, XP_005268781.1, XP_005268782.1, XP_005268784.1, XP_005268785.1, XP_005268786.1, XP_005268883.1, XP_005268916.1, XP_005268921.1, XP_005268923.1, XP_005269107.1, XP_005269219.1, XP_005269272.1, XP_005269419.1, XP_005269420.1, XP_005269422.1, XP_005269424.1, XP_005269705.1, XP_005269706.1, XP_005269708.1, XP_005269709.1, XP_005269710.1, XP_005269712.1, XP_005269795.1, XP_005269796.1, XP_005270123.1, XP_005270279.1, XP_005270280.1, XP_005270281.1, XP_005270359.1, XP_005270398.1, XP_005270400.1, XP_005270402.1, XP_005270404.1, XP_005270637.1, XP_005270836.1, XP_005270837.1, XP_005271013.1, XP_005271168.1, XP_005271169.1, XP_005271170.1, XP_005271172.1, XP_005271173.1, XP_005271321.1, XP_005271322.1, XP_005271323.1, XP_005271345.1, XP_005271568.1, XP_005271569.1, XP_005271570.1, XP_005271923.1, XP_005271924.1, XP_005272094.1, XP_005272095.1, XP_005272096.1, XP_005272097.1, XP_005272270.1, XP_005272273.1, XP_005272274.1, XP_005272275.1, XP_005272276.1, XP_005272426.2, XP_005272427.2, XP_005272441.1, XP_005272442.1, XP_005272443.1, XP_005272517.1, XP_005272535.1, XP_005272536.1, XP_005272537.1, XP_005272538.1, XP_005272539.1, XP_005272540.1, XP_005272815.1, XP_005273028.1, XP_005273194.1, XP_005273518.1, XP_005273670.1, XP_005273672.1, XP_005273739.1, XP_005274051.1, XP_005274058.1, XP_005274059.1, XP_005274060.1, XP_005274114.1, XP_005274115.1, XP_005274116.1, XP_005274117.1, XP_005274118.1, XP_005274119.1, XP_005274120.1, XP_005274123.1, XP_005274124.1, XP_005274288.1, XP_005274297.1, XP_005274298.1, XP_005274299.1, XP_005274300.1, XP_005274301.1, XP_005274302.1, XP_005274303.1, XP_005274557.2, XP_005274568.1, XP_005274596.1, XP_005274635.1, XP_005274636.1, XP_005274885.1, XP_005274965.1, XP_005275259.1, XP_005275391.1, XP_005275558.1, XP_005275701.1, XP_005275702.1, XP_005275778.1, XP_005275793.1, XP_005275896.1, XP_005276050.2, XP_005276205.1, XP_005276486.1, XP_005276496.1, XP_005276920.1, XP_005277012.1, XP_005277013.1, XP_005277014.1, XP_005277045.1, XP_005277474.1, XP_005277516.1, XP_005278397.1, XP_006709997.1, XP_006710003.1, XP_006710023.1, XP_006710031.1, XP_006710103.1, XP_006710108.1, XP_006710128.1, XP_006710187.1, XP_006710281.1, XP_006710282.1, XP_006710283.1, XP_006710295.1, XP_006710338.1, XP_006710339.1, XP_006710340.1, XP_006710341.1, XP_006710468.1, XP_006710740.1, XP_006710742.1, XP_006711024.1, XP_006711025.1, XP_006711026.1, XP_006711027.1, XP_006711028.1, XP_006711029.1, XP_006711030.1, XP_006711031.1, XP_006711032.1, XP_006711033.1, XP_006711091.1, XP_006711092.1, XP_006711301.1, XP_006711302.1, XP_006711317.1, XP_006711318.1, XP_006711319.1, XP_006711320.1, XP_006711321.1, XP_006711456.1, XP_006711457.1, XP_006711458.1, XP_006711459.1, XP_006711546.1, XP_006711588.1, XP_006711670.1, XP_006711671.1, XP_006711672.1, XP_006711673.1, XP_006711674.1, XP_006711680.1, XP_006711681.1, XP_006711757.1, XP_006711785.1, XP_006711959.1, XP_006711988.1, XP_006711989.1, XP_006711990.1, XP_006712035.1, XP_006712036.1, XP_006712037.1, XP_006712038.1, XP_006712093.1, XP_006712094.1, XP_006712095.1, XP_006712428.1, XP_006712434.1, XP_006712436.1, XP_006712437.1, XP_006712438.1, XP_006712439.1, XP_006712440.1, XP_006712441.1, XP_006712442.1, XP_006712443.1, XP_006712444.1, XP_006712445.1, XP_006712446.1, XP_006712447.1, XP_006712448.1, XP_006712478.1, XP_006712479.1, XP_006712480.1, XP_006712516.1, XP_006712623.1, XP_006712624.1, XP_006712625.1, XP_006712626.1, XP_006712656.1, XP_006712743.1, XP_006712744.1, XP_006712745.1, XP_006712750.1, XP_006712751.1, XP_006712880.1, XP_006712881.1, XP_006712903.1, XP_006712904.1, XP_006712905.1, XP_006712906.1, XP_006712913.1, XP_006712927.1, XP_006712928.1, XP_006712929.1, XP_006712930.1, XP_006712931.1, XP_006712932.1, XP_006712950.1, XP_006713354.1, XP_006713443.1, XP_006713487.1, XP_006713488.1, XP_006713716.1, XP_006713854.1, XP_006713855.1, XP_006713856.1, XP_006713888.1, XP_006713889.1, XP_006713910.1, XP_006713928.1, XP_006713929.1, XP_006713930.1, XP_006714214.1, XP_006714215.1, XP_006714290.1, XP_006714291.1, XP_006714301.1, XP_006714312.1, XP_006714313.1, XP_006714314.1, XP_006714351.1, XP_006714408.1, XP_006714409.1, XP_006714410.1, XP_006714411.1, XP_006714412.1, XP_006714413.1, XP_006714414.1, XP_006714415.1, XP_006714416.1, XP_006714417.1, XP_006714418.1, XP_006714419.1, XP_006714420.1, XP_006714421.1, XP_006714422.1, XP_006714423.1, XP_006714424.1, XP_006714425.1, XP_006714427.1, XP_006714428.1, XP_006714429.1, XP_006714430.1, XP_006714431.1, XP_006714432.1, XP_006714433.1, XP_006714434.1, XP_006714435.1, XP_006714436.1, XP_006714437.1, XP_006714438.1, XP_006714439.1, XP_006714440.1, XP_006714441.1, XP_006714442.1, XP_006714443.1, XP_006714444.1, XP_006714445.1, XP_006714446.1, XP_006714447.1, XP_006714448.1, XP_006714449.1, XP_006714450.1, XP_006714695.1, XP_006714696.1, XP_006714697.1, XP_006714698.1, XP_006714699.1, XP_006714700.1, XP_006714701.1, XP_006714702.1, XP_006714703.1, XP_006714724.1, XP_006714725.1, XP_006714866.1, XP_006714867.1, XP_006714877.1, XP_006715054.1, XP_006715055.1, XP_006715056.1, XP_006715057.1, XP_006715058.1, XP_006715193.1, XP_006715194.1, XP_006715195.1, XP_006715196.1, XP_006715254.1, XP_006715412.1, XP_006715456.1, XP_006715457.1, XP_006715458.1, XP_006715459.1, XP_006715460.1, XP_006715461.1, XP_006715462.1, XP_006715463.1, XP_006715464.1, XP_006715665.1, XP_006715937.1, XP_006715938.1, XP_006715939.1, XP_006715940.1, XP_006715941.1, XP_006715970.1, XP_006715971.1, XP_006716063.1, XP_006716065.1, XP_006716087.1, XP_006716138.1, XP_006716157.1, XP_006716297.1, XP_006716298.1, XP_006716311.1, XP_006716378.1, XP_006716381.1, XP_006716453.1, XP_006716454.1, XP_006716475.1, XP_006716476.1, XP_006716651.1, XP_006716652.1, XP_006716653.1, XP_006716654.1, XP_006716771.1, XP_006716795.1, XP_006716828.1, XP_006716948.1, XP_006716949.1, XP_006716950.1, XP_006716951.1, XP_006716952.1, XP_006716953.1, XP_006716954.1, XP_006716955.1, XP_006716984.1, XP_006717059.1, XP_006717212.1, XP_006717336.1, XP_006717337.1, XP_006717338.1, XP_006717340.1, XP_006717355.1, XP_006717381.1, XP_006717414.1, XP_006717555.1, XP_006717556.1, XP_006717557.1, XP_006717558.1, XP_006717559.1, XP_006717788.1, XP_006717999.1, XP_006718003.1, XP_006718162.1, XP_006718185.1, XP_006718186.1, XP_006718187.1, XP_006718188.1, XP_006718189.1, XP_006718190.1, XP_006718191.1, XP_006718192.1, XP_006718193.1, XP_006718194.1, XP_006718211.1, XP_006718303.1, XP_006718305.1, XP_006718332.1, XP_006718464.1, XP_006718474.1, XP_006718479.1, XP_006718533.1, XP_006718549.1, XP_006718550.1, XP_006718619.1, XP_006718620.1, XP_006718652.1, XP_006718653.1, XP_006718749.1, XP_006718750.1, XP_006718751.1, XP_006718869.1, XP_006718878.1, XP_006718882.1, XP_006718884.1, XP_006719149.1, XP_006719280.1, XP_006719281.1, XP_006719340.1, XP_006719341.1, XP_006719342.1, XP_006719343.1, XP_006719344.1, XP_006719345.1, XP_006719346.1, XP_006719347.1, XP_006719348.1, XP_006719349.1, XP_006719350.1, XP_006719351.1, XP_006719352.1, XP_006719365.1, XP_006719366.1, XP_006719367.1, XP_006719438.1, XP_006719455.1, XP_006719456.1, XP_006719536.1, XP_006719537.1, XP_006719587.1, XP_006719588.1, XP_006719677.1, XP_006719678.1, XP_006719679.1, XP_006719738.1, XP_006719739.1, XP_006719862.1, XP_006719863.1, XP_006719907.1, XP_006719918.1, XP_006719919.1, XP_006719920.1, XP_006719921.1, XP_006719922.1, XP_006719923.1, XP_006719932.1, XP_006719933.1, XP_006720016.1, XP_006720024.1, XP_006720345.1, XP_006720384.1, XP_006720457.1, XP_006720458.1, XP_006720459.1, XP_006720460.1, XP_006720494.1, XP_006720503.1, XP_006720523.1, XP_006720535.1, XP_006720593.1, XP_006720630.1, XP_006720631.1, XP_006720644.1, XP_006720729.1, XP_006720867.1, XP_006720868.1, XP_006720869.1, XP_006720870.1, XP_006721026.1, XP_006721027.1, XP_006721028.1, XP_006721029.1, XP_006721030.1, XP_006721031.1, XP_006721032.1, XP_006721070.1, XP_006721071.1, XP_006721072.1, XP_006721073.1, XP_006721077.1, XP_006721078.1, XP_006721135.1, XP_006721211.1, XP_006721212.1, XP_006721213.1, XP_006721214.1 → *cf.* page suivante

Table S3.2. (suite)

Output proteins: 2000 new human EGZs (suite)

XP_006721215.1, XP_006721300.1, XP_006721513.1, XP_006721514.1, XP_006721531.1, XP_006721566.1, XP_006721620.1, XP_006721621.1, XP_006721622.1, XP_006721623.1, XP_006721678.1, XP_006721801.1, XP_006721886.1, XP_006721887.1, XP_006721888.1, XP_006721889.1, XP_006721927.1, XP_006721963.1, XP_006721964.1, XP_006721965.1, XP_006722059.1, XP_006722060.1, XP_006722061.1, XP_006722062.1, XP_006722063.1, XP_006722180.1, XP_006722181.1, XP_006722182.1, XP_006722183.1, XP_006722184.1, XP_006722185.1, XP_006722186.1, XP_006722548.1, XP_006722550.1, XP_006722551.1, XP_006722552.1, XP_006722553.1, XP_006722554.1, XP_006722555.1, XP_006722588.1, XP_006722679.1, XP_006722680.1, XP_006722681.1, XP_006722805.1, XP_006722891.1, XP_006722908.1, XP_006722909.1, XP_006722910.1, XP_006722929.1, XP_006722930.1, XP_006722931.1, XP_006722932.1, XP_006722933.1, XP_006722986.1, XP_006723058.1, XP_006723153.1, XP_006723243.1, XP_006723267.1, XP_006723268.1, XP_006723296.1, XP_006723297.1, XP_006723380.1, XP_006723381.1, XP_006723382.1, XP_006723398.1, XP_006723399.1, XP_006723455.1, XP_006723456.1, XP_006723457.1, XP_006723467.1, XP_006723539.1, XP_006723540.1, XP_006723541.1, XP_006723542.1, XP_006723543.1, XP_006723544.1, XP_006723545.1, XP_006723546.1, XP_006723547.1, XP_006723548.1, XP_006723549.1, XP_006723550.1, XP_006723551.1, XP_006723552.1, XP_006723553.1, XP_006723554.1, XP_006723555.1, XP_006723556.1, XP_006723565.1, XP_006723566.1, XP_006723569.1, XP_006723576.1, XP_006723579.1, XP_006723678.1, XP_006723720.1, XP_006723722.1, XP_006723810.1, XP_006723811.1, XP_006723812.1, XP_006723938.1, XP_006724039.1, XP_006724040.1, XP_006724041.1, XP_006724042.1, XP_006724228.1, XP_006724242.1, XP_006724243.1, XP_006724244.1, XP_006724245.1, XP_006724246.1, XP_006724304.1, XP_006724371.1, XP_006724376.1, XP_006724377.1, XP_006724499.1, XP_006724551.1, XP_006724571.1, XP_006724572.1, XP_006724662.1, XP_006724663.1, XP_006724664.1, XP_006724665.1, XP_006724666.1, XP_006724667.1, XP_006724780.1, XP_006724781.1, XP_006725156.1, XP_006725158.1, XP_006725222.1, XP_006725224.1, XP_006725326.1, XP_006725327.1, XP_006725419.1, XP_006725420.1, XP_006725421.1, XP_006725422.1, XP_006725423.1, XP_006725424.1, XP_006725425.1, XP_006725455.1, XP_006725470.1, XP_006725552.1, XP_006725553.1, XP_006725554.1, XP_006725555.1, XP_006725570.1, XP_006725571.1, XP_006725766.1, XP_006725767.1, XP_006725768.1, XP_006725769.1, XP_006725786.1, XP_006725787.1, XP_006725844.1, XP_006725845.1, XP_006725846.1, XP_006725883.1, XP_006725884.1, XP_006725885.1, XP_006725886.1, XP_006725896.1, XP_006725897.1, XP_006725975.1, XP_006725976.1, XP_006725977.1, XP_006725978.1, XP_006725985.1, XP_006726068.1, XP_006726069.1, XP_006726070.1, XP_006726071.1, XP_006726086.1, XP_006726087.1, XP_006726163.1, XP_006726164.1, XP_006726165.1, XP_006726166.1, XP_006726174.1, XP_006726175.1, XP_006726387.1, XP_006726409.1, XP_006726523.1, XP_006726570.1, XP_006726571.1, XP_006726572.1, XP_006726577.1, XP_006726581.1, XP_006726640.1, XP_006726641.1, XP_006726840.1, XP_006726901.1

(h) *H. sapiens* EGZs against the *M. musculus* proteome → Figure S3.13

Input proteins: 8 known human EGZs

NP_000521.2, NP_004354.2, NP_005579.2, NP_060389.2, NP_079045.1, NP_714928.1, NP_001073592.1, NP_001129602.1

Output proteins: 1917 new murine EGZs

NP_031406.2, NP_031465.2, NP_031497.2, NP_031499.1, NP_031593.2, NP_031604.3, NP_031618.2, NP_031716.2, NP_031750.2, NP_031758.2, NP_031852.1, NP_031861.2, NP_031866.2, NP_031905.1, NP_031907.2, NP_031916.1, NP_031951.2, NP_032039.2, NP_032119.2, NP_032266.3, NP_032268.2, NP_032285.2, NP_032370.2, NP_032419.2, NP_032478.2, NP_032480.1, NP_032497.1, NP_032500.2, NP_032501.2, NP_032588.1, NP_032611.2, NP_032618.2, NP_032649.2, NP_032722.2, NP_032739.3, NP_032812.2, NP_033008.3, NP_033011.2, NP_033076.2, NP_033127.1, NP_033135.2, NP_033157.1, NP_033175.1, NP_033233.2, NP_033251.1, NP_033340.3, NP_033350.2, NP_033400.2, NP_033411.3, NP_033469.1, NP_033526.1, NP_033545.1, NP_033598.2, NP_033605.1, NP_033869.1, NP_034006.3, NP_034016.2, NP_034058.2, NP_034159.1, NP_034175.2, NP_034225.1, NP_034232.1, NP_034326.1, NP_034556.2, NP_034577.1, NP_034691.2, NP_034758.2, NP_034792.1, NP_034793.1, NP_034797.1, NP_034799.2, NP_034803.2, NP_034805.1, NP_034861.3, NP_034918.1, NP_034950.2, NP_034951.1, NP_034992.2, NP_035077.1, NP_035195.2, NP_035247.2, NP_035289.2, NP_035300.1, NP_035304.4, NP_035524.2, NP_035546.2, NP_035547.2, NP_035657.1, NP_035695.1, NP_035701.3, NP_035847.2, NP_035921.2, NP_035927.1, NP_035944.2, NP_035946.2, NP_035971.1, NP_036056.2, NP_036075.2, NP_038488.4, NP_038505.2, NP_038535.2, NP_038662.2, NP_038698.1, NP_038744.1, NP_038763.3, NP_038774.1, NP_038785.1, NP_038825.1, NP_038864.3, NP_038887.2, NP_038951.1, NP_056640.1, NP_056641.2, NP_058040.2, NP_058085.2, NP_058575.2, NP_058593.2, NP_058655.3, NP_059064.2, NP_059087.2, NP_059088.2, NP_059091.2, NP_059507.2, NP_061211.1, NP_061245.3, NP_061249.1, NP_061263.2, NP_061278.1, NP_061293.2, NP_061361.2, NP_062307.2, NP_062387.2, NP_062421.2, NP_062510.2, NP_062528.2, NP_062663.1, NP_062748.2, NP_062784.3, NP_062792.2, NP_063920.2, NP_063933.1, NP_064311.2, NP_064328.2, NP_064656.2, NP_065239.1, NP_065262.1, NP_065264.2, NP_065650.1, NP_066302.2, NP_067267.2, NP_067321.2, NP_067329.1, NP_067411.1, NP_068360.4, NP_068677.2, NP_075023.2, NP_075355.1, NP_075553.2, NP_075708.2, NP_075817.2, NP_076092.1, NP_076207.1, NP_076331.2, NP_076397.2, NP_077128.2, NP_077235.2, NP_077800.3, NP_079600.1, NP_079696.1, NP_079849.1, NP_079873.1, NP_079906.2, NP_079958.3, NP_080218.1, NP_080222.1, NP_080249.2, NP_080379.1, NP_080397.1, NP_080593.1, NP_080611.1, NP_080614.1, NP_080619.2, NP_080708.3, NP_080717.2, NP_080769.3, NP_080889.3, NP_081118.4, NP_081134.2, NP_081518.1, NP_081665.2, NP_081703.1, NP_081931.1, NP_082003.1, NP_082158.2, NP_082176.4, NP_082260.1, NP_082272.1, NP_082317.1, NP_082327.2, NP_082368.1, NP_082442.1, NP_082460.2, NP_082567.3, NP_082722.1, NP_082730.2, NP_082810.1, NP_082897.2, NP_082903.2, NP_082985.2, NP_083045.4, NP_083120.2, NP_083142.2, NP_083147.1, NP_083325.1, NP_083507.3, NP_083550.2, NP_083625.1, NP_083669.1, NP_083723.1, NP_083805.1, NP_083896.1, NP_083964.2, NP_084080.1, NP_084148.1, NP_084163.2, NP_084209.3, NP_084367.1, NP_084388.2, NP_084457.2, NP_084504.2, NP_084532.2, NP_084539.2, NP_109612.1, NP_109647.2, NP_109648.2, NP_112539.2, NP_113571.2, NP_113687.2, NP_114084.2, NP_149063.2, NP_203505.3, NP_443720.1, NP_444324.2, NP_444354.2, NP_444493.1, NP_542374.2, NP_569715.3, NP_570931.1, NP_573447.2, NP_573476.2, NP_573500.2, NP_573517.1, NP_598459.2, NP_598484.2, NP_598487.1, NP_598494.2, NP_598508.2, NP_598590.2, NP_598644.1, NP_598718.2, NP_598838.3, NP_598863.3, NP_612178.2, NP_619594.1, NP_619618.1, NP_620084.2, NP_620570.2, NP_653108.2, NP_653146.1, NP_659058.2, NP_659061.2, NP_659070.2, NP_659074.2, NP_659155.1, NP_660262.2, NP_663328.2, NP_663434.2, NP_663478.2, NP_663491.1, NP_663498.1, NP_663526.3, NP_663531.1, NP_665831.1, NP_666098.2, NP_666109.2, NP_666112.1, NP_666132.1, NP_666175.1, NP_666242.2, NP_666287.3, NP_666310.1, NP_689420.1, NP_694694.1, NP_694708.2, NP_694793.1, NP_694804.3, NP_694816.3, NP_703189.3, NP_705728.1, NP_705734.4, NP_705761.2, NP_705793.1, NP_705797.1, NP_705805.1, NP_705811.8, NP_722482.1, NP_733479.1, NP_739565.2, NP_758472.2, NP_758500.2, NP_758512.3, NP_766034.2, NP_766084.3, NP_766111.1, NP_766137.2, NP_766284.2, NP_766338.1, NP_766396.2, NP_766464.1, NP_766470.2, NP_766527.3, NP_766567.1, NP_766625.1, NP_775279.2, NP_775545.1, NP_775552.1, NP_775578.2, NP_775613.2, NP_775615.2, NP_776099.1, NP_777363.1, NP_777364.3, NP_778194.3, NP_780421.2, NP_780520.3, NP_780538.2, NP_780668.3, NP_780730.2, NP_780760.2, NP_783615.2, NP_786927.2, NP_796098.3, NP_796158.2, NP_796182.2, NP_796237.2, NP_796286.2, NP_796363.2, NP_808270.2, NP_808310.2, NP_808311.1, NP_808328.2, NP_808376.1, NP_808394.1, NP_808411.1, NP_808489.4, NP_835179.1, NP_835182.1, NP_839990.2, NP_848745.2, NP_848822.2, NP_849205.1, NP_849231.1, NP_849257.1, NP_849265.2, NP_853624.1, NP_853625.1, NP_874357.2, NP_898911.2, NP_898949.2, NP_898970.1, NP_899022.2, NP_899071.2, NP_899248.1, NP_906271.1, NP_919324.2, NP_922936.3, NP_932139.2, NP_932758.1, NP_932775.2, NP_941020.1, NP_941033.2, NP_941037.2, NP_941955.1, NP_957707.2, NP_958763.1, NP_958787.2, NP_958788.2, NP_958789.3, NP_958790.2, NP_958791.2, NP_958792.2, NP_958793.2, NP_958794.2, NP_958795.2, NP_958796.2, NP_996738.2, NP_996992.1, NP_997145.2, NP_997554.2, NP_997597.1, NP_997603.3, NP_998893.1, XP_355890.5, XP_485502.1, NP_001001489.1, NP_001001884.1, NP_001002272.1, NP_001003667.1, NP_001003668.2, NP_001003670.1, NP_001004177.2, NP_001005511.2, NP_001005784.1, NP_001011874.1, NP_001013409.1, NP_001013826.1, NP_001017955.2, NP_001020563.1, NP_001025150.2, NP_001028349.1, NP_001028399.1, NP_001028440.2, NP_001028448.3, NP_001028471.1, NP_001028620.1, NP_001028805.2, NP_001029287.1, NP_001032814.1, NP_001033046.1, NP_001033699.1 → *cf.* page suivante

Table S3.2. (suite)

Output proteins: 1917 new murine EGZs (suite)

NP_001033732.1, NP_001034127.1, NP_001034218.1, NP_001034274.1, NP_001034275.1, NP_001034276.1, NP_001035028.1, NP_001035174.1, NP_001035175.1, NP_001035176.1, NP_001035177.1, NP_001035221.1, NP_001035489.1, NP_001035549.1, NP_001035903.1, NP_001036125.1, NP_001036172.1, NP_001036173.1, NP_001036174.1, NP_001070660.1, NP_001073587.1, NP_001073853.1, NP_001073854.1, NP_001073855.1, NP_001074240.1, NP_001074249.1, NP_001074263.1, NP_001074264.1, NP_001074265.1, NP_001074266.1, NP_001074504.1, NP_001074510.1, NP_001074514.1, NP_001074553.1, NP_001074561.1, NP_001074672.1, NP_001074674.1, NP_001074719.1, NP_001074724.1, NP_001074759.1, NP_001074787.1, NP_001074839.2, NP_001075124.1, NP_001075448.1, NP_001075450.1, NP_001076430.1, NP_001076431.1, NP_001078845.1, NP_001078854.1, NP_001093098.1, NP_001096033.1, NP_001096641.1, NP_001103213.1, NP_001103383.1, NP_001103601.1, NP_001103602.1, NP_001103671.1, NP_001104518.1, NP_001104784.1, NP_001106844.2, NP_001106856.1, NP_001106859.1, NP_001106860.1, NP_001107612.1, NP_001116075.1, NP_001121558.2, NP_001121642.1, NP_001121643.1, NP_001123659.1, NP_001123997.1, NP_001123998.1, NP_001129129.1, NP_001129709.1, NP_001135437.1, NP_001139116.1, NP_001139357.1, NP_001139439.1, NP_001139670.1, NP_001152839.1, NP_001153008.1, NP_001153009.1, NP_001153031.1, NP_001153075.1, NP_001153076.1, NP_001153077.1, NP_001153078.1, NP_001153155.1, NP_001153458.1, NP_001153691.1, NP_001153692.1, NP_001153693.1, NP_001153694.1, NP_001153723.1, NP_001153872.1, NP_001154810.1, NP_001155096.1, NP_001155137.1, NP_001155890.1, NP_001156005.1, NP_001156442.1, NP_001156654.1, NP_001156655.1, NP_001156656.1, NP_001156657.1, NP_001156658.1, NP_001156659.1, NP_001156999.1, NP_001157000.1, NP_001157012.1, NP_001157014.1, NP_001157021.1, NP_001157041.1, NP_001157227.1, NP_001157229.1, NP_001157230.1, NP_001157290.1, NP_001157675.1, NP_001157688.1, NP_001157689.1, NP_001158181.1, NP_001159424.1, NP_001159629.1, NP_001159946.1, NP_001161468.1, NP_001161764.1, NP_001161765.1, NP_001163866.1, NP_001164256.1, NP_001164257.1, NP_001164318.1, NP_001167549.1, NP_001167550.1, NP_001170860.1, NP_001171320.1, NP_001171321.1, NP_001171322.1, NP_001171323.1, NP_001171324.1, NP_001171325.1, NP_001171326.1, NP_001171327.1, NP_001171372.1, NP_001177335.1, NP_001177337.1, NP_001185752.1, NP_001185753.1, NP_001185754.1, NP_001185755.1, NP_001185795.1, NP_001185796.1, NP_001186153.1, NP_001186263.1, NP_001186856.1, NP_001188252.1, NP_001189429.2, NP_001192155.1, NP_001192170.1, NP_001192241.1, NP_001193780.1, NP_001229301.1, NP_001230766.1, NP_001231820.1, NP_001231832.1, NP_001238819.1, NP_001238988.1, NP_001238989.1, NP_001238990.1, NP_001239010.1, NP_001239011.1, NP_001239129.1, NP_001239130.1, NP_001239131.1, NP_001239387.1, NP_001240751.1, NP_001240796.1, NP_001240821.1, NP_001241653.1, NP_001254636.1, NP_001257575.1, NP_001258299.1, NP_001263208.1, NP_001263250.1, NP_001263316.1, NP_001263394.1, NP_001263431.1, NP_001264050.1, NP_001264051.1, NP_001264052.1, NP_001264250.1, NP_001264251.1, NP_001265185.1, NP_001268980.1, NP_001268981.1, NP_001269922.1, NP_001269923.1, NP_001271148.1, NP_001271150.1, NP_001272427.1, NP_001272429.1, NP_001272721.1, NP_001272723.1, NP_001272781.1, NP_001273189.1, NP_001274093.1, NP_001274094.1, NP_001274095.1, NP_001274096.1, XP_003689232.1, XP_006495542.1, XP_006495543.1, XP_006495544.1, XP_006495545.1, XP_006495546.1, XP_006495547.1, XP_006495549.1, XP_006495550.1, XP_006495551.1, XP_006495613.1, XP_006495961.1, XP_006496104.1, XP_006496105.1, XP_006496106.1, XP_006496107.1, XP_006496108.1, XP_006496109.1, XP_006496110.1, XP_006496282.1, XP_006496425.1, XP_006496426.1, XP_006496437.1, XP_006496438.1, XP_006496507.1, XP_006496508.1, XP_006496757.1, XP_006496829.1, XP_006496830.1, XP_006496831.1, XP_006496832.1, XP_006496833.1, XP_006496834.1, XP_006496835.1, XP_006496885.1, XP_006496886.1, XP_006496887.1, XP_006496888.1, XP_006497107.1, XP_006497108.1, XP_006497109.1, XP_006497110.1, XP_006497111.1, XP_006497112.1, XP_006497221.1, XP_006497680.1, XP_006497683.1, XP_006497917.1, XP_006497918.1, XP_006497919.1, XP_006497920.1, XP_006497921.1, XP_006498008.1, XP_006498009.1, XP_006498564.1, XP_006498565.1, XP_006498566.1, XP_006498567.1, XP_006498568.1, XP_006498651.1, XP_006498723.1, XP_006498724.1, XP_006498725.1, XP_006498726.1, XP_006498727.1, XP_006498728.1, XP_006498729.1, XP_006498730.1, XP_006498731.1, XP_006498732.1, XP_006498935.1, XP_006499054.1, XP_006499055.1, XP_006499056.1, XP_006499057.1, XP_006499121.1, XP_006499122.1, XP_006499214.1, XP_006499393.1, XP_006499490.1, XP_006499491.1, XP_006499492.1, XP_006499493.1, XP_006499494.1, XP_006499495.1, XP_006499496.1, XP_006499497.1, XP_006499565.1, XP_006499566.1, XP_006499567.1, XP_006499568.1, XP_006499660.1, XP_006499724.1, XP_006499725.1, XP_006499726.1, XP_006499727.1, XP_006499728.1, XP_006499729.1, XP_006499730.1, XP_006499731.1, XP_006499732.1, XP_006500318.1, XP_006500357.1, XP_006500358.1, XP_006500572.1, XP_006500573.1, XP_006500574.1, XP_006500575.1, XP_006500600.1, XP_006500617.1, XP_006500651.1, XP_006500671.1, XP_006500672.1, XP_006500673.1, XP_006500674.1, XP_006500675.1, XP_006500676.1, XP_006500677.1, XP_006500678.1, XP_006500679.1, XP_006500834.1, XP_006500835.1, XP_006500836.1, XP_006500852.1, XP_006501151.1, XP_006501505.1, XP_006501506.1, XP_006501754.1, XP_006501755.1, XP_006501756.1, XP_006501757.1, XP_006501758.1, XP_006501759.1, XP_006501760.1, XP_006501761.1, XP_006501762.1, XP_006501763.1, XP_006501940.1, XP_006502049.1, XP_006502050.1, XP_006502051.1, XP_006502052.1, XP_006502300.1, XP_006502301.1, XP_006502302.1, XP_006502303.1, XP_006502304.1, XP_006502305.1, XP_006502306.1, XP_006502307.1, XP_006502308.1, XP_006502309.1, XP_006502310.1, XP_006502311.1, XP_006502312.1, XP_006502351.1, XP_006502352.1, XP_006502353.1, XP_006502354.1, XP_006502424.1, XP_006502637.1, XP_006502679.1, XP_006502762.1, XP_006502763.1, XP_006502764.1, XP_006502765.1, XP_006502766.1, XP_006502767.1, XP_006502768.1, XP_006502769.1, XP_006502972.1, XP_006502973.1, XP_006502974.1, XP_006502975.1, XP_006502976.1, XP_006502977.1, XP_006502978.1, XP_006502979.1, XP_006503063.1, XP_006503064.1, XP_006503074.1, XP_006503134.1, XP_006503135.1, XP_006503136.1, XP_006503156.1, XP_006503157.1, XP_006503166.1, XP_006503167.1, XP_006503354.1, XP_006503355.1, XP_006503357.1, XP_006503438.1, XP_006503468.1, XP_006503469.1, XP_006503490.1, XP_006503491.1, XP_006503492.1, XP_006503493.1, XP_006503720.1, XP_006504105.1, XP_006504106.1, XP_006504183.1, XP_006504184.1, XP_006504185.1, XP_006504186.1, XP_006504187.1, XP_006504423.1, XP_006504563.1, XP_006504564.1, XP_006504638.1, XP_006504639.1, XP_006504640.1, XP_006504678.1, XP_006504931.1, XP_006504932.1, XP_006504933.1, XP_006504934.1, XP_006504935.1, XP_006504936.1, XP_006505128.1, XP_006505129.1, XP_006505132.1, XP_006505280.1, XP_006505416.1, XP_006505417.1, XP_006505418.1, XP_006505419.1, XP_006505420.1, XP_006505421.1, XP_006505422.1, XP_006505423.1, XP_006505424.1, XP_006505425.1, XP_006505548.1, XP_006505549.1, XP_006505550.1, XP_006505551.1, XP_006505552.1, XP_006505553.1, XP_006505554.1, XP_006505555.1, XP_006505556.1, XP_006505557.1, XP_006505558.1, XP_006505809.1, XP_006506094.1, XP_006506300.1, XP_006506302.1, XP_006506511.1, XP_006506588.1, XP_006506589.1, XP_006506794.1, XP_006506795.1, XP_006506796.1, XP_006506797.1, XP_006506863.1, XP_006506920.1, XP_006506921.1, XP_006507021.1, XP_006507022.1, XP_006507023.1, XP_006507024.1, XP_006507025.1, XP_006507027.1, XP_006507028.1, XP_006507649.1, XP_006507774.1, XP_006507775.1, XP_006507776.1, XP_006507777.1, XP_006507783.1, XP_006507784.1, XP_006507787.1, XP_006507789.1, XP_006507790.1, XP_006507791.1, XP_006507825.1, XP_006507826.1, XP_006507828.1, XP_006507958.1, XP_006507959.1, XP_006507969.1, XP_006508146.1, XP_006508147.1, XP_006508237.1, XP_006508418.1, XP_006508419.1, XP_006508420.1, XP_006508421.1, XP_006508422.1, XP_006508424.1, XP_006508425.1, XP_006508426.1, XP_006508508.1, XP_006508530.1, XP_006508531.1, XP_006508532.1, XP_006508533.1, XP_006508534.1, XP_006508582.1, XP_006508583.1, XP_006508584.1, XP_006508585.1, XP_006508586.1, XP_006508587.1, XP_006508588.1, XP_006508590.1, XP_006508591.1, XP_006508592.1, XP_006508593.1, XP_006508594.1, XP_006508595.1, XP_006508596.1, XP_006508597.1, XP_006508598.1, XP_006508599.1, XP_006508712.1, XP_006508713.1, XP_006508714.1, XP_006508715.1, XP_006509106.1, XP_006509107.1, XP_006509597.1, XP_006509598.1, XP_006509599.1, XP_006509673.1, XP_006509674.1, XP_006509675.1, XP_006509676.1, XP_006509677.1, XP_006509678.1, XP_006509679.1, XP_006509680.1, XP_006509681.1, XP_006509682.1, XP_006509683.1, XP_006509684.1, XP_006509685.1, XP_006509686.1, XP_006510092.1, XP_006510093.1, XP_006510094.1, XP_006510095.1, XP_006510096.1, XP_006510097.1, XP_006510098.1, XP_006510099.1, XP_006510100.1, XP_006510180.1, XP_006510181.1, XP_006510182.1, XP_006510183.1, XP_006510184.1, XP_006510185.1, XP_006510247.1, XP_006510248.1, XP_006510555.1, XP_006510556.1, XP_006510557.1, XP_006510558.1, XP_006510569.1, XP_006510665.1, XP_006510666.1, XP_006510762.1, XP_006510763.1, XP_006510764.1, XP_006510765.1, XP_006510766.1, XP_006510767.1, XP_006510768.1, XP_006510787.1, XP_006510972.1, XP_006510973.1, XP_006511287.1, XP_006511288.1, XP_006511289.1, XP_006511291.1, XP_006511298.1, XP_006511299.1, XP_006511300.1, XP_006511301.1, XP_006511302.1, XP_006511303.1, XP_006511304.1, XP_006511354.1, XP_006511355.1, XP_006511697.1, XP_006512004.1, XP_006512005.1, XP_006512006.1, XP_006512007.1, XP_006512008.1, XP_006512193.1, XP_006512194.1, XP_006512195.1, XP_006512512.1, XP_006512513.1, XP_006512651.1, XP_006512652.1, XP_006512653.1, XP_006512659.1, XP_006512836.1, XP_006512837.1, XP_006512838.1, XP_006512929.1, XP_006513056.1, XP_006513082.1, XP_006513083.1, XP_006513084.1, XP_006513085.1, XP_006513138.1, XP_006513342.1, XP_006513343.1, XP_006513344.1, XP_006513345.1, XP_006513508.1, XP_006513555.1, XP_006513556.1, XP_006513557.1, XP_006513558.1, XP_006513559.1, XP_006513560.1, XP_006513738.1, XP_006513739.1, XP_006513740.1, XP_006513741.1, XP_006513742.1, XP_006513743.1, XP_006513832.1, XP_006513854.1, XP_006513855.1, XP_006513856.1, XP_006513857.1, XP_006513858.1, XP_006513914.1, XP_006513915.1, XP_006513916.1, XP_006513917.1, XP_006513918.1 → *cf.* page suivante

Table S3.2. (suite)

Output proteins: 1917 new murine EGZs (suite)

XP_006513919.1, XP_006513920.1, XP_006513921.1, XP_006513958.1, XP_006514007.1, XP_006514008.1, XP_006514009.1, XP_006514036.1, XP_006514100.1, XP_006514101.1, XP_006514215.1, XP_006514216.1, XP_006514217.1, XP_006514218.1, XP_006514219.1, XP_006514220.1, XP_006514221.1, XP_006514222.1, XP_006514223.1, XP_006514224.1, XP_006514225.1, XP_006514226.1, XP_006514227.1, XP_006514235.1, XP_006514236.1, XP_006514293.1, XP_006514364.1, XP_006514365.1, XP_006514366.1, XP_006514367.1, XP_006514368.1, XP_006514369.1, XP_006514370.1, XP_006514371.1, XP_006514372.1, XP_006514373.1, XP_006514374.1, XP_006514544.1, XP_006514545.1, XP_006514546.1, XP_006514547.1, XP_006514548.1, XP_006514768.1, XP_006514769.1, XP_006514770.1, XP_006514974.1, XP_006514995.1, XP_006514996.1, XP_006514997.1, XP_006514998.1, XP_006515178.1, XP_006515179.1, XP_006515180.1, XP_006515340.1, XP_006515341.1, XP_006515342.1, XP_006515343.1, XP_006515344.1, XP_006515345.1, XP_006515346.1, XP_006515347.1, XP_006515348.1, XP_006515349.1, XP_006515445.1, XP_006515446.1, XP_006515542.1, XP_006515543.1, XP_006515544.1, XP_006515545.1, XP_006515546.1, XP_006515720.1, XP_006515721.1, XP_006515722.1, XP_006515723.1, XP_006515724.1, XP_006515725.1, XP_006515726.1, XP_006515727.1, XP_006515853.1, XP_006515854.1, XP_006516004.1, XP_006516005.1, XP_006516006.1, XP_006516595.1, XP_006517398.1, XP_006517399.1, XP_006517400.1, XP_006517401.1, XP_006517402.1, XP_006517403.1, XP_006517405.1, XP_006517406.1, XP_006517587.1, XP_006517588.1, XP_006517589.1, XP_006517694.1, XP_006517695.1, XP_006517727.1, XP_006517728.1, XP_006517729.1, XP_006517730.1, XP_006517817.1, XP_006517845.1, XP_006517846.1, XP_006517869.1, XP_006518572.1, XP_006518582.1, XP_006519027.1, XP_006519028.1, XP_006519029.1, XP_006519085.1, XP_006519086.1, XP_006519087.1, XP_006519097.1, XP_006519098.1, XP_006519099.1, XP_006519100.1, XP_006519101.1, XP_006519102.1, XP_006519160.1, XP_006519326.1, XP_006519327.1, XP_006519328.1, XP_006519329.1, XP_006519330.1, XP_006519331.1, XP_006519332.1, XP_006519333.1, XP_006519334.1, XP_006519456.1, XP_006519649.1, XP_006519650.1, XP_006519992.1, XP_006519993.1, XP_006520442.1, XP_006520443.1, XP_006520444.1, XP_006520577.1, XP_006520578.1, XP_006520751.1, XP_006520752.1, XP_006520753.1, XP_006520754.1, XP_006520755.1, XP_006520756.1, XP_006520757.1, XP_006520788.1, XP_006520789.1, XP_006520790.1, XP_006520791.1, XP_006520792.1, XP_006520793.1, XP_006520794.1, XP_006520795.1, XP_006520796.1, XP_006520948.1, XP_006521152.1, XP_006521153.1, XP_006521154.1, XP_006521185.1, XP_006521186.1, XP_006521203.1, XP_006521204.1, XP_006521229.1, XP_006521230.1, XP_006521231.1, XP_006521281.1, XP_006521284.1, XP_006521454.1, XP_006521455.1, XP_006521567.1, XP_006521920.1, XP_006521921.1, XP_006521922.1, XP_006521923.1, XP_006522218.1, XP_006522296.1, XP_006522297.1, XP_006522299.1, XP_006522587.1, XP_006522588.1, XP_006522589.1, XP_006522590.1, XP_006522591.1, XP_006522592.1, XP_006522593.1, XP_006522594.1, XP_006522595.1, XP_006522596.1, XP_006522597.1, XP_006522834.1, XP_006522835.1, XP_006522836.1, XP_006522837.1, XP_006522936.1, XP_006522951.1, XP_006523196.1, XP_006523219.1, XP_006523387.1, XP_006523472.1, XP_006523568.1, XP_006523814.1, XP_006524020.1, XP_006524050.1, XP_006524051.1, XP_006524052.1, XP_006524053.1, XP_006524054.1, XP_006524116.1, XP_006524117.1, XP_006524118.1, XP_006524119.1, XP_006524120.1, XP_006524121.1, XP_006524122.1, XP_006524123.1, XP_006524124.1, XP_006524125.1, XP_006524126.1, XP_006524127.1, XP_006524128.1, XP_006524129.1, XP_006524130.1, XP_006524131.1, XP_006524134.1, XP_006524136.1, XP_006524219.1, XP_006524220.1, XP_006524221.1, XP_006524222.1, XP_006524396.1, XP_006524440.1, XP_006524441.1, XP_006524679.1, XP_006524680.1, XP_006524681.1, XP_006524682.1, XP_006524726.1, XP_006524727.1, XP_006524728.1, XP_006524729.1, XP_006524731.1, XP_006524732.1, XP_006524733.1, XP_006525040.1, XP_006525157.1, XP_006525240.1, XP_006525241.1, XP_006525267.1, XP_006525268.1, XP_006525269.1, XP_006525270.1, XP_006525271.1, XP_006525272.1, XP_006525273.1, XP_006525274.1, XP_006525275.1, XP_006525276.1, XP_006525277.1, XP_006525278.1, XP_006525279.1, XP_006525280.1, XP_006525281.1, XP_006525282.1, XP_006525283.1, XP_006525284.1, XP_006525504.1, XP_006525505.1, XP_006525766.1, XP_006525844.1, XP_006525845.1, XP_006525846.1, XP_006525847.1, XP_006525848.1, XP_006525849.1, XP_006525850.1, XP_006525851.1, XP_006525852.1, XP_006525853.1, XP_006525854.1, XP_006525855.1, XP_006525856.1, XP_006525857.1, XP_006525858.1, XP_006525859.1, XP_006525860.1, XP_006525861.1, XP_006525862.1, XP_006525863.1, XP_006525864.1, XP_006525938.1, XP_006526020.1, XP_006526021.1, XP_006526022.1, XP_006526023.1, XP_006526024.1, XP_006526025.1, XP_006526026.1, XP_006526064.1, XP_006526065.1, XP_006526066.1, XP_006526067.1, XP_006526279.1, XP_006526280.1, XP_006526281.1, XP_006526282.1, XP_006526364.1, XP_006526499.1, XP_006526689.1, XP_006526690.1, XP_006526691.1, XP_006526692.1, XP_006526797.1, XP_006527047.1, XP_006527048.1, XP_006527049.1, XP_006527067.1, XP_006527344.1, XP_006527345.1, XP_006527346.1, XP_006527347.1, XP_006527348.1, XP_006527448.1, XP_006527449.1, XP_006527450.1, XP_006527451.1, XP_006527452.1, XP_006527453.1, XP_006527454.1, XP_006527455.1, XP_006527474.1, XP_006527478.1, XP_006527610.1, XP_006527611.1, XP_006527612.1, XP_006527824.1, XP_006527825.1, XP_006527951.1, XP_006528428.1, XP_006528531.1, XP_006528532.1, XP_006528533.1, XP_006528534.1, XP_006528535.1, XP_006528536.1, XP_006528614.1, XP_006528693.1, XP_006528694.1, XP_006528979.1, XP_006528980.1, XP_006528981.1, XP_006528996.1, XP_006528997.1, XP_006528998.1, XP_006528999.1, XP_006529000.1, XP_006529001.1, XP_006529002.1, XP_006529003.1, XP_006529004.1, XP_006529005.1, XP_006529006.1, XP_006529007.1, XP_006529008.1, XP_006529009.1, XP_006529010.1, XP_006529011.1, XP_006529013.1, XP_006529371.1, XP_006529372.1, XP_006529432.1, XP_006529433.1, XP_006529434.1, XP_006529435.1, XP_006529436.1, XP_006529437.1, XP_006529438.1, XP_006529439.1, XP_006529440.1, XP_006529441.1, XP_006529442.1, XP_006529443.1, XP_006529501.1, XP_006529502.1, XP_006529503.1, XP_006529504.1, XP_006529507.1, XP_006529508.1, XP_006529564.1, XP_006529565.1, XP_006529566.1, XP_006529567.1, XP_006529568.1, XP_006529569.1, XP_006529570.1, XP_006529571.1, XP_006529574.1, XP_006529856.1, XP_006529915.1, XP_006529916.1, XP_006529917.1, XP_006530071.1, XP_006530072.1, XP_006530073.1, XP_006530074.1, XP_006530084.1, XP_006530380.1, XP_006530381.1, XP_006530382.1, XP_006530383.1, XP_006530384.1, XP_006530385.1, XP_006530386.1, XP_006530387.1, XP_006530388.1, XP_006530389.1, XP_006530650.1, XP_006530651.1, XP_006530652.1, XP_006530653.1, XP_006530654.1, XP_006530655.1, XP_006530656.1, XP_006530657.1, XP_006530658.1, XP_006530659.1, XP_006530660.1, XP_006530661.1, XP_006530662.1, XP_006530663.1, XP_006530664.1, XP_006530665.1, XP_006530666.1, XP_006530667.1, XP_006530668.1, XP_006530669.1, XP_006530670.1, XP_006530671.1, XP_006530672.1, XP_006530673.1, XP_006530674.1, XP_006530682.1, XP_006530683.1, XP_006530684.1, XP_006530818.1, XP_006530819.1, XP_006530820.1, XP_006530821.1, XP_006530824.1, XP_006530825.1, XP_006530826.1, XP_006530827.1, XP_006530933.1, XP_006530934.1, XP_006530935.1, XP_006530936.1, XP_006530937.1, XP_006530938.1, XP_006530964.1, XP_006530965.1, XP_006530966.1, XP_006530967.1, XP_006530968.1, XP_006531125.1, XP_006531254.1, XP_006531255.1, XP_006531256.1, XP_006531257.1, XP_006531258.1, XP_006531259.1, XP_006531260.1, XP_006531432.1, XP_006531477.1, XP_006531478.1, XP_006531479.1, XP_006531480.1, XP_006531481.1, XP_006531482.1, XP_006531483.1, XP_006531484.1, XP_006531485.1, XP_006531846.1, XP_006531847.1, XP_006531913.1, XP_006532040.1, XP_006532041.1, XP_006532042.1, XP_006532043.1, XP_006532198.1, XP_006532199.1, XP_006532200.1, XP_006532216.1, XP_006532243.1, XP_006532244.1, XP_006532245.1, XP_006532332.1, XP_006532482.1, XP_006532483.1, XP_006532484.1, XP_006532485.1, XP_006532486.1, XP_006532487.1, XP_006532488.1, XP_006532489.1, XP_006532490.1, XP_006532491.1, XP_006532543.1, XP_006532544.1, XP_006532565.1, XP_006532632.1, XP_006532633.1, XP_006532634.1, XP_006532635.1, XP_006532709.1, XP_006533020.1, XP_006533021.1, XP_006533022.1, XP_006533023.1, XP_006533172.1, XP_006533173.1, XP_006533174.1, XP_006533288.1, XP_006533289.1, XP_006533340.1, XP_006533341.1, XP_006533342.1, XP_006533343.1, XP_006533344.1, XP_006533345.1, XP_006533346.1, XP_006533347.1, XP_006533348.1, XP_006533349.1, XP_006533507.1, XP_006533825.1, XP_006533826.1, XP_006533827.1, XP_006533828.1, XP_006533829.1, XP_006533830.1, XP_006533831.1, XP_006533832.1, XP_006533833.1, XP_006533834.1, XP_006533835.1, XP_006533836.1, XP_006533837.1, XP_006533838.1, XP_006533839.1, XP_006533840.1, XP_006533841.1, XP_006533842.1, XP_006533843.1, XP_006533844.1, XP_006533845.1, XP_006533846.1, XP_006533854.1, XP_006533924.1, XP_006533925.1, XP_006533926.1, XP_006533931.1, XP_006533932.1, XP_006533933.1, XP_006533935.1, XP_006533936.1, XP_006534199.1, XP_006534200.1, XP_006534203.1, XP_006534204.1, XP_006534246.1, XP_006534247.1, XP_006534248.1, XP_006534249.1, XP_006534250.1, XP_006534329.1, XP_006534330.1, XP_006534446.1, XP_006534447.1, XP_006534508.1, XP_006534509.1, XP_006534510.1, XP_006534511.1, XP_006534512.1, XP_006534513.1, XP_006534514.1, XP_006534515.1, XP_006534516.1, XP_006534517.1, XP_006534586.1, XP_006534587.1, XP_006534615.1, XP_006534616.1, XP_006534815.1, XP_006534968.1, XP_006534969.1, XP_006534970.1, XP_006534990.1, XP_006534991.1, XP_006535019.1, XP_006535020.1, XP_006535283.1, XP_006535284.1, XP_006535285.1, XP_006535286.1, XP_006535287.1, XP_006535288.1, XP_006535289.1, XP_006535290.1, XP_006535291.1, XP_006535627.1, XP_006535713.1, XP_006535714.1, XP_006535715.1, XP_006535716.1, XP_006535717.1, XP_006535718.1, XP_006535719.1, XP_006536003.1, XP_006536720.1, XP_006536767.1, XP_006536772.1, XP_006536773.1, XP_006536829.1, XP_006536830.1, XP_006537094.1, XP_006537095.1, XP_006537096.1, XP_006537097.1, XP_006537098.1, XP_006537099.1, XP_006537100.1, XP_006537101.1, XP_006537102.1, XP_006537103.1, XP_006537104.1, XP_006537105.1, XP_006537106.1, XP_006537107.1, XP_006537108.1, XP_006537109.1, XP_006537110.1, XP_006537111.1, XP_006537112.1, XP_006537113.1, XP_006537114.1, XP_006537115.1 → *cf.* page suivante

Table S3.2. (suite)

Output proteins: 1917 new murine EGZs (suite)

XP_006537116.1, XP_006537117.1, XP_006537451.1, XP_006537498.1, XP_006537503.1, XP_006537504.1, XP_006537622.1, XP_006537623.1, XP_006537624.1, XP_006537625.1, XP_006537626.1, XP_006537650.1, XP_006537651.1, XP_006537652.1, XP_006537653.1, XP_006537663.1, XP_006537664.1, XP_006537665.1, XP_006537705.1, XP_006537706.1, XP_006537707.1, XP_006537708.1, XP_006537709.1, XP_006537710.1, XP_006537711.1, XP_006537712.1, XP_006537713.1, XP_006537739.1, XP_006537868.1, XP_006537869.1, XP_006537870.1, XP_006537874.1, XP_006537875.1, XP_006537876.1, XP_006537877.1, XP_006537878.1, XP_006537879.1, XP_006537880.1, XP_006537901.1, XP_006538166.1, XP_006538167.1, XP_006538168.1, XP_006538169.1, XP_006538211.1, XP_006538212.1, XP_006538213.1, XP_006538214.1, XP_006538215.1, XP_006538216.1, XP_006538217.1, XP_006538221.1, XP_006538222.1, XP_006538271.1, XP_006538321.1, XP_006538455.1, XP_006538811.1, XP_006538812.1, XP_006538813.1, XP_006538814.1, XP_006538867.1, XP_006538868.1, XP_006538878.1, XP_006538914.1, XP_006539284.1, XP_006539285.1, XP_006539286.1, XP_006539287.1, XP_006539288.1, XP_006539289.1, XP_006539290.1, XP_006539291.1, XP_006539354.1, XP_006539355.1, XP_006539356.1, XP_006539357.1, XP_006539358.1, XP_006539359.1, XP_006539360.1, XP_006539361.1, XP_006539362.1, XP_006539363.1, XP_006539364.1, XP_006539365.1, XP_006539366.1, XP_006539367.1, XP_006539368.1, XP_006539369.1, XP_006539370.1, XP_006539371.1, XP_006539372.1, XP_006539373.1, XP_006539374.1, XP_006539375.1, XP_006539376.1, XP_006539377.1, XP_006539507.1, XP_006539746.1, XP_006539747.1, XP_006539748.1, XP_006539749.1, XP_006539750.1, XP_006539751.1, XP_006539752.1, XP_006539753.1, XP_006539755.1, XP_006539842.1, XP_006539843.1, XP_006539844.1, XP_006539854.1, XP_006539855.1, XP_006539856.1, XP_006539857.1, XP_006539948.1, XP_006540165.1, XP_006540166.1, XP_006540167.1, XP_006540168.1, XP_006540169.1, XP_006540170.1, XP_006540171.1, XP_006540305.1, XP_006540306.1, XP_006540307.1, XP_006540339.1, XP_006540340.1, XP_006540341.1, XP_006540426.1, XP_006540427.1, XP_006540428.1, XP_006540429.1, XP_006540430.1, XP_006540431.1, XP_006540470.1, XP_006540471.1, XP_006540492.1, XP_006540493.1, XP_006540839.1, XP_006540840.1, XP_006540927.1, XP_006540928.1, XP_006540929.1, XP_006540930.1, XP_006540931.1, XP_006540932.1, XP_006540933.1, XP_006540934.1, XP_006540935.1, XP_006540936.1, XP_006540937.1, XP_006540938.1, XP_006540939.1, XP_006541023.1, XP_006541024.1, XP_006541025.1, XP_006541026.1, XP_006541027.1, XP_006541028.1, XP_006541029.1, XP_006541395.1, XP_006541494.1, XP_006541495.1, XP_006543392.1, XP_006543394.1, XP_006543590.1, XP_006543692.1, XP_006543693.1, XP_006543694.1, XP_006543695.1, XP_006543730.1, XP_006543731.1, XP_006543732.1, XP_006543766.1, XP_006543767.1, XP_006543954.1, XP_006543972.1, XP_006544243.1, XP_006544244.1, XP_006544461.1, XP_006544476.1, XP_006544479.1, XP_006544608.1, XP_006544659.1, XP_006544891.1, XP_006544892.1, XP_006544893.1, XP_006544894.1, XP_006544895.1, XP_006544896.1, XP_006544897.1, XP_006544898.1, XP_006544899.1, XP_006544900.1, XP_006544901.1, XP_006544902.1, XP_006544903.1

(i) *H. sapiens* SPRR2s against the *H. sapiens* proteome → Figure S3.14

Input proteins: 8 known human SPRR2s

NP_005979.1, NP_008876.3, NP_001014313.1, NP_001014450.1, NP_001017418.1, NP_001019380.2

Output proteins: 8 known human SPRR2s

NP_005979.1, NP_008876.3, NP_001014313.1, NP_001014450.1, NP_001017418.1, NP_001019380.2

(j) *H. sapiens* SPRR2s against the *M. musculus* proteome → Figure S3.15

Input proteins: 8 known human SPRR2s

NP_005979.1, NP_008876.3, NP_001014313.1, NP_001014450.1, NP_001017418.1, NP_001019380.2

Output proteins: 8 known murine SPRR2s

NP_035598.2, NP_035599.2, NP_035602.1, NP_035604.1, NP_001158259.1

Table S3.3. Correspondance between AtLTP clusters formed by Edstam *et al.* (2011) and KAPPA. Numbers indicate the number of sequences in common between two groups. Edstam's groups AtLTP1, AtLTP2, AtLTPc and AtLTPe were exactly found by KAPPA. Groups AtLTPd and AtLTPg were fractioned into 2 and 3 KAPPA clusters, respectively. Three out of 4 proteins from group AtLTPx were attributed to existing clusters.

KAPPA's clusters	Clusters from Edstam *et al.* (2011)							New LTPs	Total
	1	2	c	d	e	g	x		
Cluster 01						21		37	58
Cluster 02		14					1	8	23
Cluster 03	12						2	6	20
Cluster 04				9				3	12
Cluster 05						7		2	9
Cluster 06						2		3	5
Cluster 07								4	4
Cluster 08				3				1	4
Cluster 09			3					0	3
Cluster 10								2	2
Cluster 11								2	2
Cluster 12								2	2
Cluster 13								2	2
Cluster 14								2	2
Cluster 15								2	2
Cluster 16					2			0	2
Cluster 17								2	2
Cluster 18								2	2
Cluster 19								2	2
Cluster 20								2	2
Cluster 21								2	2
Singletons						1	1	46	48
Total	**12**	**14**	**3**	**12**	**2**	**31***	**4**	**0**	**0**

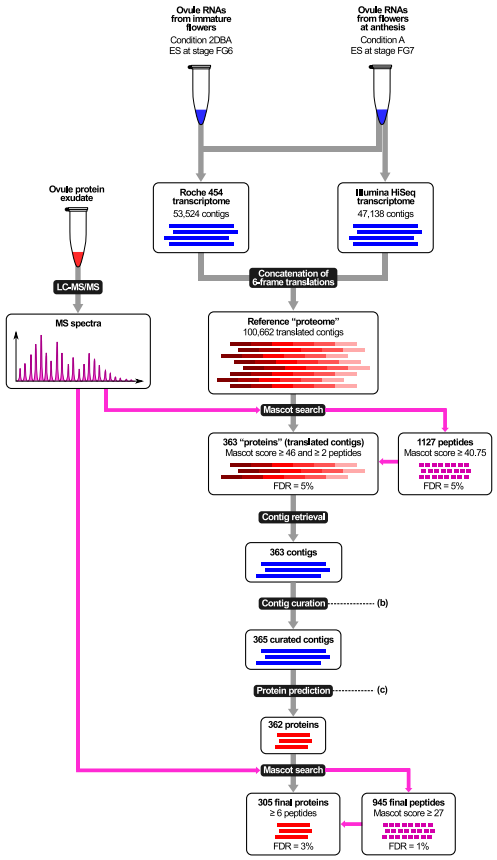
*One AtLTPg was not considered by KAPPA since it did not contain any cysteine (AT1G05450.1). The total number of AtLTPg described by Edstam *et al.* (2011) is then 32.

Table S3.4. KAPPA advantages with respect to conventional methods for discovery and clustering of key aminoacid patterned proteins. → *cf.* Fichier XLSX disponible en ligne à https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv047

Characteristics	KAPPA	Conventional methods
Automated	Fully automated	Curation steps needed
Speed	Quick, multithreadable	Time-consuming
Analysis criteria	Quantitative and objective (κ-score)	Partially subjective
Mapping and clustering	Focused on the key residue pattern	Pattern and rest of the sequence analysed at once
Subclustering	Yes	No
De novo search offered	Yes	No
Sequence pre-filtering	Yes	No
Detection of distant CRPs	Easy	Difficult
CRPs/non CRPs discrimination	Efficient, even for distant relatives	Works only for close CRPs
Optimisation	Easy (optimisation functionality)	Tedious
Connection to Cytoscape	Yes	No

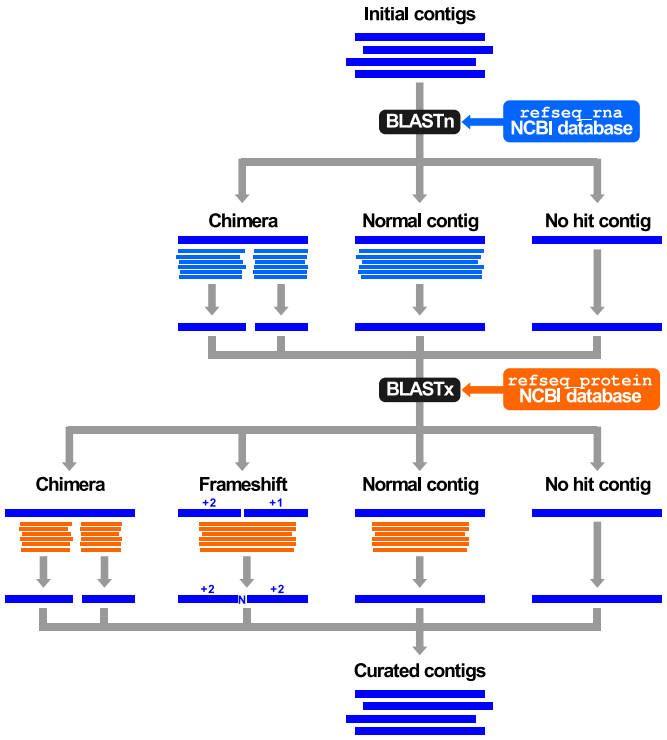
ANNEXE B

Matériel supplémentaire du chapitre 4

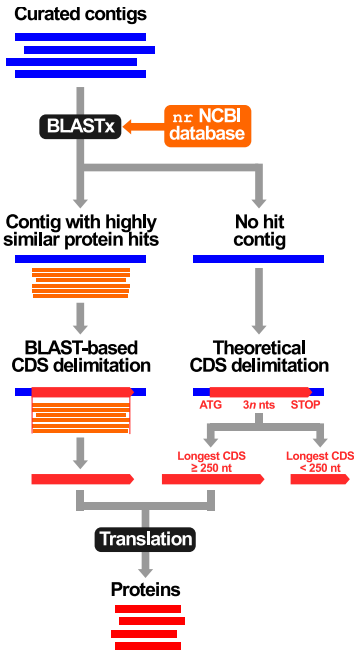


(a) Main workflow

Figure S4.1. Protein identification using customized protein database derived from RNA-seq data and its optimization through contig curation protocol. Ovule RNAs were sequenced in parallel using 454 and Illumina NGS platforms. Two *de novo* assembled ovule transcriptomes were generated, comprising 100 662 sequences in total. MS/MS spectra obtained from ovule exudations were searched against six-frame translations of the concatenated database. Contig curation was applied to the initially identified 363 proteins to fix frameshifts and chimera stemming from RNA-seq or assembly (module contig curation). Protein sequences in the ovule secretome were deduced after curation (module protein prediction).

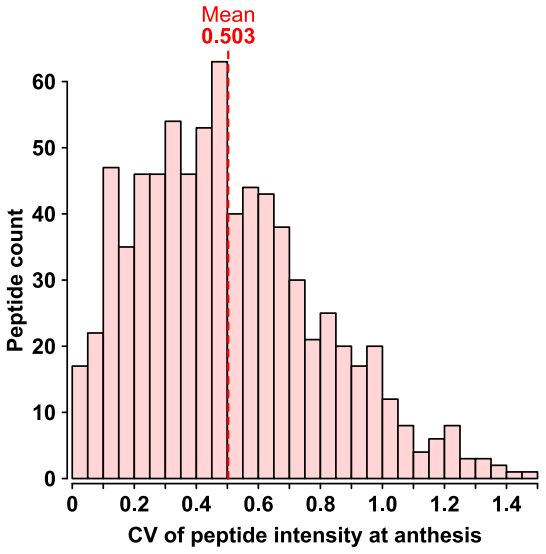


(b) Contig curation

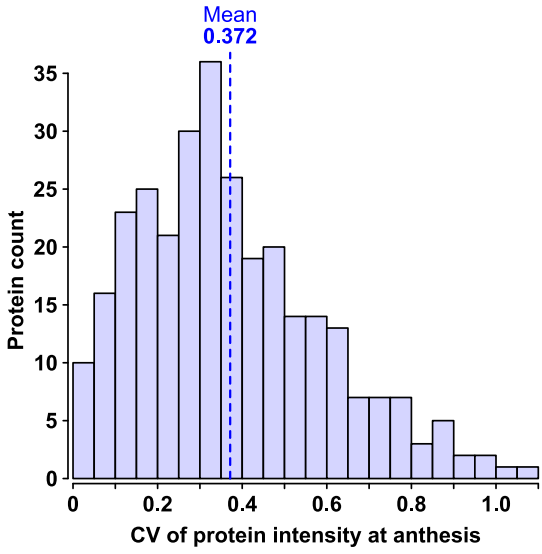


(c) Protein prediction

Figure S4.1. (suite)

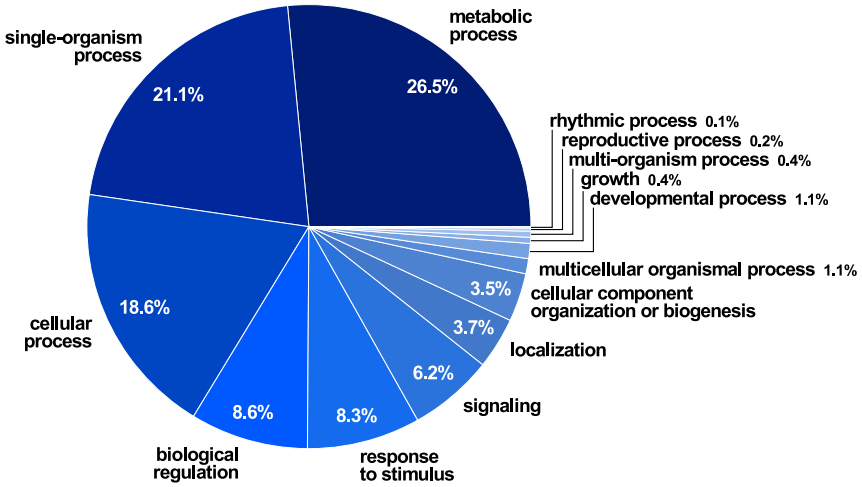


(a) Peptide intensity CV vs. the number of relative peptides in the reference condition (Anthesis).

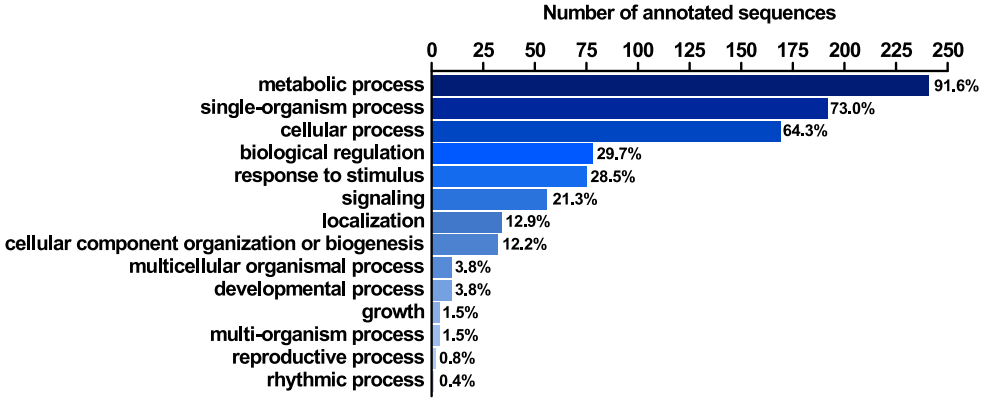


(b) Protein intensity CV vs. the number of relative proteins.

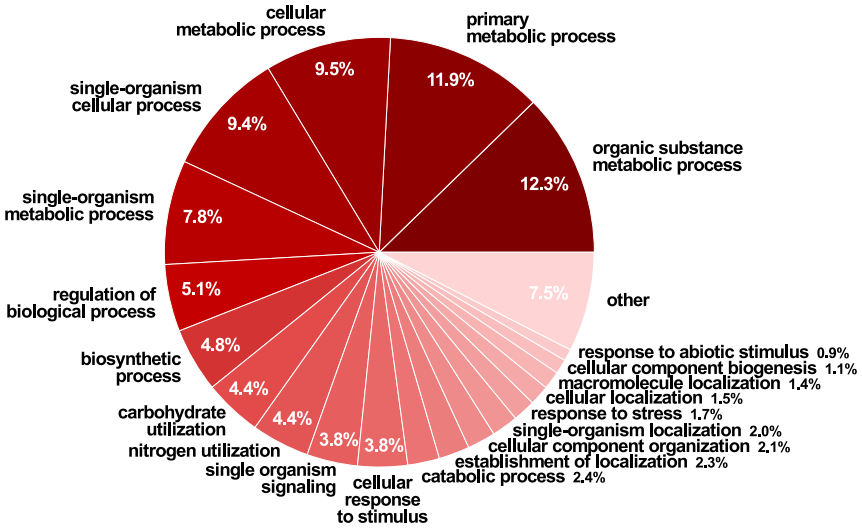
Figure S4.2. Reproducibility of label-free quantification of ovule secretome based on three biological replicates. Dotted lines mark 50 % of the peptide (a) and protein (b) population.



(a) A pie chart showing GO classification under biological process

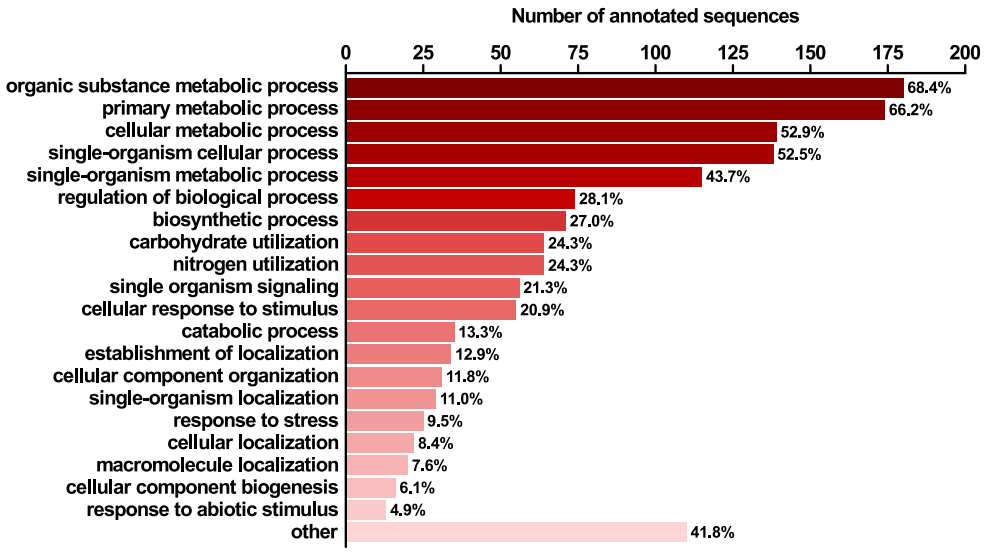


(b) A bar plot displaying the proportion of sequences associated with each GO-term under biological process



(c) A pie chart showing GO classification under molecular function

Figure S4.3. GO classification of ovule secretome in *S. chacoense*.



(d) A bar plot displaying the proportion of sequences associated with each GO-term under molecular function

Figure S4.3. (suite)

Table S4.1. Main characteristics of ovule secretome in *S. chacoense*. → *cf.* Fichier XLSX disponible en ligne à : http://pubs.acs.org/doi/suppl/10.1021/acs.jproteome.5b00618/suppl_file/pr5b00618_si_002.xlsx

Table S4.2. PFAM domain and family annotation associated with each secreted protein in ovule secretome. → *cf.* Fichier XLSX disponible en ligne à : http://pubs.acs.org/doi/suppl/10.1021/acs.jproteome.5b00618/suppl_file/pr5b00618_si_002.xlsx

Table S4.3. Determination of OSP specificity by comparison of ovule secretome to other secretomic data sets. → *cf.* Fichier XLSX disponible en ligne à : http://pubs.acs.org/doi/suppl/10.1021/acs.jproteome.5b00618/suppl_file/pr5b00618_si_002.xlsx

Table S4.4. Ortholog survey of novel ScCRPs detected from the ovule secretome against other solanaceous species. → *cf.* Fichier XLSX disponible en ligne à : http://pubs.acs.org/doi/suppl/10.1021/acs.jproteome.5b00618/suppl_file/pr5b00618_si_002.xlsx

ANNEXE C

Matériel supplémentaire du chapitre 5

Dataset S5.1. Detailed view of microarray data, differential expression analysis, and *k*-means clustering results. (a) Conditions and samples under study. (b) Raw expression values. (c) Expression values. (d) Average expression values. (e) Differential expression: Log-ratios. (f) Differential expression: Fold-changes. (g) Differential expression: p-values. (h) Differential expression: Regulations. (i) Clustering results. → *cf.* Fichier XLSX disponible en ligne à : https://www.mdpi.com/2223-7747/8/6/185/s1

Dataset S5.2. Detailed view of *in silico* annotations made on genes on the microarray: BLAST results, GO functional classification, enzyme and transcription factor predictions, and predictions on secreted proteins. (a) Nucleotide BLAST results. (b) Protein BLAST results. (c) Sequence descriptions. (d) GO functional annotations. (e) Enzyme predictions. (f) Transcription factor predictions. (g) Secretion predictions. (h) Predictions on signal peptide-containing proteins. → *cf.* Fichier XLSX disponible en ligne à : https://www.mdpi.com/2223-7747/8/6/185/s1

Dataset S5.3. Detailed view of differential expression data for ovary genes modulated by pollination. (a) Detailed view of genes regulated 6 HAP. (b) Detailed view of genes regulated 24 HAP.(c) Detailed view of genes regulated 48 HAP.(d) Detailed view of genes belonging to each cluster. → *cf.* Fichier XLSX disponible en ligne à : https://www.mdpi.com/2223-7747/8/6/185/s1

Dataset S5.4. Statistical analysis on GO functional categories enriched in each pollination condition. (a) Genes specific to CCP 6 HAP. (b) Genes specific to CIP 6 HAP. (c) Genes specific to HCP 6 HAP. (d) Genes common to CCP and CIP 6 HAP. (e) Genes common to CCP and HCP 6 HAP. (f) Genes common to CIP and HCP 6 HAP. (g) Genes common to CCP, CIP, and HCP 6 HAP. (h) Genes specific to CCP 24 HAP. (i) Genes specific to CIP 24 HAP. (j) Genes specific to HCP 24 HAP. (k) Genes common to CCP and CIP 24 HAP. (l) Genes common to CCP and HCP 24 HAP. (m) Genes common to CIP and HCP 24 HAP. (n) Genes common to CCP, CIP, and HCP 24 HAP. (o) Genes specific to CCP 48 HAP. (p) Genes specific to CIP 48 HAP. (q) Genes specific to HCP 48 HAP. (r) Genes common to CCP and CIP 48 HAP. (s) Genes common to CCP and HCP 48 HAP. (t) Genes common to CIP and HCP 48 HAP. (u) Genes common to CCP, CIP, and HCP 48 HAP. → *cf.* Fichier XLSX disponible en ligne à : https://www.mdpi.com/2223-7747/8/6/185/s1

Dataset S5.5. Statistical analysis on GO functional categories enriched in each cluster. (a) Cluster 1. (b) Cluster 2. (c) Cluster 3. (d) Cluster 4. (e) Cluster 5. (f) Cluster 6. (g) Cluster 7. (h) Cluster 8. (i) Cluster 9. (j) Cluster 10. (k) Cluster 11. (l) Cluster 12. (m) Cluster 13. (n) Cluster 14. (o) Cluster 15. (p) Cluster 16. (q) Cluster 17. (r) Cluster 18. (s) Cluster 19. (t) Cluster 20. (u) Cluster 21. (v) Cluster 22. (w) Cluster 23. (x) Cluster 24. (y) Cluster 25. → *cf.* Fichier XLSX disponible en ligne à : https://www.mdpi.com/2223-7747/8/6/185/s1

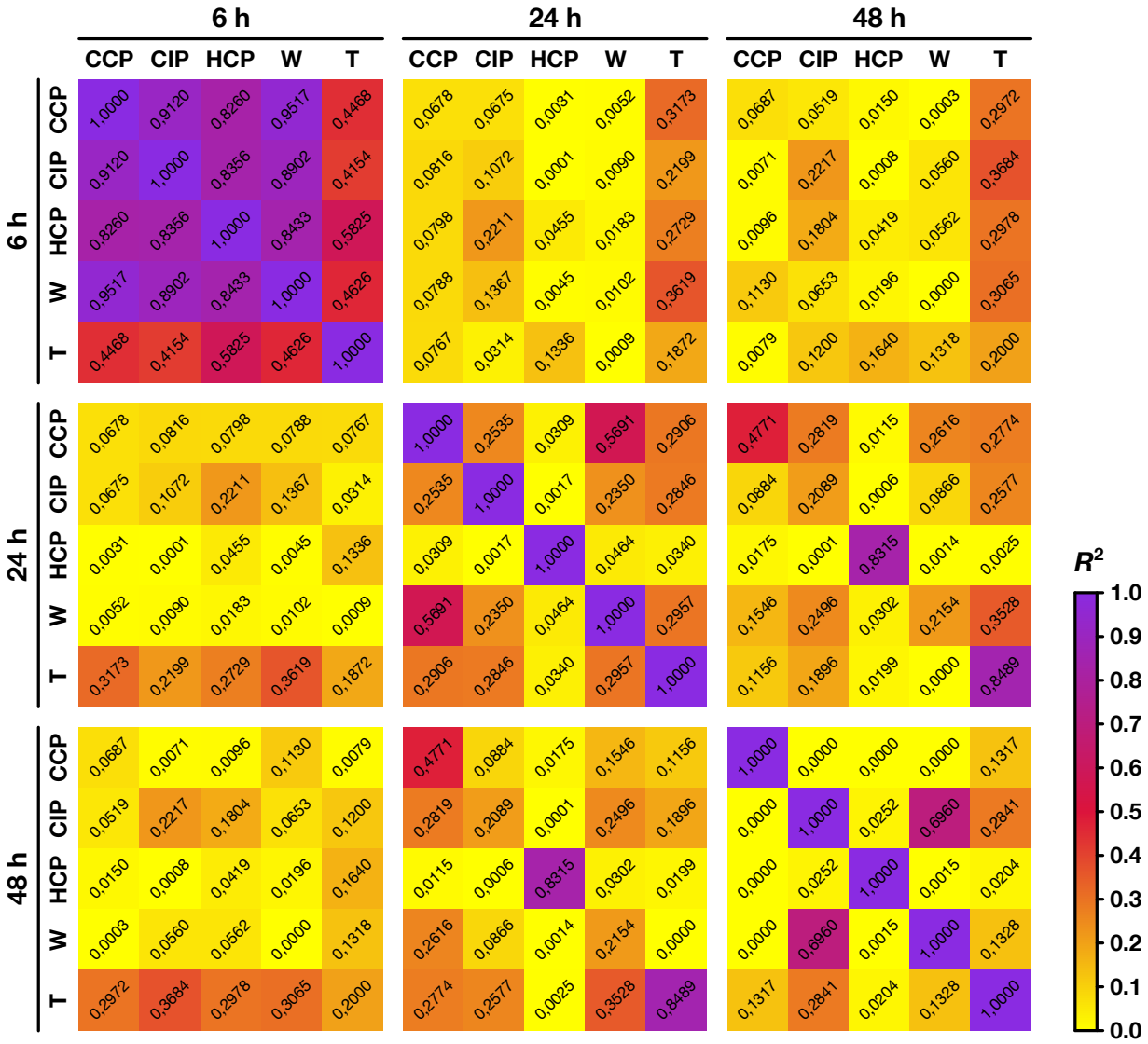


Figure S5.1. Heatmap describing pairwise correlation coefficients of expression ratios between samples. Squared Pearson's correlation coefficients (R^2) for each pair of samples under study. For two samples X and Y, the linear regression was performed on $\log_2$ values of expression changes (X/UO vs. Y/UO) of genes regulated in X and/or Y.

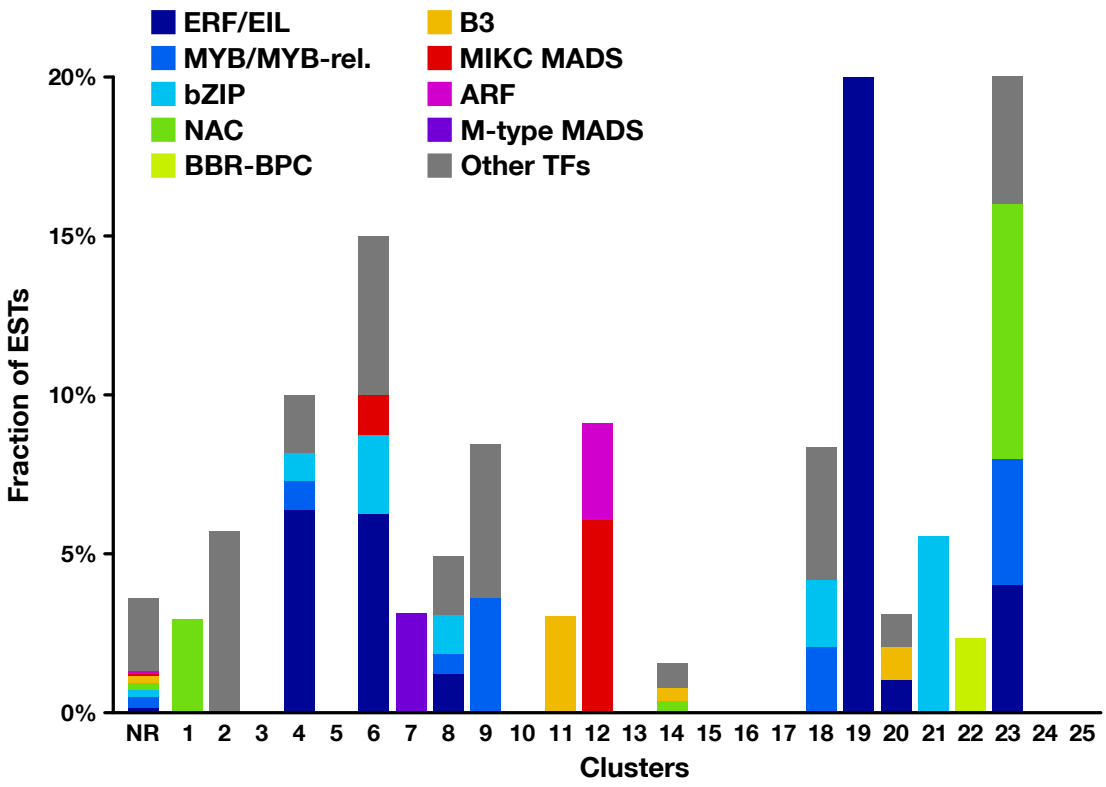


Figure S5.2. Graphical summary of *in silico* transcription factor predictions made in each cluster. The PlantTFDB prediction tool was used on the best BLASTx hit for each EST to predict and classify sequences into transcription factor (TF) families.

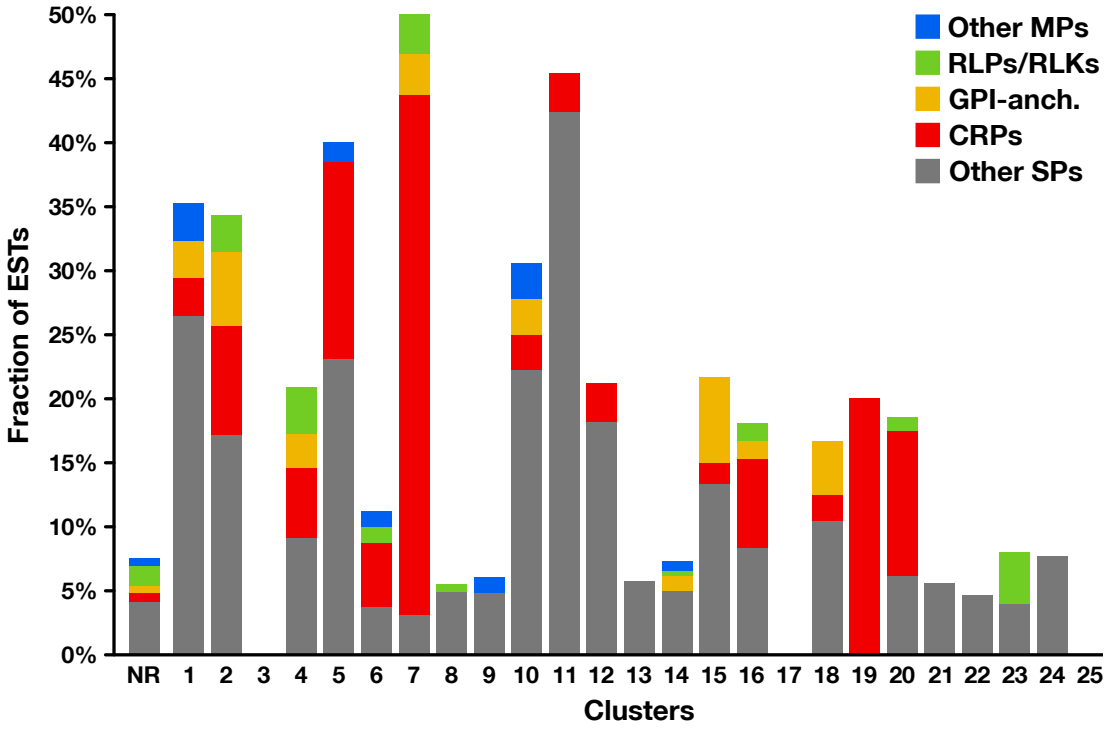


Figure S5.3. Graphical summary of *in silico* predictions made on secreted proteins in each cluster. SignalP was used on the best BLASTx hit for each EST to predict the presence of a signal peptide. GPI anchors were predicted using PredGPI. Remaining sequences were inspected for the presence of transmembrane helices with TMHMM. Sequences with one such helix were classified as potential receptor-like proteins or kinases (RLPs/RLKs); sequences with multiple transmembrane helices were classified as other membrane proteins. Sequences without a transmembrane helix nor a GPI anchor were split into cysteine-rich proteins (CRPs; mature peptide ≤ 150 aa, 6+ cysteines) and other non-membrane proteins.

Table S5.1. Number and proportion of genes regulated after each treatment, along with coregulation statistics across conditions and time points.. The number of up- and down-regulated genes in each category/overlap is given in columns "↑" and "↓", respectively. The number of coregulated genes, i.e. having the same regulation in the two conditions compared, is given in columns "↑↑" (up/up) and "↓↓" (down/down). The number of genes in the overlap having opposite regulations are shown in column "Opp.". Values given in column "*p*-values" express the significance of overlaps between gene sets; they were computed with a Fisher's exact test. Local percentages (in black) were computed according to the total number of genes in each row. Global percentages (in gray) were computed according to the total number of genes regulated in the condition. In this table, "specific" refers to genes that are regulated in the considered condition but not in the other pollination conditions at the same time point.

(a) Genes regulated 6 h after conspecific compatible pollination (CCP)

Category	Regulated			Coregulated			Opp.	*p*-value
	all	↑	↓	↑↑	↓↓	total		
Total	**125**	**68**	**57**					
	100%	54.4%	45.6%					
	100%	54.4%	45.6%					
Specific	**52**	**22**	**30**					
	100%	42.3%	57.7%					
	41.6%	17.6%	24.0%					
Non specific	**73**	**46**	**27**					
	100%	63.0%	37.0%					
	58.4%	36.8%	21.6%					
Common to CIP 6 h	**63**	**40**	**23**	**40**	**23**	**63**	**0**	1.2×10^{-94}
	100%	63.5%	36.5%	63.5%	36.5%	100%	0.0%	
	50.4%	32.0%	18.4%	32.0%	18.4%	50.4%	0.0%	
Common to HCP 6 h	**48**	**30**	**18**	**30**	**18**	**48**	**0**	7.8×10^{-130}
	100%	62.5%	37.5%	62.5%	37.5%	100%	0.0%	
	38.4%	24.0%	14.4%	24.0%	14.4%	38.4%	0.0%	
Common to W 6 h	**90**	**49**	**41**	**49**	**41**	**90**	**0**	1.0×10^{-42}
	100%	54.4%	45.6%	54.4%	45.6%	100%	0.0%	
	72.0%	39.2%	32.8%	39.2%	32.8%	72.0%	0.0%	
Common to T 6 h	**55**	**35**	**20**	**35**	**18**	**53**	**2**	1.0×10^{-90}
	100%	63.6%	36.4%	63.6%	32.7%	96.4%	3.6%	
	44.0%	28.0%	16.0%	28.0%	14.4%	42.4%	1.6%	
Common to CCP 24 h	**30**	**16**	**14**	**11**	**11**	**22**	**8**	4.0×10^{-92}
	100%	53.3%	46.7%	36.7%	36.7%	73.3%	26.7%	
	24.0%	12.8%	11.2%	8.8%	8.8%	17.6%	6.4%	
Common to CCP 48 h	**54**	**32**	**22**	**16**	**20**	**36**	**18**	3.3×10^{-27}
	100%	59.3%	40.7%	29.6%	37.0%	66.7%	33.3%	
	43.2%	25.6%	17.6%	12.8%	16.0%	28.8%	14.4%	

Table S5.1. (continued)

(b) Genes regulated 6 h after conspecific incompatible pollination (CIP)

Category	Regulated			Coregulated			Opp.	*p*-value
	all	↑	↓	↑↑	↓↓	total		
Total	**88**	**57**	**31**					
	100%	64.8%	35.2%					
	100%	64.8%	35.2%					
Specific	**22**	**15**	**7**					
	100%	68.2%	31.8%					
	25.0%	17.0%	8.0%					
Non specific	**66**	**42**	**24**					
	100%	63.6%	36.4%					
	75.0%	47.7%	27.3%					
Common to CCP 6 h	**63**	**40**	**23**	**40**	**23**	**63**	**0**	4.0×10^{-28}
	100%	63.5%	36.5%	63.5%	36.5%	100%	0.0%	
	71.6%	45.5%	26.1%	45.5%	26.1%	71.6%	0.0%	
Common to HCP 6 h	**41**	**26**	**15**	**26**	**15**	**41**	**0**	8.6×10^{-78}
	100%	63.4%	36.6%	63.4%	36.6%	100%	0.0%	
	46.6%	29.5%	17.0%	29.5%	17.0%	46.6%	0.0%	
Common to W 6 h	**63**	**39**	**24**	**39**	**24**	**63**	**0**	1.4×10^{-27}
	100%	61.9%	38.1%	61.9%	38.1%	100%	0.0%	
	71.6%	44.3%	27.3%	44.3%	27.3%	71.6%	0.0%	
Common to T 6 h	**46**	**31**	**15**	**31**	**14**	**45**	**1**	9.1×10^{-52}
	100%	67.4%	32.6%	67.4%	30.4%	97.8%	2.2%	
	52.3%	35.2%	17.0%	35.2%	15.9%	51.1%	1.1%	
Common to CIP 24 h	**16**	**10**	**6**	**9**	**6**	**15**	**1**	4.5×10^{-113}
	100%	62.5%	37.5%	56.2%	37.5%	93.8%	6.2%	
	18.2%	11.4%	6.8%	10.2%	6.8%	17.0%	1.1%	
Common to CIP 48 h	**17**	**14**	**3**	**13**	**2**	**15**	**2**	1.5×10^{-97}
	100%	82.4%	17.6%	76.5%	11.8%	88.2%	11.8%	
	19.3%	15.9%	3.4%	14.8%	2.3%	17.0%	2.3%	

Table S5.1. (continued)

(c) Genes regulated 6 h after heterospecific compatible pollination (HCP)

Category	Regulated			Coregulated			Opp.	*p*-value
	all	↑	↓	↑↑	↓↓	total		
Total	**64**	**43**	**21**					
	100%	67.2%	32.8%					
	100%	67.2%	32.8%					
Specific	**13**	**11**	**2**					
	100%	84.6%	15.4%					
	20.3%	17.2%	3.1%					
Non specific	**51**	**32**	**19**					
	100%	62.7%	37.3%					
	79.7%	50.0%	29.7%					
Common to CCP 6 h	**48**	**30**	**18**	**30**	**18**	**48**	**0**	1.2×10^{-16}
	100%	62.5%	37.5%	62.5%	37.5%	100%	0.0%	
	75.0%	46.9%	28.1%	46.9%	28.1%	75.0%	0.0%	
Common to CIP 6 h	**41**	**26**	**15**	**26**	**15**	**41**	**0**	4.1×10^{-31}
	100%	63.4%	36.6%	63.4%	36.6%	100%	0.0%	
	64.1%	40.6%	23.4%	40.6%	23.4%	64.1%	0.0%	
Common to W 6 h	**49**	**30**	**19**	**30**	**19**	**49**	**0**	4.6×10^{-15}
	100%	61.2%	38.8%	61.2%	38.8%	100%	0.0%	
	76.6%	46.9%	29.7%	46.9%	29.7%	76.6%	0.0%	
Common to T 6 h	**43**	**31**	**12**	**31**	**12**	**43**	**0**	4.4×10^{-23}
	100%	72.1%	27.9%	72.1%	27.9%	100%	0.0%	
	67.2%	48.4%	18.8%	48.4%	18.8%	67.2%	0.0%	
Common to HCP 24 h	**21**	**15**	**6**	**13**	**3**	**16**	**5**	1.5×10^{-43}
	100%	71.4%	28.6%	61.9%	14.3%	76.2%	23.8%	
	32.8%	23.4%	9.4%	20.3%	4.7%	25.0%	7.8%	
Common to HCP 48 h	**13**	**10**	**3**	**7**	**2**	**9**	**4**	3.1×10^{-68}
	100%	76.9%	23.1%	53.8%	15.4%	69.2%	30.8%	
	20.3%	15.6%	4.7%	10.9%	3.1%	14.1%	6.2%	

Table S5.1. (continued)

(d) Genes regulated 24 h after conspecific compatible pollination (CCP)

Category	Regulated			Coregulated			Opp.	*p*-value
	all	↑	↓	↑↑	↓↓	total		
Total	**354**	**151**	**203**					
	100%	42.7%	57.3%					
	100%	42.7%	57.3%					
Specific	**269**	**128**	**141**					
	100%	47.6%	52.4%					
	76.0%	36.2%	39.8%					
Non specific	**85**	**23**	**62**					
	100%	27.1%	72.9%					
	24.0%	6.5%	17.5%					
Common to CIP 24 h	**41**	**14**	**27**	**14**	**23**	**37**	**4**	0
	100%	34.1%	65.9%	34.1%	56.1%	90.2%	9.8%	
	11.6%	4.0%	7.6%	4.0%	6.5%	10.5%	1.1%	
Common to HCP 24 h	**58**	**14**	**44**	**8**	**17**	**25**	**33**	6.1×10^{-307}
	100%	24.1%	75.9%	13.8%	29.3%	43.1%	56.9%	
	16.4%	4.0%	12.4%	2.3%	4.8%	7.1%	9.3%	
Common to W 24 h	**61**	**24**	**37**	**24**	**37**	**61**	**0**	0
	100%	39.3%	60.7%	39.3%	60.7%	100%	0.0%	
	17.2%	6.8%	10.5%	6.8%	10.5%	17.2%	0.0%	
Common to T 24 h	**56**	**30**	**26**	**30**	**23**	**53**	**3**	0
	100%	53.6%	46.4%	53.6%	41.1%	94.6%	5.4%	
	15.8%	8.5%	7.3%	8.5%	6.5%	15.0%	0.8%	
Common to CCP 6 h	**30**	**14**	**16**	**11**	**11**	**22**	**8**	0
	100%	46.7%	53.3%	36.7%	36.7%	73.3%	26.7%	
	8.5%	4.0%	4.5%	3.1%	3.1%	6.2%	2.3%	
Common to CCP 48 h	**222**	**97**	**125**	**95**	**122**	**217**	**5**	1.2×10^{-30}
	100%	43.7%	56.3%	42.8%	55.0%	97.7%	2.3%	
	62.7%	27.4%	35.3%	26.8%	34.5%	61.3%	1.4%	

Table S5.1. (continued)

(e) Genes regulated 24 h after conspecific incompatible pollination (CIP)

Category	Regulated			Coregulated			Opp.	*p*-value
	all	↑	↓	↑↑	↓↓	total		
Total	**89**	**52**	**37**					
	100%	58.4%	41.6%					
	100%	58.4%	41.6%					
Specific	**17**	**9**	**8**					
	100%	52.9%	47.1%					
	19.1%	10.1%	9.0%					
Non specific	**72**	**43**	**29**					
	100%	59.7%	40.3%					
	80.9%	48.3%	32.6%					
Common to CCP 24 h	**41**	**18**	**23**	**14**	**23**	**37**	**4**	1.9×10^{-38}
	100%	43.9%	56.1%	34.1%	56.1%	90.2%	9.8%	
	46.1%	20.2%	25.8%	15.7%	25.8%	41.6%	4.5%	
Common to HCP 24 h	**45**	**34**	**11**	**15**	**6**	**21**	**24**	8.4×10^{-38}
	100%	75.6%	24.4%	33.3%	13.3%	46.7%	53.3%	
	50.6%	38.2%	12.4%	16.9%	6.7%	23.6%	27.0%	
Common to W 24 h	**33**	**24**	**9**	**23**	**9**	**32**	**1**	6.6×10^{-78}
	100%	72.7%	27.3%	69.7%	27.3%	97.0%	3.0%	
	37.1%	27.0%	10.1%	25.8%	10.1%	36.0%	1.1%	
Common to T 24 h	**30**	**24**	**6**	**24**	**5**	**29**	**1**	2.4×10^{-70}
	100%	80.0%	20.0%	80.0%	16.7%	96.7%	3.3%	
	33.7%	27.0%	6.7%	27.0%	5.6%	32.6%	1.1%	
Common to CIP 6 h	**16**	**9**	**7**	**9**	**6**	**15**	**1**	6.0×10^{-115}
	100%	56.2%	43.8%	56.2%	37.5%	93.8%	6.2%	
	18.0%	10.1%	7.9%	10.1%	6.7%	16.9%	1.1%	
Common to CIP 48 h	**28**	**10**	**18**	**9**	**18**	**27**	**1**	6.0×10^{-79}
	100%	35.7%	64.3%	32.1%	64.3%	96.4%	3.6%	
	31.5%	11.2%	20.2%	10.1%	20.2%	30.3%	1.1%	

Table S5.1. (continued)

(f) Genes regulated 24 h after heterospecific compatible pollination (HCP)

Category	Regulated			Coregulated			Opp.	*p*-value
	all	↑	↓	↑↑	↓↓	total		
Total	**285**	**140**	**145**					
	100%	49.1%	50.9%					
	100%	49.1%	50.9%					
Specific	**196**	**93**	**103**					
	100%	47.4%	52.6%					
	68.8%	32.6%	36.1%					
Non specific	**89**	**47**	**42**					
	100%	52.8%	47.2%					
	31.2%	16.5%	14.7%					
Common to CCP 24 h	**58**	**35**	**23**	**8**	**17**	**25**	**33**	8.1×10^{-217}
	100%	60.3%	39.7%	13.8%	29.3%	43.1%	56.9%	
	20.4%	12.3%	8.1%	2.8%	6.0%	8.8%	11.6%	
Common to CIP 24 h	**45**	**20**	**25**	**15**	**6**	**21**	**24**	0
	100%	44.4%	55.6%	33.3%	13.3%	46.7%	53.3%	
	15.8%	7.0%	8.8%	5.3%	2.1%	7.4%	8.4%	
Common to W 24 h	**37**	**10**	**27**	**5**	**5**	**10**	**27**	0
	100%	27.0%	73.0%	13.5%	13.5%	27.0%	73.0%	
	13.0%	3.5%	9.5%	1.8%	1.8%	3.5%	9.5%	
Common to T 24 h	**41**	**4**	**37**	**4**	**9**	**13**	**28**	1.8×10^{-298}
	100%	9.8%	90.2%	9.8%	22.0%	31.7%	68.3%	
	14.4%	1.4%	13.0%	1.4%	3.2%	4.6%	9.8%	
Common to HCP 6 h	**21**	**16**	**5**	**13**	**3**	**16**	**5**	0
	100%	76.2%	23.8%	61.9%	14.3%	76.2%	23.8%	
	7.4%	5.6%	1.8%	4.6%	1.1%	5.6%	1.8%	
Common to HCP 48 h	**128**	**70**	**58**	**70**	**58**	**128**	**0**	3.0×10^{-199}
	100%	54.7%	45.3%	54.7%	45.3%	100%	0.0%	
	44.9%	24.6%	20.4%	24.6%	20.4%	44.9%	0.0%	

Table S5.1. (continued)

(g) Genes regulated 48 h after conspecific compatible pollination (CCP)

Category	Regulated			Coregulated			Opp.	*p*-value
	all	↑	↓	↑↑	↓↓	total		
Total	**1018**	**515**	**503**					
	100%	50.6%	49.4%					
	100%	50.6%	49.4%					
Specific	**893**	**464**	**429**					
	100%	52.0%	48.0%					
	87.7%	45.6%	42.1%					
Non specific	**125**	**51**	**74**					
	100%	40.8%	59.2%					
	12.3%	5.0%	7.3%					
Common to CIP 48 h	**76**	**35**	**41**	**14**	**28**	**42**	**34**	0
	100%	46.1%	53.9%	18.4%	36.8%	55.3%	44.7%	
	7.5%	3.4%	4.0%	1.4%	2.8%	4.1%	3.3%	
Common to HCP 48 h	**60**	**24**	**36**	**13**	**11**	**24**	**36**	0
	100%	40.0%	60.0%	21.7%	18.3%	40.0%	60.0%	
	5.9%	2.4%	3.5%	1.3%	1.1%	2.4%	3.5%	
Common to W 48 h	**80**	**43**	**37**	**29**	**24**	**53**	**27**	0
	100%	53.8%	46.2%	36.2%	30.0%	66.2%	33.8%	
	7.9%	4.2%	3.6%	2.8%	2.4%	5.2%	2.7%	
Common to T 48 h	**43**	**16**	**27**	**16**	**21**	**37**	**6**	0
	100%	37.2%	62.8%	37.2%	48.8%	86.0%	14.0%	
	4.2%	1.6%	2.7%	1.6%	2.1%	3.6%	0.6%	
Common to CCP 6 h	**54**	**18**	**36**	**16**	**20**	**36**	**18**	0
	100%	33.3%	66.7%	29.6%	37.0%	66.7%	33.3%	
	5.3%	1.8%	3.5%	1.6%	2.0%	3.5%	1.8%	
Common to CCP 24 h	**222**	**98**	**124**	**95**	**122**	**217**	**5**	0
	100%	44.1%	55.9%	42.8%	55.0%	97.7%	2.3%	
	21.8%	9.6%	12.2%	9.3%	12.0%	21.3%	0.5%	

Table S5.1. (continued)

(h) Genes regulated 48 h after conspecific incompatible pollination (CIP)

Category	Regulated			Coregulated			Opp.	*p*-value
	all	↑	↓	↑↑	↓↓	total		
Total	**147**	**65**	**82**					
	100%	44.2%	55.8%					
	100%	44.2%	55.8%					
Specific	**60**	**33**	**27**					
	100%	55.0%	45.0%					
	40.8%	22.4%	18.4%					
Non specific	**87**	**32**	**55**					
	100%	36.8%	63.2%					
	59.2%	21.8%	37.4%					
Common to CCP 48 h	**76**	**27**	**49**	**14**	**28**	**42**	**34**	3.4×10^{-22}
	100%	35.5%	64.5%	18.4%	36.8%	55.3%	44.7%	
	51.7%	18.4%	33.3%	9.5%	19.0%	28.6%	23.1%	
Common to HCP 48 h	**22**	**14**	**8**	**8**	**0**	**8**	**14**	8.7×10^{-165}
	100%	63.6%	36.4%	36.4%	0.0%	36.4%	63.6%	
	15.0%	9.5%	5.4%	5.4%	0.0%	5.4%	9.5%	
Common to W 48 h	**77**	**40**	**37**	**40**	**36**	**76**	**1**	2.0×10^{-75}
	100%	51.9%	48.1%	51.9%	46.8%	98.7%	1.3%	
	52.4%	27.2%	25.2%	27.2%	24.5%	51.7%	0.7%	
Common to T 48 h	**19**	**12**	**7**	**12**	**6**	**18**	**1**	2.3×10^{-197}
	100%	63.2%	36.8%	63.2%	31.6%	94.7%	5.3%	
	12.9%	8.2%	4.8%	8.2%	4.1%	12.2%	0.7%	
Common to CIP 6 h	**17**	**14**	**3**	**13**	**2**	**15**	**2**	3.5×10^{-201}
	100%	82.4%	17.6%	76.5%	11.8%	88.2%	11.8%	
	11.6%	9.5%	2.0%	8.8%	1.4%	10.2%	1.4%	
Common to CIP 24 h	**28**	**9**	**19**	**9**	**18**	**27**	**1**	1.1×10^{-180}
	100%	32.1%	67.9%	32.1%	64.3%	96.4%	3.6%	
	19.0%	6.1%	12.9%	6.1%	12.2%	18.4%	0.7%	

Table S5.1. (continued)

(i) Genes regulated 48 h after heterospecific compatible pollination (HCP)

Category	Regulated			Coregulated			Opp.	*p*-value
	all	↑	↓	↑↑	↓↓	total		
Total	**166**	**97**	**69**					
	100%	58.4%	41.6%					
	100%	58.4%	41.6%					
Specific	**95**	**51**	**44**					
	100%	53.7%	46.3%					
	57.2%	30.7%	26.5%					
Non specific	**71**	**46**	**25**					
	100%	64.8%	35.2%					
	42.8%	27.7%	15.1%					
Common to CCP 48 h	**60**	**38**	**22**	**13**	**11**	**24**	**36**	5.1×10^{-46}
	100%	63.3%	36.7%	21.7%	18.3%	40.0%	60.0%	
	36.1%	22.9%	13.3%	7.8%	6.6%	14.5%	21.7%	
Common to CIP 48 h	**22**	**16**	**6**	**8**	**0**	**8**	**14**	1.1×10^{-195}
	100%	72.7%	27.3%	36.4%	0.0%	36.4%	63.6%	
	13.3%	9.6%	3.6%	4.8%	0.0%	4.8%	8.4%	
Common to W 48 h	**22**	**15**	**7**	**12**	**0**	**12**	**10**	5.2×10^{-184}
	100%	68.2%	31.8%	54.5%	0.0%	54.5%	45.5%	
	13.3%	9.0%	4.2%	7.2%	0.0%	7.2%	6.0%	
Common to T 48 h	**13**	**2**	**11**	**2**	**1**	**3**	**10**	1.3×10^{-236}
	100%	15.4%	84.6%	15.4%	7.7%	23.1%	76.9%	
	7.8%	1.2%	6.6%	1.2%	0.6%	1.8%	6.0%	
Common to HCP 6 h	**13**	**8**	**5**	**7**	**2**	**9**	**4**	2.0×10^{-249}
	100%	61.5%	38.5%	53.8%	15.4%	69.2%	30.8%	
	7.8%	4.8%	3.0%	4.2%	1.2%	5.4%	2.4%	
Common to HCP 24 h	**128**	**70**	**58**	**70**	**58**	**128**	**0**	1.5×10^{-24}
	100%	54.7%	45.3%	54.7%	45.3%	100%	0.0%	
	77.1%	42.2%	34.9%	42.2%	34.9%	77.1%	0.0%	

Table S5.2. Summary of *in silico* transcription factor predictions made on genes modulated at a distance by pollination, with comparative enrichment analyses across conditions and clusters. The PlantTFDB prediction tool was used on the best BLASTx hit for each EST to predict and classify sequences into transcription factor (TF) families. Columns *Poll.* and *Rest* give the total number of genes regulated in at least one pollination condition, and the number of remaining genes, respectively. Asterisks (*) indicate a significant enrichment for a given prediction in a given sample (Fisher's exact test, $P < 0.05$.

(a) Genes regulated 6 h after pollination

Prediction	CCP 6 h		CIP 6 h		HCP 6 h		Poll.	Rest
	↑	↓	↑	↓	↑	↓		
ERF/EIL	1		*2	1	1		*17	8
MYB/MYB-rel.							7	18
bZIP							7	12
NAC							4	11
BBR-BPC							1	1
B3	1		1		1		3	12
MIKC MADS	1		*1				3	3
ARF	1		1				1	6
M-type MADS		*1					1	
Other TFs		1					21	121
Total TFs	4	2	5	1	2		65	192
Total	**68**	**57**	**57**	**31**	**43**	**21**	**1441**	**5340**

(b) Genes regulated 24 h after pollination

Prediction	CCP 24 h		CIP 24 h		HCP 24 h		Poll.	Rest
	↑	↓	↑	↓	↑	↓		
ERF/EIL	1	*10	*4	1	*9	2	*17	8
MYB/MYB-rel.	1	1			1	1	7	18
bZIP	1				*4	1	7	12
NAC	1	1			1		4	11
BBR-BPC				*1			1	1
B3							3	12
MIKC MADS		1				1	3	3
ARF							1	6
M-type MADS							1	
Other TFs	1	3			5	3	21	121
Total TFs	5	*16	4	2	*20	8	65	192
Total	**151**	**203**	**52**	**37**	**140**	**145**	**1441**	**5340**

Table S5.2. (continued)

(c) Genes regulated 48 h after pollination

Prediction	CCP 48 h		CIP 48 h		HCP 48 h		Poll.	Rest
	↑	↓	↑	↓	↑	↓		
ERF/EIL		*11	*3	1	*7	1	*17	8
MYB/MYB-rel.		3	*3		1	1	7	18
bZIP		*4			*3		7	12
NAC	2	1	*2			1	4	11
BBR-BPC							1	1
B3	2		1				3	12
MIKC MADS		1				1	3	3
ARF							1	6
M-type MADS		1					1	
Other TFs	2	12	1		1	3	21	121
Total TFs	6	*33	*10	1	*12	*7	65	192
Total	**515**	**503**	**65**	**82**	**97**	**69**	**1441**	**5340**

(d) Pollination-induced genes belonging to clusters 1–12

Prediction	Clusters												Poll.	Rest
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**		
ERF/EIL				*7		*5		2					*17	8
MYB/MYB-rel.				1				1	*3				7	18
bZIP				1		*2		2					7	12
NAC	1												4	11
BBR-BPC													1	1
B3											1		3	12
MIKC MADS						1						*2	3	3
ARF												*1	1	6
M-type MADS							*1						1	
Other TFs		2		2		4		3	4				21	121
Total TFs	1	2		*11		*12	1	8	*7		1	3	65	192
Total	**34**	**35**	**15**	**110**	**65**	**80**	**32**	**163**	**83**	**36**	**33**	**33**	**1441**	**5340**

Table S5.2. (continued)

(e) Pollination-induced genes belonging to clusters 13–25

Prediction	Clusters													Poll.	Rest
	13	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**		
ERF/EIL							*1	1			1			*17	8
MYB/MYB-rel.						1					1			7	18
bZIP						1			*1					7	12
NAC		1									*2			4	11
BBR-BPC										*1				1	1
B3		1						1						3	12
MIKC MADS														3	3
ARF														1	6
M-type MADS														1	
Other TFs		2				2		1			1			21	121
Total TFs		4				4	1	3	1	1	*5			65	192
Total	**52**	**260**	**60**	**72**	**10**	**48**	**5**	**97**	**18**	**43**	**25**	**13**	**19**	**1441**	**5340**

Table S5.3. Summary of *in silico* signal peptide and subsequent predictions made on genes modulated at a distance by pollination, with comparative enrichment analyses across conditions and clusters. SignalP was used on the best BLASTx hit for each EST to predict the presence of a signal peptide. GPI anchors were predicted using PredGPI. Remaining sequences were inspected for the presence of transmembrane helices with TMHMM. Sequences with one such helix were classified as potential receptor-like proteins or kinases (RLPs/RLKs); sequences with multiple transmembrane helices were classified as other membrane proteins. Sequences without a transmembrane helix nor a GPI anchor were split into cysteine-rich proteins (CRPs; mature peptide ≤ 150 aa, 6+ cysteines) and other non-membrane proteins. Columns *Poll.* and *Rest* give the total number of genes regulated in at least one pollination condition, and the number of remaining genes, respectively. Asterisks (*) indicate a significant enrichment for a given prediction in a given sample (Fisher's exact test, $P < 0.05$).

(a) Genes regulated 6 h after pollination

Prediction	CCP 6 h		CIP 6 h		HCP 6 h		Poll.	Rest
	↑	↓	↑	↓	↑	↓		
CRPs	*9	*6	*6	*5	1	*2	*59	40
Other non-membrane	*18	1	*12		*14		*130	219
GPI-anchored		1		1			*18	28
RLPs/RLKs							12	*84
Other membrane					1		7	32
Total secreted	*27	8	*18	6	*16	2	*226	403
Total	**68**	**57**	**57**	**31**	**43**	**21**	**1441**	**5340**

(b) Genes regulated 24 h after pollination

Prediction	CCP 24 h		CIP 24 h		HCP 24 h		Poll.	Rest
	↑	↓	↑	↓	↑	↓		
CRPs	4	*14	*5	*6	4	*11	*59	40
Other non-membrane	12	*20	4	2	*20	13	*130	219
GPI-anchored		1			*5	3	*18	28
RLPs/RLKs		1			2		12	*84
Other membrane					1		7	32
Total secreted	16	*36	9	*8	*32	*27	*226	403
Total	**151**	**203**	**52**	**37**	**140**	**145**	**1441**	**5340**

Table S5.3. (continued)

(c) Genes regulated 48 h after pollination

Prediction	CCP 48 h		CIP 48 h		HCP 48 h		Poll.	Rest
	↑	↓	↑	↓	↑	↓		
CRPs	*14	*30		*7	3	*5	*59	40
Other non-membrane	*37	*52	6	*16	*12	*8	*130	219
GPI-anchored	*9	1		1	2	*3	*18	28
RLPs/RLKs	3	5	1	2	3		12	*84
Other membrane	2	4					7	32
Total secreted	*65	*92	7	*26	*20	*16	*226	403
Total	**515**	**503**	**65**	**82**	**97**	**69**	**1441**	**5340**

(d) Pollination-induced genes belonging to clusters 1–12

Prediction	Clusters												Poll.	Rest
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**		
CRPs	1	*3		*6	*10	*4	*13			1	1	1	*59	40
Other non-mb	*9	*6		10	*15	3	1	8	4	*8	*14	*6	*130	219
GPI-anchored	1	*2		*3			1			1			*18	28
RLPs/RLKs		1		4		1	1	1					12	*84
Other mb	1				1	1			1	1			7	32
Total secreted	*12	*12		*23	*26	9	*16	9	5	*11	*15	*7	*226	403
Total	**34**	**35**	**15**	**110**	**65**	**80**	**32**	**163**	**83**	**36**	**33**	**33**	**1441**	**5340**

(e) Pollination-induced genes belonging to clusters 13–25

Prediction	Clusters													Poll.	Rest
	13	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**		
CRPs			1	*5		1	1	*11						*59	40
Other non-mb	3	13	*8	6		5		6	1	2	1	1		*130	219
GPI-anchored		3	*4	1		*2								*18	28
RLPs/RLKs		1		1				1			1			12	*84
Other mb		2												7	32
Total secreted	3	19	*13	*13		8	1	*18	1	2	2	1		*226	403
Total	**52**	**260**	**60**	**72**	**10**	**48**	**5**	**97**	**18**	**43**	**25**	**13**	**19**	**1441**	**5340**

Table S5.4. Summary of *in silico* metabolic pathway predictions made on genes modulated at a distance by pollination, with comparative enrichment analyses across conditions and clusters. The enzymes codes retrieved from BLAST and Blast2GO analyses were mapped to KEGG metabolic pathways. Columns *Poll.* and *Rest* give the total number of genes regulated in at least one pollination condition, and the number of remaining genes, respectively. Asterisks (*) indicate a significant enrichment for a given prediction in a given sample (Fisher's exact test, $P < 0.05$).

(a) Detailed view of metabolic pathway enrichment in each pollination condition

Metabolic pathway	6 HAP						24 HAP						48 HAP						Poll.	Rest
	CCP		CIP		HCP		CCP		CIP		HCP		CCP		CIP		HCP			
	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓		
Total ESTs	**68**	**57**	**57**	**31**	**43**	**21**	**151**	**203**	**52**	**37**	**140**	**145**	**515**	**503**	**65**	**82**	**97**	**69**	**1441**	**5340**
ESTs involved in metabolic pathways	12	6	8	4	4	3	10	16	9	1	16	24	51	66	9	12	9	8	180	736
1. Carbohydrate metabolism																				
Amino sugar and nucleotide sugar metabolism [00520]	1	1	0	2	0	1	2	1	2	0	2	4	7	4	0	1	2	0	21	64
Ascorbate and aldarate metabolism [00053]	0	1	0	***2**	0	1	1	1	0	0	0	0	1	1	0	0	1	0	6	28
Butanoate metabolism [00650]	0	0	0	0	0	0	0	0	0	0	0	2	0	3	0	0	0	1	5	15
C5-Branched dibasic acid metabolism [00660]	0	0	0	0	0	0	0	0	0	0	0	***2**	2	0	0	0	0	0	4	8
Citrate cycle (TCA cycle) [00020]	0	0	0	0	0	0	0	0	0	0	0	***3**	5	1	0	0	0	1	8	26
Fructose and mannose metabolism [00051]	0	0	0	0	0	0	1	0	0	0	0	0	1	3	0	1	0	0	5	29
Galactose metabolism [00052]	0	0	0	0	0	0	1	0	0	0	0	0	3	1	0	0	0	0	4	***53**
Glycolysis / Gluconeogenesis [00010]	1	1	0	2	0	1	1	0	1	0	2	1	3	7	1	1	0	0	13	79
Glyoxylate and dicarboxylate metabolism [00630]	0	0	0	0	0	0	1	0	1	0	2	0	3	5	2	0	0	0	9	35
Inositol phosphate metabolism [00562]	0	0	0	1	0	0	0	0	0	0	0	2	1	1	0	0	0	0	4	24
Pentose and glucuronate interconversions [00040]	0	1	0	***2**	0	1	0	1	0	0	1	0	2	2	0	0	1	0	7	32
Pentose phosphate pathway [00030]	1	0	0	0	0	0	0	0	0	0	0	0	2	2	0	1	0	0	4	34
Propanoate metabolism [00640]	0	0	0	0	0	0	0	0	1	0	2	2	2	***8**	1	0	0	1	***12**	18
Pyruvate metabolism [00620]	0	0	0	0	0	0	0	1	1	0	2	1	4	7	1	0	0	0	12	62
Starch and sucrose metabolism [00500]	1	0	0	0	0	1	0	3	1	0	1	1	2	3	0	1	1	0	10	58
2. Energy metabolism																				
Carbon fixation in photosynthetic organisms [00710]	1	0	0	0	0	0	0	0	0	0	0	0	2	4	0	1	0	0	6	33
Carbon fixation pathways in prokaryotes [00720]	0	0	0	0	0	0	0	0	1	0	2	2	5	***7**	1	0	0	1	13	27
Methane metabolism [00680]	0	1	0	***2**	0	1	1	0	1	0	2	0	3	5	1	1	0	0	10	34
Nitrogen metabolism [00910]	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	3	13
Oxidative phosphorylation [00190]	0	0	0	1	0	1	1	1	0	1	0	0	0	2	0	0	0	0	4	35
Photosynthesis [00195]	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	***1**	0	0	2	1
Sulfur metabolism [00920]	0	0	1	0	1	0	0	0	0	0	1	0	0	3	0	0	0	0	3	17
3. Lipid metabolism																				
alpha-Linolenic acid metabolism [00592]	0	***2**	0	***2**	0	1	1	2	0	0	0	1	4	4	1	2	0	1	***13**	22
Arachidonic acid metabolism [00590]	0	0	0	0	0	0	0	2	0	0	0	2	0	3	0	1	0	0	6	14
Ether lipid metabolism [00565]	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	4
Fatty acid biosynthesis [00061]	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	3	18
Fatty acid degradation [00071]	0	***2**	0	***2**	0	1	0	0	0	0	0	0	0	***7**	0	0	0	0	10	30
Fatty acid elongation [00062]	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	19
Glycerolipid metabolism [00561]	0	0	0	0	0	0	0	0	0	0	1	2	0	1	0	0	1	0	4	28
Glycerophospholipid metabolism [00564]	0	***2**	0	1	0	1	0	2	0	1	2	2	0	2	1	1	2	1	9	30
Linoleic acid metabolism [00591]	0	0	0	0	0	0	0	1	0	0	0	1	0	2	0	***2**	0	0	4	9
Sphingolipid metabolism [00600]	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	***47**
Steroid biosynthesis [00100]	0	0	0	0	0	0	0	***2**	1	0	0	***3**	***6**	0	0	0	0	0	***8**	4
Steroid hormone biosynthesis [00140]	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	2	17
4. Nucleotide metabolism																				
Purine metabolism [00230]	0	0	1	0	1	0	2	2	0	0	3	1	9	7	1	***5**	1	1	21	69
Pyrimidine metabolism [00240]	0	0	0	0	0	0	1	0	0	0	0	0	4	1	***2**	1	0	0	6	24
5. Amino acid metabolism																				
Alanine, aspartate and glutamate metabolism [00250]	0	0	0	0	0	0	0	2	0	0	0	1	0	4	0	0	0	1	6	21
Arginine and proline metabolism [00330]	0	0	0	0	0	0	0	0	1	0	3	0	1	5	0	0	***3**	0	9	41
Arginine biosynthesis [00220]	0	0	0	0	0	0	0	0	0	0	0	1	1	2	0	0	0	1	4	16
Cysteine and methionine metabolism [00270]	1	0	1	0	0	0	0	1	0	0	1	1	3	8	0	1	0	0	12	48
Glycine, serine and threonine metabolism [00260]	0	1	0	***2**	0	1	1	1	0	0	0	0	3	1	0	0	0	1	7	44
Histidine metabolism [00340]	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	4	17
Lysine biosynthesis [00300]	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	10
Lysine degradation [00310]	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	3	32
Phenylalanine metabolism [00360]	0	0	0	0	0	0	0	0	1	0	2	2	2	***11**	1	0	0	2	15	38
Phenylalanine, tyrosine and tryptophan biosynthesis [00400]	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	1	0	5	19
Tryptophan metabolism [00380]	0	0	0	0	0	0	0	0	0	0	0	0	1	6	1	0	0	0	8	34
Tyrosine metabolism [00350]	2	1	***2**	***2**	0	1	0	0	0	0	0	0	1	2	1	0	0	1	8	31
Valine, leucine and isoleucine biosynthesis [00290]	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	15
Valine, leucine and isoleucine degradation [00280]	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	4	25
6. Metabolism of other amino acids																				
beta-Alanine metabolism [00410]	0	0	0	0	0	0	0	0	0	0	0	1	1	3	0	0	0	***2**	6	20
Cyanoamino acid metabolism [00460]	0	0	0	0	0	0	1	1	0	0	2	0	3	0	0	1	1	0	6	16
Glutathione metabolism [00480]	0	0	0	0	0	0	2	1	1	0	2	***4**	3	8	0	1	1	2	17	43
Selenocompound metabolism [00450]	0	0	1	0	1	0	0	0	0	0	1	0	2	3	0	0	0	0	5	9
Taurine and hypotaurine metabolism [00430]	0	0	0	0	0	0	0	0	0	0	***2**	0	1	0	0	0	1	0	2	7

→ *cf.* page suivante

Table S5.4. (continued)

Metabolic pathway	6 HAP						24 HAP						48 HAP						Poll.	Rest
	CCP		CIP		HCP		CCP		CIP		HCP		CCP		CIP		HCP			
	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓		
7. Glycan biosynthesis and metabolism																				
Glycosaminoglycan degradation [00531]	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	27
Glycosphingolipid biosynthesis - ganglio series [00604]	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	26
N-Glycan biosynthesis [00510]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	9
Other glycan degradation [00511]	0	0	0	0	0	0	1	0	0	0	0	0	3	0	1	1	1	0	4	38
8. Metabolism of cofactors and vitamins																				
Folate biosynthesis [00790]	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	15
Nicotinate and nicotinamide metabolism [00760]	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	2	12
One carbon pool by folate [00670]	0	0	0	0	0	0	1	0	0	0	0	0	3	2	0	0	0	0	5	14
Pantothenate and CoA biosynthesis [00770]	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	2	19
Porphyrin and chlorophyll metabolism [00860]	0	**2	0	0	0	0	0	2	1	0	0	2	2	2	0	1	0	1	9	21
Retinol metabolism [00830]	0	1	0	**2	0	1	0	0	0	0	0	0	0	1	0	0	0	0	3	17
Riboflavin metabolism [00740]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	13
Thiamine metabolism [00730]	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	**2**	1	0	4	18
Ubiquinone and other terpenoid-quinone biosynthesis [00130]	**5	0	**4	0	**3	0	0	0	1	0	2	0	0	**6	1	0	0	0	11	22
9. Metabolism of terpenoids and polyketides																				
Biosynthesis of vancomycin group antibiotics [01055]	0	0	0	0	0	0	**2	0	**1	0	0	0	**2	0	0	0	0	0	**3	0
Carotenoid biosynthesis [00906]	0	0	0	0	0	0	0	0	0	0	**2	0	0	0	0	0	0	0	2	11
Diterpenoid biosynthesis [00904]	0	0	0	0	0	0	**3	0	0	0	0	0	0	0	**4	0	0	0	**4	1
Geraniol degradation [00281]	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	3	12
Insect hormone biosynthesis [00981]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	7
Limonene and pinene degradation [00903]	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	3	10
Polyketide sugar unit biosynthesis [00523]	0	0	0	0	0	0	**2	0	**1	0	0	0	**2	0	0	0	0	0	**3	0
Terpenoid backbone biosynthesis [00900]	0	0	0	0	0	0	0	1	1	0	0	1	2	3	0	1	0	0	5	9
10. Biosynthesis of other secondary metabolites																				
Acarbose and validamycin biosynthesis [00525]	0	0	0	0	0	0	**2	0	**1	0	0	0	**2	0	0	0	0	0	**3	0
Anthocyanin biosynthesis [00942]	0	0	0	0	0	0	0	0	**2	0	0	**2	0	0	0	0	0	**1	2	2
Betalain biosynthesis [00965]	**2	0	**2	0	0	0	0	0	0	0	0	0	0	0	**1	0	0	0	2	1
Caffeine metabolism [00232]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	10
Flavone and flavonol biosynthesis [00944]	1	0	1	0	0	0	0	0	1	0	0	**2	0	0	1	0	0	1	3	4
Indole alkaloid biosynthesis [00901]	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	6
Isoquinoline alkaloid biosynthesis [00950]	1	0	1	0	0	0	0	0	0	0	0	1	1	2	1	0	0	**2	7	11
Monobactam biosynthesis [00261]	0	0	**1	0	**1	0	0	0	0	0	1	0	0	**2	0	0	0	0	2	2
Novobiocin biosynthesis [00401]	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	5
Penicillin and cephalosporin biosynthesis [00311]	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2	14
Phenylpropanoid biosynthesis [00940]	1	1	0	0	0	0	1	2	1	0	2	2	4	8	1	0	0	2	16	46
Streptomycin biosynthesis [00521]	0	0	0	0	0	0	**2	0	1	0	0	0	2	0	0	0	0	0	3	9
Tropane, piperidine and pyridine alkaloid biosynthesis [00960]	0	0	0	0	0	0	0	0	1	0	0	0	1	2	0	1	0	1	5	13
11. Xenobiotics biodegradation and metabolism																				
Aminobenzoate degradation [00627]	**5	0	**4	0	**3	0	0	0	1	0	0	0	2	4	0	0	0	0	12	21
Benzoate degradation [00362]	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	3	12
Caprolactam degradation [00930]	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	10
Chloroalkane and chloroalkene degradation [00625]	0	1	0	**2	0	1	0	1	0	0	0	1	0	1	0	0	0	0	4	13
Chlorocyclohexane and chlorobenzene degradation [00361]	**6	0	**4	0	**3	0	0	0	0	0	0	0	0	0	0	0	0	0	**6	5
Drug metabolism - cytochrome P450 [00982]	0	1	0	**2	0	1	1	1	0	0	0	3	1	3	0	0	0	2	8	32
Drug metabolism - other enzymes [00983]	0	0	0	0	0	0	2	1	0	0	1	3	6	4	2	1	0	2	14	41
Fluorobenzoate degradation [00364]	**1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
Metabolism of xenobiotics by cytochrome P450 [00980]	0	1	0	**2	0	1	1	1	0	0	0	3	1	3	0	0	0	2	8	34
Naphthalene degradation [00626]	0	1	0	**2	0	**1	0	0	0	0	0	0	0	0	0	0	0	0	2	5
Styrene degradation [00643]	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2	9
Toluene degradation [00623]	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7

Table S5.4. (continued)

(b) Detailed view of metabolic pathway enrichment in each cluster

Prediction	Clusters																									Poll.	Rest
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25		
Total ESTs	**34**	**35**	**15**	**110**	**65**	**80**	**32**	**163**	**83**	**36**	**33**	**33**	**52**	**260**	**60**	**72**	**10**	**48**	**5**	**97**	**18**	**43**	**25**	**13**	**19**	**1441**	**5340**
ESTs involved in metabolic pathways	5	7	0	11	10	5	3	28	10	8	2	7	6	25	6	0	3	5	1	***24**	1	0	7	3	3	180	736
1. Carbohydrate metabolism																											
Amino sugar and nucleotide sugar metabolism [00520]	2	2	0	0	0	0	0	2	2	0	0	0	0	4	2	0	0	***3**	0	2	0	0	0	***2**	0	21	64
Ascorbate and aldarate metabolism [00053]	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	***2**	0	6	28
Butanoate metabolism [00650]	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	5	15
C5-Branched dibasic acid metabolism [00660]	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	***2**	0	0	0	0	0	4	8
Citrate cycle (TCA cycle) [00020]	0	0	0	0	1	0	0	0	0	1	0	0	***2**	1	1	0	0	0	0	2	0	0	0	0	0	8	26
Fructose and mannose metabolism [00051]	0	0	0	1	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	5	29
Galactose metabolism [00052]	0	0	0	0	0	0	0	0	0	1	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	4	***53**
Glycolysis / Gluconeogenesis [00010]	0	0	0	1	0	0	0	5	0	2	0	0	2	0	1	0	0	0	0	0	0	0	0	***2**	0	13	79
Glyoxylate and dicarboxylate metabolism [00630]	0	0	0	0	0	0	0	***4**	0	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	1	9	35
Inositol phosphate metabolism [00562]	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	***1**	1	0	0	0	0	0	4	24
Pentose and glucuronate interconversions [00040]	0	1	0	0	1	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	***2**	0	7	32
Pentose phosphate pathway [00030]	0	0	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	4	34
Propanoate metabolism [00640]	0	0	0	0	1	0	0	***7**	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	***12**	18
Pyruvate metabolism [00620]	0	0	0	0	1	0	0	***6**	0	1	0	0	1	2	1	0	0	0	0	0	0	0	0	0	0	12	62
Starch and sucrose metabolism [00500]	***3**	0	0	1	1	1	0	1	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	10	58
2. Energy metabolism																											
Carbon fixation in photosynthetic organisms [00710]	0	0	0	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	6	33
Carbon fixation pathways in prokaryotes [00720]	0	0	0	0	1	0	0	***6**	0	0	0	0	***2**	1	1	0	0	0	0	2	0	0	0	0	0	13	27
Methane metabolism [00680]	0	0	0	1	0	0	0	***4**	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	***2**	0	10	34
Nitrogen metabolism [00910]	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	3	13
Oxidative phosphorylation [00190]	0	0	1	0	1	0	0	0	0	***2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	35
Photosynthesis [00195]	0	0	0	0	0	***1**	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	1
Sulfur metabolism [00920]	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3	17
3. Lipid metabolism																											
alpha-Linolenic acid metabolism [00592]	0	0	0	1	0	0	0	3	0	1	0	0	0	2	0	0	0	1	0	2	0	0	1	***2**	0	***13**	22
Arachidonic acid metabolism [00590]	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	***2**	0	0	0	0	0	0	0	0	6	14
Ether lipid metabolism [00565]	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4
Fatty acid biosynthesis [00061]	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	18
Fatty acid degradation [00071]	0	0	0	0	2	1	0	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	***2**	1	10	30
Fatty acid elongation [00062]	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	19
Glycerolipid metabolism [00561]	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	4	28
Glycerophospholipid metabolism [00564]	0	1	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2	1	0	1	0	0	9	30
Linoleic acid metabolism [00591]	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	4	9
Sphingolipid metabolism [00600]	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2	***47**
Steroid biosynthesis [00100]	0	0	0	***2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	***6**	0	0	0	0	0	***8**	4
Steroid hormone biosynthesis [00140]	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	17
4. Nucleotide metabolism																											
Purine metabolism [00230]	0	0	0	4	1	0	0	3	2	0	0	0	0	6	1	1	0	1	0	0	1	1	0	0	0	21	69
Pyrimidine metabolism [00240]	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	1	0	0	1	0	0	6	24
5. Amino acid metabolism																											
Alanine, aspartate and glutamate metabolism [00250]	0	0	0	0	0	0	0	2	***2**	0	0	0	0	0	0	0	***1**	1	0	0	0	0	0	0	0	6	21
Arginine and proline metabolism [00330]	1	0	0	2	0	1	0	2	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	9	41
Arginine biosynthesis [00220]	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	4	16
Cysteine and methionine metabolism [00270]	0	1	0	0	0	0	0	3	2	1	1	0	0	2	0	0	0	0	0	1	0	0	1	0	0	12	48
Glycine, serine and threonine metabolism [00260]	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	2	0	0	0	***2**	0	7	44
Histidine metabolism [00340]	0	0	0	0	0	0	0	2	***2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	17
Lysine biosynthesis [00300]	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	10
Lysine degradation [00310]	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	32
Phenylalanine metabolism [00360]	1	0	0	0	0	0	0	***7**	2	1	0	0	0	1	0	0	0	1	0	2	0	0	0	0	0	15	38
Phenylalanine, tyrosine and tryptophan biosynth. [00400]	0	0	0	1	0	0	***2**	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	19
Tryptophan metabolism [00380]	1	0	0	0	1	0	0	3	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	8	34
Tyrosine metabolism [00350]	0	0	0	0	0	0	0	1	1	0	0	***2**	0	0	0	0	0	0	0	2	0	0	0	***2**	0	8	31
Valine, leucine and isoleucine biosynthesis [00290]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	15
Valine, leucine and isoleucine degradation [00280]	0	0	0	0	1	0	0	***3**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	25
6. Metabolism of other amino acids																											
beta-Alanine metabolism [00410]	0	0	0	0	0	0	0	***3**	0	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	6	20
Cyanoamino acid metabolism [00460]	***2**	***2**	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	6	16
Glutathione metabolism [00480]	0	***2**	0	0	2	0	0	4	0	1	***2**	0	1	1	0	0	1	0	0	2	0	1	0	0	0	17	43
Selenocompound metabolism [00450]	0	0	0	0	0	0	0	***2**	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	5	9
Taurine and hypotaurine metabolism [00430]	0	***2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	7

Table S5.4. (continued)

Prediction	Clusters																									Poll.	Rest
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25		
7. Glycan biosynthesis and metabolism																											
Glycosaminoglycan degradation [00531]	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2	27
Glycosphingolipid biosynthesis - ganglio series [00604]	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2	26
N-Glycan biosynthesis [00510]	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9
Other glycan degradation [00511]	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	4	38
8. Metabolism of cofactors and vitamins																											
Folate biosynthesis [00790]	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	15
Nicotinate and nicotinamide metabolism [00760]	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	2	12
One carbon pool by folate [00670]	0	0	0	0	0	0	0	2	0	0	0	0	0	***3**	0	0	0	0	0	0	0	0	0	0	0	5	14
Pantothenate and CoA biosynthesis [00770]	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	19
Porphyrin and chlorophyll metabolism [00860]	1	0	0	1	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	***2**	9	21
Retinol metabolism [00830]	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	***2**	0	3	17
Riboflavin metabolism [00740]	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	13
Thiamine metabolism [00730]	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	4	18
Ubiquinone and other terpenoid-quinone biosynth. [00130]	0	0	0	0	0	0	0	***5**	1	0	0	***4**	0	0	0	0	0	0	0	1	0	0	0	0	0	11	22
9. Metabolism of terpenoids and polyketides																											
Biosynthesis of vancomycin group antibiotics [01055]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	***1**	0	0	0	0	***1**	0	0	0	0	0	***3**	0
Carotenoid biosynthesis [00906]	0	***2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	11
Diterpenoid biosynthesis [00904]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	***4**	0	0	***4**	1
Geraniol degradation [00281]	0	0	0	0	1	0	0	***2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	12
Insect hormone biosynthesis [00981]	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7
Limonene and pinene degradation [00903]	0	0	0	0	0	0	0	***3**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	10
Polyketide sugar unit biosynthesis [00523]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	***1**	0	0	0	0	***1**	0	0	0	0	0	***3**	0
Terpenoid backbone biosynthesis [00900]	0	0	0	0	0	0	0	0	***3**	0	0	0	0	0	0	0	0	0	0	***2**	0	0	0	0	0	5	9
10. Biosynthesis of other secondary metabolites																											
Acarbose and validamycin biosynthesis [00525]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	***1**	0	0	0	0	***1**	0	0	0	0	0	***3**	0
Anthocyanin biosynthesis [00942]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	***2**	0	0	0	0	0	2	2
Betalain biosynthesis [00965]	0	0	0	0	0	0	0	0	0	0	0	***2**	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
Caffeine metabolism [00232]	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	10
Flavone and flavonol biosynthesis [00944]	0	0	0	0	0	0	0	0	0	0	0	***1**	0	0	0	0	0	0	0	***2**	0	0	0	0	0	3	4
Indole alkaloid biosynthesis [00901]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	6
Isoquinoline alkaloid biosynthesis [00950]	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	***3**	0	0	1	0	0	7	11
Monobactam biosynthesis [00261]	0	0	0	0	0	0	0	***2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
Novobiocin biosynthesis [00401]	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5
Penicillin and cephalosporin biosynthesis [00311]	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	14
Phenylpropanoid biosynthesis [00940]	1	0	0	0	2	0	0	***6**	0	1	0	0	0	3	1	0	0	0	0	1	0	0	0	1	0	16	46
Streptomycin biosynthesis [00521]	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	3	9
Tropane, piperidine and pyridine alkaloid biosynth. [00960]	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	***1**	0	0	***2**	0	0	0	0	0	5	13
11. Xenobiotics biodegradation and metabolism																											
Aminobenzoate degradation [00627]	1	0	0	0	1	0	0	2	0	1	0	***4**	0	1	1	0	0	0	0	1	0	0	0	0	0	12	21
Benzoate degradation [00362]	0	0	0	0	0	0	0	***2**	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	12
Caprolactam degradation [00930]	0	0	0	0	0	0	0	***2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	10
Chloroalkane and chloroalkene degradation [00625]	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	***1**	0	0	0	0	0	0	***2**	0	4	13
Chlorocyclohexane and chlorobenzene degradation [00361]	0	0	0	0	0	0	0	0	0	0	0	***5**	0	0	0	0	0	0	0	1	0	0	0	0	0	***6**	5
Drug metabolism - cytochrome P450 [00982]	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	***2**	0	8	32
Drug metabolism - other enzymes [00983]	0	1	0	0	1	0	0	1	1	0	0	0	0	3	0	0	0	0	0	***4**	0	1	***2**	0	0	14	41
Fluorobenzoate degradation [00364]	0	0	0	0	0	0	0	0	0	0	0	***1**	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
Metabolism of xenobiotics by cytochrome P450 [00980]	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	***2**	0	8	34
Naphthalene degradation [00626]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	***2**	0	2	5
Styrene degradation [00643]	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	9
Toluene degradation [00623]	0	0	0	0	0	0	0	0	0	0	0	***1**	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7

ANNEXE D

Matériel supplémentaire du chapitre 6

Dataset S6.1. Details on assembled sequences. (a) Details on assembled genes. **(b)** Details on assembled transcripts. → *cf.* Fichier XLSX annexé à la présente thèse.

Dataset S6.2. Differential gene expression analysis. (a) Normalized read counts. **(b)** Differential expression analysis: "Anth" vs. "Leaf". **(c)** Differential expression analysis: "2 DBA" vs. "Anth". **(d)** Differential expression analysis: "*frk1*" vs. "Anth". → *cf.* Fichier XLSX annexé à la présente thèse.

Dataset S6.3. *In silico* predictions on assembled peptide sequences. (a) BLASTp comparison to RefSeq protein sequences. **(b)** Domain annotation results obtained with InterProScan. **(c)** Functional classification into Gene Ontology categories. **(d)** Enzyme code predictions. **(e)** Transcription factor predictions. **(f)** Signal peptide predictions. **(g)** Cysteine-rich protein predictions. **(h)** *N*-glycosylation predictions. **(i)** Glycosylphosphatidylinositol anchor predictions. **(j)** Transmembrane helix predictions. **(k)** Classification of putative receptor-like kinases and proteins. → *cf.* Fichier XLSX annexé à la présente thèse.

Dataset S6.4. GO enrichment analyses. (a–c) GO-terms enriched in ovule-enriched transcripts. **(d–f)** GO-terms enriched in 2 DBA up-regulated transcripts. **(g–i)** GO-terms enriched in 2 DBA down-regulated transcripts. **(j–l)** GO-terms enriched in *frk1* up-regulated transcripts. **(m–o)** GO-terms enriched in *frk1* down-regulated transcripts. **(p–r)** GO-terms enriched in ovule-enriched transcripts down-regulated in both 2 DBA and *frk1*. → *cf.* Fichier XLSX annexé à la présente thèse.

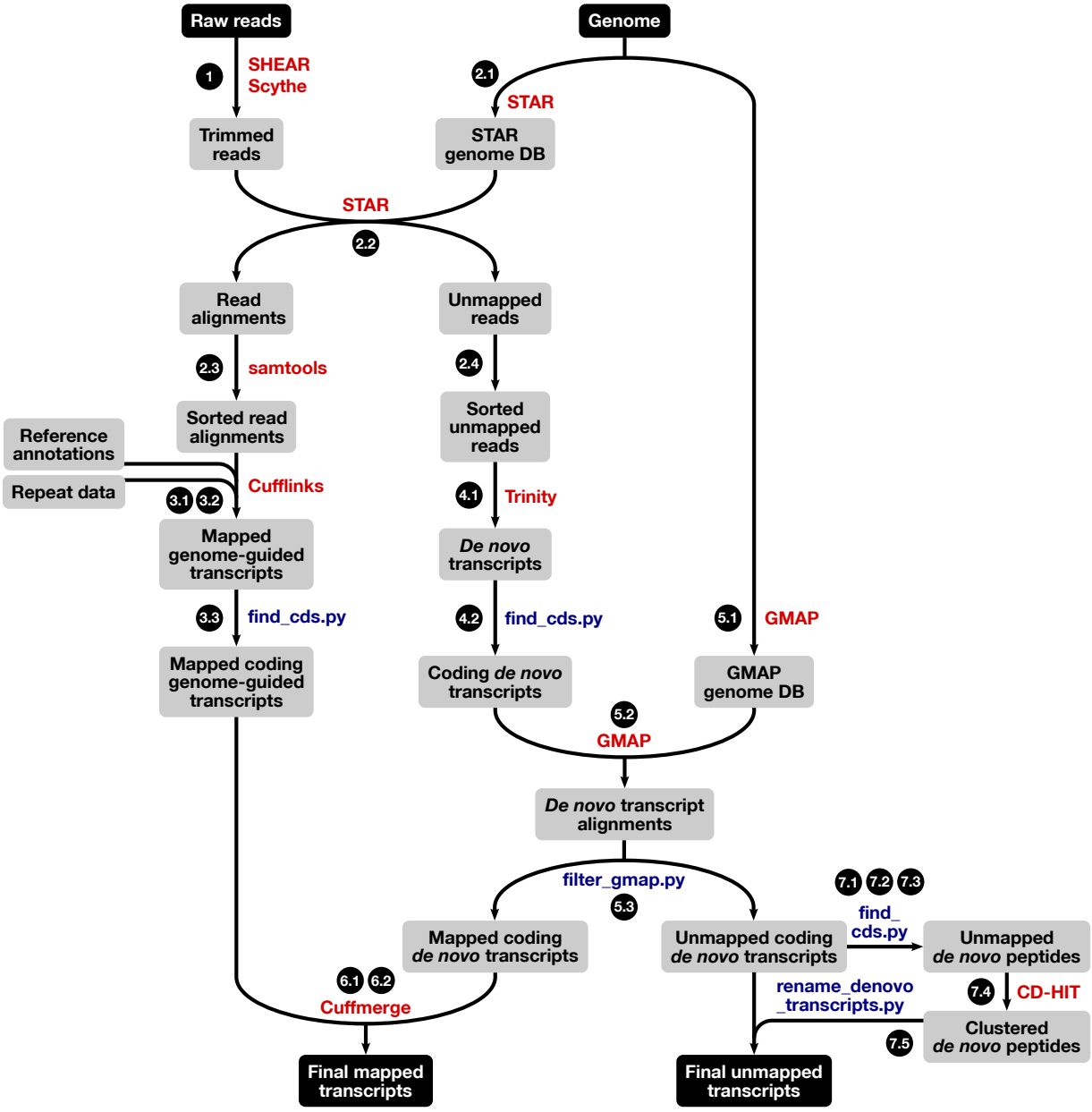


Figure S6.1. Assembly workflow. Names in red and blue refer to external and home-made programs, respectively. Circled numbers refer to steps depicted in the detailed commands file available on this paper's Dryad repository.

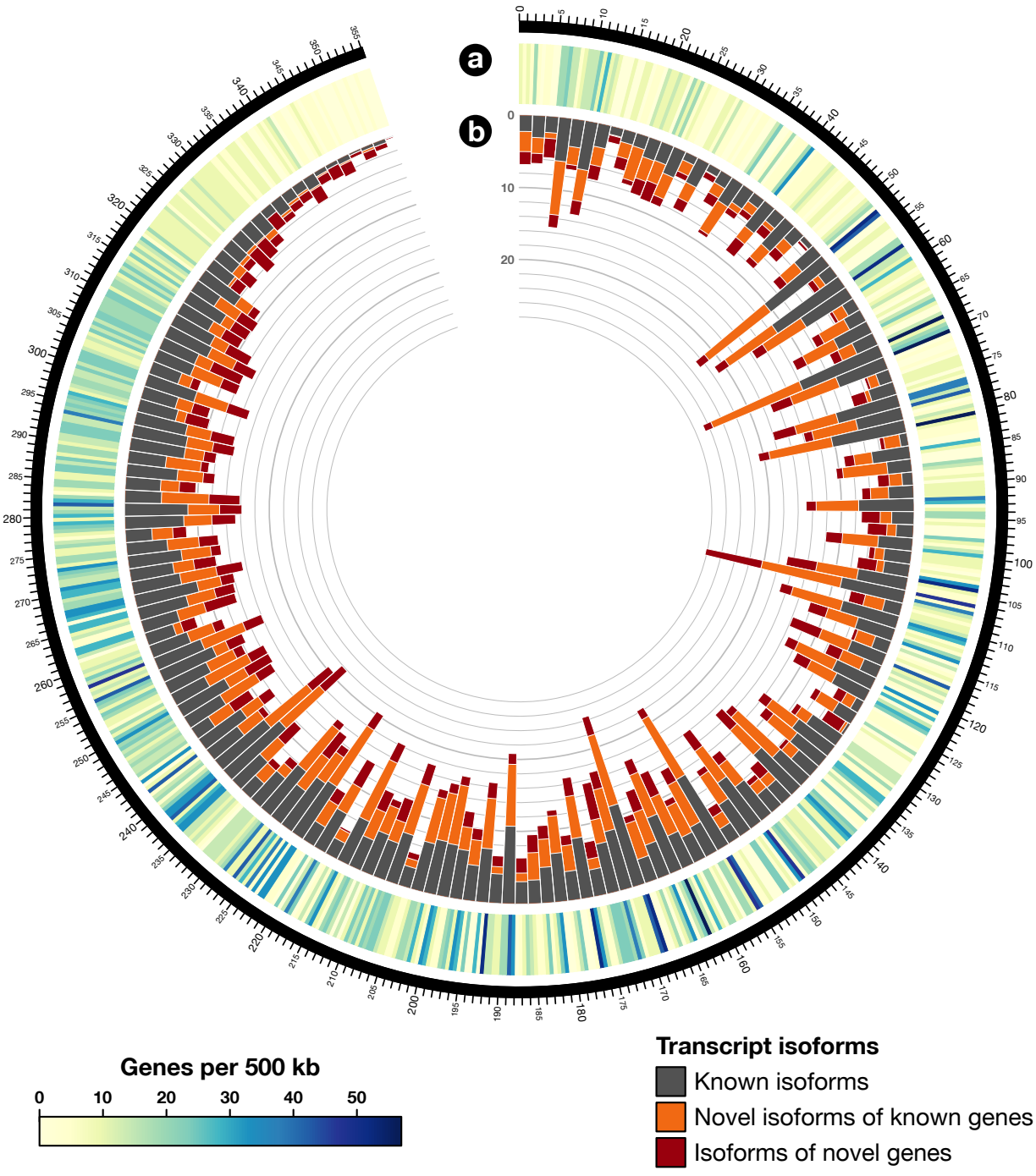
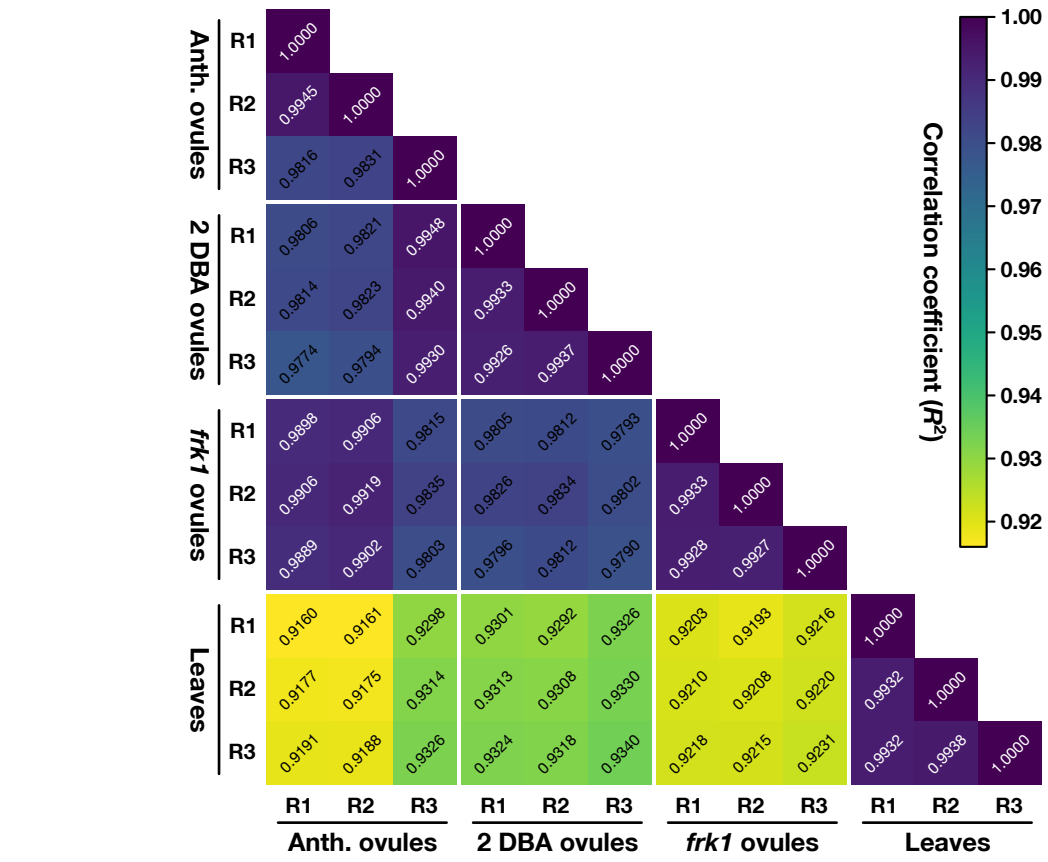
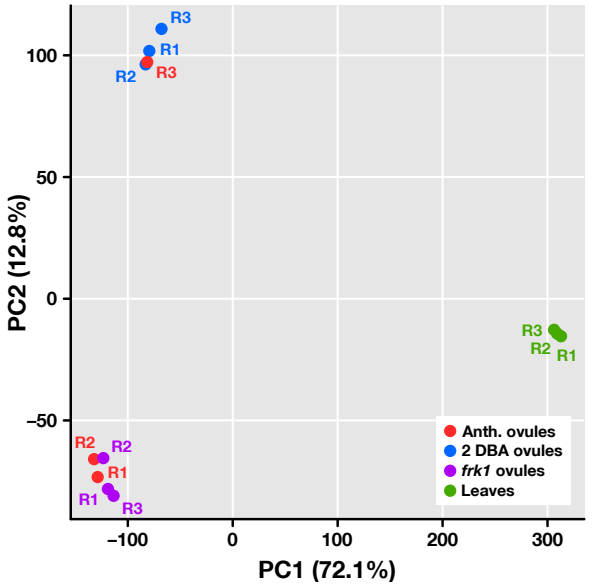


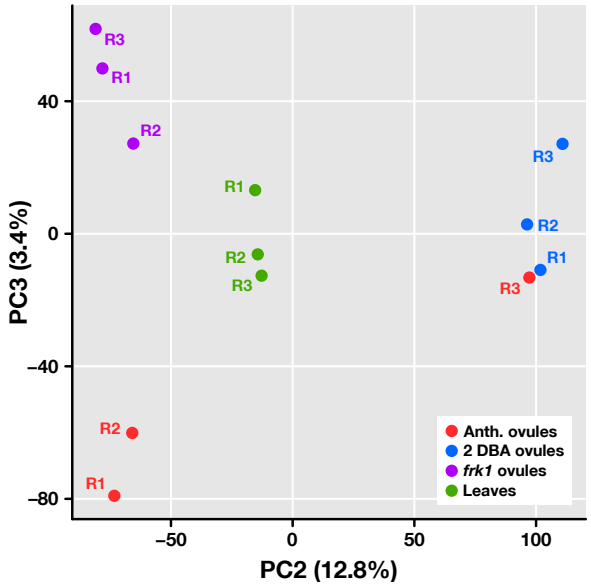
Figure S6.2. Distribution of gene and transcript annotations along *S. chacoense* chrUn (unanchored scaffolds). **(a)** Heatmap depicting gene density, expressed in number of genes per 500 kb. **(b)** Histogram view of transcript annotations, in terms of number of transcripts per 1,5 Mb: in gray, transcript isoforms described by Leisner *et al.*; in orange, new isoforms found for known genes; in red, isoforms of novel genes detected original to our assembly.



(a) Heatmap of pairwise correlation coefficients

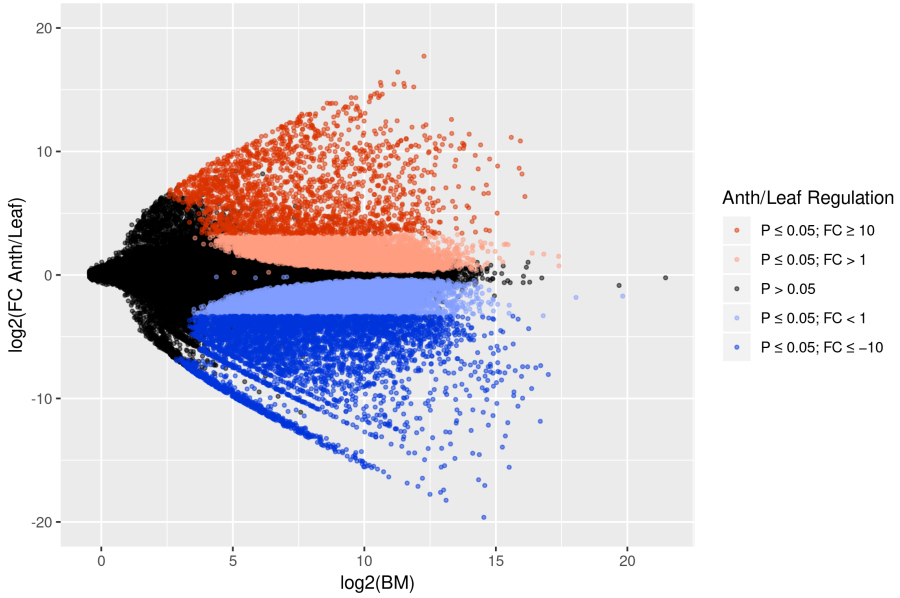


(b) Principal component analysis: PC1 vs. PC2

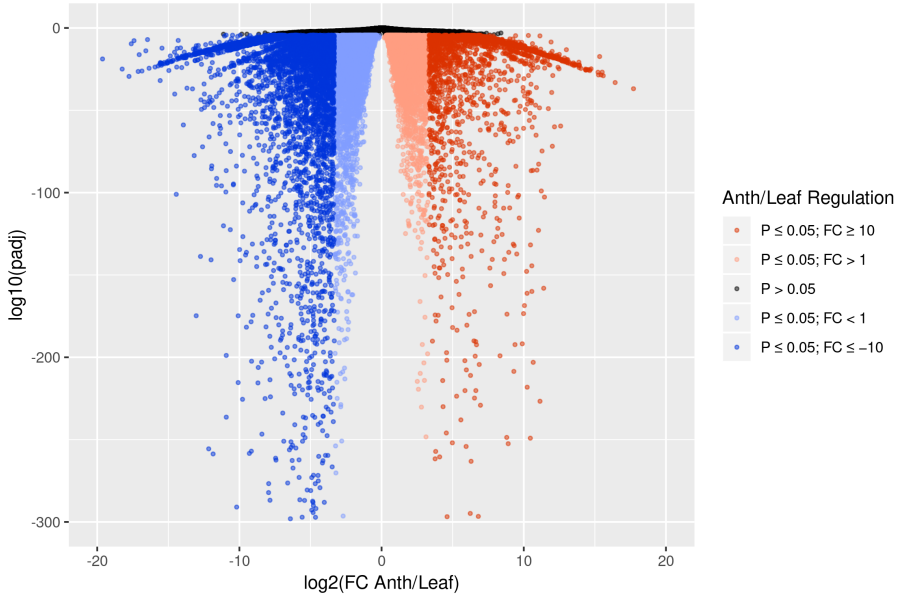


(c) Principal component analysis: PC2 vs. PC3

Figure S6.3. Preliminary analyses on gene expression data. Pairwise correlation coefficients (a) and principal component analysis (b, c) were computed on DESeq2 normalized read counts across samples.

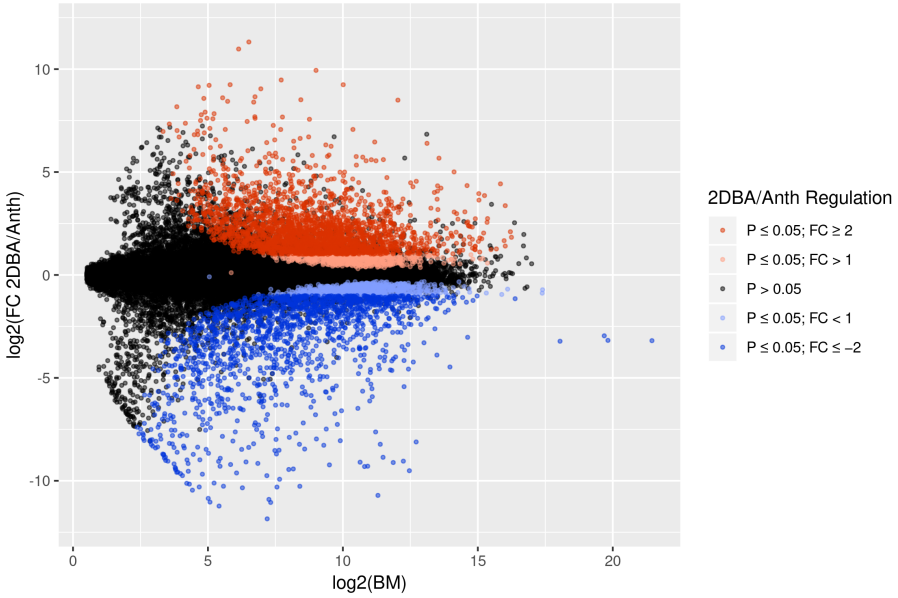


(a) MA plot

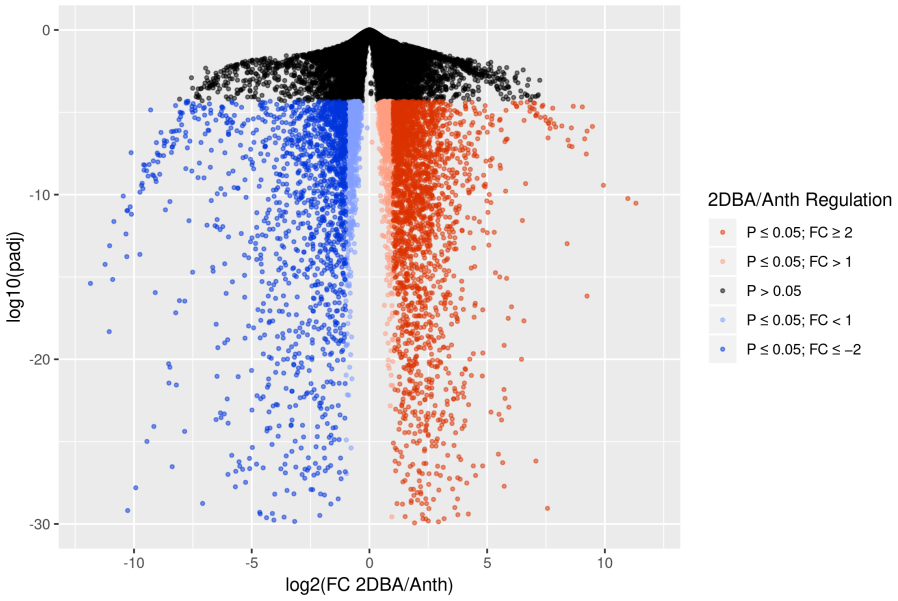


(b) Volcano plot

Figure S6.4. Differential expression analysis plots: Anth vs. Leaf. Differential gene expression (DGE) data were obtained with the DESeq2 package after read quantification by Salmon. For each gene, the MA plot represents $\log_2(\text{FC})$, the binary logarithm of the Anth/Leaf expression fold-change, as a function of $\log_2(\text{BM})$, the binary logarithm of the expression base mean, i.e. the mean of normalized counts across all samples under study. The Volcano plot represents $\log_{10}(p_{\text{adj}})$, the decimal logarithm of the adjusted p-value, as a function of $\log_2(\text{FC})$. Black points represent genes that were not significantly DE ($P > 0.05$). Light blue/red points represent genes with $P \leq 0.05$ but $|\text{FC}| < 5$. Dark blue/red points represent genes that were considered differentially expressed in this study ($P \leq 0.05$ and $|\text{FC}| \geq 5$).

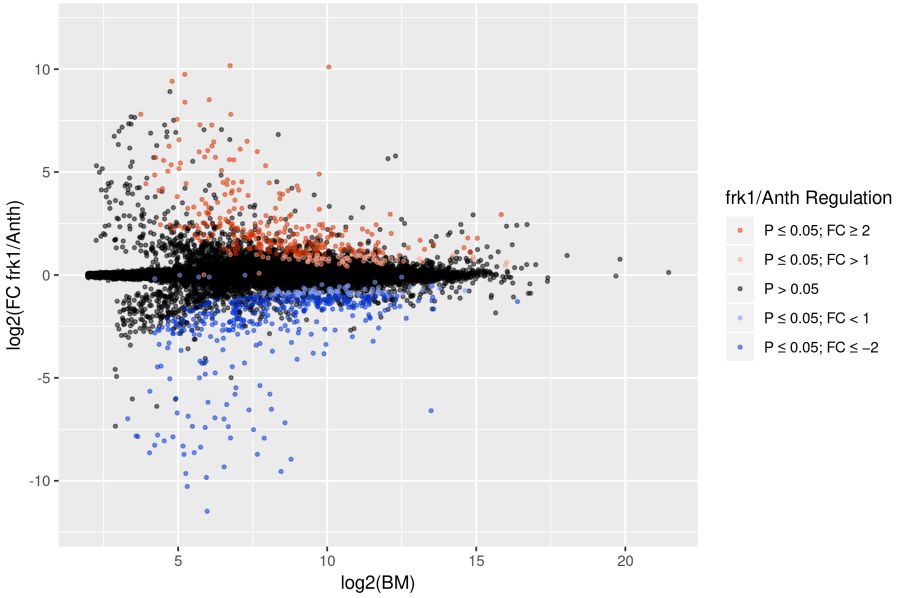


(a) MA plot

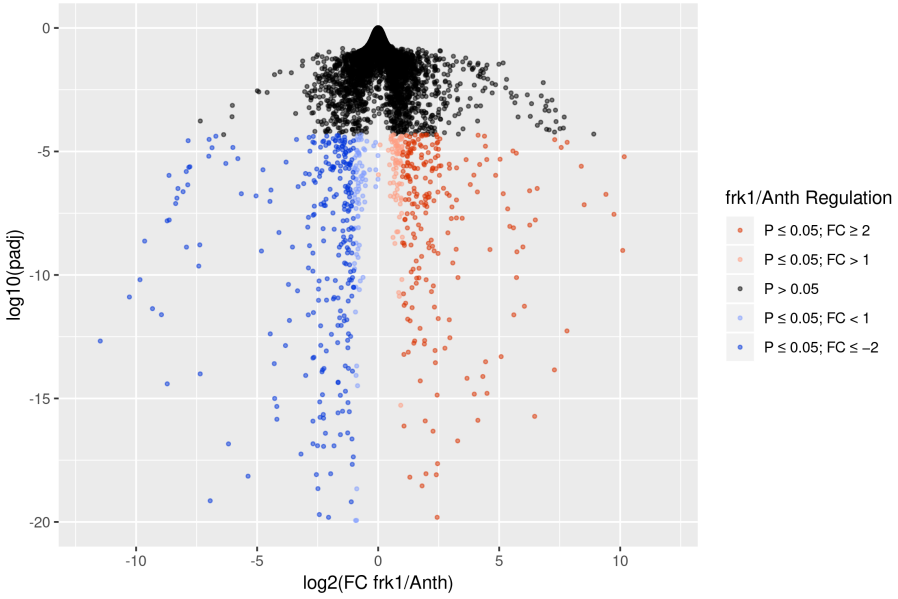


(b) Volcano plot

Figure S6.5. Differential expression analysis plots: 2 DBA vs. Anth. Differential gene expression (DGE) data were obtained with the DESeq2 package after read quantification by Salmon. For each gene, the MA plot represents $\log_2(\text{FC})$, the binary logarithm of the 2 DBA/Anth expression fold-change, as a function of $\log_2(\text{BM})$, the binary logarithm of the expression base mean, i.e. the mean of normalized counts across all samples under study. The Volcano plot represents $\log_{10}(p_{\text{adj}})$, the decimal logarithm of the adjusted p-value, as a function of $\log_2(\text{FC})$. Black points represent genes that were not significantly DE ($P > 0.05$). Light blue/red points represent genes with $P \leq 0.05$ but $|\text{FC}| < 5$. Dark blue/red points represent genes that were considered differentially expressed in this study ($P \leq 0.05$ and $|\text{FC}| \geq 5$).



(a) MA plot



(b) Volcano plot

Figure S6.6. Differential expression analysis plots: *frk1* vs. Anth. Differential gene expression (DGE) data were obtained with the DESeq2 package after read quantification by Salmon. For each gene, the MA plot represents $\log_2(\text{FC})$, the binary logarithm of the *frk1*/Anth expression fold-change, as a function of $\log_2(\text{BM})$, the binary logarithm of the expression base mean, i.e. the mean of normalized counts across all samples under study. The Volcano plot represents $\log_{10}(p_{\text{adj}})$, the decimal logarithm of the adjusted p-value, as a function of $\log_2(\text{FC})$. Black points represent genes that were not significantly DE ($P > 0.05$). Light blue/red points represent genes with $P \leq 0.05$ but $|\text{FC}| < 5$. Dark blue/red points represent genes that were considered differentially expressed in this study ($P \leq 0.05$ and $|\text{FC}| \geq 5$).

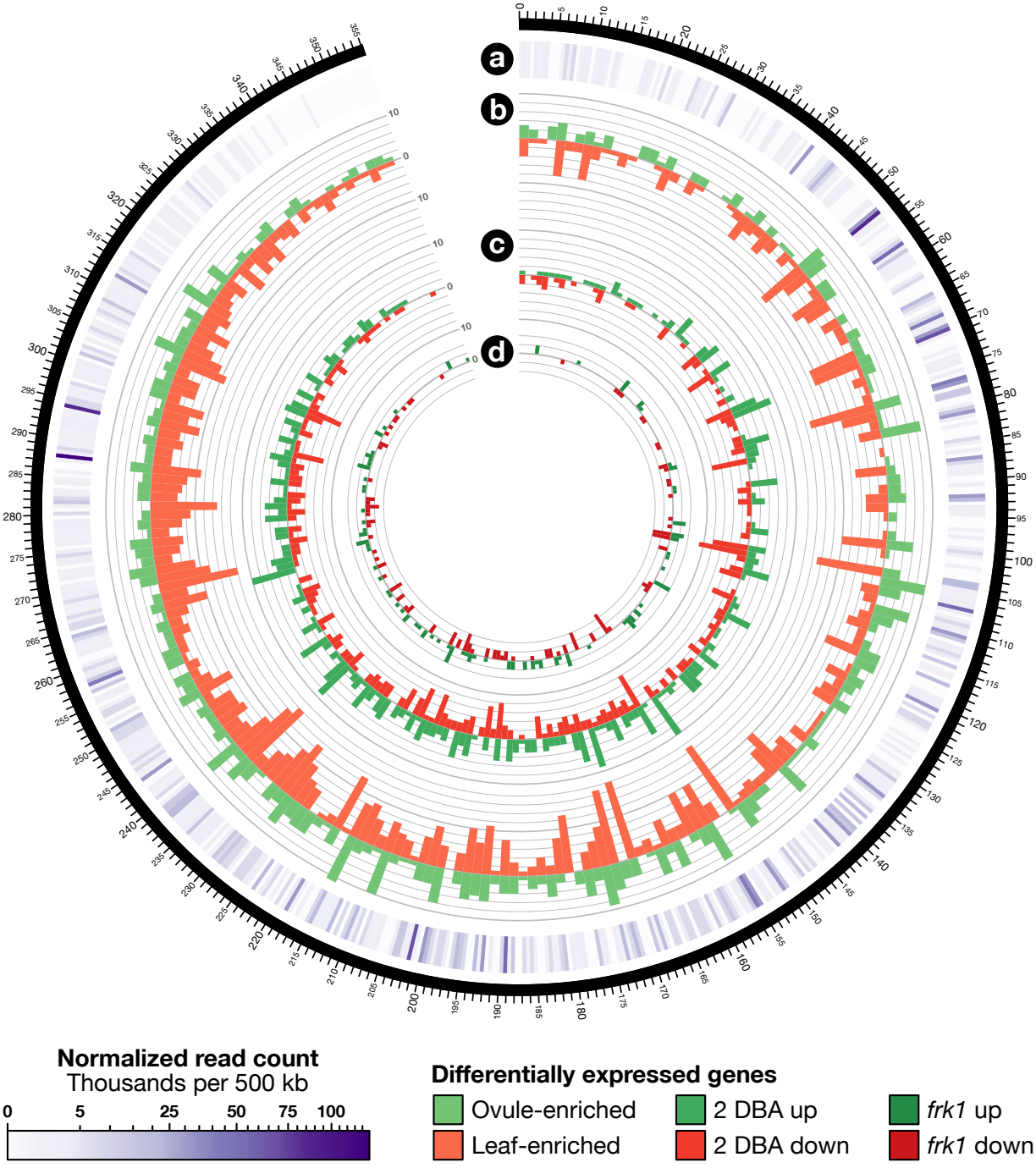


Figure S6.7. Distribution of differentially expressed genes along *S. chacoense* chrUn (unanchored scaffolds). **(a)** Heatmap representing the expression level, in terms of normalized read count per 500 kb. **(b–d)** Histograms depicting the number of genes up- and down-regulated in Anth vs. Leaf, 2 DBA vs. Anth, and *frk1* vs. Anth, respectively, per 1,5 Mb.

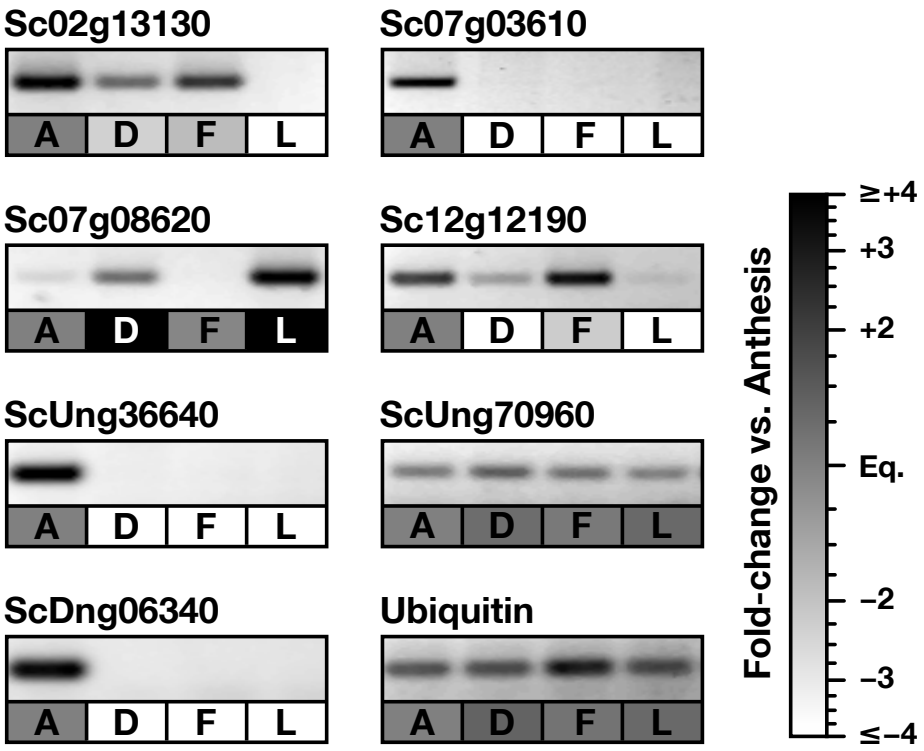


Figure S6.8. Validation of DGE analysis by semi-quantitative PCR. Seven genes displaying different expression patterns in WT ovules at anthesis (A), WT ovules 2 DBA (D), *frk1* ovules at anthesis (F) and WT leaves (L) were randomly chosen. Ubiquitin (ScDng00525) was used as the internal control.

Table S6.1. Details about Illumina read libraries used in this study.

Cond.	Rep.	SRA Run	Raw reads	Trimmed reads	Mapped reads		Unmapped reads
					Total	Unique	
Anth	R1	SRR2782690	93 204 830	91 896 894	80 280 200 87.4%	76 589 004 83.3%	11 616 694 12.6%
	R2	SRR2782691	117 704 070	116 202 906	102 534 190 88.2%	97 674 236 84.1%	13 668 716 11.8%
	R3	SRR2782692	87 553 120	86 494 284	83 140 892 96.1%	79 120 416 91.5%	3 353 392 3.9%
2 DBA	R1	SRR2782581	100 419 074	99 277 938	95 655 420 96.4%	91 071 884 91.7%	3 622 518 3.6%
	R2	SRR2782415	112 147 762	110 320 024	107 052 532 97.0%	101 945 802 92.4%	3 267 492 3.0%
	R3	SRR2782416	107 621 552	105 805 002	101 742 852 96.2%	96 550 658 91.3%	4 062 150 3.8%
frk1	R1	SRR2782417	107 315 132	105 395 004	89 440 274 84.9%	85 183 556 80.8%	15 954 730 15.1%
	R2	SRR2782418	128 173 852	126 116 274	106 620 332 84.5%	101 552 284 80.5%	19 495 942 15.5%
	R3	SRR2782419	134 662 860	132 502 456	112 873 190 85.2%	107 764 928 81.3%	19 629 266 14.8%
Leaf	R1	SRR2782438	109 994 694	108 176 048	94 945 558 87.8%	89 250 196 82.5%	13 230 490 12.2%
	R2	SRR2782462	101 439 608	99 784 606	88 865 460 89.1%	83 588 436 83.8%	10 919 146 10.9%
	R3	SRR2782513	100 726 756	98 864 430	85 448 000 86.4%	80 503 618 81.4%	13 416 430 13.6%
Total			**1 300 963 310**	**1 280 835 866**	**1 148 598 900** **89.7%**	**1 090 795 018** **85.2%**	**132 236 966** **10.3%**

Table S6.2. Primers used for semi-quantitative RT-PCR DGE validation.

Target	Forward primer	Reverse primer
Sc02g13130	TCTCCAATTGGTGTTGGCTT	AGCAGGATTCCTCTGAGATCTATG
Sc07g03610	CCCGGATCCGGTCCAAGGTTTCGGGCC	CCCAAGCTTTTAATCTCCTTTATTTGCAAATGAAAA
Sc07g08620	GAAAACTTCATCGCAAGCC	ACATTCTTGTAAGGTTTAAATTCCCT
Sc12g12190	TTCCATATGGTCCATCTCCTTG	GTTCCGTTCTCTCATTCCCG
ScDng06340	CCCGGATCCGTATCCTCCAAGAATAGAAAGTATTTGT	CCCAAGCTTTCACCCACTAATATAACCCCCA
ScUng36640	CCCGGATCCGCGACCTTTAATCACATTTAGTAGTGC	CCCAAGCTTTTATGGAGCCAAATTAGGAGTTG
ScUng70960	TGCAGATCATTCATTCGTATGTATG	CATGCATTTTTGTGTATTCCAACT
Ubiquitin	GCTGGCAAGCAGTTGGAA	TCTGCATACCTCCCCTGAG

Table S6.3. Summary of signal peptide and subsequent predictions, with enrichment analyses. *Solanum chacoense* protein sequences were checked for the presence of a signal peptide with SignalP v. 5.0. Presumably secreted proteins (SPs) were further checked for the presence of transmembrane helices (TMHs) with TMHMM v. 2.0. SPs with one TMH were considered to be potential RLKs/RLPs, while those with two TMHs or more were tagged as other membrane proteins. SPs with no TMH were checked for the presence of glycosylphosphatidylinositol (GPI) anchors with PredGPI. Those with no GPI were considered to be non-membrane proteins. Furthermore, cysteine-rich proteins (CRPs; 6+ cysteines, mature part ≤ 150 aa) were quantified, and the presence of potential *N*-glycosylated residues was predicted with NetNGlyc-1.0 on all SPs. Proteins with at least one residue associated to a score of "++" or better were considered *N*-glycosylated. Transcripts associated to each prediction were counted in the whole dataset, as well as in ovule-enriched (O↑), 2 DBA up-regulated (D↑), 2 DBA down-regulated (D↓), *frk1* up-regulated (F↑), and *frk1* down-regulated (F↓) transcripts, as well as ovule-enriched transcripts that were down-regulated in both 2 DBA and *frk1* conditions (O↑D↓F↓). Asterisks (*) indicate a significant enrichment (Fisher's exact test, $P \leq 0.05$).

Prediction	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Membrane-bound SPs	1172	69	*184	57	*23	13	6
GPI-anchored proteins	128	13	*23	*11	1	*5	*4
Potential RLKs/RLPs	795	44	*136	29	*18	5	2
Other membrane SPs	249	12	*25	17	4	3	0
Non-membrane SPs	3578	*509	*503	*306	*99	*113	*63
Cysteine-rich proteins	478	*138	*101	*71	*16	*30	*22
N-glycosylated proteins	2469	*264	*383	*169	*63	*57	*27
Total secreted proteins	**4750**	***578**	***687**	***363**	***122**	***126**	***69**
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**

Table S6.4. Summary of RLK/RLP predictions, with enrichment analyses. *Solanum chacoense* protein sequences predicted to contain a signal peptide and one transmembrane helix were considered to be putative RLKs/RLPs and were checked for the presence of Pfam motifs with HMMER. Receptors with a predicted kinase domain in the intracellular (C-terminal) part were classified as receptor-like kinases (RLKs); the rest were classified as receptor-like proteins (RLPs). RLKs and RLPs were then divided into subcategories based for the presence of predicted LRR, lectin, and malectin domains in their extracellular part. Transcripts associated to each prediction were counted in the whole dataset, as well as in ovule-enriched (O↑), 2 DBA up-regulated (D↑), 2 DBA down-regulated (D↓), *frk1* up-regulated (F↑), and *frk1* down-regulated (F↓) transcripts, as well as ovule-enriched transcripts that were down-regulated in both 2 DBA and *frk1* conditions (O↑D↓F↓). Asterisks (*) indicate a significant enrichment (Fisher's exact test, $P \leq 0.05$).

Prediction	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Receptor-like kinases (RLKs)	392	21	*85	20	*11	3	1
LRR-RLKs	153	8	*39	5	3	1	0
Lectin-RLKs	84	5	*20	*9	3	1	0
Malectin-RLKs	33	3	5	2	1	1	1
Other RLKs	122	5	*21	4	*4	0	0
Receptor-like proteins (RLPs)	403	23	*51	9	7	2	1
LRR-RLPs	70	3	6	2	*3	1	1
Other RLPs	333	20	*45	7	4	1	0
Total RLKs/RLPs	**795**	**44**	***136**	**29**	***18**	**5**	**2**
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**

Table S6.5. Summary of transcription factor predictions, with enrichment analyses. *Solanum chacoense* protein sequences were scanned for potential transcription factors (TFs) with the PlantTFDB v. 4.0 prediction tool. Transcripts predicted to encode TFs were counted in the whole dataset, as well as in ovule-enriched (O↑), 2 DBA up-regulated (D↑), 2 DBA down-regulated (D↓), *frk1* up-regulated (F↑), and *frk1* down-regulated (F↓) transcripts, as well as ovule-enriched transcripts that were down-regulated in both 2 DBA and *frk1* conditions (O↑D↓F↓). Asterisks (*) indicate a significant enrichment (Fisher's exact test, $P \leq 0.05$).

Prediction	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
AP2	28	2	*5	1	0	1	0
ARF	24	4	4	2	0	0	0
ARR-B	20	1	1	0	0	0	0
B3	85	*15	4	3	1	1	0
BBR-BPC	6	0	0	0	0	0	0
BES1	9	0	1	0	0	0	0
bHLH	166	*21	*25	11	4	3	0
bZIP	79	*10	6	3	0	0	0
C2H2	117	*14	8	8	0	0	0
C3H	53	3	2	2	0	1	0
CAMTA	9	0	0	0	0	0	0
CO-like	15	3	3	1	*2	1	0
CPP	4	0	0	0	0	0	0
DBB	7	2	0	0	0	0	0
Dof	37	2	*10	1	0	0	0
E2F/DP	9	0	1	0	0	0	0
EIL	9	0	0	1	0	0	0
ERF	188	9	12	*17	1	0	0
FAR1	29	0	0	0	0	0	0
G2-like	63	4	3	3	0	0	0
GATA	35	3	*7	1	1	1	0
GRAS	71	6	*9	3	2	0	0
GRF	11	2	*6	0	0	0	0
GeBP	16	2	0	0	0	0	0
HB-PHD	2	0	1	0	0	0	0
HB-other	11	0	0	0	0	0	0
HD-ZIP	59	*14	7	*10	1	0	0
HRT-like	2	0	0	0	0	0	0
HSF	32	3	4	2	0	0	0
→ *continued on next page*							
Total transcription factors	**2124**	***212**	***196**	***130**	**24**	**21**	**5**
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**

Table S6.5. (continued)

Prediction	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
LBD	52	*8	2	1	0	0	0
LFY	1	0	1	0	0	0	0
LSD	5	0	1	0	0	0	0
M-type MADS	130	7	2	4	1	1	1
MIKC-MADS	41	*18	3	3	1	0	0
MYB	143	13	13	*19	1	*5	1
MYB-related	81	8	5	2	3	1	0
NAC	140	9	11	*15	2	1	1
NF-X1	2	0	0	0	0	0	0
NF-YA	11	*3	0	1	0	0	0
NF-YB	27	3	1	1	0	*2	1
NF-YC	21	2	1	0	0	0	0
NZZ/SPL	1	1	1	0	0	0	0
Nin-like	20	2	0	2	0	0	0
RAV	2	1	1	1	0	0	0
S1Fa-like	2	0	0	0	0	0	0
SAP	2	1	1	0	*1	0	0
SBP	16	0	3	0	0	0	0
SRS	9	*3	1	0	0	0	0
STAT	1	0	0	0	0	0	0
TALE	23	2	3	*4	0	0	0
TCP	31	0	*7	1	0	1	0
Trihelix	35	4	2	0	0	0	0
VOZ	2	0	0	0	0	0	0
WOX	11	2	*6	0	0	0	0
WRKY	88	2	4	6	3	1	1
Whirly	2	0	0	0	0	0	0
YABBY	9	2	1	0	0	1	0
ZF-HD	20	1	*7	1	0	0	0
Total transcription factors	**2124**	***212**	***196**	***130**	**24**	**21**	**5**
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**

Table S6.6. Enzymatic pathway enrichment analysis. *Solanum chacoense* protein sequences were compared to the RefSeq protein database with BLASTp. Enzyme codes were then retrieved with Blast2GO and match to metabolic pathways using the KEGG database. Transcripts predicted to be involved in each metabolic pathway were counted in the whole dataset, as well as in ovule-enriched (O↑), 2 DBA up-regulated (D↑), 2 DBA down-regulated (D↓), *frk1* up-regulated (F↑), and *frk1* down-regulated (F↓) transcripts, as well as ovule-enriched transcripts that were down-regulated in both 2 DBA and *frk1* conditions (O↑D↓F↓). Asterisks (*) indicate a significant enrichment (Fisher's exact test, $P \leq 0.05$).

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
0. Global and overview maps							
Metabolic pathways (01100)	2261	*175	*214	*152	34	33	11
Biosynthesis of secondary metabolites (01110)	1039	68	*97	*67	15	9	4
Biosynthesis of antibiotics (01130)	513	20	39	20	6	4	0
Microbial metabolism in diverse environments (01120)	460	19	34	23	7	3	0
1. Carbohydrate metabolism							
Amino sugar and nucleotide sugar metabolism (00520)	146	*16	8	*13	1	1	0
Ascorbate and aldarate metabolism (00053)	24	2	3	0	0	1	0
Butanoate metabolism (00650)	30	1	*5	1	1	0	0
C5-Branched dibasic acid metabolism (00660)	4	0	0	1	0	0	0
Citrate cycle (TCA cycle) (00020)	49	1	0	0	0	1	0
Fructose and mannose metabolism (00051)	93	3	8	2	3	0	0
Galactose metabolism (00052)	64	4	8	3	1	0	0
Glycolysis / Gluconeogenesis (00010)	116	3	9	4	2	2	0
Glyoxylate and dicarboxylate metabolism (00630)	88	8	8	7	0	0	0
Inositol phosphate metabolism (00562)	54	3	4	1	2	1	0
Pentose and glucuronate interconversions (00040)	202	*25	*37	13	*8	*7	*3
Pentose phosphate pathway (00030)	83	2	4	2	1	0	0
Propanoate metabolism (00640)	42	6	1	0	0	0	0
Pyruvate metabolism (00620)	105	7	4	6	0	*4	0
Starch and sucrose metabolism (00500)	158	11	*17	11	2	4	1
2. Energy metabolism							
Carbon fixation in photosynthetic organisms (00710)	93	7	9	4	1	0	0
Carbon fixation pathways in prokaryotes (00720)	66	7	2	0	1	0	0
Methane metabolism (00680)	94	8	8	2	2	0	0
Nitrogen metabolism (00910)	32	5	2	*4	1	*3	*2
Oxidative phosphorylation (00190)	39	1	0	3	0	0	0
Sulfur metabolism (00920)	35	5	1	4	0	0	0
3. Lipid metabolism							
alpha-Linolenic acid metabolism (00592)	23	0	3	1	1	1	0
Arachidonic acid metabolism (00590)	22	0	2	0	0	0	0
Biosynthesis of unsaturated fatty acids (01040)	28	1	2	0	0	0	0
Cutin, suberine and wax biosynthesis (00073)	41	3	*7	*6	1	0	0
Ether lipid metabolism (00565)	27	1	1	2	0	0	0
Fatty acid biosynthesis (00061)	44	1	*8	1	1	0	0
Fatty acid degradation (00071)	24	1	2	1	1	1	0
Fatty acid elongation (00062)	17	1	1	0	1	0	0
Glycerolipid metabolism (00561)	57	1	7	4	0	1	0
Glycerophospholipid metabolism (00564)	63	1	4	3	1	0	0
Linoleic acid metabolism (00591)	5	0	1	0	1	0	0
Primary bile acid biosynthesis (00120)	4	0	0	0	0	0	0
Sphingolipid metabolism (00600)	17	0	1	1	0	0	0
Steroid biosynthesis (00100)	18	0	0	0	0	0	0
Steroid hormone biosynthesis (00140)	22	1	1	2	0	1	0
Synthesis and degradation of ketone bodies (00072)	6	0	2	0	0	0	0
4. Nucleotide metabolism							
Purine metabolism (00230)	275	*30	13	*23	3	1	0
Pyrimidine metabolism (00240)	49	1	*9	1	0	1	0

→ *continued on next page*

Table S6.6. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
5. Amino acid metabolism							
Alanine, aspartate and glutamate metabolism (00250)	41	2	3	1	0	0	0
Arginine and proline metabolism (00330)	39	0	4	0	0	1	0
Arginine biosynthesis (00220)	34	1	1	0	0	1	0
Cysteine and methionine metabolism (00270)	103	11	8	8	0	0	0
Glycine, serine and threonine metabolism (00260)	74	5	9	3	0	1	0
Histidine metabolism (00340)	11	0	1	0	0	0	0
Lysine biosynthesis (00300)	14	0	0	0	0	0	0
Lysine degradation (00310)	66	4	5	0	1	0	0
Phenylalanine metabolism (00360)	35	3	2	0	0	0	0
Phenylalanine, tyrosine and tryptophan biosynthesis (00400)	55	0	*9	2	2	0	0
Tryptophan metabolism (00380)	56	6	4	5	1	0	0
Tyrosine metabolism (00350)	63	3	*11	1	1	1	0
Valine, leucine and isoleucine biosynthesis (00290)	33	5	3	4	0	1	1
Valine, leucine and isoleucine degradation (00280)	59	*11	5	2	1	0	0
6. Metabolism of other amino acids							
beta-Alanine metabolism (00410)	53	*8	6	0	0	0	0
Cyanoamino acid metabolism (00460)	11	0	1	0	0	0	0
D-Alanine metabolism (00473)	1	0	0	0	0	0	0
D-Glutamine and D-glutamate metabolism (00471)	3	0	1	0	0	0	0
Glutathione metabolism (00480)	61	2	4	3	0	0	0
Phosphonate and phosphinate metabolism (00440)	4	0	1	0	0	0	0
Selenocompound metabolism (00450)	24	3	1	3	0	0	0
Taurine and hypotaurine metabolism (00430)	5	0	*2	0	0	0	0
7. Glycan biosynthesis and metabolism							
Glycosaminogl. bios. - chondroitin sulfate / dermatan sulfate (00532)	6	0	0	0	0	0	0
Glycosaminoglycan biosynthesis - heparan sulfate / heparin (00534)	7	0	0	0	0	0	0
Glycosaminoglycan biosynthesis - keratan sulfate (00533)	1	0	1	0	0	0	0
Glycosaminoglycan degradation (00531)	8	0	1	0	0	0	0
Glycosphingolipid biosynthesis - ganglio series (00604)	7	0	1	0	0	0	0
Glycosphingolipid biosynthesis - globo and isoglobo series (00603)	10	1	1	1	0	1	*1
Glycosphingolipid biosynthesis - lacto and neolacto series (00601)	4	1	0	1	0	*1	*1
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis (00563)	14	0	0	0	0	0	0
Lipopolysaccharide biosynthesis (00540)	8	0	0	0	0	0	0
Mannose type O-glycan biosynthesis (00515)	6	0	0	0	0	0	0
N-Glycan biosynthesis (00510)	22	0	1	1	0	0	0
Other glycan degradation (00511)	18	2	2	1	0	0	0
Other types of O-glycan biosynthesis (00514)	2	0	0	0	0	0	0
Peptidoglycan biosynthesis (00550)	11	0	1	0	0	0	0
Various types of N-glycan biosynthesis (00513)	14	0	2	0	0	0	0
8. Metabolism of cofactors and vitamins							
Biotin metabolism (00780)	12	0	2	1	1	0	0
Folate biosynthesis (00790)	25	0	0	1	0	1	0
Lipoic acid metabolism (00785)	6	0	0	0	0	0	0
Nicotinate and nicotinamide metabolism (00760)	42	0	1	2	0	0	0
One carbon pool by folate (00670)	29	0	2	0	0	0	0
Pantothenate and CoA biosynthesis (00770)	38	5	0	4	0	0	0
Porphyrin and chlorophyll metabolism (00860)	56	1	*11	1	0	0	0
Retinol metabolism (00830)	6	1	0	*2	0	1	0
Riboflavin metabolism (00740)	52	3	*9	1	0	0	0
Thiamine metabolism (00730)	61	4	8	1	0	0	0
Ubiquinone and other terpenoid-quinone biosynthesis (00130)	12	*3	1	2	0	1	0
Vitamin B6 metabolism (00750)	10	0	1	0	0	0	0

→ *continued on next page*

Table S6.6. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
9. Metabolism of terpenoids and polyketides							
Biosynthesis of ansamycins (01051)	2	0	0	0	0	0	0
Biosynthesis of siderophore group nonribosomal peptides (01053)	1	0	1	0	0	0	0
Biosynthesis of vancomycin group antibiotics (01055)	5	0	1	0	0	0	0
Carotenoid biosynthesis (00906)	9	0	0	0	0	0	0
Diterpenoid biosynthesis (00904)	4	0	0	0	0	0	0
Geraniol degradation (00281)	13	1	1	0	1	0	0
Insect hormone biosynthesis (00981)	8	0	1	0	0	0	0
Limonene and pinene degradation (00903)	3	0	1	0	0	0	0
Monoterpenoid biosynthesis (00902)	2	0	0	0	0	0	0
Polyketide sugar unit biosynthesis (00523)	7	0	1	0	0	0	0
Sesquiterpenoid and triterpenoid biosynthesis (00909)	12	0	0	0	0	0	0
Terpenoid backbone biosynthesis (00900)	48	0	3	2	0	0	0
Zeatin biosynthesis (00908)	8	2	1	0	1	0	0
10. Biosynthesis of other secondary metabolites							
Acarbose and validamycin biosynthesis (00525)	5	0	1	0	0	0	0
Aflatoxin biosynthesis (00254)	7	0	0	0	0	0	0
Anthocyanin biosynthesis (00942)	1	0	0	0	*1	0	0
Biosynthesis of secondary metabolites - unclassified (00999)	4	0	0	0	0	0	0
Caffeine metabolism (00232)	29	1	3	*8	*2	0	0
Carbapenem biosynthesis (00332)	2	0	0	0	0	0	0
Flavone and flavonol biosynthesis (00944)	1	0	0	0	0	0	0
Flavonoid biosynthesis (00941)	5	*2	0	1	0	0	0
Glucosinolate biosynthesis (00966)	9	*4	0	2	0	0	0
Indole alkaloid biosynthesis (00901)	31	3	1	*5	*2	0	0
Isoflavonoid biosynthesis (00943)	4	0	*2	0	0	0	0
Isoquinoline alkaloid biosynthesis (00950)	49	3	*10	0	0	0	0
Monobactam biosynthesis (00261)	21	3	0	3	0	0	0
Neomycin, kanamycin and gentamicin biosynthesis (00524)	12	1	*3	0	1	0	0
Novobiocin biosynthesis (00401)	4	0	0	0	*1	0	0
Penicillin and cephalosporin biosynthesis (00311)	1	0	0	0	0	0	0
Phenazine biosynthesis (00405)	1	0	1	0	0	0	0
Phenylpropanoid biosynthesis (00940)	202	*31	*26	*25	4	3	*3
Streptomycin biosynthesis (00521)	26	1	*6	0	*2	0	0
Tropane, piperidine and pyridine alkaloid biosynthesis (00960)	18	3	1	0	0	0	0
11. Xenobiotics biodegradation and metabolism							
Aminobenzoate degradation (00627)	18	0	1	0	0	0	0
Atrazine degradation (00791)	1	0	0	0	0	0	0
Benzoate degradation (00362)	12	1	1	0	1	0	0
Caprolactam degradation (00930)	12	1	1	0	1	0	0
Chloroalkane and chloroalkene degradation (00625)	7	0	1	1	0	1	0
Chlorocyclohexane and chlorobenzene degradation (00361)	1	0	0	0	0	0	0
Drug metabolism - cytochrome P450 (00982)	60	4	6	*6	0	2	1
Drug metabolism - other enzymes (00983)	125	5	*18	10	*5	0	0
Metabolism of xenobiotics by cytochrome P450 (00980)	17	1	0	2	0	1	0
Naphthalene degradation (00626)	4	0	0	1	0	*1	0
Nitrotoluene degradation (00633)	28	0	3	*8	*2	0	0
Steroid degradation (00984)	23	1	3	2	0	1	0
Styrene degradation (00643)	23	0	2	0	1	0	0
Toluene degradation (00623)	12	1	1	0	1	0	0

Table S6.7. Pfam domain enrichment analysis. *Solanum chacoense* protein sequences were compared to the Pfam domain database with HMMER. Transcripts predicted to be associated to each domain were counted in the whole dataset, as well as in ovule-enriched (O↑), 2 DBA up-regulated (D↑), 2 DBA down-regulated (D↓), *frk1* up-regulated (F↑), and *frk1* down-regulated (F↓) transcripts, as well as ovule-enriched transcripts that were down-regulated in both 2 DBA and *frk1* conditions (O↑D↓F↓). Asterisks (*) indicate a significant enrichment (Fisher's exact test, $P \leq 0.05$). Only Pfam domains enriched in at least on conditions are presented in this table.

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
2Fe-2S iron-sulfur cluster binding domain (PF00111)	22	1	*4	*4	0	0	0
2OG-Fe(II) oxygenase superfamily (PF03171)	225	*31	*26	*31	3	*6	2
3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III C terminal (PF08541)	27	1	*8	0	*2	0	0
3' exoribonuclease family, domain 2 (PF03725)	8	2	0	*2	0	0	0
4Fe-4S single cluster domain (PF13394)	3	0	0	0	0	*1	0
4Fe-4S single cluster domain of Ferredoxin I (PF13370)	3	1	0	0	*1	0	0
AAA domain (PF13245)	1	1	0	*1	0	*1	*1
ABA/WDS induced protein (PF02496)	8	*3	0	*4	1	0	0
ABC transporter (PF00005)	156	*22	*19	*19	4	0	0
ABC transporter transmembrane region (PF00664)	54	2	*10	3	1	0	0
ABC-2 family transporter protein (PF12698)	8	2	0	*2	0	0	0
ABC-2 type transporter (PF01061)	69	*17	8	*12	*3	0	0
ABC-transporter N-terminal (PF14510)	24	*7	2	*6	*3	0	0
Acetyltransferase (GNAT) family (PF00583)	50	0	2	*8	1	1	0
Adenosine/AMP deaminase (PF00962)	3	0	0	0	0	*1	0
AIG1 family (PF04548)	16	3	1	*3	0	0	0
Alcohol dehydrogenase GroES-like domain (PF08240)	75	4	*11	2	*3	2	0
Aldose 1-epimerase (PF01263)	19	2	*5	1	*2	0	0
Alginate lyase (PF08787)	2	*2	0	*2	0	*1	*1
Alpha 1,4-glycosyltransferase conserved region (PF04572)	5	*3	0	*2	0	0	0
Alpha-L-arabinofuranosidase C-terminal domain (PF06964)	3	*2	0	1	0	0	0
Alpha-L-fucosidase (PF01120)	3	*2	0	0	0	0	0
alpha/beta hydrolase fold (PF00561)	50	1	*10	1	2	1	0
alpha/beta hydrolase fold (PF07859)	54	*8	2	*17	2	0	0
Aluminium activated malate transporter (PF11744)	24	*7	1	3	0	0	0
Amino acid permease (PF13520)	21	3	1	*4	0	0	0
Amino-transferase class IV (PF01063)	12	*5	0	2	0	0	0
Ammonium Transporter Family (PF00909)	9	1	2	1	*2	1	*1
AMP-binding enzyme (PF00501)	73	3	*10	2	2	1	0
Annexin (PF00191)	18	3	3	*3	0	*2	*2
Anticodon binding domain (PF03129)	13	0	0	0	0	*2	0
AP2 domain (PF00847)	223	12	18	*20	1	1	0
Arginase family (PF00491)	3	0	0	0	0	*1	0
Argonaute linker 1 domain (PF08699)	17	*5	2	2	0	0	0
Argonaute linker 2 domain (PF16488)	15	*5	2	0	0	0	0
Aromatic amino acid lyase (PF00221)	26	4	2	*6	1	*4	1
Aspartic acid proteinase inhibitor (PF16845)	20	*4	2	1	*2	1	0
Associated with HOX (PF07526)	15	0	3	*3	0	0	0
ATP synthase B/B' CF(0) (PF00430)	2	0	1	0	*1	0	0
ATP-sulfurylase (PF01747)	6	2	0	*2	0	0	0
Auxin response factor (PF06507)	26	*5	4	3	0	0	0
AWPM-19-like family (PF05512)	4	0	*2	0	0	0	0
B3 DNA binding domain (PF02362)	125	*29	9	7	1	1	0
B-box zinc finger (PF00643)	29	4	3	1	*2	1	0
Bacterial extracellular solute-binding proteins, family 3 (PF00497)	22	0	*4	0	0	0	0
Barwin family (PF00967)	4	1	0	*2	0	*1	0
BED zinc finger (PF02892)	18	0	0	0	*2	0	0
Berberine and berberine like (PF08031)	27	*7	*5	2	1	*3	*2
Beta-sandwich domain in beta galactosidase (PF17834)	17	1	*5	1	0	*2	1
bHLH-MYC and R2R3-MYB transcription factors N-terminal (PF14215)	22	4	2	3	0	*2	0
Blue/Ultraviolet sensing protein C terminal (PF12546)	3	0	1	0	*1	0	0
BRCA1 C Terminus (BRCT) domain (PF00533)	14	1	*4	1	0	0	0
BURP domain (PF03181)	30	*12	*12	*4	1	0	0

→ *continued on next page*

Table S6.7. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
C1 domain (PF03107)	42	*7	0	3	1	0	0
C2H2-type zinc finger (PF13912)	63	*9	4	6	0	0	0
C-terminus of AA_permease (PF13906)	10	*3	1	*4	0	0	0
C-terminus of histone H2A (PF16211)	23	0	*11	0	0	0	0
Calcium-binding EGF domain (PF07645)	27	0	*8	0	0	0	0
Calcium-dependent channel, 7TM region, putative phosphate (PF02714)	14	2	1	1	0	*2	*1
Calmodulin binding protein-like (PF07887)	22	0	*6	0	*2	0	0
Calponin homology (CH) domain (PF00307)	15	1	*6	1	0	0	0
Carbohydrate binding domain (PF02018)	9	*5	*3	*3	1	0	0
Carbohydrate binding domain CBM49 (PF09478)	3	1	*2	0	*1	0	0
Carbohydrate esterase, sialic acid-specific acetylesterase (PF03629)	17	*4	*5	*4	1	*4	*4
Carboxypeptidase A inhibitor (PF02977)	9	*4	2	*4	0	0	0
Carlavirus endopeptidase (PF05379)	2	0	0	1	*1	0	0
Catalase (PF00199)	8	*5	0	*4	0	0	0
Catalase-related immune-responsive (PF06628)	7	*3	0	*3	0	0	0
CCAAT-binding transcription factor (CBF-B/NF-YA) subunit B (PF02045)	11	*3	0	1	0	0	0
CCT motif (PF06203)	43	*9	5	1	2	1	0
Cell cycle regulated microtubule associated protein (PF12214)	6	0	*5	0	0	0	0
Cellulase (glycosyl hydrolase family 5) (PF00150)	20	3	*7	1	*2	1	1
Cellulose synthase (PF03552)	37	2	5	*6	0	2	1
Centromere kinetochore component CENP-T histone fold (PF15511)	17	0	*11	0	0	0	0
Chitin recognition protein (PF00187)	26	3	3	*4	1	*2	0
Chitinase class I (PF00182)	23	4	1	*5	0	1	0
Chlorophyll A-B binding protein (PF00504)	54	0	*35	0	*19	0	0
Chlorophyllase (PF07224)	2	1	*2	0	0	0	0
Chlorophyllase enzyme (PF12740)	2	0	0	1	0	*1	0
CNH domain (PF00780)	3	0	0	0	*1	0	0
COBRA-like protein (PF04833)	18	2	*4	0	0	0	0
Coiled-coil region of Oberon (PF16312)	7	1	0	1	0	1	*1
Common central domain of tyrosinase (PF00264)	24	0	*7	0	0	0	0
Conserved in the green lineage and diatoms 27 (PF06799)	2	0	0	0	*1	0	0
Core histone H2A/H2B/H3/H4 (PF00125)	68	1	*35	0	0	0	0
Cotton fibre expressed protein (PF05553)	44	1	*7	*5	0	0	0
CRAL/TRIO domain (PF00650)	39	4	*6	0	1	1	0
CRAL/TRIO, N-terminal domain (PF03765)	21	1	*4	0	1	1	0
Cullin protein neddylation domain (PF10557)	14	1	0	1	0	1	*1
Cyanobacterial and plastid NDH-1 subunit M (PF10664)	1	0	0	*1	0	0	0
Cyclic nucleotide-binding domain (PF00027)	34	3	3	*5	2	0	0
Cyclin, C-terminal domain (PF02984)	41	1	*26	0	0	0	0
Cyclin, N-terminal domain (PF00134)	58	2	*26	1	0	0	0
Cystatin domain (PF00031)	23	*6	2	0	0	0	0
Cysteine-rich antifungal protein 2, defensin-like (PF10868)	5	*2	1	0	0	0	0
Cysteine-rich TM module stress tolerance (PF12734)	4	0	0	1	*1	0	0
Cytochrome P450 (PF00067)	679	*73	*60	*74	13	12	4
Cytosol aminopeptidase family, catalytic domain (PF00883)	6	0	*4	0	0	1	0
Cytosol aminopeptidase family, N-terminal domain (PF02789)	4	0	*2	0	0	*1	0
Cytosolic domain of 10TM putative phosphate transporter (PF14703)	14	2	1	1	0	*2	*1
D-mannose binding lectin (PF01453)	150	5	*26	*13	*8	1	0
Dehydrogenase E1 component (PF00676)	20	1	1	0	0	*2	0
Dimerisation domain (PF08100)	38	4	*6	*6	0	1	1
Dirigent-like protein (PF03018)	39	*6	*6	3	0	0	0
Divergent CCT motif (PF09425)	23	1	4	2	0	*3	0
DnaJ C terminal domain (PF01556)	30	3	0	*4	0	1	0
DnaJ central domain (PF00684)	17	3	0	*4	0	0	0
Dof domain, zinc finger (PF02701)	37	2	*10	1	0	0	0
Domain associated at C-terminal with AAA (PF14363)	27	*5	2	3	0	1	1
Domain of unknown function (DUF588) (PF04535)	52	*11	5	*9	1	*3	1
Domain of unknown function (DUF702) (PF05142)	8	*3	1	0	0	0	0
Domain of unknown function (DUF966) (PF06136)	7	2	*3	0	0	0	0
Domain of unknown function (DUF1338) (PF07063)	1	1	0	0	0	*1	0

→ *continued on next page*

Table S6.7. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
Domain of unknown function (DUF3403) (PF11883)	35	1	5	*5	*3	1	0
Domain of unknown function (DUF4033) (PF13225)	2	0	1	0	*1	0	0
Domain of unknown function (DUF4094) (PF13334)	13	1	1	*3	0	0	0
Domain of unknown function (DUF4281) (PF14108)	4	*3	1	0	0	0	0
Domain of unknown function (DUF4371) (PF14291)	7	*3	0	*3	0	0	0
Domain of unknown function (DUF4378) (PF14309)	22	2	*4	1	1	0	0
Domain of unknown function (DUF4408) (PF14364)	7	0	*3	0	0	0	0
Dormancy/auxin associated protein (PF05564)	5	*4	0	1	0	0	0
DREPP plasma membrane polypeptide (PF05558)	1	0	1	0	*1	0	0
Drought induced 19 protein (Di19), zinc-binding (PF05605)	8	1	0	*2	0	0	0
DUF761-associated sequence motif (PF14383)	14	0	*4	0	1	0	0
dUTPase (PF00692)	3	0	*2	0	0	0	0
DVL family (PF08137)	7	*3	*3	1	0	0	0
EamA-like transporter family (PF00892)	53	*8	*7	2	0	*3	1
EGF domain (PF12947)	3	0	*2	0	0	0	0
eIF-6 family (PF01912)	4	*2	0	0	0	0	0
Enoyl-(Acyl carrier protein) reductase (PF13561)	68	3	*12	4	*3	1	0
Enoyl-CoA hydratase/isomerase (PF16113)	24	*5	1	0	0	0	0
Epidermal patterning factor proteins (PF17181)	14	1	*5	1	0	0	0
ESCO1/2 acetyl-transferase (PF13880)	2	0	*2	0	0	0	0
Eukaryotic aspartyl protease (PF00026)	17	0	1	1	0	*2	0
Eukaryotic glutathione synthase (PF03199)	1	0	0	*1	0	0	0
Eukaryotic glutathione synthase, ATP binding domain (PF03917)	1	0	0	*1	0	0	0
Eukaryotic-type carbonic anhydrase (PF00194)	7	*3	1	*3	1	*2	*2
Exostosin family (PF03016)	47	*8	6	2	0	1	1
EXS family (PF03124)	12	1	2	1	*2	0	0
F-box associated (PF07734)	91	8	1	4	0	2	*2
FAD binding domain (PF01565)	58	*9	*8	4	*3	*7	*2
FAD-binding domain (PF08022)	20	*5	2	*7	0	0	0
FAE1/Type III polyketide synthase-like protein (PF08392)	28	1	*9	0	*3	0	0
Fasciclin domain (PF02469)	27	0	*10	0	0	0	0
Fatty acid desaturase (PF00487)	33	0	*7	2	1	1	0
Fatty acid hydroxylase superfamily (PF04116)	29	1	*6	2	0	*2	0
Ferric reductase like transmembrane component (PF01794)	16	*5	2	*7	0	0	0
Ferric reductase NAD binding domain (PF08030)	20	*4	2	*7	0	0	0
Ferritin-like domain (PF13668)	6	*3	0	*5	0	*2	0
Fibronectin type III-like domain (PF14310)	11	*4	*3	1	0	1	*1
Fibronectin type-III domain (PF17766)	110	13	*23	5	3	2	0
Filament-like plant protein, long coiled-coil (PF05911)	23	0	*7	0	0	0	0
FIST N domain (PF08495)	2	0	0	0	0	*1	0
Flavin containing amine oxidoreductase (PF01593)	39	2	2	*7	0	1	0
Flavin-binding monooxygenase-like (PF00743)	35	3	*6	4	0	1	1
FMN-dependent dehydrogenase (PF01070)	12	*3	0	*4	0	*2	0
Formin Homology 2 Domain (PF02181)	25	1	4	1	*2	0	0
GAF domain (PF01590)	12	*3	0	2	0	0	0
gag-polypeptide of LTR copia-type (PF14223)	72	*24	2	*22	2	*5	*4
GAG-pre-integrase domain (PF13976)	34	*6	0	3	0	1	1
Galactose binding lectin domain (PF02140)	12	1	*4	1	0	*2	*1
Galactosyltransferase (PF01762)	22	2	1	*4	0	0	0
Gamma interferon inducible lysosomal thiol reductase (GILT) (PF03227)	3	0	*2	0	0	0	0
Gamma-thionin family (PF00304)	19	*9	*6	*3	*2	0	0
GATA zinc finger (PF00320)	35	3	*7	1	1	1	0
GDA1/CD39 (nucleoside phosphatase) family (PF01150)	15	0	*5	0	0	*2	0
GDP-mannose 4,6 dehydratase (PF16363)	30	2	3	*5	0	0	0
GDSL-like Lipase/Acylhydrolase (PF00657)	106	11	*25	7	3	*4	*2
GDSL-like Lipase/Acylhydrolase family (PF13472)	4	1	0	1	*1	0	0
GDSL/SGNH-like Acyl-Esterase family found in Pmr5 and Cas1p (PF13839)	77	*13	*13	*10	0	*6	*6
GH3 auxin-responsive promoter (PF03321)	36	3	*7	*5	0	0	0
Gibberellin regulated protein (PF02704)	27	*8	*13	3	*3	*2	1
Globin (PF00042)	2	0	*2	0	0	0	0

→ *continued on next page*

Table S6.7. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
Glu/Leu/Phe/Val dehydrogenase, dimerisation domain (PF02812)	5	*2	0	1	0	0	0
Glucose / Sorbosone dehydrogenase (PF07995)	5	1	0	*2	0	1	0
Glutamate/Leucine/Phenylalanine/Valine dehydrogenase (PF00208)	5	*2	0	1	0	0	0
Glutathione S-transferase, C-terminal domain (PF13410)	26	2	*5	*5	0	0	0
Glutathione S-transferase, N-terminal domain (PF02798)	72	4	*9	*8	2	1	0
Glutathione S-transferase, N-terminal domain (PF13409)	4	*2	0	*2	0	0	0
Glyceraldehyde 3-phosphate dehydrogenase, C-terminal domain (PF02800)	11	1	*3	1	*2	0	0
Glycosyl hydrolase family 1 (PF00232)	29	0	*8	0	*2	0	0
Glycosyl hydrolase family 3 C-terminal domain (PF01915)	27	*9	*7	1	0	1	1
Glycosyl hydrolase family 3 N terminal domain (PF00933)	25	*6	*5	0	0	0	0
Glycosyl hydrolase family 9 (PF00759)	30	4	*10	3	*5	1	0
Glycosyl hydrolase family 10 (PF00331)	14	*7	*4	*5	1	*2	*2
Glycosyl hydrolase family 14 (PF01373)	10	0	*3	1	1	0	0
Glycosyl hydrolases family 16 (PF00722)	56	*14	7	*9	2	*3	*3
Glycosyl hydrolases family 17 (PF00332)	93	8	*13	*9	0	0	0
Glycosyl hydrolases family 18 (PF00704)	18	1	2	*4	1	0	0
Glycosyl hydrolases family 28 (PF00295)	70	*11	*11	*7	*3	*3	*2
Glycosyl hydrolases family 32 C terminal (PF08244)	16	3	*4	1	1	0	0
Glycosyl hydrolases family 35 (PF01301)	19	1	*5	1	0	*2	1
Glycosyl transferase family group 2 (PF13632)	5	0	*2	0	0	0	0
Glycosyltransferase like family 2 (PF13641)	7	1	*4	0	0	0	0
Glycosyltransferase sugar-binding region containing DXD motif (PF04488)	4	*2	0	*2	0	0	0
Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily (PF00903)	21	2	0	2	0	*4	0
GMC oxidoreductase (PF00732)	11	1	*5	*3	1	0	0
GMC oxidoreductase (PF05199)	11	0	*5	2	1	0	0
Golgi complex component 7 (COG7) (PF10191)	1	0	0	0	0	*1	0
GRAM domain (PF02893)	21	0	*11	0	0	0	0
GUCT (NUC152) domain (PF08152)	4	0	0	*2	0	0	0
HAD superfamily, subfamily IIIB (Acid phosphatase) (PF03767)	12	0	*5	0	0	0	0
HCO3- transporter family (PF00955)	8	*4	0	*5	0	0	0
HD domain (PF01966)	3	1	0	1	0	*1	*1
HD domain (PF13328)	8	0	0	*2	0	0	0
HD-ZIP protein N terminus (PF04618)	4	0	*2	0	0	0	0
Heavy-metal-associated domain (PF00403)	71	3	*15	6	0	1	0
Helix-hairpin-helix motif (PF12836)	2	0	*2	0	0	0	0
Helix-loop-helix DNA-binding domain (PF00010)	158	*21	*22	*13	4	3	0
Hemerythrin HHE cation binding domain (PF01814)	5	1	*2	0	0	0	0
High-affinity nitrate transporter accessory (PF16974)	2	0	0	1	*1	0	0
His Kinase A (phospho-acceptor) domain (PF00512)	16	*4	0	*3	0	0	0
Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase (PF13589)	12	2	0	1	0	1	*1
Homeobox associated leucine zipper (PF02183)	29	*9	3	*6	1	1	0
Homeobox KN domain (PF05920)	25	2	3	*4	0	0	0
Homeodomain (PF00046)	82	*18	*14	*10	1	1	0
HORMA domain (PF02301)	4	0	1	1	0	*1	0
Hydrophobic seed protein (PF14547)	36	*7	*9	4	2	1	0
Hydroxymethylglutaryl-coenzyme A synthase C terminal (PF08540)	5	0	*2	0	0	0	0
Integrase core domain (PF00665)	48	3	1	3	0	*3	0
Integrase zinc binding domain (PF17921)	21	0	2	*4	0	*2	0
Ion transport protein (PF00520)	36	3	3	*5	2	1	0
IQ calmodulin-binding motif (PF00612)	60	2	*13	1	1	0	0
Jacalin-like lectin domain (PF01419)	17	0	*4	0	1	0	0
K-box region (PF01486)	58	*25	4	3	1	0	0
K+ potassium transporter (PF02705)	21	*5	0	*4	0	1	1
KHA, dimerisation domain of potassium ion channel (PF11834)	11	2	*3	1	1	0	0
Kinesin motor domain (PF00225)	66	3	*32	1	0	0	0
Kinesin-associated protein (KAP) (PF05804)	1	0	0	0	*1	0	0
KIP1-like protein (PF07765)	21	1	*5	1	0	0	0
Late embryogenesis abundant protein (PF03168)	46	2	5	*6	0	0	0
Late embryogenesis abundant protein (PF03242)	10	2	0	*4	0	0	0
Late exocytosis, associated with Golgi transport (PF13967)	15	3	1	1	0	*2	*1

→ *continued on next page*

Table S6.7. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
Lateral organ boundaries (LOB) domain (PF03195)	54	*8	2	1	0	0	0
Lecithin:cholesterol acyltransferase (PF02450)	12	*3	0	0	0	0	0
Leucine Rich Repeat (PF00560)	331	12	*33	10	*10	3	1
Leucine rich repeat (PF13855)	670	26	*80	19	*19	4	2
Leucine rich repeat N-terminal domain (PF08263)	387	17	*65	14	*13	1	1
Leucine Rich repeats (2 copies) (PF12799)	32	2	2	2	0	*2	*2
Ligand-gated ion channel (PF00060)	21	0	*4	0	0	0	0
linker histone H1 and H5 family (PF00538)	23	3	*7	2	0	1	1
Lipase (class 3) (PF01764)	79	3	*16	2	1	0	0
Lipoxygenase (PF00305)	34	*7	5	*6	*3	*2	1
LysM domain (PF01476)	21	0	*4	1	0	0	0
Lytic transglycolase (PF03330)	44	*7	*14	2	*4	1	1
Mad3/BUB1 homology region 1 (PF08311)	3	0	*3	0	0	0	0
Major intrinsic protein (PF00230)	63	6	*10	4	*4	1	0
Male sterility protein (PF03015)	17	2	*4	2	1	0	0
Malectin domain (PF11721)	26	3	*8	2	0	*2	*2
Malectin-like domain (PF12819)	48	1	*7	2	1	0	0
MatE (PF01554)	95	*14	*12	7	1	3	0
MazG-like family (PF12643)	4	*2	1	0	0	0	0
MCM AAA-lid domain (PF17855)	8	0	*5	0	0	0	0
MCM N-terminal domain (PF14551)	7	0	*5	0	0	0	0
MCM OB domain (PF17207)	9	0	*6	0	0	0	0
MCM P-loop domain (PF00493)	8	0	*5	0	0	0	0
Membrane bound O-acyl transferase family (PF13813)	7	0	*3	0	0	0	0
Membrane transport protein (PF03547)	21	1	*4	0	0	0	0
Met-10+ like-protein (PF02475)	3	0	0	0	0	*1	0
Metallothionein (PF01439)	8	1	2	*2	0	0	0
Methyltransferase domain (PF13578)	1	1	0	*1	0	0	0
Microtubule associated protein (MAP65/ASE1 family) (PF03999)	12	0	*6	1	0	0	0
Mitochondrial pyruvate carriers (PF03650)	11	*6	0	*4	0	0	0
MoaE protein (PF02391)	1	0	0	*1	0	0	0
Morc6 ribosomal protein S5 domain 2-like (PF17942)	7	2	0	1	0	1	*1
Mpv17 / PMP22 family (PF04117)	17	*4	0	2	0	0	0
Multicopper oxidase (PF00394)	68	*13	*11	2	0	*3	0
Multicopper oxidase (PF07731)	55	3	*11	2	0	*3	0
Multicopper oxidase (PF07732)	63	*13	*11	2	0	*3	0
Myb-like DNA-binding domain (PF00249)	266	*28	22	*27	5	5	1
Myb-like DNA-binding domain (PF13921)	30	1	2	*5	0	1	0
Mycolic acid cyclopropane synthetase (PF02353)	22	1	0	*7	0	*6	1
Myo-inositol-1-phosphate synthase (PF01658)	2	0	*2	0	*1	0	0
Myo-inositol-1-phosphate synthase (PF07994)	2	0	*2	0	*1	0	0
N-terminal domain of argonaute (PF16486)	17	*5	2	2	0	0	0
N-terminal domain of oxidoreductase (PF16884)	14	2	*4	0	1	1	0
NAD(P)-binding Rossmann-like domain (PF13450)	18	0	0	*5	0	*3	0
NADH:flavin oxidoreductase / NADH oxidase family (PF00724)	20	0	*6	0	*4	0	0
NADPH-dependent FMN reductase (PF03358)	9	*4	0	*3	0	0	0
NAF domain (PF03822)	32	5	2	*4	1	1	1
Neprosin (PF03080)	40	*12	5	2	0	1	1
Niemann-Pick C1 N terminus (PF16414)	1	0	0	*1	0	0	0
Nitronate monooxygenase (PF03060)	1	0	0	*1	0	0	0
NmrA-like family (PF05368)	26	3	*12	3	0	0	0
No apical meristem (NAM) protein (PF02365)	139	9	11	*17	2	1	1
non-haem dioxygenase in morphine synthesis N-terminal (PF14226)	210	*29	19	*31	3	4	1
non-SMC mitotic condensation complex subunit 1 (PF12717)	3	0	*2	0	0	0	0
NPH3 family (PF03000)	35	2	*10	2	1	0	0
NPR1/NIM1 like defence protein C terminal (PF12313)	4	0	0	*2	0	0	0
Nse4 C-terminal (PF08743)	4	1	*2	1	0	0	0
Nucleotide-diphospho-sugar transferase (PF03407)	15	3	1	*3	0	0	0
NUDIX domain (PF00293)	38	1	*6	1	0	1	0
Nudix hydrolase domain (PF18290)	7	0	*3	0	0	0	0

→ *continued on next page*

Table S6.7. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
O-methyltransferase (PF01596)	14	1	*4	0	0	0	0
O-methyltransferase domain (PF00891)	61	4	8	*7	0	1	1
OPT oligopeptide transporter protein (PF03169)	29	*7	2	*4	0	0	0
OST3 / OST6 family, transporter family (PF04756)	5	1	0	1	0	*2	0
OTU-like cysteine protease (PF02338)	17	2	0	*3	0	0	0
Oxygen evolving enhancer protein 3 (PsbQ) (PF05757)	5	0	*2	0	0	0	0
P21-Rho-binding domain (PF00786)	15	0	3	*3	1	0	0
PA domain (PF02225)	76	9	*13	3	3	1	0
Palmitoyl protein thioesterase (PF02089)	4	1	*2	0	0	*1	0
PAN domain (PF00024)	2	0	*2	0	0	0	0
PAN-like domain (PF08276)	96	3	*13	*9	*4	1	0
Patatin-like phospholipase (PF01734)	26	2	*5	1	0	0	0
Patched family (PF02460)	1	0	0	*1	0	0	0
Pathogenesis-related protein Bet v I family (PF00407)	111	9	7	9	1	*4	2
PAZ domain (PF02170)	22	*5	3	2	0	0	0
PDDEXK-like family of unknown function (PF04720)	18	2	*4	2	0	0	0
Pectate lyase (PF00544)	27	1	*8	1	*2	0	0
Pectate lyase superfamily protein (PF12708)	1	1	0	*1	0	0	0
Pectinacetylesterase (PF03283)	44	*22	5	*13	*3	1	1
Pectinesterase (PF01095)	81	*11	*15	4	3	*4	1
Peptidase C13 family (PF01650)	19	2	0	*3	0	0	0
Peptidase inhibitor I9 (PF05922)	98	*13	*22	5	3	3	1
Permease family (PF00860)	17	2	*6	1	1	0	0
Peroxidase (PF00141)	164	*24	*23	*18	4	3	*3
PetM family of cytochrome b6f complex subunit 7 (PF08041)	3	0	*2	0	0	0	0
PGAP1-like protein (PF07819)	2	0	0	0	*1	0	0
PHD-like zinc-binding domain (PF13771)	5	0	*3	0	0	0	0
Phloem protein 2 (PF14299)	54	5	6	*6	0	2	*2
Phosphate-induced protein 1 conserved region (PF04674)	21	*5	2	3	0	0	0
Phosphoenolpyruvate carboxylase (PF00311)	8	*5	0	0	0	0	0
Phosphorylase superfamily (PF01048)	7	2	*3	0	0	0	0
Photosystem I psaG / psaK (PF01241)	4	0	*3	0	0	0	0
Photosystem I reaction centre subunit IV / PsaE (PF02427)	3	0	*3	0	0	0	0
Photosystem I reaction centre subunit VI (PF03244)	4	0	*3	0	0	0	0
Photosystem I reaction centre subunit XI (PF02605)	2	0	*2	0	0	0	0
Photosystem II reaction centre X protein (PsbX) (PF06596)	5	0	*4	0	0	0	0
Phytanoyl-CoA dioxygenase (PhyH) (PF05721)	1	0	0	*1	0	0	0
Phytosulfokine precursor protein (PSK) (PF06404)	7	*3	2	*3	1	1	*1
Piwi domain (PF02171)	23	*7	*5	3	0	1	0
Plant invertase/pectin methylesterase inhibitor (PF04043)	125	*31	*29	*16	*4	*4	2
Plant PDR ABC transporter associated (PF08370)	25	*7	2	*6	*3	0	0
Plant phosphoribosyltransferase C-terminal (PF08372)	19	3	*10	0	0	0	0
Plant pleckstrin homology-like region (PF08458)	7	*3	1	1	0	0	0
Plant protein of unknown function (DUF639) (PF04842)	5	1	0	*2	0	1	*1
Plant protein of unknown function (DUF936) (PF06075)	9	0	*3	0	0	0	0
Plant protein of unknown function (PF03140)	112	0	*15	3	0	2	0
Plant self-incompatibility protein S1 (PF05938)	63	*34	*10	*12	1	*9	*8
Plants and Prokaryotes Conserved (PCC) domain (PF03479)	40	*8	3	4	0	0	0
Plastocyanin-like domain (PF02298)	59	*13	*11	*9	2	1	1
PLAT/LH2 domain (PF01477)	29	*7	*5	*5	*2	*2	1
PLATZ transcription factor (PF04640)	17	3	1	*6	0	0	0
PMR5 N terminal Domain (PF14416)	59	*10	*11	*8	0	*4	*4
Pollen allergen (PF01357)	39	5	*13	2	*3	1	1
Pollen proteins Ole e I like (PF01190)	44	*8	6	*6	*4	1	0
Polyketide cyclase / dehydrase and lipid transport (PF10604)	21	1	*6	2	0	0	0
Polyphenol oxidase middle domain (PF12142)	19	0	*7	0	0	0	0
Polysaccharide biosynthesis (PF04669)	12	1	*5	1	0	0	0
Polysaccharide lyase family 4, domain II (PF14686)	12	1	*4	0	0	0	0
Polysaccharide lyase family 4, domain III (PF14683)	15	1	*7	0	0	0	0
POT family (PF00854)	128	10	*17	*15	1	*14	0

→ *continued on next page*

Table S6.7. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
Potato inhibitor I family (PF00280)	44	1	*11	0	0	0	0
Potato type II proteinase inhibitor family (PF02428)	23	1	*9	1	0	*4	0
PQ loop repeat (PF04193)	6	0	*3	0	0	0	0
Prephenate dehydratase (PF00800)	4	0	*2	0	*1	0	0
Prephenate dehydrogenase (PF02153)	3	0	0	0	*1	0	0
Probable lipid transfer (PF14368)	59	7	*10	*6	1	1	0
Prolamin-like (PF05617)	23	*11	1	*11	0	*10	*10
PRONE (Plant-specific Rop nucleotide exchanger) (PF03759)	14	1	*6	1	0	0	0
Protease inhibitor/seed storage/LTP family (PF00234)	40	*10	*6	*6	1	0	0
Protein kinase C terminal domain (PF00433)	8	0	0	*2	0	0	0
Protein kinase domain (PF00069)	1033	67	*104	56	13	9	4
Protein of unknown function (DUF563) (PF04577)	10	*3	0	2	0	0	0
Protein of unknown function (DUF568) (PF04526)	14	1	*5	1	0	0	0
Protein of unknown function (DUF640) (PF04852)	13	3	0	*3	0	1	0
Protein of unknown function (DUF642) (PF04862)	14	1	*4	0	1	0	0
Protein of unknown function (DUF740) (PF05340)	4	0	*2	0	0	0	0
Protein of unknown function (DUF789) (PF05623)	17	*9	0	*4	0	0	0
Protein of unknown function (DUF1298) (PF06974)	11	0	*3	2	0	0	0
Protein of unknown function (DUF1635) (PF07795)	5	0	*2	0	0	0	0
Protein of unknown function (DUF1666) (PF07891)	8	0	*3	0	0	0	0
Protein of unknown function (DUF3007) (PF11460)	3	0	0	0	0	*2	0
Protein of unknown function (DUF3675) (PF12428)	15	*4	0	2	0	0	0
Protein of unknown function (DUF4005) (PF13178)	24	1	*9	0	1	0	0
Protein of unknown function (DUF4101) (PF13355)	5	*2	0	1	0	0	0
Protein of unknown function (DUF_B2219) (PF12143)	23	0	*8	0	0	0	0
Protein tyrosine kinase (PF07714)	545	23	*66	23	9	7	4
Pterin 4 alpha carbinolamine dehydratase (PF01329)	2	0	0	0	0	*1	0
PUA-like domain (PF14306)	5	*3	0	*3	0	0	0
Pumilio-family RNA binding repeat (PF00806)	12	*3	1	0	0	0	0
Purine nucleobase transmembrane transport (PF16913)	32	5	1	*7	0	1	1
Putative AtpZ or ATP-synthase-associated (PF16594)	1	1	0	*1	0	0	0
Putative Phosphatase (PF06888)	15	*7	0	*7	0	0	0
Putative small multi-drug export protein (PF06695)	1	0	0	0	0	*1	0
Pyridoxal-dependent decarboxylase conserved domain (PF00282)	41	0	*6	0	1	2	0
Pyridoxal-phosphate dependent enzyme (PF00291)	29	4	*5	3	0	0	0
Pyroglutamyl peptidase (PF01470)	4	0	0	0	0	*1	0
QLQ (PF08880)	12	2	*6	0	0	0	0
Rapid ALkalinization Factor (RALF) (PF05498)	14	*5	2	*6	0	1	*1
Receptor family ligand binding region (PF01094)	22	0	*4	0	0	0	0
Region found in RelA / SpoT proteins (PF04607)	7	0	0	*2	0	0	0
Regulator of Vps4 activity in the MVB pathway (PF03398)	14	1	2	*3	*2	0	0
Remorin, C-terminal region (PF03763)	20	0	*4	*5	1	0	0
Respiratory burst NADPH oxidase (PF08414)	9	*3	2	*4	0	0	0
Response regulator receiver domain (PF00072)	51	*8	3	3	0	0	0
Reticulon (PF02453)	28	1	*5	1	0	0	0
Retinal pigment epithelial membrane protein (PF03055)	23	0	0	*6	0	*2	0
Retroviral aspartyl protease (PF08284)	15	1	1	2	0	*2	0
Rhamnogalacturonate lyase family (PF06045)	11	1	*4	0	0	0	0
RHO protein GDP dissociation inhibitor (PF02115)	5	0	*2	0	0	0	0
Rhodanese-like domain (PF00581)	16	1	3	1	*2	0	0
Ribonuclease T2 family (PF00445)	43	*24	3	*18	0	*6	*5
Ribonucleotide reductase, barrel domain (PF02867)	4	0	*2	0	0	0	0
Ribosomal protein S8 (PF00410)	19	*4	0	0	0	0	0
Ribosome associated membrane protein RAMP4 (PF06624)	3	*2	0	0	0	0	0
Ribulose bisphosphate carboxylase, small chain (PF00101)	8	0	*3	0	0	0	0
Ribulose-1,5-bisphosphate carboxylase small subunit (PF12338)	7	0	*3	0	0	0	0
RNA dependent RNA polymerase (PF00978)	4	0	0	*3	*1	0	0
Rrp15p (PF07890)	1	0	0	0	0	*1	0
Rx N-terminal domain (PF18052)	198	8	15	7	*6	1	0
S locus-related glycoprotein 1 binding pollen coat protein (SLR1-BP) (PF07333)	4	*2	*3	0	0	0	0

→ *continued on next page*

Table S6.7. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
S-locus glycoprotein domain (PF00954)	106	3	*20	*9	*4	1	0
Salt stress response/antifungal (PF01657)	71	4	*10	4	*3	1	0
SAM dependent carboxyl methyltransferase (PF03492)	40	3	1	*6	1	1	0
Saposin-like type B, region 1 (PF05184)	12	0	0	1	0	*2	0
Saposin-like type B, region 2 (PF03489)	12	0	0	1	0	*2	0
Sar8.2 family (PF03058)	3	*2	0	*2	0	0	0
Sec61beta family (PF03911)	5	*2	0	0	0	0	0
Senescence regulator (PF04520)	20	*5	0	*4	0	1	1
Serine carboxypeptidase (PF00450)	86	*11	*11	*10	2	*5	0
Serine carboxypeptidase S28 (PF05577)	7	*4	1	0	1	0	0
Serine hydrolase (FSH1) (PF03959)	27	0	*5	0	0	1	0
Shikimate 5'-dehydrogenase C-terminal domain (PF18317)	4	0	*2	0	0	0	0
short chain dehydrogenase (PF00106)	86	5	*11	1	2	0	0
Sieve element occlusion C-terminus (PF14577)	5	0	*5	0	0	0	0
Sieve element occlusion N-terminus (PF14576)	6	0	*5	0	0	0	0
Sodium:sulfate symporter transmembrane region (PF00939)	6	0	1	0	*2	1	0
SPX domain (PF03105)	16	3	1	1	*2	1	0
SRF-type transcription factor (DNA-binding and dimerisation domain) (PF00319)	190	*35	6	7	2	1	1
START domain (PF01852)	58	7	3	*6	0	2	*2
STAS domain (PF01740)	18	2	2	1	1	*2	0
Strictosidine synthase (PF03088)	22	2	1	3	*2	0	0
Subtilase family (PF00082)	127	*16	*26	6	3	3	0
Sucrose synthase (PF00862)	16	*4	1	1	0	1	0
Sugar (and other) transporter (PF00083)	108	*21	9	*12	2	2	2
Sugar efflux transporter for intercellular exchange (PF03083)	35	*9	4	*8	1	0	0
Sugar-tranasporters, 12 TM (PF05631)	4	0	*3	0	0	0	0
Sulfate permease family (PF00916)	16	2	2	1	1	*2	0
Sulfite exporter TauE/SafE (PF01925)	10	0	0	*3	0	*3	0
Sulfotransferase domain (PF00685)	57	4	*8	1	1	1	0
Targeting protein for Xklp2 (TPX2) (PF06886)	19	1	*8	2	0	0	0
TATA-binding related factor (TRF) of subunit 20 of Mediator complex (PF08612)	5	*4	0	1	0	0	0
TB2/DP1, HVA22 family (PF03134)	18	2	3	*3	0	0	0
TCP family transcription factor (PF03634)	32	0	*7	1	0	1	0
Terpene synthase family, metal binding domain (PF03936)	104	2	9	1	1	*4	0
Terpene synthase, N-terminal domain (PF01397)	103	4	5	1	0	*4	0
Tetratricopeptide repeat (PF13371)	1	1	0	*1	0	0	0
Thaumatin family (PF00314)	34	3	*9	1	0	*2	1
Thi4 family (PF01946)	4	0	0	*3	0	0	0
Thymidine kinase (PF00265)	5	1	*3	0	0	0	0
tify domain (PF06200)	33	1	4	2	0	*3	0
TMEM214, C-terminal, caspase 4 activator (PF10151)	3	*2	0	0	0	0	0
Transcriptional repressor, ovate (PF04844)	27	*6	*6	2	0	1	0
Transferase family (PF02458)	229	18	*37	*18	5	3	0
Transmembrane amino acid transporter protein (PF01490)	71	9	*12	4	*3	0	0
Trehalose-phosphatase (PF02358)	22	3	2	3	0	*2	1
Trypsin and protease inhibitor (PF00197)	66	*10	5	*9	1	*3	0
Tubulin C-terminal domain (PF03953)	28	0	*9	2	0	0	0
Tubulin/FtsZ family, GTPase domain (PF00091)	34	0	*8	1	0	0	0
Type I 3-dehydroquinase (PF01487)	6	0	*3	0	0	0	0
Type IIB DNA topoisomerase (PF04406)	4	0	0	0	0	*1	0
Ubiquitin-2 like Rad60 SUMO-like (PF11976)	12	*3	2	0	0	0	0
UDP-glucoronosyl and UDP-glucosyl transferase (PF00201)	315	23	*63	*28	6	*7	1
Uncharacterised protein family (UPF0014) (PF03649)	1	1	0	*1	0	0	0
Uncharacterized conserved protein (DUF2358) (PF10184)	6	1	0	*2	0	0	0
Uncharacterized protein family, UPF0114 (PF03350)	5	0	0	*4	0	0	0
Universal stress protein family (PF00582)	48	*7	2	*7	0	0	0
Vacuolar sorting protein 39 domain 1 (PF10366)	3	0	0	0	*1	0	0
Vacuolar sorting protein 39 domain 2 (PF10367)	2	0	0	0	*1	0	0
Viral (Superfamily 1) RNA helicase (PF01443)	6	0	0	*5	1	1	0
Viral coat protein (PF00286)	4	0	0	*2	0	*1	0

→ *continued on next page*

Table S6.7. (continued)

Metabolic pathway	Total	O↑	D↑	D↓	F↑	F↓	O↑D↓F↓
Total sequences	**65287**	**4353**	**4117**	**2815**	**708**	**676**	**218**
Viral methyltransferase (PF01660)	4	0	0	*3	*1	0	0
Wall-associated receptor kinase galacturonan-binding (PF13947)	82	0	*16	4	2	0	0
WAX2 C-terminal domain (PF12076)	13	2	*3	1	0	0	0
Wax ester synthase-like Acyl-CoA acyltransferase domain (PF03007)	10	0	*3	2	0	0	0
Weak chloroplast movement under blue light (PF05701)	12	1	*4	1	0	0	0
Wound-induced protein (PF12609)	20	*5	0	2	0	0	0
WRC (PF08879)	16	3	*6	0	0	0	0
X8 domain (PF07983)	53	1	*19	1	0	0	0
XPG I-region (PF00867)	6	0	*3	0	1	0	0
XPG N-terminal domain (PF00752)	6	0	*3	0	1	0	0
XS zinc finger domain (PF03470)	7	2	0	*2	0	0	0
Xylanase inhibitor C-terminal (PF14541)	116	*16	*28	9	2	2	1
Xylanase inhibitor N-terminal (PF14543)	106	*15	*25	8	2	0	0
Xyloglucan endo-transglycosylase (XET) C-terminus (PF06955)	48	*12	*7	*9	2	*3	*3
Xyloglucan fucosyltransferase (PF03254)	3	1	0	1	0	*1	*1
YABBY protein (PF04690)	11	*3	1	0	0	1	0
YgbB family (PF02542)	1	0	0	*1	0	0	0
ZF-HD protein dimerisation region (PF04770)	21	2	*7	2	0	0	0
Zinc finger, C3HC4 type (RING finger) (PF00097)	26	*5	3	3	0	1	0
Zinc knuckle (PF00098)	52	5	3	*7	0	2	1
Zinc-binding dehydrogenase (PF00107)	82	8	*17	2	*6	3	0
Zinc-finger domain of monoamine-oxidase A repressor R1 (PF10497)	10	0	*3	0	0	0	0
zinc-finger of acetyl-transferase ESCO (PF13878)	2	0	*2	0	0	0	0
zinc-finger of the FCS-type, C2-C2 (PF04570)	22	*7	1	0	0	0	0
Zinc-ribbon (PF14599)	10	*3	1	0	0	0	0

Table S6.8. Detailed view of ovule-enriched proteins containing domains of unknown function (DUFs). DUFs were identified in *S. chacoense* protein sequences by querying the Pfam databse with HMMER. Positions are given in terms of aminoacids. "Expr. BM" designates the expression base mean computed with DESeq2. "A/Leaf", "2 DBA/A" and "*frk1*/A" designate the expression fold-changes in Anth vs. Leaf, 2 DBA vs. Anth, and *frk1* vs. Anth, respectively.

Gene ID	DUF	Pfam ID	Positions	Expr. BM	Expression fold-changes		
					A/Leaf	2 DBA/A	*frk1*/A
Sc01g14900	DUF588	PF04535	35-178	10.07	+48,603	−8,128	+1,031
Sc02g05640	DUF588	PF04535	9-139	668.52	+47,315	−2,554	−1,093
Sc03g16370	DUF588	PF04535	22-164	418.54	+84,747	−2,112	+1,311
Sc03g27260	DUF588	PF04535	32-185	275.54	+10,261	+4,828	+1,483
Sc05g19750	DUF588	PF04535	39-179	578.68	+17,536	−4,301	−1,990
Sc06g18950	DUF588	PF04535	47-194	61.20	+37,437	−5,839	−2,556
Sc06g20810	DUF588	PF04535	27-169	40.80	+5,039	−1,444	−1,100
Sc08g12630	DUF588	PF04535	56-199	5.27	+9,402	−1,217	−1,041
Sc09g20710	DUF588	PF04535	42-189	4.52	+20,335	−1,434	+1,000
ScUng75055	DUF588	PF04535	22-88	138.67	+26,842	−1,979	+1,691
ScUng75060	DUF588	PF04535	3-121	14.27	+164,859	−17,047	−1,021
Sc02g21580	DUF702	PF05142	121-279	764.55	+8,989	+1,457	+1,157
Sc04g22440	DUF702	PF05142	150-307	633.82	+33,215	+1,908	+1,149
Sc11g14150	DUF702	PF05142	143-307	703.65	+7,087	+2,159	+1,035
Sc11g21300	DUF966	PF06136	46-392	395.24	+9,789	+1,761	+1,153
ScUng39570	DUF966	PF06136	39-393	974.23	+5,560	+5,420	−1,044
Sc05g07470	DUF1338	PF07063	82-368	657.33	+15,301	−1,697	−2,693
Sc07g07282	DUF1985	PF09331	20-79	2.19	+10,727	−4,875	−1,023
Sc10g07440	DUF1985	PF09331	225-358	7.47	+6,208	−2,474	−1,005
Sc11g09915	DUF1985	PF09331	2-124	367.73	+7,539	−1,720	−1,151
ScUng00150	DUF1985	PF09331	4-131	11.44	+244,054	−528,991	−1,114
ScUng02580	DUF1985	PF09331	156-272	5.23	+6,628	−60,000	+1,006
ScUng19725	DUF1985	PF09331	166-329	18.42	+6,081	−1,298	−1,046
ScUng41480	DUF1985	PF09331	127-263	14.78	+5,645	−15,435	−3,179
Sc07g19880	DUF2431	PF10354	34-199	218.53	+27,613	+1,539	+1,405
Sc10g17850	DUF3336	PF11815	100-229	246.10	+159,046	+1,370	+1,023
Sc07g22370	DUF3403	PF11883	1470-1513	392.12	+5,890	−3,047	−1,017
Sc01g14400	DUF3444	PF11926	428-576, 651-855	93.71	+11,564	+1,132	−1,021
Sc11g11800	DUF4094	PF13334	9-105	333.73	+624,462	−18,218	−1,036
ScUng56587	DUF4216	PF13952	137-185	8.30	+48,442	−1,018	−1,034
Sc02g05630	DUF4228	PF14009	4-102	41.43	+41,821	−1,410	+1,030
Sc02g11520	DUF4228	PF14009	9-188	107.18	+6,045	−1,354	−1,359
Sc02g23150	DUF4228	PF14009	1-174	3628.65	+12,048	−2,992	−1,103
Sc06g02190	DUF4228	PF14009	1-166	174.62	+16,734	−18,568	−1,041
Sc02g04600	DUF4281	PF14108	97-199	743.56	+17,690	−1,108	+1,061
Sc03g03250	DUF4281	PF14108	93-221	77.37	+5,047	+1,354	+1,147
ScUng68660	DUF4281	PF14108	97-225	142.08	+6,495	−1,226	+1,021
Sc10g09822	DUF4283	PF14111	1-67	58.94	+77,510	+2,090	−1,110
Sc11g10112	DUF4283	PF14111	23-123	9.90	+44,236	+1,136	+1,004
Sc04g06975	DUF4371	PF14291	1-113	3.52	+24,393	−60,911	−1,018
Sc08g03808	DUF4371	PF14291	1-155	5.19	+56,222	−30,951	−1,032
Sc10g08135	DUF4371	PF14291	46-232	116.46	+506,590	−39,997	−1,567
Sc02g24990	DUF4378	PF14309	755-922	1385.31	+5,060	−1,702	+1,032
Sc09g02660	DUF4378	PF14309	399-471	1307.79	+5,386	−1,154	−1,739
Sc12g04550	DUF4666	PF15697	67-120, 126-152	619.85	+46,305	−30,660	−1,009

ANNEXE E

Matériel supplémentaire du chapitre 7

Tab. S7.1. Coordonnées génomiques des séquences codant pour la portion mature des DEFL identifiées *S. chacoense*, avec celles de leurs homologues chez *S. tuberosum*. Les coordonnées des homologues ont été identidiées dans le génome de *S. tuberosum* v. 4.03 par BLAST.

Gène	*S. chacoense*			*S. tuberosum*		
	Chrom.	Start	End	Chrom.	Start	End
Sc01g00040	chr. 1	172 010	171 870	chr. 1	152 561	152 421
Sc01g06482	chr. 1	9 790 874	9 790 716	chr. 00	19 770 161	19 770 003
Sc01g09732	chr. 1	27 767 066	27 767 224	chr. 1	20 951 669	20 951 514
Sc01g09738	chr. 1	27 784 861	27 785 019	chr. 1	20 951 669	20 951 514
Sc03g15090	chr. 3	34 387 277	34 386 975	chr. 3	46 610 992	46 610 690
Sc04g00195	chr. 4	243 128	242 916	chr. 00	12 501 582	12 501 370
Sc04g02195	chr. 4	2 302 340	2 302 480	chr. 4	3 435 712	3 435 852
Sc04g14918	chr. 4	29 200 486	29 200 313	chr. 4	62 201 397	62 201 222
Sc05g00860	chr. 5	632 307	632 104	chr. 5	635 329	635 126
Sc05g01592	chr. 5	1 190 207	1 190 090	chr. 5	1 165 327	1 165 210
	chr. 5	1 189 897	1 189 800	chr. 5	1 165 021	1 164 924
Sc05g07242	chr. 5	7 028 772	7 028 539	chr. 5	6 238 070	6 238 303
Sc05g07245	chr. 5	7 042 159	7 041 941	chr. 5	6 222 758	6 222 976
Sc06g11955	chr. 6	27 560 939	27 560 942	chr. 6	44 314 997	44 315 000
	chr. 6	27 561 054	27 561 250	chr. 6	44 315 112	44 315 300
Sc06g20400	chr. 6	35 493 993	35 493 774	chr. 6	53 614 179	53 614 398
	chr. 6	35 492 673	35 492 588	chr. 6	53 615 493	53 615 578
Sc07g03390	chr. 7	4 027 443	4 027 446	chr. 00	29 280 398	29 280 395
	chr. 7	4 028 051	4 028 220	chr. 00	29 279 731	29 279 562
Sc07g03400	chr. 7	4 034 466	4 034 606	chr. 00	29 275 500	29 275 360
Sc07g03405	chr. 7	4 046 852	4 046 992	chr. 7	3 688 442	3 688 302
Sc07g03445	chr. 7	4 133 374	4 133 514	chr. 7	3 645 550	3 645 410
Sc07g03465	chr. 7	4 157 951	4 157 799	chr. 7	3 619 508	3 619 660
Sc07g04620	chr. 7	5 913 916	5 914 065	chr. 7	7 365 796	7 365 647
Sc07g04630	chr. 7	5 972 542	5 972 408	chr. 7	7 208 962	7 209 096
Sc07g04920	chr. 7	6 902 079	6 901 840	chr. 7	9 042 227	9 041 988
Sc07g11045	chr. 7	30 606 796	30 606 629	chr. 7	46 851 766	46 851 599

→ *voir page suivante*

Tab. S7.1. (suite)

Gène	*S. chacoense*			*S. tuberosum*		
	Chrom.	Start	End	Chrom.	Start	End
Sc08g02255	chr. 8	2 547 233	2 547 099	chr. 8	52 896 011	52 896 163
Sc09g04835	chr. 9	4 936 578	4 936 483	chr. 9	6 282 267	6 282 362
Sc09g10288	chr. 9	22 385 442	22 385 603	chr. 9	47 960 401	47 960 562
Sc09g19015	chr. 9	32 397 466	32 397 660	chr. 9	59 275 946	59 276 140
Sc09g21162	chr. 9	34 370 108	34 369 965	chr. 9	2 863 872	2 864 018
Sc09g21164	chr. 9	34 375 046	34 374 885	chr. 9	2 856 099	2 856 260
Sc09g21168	chr. 9	34 379 816	34 379 646	chr. 9	2 851 473	2 851 643
Sc09g21242	chr. 9	34 493 803	34 493 561	chr. 9	2 742 282	2 742 524
Sc09g21245	chr. 9	34 496 026	34 495 802	chr. 9	2 740 328	2 740 552
Sc09g21248	chr. 9	34 499 276	34 499 273	chr. 9	2 737 223	2 737 226
	chr. 9	34 498 306	34 498 068	chr. 9	2 738 092	2 738 330
Sc10g03055	chr. 10	2 951 783	2 951 932	chr. 10	4 235 762	4 235 619
Sc10g08645	chr. 10	22 006 436	22 006 260	chr. 10	47 140 587	47 140 411
Sc10g09020	chr. 10	22 868 401	22 868 228	chr. 10	48 005 423	48 005 250
Sc10g13555	chr. 10	28 416 895	28 416 743	chr. 10	52 896 011	52 896 163
Sc10g13665	chr. 10	28 550 282	28 550 130	chr. 00	27 913 836	27 913 684
Sc10g18935	chr. 10	34 821 546	34 821 403	chr. 10	58 830 581	58 830 724
Sc11g17165	chr. 11	33 731 811	33 731 972	chr. 11	41 482 662	41 482 823
Sc11g20045	chr. 11	36 534 587	36 534 835	chr. 11	43 784 337	43 784 585
Sc12g08000	chr. 12	11 115 838	11 115 530	chr. 12	31 546 786	31 547 094
Sc12g12265	chr. 12	33 134 178	33 133 978	chr. 12	51 355 369	51 355 169
Sc12g18185	chr. 12	44 447 279	44 447 085	chr. 12	53 725 163	53 724 969
ScUng03135	scaff. 105	539 985	539 827	chr. 1	23 835 866	23 836 024
ScUng12282	scaff. 344	478 752	478 561	chr. 1	23 145 704	23 145 513
ScUng35555	scaff. 826	19 179	19 346	chr. 7	4 886 090	4 886 257
ScUng37761	scaff. 875	262 828	262 679	chr. 7	289 930	289 781
ScUng60355	scaff. 1563	14 162	14 001	chr. 6	58 971 579	58 971 740
ScUng69230	scaff. 2060	23 898	24 032	chr. 7	7 208 962	7 209 096
ScDng02289	ScDng02289	150	201	—	—	—
ScDng03407	ScDng03407	104	361	—	—	—
ScDng04274	ScDng04274	170	409	chr. 11	16 089 481	16 089 720
ScDng04736	ScDng04736	248	481	chr. 7	878 977	878 744
ScDng08283	ScDng08283	133	273	chr. 00	29 279 702	29 279 562
ScDng08860	ScDng08860	183	326	chr. 4	3 435 718	3 435 857
ScDng08953	ScDng08953	106	255	chr. 1	50 807 715	50 807 575
ScDng09296	ScDng09296	113	253	chr. 1	160 642	160 502
ScDng13861	ScDng13861	86	214	chr. 3	37 559 850	37 559 972

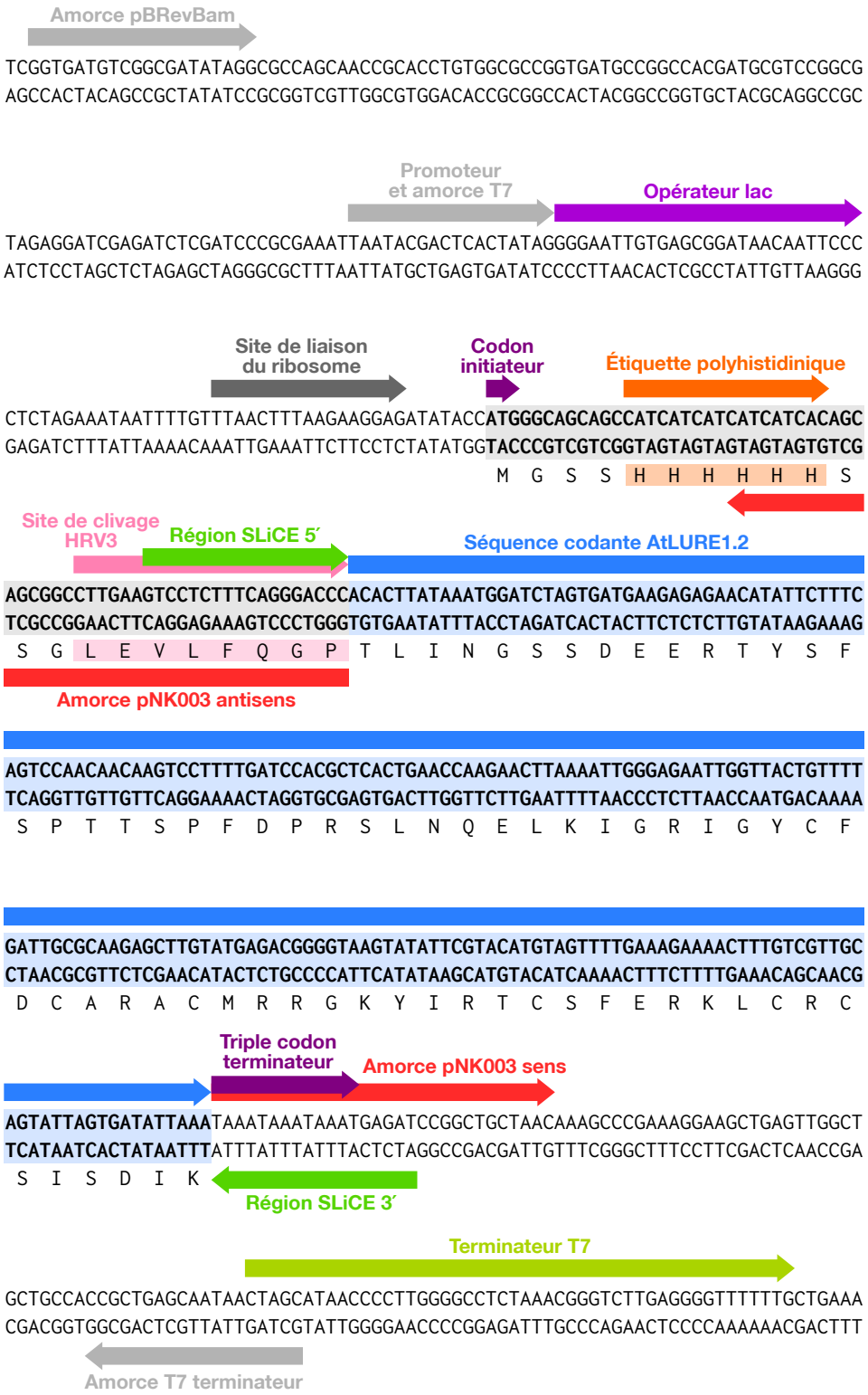


Fig. S7.1. Détail de la cassette d'expression du plasmide PNK003 natif.

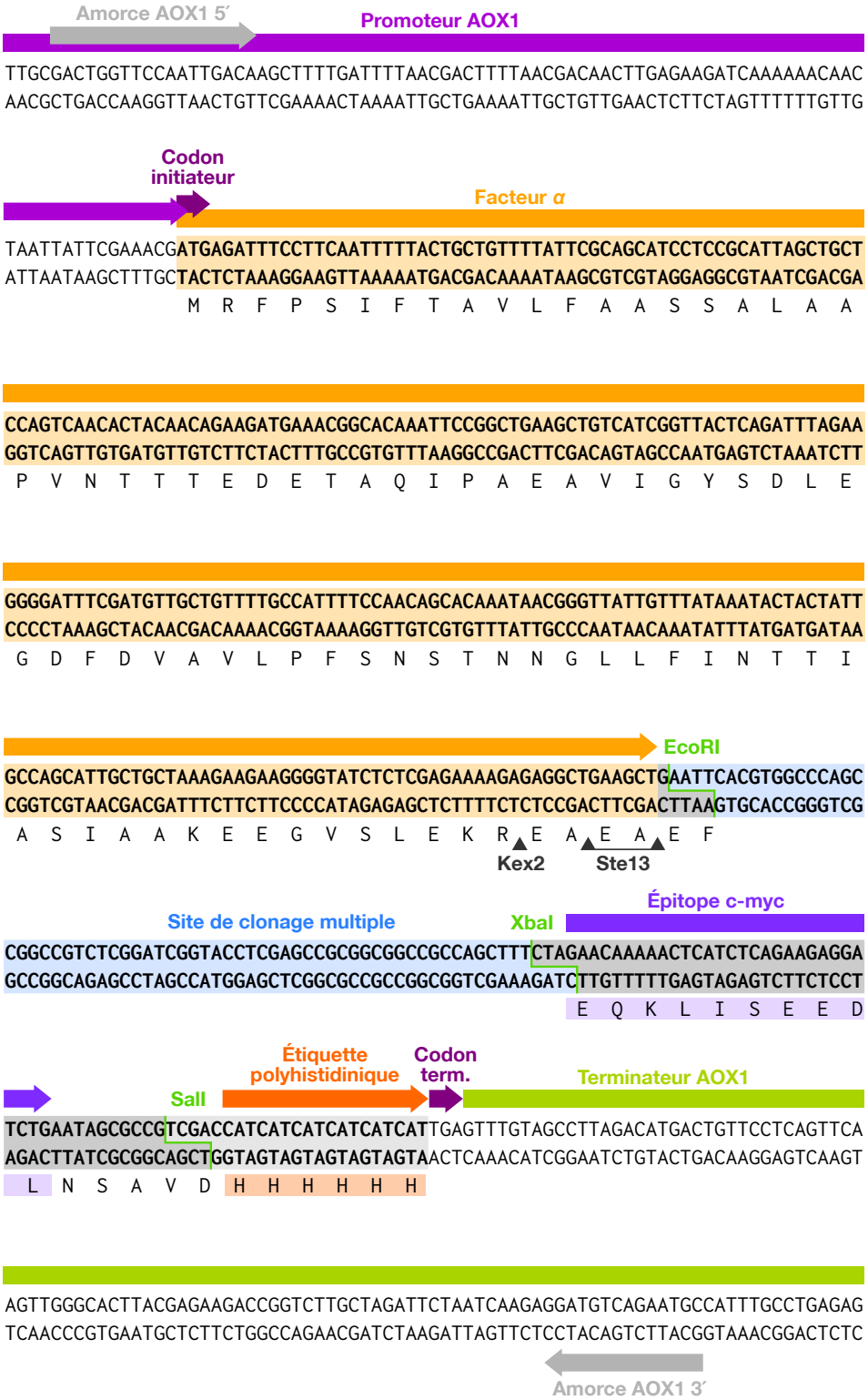


Fig. S7.2. Détail de la cassette d'expression du plasmide pPICZαA natif.

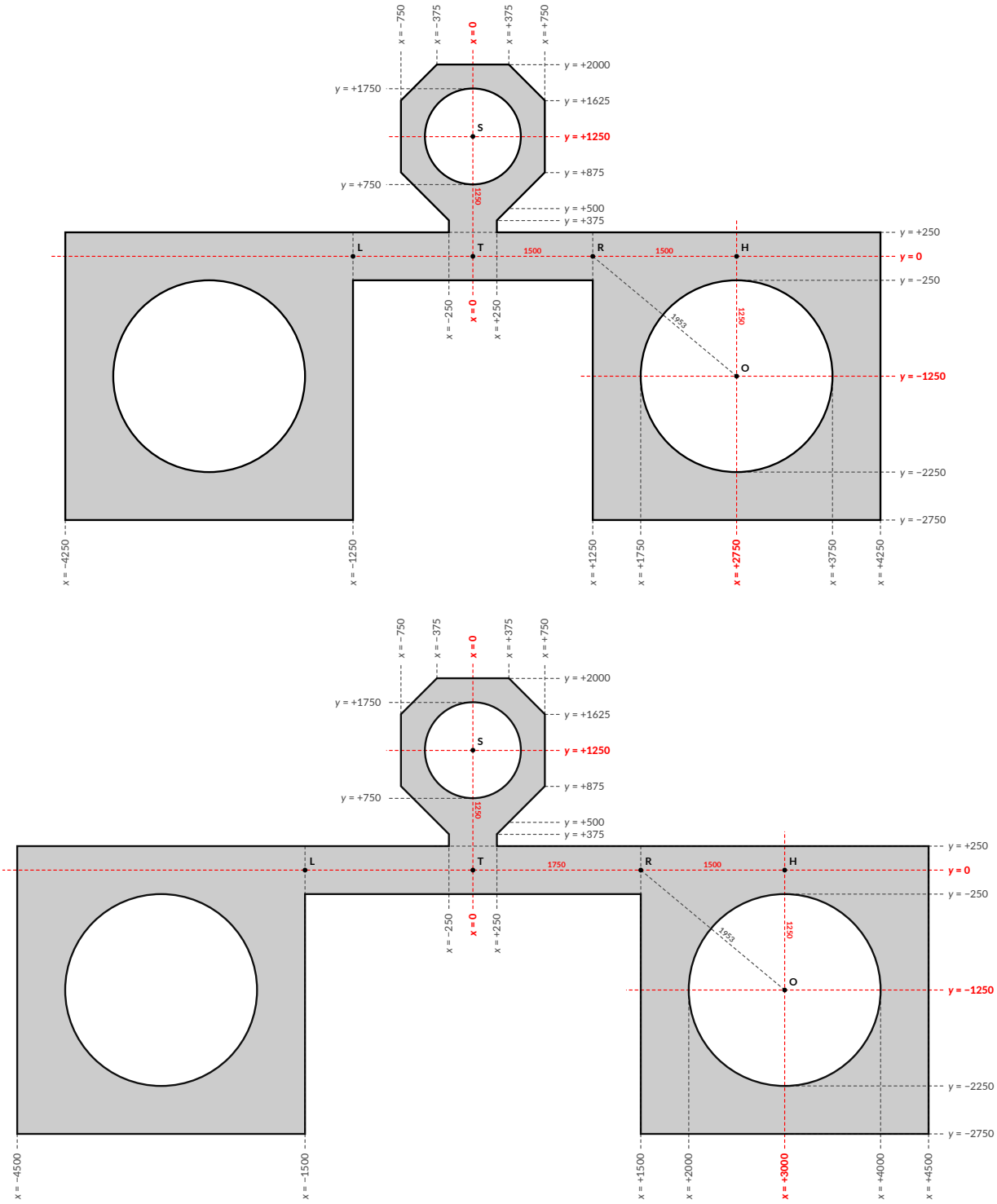


Fig. S7.3. Schémas des dispositifs microfluidiques alternatifs pour tests de guidage semi-*in vivo* à longue distance chez *Solanum*. Les coordonnées sont données en micromètres. Profondeur : 100 µm.

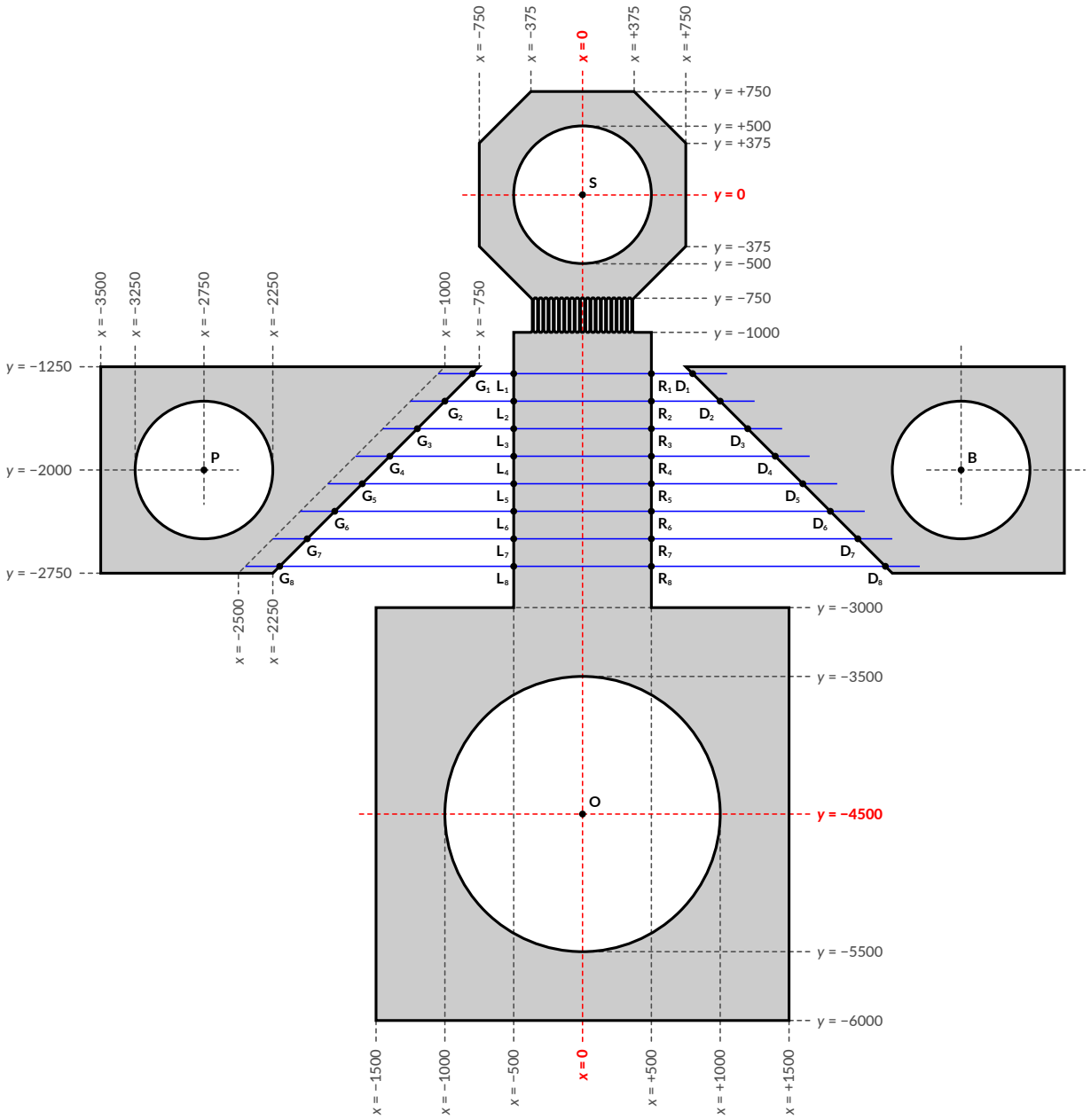


Fig. S7.4. Schéma du dispositif microfluidique alternatif pour tests de guidage semi-*in vivo* à courte distance chez *Solanum*, avec test simultané de différentes concentrations. Les coordonnées sont données en micromètres. Profondeur : 30 µm pour la couche inférieure, 1 µm pour les canaux de la couche supérieure (traits bleus).