Université de Montréal

Génomique fonctionelle de la transduction de signal, isolement et caractérisation de récepteurs kinases chez Solanum chacoense.

par
Hugo Germain

Département de Sciences Biologiques
Faculté des Arts et Sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Doctorat
en Sciences Biologiques
option Biologie Végétale

avril, 2007

© Hugo Germain, 2007

Faculté des études supérieures
Grade conféré à compter du
2007 OCT 04
Université de Montréal

QK
3
U54
2007
V.008

Université de Montréal

Direction des bibliothèques

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Génomique fonctionelle de la transduction de signal, isolement et caractérisation de récepteurs kinases chez Solanum chacoense.

présentée par :
Hugo Germain

a été évaluée par un jury composé des personnes suivantes :

Jean Rivoal, président-rapporteur et représentant du doyen de la FES
Daniel Philippe Matton, directeur de recherche
Patrick Gullick, membre du jury
Daphne Goring, examinateur externe

Résumé

Dans le but de caractériser des gènes régulatoires qui sont exprimés dans les ovules suite à la fécondation nous avons entrepris un projet de séquençage d'EST chez *Solanum chacoense*, une espèce autoincompatible de patate sauvage. Deux banques de cDNA faites à partir de tissu d'ovules couvrant les stades de zygote à torpille ont été construites et étalées à haute densité sur des membranes de nylon. Pour diminuer la redondance des EST et pour enrichir en transcrits peu abondant, nous avons utilisé une soustraction avec une sonde radio-active faite à partir des ARNm ayant servis à construire les banques afin de sélectionner les ADNc à séquencer. 7741 bonnes séquences ont été obtenues, de ces séquences et nous avons isolé 6374 unigènes. La méthode de soustraction a donc permis un fort enrichissement en singletons, une diminution du nombre de gènes par contigs, et un fort enrichissement du nombre d'unigènes obtenus (82%). Afin de mieux comprendre les mécanismes de transduction de signaux qui ont lieu pendant le développement de l'embryon, le niveau d'expression relatif d'ARNm de tous les récepteurs a été analysé à partir de tissus reproducteur mâle, femelle et de tissus végétatif. 28 des 30 récepteurs kinases analysés étaient exprimés de façon prédominante dans les ovaires et 23 étaient induits suite à la fécondation. Donc, notre méthode de soustraction n'a pas apporté de biais en sélectionnant les gènes faiblement exprimés dans les ovaires et fortement exprimés dans les autres tissus. De tous les récepteurs kinases analysés, la famille des *leucine-rich repeat* était de loin la plus abondante avec 25 membres de 11 sous-familles différentes.

Des trentes récepteurs kinases analysés en RT-PCR nous en avons sélectionnés deux dont le niveau d'expression était particulièrement restreint au niveau de l'ovule, soit : *ScORK17* et *ScORK28*. *Solanum chacoense* Ovule Receptor Kinase 17 est un récepteur kinase de la sous-famille LRR-VI et son expression est spécifique aux tissus reproducteurs femelles. La sous-famille LRR-VI compte 11 membres chez *A. thaliana* et à notre connaissance, à ce jour, aucun des membres ne s'est vu attribué une fonction. Nous avons observé qu'un arbre phylogénétique fait à partir du domaine kinase des membres de la

famille LRR-VI chez *Arabidopsis* et le riz sépare la famille en deux clades, soit un clade contenant les membres ayant une moyenne de 8,2 LRR par protéine et ceux ayant 2,7 LRR par protéine. Les hybridations *in situ* ont démontrés que le transcrit de *ScORK17* était principalement présent dans le tégument de l'ovule et dans l'endothélium. Nous avons aussi démontré par expression transitoire que *ScORK17* est N-glycosylé *in planta*. Puisqu'aucune lignée t-DNA n'était disponible chez SALK pour l'orthologue de *ScORK17* chez *A. thaliana,* nous avons commandé la seule lignée disponible chez GABI, mais aucun défaut phénotypique n'a été observé de même que chez les transgéniques surexprimants chez *S. chacoense*. Par contre lorsque nous avons introduit une version tronquée de *ScORK17* dans *Arabidopsis thaliana*, les plantes transgéniques résultantes ont un nombre de graines fortement réduit ce qui supporte l'hypothèse que ScORK17 joue un rôle dans le développement.

L'expression tissulaire de *ScORK28* au niveau des ARN messagers et au niveau protéique est restreinte aux structures reproductrices femelles. Des analyses d'expression protéiques *in planta* ont démontrées que ScORK28 est N-glycosylé et que les fusions ScORK28:GFP sont localisé à la membrane plasmique. L'expression du domaine kinase de ScORK28 en bactérie nous a permis de déterminer que celui-ci est actif et nécessite du Mg^{2+} pour pouvoir s'auto ou se transphosphoryler. Ces analyses ont aussi démontré que le domaine juxtamembranaire intracellulaire est nécessaire à l'activité kinase et est probablement un site de phosphorylation. Les lignées transgéniques obtenues chez *S. chacoense* ont démontré que le niveau d'expression de la protéine ScORK28 était régulé de façon très stringente puisqu'aucune plante démontrant un niveau d'expression de protéine plus élevé que les types sauvages n'a été obtenu malgré une forte accumulation de l'ARN messager.

Mots-clés : récepteur kinase, ligand, peptide de signalisation, communication intercellulaire, développement embryonnaire, *express sequence tag*, soustraction virtuel, fertilisation, PCR en temps réel, localisation cellulaire, membrane plasmique, activité kinase, transformation hétérologue.

Abstract

In order to characterize regulatory genes that are expressed in ovule tissues after fertilization we have undertaken an EST sequencing project in *Solanum chacoense*, a self-incompatible wild potato species. Two cDNA libraries made from ovule tissues covering embryo development from zygote to late torpedo-stage were constructed and plated at high density on nylon membranes. To decrease EST redundancy and enrich for transcripts corresponding to weakly expressed genes a self-probe subtraction method was used to select the colonies harboring the genes to be sequenced. 7741 good sequences were obtained and, from these, 6374 unigenes were isolated. Thus, the self-probe subtraction resulted in a strong enrichment in singletons, a decrease in the number of clones per contigs, and concomitantly, an enrichment in the total number of unigenes obtained (82%). To gain insight into signal transduction events occurring during embryo development all the receptor-like kinases (or protein receptor kinases) were analyzed by quantitative realtime RT-PCR. Interestingly, 28 out of the 30 RLK isolated were predominantly expressed in ovary tissues or young developing fruits, and 23 were transcriptionaly induced following fertilization. Thus, the selfprobe subtraction did not select for genes weakly expressed in the target tissue while being highly expressed elsewhere in the plant. Of the receptor-like kinases (RLK) genes isolated, the leucine-rich repeat (LRR) family of RLK was by far the most represented with 25 members covering 11 LRR classes.

Of the thirty receptor kinases that were analyzed by RT-PCR, two that had an expression pattern that was mostly predominant in the female tissue, were selected for further caracterization, these are : *ScORK17* and *ScORK28*. *Solanum chacoense Ovule Receptor Kinase 17* is a receptor kinase of the LRR-VI subfamily and its expression is highly specific to the female reproductive tissue. The LRR-VI subfamily counts eleven members in *Arabidopsis thaliana* and, to this day, no members has been attributed a function. We observed that a phylogenetic tree inferred using the kinase domain of LRR-VI

subfamily members separates the family into two clades ; one containing an average of 8,2 LRR per protein and a second clade containing an average of 2.7. *In situ* hybridization was analyzed for *ScORK17* and the signal was shown to be mainly present in single ovule integument that is developing into the seed coat and in the endothelium. Transient expression also revealed that *ScORK17* is N-glycosylated *in planta*. Since no t-DNA containing SALK lines were available for *ScORK17* putative ortolog we ordered and analysed the only available GABI-line but no phenotypic abnormalities were observed in the homozygous population or in overexpressing lines in *S. chacoense*. However when a truncated *S.chacoense* clone was transformed in *A. thaliana* the resulting transgenic plants harbored a much reduced seed set, supporting a developmental role for *ScORK17*.

Solanum chacoense Ovule Receptor Kinase 28 displayed high level of tissue specificity at the RNA and protein levels and was predominantly expressed in female reproductive tissues. Protein expression analyses *in planta* revealed that ScORK28 was N-glycosylated and ScORK28:GFP fusion analyses showed that it was localized at the plasma membrane. Bacterial expression of ScORK28 kinase catalytic domain revealed that ScORK28 is an active Mg^{2+}-dependent kinase that can either auto or transphosphorylate. It also showed that the intracellular juxtamembrane domain is necessary for kinase activity and is most probably a major phosphorylation site. Transgenic overexpression lines with full-length or dominant negative constructs showed that ScORK28 protein accumulation was tightly controlled as no plants with significantly higher ScORK28 protein levels could be obtained even in plants strongly overexpressing *ScORK28* mRNAs.

Keywords : Receptor kinase, ligand, signal peptide, cell-cell communication, embryo development, express sequence tag, virtual subtraction, real-time PCR, fertilization, subcellular localisation, plasma membrane, kinase activity, heterologous transformation.

Table des matières

Liste des tableaux

Liste des figures

Liste des sigles et abbréviations

Sauf si mentionné, le gène et son abréviation proviennent de l'espèce *Arabidopsis thaliana*.

χ^2	test statistique
°C	degré Celcius
%	pourcentage
18S	*ARN ribosomal 18S* de *Solanum chacoense* ou d'*Arabidopsis thaliana*
3'	extrémité 3' hydroxyl
5'	extrémité 5' phosphate
Å	Amstrong, unité de mesure
a/A	adénine
A	alanine
ABI	Applied Biosystems
ACF	Alternative catalytic function
ACR4	Gène CRINKLY4 chez *Arabidopsis thaliana*
ANP	atrial natriuretic peptide
Approx	approximately
ARN	acide ribonucléique
ARNm	ARN messager
ATP	adénosine triphosphate
ATPase	adénosine triphosphatase
AtPNP-A	protéine PNP-A de *Arabidopsis thaliana*

AZ	abcission zone
BAK1	BRI1 associated kinase 1
Bg	Basic 7S globulin protein
BiFC	bimolecular fluorecence complementation
BKI1	BRI1 kinase inhibitor 1
BLAST	Basic Local Alignment Search Tool
BLASTN	Basic Local Alignment Search Tool nucleotide
BoSRK	*Brassica oleracea* S-locus receptor kinase
bp	base pair
BRI1	gène *Brassinolide receptor insensitive 1*
BRICK	gène de la famille BRICK
BRICK1	gène *BRICK1*
BRICK2	gène *BRICK2*
BRICK3	gène *BRICK3*
C	cytosine
C	cystéine
CYS	cystéine
CA	California
CaMV 35S	cauliflower mosaic virus 35S promoter
CAX	Calcium exchanger
cDNA	complementary DNA
CDPK	calcium dependent protein kinase

cGMP	cyclic guanosine-3'-5'-monophosphate
CLE	CLAVATA/ESR
CLV1	gène *CLAVATA1*
CLV3	gène *CLAVATA3*
cm	centimètre
CPP	cystein cluster protein
CR4	gène CRINKLY4
C-terminal	domaine carboxyl terminal d'une protéine
C-terminus	extrémité carboxy-terminale d'une protéine
CTR1	gène *CONSTITUTIVE TRIPLE RESPONSE 1*
d	day
D	acide aspartique
DAP	days after pollination
dATP	déoxyadénosine triphosphate
DIC	differential interference contrast
DMSO	Dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate/désoxyribonucléotide triphosphate
Dr.	Docteur
d(T)	deoxythymidine
DTT	dithiothreitol
DUF	domain of unknown function

DVL	gène *DEVIL*
E	acide glutamique
EGF	epidermal growth factor
email	electronic mail
EMBL	European Molecular Biology Laboratory
EMS	Ethyl Methhyl Sulfonate
EMS1	gène EXCESS MICROSPOROGENOUS 1 même que EXS
ENOD40	gène Early nodulin 40
ESR	Embryo surrounding region
EST	expressed sequence tag
E-value	valeur statistique
EXS	extrasporogenous cells
F	phénylalanine
F58C	domaine des facteurs de coagulation V and VIII
FAX	facimile
Fig.	figure
FLS2	gène *FLAGELLIN SENSITIVE*
FN3	fibronectine type III
FQRNT	Fonds Québécois de la Recherche sur la Nature et les Technologies
FRET	Fluorescence resonance energy transfer

FU	Furin-like domain
FZ	Frizzled cystein-rich domain
G	guanine
G	glycine
G4	lignée G4 de *Solanum chacoense*
GE	General Electric
GFP	Green Fluorescent Protein
GMC	guard mother cell
GmENOD	gène ENOD chez *Glycine max*
GSI	gametophytic self-incompatibility
GUS	gène glucoronidase permettant une coloration
h	hour/heure
H	histidine
^{3}H	tritium
H^{+}	ion hydrogène
HAP	hours after pollination
HAR1	hypernodulation autoregulatory receptor kinase
H. glycines	espèce *Heterodera glycines*
HIS	histidine
HgCLE	gène CLE chez *Heterodera glycines*
HT	gene HT
I	isoleucine

IDA	gène *Inflorescence deficient in abcission*
IDL	IDA-like
Ig	domaine immunoglobuline-like
IgC2	domaine immunoglobuline-like C2
IGF	insulin growth factor
IPT	immunoglobin-like fold shared by plexins and transcription factors
IRAK	interleukin-1 receptor associated kinase
IRBV	Institut de Recherche en Biologie Végétale
irPNP	immunoreactive PNP
JM	domaine juxtamembranaire
K	lysine
K^+	ion potassium
KAPP	gène kinase associated protein phosphatase
kb	kilobase
Kd	constante de dissociation
kDa	kiloDalton
L/l	litre
L	leucine
LB	Luria-Bertani
LBA4404	souche d'*Agrobacterium tumefaciens*
LCR	low molecular weight cystein-rich protein

LDLA	low density lipoprotein A
LiAc	acétate de lithium
LRR	leucine-rich repeat
LYK3	LysM domain containing receptor kinase
LysM	domaine de ~40 résidus impliqué dans la liaison à des peptidoglycans
M	méthionine
M	molaire
MALDI-TOF	Matrix assisted laser desorption assay time-of-flight
MAM	meprin antigen domain
MAPK	protéine Mitogen activated protein kinase
MARK	maize atypical receptor kinase
MBP	myelin basic priotein
MD	Maryland
min	minute
MIPS	Munich Information Center for Protein Sequence
mg	miligramme
mJ	miliJoule
ml	milliliter
mm	milimètre
mM	milimolaire
MPK4	mutant du gene MPK4

mRNA	messenger RNA
MS	mass spectrometry
MS	Murashige and Skoog
MW	molecular weight
n	nombre d'échantillon/ d'acide nucléique
N	unité de concentration; Normal
N	asparagine
Na^+	ion sodium
NARK	Nodule autoregulation receptor kinase
NaHT	gène HT de *Nicotiana alata*
NCBI	National Center for Biological Information
NCR	nodule cysteine-rich
NFR1-5	Nod factor recognition
ng	nanogramme
nm	nanomètre
nmol	nanomole
no.	number/numéro
NORK	Nodule receptor kinase
NP	natriuretic peptide
NRC	National Research Council
NSERC	Natural Sciences and Engineering Research Council of Canada
nt	nucléotide

NtRALF	gène RALF chez *Nicotiana tabacum*
N-terminal	domaine amino-terminal d'une protéine
N-terminus	extrémité amino-terminale d'une protéine
OFP	oligonucleotide fingerprinting
ON/ONT	Ontario
ORF	open reading frame
ORK	ovule receptor kinase
OsEAL	*Oryza sativa* Egg apparatus-like
OsTMK	*Oryza sativa* transmembrane receptor kinase
OX	ligné surexprimante
P	proline
^{32}P	phosphate radioactif
pANP	plant atrial natriuretic peptide
PCR	polymerase chain reaction
pH	potentiel hydrogène
Ph D	Philosophiæ doctor
PE	Perkin Elmer
PERK	Proline-rich extensin-like receptor kinase
pI	point isoélectrique
PLS	gène *POLARIS*
PNGaseF	peptide N-glycosidase F
PNP	plant natriuretic peptide

poly(A)	polyadénine
PSI	domaine retrouvé chez les plexines
PRK	protéine récepteur kinase
PRK1	pollen receptor kinase 1
PSK	phytosulfokine
Pwo	ADN polymérase de *Pyrococcus woesei*
Q	glutamine
QC	Québec
QPCR	quantitative PCR
R	purine : soit les nucléotides adénine ou guanine
RALF	gène Rapid alcalinization factor
rANP	rat atrial natriuretic peptide
RD	type de protéine kinase
RE	reticulum endoplasmique
REC	domaine contenant un site phosphoaccepteur cible d'une histidine kinase
RKF3	receptor kinase in flower
RLK	receptor-like kinase
RLK4	gène RLK4
RLK5	gène RLK5
RK	récepteur kinase

RNA	ribonucleic acid
RNAi	RNA interference
ROP	Rho-of-plants
ROT	gene *Rotundifolia*
RT-PCR	reverse transcription polymerase chain reaction
RTF	domain situé en C-terminal du gène *ROT*
S/s	sérine
S	unité de sédimentation d'une ou d'un ensemble de molécules
S-domain	domaine protéique associé au locus S d'incompatibilité
SAM	shoot apical meristem
SC	self-compatible
ScHT A	gène HT A chez *Solanum chacoense*
ScHT B	gène HT B chez *Solanum chacoense*
SCR	gène S-locus cystein rich protein
SD	standard deviation
SEMA	domaine retrouvé chez les semaphorines
sec	secondes
SERK	gène somatic embryogenesis receptor kinase
SI	self-incompatibility
SMC	subsidiary mother cell
SOL	supressor of ligand-like protein
SP11	S-locus protein 11

SSI	self-incompatibility system
S-RNA	RNAse associé au locus S
SRK	S-locus receptor kinase
SSC	salt sodium citrate buffer
SUB	Strubellig
SRF	Strubellig receptor family
SUS	sucrose synthase
SV	soustraction virtuel
SYMRK	symbiosis receptor kinase
T	thymine
T/t	thréonine
TBS	Tris buffered saline
TBST	Tris buffered saline +Tween
T-DNA	transfer DNA
TBLASTN	BLAST qui compare une séquence protéique contre une séquence d'ADN traduite dans tous les cadres de lecture
tel	téléphone
TGFβ	récepteur transforming growth factor
THL	thioredoxin-h-like
TLA	gène *transparent leaf area 1*
TNFR	tumor necrosis factor receptor

TOAD	gène TOADSTOOL
TobHypSys	systémine du tabac contenant des hydroxyprolines
TomhypSys	systémine de tomate contenant des hydroxyprolines
TomproHypSys	précurseur de la systémine de tomate contenant des hydroxyprolines
TRIP	TGFβ-receptor interacting protein
TTL	gène transthyretin-like
U	uracile
U	unité d'activité d'une enzyme
URK1	domaine retrouvé chez les uridine kinases
USA	United States of America
UTR	untranslated region
UV	ultraviolet
V	volt
VH1	Vascular highway
VIRSC	virtual subtraction control
V22	génotype comportant les allèles d'incompatibilité S_{12},S_{14}
W	tryptophane
WAK	wall associated kinase
WIF	domaine chez les Ryk récepteur tyrosine kinases
WT	wild type (témoin, type sauvage)
WUS	gène *WUSCHEL*

x/X	n'importe quel acide aminé
X	unité de concentration d'une solution
Xa21	récepteur conférant la résistance à un pathovar de *Xanthomonas*
Y	tyrosine
YFP	yellow fluorescent protein
ZmEA1	gène *Zea mays EGG APPARATUS 1* de *Zea mays*
ZmKIK1	gène *Zea mays kinase domain interacting kinase 1*
µg	microgramme
µl	microlitre
µm	micromètre
µM	micromolaire

Remerciements

Je dois un énorme merci à Daniel, mon directeur. Lorsque je me suis présenté à lui il y près de sept ans en lui disant que je voulais faire un doctorat dans son laboratoire, je n'avais utilisé à peu près aucune des techniques qui sont utlisées de routine dans son laboratoire et je ne comprenais pas grand chose à ce qu'il disait. Bref, je partais de loin! Daniel m'a fait confiance et m'a offert une formation plus que complète. En plus d'assurer la direction de mon projet Daniel m'a permi de participer à des congrès internationaux, de faire un stage au Japon, d'enseigner des cours, de superviser des stagiaires, il m'a donné la chance d'être auxiliaire d'enseignement à plusieurs reprises et m'a formé à écrire et évaluer des articles. Parce qu'il a été beaucoup plus qu'un directeur...merci Daniel!

Je dois aussi avouer que j'ai, en partie, choisi le laboratoire de Daniel à cause du dynamisme des étudiants qui s'y trouvaient lorsque je l'ai visité. Martin, Sier-Ching et Sébastien je sais que je serais peut-être toujours 'le petit nouveau' à vos yeux mais je vous remercie de tout ce que vous m'avez appris et surtout des bons moments que nous avons passé ensemble. Eric…tu as été un stagiaire de qualité supérieure, ta joie de vivre met une belle atmosphère dans le laboratoire. Je te souhaite que la fin de ton doctorat se termine bien, dans les délais que tu souhaites et que tu obtiennes des publications de qualité supérieure. À tous les autres étudiants qui ont été au labo en même temps que moi, et aux étudiants de l'IRBV qui ont cheminé avec moi, merci d'avoir fait rendu mon long séjour ici aussi agréable.

Je dois aussi remercier mes parents. Mes parents étaient tous les deux dans l'enseignement, et ils ont toujours fait de l'éducation quelque chose de très important. Ils m'ont encouragé dès la maternelle alors que j'avais de la difficulté à faire mes boucles, au primaire lorsque je voulais devenir livreur de pizza, au secondaire lorsque je voulais devenir astronaute et plus tard lorsque je ne savais plus trop exactement ce que je voulais devenir. Merci pour votre support indéfectible.

Et finalement merci à ma femme Fannie. Elle qui a compris qu'il ne faut pas demander : Quand est-ce que tu vas terminer ? et qui m'a supporté pendant toutes ces années me demandant régulièrement sur quoi exactement je travaillais. Merci pour tout et merci pour ce beau petit garçon que nous avons.

Chapitre 1

Introduction

Dans les organismes biologiques complexes, hautement spécialisés et sessiles comme les plantes, la communication cellule-cellule et environnement-cellule est primordiale pour assurer un développement optimal de l'organisme. Cette information peut être perçue par les récepteurs kinases de plantes (PRK) (Cock et al., 2002)[1]. La fonction des PRKs est de percevoir de l'information (ligand[2]) par le biais de leur domaine extracellulaire et de la transmettre vers l'intérieur de la cellule grâce à leur domaine kinase intracellulaire. Ce signal pourra ensuite être relayé au noyau par divers moyens ou cascades de signalisations permettant la régulation de l'expression génique et une réponse appropriée de la plante face au stimulus.

Selon les prédictions bioinformatiques, le génome d'*Arabidopsis thaliana* compterait pas moins de 610 PRK (Shiu et Bleecker, 2001a, b). Cependant, seulement 417 seraient de vrais PRK, c'est-à-dire des protéines ayant un domaine extracellulaire, un domaine transmembranaire et un domaine sérine/thréonine kinase. De ces 417 PRK, 216 seraient des 'leucine-rich repeat' PRK (LRR-PRK). Il s'agit de loin de la plus grande famille. Le nombre de PRK retrouvé chez les plantes est très élevé si on le compare à celui retrouvé chez l'humain où seulement 70 récepteurs kinases sont présents et où 58 sont des récepteurs tyrosines kinases, un domaine absent chez les récepteurs de plantes (Manning et al., 2002). Il faut noter que ce type de récepteur est très largement distribué à travers les espèces et on a longtemps cru qu'il était absent des eucaryotes unicellulaires (Hunter et Plowman, 1997). Cependant des récepteurs de type tyrosine kinase ont récemment été

---

[1] L'appellation 'recepteur-like kinase' (RLK) est souvent utilisé mais depuis la découverte de plusieurs ligands la littérature identifie maintenant ces protéines comme des 'plant receptor kinase' (PRK) au même titre que les récepteurs kinases (RK) animaux. Dans les chapîtres en anglais l'abbréviation RLK est généralement utilisée.

[2] La section couvrant les ligands des récepteurs kinases se trouve dans le chapître deux.

découvert chez un organisme unicellulaire de la famille des choanoflagellées (King et al., 2003).

Les PRK ont été associés à plusieurs aspects du développement de la plante: croissance du tube pollinique (Wengier et al., 2003), incompatibilité sporophytique (Takasaki et al., 2000), différentiation du méristème (Clark et al., 1993a), relations symbiotiques (Endre et al., 2002), perception de régulateur de croissance (Li et Chory, 1997; Matsubayashi et al., 2002), détection d'agents phytopathogènes (Song et al., 1995; Asai et al., 2002), et croissance de l'albumen (Becraft et al., 1996) (voir table 1 pour une vue plus complète). Intuitivement on pourrait s'attendre à ce que chaque récepteur soit associé à un ligand mais il semble qu'un même récepteur puisse être associé à deux ligands, c'est le cas de BRI1 (Szekeres, 2003). Initialement identifié comme percepteur des brassinostéroides, ce récepteur peut aussi lier chez les solanacées la systémine, la première hormone peptidique identifiée chez les végétaux (Pearce et al., 1991).

Tableau 1. Fonctions biologiques des différents récepteurs kinase de plantes. Modifié de Haffani et al. 2004.

Nom du gène	Fonction biologique	Référence(s)
Crinlky4-like		
Crinkly4	Différentiation cellulaire.	Becraft et al. 1996
LRR-PRK		
CLV1	Maintenance du méristème apical.	Clark et al. 1993, 1997
ERECTA	Initiation et élongation d'organe.	Torii et al. 1996
EXS	Développement des anthères et de l'embryon.	Canales et al. 2002
EMS1	Microsporogenèse et développement des cellules du tapetum.	Zhao et al. 2002
HAESA	Abscission des organes floraux.	Jinn et al. 2000
PRK1	Développement du pollen et du sac embryonnaire	Mu et al. 1994(Lee et al., 1997)
VH1	Développement vasculaire.	Clay et Nelson 2002
BAK1	Signalisation des brassinostéroïdes.	Li et al.2002 ; Nam et Li 2002
BRI1	Signallisation des brassinostéroïdes.	Li and Chory 1997
PSKR	Perception de la phytosulfokine.	Matsubayashi et al. 2002
Xa21	Résistance race-spécifique à la pourriture bactérienne du riz.	Song et al. 1995
FLS2	Perception de la flagelline dans la réponse immunitaire.	Gomez-Gomez and Boller 2000
HAR1/NARK	Symbiose rhizobienne et contrôle de la prolifération des nodules.	Krusell et al. 2002, Nishimura et. 2002,
SYMRK/NORK	Symbiose rhizobienne et contrôle de la prolifération des nodules.	Endre et al. 2002, Stracke et al. 2002
LRK10-LIKE		
LRK10	Résistance à la rouille du blé.	Feuillet et al. 1997
LysM		
NFR1/LYK3, NFR5	Résistance à la rouille du blé.	Limpens et al 2003 ; Madsen et al. 2003 ; Radutoiu et al. 2003
S-domain		
SRK	Déterminant femelle de l'auto-incompatibilité chez Brassica.	Takasaki et al. 2000 ; Silva et al. 2001
WAK-like		
WAKs	Expansion cellulaire.	Lally et al. 2001 ; Anderson et al. 2001

Structure des PRK

Le domaine extracellulaire

Les PRK ont une structure très caractéristique et facilement reconnaissable (figure 1A). Leur portion extracellulaire est très variable, ce qui leurs confèrent leurs spécificités. Bien qu'il existe près d'une quinzaine de types de domaines extracellulaires chez les plantes (lectine, extensine, EGF, TNFR, LysM, S-*domain*, thaumatine, WAK-*like*, etc) près de 50% des récepteurs ont des domaines extracellulaires riches en leucine (LRR) (Shiu et Bleecker, 2001b). Chacune des unités de ces répétition (LRR) est composée de la séquence consensus suivante: LxxLxxLxLxxNxLt/sgxIpxxLGx (Dievart et Clark, 2003), et celle-ci peut être répétée jusqu'à 28 fois dans certains cas (Aderem et Ulevitch, 2000). La présence de paires de cystéines, une en N-terminal et une en C-terminal des LRR est aussi fréquemment rencontrée. La paire en N-terminal est située a une soixantaine de résidus du codon de départ et est souvent située dans un contexte très conservé (Cx_nWxGVt/sC) alors que la paire C-terminal est située entre 20 et 30 résidus du domaine transmembranaire, est peu conservée voire parfois même absente (Dievart et Clark, 2003). Ces paires de cystéines seraient impliquées dans la formation de ponts disulfures intra- ou intermoléculaires. Par exemple, l'hétérodimère *CLAVATA1/CLAVATA2* (Trotochaud et al., 1999) est stabilisé par un pont disulfure ce qui implique que l'assemblage de l'hétérodimère aurait lieu dans le RE. Parmi les autres domaines extracellulaires aussi présents, notons le S-domain, retrouvé au locus S et conférant la reconnaissance du *self-pollen*; les motifs EGF (epidermal growth factor) et TNFR (tumor necrosis factor receptor) retrouvés chez les mammifères et les *wall-associated kinase* (WAK); et des récepteurs à domaine thaumatine, lectine et riche en proline, auxquels aucune fonction n'a été associé pour l'instant.

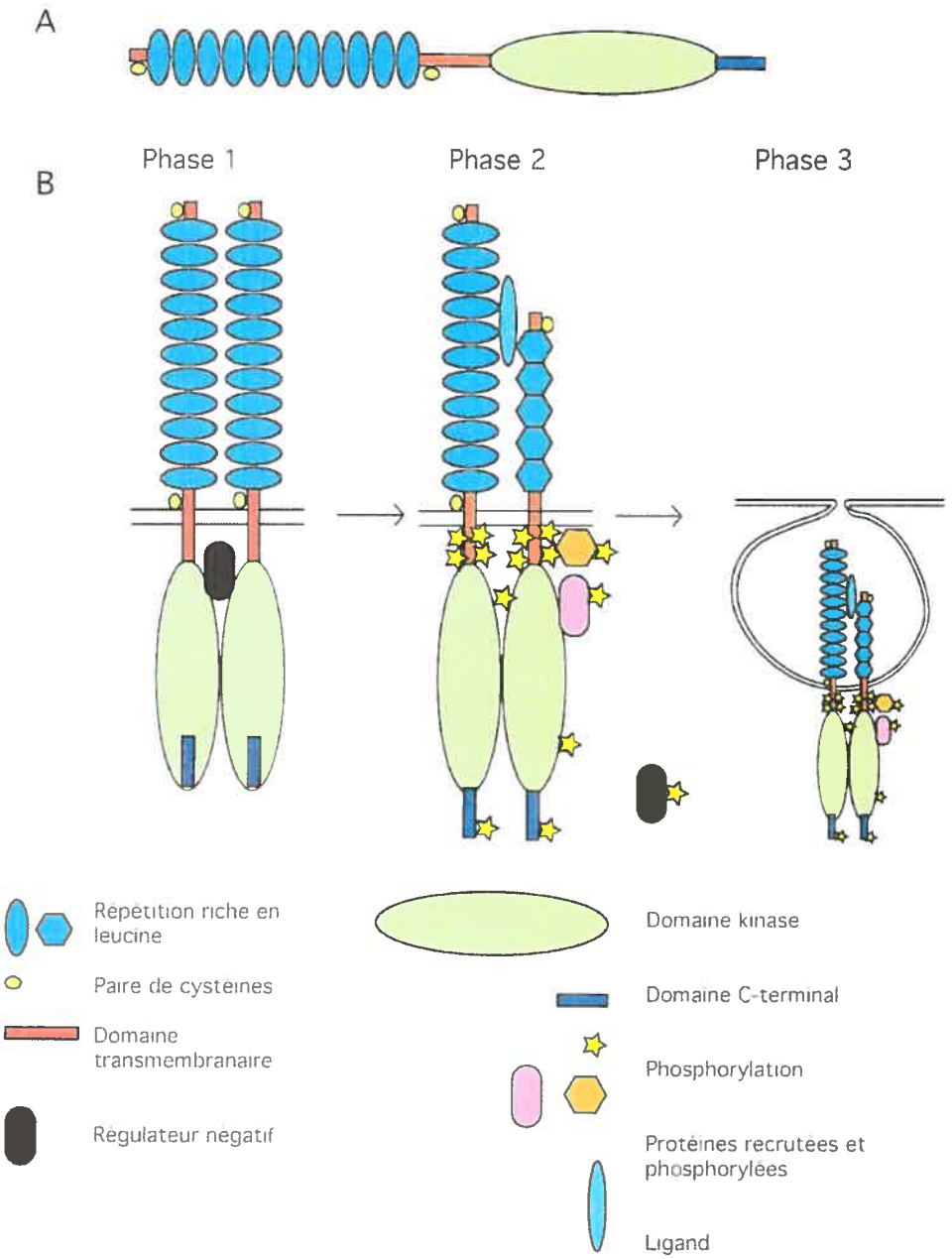
Chez les animaux, les domaines extracellulaires des récepteurs kinases sont bien différents de ceux des plantes. Alors que plus de 50% des domaines extracellulaires de plantes sont constitués de domaines LRR, seulement 3 récepteurs humains contiennent des

domaines LRR (*Neurotrophic tyrosine kinase receptor A* 1, 2 et 3, Human Protein Reference Database, www.hprd.org). De plus, ce domaine qui est présent de façon répétée chez les plantes est habituellement solitaire chez l'humain. À titre d'exemple, le *Neurotrophic tyrosine kinase A* contient une seule répétition de LRR, alors que CLAVATA1, un récepteur de plante, compte vingt-et-une répétitions (Clark et al., 1993a). Les motifs les plus observés dans les domaines extracellulaires chez l'humain sont les domaines Ig, IgC2, et Ig-*like* (immunoglobuline) et le domaine FN3 (fibronectine type III), il s'agit de domaines absents chez les PRK. Bref, les domaines extracellulaires retrouvés dans le règne animal sont extrêmement différents de ceux retrouvés chez les végétaux, ce qui implique, forcément, une différence structurale importante au niveau des ligands reconnus par les PRK.

Tableau 2. Tableau comparatif des domaines extracellulaires retrouvés chez la plante modèle *Arabidopsis thaliana* et chez l'humain. Source Human Protein Reference Database, www.hprd.org.

Arabidopsis thaliana	Humain
C-type lectine (peuvent lier des sucres en présence de Ca^{2+})	LDLA (*low density lipoprotein A*)
Crinkly4-*like*	MAM (*meprin antigen domain*)
CrRLK1-*like*	Ig, IG-like, IgC2 (immunoglobuline)
DUF26 (domaine de fonction inconnu)	FN3 (fibronectine type III)
Extensine (protéine de la paroi, fortement glycosylée, impliqué dans la rigidité de la paroi.	F58C (domaine des facteurs de coagulation V and VIII)
Légume lectine (possède 2 à 4 domaines pour lier un carbohydrate)	REC (domaine contenant un site phosphoaccepteur cible d'une histidine kinase)
LRK10-*like*	FU (*Furin-like cystein-rich*)
LysM (motif riche en lysine, approx. 40 a.a. de long)	SEMA (domaine retrouvé chez les sémaphorines)
URK I	PSI (domaine retrouvé chez les plexines)
PERK-*like*	IPT (domaine à repliement similaire à une immunoglobuline)
RKF3-*like*	FZ (*Frizzled cystein-rich domain*)
S-Domain (domaine conférence la spécificité à l'haplotype lors de la réponse d'autoimcompatibilité).	WIF (domaine extracellulaire chez les Ryk récepteur tyrosine kinases)
Thaumatine (souvent trouvé dans les gènes de réponses aux pathogènes)	Kringle
WAK-*like* (*wall associated kinase*)	EGF (motif répété riche en cystéine)
TNFR (*tumor necrosis factor*)	TNFR (*tumor necrosis factor*)
LRR	LRR (*leucine rich-repeat*)

Figure 1. A) Structure d'un récepteur kinase de type riche en leucine B) Mode d'action des récepteurs kinases



Phase 1. Le récepteur est sous-forme homodimérique non-phosphorylé en absence de ligand. Le complexe est maintenu inactif par une protéine intracellulaire et le domaine C-terminale.

Phase 2. En présence du ligand il y a auto/transphosphorylation du domaine intracellulaire, phosphorylation de régulateurs négatifs et phosphorylation de substrats.

Phase 3. Le récepteur est ensuite inactivé, soit par une déphosphorylation (activité phosphatase) ou alors une internalisation dans une vésicule.

Le domaine intracellulaire

Le domaine intracellulaire peut lui-même être scindé en trois sections ayant chacune un rôle important : le domaine juxtamembranaire, le domaine kinase et le domaine C-terminal. Le domaine juxtamembranaire est situé entre le domaine transmembranaire et le domaine kinase. Souvent négligé, il a cependant des fonctions très importantes. Tout d'abord, le domaine juxtamembranaire (JM) est le site d'une bonne partie des événements de phosphorylation ayant lieu dans le domaine intracellulaire (Nuhse et al., 2004). Les premiers travaux ayant permis de mettre en valeur le rôle du domaine juxtamembranaire ont été réalisés dans le laboratoire de Pamela Ronald à l'université de Californie. Dans une série d'expériences voulant démontrer le rôle du domaine extracellulaire de BRI1 dans la liaison des brasinostéroïdes, l'équipe de Ronald a produit des récepteurs chimériques possédant différentes combinaisons de domaines entre les récepteurs Xa21 et BRI1. Ces deux récepteurs permettent de percevoir une souche de *Xanthomonas* et de déclencher la réponse de défense dans le cas de Xa21, et de percevoir les brassinostéroides dans le cas de BRI1. Lors d'une fusion entre le domaine extracellulaire de BRI1 (perception du brassinolide) et les domaines transmembranaires et intracellulaires de Xa21, il est possible d'activer les mécanismes de défense en présence de brassinolide, indiquant que la chimère est fonctionnelle. Cependant, si le domaine JM de BRI1 est utilisé, en lieu et place du domaine JM de Xa21, aucune activation des gènes de défense est détectée, indiquant que le domaine JM de Xa21 est nécessaire pour que la signalisation en aval de la perception puisse s'effectuer et que le domaine kinase n'est pas suffisant (He et al., 2000). Des travaux impliquant des expériences de phosphorylation *in vitro* ont aussi démontré que le domaine intracellulaire de BRI1 était principalement phosphorylé dans son domaine JM (Oh et al., 2000). De plus, une analyse par spectrométrie de masse a permis d'évaluer que les sites de phosphorylations *in vivo* des récepteurs kinases de plantes sont à plus de 50% situés dans le domaine JM (Nuhse et al., 2004). La présence de sites de phosphorylation dans le domaine

juxtamembranaire causerait un changement de charge et qui permettrait à ce domaine d'agir de site d'ancrage (*docking site*) pour d'autres protéines (Holland et al., 1997; Sachsenmaier et al., 1999). Les protéines qui vont se lier au site d'ancrage peuvent agir comme répresseur ou régulateur négatif (phosphatase) du récepteur ou comme modulateur de l'activité de protéine(s) en aval.

Le domaine kinase et son activité

Le domaine kinase des PRK est évidemment très important grâce à sa fonction catalytique qui consiste à transférer un groupement phosphate d'une molécule d'ATP vers la portion hydroxyle d'une sérine ou d'une thréonine. Il diffère peu des domaines kinases de MAPK ('mitogen activated protein kinase') et compte les douze mêmes sous-domaines. Au niveau phylogénétique, il est cependant possible de différencier les domaines kinases des PRK des domaines kinases des MAPK (Shiu et Bleecker, 2001b). Il est aussi important de noter que certains récepteurs kinases de plante ont un domaine kinase inactif, manquant certains des résidus essentiels, mais sont néanmoins nécessaires pour l'activation de protéines en aval (Llompart et al., 2003). Le domaine kinase peut lui-même être le site de phosphorylation(s) quoiqu'en nombre celles-ci sont plus abondantes dans les domaines JM et C-terminal (Oh et al., 2000). Il faut noter que peu d'études se sont attardées à l'activité kinase des PRKs. Parmi les études qui s'y sont attardées, celles de Schaller et Bleecker (1993) ont démontrée que l'activité kinase de fraction microsomale était plus importante en présence de manganèse qu'en présence de magnésium indiquant, selon eux, une préférence générale des PRK pour le manganèse (Schaller et Bleecker, 1993). Bien que les essais kinases soient généralement faits avec les deux ions Mn^{2+} et Mg^{2+} (Friedrichsen et al., 2000; He et al., 2000; Jin et al., 2000), l'activité kinase de certains récepteurs a été observé en présence de Mg^{2+} seul (Shah et al., 2001; Llompart et al., 2003) ou en présence de Mn^{2+} seul (Charpenteau et al., 2004).

Un point qui a reçu peu d'attention jusqu'à présent est le rôle du domaine juxtamembranaire dans l'activité kinase des PRKs. Les travaux de Pamela Ronald mentionnés ci-dessus indiquaient que le domaine kinase de Xa21 ne pouvait signaler lorsque fusionné avec le domaine juxtamembranaire de BRI1 (He et al., 2000). Cependant les expérimentateurs n'ont pas porté une attention particulière à la présence ou à l'intégrité du domaine juxtamembranaire lors de leurs constructions pour les essais kinase. Les essais kinase réalisés avec CLV1, BRI1, Xa21, et CR4 ont fonctionné et contenaient le domaine juxtamembranaire (Williams et al., 1997; Friedrichsen et al., 2000; He et al., 2000; Jin et al., 2000). Les essais faits avec le récepteur MARK se sont révélés infructueux, alors que celui-ci était amputé de la majeure partie de son domaine juxtamembranaire, et a été déclaré non-fonctionnel (Llompart et al., 2003).

Sur la base de caractéristiques purement associées à la séquence primaire des protéines kinases, celles-ci peuvent être divisé en trois sous-famille. RD, non-RD ou ACF (Johnson et al., 1996b; Dardick et Ronald, 2006). Les RD kinases sont régulées par une phosphorylation dans la boucle activation et ont un résidu arginine (R) précédant le résidu invariable aspartate dans le sous-domaine VI. De façon générale, les RD-kinases sont beaucoup plus actives que les non-RD kinases ou les ACF (Johnson et al., 1996b), quoique des exemples d'autophosphorylation fortes aient été démontrés pour des non-RD (Oh et al., 2000) et des ACF (Mu et al., 1994; Li et al., 2005). Les RD-kinases sont aussi beaucoup plus nombreuses et représentent 70% du kinome d'*Arabidopsis* alors que les non-RD kinases représentent seulement 10% du kinome d'*Arabidopsis* (Dardick et Ronald, 2006). Les non-RD kinases, qui comprennent chez les plantes Xa21 et FLS2, et chez l'humain IRAK, ont été associées à la perception des pathogènes (Dardick and Ronald, 2006). La troisième classe de protéines kinases, les ACF, pour *alternative catalytic function*, comprend 20% du kinome d'*Arabidopsis*. Il manque à ces protéines un ou plusieurs de trois résidus conservé (K/D/D) jugés importants pour l'activité catalytique de ces protéines (Dardick et Ronald, 2006).

Le domaine C-terminal

De longueur très variable, le domaine C-terminal est aussi (comme le domaine JM) souvent sous-estimé, et est parfois considéré à tort comme une série d'acide aminé sans fonction entre le domaine kinase et le codon stop. Bien que la fonction du domaine ou de la région C-terminal n'ait été étudiée que chez un très petit nombre de PRK nous savons maintenant qu'il joue un rôle régulateur de l'activité du récepteur. En effet, chez le récepteur BRI1, une délétion du domaine C-terminal cause une activation du récepteur et de l'aval de la voie (Wang et al., 2005a) indiquant que celui-ci agit comme régulateur négatif de l'activité de BRI1. La présence de phosphorylation *in vivo* a aussi été démontrée dans ce domaine (Nuhse et al., 2004), ce qui pourrait permettre moduler la fonction de ce domaine, ou de modifier son interaction avec le domain kinase, dans le cas d'une autoinhibition.

Mode d'action des PRK

Le mode d'action des PRKs chez les végétaux n'a été élucidé que pour un petit nombre de récepteurs, les mieux étudiés étant CLV1, BRI1, SERK et SRK impliqués respectivement dans la maintenance des méristèmes, dans la perception des brassinostéroïde, dans la différentiation cellulaire et dans la réponse d'autoincompatibilité. Le système le mieux caractérisé actuellement chez les végétaux, celui de la perception des brassinostéroïde, sera utilisé comme archétype du fonctionnement des PRKs. Non seulement il est bien caractérisé au niveau de la perception, mais il s'agit d'un des seuls PRK dont la voie de signalisation en aval est bien caractérisée.

Le PRK *BRI1* (*brassinosteroid insensitive 1*) a été découvert par mutagénèse à l'éthylméthanesulfonate (Clouse et al., 1996) et confère un phénotype de nanisme irréversible même suite à l'application exogène de brassinolide. Ce dernier résultat

confirme qu'il s'agit bien d'un problème de perception et non de biosynthèse du ligand. L'interaction entre le brassinolide et le récepteur a d'abord été démontrée à l'aide des constructions chimériques BRI1/Xa21 mentionnées précédemment et dans lesquelles l'application de brassinolide permettait d'activer les gènes de défense lorsque l'on fusionnait le domaine extracellulaire de BRI1 avec le domaine intracellulaire de Xa21. Des évidences d'interactions directes ont par la suite été démontrés en utilisant du ^{3}H-brassinolide, ces expériences ont aussi montré qu'un îlot situé entre les LRR 21 et 22 est nécessaire à l'interaction (Kinoshita et al., 2005).

Par la suite, il a été démontré que BRI1 perçoit le brassinolide en hétérodimère avec BAK1 (BRI1 *associated kinase*) un autre LRR-PRK et qu'ils peuvent s'auto/transphoryler lorsqu'ils sont co-exprimés dans un système bactérien (Li et al., 2002; Nam et Li, 2002). Des expériences d'interaction protéine-protéine entre BRI1 et BAK1 par FRET (*fluorescence resonance energie transfer*) ont démontré l'existence de l'homodimère BRI1 et de la formation de l'hétérodimère avec BAK1 juste avant une internalisation par endocytose (Russinova et al., 2004). Cette dernière expérience suggère que BRI1 existe sous forme homodimérique au repos (en absence de brassinostéroïde) et qu'il forme un hétérodimère avec BAK1 en présence de brassinostéroïde, puis serait rapidement inactivé par endocytose (pour une revue récente voir (Li et Jin, 2007)).

Régulation de l'activité de BRI1

À ce jour, quatre protéines ont démontré une interaction avec le domaine intracellulaire de BRI1 : KAPP (*kinase associated protein phosphatase*) (Friedrichsen et al., 2000), TTL (*transthyrétine-like*) (Nam et Li, 2004), BKI1 (Wang et Chory, 2006) et TRIP1 (TGFβ-RECEPTOR-INTERACTING PROTEIN 1).

KAPP est une protéine phosphatase qui interagit avec plusieurs récepteurs kinases, dont CLAVATA1 (Williams et al., 1997), FLS2 (Gomez-Gomez et al., 2001), SERK1 (Shah et al., 2002), WAK1 (Park et al., 2001), BoSRK (Vanoosthuyse et al., 2003), RLK5

(van der Knaap et al., 1999), OsTMK (Braun et al., 1997), ZmKIK1 (Braun et al., 1997) et RLK4 (Braun et al., 1997), soit une bonne partie des récepteurs de plantes connus chez les plantes. Les travaux sur KAPP, particulièrement chez CLAVATA et SERK, ont démontré que KAPP est phosphorylé par CLV1 et, en retour, elle se lie à CLV1 et le déphosphoryle (Williams et al., 1997). Les travaux sur SERK ont démontré que KAPP peut lier SERK si celui-ci est phosphorylé, que les deux protéines se retrouvent à la membrane plasmique, mais que l'interaction physique entre les deux (observée par FRET) n'a lieu que dans des vésicules intracellulaires (Shah et al., 2002). Ces résultats suggèrent que KAPP lie le récepteur lorsque celui-ci est phosphorylé, inactive le récepteur en le déphosphorylant et provoque son internalisation (figure 1B).

Pour ce qui est de TTL, moins de données ont été recueillies. On sait que TTL est localisé à la membrane plasmique, qu'il peut être phosphorylé par BRI1 *in vitro* et que l'interaction est dépendante de la phosphorylation de BRI1. Par contre les données génétiques recueillies ont démontré qu'il s'agit d'un régulateur négatif (comme KAPP) puisque la surexpression de *TTL* entraîne en phénotype similaire aux mutants faibles *BRI1* ou aux mutants nuls *BAK1* (Nam et Li, 2004).

Les données sur *BKI1* sont semblables à celles recueillies pour *TTL*. BKI1 a été isolé de façon récurrente lors de criblages de double hybrides chez la levure. Une série de fusion avec la protéine YFP (*yellow fluorecent protein)* a confirmé que le domaine C-terminal de BKI1 est nécessaire et suffisant à l'interaction avec le domaine kinase de BRI1. Tout comme pour TTL, les données génétiques ont démontré qu'il s'agit d'un régulateur négatif, la surexpression de *BKI1* entraînant un phénotype similaire aux mutants faible *BRI1* ou aux mutants nuls *BAK1*. De plus, en absence de brassinolide, BKI1 est localisé à la membrane plasmique, cependant, immédiatement après l'addition de brassinolide, BKI1 se dissocie de la membrane plasmique (Wang et Chory, 2006). Finalement BKI1 semble spécifique à BRI1 car des essais de double hybrides indiquent qu'il ne peut interagir avec d'autre récepteur phylogénétiquement proche de BRI1 (BRI1-like) ou avec d'autre

récepteurs plus éloignés (CLV1). Il a été postulé que BKI1 est associé avec BRI1 en absence de brassinolide pour empêcher l'interaction BRI1/BAK1 (Wang et Chory, 2006).

Dans le cas de TRIP1, nous savons que les plantes exprimants un ARNm anti-sens de TRIP1 présentent des caractéristiques morphologiques ressemblant aux mutants BRI1. De plus, nous savons que le domaine kinase de BRI1 peut phosphorylé TRIP1, nous savons aussi que BRI1 et TRIP1 interagissent *in vivo* car ils peuvent être co-immunoprécipité à partir d'un extrait protéique de plante (Ehsan et al., 2005).

Bien qu'il s'agisse (selon moi) du récepteur kinase de plante le mieux caractérisé, je vais compléter l'état des connaissances sur ce sujet avec des données issues de l'étude de d'autres récepteurs kinases (CLAVATA1 et SRK) afin de combler les lacunes laisser par les travaux sur BRI1. CLAVATA1, un récepteur kinase de type LRR, a été l'un des premier PRK découvert chez les plantes, et, est impliqué dans la différentiation des cellules des méristèmes (Clark et al., 1993a). Les mutants insertionnels *clavata1* présentent des méristèmes plus gros que les méristèmes de type sauvage et développent un trop grand nombre d'organes (Clark et al., 1993a). *In vivo,* CLAVATA1, dont le poids moléculaire est de ~120 kDa, se retrouve soit dans un complexe protéique de 185 kDa ou dans un complexe de 450 kDa et il est lié à ce complexe par des ponts disulfures (Trotochaud et al., 1999). Ce complexe contient la protéine KAPP mais aussi une protéine ROP GTPase (*Rho of plants*) (Trotochaud et al., 1999). Le même groupe de recherche a plus tard démontré que CLAVATA2, faisait aussi probablement partie de ce complexe car il est nécessaire à la liaison de CLAVATA3, le ligand (Kayes et Clark, 1998). CLAVATA1 agit donc de façon hétérodimérique avec un autre récepteur (comme BRI1), à la différence près que CLAVATA2 n'est pas un récepteur kinase mais bien une *receptor-like protein*, c'est-à-dire que sa structure est en tout point semblable à CLAVATA1 ou BRI1 mais il ne possède pas de domaine intracellulaire. Donc CLAVATA1 forme un complexe multimérique incluant CLAVATA2, CLAVATA3, ROP, KAPP et possiblement d'autres protéines car la somme de la taille de ces protéines ne totalise pas 450 kDa. Il s'agissait de la première

démonstration qu'un PRK peut former un hétérodimère avec une *receptor-like protein,* ce qui n'a toujours pas été démontré, à ce jour, pour d'autre PRK. Il existe plus de 178 *receptor-like protein* chez *Arabidopsis* (Shiu et Bleecker, 2003). *CLAVATA2* était le premier à se voir attribuer une fonction, puisque les mutants *clavata2* démontre le même phénotype que les mutants *clv1* (Kayes et Clark, 1998).

Le récepteur SRK (S-locus receptor kinase) est responsable de la perception du pollen du soi et déclenche la réaction d'auto-incompatibilité chez *Brassica* en présence de pollen du soi. Le récepteur SRK est situé dans le stigmate et lie le déterminant mâle (SCR) qui se trouve à la surface du grain de pollen. Lorsque le pollen est issu d'un génotype compatible (donc différent) la réaction d'incompatiblité n'est pas enclenchée, lorsque le pollen est issu d'une plante portant les mêmes allèles d'incompatibilité la réaction est enclenchée. Les travaux de Giranton et al. (2000) ont démontré que lorsque SRK est produit en système de cellule d'insecte/baculovirus, et partiellement purifiée en microsome, la protéine est constitutivement active (Giranton et al., 2000). Ces résultats paraissent incompatibles avec des résultats publiés plus tard, indiquant que la protéine est non phosphorylée *in vivo* lors d'une pollinisation compatible et devient phosphorylée environ 60 minutes suite à une pollinisation incompatible (Cabrillac et al., 2001). Par ailleurs, Cabrillac et al. (2001) ont remarqué que l'addition d'extraits de stigmates entraînait une déphosphorylation du récepteur. Ils ont soupçonné l'agent responsable de cette déphosphorylation d'être soit la protéine *thioredoxin-h-like* THL-1 ou la THL2. Ces deux protéines avaient préalablement été identifiées comme interactants en double hybride dans la levure et l'interaction a par la suite été confirmée par co-immunoprécipitation (Bower et al., 1996). De plus, il avait été démontré que THL1 pouvait être phosphorylé par SRK et qu'une thioredoxin de mammifères pouvait réguler l'activité de *apoptosis signal-regulating kinase 1* (ASK1) et de la MAP kinase p38 (Saitoh et al., 1998; Hashimoto et al., 1999). De plus, lorsque Cabrillac et al.(2001) ont ajouté la protéine recombinante THL1 exprimée en bactérie à un essai kinase *in vitro* avec SRK, SRK ne se retrouvait plus sous sa forme phosphorylé, par contre des versions mutées de THL1 ne pouvaient empêcher la

phosphorylation de SRK. L'addition d'un compétiteur, c'est-à-dire une version non-fonctionelle de SRK en excès dans l'essai (avec SRK et THL) a permis de rétablir la phosphorylation (possiblement en déplacant THL1) indiquant qu'il s'agit d'une interaction directe servant à prévenir un activation spontanée de la voie d'activation de la réponse d'autoi-incompatibilité (Cabrillac et al., 2001).

Donc à partir des données recueillies chez le BRI1, SRK et CLAVATA1 nous pouvons comprendre que les récepteurs interagissent avec d'autres protéines transmembranaires qui sont soit des récepteurs kinases (BRI1/BAK1) ou des récepteurs sans domaine kinase (CLV1/CLV2). Lorsque inactifs, ils interagissent aussi avec des protéines intracellulaires dont la fonction est de les maintenir inactifs et d'empêcher une activation spontanée du domaine kinase (BKI1, THL). La liaison avec un ligand semble entraîner (dans un ordre plus ou moins certains) la phosphorylation de substrats (BKI1, THL) et l'auto/transphosphorylation du domaine kinase. Par la suite, des protéines comme ROP (une Rho GTPase) et KAPP (une protéine phosphatase) pourrait servir à moduler le signal. La liaison avec KAPP désactive le récepteur et provoque son internalisation par endocytose, possiblement pour retirer le récepteur lié à son ligand de la surface cellulaire. Bien que nous présumions que le récepteur soit ensuite dissocié de son ligand dans une vésicule endosomale et soit ensuite recyclé vers la surface (si nécessaire), comme c'est le cas chez les animaux (Dautry-Varsat et al., 1983), aucune observation en ce sens n'a à ce jour été rapporté chez les plantes.

Mode d'action des récepteurs chez les animaux

Les récepteurs animaux les plus étudiés, et de loin, sont les récepteurs de la famille ERBB dont le membre le plus connu est le récepteur du facteur de croissance épidermale (EGFR). J'utiliserai donc ce groupe de récepteur, un récepteur tyrosine kinase, et des

exemples du récepteur du facteur de transformation (TGFR), un récepteur sérine/thréonine kinase comme exemples des mécanismes de fonctionnement des récepteurs animaux.

Le cas de la famille ERBB

La famille ERBB est composée de quatre membres, soit ERBB1 (EGFR), ERBB2 (HER2), ERBB3 (HER3) et ERBB4 (HER4). Les quatre récepteurs ont un domaine extracellulaire très semblable contenant un motif EGF. Cependant au niveau du domaine kinase seul ERBB1 et ERBB4 sont pleinement fonctionnel. Bien qu'ERBB2 et ERBB3 n'aient pas d'activité kinase intrinsèque, ils permettent d'amplifier le signal d'ERBB1 et 4 lorsqu'ils forment des hétérodimères. ERBB1 peut lier différents ligands seul ou en formant différentes combinaisons d'hétérodimères, par exemple ERBB1 peut lier l'EGF, le facteur de croissance EGF-like, l'épiréguline, l'amphireguline, TGFα, et la betacelluline (Warren et Landgraf, 2006). Il est intéressant de noter que si les résidus qui sont les cibles d'autophophorylation dans la portion intracellulaire de ERBB1 sont mutés, ERBB1 ne perd aucune activité de phosphorylation de ses substrats. Ce dernier résultat indique qu'ERBB1 est constitutivement actif (Honegger et al., 1988). Au repos, ERBB1 est majoritairement sous forme monomérique (Burgess et al., 2003), la présence de ligand favorise la dimérisation des monomères et stabilise les dimères (instables déjà présents). Suite à la dimérisation, le domaine C-terminal sera abondamment phosphorylé et agira comme appât pour le recrutement de protéines substrats (Yarden et Schlessinger, 1987; Honegger et al., 1990). À titre indicatif, 17 des 23 sites de phosphorylation démontrés chez EGFR sont dans le domaine C-terminal, 4 sont dans le domaine kinase et 2 sont dans le domaine juxtamembranaire (Human Protein Reference Database, www.hprd.org). Donc, la phosphorylation non désirée de protéines substrats par le domaine kinase constitutivement actif de ERBB1 pourrait être empêché par l'absence physique des protéines substrats qui sont recrutées seulement lors de la phosphorylation du domaine C-terminal. La dimérisation d'ERBB1, c'est-à-dire, l'interaction physique entre deux molécules ERBB1, bien qu'elle se fasse suite à la liaison d'un ligand n'est pas effectuée par le ligand. Deux ligands lient

chacun un monomère d'ERBB1, à l'extérieur de la protéine et la dimérisation est maintenu par une boucle de dimérisation situé dans le côté intérieur de la protéine dans le domaine extracellulaire. Des mutations dans cette boucle empêchent la dimérisation même en présence du ligand (Burgess et al., 2003). Suite à la signalisation le récepteur sera désactivé par déphosphorylation ou par internalisation dans une vésicule endosomale (Opresko et al., 1995) ou vers des endosomes de recyclage pour ensuite être renvoyé vers la surface (Wilde et al., 1999). Il est important de noter, bien que cela soit souvent oublié, que EGFR (comme possiblement bien d'autres récepteurs) peut continuer à signaler lorsqu'il se trouve dans les endosomes (Wang et al., 2002). La signalisation à partir des endosomes est cependant différentes au niveau qualitatif car l'activation de voie de MAPK ne peut se faire que s'il y a endocytose alors que l'activation de la phospholipase C ne nécessite pas l'endocytose et peut se faire lorsque le récepteur est maintenu au niveau de la membrane plasmique (Teis et al., 2002).

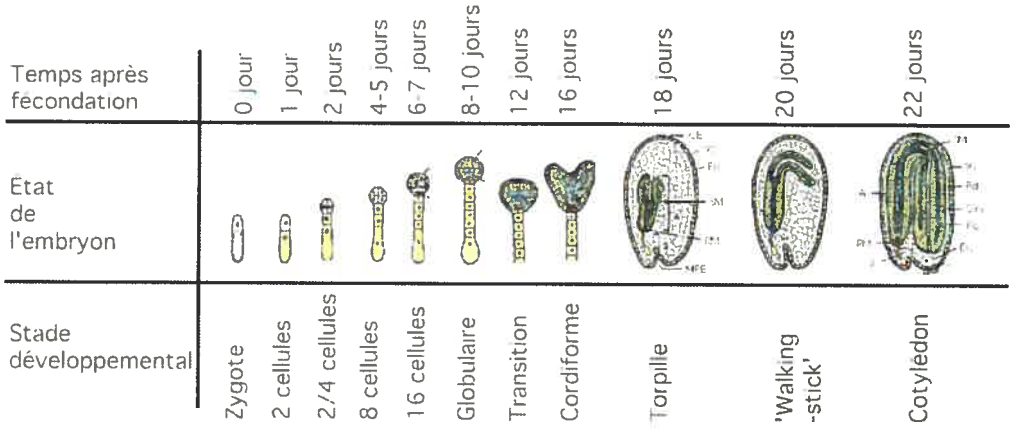
Le mécanisme de la signalisation par le récepteur de TGFβ est similaire, mais présente néanmoins des différences importantes avec les ERBB. Au repos, le récepteur kinase de TGFβ, qui est un récepteur sérine kinase, donc structurellement plus près des récepteurs de plantes que les ERBB, et existe sous forme homodimérique (Derynck et Zhang, 2003). La liaison du ligand induit la formation d'un complexe protéique bi-dimérique (formé de deux homodimères). Le domaine juxtamembranaire qui sert de site d'ancrage à la protéine inhibitrice FKBP12 sera phosphorylé en deux endroits suite à la liaison du ligand et deviendra un site d'ancrage pour les protéines substrats des facteurs de transcriptions de la famille SMAD (Huse et al., 2001). La phosphorylation des facteurs de transcriptions de type SMAD les relocalisera vers le noyau. Les exemples discutés précédemment suggèrent que le mode d'action des RK et PRK est très similaires, mais que les partenaires, de même que les voies d'activation en aval sont très différentes.

Anatomie et développpement de la graine.

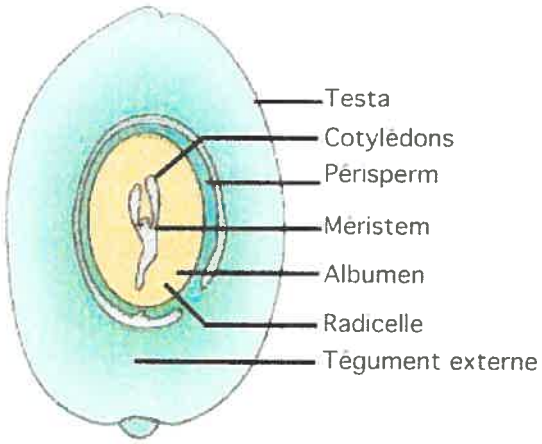
Puisque la présente thèse porte sur la caractérisation de PRK exprimés lors du développement de l'ovaire, les structures de la graine et les étapes du développement de l'embryon seront succinctement présentés afin de pouvoir bien comprendre les chapitres 3, 4, 5. La figure 2 présente en A les différents stades du développement embryonnaire et les jours suivant la fécondation correspondant aux différents stades. Pour les RT-PCR du chapitre trois, nous avons utilisé des fruits de 0 à 12 jours post-pollinisation (différent de post-fécondation). Or, comme la fécondation se produit à approximativement 2 jours après la pollinisation, tous les stades doivent être décalés de 2 jours. Donc, dans le chapitre trois, les fruits 12 jours sont au stade globulaire, c'est-à-dire 10 jours après la fécondation. La figure 2B présente les différentes structures de la graine et peut être utilisé pour analyser les résultats d'hybridation *in situ* du chapitre 5.

Figure 2. Schéma représentant le développement embryonnaire et les structures de la graine chez *S. chacoense*.

A



B



Chapitre 2

Plant bioactive peptides: an expanding class of signaling molecules

Hugo Germain, Eric Chevalier, and Daniel P. Matton

Publié dans Canadian Journal of Botany **84**(1): 1–19 (2006)

Contribution des co-auteurs

Cette publication est un papier de synthèse de la littérature disponible à ce jour sur les petits peptides de signalisation chez les plantes. Il s'agit d'un papier invité suite à l'obtention de la bourse CD Nelson Award par Daniel P. Matton. J'ai écrit et effectué la revue de la littérature de presque tout le papier sauf les sections sur RALF et HT. Éric Chevalier a rédigé la section sur RALF et Daniel Matton a rédigé la section sur HT.

Page titre

REVIEW / SYNTHÈSE

Plant bioactive peptides: an expanding class of signaling molecules

Hugo Germain, Eric Chevalier, and Daniel P. Matton

Institut de Recherche en Biologie Végétale, Département de sciences biologiques, Université de Montréal, 4101, rue Sherbrooke est, Montréal, QC H1X 2B2, Canada.

Corresponding author: *D.P. Matton* [redacted]

Received 28 September 2005. Published on the NRC Research Press Web site at http://canjbot.nrc.ca on 8 March 2006.

Source de financement

This work was supported by NSERC, FQRNT and by the Canada Research Chair program.

Key words

Signaling peptide, ligand, receptor, cell–cell communication.

Abstract

Until recently, our knowledge of intercellular signaling in plants was limited to the so-called five classical plant hormones: auxin, cytokinin, gibberellin, ethylene, and abscissic acid. Other chemical compounds like sterols and lipids have also been recognized as signaling molecules in plants, but it was only recently discovered that peptides in plants, as in animal cells, play crucial roles in various aspects of growth and development, biotic and abiotic stress responses, and self/non-self recognition in sporophytic self-incompatibility. These peptides are often part of a very large gene family whose members show diverse, sometime overlapping spatial and temporal expression patterns, allowing them to regulate different aspects of plant growth and development. Only a handful of peptides have been linked to a bona fide receptor, thereby activating a cascade of events. Since these peptides have been thoroughly reviewed in the past few years, this review will focus on the small putative plant signaling peptides, some often disregarded in the plant peptide literature, which have been shown through biochemical or genetic studies to play important roles in plants.

Résumé

Jusqu'à tout récemment, nos connaissances sur la communication intercellulaire chez les végétaux étaient limitées aux cinq hormones classiques : les auxines, les cytokinines, les gibbérelines, l'éthylène et l'acide abscissique. D'autres composés, tels les stérols et les lipides, sont maintenant aussi reconnus comme molécules de signalisation, mais ce n'est que tout récemment que des peptides, comme chez les cellules animales, ont été découverts chez les plantes. Certains de ces peptides sont impliqués dans la croissance et le développement des plantes, les réponses face aux stress biotiques et abiotiques, de même que dans la reconnaissance du soi et du non-soi, dans les barrières de reproduction comme l'auto-incompatibilité sporophytique. Ces peptides sont souvent codés par de grandes familles de gènes et chaque membre démontre des patrons d'expression variés, parfois spatialement et temporellement distincts, leur permettant de réguler différents aspects de la croissance et du développement des plantes. Jusqu'à présent, des récepteurs n'ont été isolés que pour quelques peptides, et ces voies de signalisation commencent à être bien connues. Ces quelques rares couples récepteur-ligand ayant été commentés plusieurs fois en profondeur depuis quelques années, nous porterons, dans la présente revue, notre attention sur d'autres groupes de peptides, certains largement ignorés dans les revues de littérature récentes, mais dont la caractérisation biochimique, génétique ou phénotypique, laisse entrevoir qu'ils seraient aussi potentiellement impliqués dans la signalisation ou la régulation de processus variés chez les végétaux.

Introduction

The idea that plants, like animal cells, could use peptides as signaling molecules arose 15 years ago with the discovery of systemin, a plant peptide involved in the systemic response against herbivores (Pearce et al. 1991). Other processes, such as the control of shoot apical meristem development and organ initiation, were later shown to be controlled by the CLAVATA3 (CLV3) peptide (Fletcher et al. 1999). Subsequently, genomic data for *Arabidopsis thaliana* revealed that the genome encodes hundreds of protein receptor kinases (Shiu and Bleecker 2001), several of these possibly binding a single or a few related peptides, suggesting that the recently discovered peptides were probably not the exception. The changing picture in plant peptide signaling is due to the emergence of peptide families. CLV3, which has been extensively studied (Fletcher et al. 1999; Trotochaud et al. 1999; Trotochaud et al. 2000; Lenhard and Laux 2003), is now known to be part of one of these families (Cock and McCormick 2001). These discoveries are due in great part to the completion of the *Arabidopsis* genome (The Arabidopsis Genome Initiative 2000) and to the massive number of EST sequences gathered over the last few years in various plant species.

Assigning a ligand to a receptor remains a very difficult task. Even now, putative ligands are found randomly because they are often not annotated using conventional computational algorithms. Ligand-receptor direct interaction remains state-of-the-art biochemistry work: receptors are difficult to express in soluble form and ligands are generally only available in planta at nanomolar to femtomolar concentrations. Furthermore, a synthetic ligand may not bear the proper post-translational modifications or display proper folding, resulting in nonfunctional molecules, and T-DNA insertion lines are often aphenotypic owing to the redundancy found in large family of peptides harboring very conserved functional domains.

Despite all the above-mentioned hurdles, some investigators have succeeded in their quest and a few plant peptide ligands have been assigned to their respective receptors. For examples, the mechanisms of action for CLV3, phytosulfokine, systemin, and the incompatibility determinant SCR/SP11 has been partly unraveled; recently, these receptor–ligand pairs were thoroughly reviewed and so will not be considered here (Becraft 2002; Ryan et al. 2002; Chilley 2003; Haffani et al. 2004; Torii 2004). Other non-peptide molecules such as ethylene, brassinosteroid, cytokinin, auxin, gibberellin, and abscisic acid have also been well characterized and shown to interact with specific receptor proteins (Schaller and Bleecker 1995; He et al. 2000; Inoue et al. 2001; Dharmasiri et al. 2005; Ueguchi-Tanaka et al. 2005; Razem et al. 2006). This review will thus focus on peptide families and (or) putative small signaling peptides for which a receptor or a function has not been assigned yet. Some of these putative peptides have generally been disregarded in recent reviews of the plant signaling peptide literature, although many have been partially characterized at the functional level through biochemical analyses or through the use of mutant plants. Undoubtedly, many more of these putative bioactive peptides are likely to be characterized in the next few years and may well turn out to be involved in important plant processes. In this review, the peptides presented will be grouped into five functional categories that are linked to their putative roles in planta.

Development

The BRICK genes

All multicellular organisms arise from one cell that divides to produce a multicellular organism with several cell types. The process by which the daughter cell acquires a different fate (cell type) is termed asymmetric cell division (Scheres and Benfey 1999). Stomata also arise from asymmetric cell divisions. To generate a normal stomata, a first asymmetric cell division must generate the guard mother cell (GMC), followed by a polarized alignment of the subsidiary mother cell (SMC) with the GMC. The SMC will

eventually divide to produce the subsidiary cell whereas the GMC will become the guard cells (Gallagher and Smith 2000). The *brick1* mutant was identified as a recessive mutation that disrupts subsidiary cell division using a screen for abnormal cell patterning in maize leaf epidermis (Gallagher and Smith 2000). In the *brick1* mutant, nearly 80% of the SMC division does not result in a daughter cell that is a subsidiary cell. The cause for this seems to be a failure of the SMC to polarize. Prior to the SMC division, the nuclei must align with the GMC and actin patches must form. The number of cells undergoing these events in the *brick1* mutant is greatly reduced. Additional phenotypes observed in the *brick1* mutant include the absence of the epidermal cell lobe, as well as shorter and blunter prickle hairs (Frank and Smith 2002). All these anatomical features of leaf epidermis arise following accumulation of F-actin patches that are never observed in *brick1* mutant. Remarkably, the pollen tube, a cell also displaying polarized growth, is not affected by the *brick1* mutation, and the mutant plants look normal, except for their approximately one-third shorter stature compared to that of wild-type plants (Frank et al. 2003). The use of maize EMS-mutagenized lines and complementation analyses revealed two other genes (designated as *brick2* and *brick3*) causing the same phenotype as *brick1* (Frank et al. 2003). All combinations of double mutants phenocopied the single-mutant, indicating that the three genes are involved in the same pathway (Frank et al. 2003). Frank et al. (2003) also established that *BRICK2* and *BRICK3* could act cell autonomously to promote pavement cell lobe formation or non-autonomously to promote polarized SMC division, whereas *BRICK1* could act non-autonomously for both functions.

The maize *BRICK1* gene (the only one yet characterized) encodes a short 252-bp cDNA and the corresponding 8 kDa peptide (Frank and Smith 2002) lacks features such as a signal peptide or the presence of conserved cysteine residues, which are often associated with diffusible peptides. The gene is highly conserved in the plant and animal kingdoms where it is involved in actin polymerization (Eden et al. 2002). It is too early to conclude that the *BRICK* genes are part of the same peptide family; it is possible that they represent different proteins acting in the same signaling pathway. It is nonetheless surprising that the

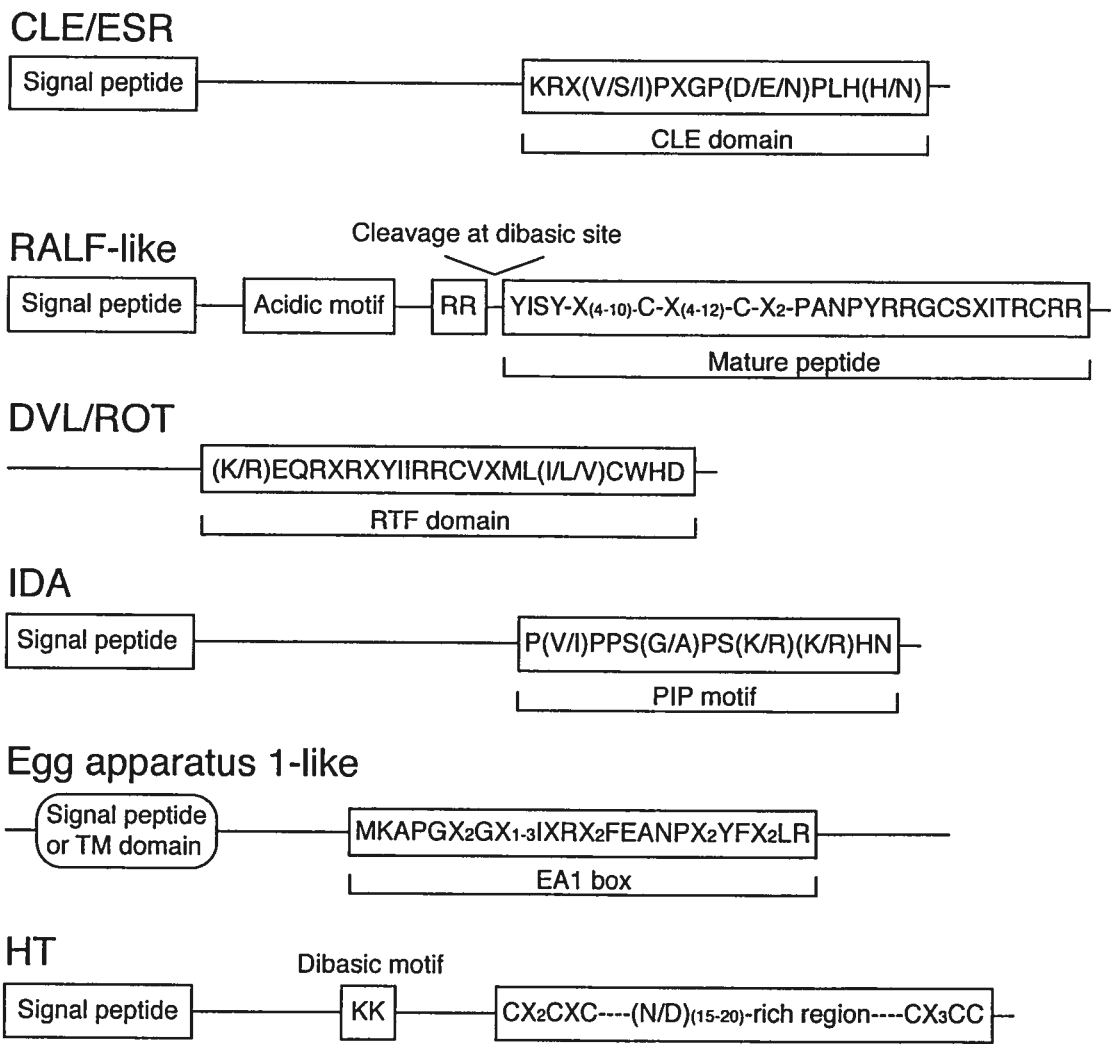
three genes can act in a non cell-autonomous manner, suggesting that their gene product is diffusible. The lack of a signal peptide and the fact that the gene product acts over a very short distance suggest that the peptide(s) can travel through plasmodesmata. *Short-root*, a transcription factor has been shown to travel (when fused to GFP) through the plasmodesmata (Nakajima et al. 2001), and proteins as big as 61 kDa have been shown to use these channels (Oparka et al. 1999). It is therefore possible that the 8 kDa peptide encoded by the *BRICK1* gene either moves through the plasmodesmata or uses a non-conventional/non-classical leaderless secretion pathway as shown for some mammalian proteins (Bendtsen et al. 2004*b*), and acts as paracrine hormones on the surrounding cell to promote proper localization of actin patches in epidermal cells.

CLAVATA3-like / ESR (CLE)

CLAVATA3 (CLV3) is one of the first and best-characterized small signaling peptide in plants. The *clv3* mutant causes the same phenotype as the *clavata1* (*clv1*) mutant (Clark et al. 1995) and encodes a small peptide of 96 amino acids (Fletcher et al. 1999) with an N-terminal signal peptide, producing a mature peptide of 76 amino acids after cleavage (Rojo et al. 2002). Direct interaction with CLV1, although previously published, has now been retracted (Nishihama et al. 2003) (for a review on *CLAVATA*s, see Clark 2001). Although CLV3 was initially described as a protein with no similarities with other proteins from public databases, it was later found to display weak similarities (E-values ranging from 0.15 to 0.45) with the embryo surrounding region proteins (ESR) (Cock and McCormick 2001). *ESR* genes are expressed only in the embryo surrounding region, and three members have been found in the maize genome. All three *ESRs* encode small peptides of 14 kDa, which contain a signal peptide (Opsahl-Ferstad et al. 1997). Cell fractionation, immunocytological results, and transient expression assays all agree that the ESR proteins are secreted proteins that remain mostly in the embryo surrounding region but are also present in the basal endosperm layer (Bonello et al. 2002).

The homology between CLV3 and the ESRs is low when the overall protein is considered but high for a stretch of 14 amino acids located in the C-terminal region. BLAST searches with this highly conserved region retrieved a total of 42 related sequences from both monocots and dicots (Cock and McCormick 2001). Twenty-six are from *Arabidopsis* and half are located on chromosome 1 (Cock and McCormick 2001; Hobe et al. 2003; Sharma et al. 2003). Sequence conservation between all these peptides is very low (except for the C-terminal domain). Nearly all these putative peptides have a signal peptide, are short, are intronless (except for *CLV3* and *CLE40*), and have a very conserved C-terminal domain (Fig. 1). Since most *A. thaliana CLEs* originate from genomic sequences and could be pseudogenes, Sharma et al. (2003) used RT-PCR to determine whether the *A. thaliana CLEs* are expressed. Expression is detectable for all but one (*CLE26*) gene. Transient expression in leek epidermal cells of CLE4, CLE9, and CLE40 fused to GFP revealed that these three CLEs have a functional signal peptide that directs the protein to the cell wall or the extracellular space (Sharma et al. 2003). Single amino acid changes in the CLE motif, as observed in the *clv3-1* and *clv3-5* mutant, disrupts CLV3 function, indicating the importance of the CLE domain for CLV3 activity (Fletcher et al. 1999).

Figure 3. Schematic representation illustrating the conserved domains and motifs found in peptide members of multigenes families.



Of these *CLE* genes, two have been shown to cause a developmental defect when their expression pattern is altered. One of the two is *CLE19*. When *CLE19* expression is driven by the CaMV 35S promoter, the transgenic plants have a short-root phenotype in which the root meristem becomes fully differentiated, pistils are pin-shaped (without carpel and ovules), and flowers have disconnected vascular bundles (Fiers et al. 2004). It should be noted that T-DNA mutants are aphenotypic (Fiers et al. 2004). A suppressor screen of *Brassica napus* for the *CLE19* phenotype revealed two loci, named *SOL1* and *SOL2* (suppressor of LIGAND-LIKE PROTEIN, the closest homolog to *AtCLE19* from *B. napus*), that fully restore the wild-type phenotype (Casamitjana-Martinez et al. 2003). It should be noted that *SOL1* and *SOL2* do not have a phenotype of their own. The cloning of *SOL1* revealed that it encodes a transmembrane Zn^{2+}-carboxypeptidase, which is possibly involved in processing of the peptide. Although processing of CLV3 has not been demonstrated, it has been shown for other putative peptide ligands (Yang et al. 1999; Pearce et al. 2001*b*). The facts that *CLE19* overexpression causes a defect in root meristem maintenance and *SOL1* encodes a processing enzyme suggest that *CLE19* is involved in a CLV-like pathway (Casamitjana-Martinez et al. 2003).

The second *CLE* gene that has a known phenotype is *CLE40*. This gene is the only *CLE* gene that has a similar genomic organization to *CLV3*, both having two introns, whereas all other *CLE* genes lack introns. Overall, the two sequences share very little sequence similarities, but their CLE domain is 70% identical, and both of these CLEs, unlike others, have a short stretch of amino acids C-terminal to the CLE domain (that terminates with a proline). Plants overexpressing *CLE40* have phenotypes similar to the ones observed in 35S::*CLV3* plants (Hobe et al. 2003). Not only are the phenotypes similar, but *CLE40* can also downregulate *WUSCHEL* expression (like *CLV3*) and *CLE40* requires

functional *CLV1* receptor complex for its shoot apical meristem activity, indicating that *CLE40* when expressed at sufficient levels can substitute for *CLV3* in regulating SAM cell differentiation. 35S::*CLV3* and 35S::*CLE40* also have the same effect on root development; both cause initial growth retardation, eventually leading to a total root growth arrest. A T-DNA insertion in the *CLE40* gene causes a root-waving phenotype, indicating that *CLE40* is required for normal root development and that other *CLEs*, albeit similar in their CLE domain, could not substitute for *CLE40*. Application of chemically synthesized partial peptides (p) of CLV3, CLE19 and CLE40, corresponding to their CLE domain, is sufficient to reduce root length (Fiers et al. 2005). *CLV2*, but not *CLV1* or *CLV3*, is also involved in CLV3p, CLE19p, and CLE40p recognition. Synthetic CLE5p is unable to reduce root length, indicating that not all *CLEs* act in this pathway.

Using a motif search Olsen and Skriver (2003) found a *CLE*-like sequence from oesophageal gland cell library of the plant parasitic nematode *Heterodera glycines* (HgCLE) (Gao et al. 2001; Wang et al. 2001). This plant-infecting nematode possesses a hollow protrusible stylet capable of piercing the plant cell wall and excretes products generated by the oesophageal gland into the plant cell, thereby transforming it into a feeding cell via the action of parasitism proteins (Davis et al. 2000). An antibody directed against HgCLE revealed that the protein is present only in secretory granules of the oesophageal gland of the parasitic stage of *H. glycines* dissected from soybean roots, and in secretory granules that migrate into the collecting reservoir at the base of the stylet (Wang et al. 2005). Transgenic plants overexpressing *HgCLE* in wild-type background display floral, SAM, and root phenotypes reminiscent of *CLE19* and *CLE40* misexpression, and also decrease *WUS* expression as does *CLV3*. 35S::*HgCLE* transformed in *clv3-1* (a strong *CLV3* mutant) background is able to partially or fully rescue the phenotype, indicating that *HgCLE*, like *CLE19* and *CLE40*, can also substitute for *CLV3* when expressed in the SAM (Wang et al. 2005).

The fact that several *CLE* members can phenocopy *CLV3* when misexpressed in plants suggests that the specificity of the different *CLE* members is conferred, at least partly, by cis-regulatory elements defining spatial and temporal domains, rather than by the protein sequence only, and further demonstrate that the C-terminal domain is important in receptor binding and recognition. Apparent molecular mimicry, here exemplified by *HgCLE*, is generally observed for genes potentially involved in infection processes between bacteria and plants or nematodes and plants, and is generally believed to be the result of horizontal gene transfer (Collmer 1998; Yan et al. 1998).

RALF

The discovery of the tomato systemin revealed that this peptide could induce the alkalinization of the medium when added to a suspension cell culture of *Lycopersicon peruvianum* (Felix and Boller 1995). Using this property as an assay for systemin in tobacco leaf extract, Pearce et al. (2001*b*) found a 5 kDa peptide that causes a stronger and faster alkalinization response than systemin itself (Pearce et al. 2001*b*). This peptide, which was called Rapid Alkalinization Factor (RALF), was also purified from tomato and alfalfa leaves (Pearce et al. 2001*b*). Meanwhile, Haruta and Constabel (2003) also purified a RALF homolog from poplar cell culture using a similar alkalinization assay.

RALF-like genes have now been found in more than 16 species (Pearce et al. 2001*b*), including 34 *RALF-like* genes in the *Arabidopsis* genome. The fact that this family has emerged from duplication events is strengthened by the fact that *AtRALF* copies often occur in tandem repeats in the *A. thaliana* genome and show high sequence similarity (Olsen et al. 2002). The *Nicotiana tabacum* RALF (NtRALF) is a 49-amino-acid peptide that is processed from a preproprotein of 115 amino acids. The preproprotein has a N-terminal signal peptide and a highly conserved mature peptide that is located at the C-terminus of the protein (Fig. 1). The activity of NtRALF requires the formation of disulfide bridges between Cys-18 and Cys-28 and between Cys-41 and Cys-47. All *RALF-like* genes

identified share this common architecture. Another common feature that is conserved among preproRALFs is an acidic segment upstream of a dibasic residue motif possibly involved in the processing of the peptide (Cleland 1995; Franssen and Bisseling 2001; Olsen et al. 2002) via the action of the subtilisin-like serine protease of the kexin family. In animal cells, these proteases are involved in the activation of peptide hormones, growth factor and viral proteins (Siezen and Leunissen 1997). Even if the cleavage specificity is unknown in plants, serine proteases are abundantly found in plants and are, therefore, good candidates for RALF processing (Siezen and Leunissen 1997; Franssen and Bisseling 2001). It should be noted that contrary to a previous report (Boller 2005), RALF has never been shown to be a glycopeptide.

The physiological effect of systemin is similar to that of erythrosin B, an inhibitor of plasma membrane proton ATPases, which causes an alkalinization of the medium, whereas suramin, an inhibitor of ligand/receptor interaction, inhibits the alkalinization of the medium. It was therefore concluded that the alkalinization of the medium mediated by systemin is the result of an inhibition of the plasma membrane proton ATPase by a receptor-mediated signaling pathway (Stratmann et al. 2000). Since RALF induces the phosphorylation of a 48 kDa MAPK and the concomitant alkalinization of the medium like systemin (Pearce et al. 2001*b*), and since this alkalinization is inhibited by suramin, it could act as a secreted ligand that controls cellular events by its interaction with a membrane receptor. Recently, transient expression of *AtRALF* in *Nicotiana benthamiana* leaves revealed that it localizes to the cell wall after being observed transiently in the endoplasmic reticulum (Escobar et al. 2003). Moreover, the use of a photoactivable analog of the tomato RALF revealed two RALF interacting proteins of 120 and 25 kDa. It appears that these two proteins act in a membrane complex, since only one high affinity site (KD = 0.8 nmol/L) was characterized (Scheer et al. 2005).

The challenge now is to uncover the biological function of the RALF peptides in plants. RALF does not induce the synthesis of tobacco trypsin inhibitors (Pearce et al.

2001*b*) and treatments with elicitors and defense or wound hormones do not induce the transcription of *RALF-like* genes in cell cultures (Haruta and Constabel 2003) or intact plant tissues (Germain et al. 2005*a*). Moreover, the expression of some *RALF-like* genes in *A. thaliana* is not altered in ethylene-constitutive-background (*ctr1*) and in systemic-acquired-resistance-background (*mpk4*) mutants (Olsen et al. 2002). Therefore, *RALF* is most probably not involved in defense responses.

However, *RALF-like* genes are found in several EST databases (Pearce et al. 2001*b*; Germain et al. 2005*b*), and tissues expression analyses of *RALF-like* genes in poplar (Haruta and Constabel 2003), *Arabidopsis* (Olsen et al. 2002) and *Solanum chacoense* (Germain et al. 2005*a*) have revealed specific and narrow expression patterns for some *RALFs* and broader expression patterns for others. Therefore, *RALFs* could control various developmental processes such as fruit development and root elongation. In fact, two *RALF-like* genes are specifically expressed in fruits after pollination (Germain et al. 2005*a*), and the addition of a synthetic tomato RALF peptide to the medium arrest root growth of tomato and *A. thaliana* seeds (Pearce et al. 2001*b*) and decreases the elongation rate of pollen tube in vitro (Bedinger et al. 2004). A detailed analysis revealed that *Arabidopsis* roots grown in presence of a tomato RALF have fewer root hairs and enlarged root meristem cells (Pearce et al. 2001*b*). Taken together, these results suggest that *RALF* could be involved in the control of cell growth, perhaps through an exchange of protons across the plasma membrane. Plant cell growth has already been shown to be modulated by auxin via a similar proton pumps mechanism (Cleland 1995).

A better understanding of the biological function of *RALF* will await the analysis of transgenic lines. However, the high sequence similarity shared by some *RALFs* and the presence of numerous *RALF* genes per genome could obscure the effect of overexpression, knock-down or knockout lines. Promoter and in situ analyses will pinpoint the specific expression domains of each *RALFs* and the identification of the protease(s) that processes RALF will allow a better understanding of the activation mechanism of this preproprotein.

Finally, the identification of the recently published RALF-interacting membrane proteins should lead to the deciphering of this signaling cascade.

DEVIL/ROTUNDIFOLIA

DEVIL1 (Wen et al. 2004) and *ROTUNDIFOLIA4* (Narita et al. 2004) are two different genes that were found nearly at the same time by two independent groups, and that are part of the same gene family. *DEVIL1* (*DVL1*, which was described first), and *ROTUNDIFOLIA4* (*ROT4*) were both discovered by activation tagging in *Arabidopsis*. In the case of *DVL*, the phenotype caused by the insertion of the tag is pleiotropic, with a shorter stature, rounder rosette leaves, clustered inflorescences, shorter pedicels and horned fruit tips (hence the name Devil). In the case of *rot4*, the phenotype has rounder and shorter leaves.

Surprisingly, in both cases, the activation tag does not induce the expression of any of the nearby annotated genes. Thus, the investigators searched for short ORFs in the genomic DNA surrounding the tag. Short ORFs encoding small peptides of 51 (*DVL1*) or 53 (*ROT4*) amino acids were highly overexpressed in the tagged line when compared to the wild type plant. When a genomic fragment encoding *DVL1* cloned downstream of four CaMV 35S transcriptional enhancers was reinserted in the wild type plant, it mimicked the phenotype observed in the activation tag line, thereby confirming that the previously unannotated ORF is responsible for the observed phenotype (Wen et al. 2004).

In total, 22 members of the *DVL/ROT* family were found (*ROT4* is *DVL16*) and all of them encode highly homologous small protein devoid of a signal peptide (Fig. 1), and most of them (16 of 22) (Narita et al. 2004) are present in EST databases but are absent from EST or genomic sequences outside the angiosperms (Wen et al. 2004). Overexpression of various family members, *DVL2*, *DVL3*, *DVL4,* and *DVL5*, also cause a similar pleiotropic phenotype. RNAi and antisense overexpression of *DVL1* are sufficient to restore the normal phenotype in the activation-tagged line, but do not confer a phenotype

when transformed in wild-type plants (Wen et al. 2004), suggesting a functional redundancy between the family members.

The conserved RTF domain found at the C-terminus (Narita et al. 2004) consists of 29 amino acids, which are highly conserved and contains several basic residues and a pair of cysteines that is present in all but one member of the group (DVL21 contains only one cysteine, H. Germain, unpublished observation). The overexpression of the ROT4 peptide without the upstream N-terminal sequence (to the RTF domain) or without the downstream C-terminal sequence (to the RTF domain) is sufficient to confer the round leaves phenotype (Narita et al. 2004), while a frameshift or point mutations in the C-terminal domain (after RTF) of *DVL1* fails to cause the phenotype (Wen et al. 2004). These results indicate that the RTF domain is necessary and sufficient to cause the *DVL/RTF* phenotype and that its integrity is needed for peptide function. The high level of conservation in the C-terminal domain is reminiscent of other families of putative signaling peptides such as CLE (Cock and McCormick 2001) and RALF (Olsen et al. 2002), which also have highly conserved C-terminal domains.

ROT/DVL family members do not have a classical signal peptide, one that can be predicted using SignalP (Nielsen et al. 1997), but ROT4 is located at the plasma membrane, possibly as part of a membrane complex or as a peripheral membrane protein (Narita et al. 2004). Because some proteins can be secreted despite the lack of a leader sequence, a process referred to as leaderless secretion or non-conventional/classical secretory pathway (Bendtsen et al. 2004*a*), we have used SecretomeP 1.0, a sequence-based software (Bendtsen et al. 2004*b*) to predict whether the peptides described in this paper and lacking a leader sequence are proteins secreted via the non-conventional secretory pathway. All DVL/ROT peptides (except DVL18) have a predicted value above threshold (0.6), indicating that they are likely secreted through a ER–Golgi independent secretory pathway (mean = 0.758, SD = 0.09, n = 22), which would explain their localization at the plasma membrane (Narita et al. 2004). Although a highly conserved cysteine pair is observed in

DVL/ROT, the fact that these proteins do not go through the default secretory pathway suggests that these cysteines are not part of a disulfide bridge. Based on GFP results, these peptides do not appear to be mobile (Narita et al. 2004) and their capacity to act as signaling peptide remains to be demonstrated. One T-DNA line was found in *Arabidopsis* and two in rice. Although these mutants should disrupt the RTF domain, none of them display a visible phenotype, supporting the possibility of domain and (or) functional redundancy among the related family members (Narita et al. 2004). Although activation tagging, which has started to be used only recently with *Arabidopsis* (Weigel et al. 2000), should prove a useful method to discover small peptides that are not currently annotated in the *Arabidopsis* genome, care should be taken when drawing conclusion with regards to the possible function of the activated gene since the strong promoters used for activation tagging could trigger ubiquitous expression at non-physiological levels and could therefore cause an artefactual phenotype.

POLARIS

POLARIS (*PLS*) was identified in a promoter trap transgenic line in which reporter gene expression is specifically detected in the basal region of the embryo and the root tip (Topping et al. 1994; Topping and Lindsey 1997). *PLS* is also expressed in the basal part of *hydra* embryo and *emb30* mutant, the latter lacking shoot and root meristems, indicating that *PLS* is not a root marker (Topping and Lindsey 1997). It was later shown that *PLS* is also expressed in the leaf vasculature (Casson et al. 2002). A mutant homozygote for a T-DNA insertion in the *PLS* gene has a short root, whose growth responds less than wild-type plants to exogenous application of cytokinin. Overexpressing lines display a more complex leaf venation than the wild-type and are trichome-less. Uncommonly, the initiation of the short *PLS* transcript (108 nt) takes place within an upstream gene. The 4.6 kDa (36 amino acids) peptide that is predicted from the 108 bp ORF has no sequence identity with any peptide in the public databases, signal peptide or cysteine pairs, and has not been shown to be mobile. Results from a complementation experiment using a *PLS* ORF carrying either a

mutated or an unmutated ATG initiation codon (ATC) suggests that *PLS* could act as a peptide and not a RNA molecule. All plants lines generated with the mutant start codon are unable to rescue the *pls* mutant, whereas the plants with a wild-type ATG could (Casson et al. 2002). Investigation of the *PLS* promoter using a GUS reporter gene showed that it is inducible by functional auxins and not by non-functional auxin analogs (Casson et al. 2002) and intense GUS staining was observed at the site of lateral root initiation, a process induced by auxin (Topping and Lindsey 1997). Casson et al. (2002) therefore concluded that auxin could be a regulator of the spatial patterning of *PLS* expression.

Whether PLS is really a secreted signaling peptide still remains hypothetical and clarification on that matter will await additional data coming from GFP fusion or immunochemistry experiments.

Reproductive development-related peptides

IDA and IDA-like genes

Abscission is the process by which the plant sheds its leaves, flowers or fruits in a controlled manner at a special separation layer named the abscission zone (AZ). The flower generally sheds all its floral parts after they became useless following the fertilization process (Bleecker and Patterson 1997).

A new and unique mutant, Inflorescence Deficient in Abscission (*ida*), was identified in *Arabidopsis* (Butenko et al. 2003). In the *ida* mutant all floral parts remain attached indefinitely to the plant body after fertilization, and this mutant, unlike other abscission mutant, is sensitive to ethylene, which has previously been shown to be involved in abscission (for a review see Roberts et al. 2002). When the GUS reporter gene is fused to the *IDA* promoter, a very specific expression pattern is observed and staining is restricted to the surrounding area of the AZ at specific developmental stages. The break strength, which is required to remove the petals at different stages following pollination, decreases in both

the *ida* mutant and in wild-type plants, without ever reaching zero in the *ida* mutant. Afterward, only in the *ida* mutant does the break strength progressively return to its initial value. Scanning electron microscopy revealed some differences in the appearance of the abscission zone between *ida* and wild-type plants. In wild-types, the cells at the fracture plane present a rounded appearance whereas the cells of the *ida* mutant display only flattened cavity (Butenko et al. 2003).

The *ida* mutant was found in a T-DNA screening for lines delayed in abscission and the identity of the mutated gene causing the *ida* phenotype was confirmed by complementation (Butenko et al. 2003). The *IDA* gene encodes a small peptide of 77 amino acids that has a functional (classical) signal peptide. TBLASTN analyses revealed that the *IDA* gene is present in several species and that other highly similar genes can be found in *Arabidopsis*, indicating the presence of a small gene family. The four *IDA*-like genes uncovered in *Arabidopsis* display a similar structure. All are very basic proteins (pI between 10.2 and 12.5) and sequence conservation in the C-terminal part is high (Fig. 1). It should be noted that no conserved cysteine pairs are observed in the IDA family. All but one *Arabidopsis IDA*-like (*IDL*) genes, as indicated by RT-PCR analysis, display organ-specific expression patterns; *IDL*-2 was detected in all tissues tested. With such a diversity in expression patterns and given the important role observed for IDA, it is tempting to speculate that IDL acts in intercellular signaling to regulate diverse aspects of plant development.

Egg apparatus 1 and transparent leaf area 1

In maize, two peptides recently identified by the Dresselhaus group have been shown to have profound effect in reproduction and leaf development (Dresselhaus et al. 2005; Marton et al. 2005). In higher plants, the pollen tube, which contains the sperm cells, grows through the transmitting tract of the style and is then believed to be guided towards the female gametophyte by unknown molecules. ZmEA1 (*Zea mays* Egg Apparatus 1) is a small predicted protein of 94 amino acids with a predicted transmembrane domain. It is

expressed (RT-PCR and GUS staining experiments) exclusively in the egg apparatus before fertilization and it disappears after fertilization (Marton et al. 2005). Transgenic plants for the *ZmEA1* gene (RNAi and antisense) display lower seed set, but wild-type plants crossed with transgenic pollen do not show any phenotype. C-terminal GFP fusion indicated that the signal originating from the egg apparatus eventually spreads to the surface of the nucellus at the micropylar opening of the ovule, suggesting that the small protein is mobile. The authors suggest that there is proteolysis of the conserved C-terminal region from the transmembrane domain to allow secretion of the mature peptide. Putative orthologs of *ZmEA1* have been found in rice, barley, millet and *Tripsicum dactyloides*, but apparently not in dicotyledonous species (McCormick and Yang 2005). Overall peptide sequence conservation is low, but the C-terminal domain is highly hydrophobic and very well conserved. In fact, when only the conserved C-terminal domain (Fig. 1) is used for similarity searches, more than 60 independent sequences are retrieved, including both monocot and dicot species, even basal angiosperm species (Gray-Mitsumune and Matton 2006). Several methionines are present in the ZmEA1 peptide; in fact, four methionines are found in frame upstream of the putative transmembrane domain. This makes selection of the proper initiation codon difficult and produces some uncertainty about the resultant peptide sequence. Taking into account the Kozak consensus sequence (RCCAUGG, R is a purine) (Joshi et al. 1997; Kozak 1999), only the third predicted methionine would have a near Kozak consensus (3/4 outside of the AUG), while the other methionines have only one or two out of four matches. Using the most likely methionine (the 3rd one) as the translation start site would yield a peptide of 76 amino acids that contain a strongly predicted signal peptide and, after cleavage, a mature protein of 49 amino acids. This observation is also supported by the fact that the second methionine of OsEAL1 has a perfect Kozak consensus sequence (4/4) and generates a peptide that also has a signal peptide. It should be noted, however, that this is not true for all EA-like proteins, including OsEAL2. The fact that the authors observed that a ZmEA1-GFP fusion could be detected away from its synthesis site would in fact be better explained if the protein started from an

alternative initiation codon, in a better translation initiation context, and had a bona fide signal peptide.

The other maize small putative signaling peptide, named *TRANSPARENT LEAF AREA1 (TLA1),* was found in a search for genes that display high expression levels and egg cell specificity (Dresselhaus et al. 2005). The transcript encoding *ZmTLA1* varies in length from 310 nt to 502 nt due to the presence of seven alternative poly(A) signal in its 3'UTR. The ORF encoded by the *ZmTLA* transcript is 27 amino acids in length, 21 of which are hydrophobic. Since transposon-tagged lines of *ZmTLA* were unavailable, the authors used antisense overexpression to investigate the function of *ZmTLA*. A pleiotropic phenotype is observed and its most striking features are dwarfism and the presence of yellow or transparent leaf areas (*tla*). The *tla* phenotype is observable on most plants displaying high transgene expression, and can be visible from the seedling stage, throughout the plant's life, and is not accompanied by necrosis. A defect in anther maturation is also observed. Overexpression of sense *ZmTLA* is lethal in maize and *Arabidopsis* (from 500 000 selected progeny seedlings derived from *Arabidopsis*, none survive the primary leaf stage). The use of a dexamethasone-inducible promoter revealed that induction of *ZmTLA-GFP* in *Arabidopsis* causes developmental arrest. In plants that strongly overexpress *ZmTLA-GFP*, the elongation of hypocotyl cells, the development of primary root, and root hair differentiation are suppressed.

In plant lines expressing a *ZmTLA-GFP* construct, fluorescence is observed only in the apoplast (Dresselhaus et al. 2005). With appropriate controls, the authors showed that ZmTLA-GFP does not accumulate in the endoplasmic reticulum, Golgi or any other vesicle, indicating that ZmTLA is secreted into the extracellular space, but bypasses the default secretory pathway. BLAST analysis against the databases does not retrieve any similar sequences, although DNA blot analysis reveals hybridizing bands in rice and other cereals.

These data taken together suggest that ZmEA1 could be a mobile signaling peptide that is involved in pollen tube guidance, at least in maize. Whether ZmEA1 binds to a receptor at the tip of the pollen tube to control its growth remains hypothetical, although pollen tube specific receptor-like kinase have already been discovered (Muschietti et al. 1998; Kim et al. 2002). As for TLA, it would represent the first plant proteolipid for which a function has been attributed.

The HT self-incompatibility modifier

In the Brassicaceae sporophytic self-incompatibiltiy system (SSI), pollen recognition and rejection is determined by the interaction between a highly polymorphic pair composed of a small pollen secreted peptide, the SCR/SP11 ligand, and a stigma-expressed receptor kinase (SRK) (Nasrallah 2002; Nasrallah 2005). A completely different mechanism is found in the gametophytic self-incompatibility (GSI) system of the *Solanaceae*, *Rosaceae* and *Scrophulariaceae*, where a polymorphic pair composed of a stylar-expressed ribonuclease (S-RNase) interacts with a pollen-expressed F-box protein, possibly targeting the S-RNase to the ubiquitin-proteasome degradation pathway (McClure 2004). Thus, in S-RNase-based GSI systems, the simplest model implies that the cytotoxic S-RNases that enter the pollen tube are targeted for degradation in a self-compatible pollination (thus the F-box protein acts as an inhibitor of non-self S-RNases by tagging them for degradation), while in a self-incompatible pollination, interaction of the S-RNase with its cognate F-box protein inhibits degradation of the S-RNase, which then can act on cellular RNAs of the pollen tube, leading to pollen tube growth arrest. Although these two proteins are necessary and sufficient for the recognition and rejection events, numerous experiments have demonstrated that other stylar factors are also necessary for the proper expression of the self-incompatibility phenotype (reviewed in McClure et al. 2000). Such factors, often considered as modifier loci, are present in the genetic background of SI plants, unlinked to the S-locus, and often lost in self-compatible (SC) relatives of self-incompatible (SI) species.

One such candidate for a modifier gene was originally cloned in *Nicotiana alata*, and named *HT* (McClure et al. 1999). The *NaHT* gene was cloned based on a differential screen between stylar expressed mRNAs from SC *Nicotiana plumbaginifolia* and a SC accession of *N. alata* that is defective in S-RNase expression, but competent to express SI (Murfett et al. 1996). Antisense *NaHT* plants with reduced level of the HT protein, but with normal levels of S-RNases are either fully or partially self-compatible, when pollen tube growth is measured in the style of these transgenic plants (McClure et al. 1999). This strongly suggested that the *NaHT* gene encodes a modifier factor that is necessary for the SI reaction to occur.

Phylogenetic analyses of numerous NaHT homologues isolated in various solanaceous species clearly demonstrated that two different HT isoforms exist (O'Brien et al. 2002). In *Solanum chacoense*, where the expression of both isoforms is suppressed either through antisense or RNA interference, only the ScHT-B isoform, the most closely related to the *NaHT* gene (now renamed *HT-B*), is involved in transforming a SI plant to a SC plant (O'Brien et al. 2002). Suppression of the HT-A isoform has no effect on the self-incompatibility phenotype, although both isoforms share extensive sequence similarities and are coordinately expressed in styles, as for the S-RNases (O'Brien et al. 2002). All the HT proteins share some common features (Fig. 1). First, a highly conserved N-terminal region that is predicted to be a signal peptide that upon cleavage would release an approximately 75–80 amino acid peptide (approx. 8.5 kDa). Experimental evidence from N-terminal sequencing of a partially-purified extract of the peptide from *N. alata* indicates that the predicted signal sequence is indeed cleaved (McClure et al. 1999). Second, and similar to the RALF peptide family (Pearce et al. 2001*a*), a dibasic residue motif is also found in the mature HT protein, suggesting that the HT protein could be subjected to further processing. Furthermore in *N. alata*, an antiserum raised against the NaHT protein also detects a faster migrating fragment, that corresponds to a protein fragment cleaved at the dibasic residue motif position as determined from N-terminal sequencing (in that region QKI is found in *N. alata* instead of QKK as in all other species determined). This fragment

would correspond to an approximately 50-amino-acid peptide of roughly 5.5 kDa if no further post-translation modification occurs; however, it migrates at a higher molecular weight, suggesting that the fragment released undergoes modification (McClure et al. 1999). Third, all HT homologues possess a C-terminal region composed of consecutive stretches of 15–20 asparagine and aspartic acid residues, flanked by conserved cysteine motifs (CX2CXC and CX3CC), which are probably involved in disulfide bridges. The HT proteins are structurally similar to most of the previously described peptides, and could thus be also considered as putative signaling peptides. Furthermore, no direct interaction between the HT-B protein and the S-RNase could be detected either biochemically (McClure et al. 1999) or in a two-hybrid system (O'Brien et al. 2002), suggesting that HT proteins and S-RNases do not interact directly. Apart from its role in self-incompatibility barrier breakdown, only one other phenotype could be associated with the loss of *HT-B* expression. In RNAi *ScHT-B* lines displaying severely reduced *HT-B* mRNA levels, flower senescence and stylar abscission are markedly retarded following a SI pollination, suggesting that *HT-B* may also have a more general role in floral abscission, like the IDA peptide described previously. This is supported by a link between *HT-B* mutant plants and ethylene, a plant hormone involved in senescence and abscission (M. O'Brien and D.P. Matton, unpublished results). Plants with reduced levels of *HT-B* (either transgenic lines or WT accessions that have a naturally low level of *HT-B* transcripts) and treated with ethylene produced small parthenogenic-like fruits that never developed to a size greater than an ovary 4–5 d after pollination.

Defense-related peptides

Systemin-like gene family

Tobacco and tomato plants release proteinase inhibitors as a defense mechanism against herbivores (or wounding) (Green and Ryan 1972). The mediator of this defense response, systemin (in tomato), is a small mobile 18-amino-acid peptide active at

femtomolar concentrations (Pearce et al. 1991). It is also found in potato, nightshade, and pepper plants (Constabel et al. 1998), and it binds to a receptor kinase of the leucine-rich repeat family (Scheer and Ryan 2002). Synthetic systemin is as active as the native systemin (for reviews on systemin see Ryan et al. 2002; Torii 2004; Wang and He 2004). Neither EST nor polypeptide corresponding to tobacco systemin could be found despite the fact that tobacco displays a very similar response to wounding, inducing the accumulation of proteinase inhibitors. However, the use of an alkalinization assay enabled the identification of two tobacco peptides that have effects similar to the tomato systemin. The two peptides, named tobacco hydroxyproline-rich systemins (TobHypSysI and TobHypSysII) are both 18 amino acids long, contain hydroxyprolines, and are linked to pentose residues (Pearce et al. 2001*a*). Notably, synthetic TobHypSysI and TobHypSysII are 10 000 times less active than their native counterpart, indicating that the sugar decoration is important for the peptide function. Both peptides are encoded as a prohormone by the same 165 amino acid coding mRNA, a phenomenon novel for a plant peptidic hormone. Tomato also contains similar systemin-like peptides, named TomHypSysI, TomHypSysII, and TomHypSysIII, which are, respectively, 18, 20, and 15 amino acids long, are glycosylated, contain hydroxyprolines, and, as for the tobacco HypSys, are all derived from the same precursor protein (Pearce and Ryan 2003). In situ mRNA hybridization revealed that *TomHypSys* is synthesized in phloem parenchyma cells in response to wounding, systemin, or methyl jasmonate (Narvaez-Vasquez et al. 2005). A GFP fusion of TomproHypSys demonstrated that the protein precursor localizes to the cell wall of epidermal cells, whereas transmission electron microscopy revealed that the precursor protein of TomproHypSys is found in the cell wall matrix of vascular parenchyma cells (Narvaez-Vasquez et al. 2005).

Striking differences between the original systemin and these systemin-like peptides include the presence of an N-terminal signal peptide, hydroxylation of proline residues, and glycosylations. These features, all absent from systemin, indicate that the peptides must go through the secretory pathway (hydroxylation of proline residues and glycosylation are

post-translational modifications), whereas systemin does not have a signal peptide and is believed to be synthesized on cytosolic ribosomes (Pearce and Ryan 2003). It should be noted that although tobacco does not have a true systemin ortholog, it responds to systemin application in an alkalinization assay when tobacco plants are engineered to overexpress the systemin receptor kinase, suggesting that downstream signaling mechanisms are conserved (Scheer et al. 2003). Therefore, these newly identified peptides could be classified as a small protein family (even though they are coded by only one transcript) that are functionally related to the original systemin, and that may also bind to a receptor kinase.

Physiological modulators

Plant atrial natriuretic peptide (pANP)

In mammals, the natriuretic peptide (NP) was originally discovered in an extract of rat atria in 1981 (de Bold et al. 1981). The atrial natriuretic peptide is encoded by a 152-amino-acid precursor protein, the preproANP. Mature ANP is obtained after removal of the signal peptide (proANP) and cleavage of the propeptide at position 98 and deletion of the two carboxy-terminal amino acids, resulting in the C-terminal ANP peptide (99–126), which is then circularized through the formation of a disulfide bond between amino acids 7 and 23. It should be mentioned that the N-terminal peptide is also processed into three distinct peptides, proANP 1–30, proANP 31–67, and proANP 79–98, which share similar properties with regards to their capacity to cause vasodilatation and increase cellular cyclic guanosine 3',5'-monophosphate (cGMP) (Vesely et al. 1987).

The possible presence of ANP in planta was first assessed by Vesely and Giordano (1991). They used a radioimmunoassay, in which the antibodies raised against animal ANP (amino acids 99–126), proANP (amino acids 1–30) and the mid-portion of the N-terminus (proANP 31–67) were used to detect the presence of these peptides in the monocot plant *Dracena godseffiana*. All three antibodies detected a peptide, and gel permeation

chromatography revealed that the recognized peptides were at the expected molecular weight.

In a follow-up paper, Vesely and collaborators showed that rat proANP 1–30, proANP 31–67, and proANP 79–98, but not the ANP (99–126), could increase solute flow, solute uptake and transpiration (Vesely et al. 1993). Why ANP did not produce the same physiological changes as the N-terminal peptides still remains unclear. The authors also showed the immunodetection of ANP, proANP 1–30, and proANP 31–67 from a wide range of very diversified plant species, including mosses, ferns, gymnosperms, and a protist (*Euglena*).

Transpiration rate and gas exchange in plant leaves are controlled via stomatal opening and closing. The state of the stomatal aperture is itself dependent on the turgor pressure of the two guard cells that surround the stomata, which is regulated by water and solute fluxes. Gehring et al. (1996) took advantage of this turgor-based mechanism to demonstrate that rat ANP (rANP) could cause stomatal opening in plants (i.e., that rANP can cause solute movement in cells) (Gehring et al. 1996). They demonstrated that rANP could cause stomatal opening at micromolar concentration in three different plant species. It was later shown (Pharmawati et al. 1998) that stomatal opening is dependent on two factors: (1) the rANP structure, since the formation of disulfide bridges is necessary, and (2) rANP signaling occurs through a cGMP mechanism (shown using cGMP signaling inhibitors), as previously demonstrated in animals.

Plant ANPs (PNPs) were found by Billington et al. (1997) using immunoaffinity chromatography to purify ANP antiserum-specific epitope from a plant extract, followed by size exclusion separation of the eluate (Billington et al. 1997). The resultant fractions were assayed for their capacity to induce stomatal opening. This method allowed the authors to find functional immunoreactive plant ANPs (irPNP), which were shown to modulate the ATP-dependent proton gradient in vesicles (proton gradient is used as the driving force for K+ uptake, which leads to increased cell turgor) (Maryani et al. 2000), cell turgor in

Solanum tuberosum mesophyll cell protoplast (Maryani et al. 2001), and cGMP levels in a manner similar to that observed previously for rANP (Pharmawati et al. 1998; Pharmawati et al. 1999).

Using N- and C-terminal partial protein sequences obtained from immunoaffinity-purified plant natriuretic peptide immunoanalogues (irPNPs) from *Solanum tuberosum*, (Ludidi et al. 2002) successfully identified and eventually isolated a plant gene encoding irPNP. The *AtPNP-A* gene encodes a small 126 amino acids protein with an N-terminal signal peptide that had not been previously annotated in the databases. AtPNP displays weak overall sequence identity with human ANP, although significant similarity (50%; and 25% identity) is found between residues 33–66 of AtPNP and the mature human ANP. This section of the peptide (33–66) is also sufficient to confer solute uptake by protoplast (Morse et al. 2004). The *AtPNP-A* gene shares weak sequence similarities with the expansins, although it lacks a carbohydrate-binding domain. Purified recombinant full-length AtPNP-A protein, full AtPNP-A without the signal peptide, and AtPNP-A (33–66) could all induce stomatal opening, and recombinant AtPNP-A without the signal peptide causes ion (K+, Na+ and H+) fluxes in *A. thaliana* roots (Morse et al. 2004).

The data presented above strongly suggest that PNP act as a plant signaling peptide. PNP, like several plant signaling peptide (see Table 2), is encoded as a preproprotein containing a signal peptide; it can act at nanomolar concentrations (Morse et al. 2004), and is present in xylem exudates (Maryani et al. 2003), which are associated with transport rather than protein synthesis. It is also possible that PNP undergoes additional processing since amino acids 33 to 66 are sufficient to induce stomatal opening (Morse et al. 2004). It should be noted that a *PNP*-like sequence is also found in the bacterium *Xanthomonas axonopodis* (Nembaware et al. 2004). The activity of the encoded protein has not been demonstrated yet, but such a case of apparent molecular mimicry would not be without precedent in the plant signaling peptides (Olsen and Skriver 2003). NPs would represent the first known conserved hormonal system that is shared between animals and plants. The

sequence similarity and homology in the signaling mechanism strongly support that this mechanism, involved in homeostasis regulation, may have evolved before the last common ancestor of the two kingdoms, 1.6 billions years ago (Meyerowitz 1999).

Table 3 Characteristics of the signaling and bioactive peptides

Name of peptide	Receptor if known	Precursor size (in amino acid)	Final size (in amino acid)	Signal peptide	Post-translationnal modification	Expression	Comment
CLAVATA3	CLAVATA1	96	78	Yes	No	Specific to layer 1 and 2 of shoot meristem central zone	Apical meristem maintenance
Systemin	SR160	200	18	No	No	Parenchyma cells of vascular bundles	Induced by wounding
Phytosulfokine	PSKR	80	5	No	Yes, sulfated	Ubiquitous	Cellular dedifferentiation and proliferation
SCR/SP11	SRK	83	59	Yes	Disulfide bridges possible	Anther, pollen coat	Pollen determinant of self-incompatibility
BRICK	Not identified	62	62	No	No	Throughout the plant	Formation of epidermal cell lobe and actin patches involved in subsidiary mother cell polarization
CLE/ESR	Not identified	75 to 125 average 95	~70	Yes (most)	Unknown	The various family members cover all plant tissues	High pI, C-terminal highly conserved
RALF	Not identified	115	49	Yes	N-terminal processing, disulfide bridges	The various family members cover all plant tissues but individual show specificity	Stop root growth, induces alcalinization of the media, C-terminus highly conserved
DEVIL1/ROTUNDIFOLIA	Not identified	51	51	No[1]	No	Most member ubiquitous	Pleiotropic, RTF domain in C-terminus is highly conserved
POLARIS	Not identified	36	36	No	Unknown	Roo tip and basal part of the plant and leaf vasculature	Knock-out has short roots, overexppressing line has more complex leaf venation
IDA	Not identified	77	77	Yes	No	IDA, in flower specific to abcission zone. IDA-like showed specificity to other tissue	High pI, and C-terminal PIP motif highly conserved

Egg apparatus	Not identified	94	94	No	No	Exclusively in the maize egg apparatus	Has a putative transmembrane domain, C-terminus highly conserved
HT	Not identified	100	75	Yes	Possible N-terminal processing and glycosylation	Style specific	Self-incompatiblity modifier
TobSys TomSys	Not identified	165 146	18 20, 18, 15	Y	Glycosylated, hydroxylated	Cell wall matrix of vascular parenchyma cells	Encoded on a single mRNA, effects similar to systemin
Plant natriuretic peptide	Not identified	126	27	Yes	Disulfide bridges possible	Peptide isolated in xylem exudate	Regulate solute flow
ENOD40	Not identified	10 to 13	10 to 13	No	Unknown	Strongest in root nodules	Binds to sucrose synthase, transcript found at high level in nodules
NCR/DEFL	Not identified	60-90	40-70	Yes (most)	Unknown	NCR specific to nodule, DEFL mostly in reproductive tissues	1100 DEFL found in various species
4 kDa peptide (leginsulin)	Not identified	119	31	Yes	3 disulfide bridges	Germinating radicles, seeds	Possibly cellular proliferation

1 Detected using SecretomeP 1.0

ENOD40

Two independent groups discovered *ENOD40* nearly at the same time. Yang et al. (1993) found two soybean clones (*GmENOD40-1*, *GmENOD40-2*) through a cDNA library screen of root and nodule cDNAs, and the *ENOD40* clones yielded a nodule-specific signal. Kouchi and Hata (1993) also found a soybean clone, which is identical to *ENOD40-2*. In both cases, expression of the gene was induced in nodules after infection of the legume host with a *Rhizobium* spp., and was therefore named *ENOD*. In situ hybridization with *ENOD40* transcript revealed that following bacterial inoculation of the roots, *ENOD40* transcript could be detected in the nodule (Kouchi and Hata 1993; Yang et al. 1993; Asad et al. 1994; Crespi et al. 1994; Matvienko et al. 1994), as well as in other tissues (Asad et al. 1994; Fang and Hirsch 1998). *ENOD40* has now been reported (in papers or GenBank) in at least 18 species, including legumes, non-legumes, monocots, and dicots. The patterns of expression, as well as the discovery of *ENOD40* in non-legume species and in monocot, confirms that *ENOD40* is not only involved in symbiotic interaction and nodule development, but also in a more general developmental process conserved throughout the plant kingdom.

Since *ENOD40* was discovered, researchers have debated whether it is active as a peptide or as a RNA molecule. It is now accepted that the *ENOD40* transcript, which is about 700 bp long, encodes very short ORFs. Region 1, which is located in the 5'-terminus of the transcript, encodes a short and highly conserved oligopeptide of 12–13 amino acids in legumes and of 10 amino acids in non-legume (van de Sande et al. 1996). The consensus sequence of the region I-encoded peptide is W-X4-HGS. Region 2 is highly conserved at the nucleotide level, but lacks a discernible ORF or coding capacity in monocots (Kouchi et al. 1999; Larsen 2003) and is therefore not believed to be biologically active.

Solid evidences regarding the function of *ENOD40* as a peptide have only recently started to accumulate. Rohrig et al. (2002) performed affinity purification of nodule extracts using biotinylated ENOD40 (ORF1) and identified a 93 kDa protein, that after MALDI-

TOF analysis, was determined to be nodulin 100, a subunit of the enzyme sucrose synthase, one of the most abundant nodule proteins (Rohrig et al. 2002). A point mutation in the synthetic peptide abolishes binding to nodulin 100, thereby confirming the binding specificity. Hardin et al. (2003) demonstrated that sucrose synthase serine 170 could be phosphorylated both in vitro and in vivo by CDPK II. Knowing that the ENOD40 peptide could bind to sucrose synthase, the authors tested whether the presence of the ENOD40 peptide affects S170 phosphorylation by CDPK II. The dodecapeptide encoded by region I could fully antagonize (97%) S170 phosphorylation (Hardin et al. 2003). When sucrose synthase is phosphorylated at position 170 (p170-SUS), it is found in the same fraction as the proteasome (unlike the non-phosphorylated form), suggesting that this phosphorylation event could target sucrose synthase to the proteasome-mediated degradation pathway (Hardin et al. 2003). This antagonistic reaction mediated by the ENOD40 peptide against the phosphorylation event on S170 of sucrose synthase is mediated via disulfide bond formation between cysteine 4 of ENOD40 peptide and cysteine 264 of the sucrose synthase subunit (Rohrig et al. 2004). The authors also showed that the cleavage activity, but not the synthesis activity, of sucrose synthase increases following ENOD40 covalent binding. Finally, the last piece of evidence that supports the role of *ENOD40* as a peptide comes from a combination of transgenic experiments and the use of a synthetic peptide. Transgenic plants of *Arabidopsis* overexpressing *ENOD40* and protoplasts transiently overexpressing *ENOD40* have a subtle, but nonetheless statistically significant, smaller cell size when measured by flow cytometry; the same phenomenon is observed when non-transformed protoplasts are treated with an ENOD40 synthetic peptide (Guzzo et al. 2005).

Sucrose synthase, which is a very abundant enzyme in nodules, could be modified by ENOD40 to catalyze the conversion of sucrose into glucose (and fructose), thereby providing the energy and carbon-based molecule required by the bacterioïd in the nodule and making *ENOD40* a key regulator of sink strength (Rohrig et al. 2002). The expression of *ENOD40* in other tissues could still be associated with sucrose synthase since sucrose is also used as a precursor of cellulose synthesis. Notably, *ENOD40* expression is observed in

rapidly growing (and dividing) tissues such as meristems, fruits and ovules (Asad et al. 1994; Fang and Hirsch 1998; Flemetakis et al. 2000), in which the balance between cellulose synthesis and sucrose degradation must be tightly regulated. Consistent with the above mentioned results, we have also observed that *ENOD40* mRNA levels are strongly induced following fertilization and in young developing *Solanum chacoense* fruits (H. Germain and D.P. Matton, unpublished results). Based upon the data mentioned above, it seems that *ENOD40* can act as a peptide (although a possible action as a RNA molecule cannot be ruled out), but there is a lack of evidence with regard to its capacity to act non cell-autonomously, which suggests that ENOD40 might be more alike to an allosteric regulator of enzymes than a signaling peptide. Therefore, we believe that at this time *ENOD40* should not be classified as a "signaling peptide" since its mobility has not been demonstrated, but it is clearly a small bioactive peptide that can modulate enzymatic activity.

Peptides with no known function

The large NCR gene family

The *NCR* (nodule cystein-rich) gene family represents (to our knowledge) the largest family of small putative signaling peptides. Members of the family were discovered as early nodulin (*ENOD*) genes in *Pisum sativum* more than 15 years ago and were believed to be putative metal-binding protein (Scheres et al. 1990; Kardailsky et al. 1993). These genes were expressed during the development of indeterminate nodule. Members of the family were later discovered as genes expressed during nodule development through EST sequencing (Gyorgyey et al. 2000). Bioinformatic searches revealed that the family was much larger than previously believed when more than 300 cysteine cluster proteins (CPP) were identified (Fedorova et al. 2002; Mergaert et al. 2003), several of which were part of the NCR family. All members have mRNA levels that are significantly increased in developing nodules, encode small peptides (approx. 70 amino acids), have a signal peptide

at the N-terminal of the protein, possess a conserved cysteine motif. All members are found in species forming indeterminate nodules and none are found in *Arabidopsis* or other legume species outside the galegoid group. Sequence identity between the family members is quite variable, ranging from 20% to 70%, but intriguingly, the signal peptide is highly conserved even between distantly related members and for at least two members, the peptide is targeted to the extracellular space. *NCR* genes are not only abundant at the genomic level, but are also highly expressed and estimated to represent 4.6% of the total nodule transcriptome. Macroarray analyses revealed that two genes, a calmodulin-like protein (containing a signal peptide) and a signal peptide peptidase are co-expressed with NCRs. In animal cells, signal peptide peptidase catalyzes intramembrane proteolysis of some signal peptides after they have been cleaved from the preprotein (Weihofen et al. 2002).

Recently, members of the family were also found in a non-galegoid species (soybean) using other BLAST algorithms (Graham et al. 2004), and the family was even extended further to a large variety of plant species using successive iteration of the hidden Markov model algorithm (Silverstein et al. 2005) (referred to as defensin-like genes). The extended family refers to short proteins with a peptide signal and conserved cysteine motifs. This rather large description also includes LCR (low-molecular-weight cysteine rich) and SCRL (S-locus cystein-rich-like) proteins (Vanoosthuyse et al. 2001), which were identified as homologs of pollen coat proteins (Doughty et al. 2000), SCR/SP11 (Watanabe et al. 2000), and S-protein homologs (Ride et al. 1999).

Until now no functions have been ascribed to the nodule-specific NCR gene family or to any of its members (except SP11/SCR and their alleles). Given the fact that there are hundreds of members in the family, some of which display high homology among themselves, it is unlikely that phenotype searches through knockout, knock-down or overexpressing mutant lines will be successful. As suggested by Mergaert et al. (2003), it is possible, given the expression pattern and the sequence variability of NCRs (and others)

that they may act as a wide-spectrum antimicrobial cocktail or that they could be involved in the complex dialog between plants and bacteria, which is triggered during the course of nodule formation or other processes.

4 kDa peptide (leginsulin)

Barbashov and Egorov (1990) found that insulin from pig kidney could inhibit DNA synthesis in human cells, and identified a soybean protein capable of competing with insulin and insulin growth factor-I (IGF-1) for binding to a plasma membrane protein (mediating this inhibition of DNA synthesis); the soybean protein immunoreacted with the anti-insulin antibody, suggesting that plants would also express insulin-like proteins. A soybean protein that binds to insulin and insulin-like growth factor was then identified and was shown to have orthologs with similar binding activity in different species, including cowpea, mung bean, azuki bean and lupin (Komatsu and Hirano 1991).

This protein, the basic 7S globulin (Bg), is a cysteine-rich glycoprotein composed of two subunits (16 and 27 kDa) (Kagawa and Hirano 1989), which is released when soybean seeds are immersed in hot water (40–60 °C) and is believed to have tyrosine kinase activity (Komatsu et al. 1994), although no domains characteristic of protein kinases are found in this protein. Since the tyrosine kinase activity detected was measured in a crude extract preparation, it is strongly suggested that activity is derived from a contaminating or co-migrating protein. Using affinity chromatography with immobilized Bg as a bait, Watanabe et al. (1994) eluted and identified by mass spectrometry a 4 kDa peptide (which was originally named leginsulin and later renamed 4 kDa peptide) from a soybean radicle extract. The binding between the 4 kDa peptide and Bg is inhibited with 0.1 nmol/L insulin, confirming that the 4 kDa peptide and insulin are structural analogs. Edman degradation revealed that the 4 kDa peptide contains 37 amino acids, six of which are cysteine residues, expected (by mass spectrometry) to form three disulfide bonds (Watanabe et al. 1994). The 357-bp ORF, which was identified from a soybean cDNA library, encodes a 119-amino-acid prepropeptide with a signal peptide and a predicted cleavage site between amino acids

19 and 20 (IEA-AD) (H. Germain, personal observation, SignalP 3.0), while amino acids 25–56 encode the 4 kDa peptide (Watanabe et al. 1994). The native peptide, but not the oxidized form, has a strong stimulatory effect on Bg phosphorylation activity (Watanabe et al. 1994), supporting the importance of disulfide bonds in the integrity of the protein.

At physiological levels the 4 kDa peptide is thought to have a role in cellular proliferation. Transgenic callus overexpressing the 4 kDa peptide divide much more rapidly than wild-type. In a similar manner, when the peptide is added to carrot cell suspension cultures, they divide faster than wild-type suspensions in a concentration-dependent manner (Yamazaki et al. 2003). Hanada et al. (2003) used size exclusion chromatography and surface plasmon resonance, combined with alanine-scanning mutagenesis, to investigate the binding mechanism between the 4 kDa peptide and Bg. Their finding revealed that Bg purifies and binds to the 4 kDa peptide as a 80 kDa dimer. The surface plasmon resonance experiment showed that the *K*d of the interaction between wild-type 4 kDa peptide and Bg was 1.86 Å~ 10-8 mol/L (or 18.6 nmol/L) (Hanada et al. 2003). Alanine scanning revealed that a mutation in the C-terminus of the peptide has a higher impact on the binding interaction than a mutation in the N-terminus, suggesting that this region of the hormone-like peptide is more important for the protein–protein interaction. It was later shown that Bg has two sites in the 27 kDa subunit that interacts with the 4 kDa peptide and that the same region of the Bg protein is recognized by the animal insulin (Hanada and Hirano 2004). A Bg ortholog from carrot is glycosylated at four sites and the interaction of this protein with the 4 kDa peptide is 20-fold higher when the protein is deglycosylated (Shang et al. 2005). Furthermore, these glycosylations are necessary for the secretion of the 43 kDa protein (the 16 and 27 kDa subunits). It is therefore possible that glycosylation could facilitate the proper folding of the protein and must later be removed once the protein as been secreted (Shang et al. 2005).

The 4 kDa peptide contains a signal peptide, is synthesized from a larger precursor, and has pairs of disulfide bonds, features that make it a good candidate to act as a small

putative signaling peptide. On the other hand, the Bg protein has no known downstream effectors, and no clear biological activities have been reported for these 7S globulin proteins, except as seed storage proteins, suggesting that its earlier classification as a receptor-like protein by Watanabe et al. (1994) is questionable. Functional analyses in transgenic plants overexpressing or impaired in the expression of the 4 kDa peptide should provide some details about its role in plants.

Conclusion

The known putative signaling peptides already add up to a few hundred peptides and surely many more will be discovered. It was initially thought that each receptor kinase would bind one ligand. Since *CLE40* can act as a signal in the SAM through *CLV1*, and *CLV3*, *CLE19* and *CLE40* all cause root meristem consumption (Fiers et al. 2004), and the solanaceous systemin receptor can also bind brassinosteroids (Scheer et al. 2003), there are doubts about this one-to-one hypothesis. This, combined with the conspicuous observation that the members of several peptide families (RALF, IDA, CLE, DVL/ROT, HT, Egg apparatus) have highly conserved C-terminal domains, shared by all the members within a family, suggest that related peptides within a family that have most probably arisen through gene conversion and (or) duplication events, may have acquired spatial and temporal specificity and interact with a more limited number of receptors.

The field of peptide signaling and the number of putative signaling peptides in plants has seen tremendous growth over the last decade. It has gone from its birth, with the discovery of systemin, to the point where several orphan putative ligands have now been found either through biochemical, genetical or database mining. Improved prediction algorithms should allow more peptides to be found from the available databases, but success will be hampered by the fact that short cDNAs are still (for good reasons) being discarded during library construction and often end-up being underrepresented. Another bioinformatics approach enabling identification of small peptide or non-coding RNAs is

comparison of large EST sets with annotated genomic resources, allowing the retrieval of ESTs that correspond to unannotated genes (Riano-Pachon et al. 2005). New methods such as mass spectrometry of apoplastic fluids or sap have already started to yield fruitful results (Hoffmann-Benning et al. 2002; Boudart et al. 2005) and not only permits the finding of new peptides, but also provides invaluable information with regards to their post-translational modification. In the case where large peptide families are involved, functional characterization will have to rely on the production of double or triple knockout mutants (or even more) or targeted interference of conserved domains. High throughput biochemical approaches involving bioassays or high density protein arrays could speed-up the process of assigning peptides to their binding partners.

Acknowledgements

We thank members of the Matton laboratory for sharing unpublished results. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Canada Research Chair program. H.G. is the recipient of Ph.D. fellowships from NSERC and from Le Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT, Québec). E.C. is the recipient of a Ph.D. fellowship from NSERC. D.P.M. holds a Canada Research Chair in Functional Genomics and Plant Signal Transduction.

References

1. Asad, S., Fang, Y., Wycoff, K.L., and Hirsch, A.M. 1994. Isolation and characterization of cDNA and genomic clones of MsENOD40; transcript are detected in meristematic cells of alfalfa. Protoplasma, **183**: 10–23.

2. Barbashov, S.F., and Egorov, T.A. 1990. The use of monoclonal antibodies against insulin for isolation of proteins inhibiting the cell growth (translation from Russian). Mol. Biol. (Mosk.), **24**: 953–961.

3. Becraft, P.W. 2002. Receptor kinase signaling in plant development. Annu. Rev. Cell Dev. Biol. **18**: 163–192.

4. Bedinger, P., Reeves, A., Arthur-Asmah, R., Parsons, R., Clark, M., Pearce, G., and Ryan, C.A. 2004. LePRALF: a potential pollen growth regulator. *In* Frontiers in Sexual Plant Reproduction II meeting, SUNY, Albany, N.Y., 15–17 October 2004. p. 72.

5. Bendtsen, J.D., Jensen, L.J., Blom, N., Von Heijne, G., and Brunak, S. 2004*a*. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel. **17**: 349–356.

6. Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. 2004*b*. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340**: 783–795.

7. Billington, T., Pharmawati, M., and Gehring, C.A. 1997. Isolation and immunoaffinity purification of biologically active plant natriuretic peptide. Biochem. Biophys. Res. Commun. **235**: 722–725.

8. Bleecker, A.B., and Patterson, S.E. 1997. Last exit: senescence, abscission, and meristem arrest in *Arabidopsis*. Plant Cell, **9**: 1169–1179.

9. Boller, T. 2005. Peptide signalling in plant development and self/non-self perception. Curr. Opin. Cell Biol. **17**: 116–122.

10. Bonello, J.-F., Sevilla-Lecoq, S., Berne, A., and Rusueño, M.-C. 2002. Esr proteins are secreted by the cells of the embryo surrounding region. J. Exp. Bot. **53**: 1559–1568.

11. Boudart, G., Jamet, E., Rossignol, M., Lafitte, C., Borderies, G., Jauneau, A., Esquerre-Tugaye, M.T., and Pont-Lezica, R. 2005. Cell wall proteins in apoplastic fluids of Arabidopsis thaliana rosettes: identification by mass spectrometry and bioinformatics. Proteomics, **5**: 212–221.

12. Butenko, M.A., Patterson, S.E., Grini, P.E., Stenvik, G.E., Amundsen, S.S., Mandal, A., and Aalen, R.B. 2003. Inflorescence deficient in abscission controls floral organ abscission in *Arabidopsis* and identifies a novel family of putative ligands in plants. Plant Cell, **15**: 2296–2307.

13. Casamitjana-Martinez, E., Hofhuis, H.F., Xu, J., Liu, C.M., Heidstra, R., and Scheres, B. 2003. Root-specific CLE19 overexpression and the sol1/2 suppressors implicate a CLV-like pathway in the control of *Arabidopsis* root meristem maintenance. Curr. Biol. **13**: 1435–1441.

14. Casson, S.A., Chilley, P.M., Topping, J.F., Evans, I.M., Souter, M.A., and Lindsey, K. 2002. The POLARIS gene of Arabidopsis encodes a predicted peptide required for correct root growth and leaf vascular patterning. Plant Cell, **14**: 1705–1721.

15. Chilley, P. 2003. Polypeptide hormones: signaling molecules in plants. Vitam. Horm. **66**: 317–344.

16. Clark, S.E. 2001. Cell signalling at the shoot meristem. Nat. Rev. Mol. Cell Biol. **2**: 276–284.

17. Clark, S.E., Running, M.P., and Meyerowitz, E.M. 1995. *Clavata3* is a specific regulator of shoot and floral meristem development affecting the same processes as *Clavata1*. Development, **121**: 2057–2067.

18. Cleland, R.E. 1995. Auxin and cell elongation. *In* Plant hormones: physiology, biochemistry and molecular biology. *Edited by* P.J. Davies P.J.Kluwer Academic Publishers, Dordrecht, Netherlands. pp. 219–222.

19. Cock, J.M., and McCormick, S. 2001. A large family of genes that share homology with CLAVATA3. Plant Physiol. **126**: 939–942.

20. Collmer, A. 1998. Determinants of pathogenicity and avirulence in plant pathogenic bacteria. Curr. Opin. Plant Biol. **1**: 329–335.

21. Constabel, C.P., Yip, L., and Ryan, C.A. 1998. Prosystemin from potato, black nightshade, and bell pepper: primary structure and biological activity of predicted systemin polypeptides. Plant Mol. Biol. **36**: 55–62.

22. Crespi, M.D., Jurkevitch, E., Poiret, M., d'Aubenton-Carafa, Y., Petrovics, G., Kondorosi, E., and Kondorosi, A. 1994. enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. EMBO J. **13**: 5099–5112.

23. Davis, E.L., Hussey, R.S., Baum, T.J., Bakker, J., Schots, A., Rosso, M.N., and Abad, P. 2000. Nematode parasitism genes. Annu. Rev. Phytopathol. **38**: 365.

24. de Bold, A.J., Borenstein, B.H., Veress, A.T., and Sonnenberg, H. 1981. A rapid and potent natriuretic response to intravenous injection of atrial myocardial extract in rats. Life Sci. **28**: 89–94.

25. Dharmasiri, N., Dharmasiri, S., and Estelle, M. 2005. The F-box protein TIR1 is an auxin receptor. Nature (London), **435**: 441–445.

26. Doughty, J., Wong, H.Y., and Dickinson, H.G. 2000. Cystein-rich pollen coat protein (PCPs) and their interactions with stigmatic S (incompatibility) and S-related proteins in *Brassica*: putative roles in SI and pollination. Ann. Bot. (London), **85**: 161–169.

27. Dresselhaus, T., Amien, S., Marton, M., Strecke, A., Brettschneider, R., and Cordts, S. 2005. TRANSPARENT LEAF AREA1 encodes a secreted proteolipid required for anther maturation, morphogenesis, and differentiation during leaf development in maize. Plant Cell, **17**: 730–745.

28. Eden, S., Rohatgi, R., Podtelejnikov, A.V., Mann, M., and Kirschner, M.W. 2002. Mechanism of regulation of WAVE1-induced actin nucleation by Rac1 and Nck. Nature, **418**: 790–793.

29. Escobar, N.M., Haupt, S., Thow, G., Boevink, P., Chapman, S., and Oparka, K. 2003. High-throughput viral expression of cDNA-green fluorescent protein fusions reveals novel subcellular addresses and identifies unique proteins that interact with plasmodesmata. Plant Cell, **15**: 1507–1523.

30. Fang, Y., and Hirsch, A.M. 1998. Studying early nodulin gene ENOD40 expression and induction by nodulation factor and cytokinin in transgenic alfalfa. Plant Physiol. **116**: 53–68.

31. Fedorova, M., van de Mortel, J., Matsumoto, P.A., Cho, J., Town, C.D., VandenBosch, K.A., Gantt, J.S., and Vance, C.P. 2002. Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. Plant Physiol. **130**: 519–537.

32. Felix, G., and Boller, T. 1995. Systemin induces rapid ion fluxes and ethylene biosynthesis in *Lycopersicon peruvianum* cells. Plant J. **7**: 381–389.

33. Fiers, M., Hause, G., Boutilier, K., Casamitjana-Martinez, E., Weijers, D., Offringa, R., van der Geest, L., van Lookeren Campagne, M., and Liu, C.M. 2004. Mis-expression of the CLV3/ESR-like gene *CLE19* in *Arabidopsis* leads to a consumption of root meristem. Gene, **327**: 37–49.

34. Fiers, M., Golemiec, E., Xu, J., van der Geest, L., Heidstra, R., Stiekema, W., and Liu, C.M. 2005. The 14-amino acid CLV3, CLE19, and CLE40 peptides trigger consumption of the root meristem in *Arabidopsis* through a CLAVATA2-dependent pathway. Plant Cell, **17**: 2542–2553.

35. Flemetakis, E., Kavroulakis, N., Quaedvlieg, N.E., Spaink, H.P., Dimou, M., Roussis, A., and Katinakis, P. 2000. *Lotus japonicus* contains two distinct ENOD40 genes that are expressed in symbiotic, nonsymbiotic, and embryonic tissues. Mol. Plant Microbe Interact. **13**: 987–994.

36. Fletcher, J.C., Brand, U., Running, M.P., Simon, R., and Meyerowitz, E.M. 1999. Signaling of cell fate decisions by Clavata3 in *Arabidopsis* shoot meristems. Science (Washigton, D.C.), **283**: 1911–1914.

37. Frank, M.J., and Smith, L.G. 2002. A small, novel protein highly conserved in plants and animals promotes the polarized growth and division of maize leaf epidermal cells. Curr. Biol. **12**: 849–853.

38. Frank, M.J., Cartwright, H.N., and Smith, L.G. 2003. Three *Brick* genes have distinct functions in a common pathway promoting polarized cell division and cell morphogenesis in the maize leaf epidermis. Development, **130**: 753–762.

39. Franssen, H.J., and Bisseling, T. 2001. Peptide signaling in plants. Proc. Natl. Acad. Sci. U.S.A., **98**: 12855–12856.

40. Gallagher, K., and Smith, L.G. 2000. Roles for polarity and nuclear determinants in specifying daughter cell fates after an asymmetric cell division in the maize leaf. Curr. Biol. **10**: 1229–1232.

41. Gao, B., Allen, R., Maier, T., Davis, E.L., Baum, T.J., and Hussey, R.S. 2001. Identification of putative parasitism genes expressed in the oesophageal gland cells of the soybean cyst nematode *Heterodera glycines*. Mol. Plant Microbe Interact. **14**: 1247–1254.

42. Gehring, C.A., Khalid, K.M., Toop, T., and Donald, J.A. 1996. Rat natriuretic peptide binds specifically to plant membranes and induces stomatal opening. Biochem. Biophys. Res. Commun. **228**: 739–744.

43. Germain, H., Chevalier, E., Caron, S., and Matton, D.P. 2005*a*. Characterization of five *RALF*-like genes from *Solanum chacoense* provides support for a developmental role in plants. Planta, **220**: 447–454.

44. Germain, H., Rudd, S., Zotti, C., Caron, S., O'Brien, M., Chantha, S.C., Lagace, M., Major, F., and Matton, D.P. 2005*b*. 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**: 515–532.

45. Graham, M.A., Silverstein, K.A., Cannon, S.B., and VandenBosch, K.A. 2004. Computational identification and characterization of novel genes from legumes. Plant Physiol. **135**: 1179–1197.

46. Gray-Mitsumune, M., and Matton, D.P. 2006. The Egg apparatus 1 gene from maize is a member of a large gene family found in both monocots and dicots. Planta.

47. Green, T.R., and Ryan, C.A. 1972. Wound-induced proteinase inhibitor in plant leaves: a possible defense mechanism against insects. Science (Washington, D.C.), **175**: 776–777.

48. Guzzo, F., Portaluppi, P., Grisi, R., Barone, S., Zampieri, S., Franssen, H., and Levi, M. 2005. Reduction of cell size induced by enod40 in *Arabidopsis thaliana*. J. Exp. Bot. **56**: 507–513.

49. Gyorgyey, J., Vaubert, D., Jimenez-Zurdo, J.I., Charon, C., Troussard, L., Kondorosi, A., and Kondorosi, E. 2000. Analysis of *Medicago truncatula* nodule expressed sequence tags. Mol. Plant–Microbe Interact. **13**: 62–71.

50. Haffani, Y.Z., Silva, N.F., and Goring, D.R. 2004. Receptor kinase signalling in plants. Can. J. Bot. **82**: 1–15.

51. Hanada, K., and Hirano, H. 2004. Interaction of a 43-kDa receptor-like protein with a 4-kDa hormone-like peptide in soybean. Biochemistry, **43**: 12105–12112.

52. Hanada, K., Nishiuchi, Y., and Hirano, H. 2003. Amino acid residues on the surface of soybean 4-kDa peptide involved in the interaction with its binding protein. Eur. J. Biochem. **270**: 2583–2592.

53. Hardin, S.C., Tang, G.Q., Scholz, A., Holtgraewe, D., Winter, H., and Huber, S.C. 2003. Phosphorylation of sucrose synthase at serine 170: occurrence and possible role as a signal for proteolysis. Plant J. **35**: 588–603.

54. Haruta, M., and Constabel, C.P. 2003. Rapid alkalinization factors in poplar cell cultures. Peptide isolation, cDNA cloning, and differential expression in leaves and methyl jasmonate-treated cells. Plant Physiol. **131**: 814–823.

55. He, Z., Wang, Z.Y., Li, J., Zhu, Q., Lamb, C., Ronald, P., and Chory, J. 2000. Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. Science, **288**: 2360–2363.

56. Hobe, M., Muller, R., Grunewald, M., Brand, U., and Simon, R. 2003. Loss of CLE40, a protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving in *Arabidopsis*. Dev. Genes Evol. **213**: 371–381.

57. Hoffmann-Benning, S., Gage, D.A., McIntosh, L., Kende, H., and Zeevaart, J.A. 2002. Comparison of peptides in the phloem sap of flowering and non-flowering Perilla and lupine plants using microbore HPLC followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Planta, **216**: 140–147.

58. Inoue, T., Higuchi, M., Hashimoto, Y., Seki, M., Kobayashi, M., Kato, T., Tabata, S., Shinozaki, K., and Kakimoto, T. 2001. Identification of CRE1 as a cytokinin receptor from Arabidopsis. Nature (London), **409**: 1060–1063.

59. Joshi, C.P., Zhou, H., Huang, X., and Chiang, V.L. 1997. Context sequences of translation initiation codon in plants. Plant Mol. Biol. **35**: 993–1001.

60. Kagawa, H., and Hirano, H. 1989. Sequence of a cDNA encoding soybean basic 7S globulin. Nucleic Acids Res. **17**: 8868.

61. Kardailsky, I., Yang, W.C., Zalensky, A., van Kammen, A., and Bisseling, T. 1993. The pea late nodulin gene PsNOD6 is homologous to the early nodulin genes PsENOD3/14 and is expressed after the leghaemoglobin genes. Plant Mol. Biol. **23**: 1029–1037.

62. Kim, H.U., Cotter, R., Johnson, S., Senda, M., Dodds, P., Kulikauska, R., Tang, W., Ezcura, I., Herzmark, P., and McCormick, S. 2002. New pollen-specific receptor kinases identified in tomato, maize and Arabidopsis: the tomato kinases show

overlapping but distinct localization patterns on pollen tubes. Plant Mol. Biol. **50**: 1–16.

63. Komatsu, S., and Hirano, H. 1991. Plant basic 7 S globulin-like proteins have insulin and insulin-like growth factor binding activity. FEBS Lett. **294**: 210–212.

64. Komatsu, S., Koshio, O., and Hirano, H. 1994. Protein kinase activity and insulin-binding activity in plant basic 7S globulin. Biosci. Biotechnol. Biochem. **58**: 1705–1706.

65. Kouchi, H., and Hata, S. 1993. Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. Mol. Gen. Genet. **238**: 106–119.

66. Kouchi, H., Takane, K., So, R.B., Ladha, J.K., and Reddy, P.M. 1999. Rice ENOD40: isolation and expression analysis in rice and transgenic soybean root nodules. Plant J. **18**: 121–129.

67. Kozak, M. 1999. Initiation of translation in prokaryotes and eukaryotes. Gene, **234**: 187–208.

68. Larsen, K. 2003. Molecular cloning and characterization of a cDNA encoding a ryegrass (*Lolium perenne*) ENOD40 homologue. J. Plant Physiol. **160**: 675–687.

69. Lenhard, M., and Laux, T. 2003. Stem cell homeostasis in the *Arabidopsis* shoot meristem is regulated by intercellular movement of CLAVATA3 and its sequestration by CLAVATA1. Development, **130**: 3163–3173.

70. Ludidi, N.N., Heazlewood, J.L., Seoighe, C., Irving, H.R., and Gehring, C.A. 2002. Expansin-like molecules: novel functions derived from common domains. J. Mol. Evol. **54**: 587–594.

71. Marton, M.L., Cordts, S., Broadhvest, J., and Dresselhaus, T. 2005. Micropylar pollen tube guidance by egg apparatus 1 of maize. Science (Washington, D.C.), **307**: 573–576.

72. Maryani, M.M., Shabala, S.N., and Gehring, C.A. 2000. Plant natriuretic peptide immunoreactants modulate plasma-membrane H(+) gradients in Solanum tuberosum L. leaf tissue vesicles. Arch. Biochem. Biophys. **376**: 456–458.

73. Maryani, M.M., Bradley, G., Cahill, D.M., and Gehring, C.A. 2001. Natriuretic peptides and immunoreactants modify osmoticum-dependent volume changes in *Solanum tuberosum* L. mesophyll cell protoplasts. Plant Sci. **161**: 443–452.

74. Maryani, M.M., Morse, M.V., Bradley, G., Irving, H.R., Cahill, D.M., and Gehring, C.A. 2003. In situ localization associates biologically active plant natriuretic peptide immuno-analogues with conductive tissue and stomata. J. Exp. Bot. **54**: 1553–1564.

75. Matvienko, M., Van de Sande, K., Yang, W.C., van Kammen, A., Bisseling, T., and Franssen, H. 1994. Comparison of soybean and pea ENOD40 cDNA clones representing genes expressed during both early and late stages of nodule development. Plant Mol. Biol. **26**: 487–493.

76. McClure, B. 2004. S-RNase and SLF determine S-haplotype-specific pollen recognition and rejection. Plant Cell, **16**: 2840–2847.

77. McClure, B., Mou, B., Canevascini, S., and 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**: 13548–13553.

78. McClure, B.A., Cruz-Garcia, F., Beecher, B., and Sulaman, W. 2000. Factors affecting inter- and intra-specific pollen rejection in *Nicotiana*. Ann. Bot. (London), **85**: 113–123.

79. McCormick, S., and Yang, H. 2005. Is there more than one way to attract a pollen tube? Trends Plant Sci. **10**: 260–263.

80. Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A., and Kondorosi, E. 2003. A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. Plant Physiol. **132**: 161–173.

81. Meyerowitz, E.M. 1999. Plants, animals and the logic of development. Trends Cell Biol. **9**: M65–M68.

82. Morse, M., Pironcheva, G., and Gehring, C. 2004. AtPNP-A is a systemically mobile natriuretic peptide immunoanalogue with a role in *Arabidopsis thaliana* cell volume regulation. FEBS Lett. **556**: 99–103.

83. Murfett, J., Strabala, T.J., Zurek, D.M., Mou, B., Beecher, B., and McClure, B.A. 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**: 943–958.

84. Muschietti, J., Eyal, Y., and McCormick, S. 1998. Pollen tube localization implies a role in pollen-pistil interactions for the tomato receptor-like protein kinases LePRK1 and LePRK2. Plant Cell, **10**: 319–330.

85. Nakajima, K., Sena, G., Nawy, T., and Benfey, P.N. 2001. Intercellular movement of the putative transcription factor SHR in root patterning. Nature (London), **413**: 307–311.

86. Narita, N.N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., Goodrich, J., and Tsukaya, H. 2004. Overexpression of a novel small peptide

ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in Arabidopsis thaliana. Plant J. **38**: 699–713.

87. Narvaez-Vasquez, J., Pearce, G., and Ryan, C.A. 2005. The plant cell wall matrix harbors a precursor of defense signaling peptides. Proc. Natl. Acad. Sci. U.S.A., **102**: 12974–12977.

88. Nasrallah, J.B. 2002. Recognition and rejection of self in plant reproduction. Science (Washington, D.C.), **296**: 305–308.

89. Nasrallah, J.B. 2005. Recognition and rejection of self in plant self-incompatibility: comparisons to animal histocompatibility. Trends Immunol. **26**: 412–418.

90. Nembaware, V., Seoighe, C., Sayed, M., and Gehring, C. 2004. A plant natriuretic peptide-like gene in the bacterial pathogen Xanthomonas axonopodis may induce hyper-hydration in the plant host: a hypothesis of molecular mimicry. BMC Evol. Biol. **4**: 10.

91. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int. J. Neural Syst. **8**: 581–599.

92. Nishihama, R., Jeong, S., DeYoung, B., and Clark, S.E. 2003. Retraction. Science (Washington, D.C.), **300**: 1370.

93. O'Brien, M., Kapfer, C., Major, G., Laurin, M., Bertrand, C., Kondo, K., Kowyama, Y., and Matton, D.P. 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**: 985–996.

94. Olsen, A.N., and Skriver, K. 2003. Ligand mimicry? Plant-parasitic nematode polypeptide with similarity to CLAVATA3. Trends Plant Sci. **8**: 55–57.

95. Olsen, A.N., Mundy, J., and Skriver, K. 2002. Peptomics, identification of novel cationic *Arabidopsis* peptides with conserved sequence motifs. In Silico Biol. **2**: 441–451.

96. Oparka, K.J., Roberts, A.G., Boevink, P., Santa Cruz, S., Roberts, I., Pradel, K.S., Imlau, A., Kotlizky, G., Sauer, N., and Epel, B. 1999. Simple, but not branched, plasmodesmata allow the nonspecific trafficking of proteins in developing tobacco leaves. Cell, **97**: 743–754.

97. Opsahl-Ferstad, H.G., Le Deunff, E., Dumas, C., and Rogowsky, P.M. 1997. ZmEsr, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. Plant J. **12**: 235–246.

98. Pearce, G., and Ryan, C.A. 2003. Systemic signaling in tomato plants for defense against herbivores. Isolation and characterization of three novel defense-signaling glycopeptide hormones coded in a single precursor gene. J. Biol. Chem. **278**: 30044–30050.

99. Pearce, G., Strydom, D., Johnson, S., and Ryan, C.A. 1991. A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science, **253**: 895–898.

100. Pearce, G., Moura, D.S., Stratmann, J., and Ryan, C.A. 2001*a*. Production of multiple plant hormones from a single polyprotein precursor. Nature (London), **411**: 817–820.

101. Pearce, G., Moura, D.S., Stratmann, J., and Ryan, C.A., Jr. 2001*b*. RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development. Proc. Natl. Acad. Sci. U.S.A., **98**: 12843–12847.

102. Pharmawati, M., Billington, T., and Gehring, C.A. 1998. Stomatal guard cell responses to kinetin and natriuretic peptides are cGMP-dependent. Cell. Mol. Life Sci. **54**: 272–276.

103. Pharmawati, M., Shabala, S.N., Newman, I.A., and Gehring, C.A. 1999. Natriuretic peptides and cGMP modulate K+, Na+, and H+ fluxes in *Zea mays* roots. Mol. Cell Biol. Res. Commun. **2**: 53–57.

104. Razem, F.A., El-Kereamy, A., Abrams, S.R., and Hill, R.D. 2006. The RNA-binding protein FCA is an abscisic acid receptor. Nature (London), **439**: 290–294.

105. Riano-Pachon, D.M., Dreyer, I., and Mueller-Roeber, B. 2005. Orphan transcripts in *Arabidopsis thaliana*: identification of several hundred previously unrecognized genes. Plant J. **43**: 205–212.

106. Ride, J.P., Davies, E.M., Franklin, F.C., and Marshall, D.F. 1999. Analysis of *Arabidopsis* genome sequence reveals a large new gene family in plants. Plant Mol. Biol. **39**: 927–932.

107. Roberts, J.A., Elliott, K.A., and Gonzalez-Carranza, Z.H. 2002. Abscission, dehiscence, and other cell separation processes. Annu. Rev. Plant Biol. **53**: 131–158.

108. Rohrig, H., Schmidt, J., Miklashevichs, E., Schell, J., and John, M. 2002. Soybean ENOD40 encodes two peptides that bind to sucrose synthase. Proc. Natl. Acad. Sci. U.S.A., **99**: 1915–1920.

109. Rohrig, H., John, M., and Schmidt, J. 2004. Modification of soybean sucrose synthase by S-thiolation with ENOD40 peptide A. Biochem. Biophys. Res. Commun. **325**: 864–870.

110. Rojo, E., Sharma, V.K., Kovaleva, V., Raikhel, N.V., and Fletcher, J.C. 2002. CLV3 is localized to the extracellular space, where it activates the *Arabidopsis* CLAVATA stem cell signaling pathway. Plant Cell, **14**: 969–977.

111. Ryan, C.A., Pearce, G., Scheer, J., and Moura, D.S. 2002. Polypeptide hormones. Plant Cell, **14**: Suppl, S251–S264.

112. Schaller, G.E., and Bleecker, A.B. 1995. Ethylene-binding sites generated in yeast expressing the *Arabidopsis* ETR1 gene. Science, **270**: 1809–1811.

113. Scheer, J.M., and Ryan, C.A., Jr. 2002. The systemin receptor SR160 from Lycopersicon peruvianum is a member of the LRR receptor kinase family. Proc. Natl. Acad. Sci. U.S.A., **99**: 9585–9590.

114. Scheer, J.M., Pearce, G., and Ryan, C.A. 2003. Generation of systemin signaling in tobacco by transformation with the tomato systemin receptor kinase gene. Proc. Natl. Acad. Sci. U.S.A., **100**: 10114–10117.

115. Scheer, J.M., Pearce, G., and Ryan, C.A. 2005. LeRALF, a plant peptide that regulates root growth and development, specifically binds to 25 and 120 kDa cell surface membrane proteins of *Lycopersicon peruvianum*. Planta, **221**: 667–674.

116. Scheres, B., and Benfey, P.N. 1999. Asymmetric cell division in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50**: 505–537.

117. Scheres, B., van Engelen, F., van der Knaap, E., van de Wiel, C., van Kammen, A., and Bisseling, T. 1990. Sequential induction of nodulin gene expression in the developing pea nodule. Plant Cell, **2**: 687–700.

118. Shang, C., Sassa, H., and Hirano, H. 2005. The role of glycosylation in the function of a 48-kDa glycoprotein from carrot. Biochem. Biophys. Res. Commun. **328**: 144–149.

119. Sharma, V.K., Ramirez, J., and Fletcher, J.C. 2003. The Arabidopsis CLV3-like (CLE) genes are expressed in diverse tissues and encode secreted proteins. Plant Mol. Biol. **51**: 415–425.

120. Shiu, S.H., and 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**: 10763–10768.

121. Siezen, R.J., and Leunissen, J.A.M. 1997. Subtilases: the superfamily of subtilisin-like serine proteases. Protein Sci. **6**: 501–523.

122. Silverstein, K.A., Graham, M.A., Paape, T.D., and Vandenbosch, K.A. 2005. Genome organization of more than 300 defensin-like genes in *Arabidopsis*. Plant Physiol. **138**: 600–610.

123. Stratmann, J., Scheer, J., and Ryan, C.A. 2000. Suramin inhibits initiation of defense signaling by systemin, chitosan, and a beta-glucan elicitor in suspension-cultured Lycopersicon peruvianum cells. Proc. Natl. Acad. Sci. U.S.A., **97**: 8862–8867.

124. The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature (London), **408**: 796–815.

125. Topping, J.F., and Lindsey, K. 1997. Promoter trap markers differentiate structural and positional components of polar development in *Arabidopsis*. Plant Cell, **9**: 1713–1725.

126. Topping, J.F., Agyeman, F., Henricot, B., and Lindsey, K. 1994. Identification of molecular markers of embryogenesis in *Arabidopsis thaliana* by promoter trapping. Plant J. **5**: 895–903.

127. Torii, K.U. 2004. Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. Int. Rev. Cytol. **234**: 1–46.

128. Trotochaud, A.E., Hao, T., Wu, G., Yang, Z., and Clark, S.E. 1999. The CLAVATA1 receptor-like kinase requires CLAVATA3 for its assembly into a signaling complex that includes KAPP and a Rho-related protein. Plant Cell, **11**: 393–406.

129. Trotochaud, A.E., Jeong, S., and Clark, S.E. 2000. CLAVATA3, a multimeric ligand for the CLAVATA1 receptor-kinase. Science (Washington, D.C.), **289**: 613–617.

130. Ueguchi-Tanaka, M., Ashikari, M., Nakajima, M., Itoh, H., Katoh, E., Kobayashi, M., Chow, T.Y., Hsing, Y.I., Kitano, H., Yamaguchi, I., and Matsuoka, M. 2005. GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature (London), **437**: 693–698.

131. van de Sande, K., Pawlowski, K., Czaja, I., Wieneke, U., Schell, J., Schmidt, J., Walden, R., Matvienko, M., Wellink, J., van Kammen, A., Franssen, H., and Bisseling, T. 1996. Modification of phytohormone response by a peptide encoded by ENOD40 of legumes and a nonlegume. Science (Washington, D.C.), **273**: 370–373.

132. Vanoosthuyse, V., Miege, C., Dumas, C., and Cock, J.M. 2001. Two large *Arabidopsis thaliana* gene families are homologous to the *Brassica* gene superfamily that encodes pollen coat proteins and the male component of the self-incompatibility response. Plant Mol. Biol. **46**: 17–34.

133. Vesely, D.L., and Giordano, A.T. 1991. Atrial natriuretic peptide hormonal system in plants. Biochem. Biophys. Res. Commun. **179**: 695–700.

134. Vesely, D.L., Norris, J.S., Walters, J.M., Jespersen, R.R., and Baeyens, D.A. 1987. Atrial natriuretic prohormone peptides 1–30, 31–67, and 79–98 vasodilate the aorta. Biochem. Biophys. Res. Commun. **148**: 1540–1548.

135. Vesely, D.L., Gower, W.R., Jr., and Giordano, A.T. 1993. Atrial natriuretic peptides are present throughout the plant kingdom and enhance solute flow in plants. Am. J. Physiol. **265**: E465–E477.

136. Wang, X., Allen, R., Ding, X., Goellner, M., Maier, T., De Boer, J.M., Baum, T.J., Hussey, R.S., and Davis, E.L. 2001. Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. Mol. Plant Microbe Interact. **14**: 536.

137. Wang, X., Mitchum, M.G., Gao, B., Li, C., Diab, H., Baum, J.T., Hussey, R.S., and Davis, E.L. 2005. A parasitism gene from a plant-parasitic nematode with function similar to CLAVATA3/ESR (CLE) of *Arabidopsis thaliana*. Mol. Plant Pathol. **6**: 187.

138. Wang, Z.Y., and He, J.X. 2004. Brassinosteroid signal transduction–choices of signals and receptors. Trends Plant Sci. **9**: 91–96.

139. Watanabe, M., Ito, A., Takada, Y., Ninomiya, C., Kakizaki, T., Takahata, Y., Hatakeyama, K., Hinata, K., Suzuki, G., Takasaki, T., Satta, Y., Shiba, H., Takayama, S., and Isogai, A. 2000. Highly divergent sequences of the pollen self-incompatibility (S) gene in class-I S haplotypes of *Brassica campestris* (syn. rapa) L. FEBS Lett. **473**: 139–144.

140. Watanabe, Y., Barbashov, S.F., Komatsu, S., Hemmings, A.M., Miyagi, M., Tsunasawa, S., and Hirano, H. 1994. A peptide that stimulates phosphorylation of the plant insulin-binding protein. Isolation, primary structure and cDNA cloning. Eur. J. Biochem. **224**: 167–172.

141. Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z.Y., Xia, Y., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F., and Chory, J. 2000. Activation tagging in *Arabidopsis*. Plant Physiol. **122**: 1003–1013.

142. Weihofen, A., Binns, K., Lemberg, M.K., Ashman, K., and Martoglio, B. 2002. Identification of signal peptide peptidase, a presenilin-type aspartic protease. Science (Washington, D.C.), **296**: 2215–2218.

143. Wen, J., Lease, K.A., and Walker, J.C. 2004. DVL, a novel class of small polypeptides: overexpression alters *Arabidopsis* development. Plant J. **37**: 668–677.

144. Yamazaki, T., Takaoka, M., Katoh, E., Hanada, K., Sakita, M., Sakata, K., Nishiuchi, Y., and Hirano, H. 2003. A possible physiological function and the tertiary structure of a 4-kDa peptide in legumes. Eur. J. Biochem. **270**: 1269–1276.

145. Yan, Y., Smant, G., Stokkermans, J., Qin, L., Helder, J., Baum, T., Schots, A., and Davis, E. 1998. Genomic organization of four beta-1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications. Gene, **220**: 61–70.

146. Yang, H., Matsubayashi, Y., Nakamura, K., and Sakagami, Y. 1999. Oryza sativa PSK gene encodes a precursor of phytosulfokine-alpha, a sulfated peptide growth factor found in plants. Proc. Natl. Acad. Sci. U.S.A., **96**: 13560–13565.

147. Yang, W.C., Katinakis, P., Hendriks, P., Smolders, A., de Vries, F., Spee, J., van Kammen, A., Bisseling, T., and Franssen, H. 1993. Characterization of GmENOD40, a gene showing novel patterns of cell-specific expression during soybean nodule development. Plant J. **3**: 573–585.

Chapitre 3

A 6374 unigene set corresponding to low abundance transcripts expressed following fertilization in *Solanum chacoense* Bitt., and characterization of 30 receptor-like kinases

Hugo Germain, Stephen Rudd, Corine Zotti, Sébastien Caron, Sier-Ching Chantha, Marie Lagacé, François Major and Daniel P. Matton.

Publié dans Plant molecular Biology 59(3), 515-532 (2005)

Contribution des co-auteurs

Cette publication est le résultat d'un important effort d'équipe entrepris avant mon arrivé au laboratoire de Daniel. Sier-Ching Chantha et Marie Lagacé ont participé aux essais sur les premières banques de soustractions. Corine Zotti a realisé l'essentiel du travail de soustraction, de préparation des ADN plasmidiques et des réactions de séquençages. François Major en collaboration avec Sébastien Caron a permis l'élaboration de la base de données contenant l'ensemble des résultats. Sébastien Caron a aussi réalisé des PCR en temps réel. Stephen Rudd, initialement depuis l'Allemagne et ensuite de la Finlande a contribué à l'analyse des résultats. J'ai fait la grande majorité des PCR en temps réel, participé à l'analyse et l'interprétation des résultats avec Stephen Rudd et finalement j'ai rédigé le manuscript. Ha oui, j'oubliais le plus important; Daniel a eu l'idée de cette soustraction et a obtenu les fonds pour réaliser son projet.

A 6374 unigene set corresponding to low abundance transcripts expressed following fertilization in *Solanum chacoense* Bitt., and characterization of 30 receptor-like kinases

Hugo Germain[1], Stephen Rudd[2], Corine Zotti[1,3], Sébastien Caron[1], Sier-Ching Chantha[1], Marie Lagacé[1], François Major[4] and Daniel P. Matton[1]*.

[1]Institut de Recherche en Biologie Végétale (IRBV). Département de sciences biologiques, Université de Montréal, 4101 Sherbrooke Est, Montréal, QC, Canada, H1X 2B2. [2]Munich Information Center for Protein Sequences, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany and Centre for Biotechnology, Tykistökatu 6, FIN-20521, Turku, FINLAND. [3]Centre d'innovation Génome Québec et Université McGill, 740 avenue du Docteur-Penfield, Montréal, QC, Canada H3A 1A4. [4]Département d'informatique et de recherche opérationelle, Université de Montréal, C. P. 6128, H3C 3J7, Montréal, QC, Canada.

*Author for correspondence:

Tel: 1-514-872-3967

Fax: 1-514-872-9406

[redacted]

Source de financement

This work was supported by NSERC, FQRNT and by the Canada Research Chair program.

Key words

Virtual subtraction, negative selection, receptor-like kinase (RLK), expressed sequence tag (EST), *Solanum chacoense*, fertilization, embryo development, real-time PCR.

Genbank Accession numbers: DN976421 to DN984171 and DR398320.

Abbreviations:

Days after pollination, DAP; Expressed sequence tag, EST; ORK, Ovule Receptor Kinase; RLK, Receptor-like Kinase

Abstract

In order to characterize regulatory genes that are expressed in ovule tissues after fertilization we have undertaken an EST sequencing project in *Solanum chacoense*, a self-incompatible wild potato species. Two cDNA libraries made from ovule tissues covering embryo development from zygote to late torpedo-stage were constructed and plated at high density on nylon membranes. To decrease EST redundancy and enrich for transcripts corresponding to weakly expressed genes a self-probe subtraction method was used to select the colonies harboring the genes to be sequenced. 7741 good sequences were obtained and, from these, 6374 unigenes were isolated. Thus, the self-probe subtraction resulted in a strong enrichment in singletons, a decrease in the number of clones per contigs, and concomitantly, an enrichment in the total number of unigenes obtained (82%). To gain insights into signal transduction events occurring during embryo development all the receptor-like kinases (or protein receptor kinases) were analyzed by quantitative real-time RT-PCR. Interestingly, 28 out of the 30 RLK isolated were predominantly expressed in ovary tissues or young developing fruits, and 23 were transcriptionaly induced following fertilization. Thus, the self-probe subtraction did not select for genes weakly expressed in the target tissue while being highly expressed elsewhere in the plant. Of the receptor-like kinases (RLK) genes isolated, the leucine-rich repeat (LRR) family of RLK was by far the most represented with 25 members covering 11 LRR classes.

In order to gain information about the signal transduction elements and other genes that are weakly expressed during plant embryogenesis, we have undertaken a medium scale EST sequencing project using ovule tissues enriched (virtually subtracted) for rare mRNAs (Li and Thomas, 1998; Nelson et al., 1999). In classical EST sequencing projects ESTs are sequenced randomly (Van der Hoeven et al., 2002; Shimada et al., 2003; Fei et al., 2004), and the occurrence of a single EST species in the dataset is based on its expression level in the cell or tissue analyzed and, to a lesser extent, on its mRNA stability within the cell. While this approach is rapid and provides valuable information on differential gene expression levels between tissues, it requires the sequencing of large numbers of clones in order to obtain information on the weakly expressed genes. Therefore, using this method, weakly expressed genes can only be obtained at a great cost. Furthermore, genes that are only expressed in a narrow temporal window and genes expressed in only a few cells would be largely underrepresented in non-normalized libraries. Several methods can be used in an attempt to normalize the relative abundance of cDNAs within the library. For example, a simple method is to hybridize the target library with known commonly occurring abundant cDNAs and to select for only the clones that do not hybridize (Adams, 1995; Gong et al., 1997). This approach has been used with success, but requires some prior knowledge regarding cDNA frequencies. Other methods, such as oligonucleotide fingerprinting (OFP) (Radelof et al., 1998), hybridization kinetics approach (Soares et al., 1994), and methods involving the selection of mRNA based on a cap-trapping technique in order to increase the number of cDNAs containing the 5' end of the mRNA (Carninci, 2000) have also been used. Although they have proven to be useful and efficient, these techniques all have important limitations that have stood in the way of their widespread use. OFP is a very expensive method, due to the requirement of a large number of synthetically synthesized oligonucleotides, while hybridization kinetics can easily lead to artifacts in the subtracted library composition if all parameters are not perfectly controlled (Bonaldo, 1996). The high number of contigs that are found when one uses the method being developed by Carninci et al., suggest that this process is still in need of some improvements (Clark, 2001).

We used *Solanum chacoense*, a self-incompatible close relative of the potato and tomato which produces a large number of easily isolated ovules inside a fleshy tomato-like fruit, to dissect the molecular events occurring at precise time points following pollination and fertilization as well as during embryogenesis and seed development. The genomic data resulting from the EST project (obtained from two cDNA libraries covering the period from fertilization to late torpedo staged embryos), combined with the data generated from the completion of the sequencing of the Arabidopsis genome (Initiative, 2000), and the availability of more than 120,000 EST sequences analyzed in tomato (Van der Hoeven et al., 2002), were used to identify receptor-like kinase (RLK) involved in seed and embryo development. The presumed function of receptor-like kinases is to perceive extracellular stimuli and to communicate this information to an intracellular signaling pathway in order to ensure appropriate cellular responses. Information such as positional cues needed by the cell to ensure proper cell fate, or the signal given by the incoming pollen tube could be mediated by RLKs. RLKs are the most abundant kinase family in plants and, in Arabidopsis, they comprise more than 2% of the proteins characterized in its genome (Shiu and Bleecker, 2001). Despite the extensive array of developmental processes affected or regulated by RLKs, such as plant defense, meristem development, pollen/pistil interactions, hormone signaling, pollen development, cell morphogenesis and differentiation, organ shape and abscission, and somatic embryogenesis, only a limited number of RLKs have been characterized as being involved in reproductive development and none have been identified as having a specific function in zygotic embryo development (reviewed in (Becraft, 2002).

In this paper we have used a modified subtraction screen based on the virtual subtraction procedure of Li and Thomas (1998), and the negative screen of Nelson et al. (1999) as a method to enrich EST pools for rare messenger RNAs expressed during embryo development with a focus on receptor kinase signaling. Using a target tissue self-probe hybridization approach, only the clones that yielded a weak radioactive signal following hybridization were selected for sequencing. This method significantly increased the

frequency of unigene sequences. Furthermore, no bias towards non-target tissue expression was observed in the study of all the receptor-like kinases isolated as determined by an extensive quantitative real-time PCR analysis.

Material and methods

cDNA library construction and mass excision of the cDNA clones

All plant material was collected from *S. chacoense* genotype G4 (self-incompatibility alleles S12S14). For fertilization-related events, *S. chacoense* genotype V22 (self-incompatibility alleles S11S13) was used as the pollen donor. The target tissues used for the cDNA libraries synthesis were obtained by hand-dissection of ovaries from 2 to 6 days after pollination (DAP) (depericarped ovaries) or from 7 to 17 DAP ovules (fertilization occurs between 36 and 42h post-pollination). The total RNA from these tissues was isolated as described previously (Jones et al., 1985) and polyadenylated RNA was prepared by Oligo(dT) spin column chromatography using oligo dT agarose beads type VII (Amersham Pharmacia Biotech, Montréal, Qc, Canada). Complementary DNA (cDNA) synthesis was done according to the manufacturer's instruction in the Uni-Zap vector, and library packaging was done with the GigaPack Gold phage extracts (Stratagene, La Jolla, CA, USA). The titer of the primary libraries were approximately 2,8 x 107 (2-6 DAP library) and 1,2 x 107 (7-17 DAP library), with a mean clone length of 1,5 kb. For both libraries, an aliquot representing ten times the primary library titer was mass excised according to the manufacturer's instruction. Mass excised plasmids were plated on an LB agar supplement with 50 µg/ml kanamycin and 100 µg/ml ampicilin (Sigma-Aldrich Canada Ltd., Oakville, ONT, Canada) to obtain approximately 150 non-overlapping colonies per 132 mm Petri dish. 20,736 individual clones were randomly hand selected from both libraries and transferred to 384-well plates (deep well plates, 200 µl per well).

Virtual subtraction

Sterile Hybond N+ nylon membranes (Amersham Pharmacia Biotech) were placed on Petri dishes containing kanamycin and ampicilin, and clones were transferred to the membranes with a 384 prong microwell plate copier. After overnight growth at 37°C, a picture of every membrane was taken in order to eliminate colonies that had grown poorly, and that would have also been selected as corresponding to weakly expressed mRNAs. Colonies were denatured on the membranes using the following protocol. Membranes were removed from Petri dishes using forceps and placed in a 15 cm X 15 cm container with 15 ml of denaturation solution (0,5N NaOH and 1,5M NaCl) and left for 5 minutes. The membranes were then transferred to the neutralizing solution (1,5M NaCl and 0,5 M Tris-HCl pH 7,4) for 5 minutes. Membranes were then soaked in 2 X SSC for 30 seconds, air-dried and UV cross-linked (120 mJ/cm^2). Probes used for the subtraction process were synthesized from double stranded cDNA obtained during the library construction procedure. These corresponded to the unused fraction of the long double stranded cDNAs obtained after size fractionation (fractions 5 to 8 spanning from 500 bp to > 4 kb) mixed with the smaller rejected cDNAs (fractions 9 to 11 spanning from 200 bp to ~2,5 kb). The amount of the probe was empirically determined as being optimal when using 100 ng per 10 hybridized membranes. Using too little of the probe rendered the subtraction less efficient as more colonies hybridized weakly, and would thus have been selected. The membranes were exposed at –85°C with one intensifying screen on Kodak Biomax MR film (Interscience, Markham, ONT). Prior to the selection of weakly hybridizing colonies, the autoradiogram obtained was compared with the corresponding picture of the bacterial colony growth. Colonies that had not grown were not selected for transfer. From densitometric scan analyses of the autoradiograms, as well as from eye inspection, only the colonies corresponding to the lowest fifth of the hybridizing signal were transferred to 96-well plates containing Terrific Broth (Difco, MD, USA) for plasmid purification and sequencing.

Plasmid DNA preparation and sequencing

Colonies were grown overnight at 37°C in Terrific Broth containing 50 µg/ml kanamycin and 100 µg/ml ampicilin and agitated at 200 rpm on a rotary shaker. Minipreps were performed using Qiagen 96-well miniprep (Qiagen) according to the manufacturer's instructions. Plasmid DNA was quantified using absorbance at 260 nm. Sequencing was done at the Centre d'Innovation Génome Québec (Montréal, Canada). Plasmid DNA was purified using 96-well plates Milipore MultiScreen Plasmid (Milipore, Ont, Canada). Approximately 200 ng (2 µl) of plasmid DNA and 8 µl of a reaction mixture containing 5µl of water, 1.5 µl 5x sequencing buffer, 0.5 µl T3 primer at 20 µM, and 1 µl Big Dye Terminator (Applied Biosystems) was used for the sequencing reaction. The sequencing reactions were performed on a GeneAmp PCR System 9700 (Applied Biosystems), and the cycling conditions were: 96°C/3 sec, 50°C/15 sec, 60°C/4 min for 30 cycles. Amplification products were analyzed using an automated ABI 3700 DNA Analyzer (Applied Biosystems).

Sequence analysis

EST sequences were assembled into the Sputnik relational database (Rudd et al., 2003) and were cleaned of vector sequence or polylinker remnants using the crossmatch application (http://www.phrap.org). The cleaned sequences were then clustered and assembled using the HarvESTer software (Biomax informatics) that utilizes a suffix tree based clustering algorithm (HPT) and the CAP3 assembly algorithm. The result of the analysis is a collection of singleton and cluster consensus sequences. These were recovered from the Harvester application and were imported back into Sputnik. A control EST dataset, named VIRSC (virtual subtraction control), was generated by randomly selecting (using the python Random function) 50 pools of 4,001 (the average number of sequences of SV5 and SV6) TAMU *Lycopersicon esculentum* Ovary ESTs (total of 12 000 EST). These were independently clustered and assembled as described above. For unique match

analysis, all plant EST sequences from the EMBL database were downloaded from the EBI (ftp.ebi.ac.uk/pub/databases/embl/release). The sequences were loaded into the open Sputnik database and were taxonomically filtered into groups that represented the eurosids, the euasterids, the caryophyllids and the monocots. Separate datasets representing the available gene sequences from the complete plant genomes were also included. Blast tables were created for each taxonomic group and the *S. chacoense* sequences were compared to each dataset using BLASTN; results were filtered arbitrarily at 1E-10. All *S. chacoense* sequences were classified for overlap with the sequence collections. The number of sequences that overlapped with each sequence collection were counted as were the number of sequences that appeared unique to the single collection. Classification of the 30 RLKs was performed based on the Blastp homology of the available kinase and extracellular domains to their closest relatives in *Arabidopsis* according to Shiu and Bleecker (2001). Both 5' and 3' sequences obtained from each RLK ESTs were blasted and retrieved the same *Arabidopsis* ortholog, strengthening the classification.

Real-time RT-PCR

For total RNA isolation, tissues were ground in liquid nitrogen, and the powder was kept at -80°C until RNA extractions were performed. RNA was extracted using the RNeasy Plant Mini Kit from Qiagen. RNA concentration was determined by measuring its absorbance at 260 nm and verified (adjusted if necessary) by agarose gel electrophoresis and ethidium bromide staining. Two micrograms of total RNA in a final volume of 11,7 µL was denatured at 65°C for 15 minutes and quickly transferred to ice. 14,3µL of an RT mix containing 1X First Strand Buffer, 0,01 M dithiothreitol (both from Invitrogen), dNTP 1mM/each (Roche), 0,1µg of oligo d(T)20 (Sigma), 10 units of M-MLV RT (Invitrogen), and 2,15 units of RNase Inhibitor (Roche) were added to the denatured RNA for a total volume of 26 µL. The following RT program was used: 25°C for 10 minutes, 37°C for 60 minutes, 99°C for 5 minutes in a GeneAmp PCR System 9700 (Aplied Biosystem). Real-time PCR was performed in triplicate and, as a control, ubiquitin primers derived from a

ubiquitin cDNA sequence obtained from the library were used. Primers for specific receptor kinase were made from the 3'-UTR of each cDNA and were designed to generate a product between 150-250 bp in length. All primer pairs used are listed in Table S1 (see annexe). Four µL of RT was added to a PCR mix of 46 µL containing 20,45 µL sterile water, 1X PCR buffer (Qiagen), 2,5 mM $MgCl_2$, 0,2 mM of each dNTP, 0,25 µM of each primers, 8% glycerol, 3% DMSO, 3,3 µL of a 10 000X dilution of SYBRGreen I (Sigma) and 0,025 unit of HotStart Taq DNA polymerase (Qiagen). Each primer was first tested for amplicon size and the presence of primer dimers through vizualization of PCR product on agarose gel, following ethidium bromide staining. For real-time PCR amplification, the following PCR program was used: 95°C for 10 minutes, 94°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec; steps 2 to 4 were repeated 40 times in a Mx4000® Multiplex QPCR System (Stratagene), and fluorescence readings were done at 60°C. All primers were obtained from Sigma-Aldrich.

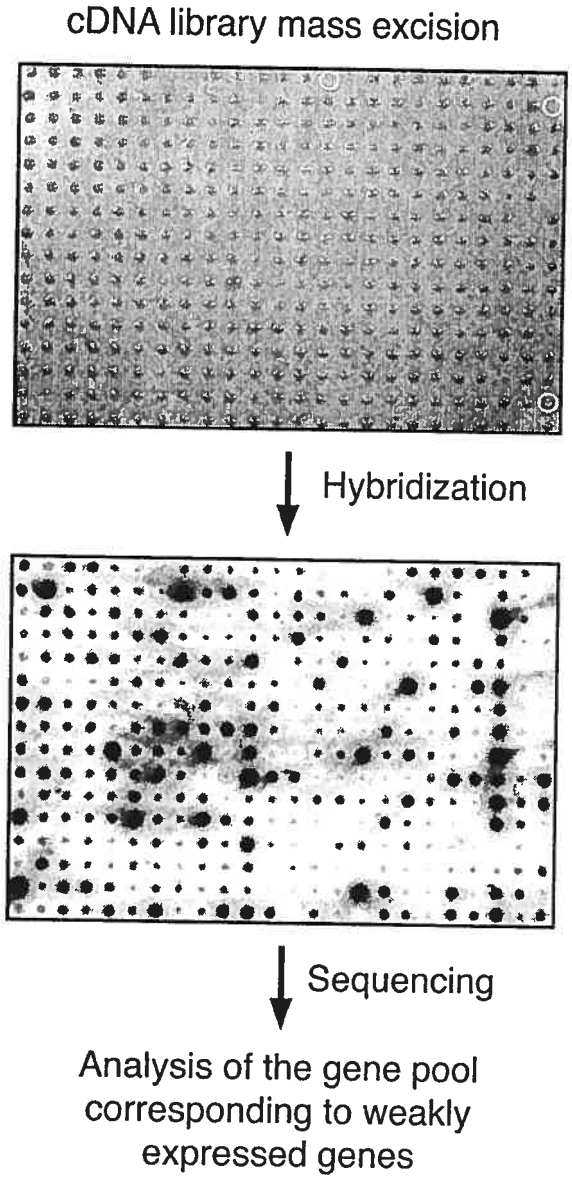
Results

Library screening for low abundance mRNAs

Functional genomics studies depends on the identification of all expressed genes, the transcriptome, within a given organ, tissue, cell type, or under specific physiological conditions. One major hurdle in EST sequencing projects is the high redundancy of sequenced clones and the poor representation of genes that are weakly expressed. We have used a modified subtraction screen based on the virtual subtraction procedure of Li and Thomas (1998), and the negative screen of Nelson et al. (1999), as a method to enrich EST pools for rare messenger RNAs. The method is based on the hybridization of arrayed cDNA clones with a radiolabeled probe made from total cDNAs prepared from the target tissues that have been used to construct the libraries to be screened. Since the probe used to screen the library corresponds exactly to the RNA profile of the tissue used for the library, without an amplification step, this results in a non-biased, sensitive probe, that represents

the actual transcriptome in a given tissue. The rationale behind the use of a "self-probe" is that any weakly abundant transcripts will also be underrepresented in the probe, and thus, a low hybridization signal will be observed for the corresponding cDNAs following membrane hybridization (Figure 4). Two cDNA libraries were made from 2 to 6 DAP depericarped ovaries (named SV6) and from isolated ovules harvested 7 to 17 DAP (named SV5), which covered seed development from early fertilization to the period when the ovules were bearing late torpedo stage embryos. After mass excision of the libraries and initial plating, bacterial colonies were transferred to 384-well plates (54 plates for each library, for roughly 20,000 colonies per library). The colonies were plated on nylon membranes and grown overnight. Each membrane was photographed and compared with the hybridization profile obtained, in order to eliminate colonies that did not grow (Figure 4, white circles). After densitometric scanning of the autoradiograms and eye inspection, colonies that displayed a hybridization signal corresponding to the lowest twenty percent were selected for further analysis. The size of the colony was also taken into account for correction purposes. The selected colonies were transferred into 96-well plates for sequencing.

Figure 4. Schematic representation of the virtual subtraction approach.



In the first step, mass-excised clones were arrayed onto nylon membranes placed on Petri dishes containing LB media and the appropriate antibiotic and incubated overnight at 37°C. Spots where no growth was observed were noted and not selected for sequencing. In the second step, plasmid DNA was cross-linked to the nylon membrane and the probe was hybridized to the membranes. After appropriate washing, the film was exposed to the membrane. All clones yielding a very weak signal or none at all (except for those that had not grown in the first step) were selected for sequencing.

Expressed sequence tag analysis

Single-pass sequences from the 5'-termini of 7741 cDNAs were obtained, with an average of more than 500 bp per EST sequence (data not shown). Table 4 shows the results from the sequence clustering analysis of the SV5 and SV6 sequence collections. In both cases, the redundancy of sequenced clones was low. This is reflected by the relatively high number of singletons and concomitantly low number of contigs obtained. A mean of only two sequences per contig was found in both cases. A direct comparison of statistical data from our EST set and an EST set generated from the large tomato EST sequencing project (Van der Hoeven et al., 2002) cannot be made due to the differences in the number of sequences to be compared. As the number of sequences increases, the likelihood of finding new unique genes decreases (diminishing returns). In order to compare the efficiency of the subtraction method used, a control EST dataset named VIRSC (virtual subtraction control) was generated by randomly selecting 50 pools of 4,001 ESTs (the average of the SV5 and SV6 sets) from the non-normalized TAMU *Lycopersicon esculentum* ovary ESTs (Van der Hoeven et al., 2002). These 4001 randomly chosen ESTs represent an equivalent control, introducing minimal bias related to sample size (saturation) and tissue sampling. Since 50 independent tomato ovary EST pools were generated and averaged, this also reduced the sampling bias within the pool of 12,000 ovary ESTs. Analysis of our EST data set (Table 4)

reveals that the number of unigenes obtained was 91% for SV5 library and 82% for SV6 library. By comparison, the non-subtracted and averaged VIRSC datasets contained only 54% of unigenes. This high percentage of unigenes obtained is reflected in the low numbers of sequences found in contigs, the number of sequences per contigs when compared to the VIRSC data set (2,23 in the pooled SV5/SV6 dataset versus 3,42 in the tomato ovary VIRSC set), and in the percentage of the gene set covered with a limited number of ESTs sequenced (numbers based on the predicted gene number (35,000) in the tomato genome, Van der Hoeven et al., 2002). Thus, the subtraction method used provides both a way to obtain weakly represented transcripts, as well as an efficient and cost effective method to obtain substantial gene set coverage per sequence run.

Tableau 4. Comparison of unsubstracted and substracted EST sequence sets from tomato and *S. chacoense*

	S. chacoense Subtracted			Tomato Non-subtracted	
	SV5	SV6	SV5/SV6	Ovary	Total
Total number of EST	3406	4335	7741	4001	120 892
ESTs in contigs	520	1327	2470	2573	106 833
Total number of contigs	227	557	1103	752	13 215
Total number of unigenes	3113	3565	6374	2180	27 274
Percentage of unigenes	91	82	82	54	23
Average number of repeat of a sequence in a contig	2,29	2,38	2,23	3,42	8,08
Percentage of the gene set covered	8,9	10,2	18	6,2	78
Percentage of the gene set covered per sequence run	0,26	0,23	0,23	0,15	0,064

In order to assess the number of totally new genes in our dataset (genes not found in any of the publicly available databases), we performed a computational comparison between our unigene set and other unigene databases, including *Arabidopsis*, rice, poplar, and *Medicago* genomes, the database of some of the core Eudicot, all Monocots, the animal and fungi kingdom and a non-redundant unigene database (see table 4). Since fully sequenced genomes represent a very large proportion of unigenes of the Eurosid and Asterid classes, all completed genome unigenes were remove from higher classes and analyzed independently. Short sequences (less than 150 bp), and sequences with low coding potential (possibly corresponding to 3'UTR) were rejected from the analysis to avoid and artificial inflation of the number of unique matches. It is to be noted that *S. chacoense* belongs to the Asterid clade. A total of 387 *S. chacoense* sequences (5% of the total EST set, 6% of the unigene set) corresponded to unique matches. These correspond to sequences that by definition have no match elsewhere in the DNA sequence databases tested, are long (>150nt) and have significant protein coding potential. A similar figure (5%) as been observed for other EST collections (Van der Hoeven et al., 2002).

Functional annotation

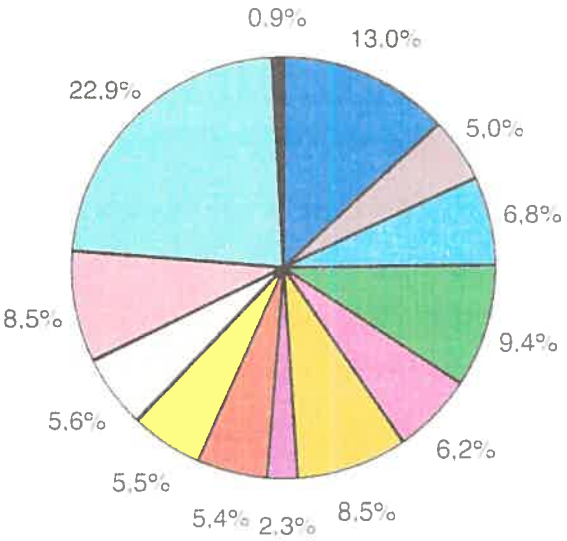
The subtracted *S. chacoense* unigenes and randomly selected EST pools from tomato ovaries (VIRSC) were compared with the *Arabidopsis thaliana* proteome using the BLASTX algorithm (Altschul et al., 1990). Homology based annotation was performed using the closest match obtained against the *Arabidopsis* proteome with an arbitrary threshold for expected values smaller than or equal to 1E-10. All ESTs with an E-value greater than 1E-10 were classified as unknown proteins. Each clone was classified in a functional category according to the criteria used by the Munich Information Center for Protein Sequences (MIPS - http:// mips.gsf.de). Using these criteria, 72% of all *S. chacoense* ESTs could be fitted into defined functional categories (Figure 5). Percentages

from EST belonging to each functional categories were compared with the tomato ovary VIRSC set (Table 5). Although overall similar, a statistical analysis (χ2- test) was performed to determine if these differences were significant and to determine if the subtraction strategy had introduced a bias into the gene pool isolated. A χ2 value of 78.3322 ($p < 1E-5$) confirmed that there were significant differences between the subtracted set versus the non subtracted set for seven of the functional categories (Figure 5B). Four functional categories were slightly more represented in the virtual subtracted pool (Cell Growth, Cell Division and DNA Synthesis; Protein Synthesis; Protein Destination; Cellular Organization) while three functional categories were slightly underrepresented (Metabolism; Cellular Communication; Development) (table 6). If we take into account the percentage of unigenes obtained both in the subtracted and non-subtracted gene set (the redundancy is quite higher in the non-subtracted EST set) and apply it per functional categories, then only the Development category would be slightly less represented in the *S. chacoense* subtracted pool.

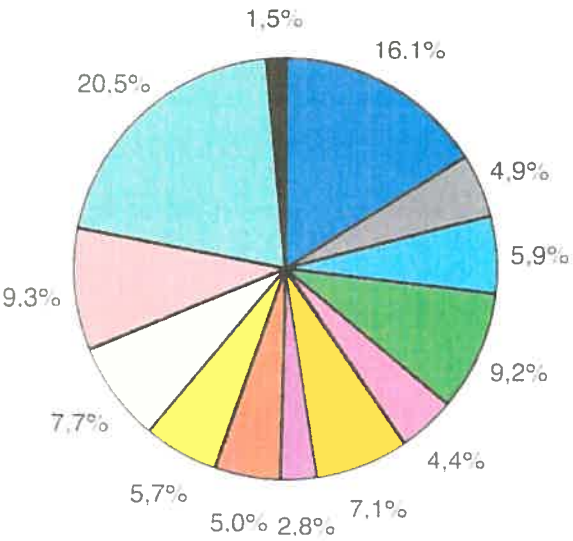
Figure 5. The subtracted *S. chacoense* unigenes and the tomato ovaries VIRSC set were Blasted (BlastX) against the *Arabidopsis thaliana* proteome.

Annotation was done based on the closest match obtained against the *Arabidopsis* proteome using a threshold of an expected value smaller or equal to 1E-10. All EST having an E-value larger than 1E-10 were classified as unknown proteins or as not yet clearly classified. Role categorization was attributed according to the criteria used by the Munich Information Center for Protein Sequences (MIPS) (http:// mips.gsf.de).

S. chacoense



S. lycopersicon



- Metabolism
- Energy
- Cell Growth, Cell Division and DNA Synthesis
- Transcription
- Protein Synthesis
- Protein Destination
- Transport Facilitation
- Intracellular transport
- Cellular Biogenesis
- Cellular Communication and Signal Transduction
- Cell Rescue, Defense, Cell Death and Ageing
- Cellular Organization
- Development

Tableau 5. Occurrence of matches (known genes) and unique matches (novel genes) against the main unigene databases

Database	Matches	Unique matches
Arabidopsis genome	4304	29
Rice genome	1612	2
Poplar genome	2951	4
Medicago genome	1475	4
Euasterid unigenes	2792	6
Eurosid unigenes	3860	26
Caryophyllid unigenes	2481	6
Monocot unigenes	2633	6
Vertebrates	1531	0
Invertebrates	2111	2
Fungi	1560	0
Non-redundant protein database	4565	70
Short sequences	28	10
Low coding potential	1323	1100
All S. chacoense sequences	6272	387

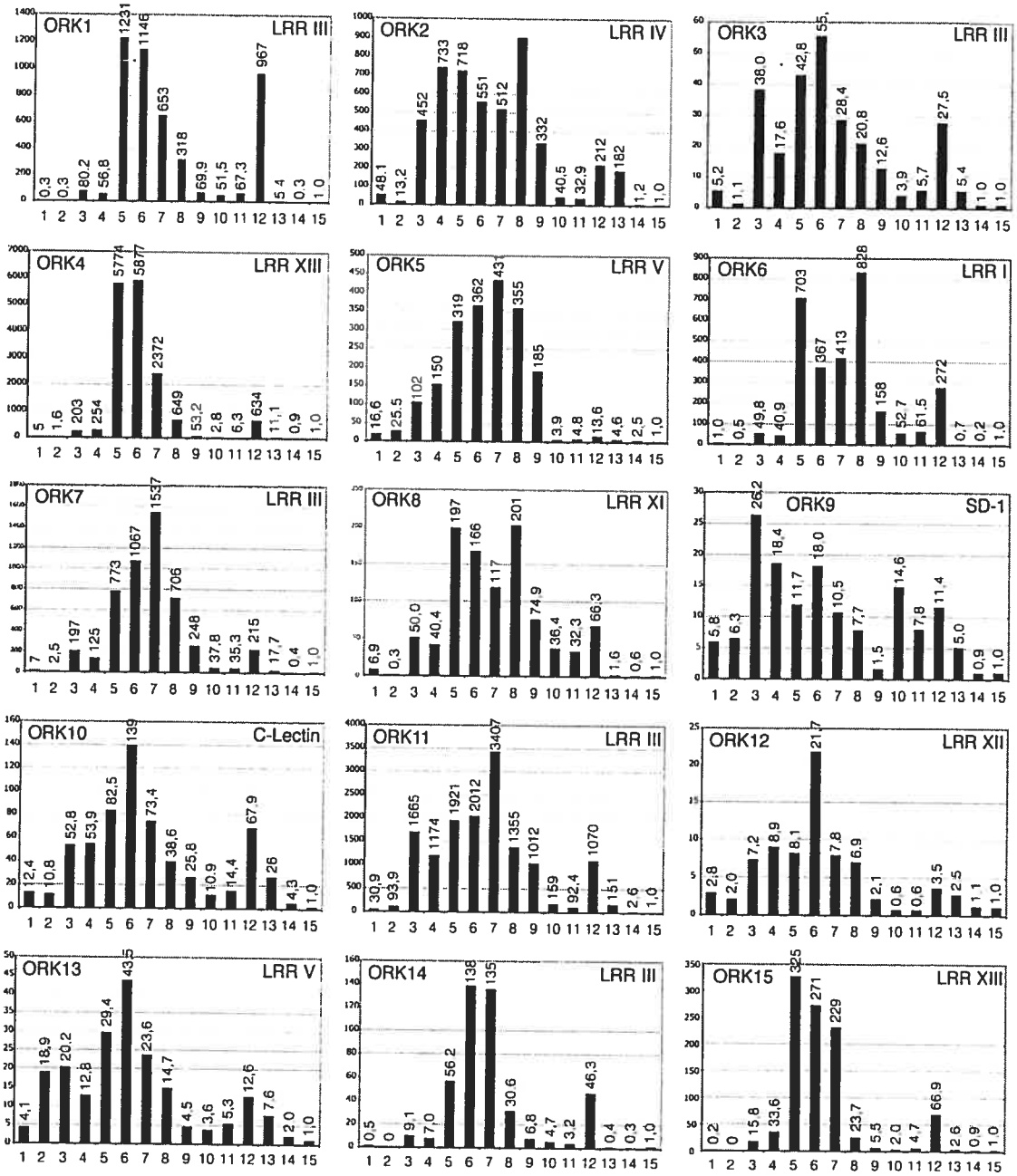
Receptor-like kinases expression analysis

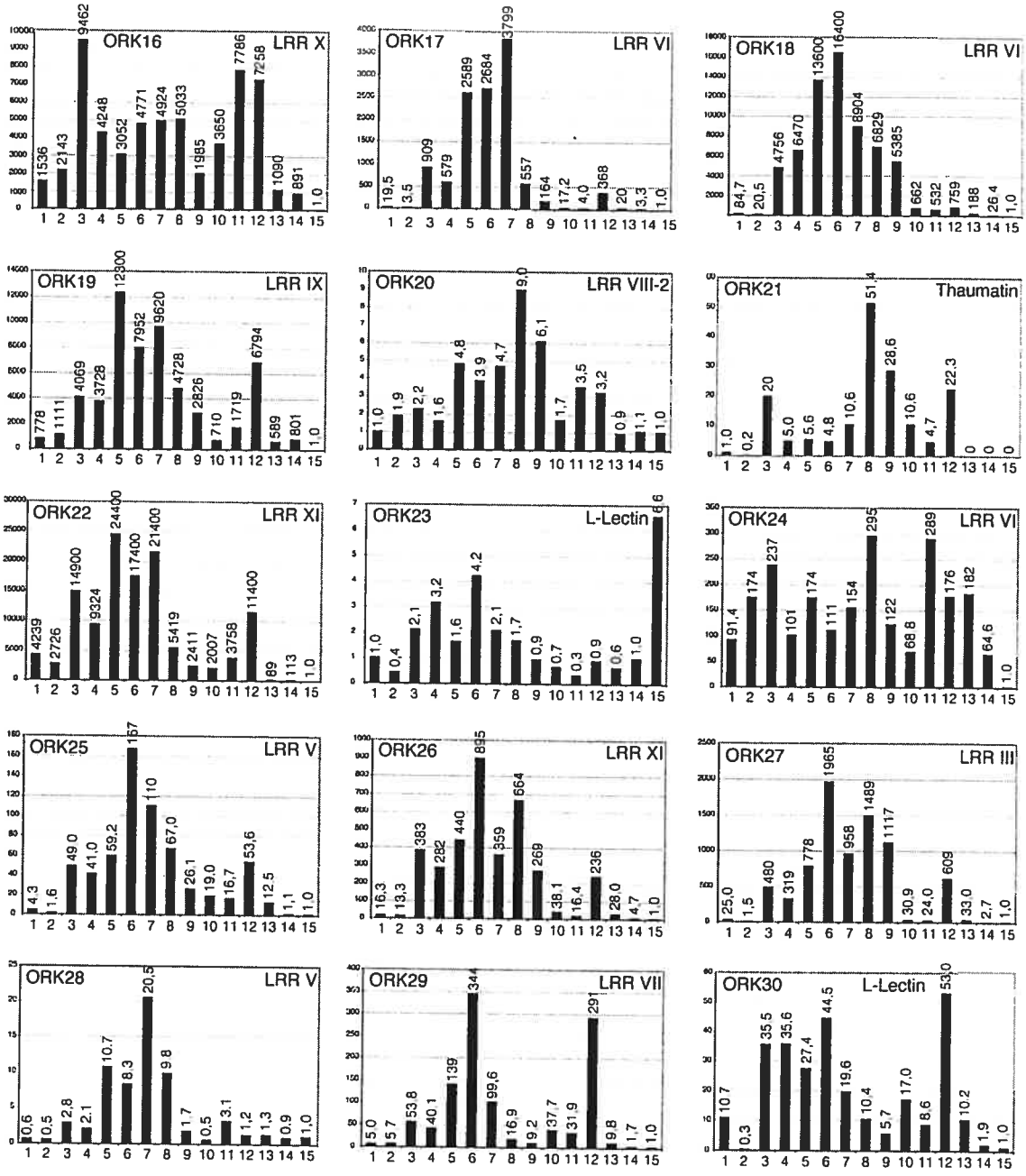
Three obvious drawbacks of the subtraction method used can be considered. Firstly, if a bacterial colony did not grow well, it could have been chosen as corresponding to a weakly expressed gene. This problem was easily resolved by comparing the profile of bacterial growth to the hybridization results (Figure 3). Secondly, colonies that grew well but which harbored a plasmid devoid of an insert would have also been selected, since they would have given a background hybridization signal. From the initial library construction, we knew that very few empty vectors were present, since restriction analysis of 50 randomly selected colonies revealed no vectors without an insert. A more thorough analysis was performed on the EST sequenced and less than 0,1% of the colonies (six in total) were found having empty vectors. Furthermore, only 28 sequenced plasmids contained inserts of less than 150 bp. Thirdly, an important bias may exist within our sampling and classification of ovule sequences. Since a "self-probe" strategy was used, there is a possibility that the genes selected are a mixture of ovule-specific and ovule-predominant genes and genes that are lowly expressed in the target tissue (ovule) but strongly expressed in other tissues. To test this hypothesis, we have selected all the receptor-like kinases identified from our gene pool in female reproductive tissues and determined their expression profile using real-time quantitative RT-PCR analyses (Figure 6). Thirty individual RLKs were retrieved from our gene pool and sequenced from both ends. No redundancy was found among theses RLKs as none was found twice. Classification of the 30 RLKs was performed based on the homology of the kinase and extracellular domains to their closest relatives in Arabidopsis according to Shiu and Bleecker (2001) (Table S3, see annexe). All the receptor kinases were named from ORK1 to ORK30 for Ovule Receptor Kinase, after the tissues in which they were found. Twenty-five corresponded to the LRR-RLK group (Table 6), the largest RLK group in *Arabidopsis*, and which represents more than 50% of all RLKs. Surprisingly, twenty-eight of the analyzed RLK showed their

strongest expression level in ovary or young fruit tissues (Figure 6). The only two RLK analyzed that did not show their strongest expression level in the female reproductive tissues were members of the L-lectin family. These RLK, ORK23 and ORK30, showed their peak expression levels in pollen and leaves respectively. With the exception of ORK21, which was undetectable in pollen and of ORK23, whose expression was highest in pollen, pollen mRNA was used as the calibrator for all quantitative RT-PCR analyses.

Figure 6. Expression profile of the 30 *S. chacoense* Ovule Receptor Kinases isolated from our subtracted pool.

Two (2) µg of total RNA from various tissues was used in a real-time quantitative PCR experiment. Ubiquitin was used as the normalizer, and each amplification was performed in triplicate. In all cases (except ORK23 and ORK30) pollen was used as the calibrator. ORK23 and 30 use 'Style 0 h' as their calibrator. Lanes: 1. Style 0 h, 2. Style 48 h, 3. Ovary 0 h, 4. Ovary 24 h, 5. Ovary 48 h, 6. Ovary 72 h, 7. Ovary 96 h, 8. Ovary 8 days, 9. Ovary 12 days, 10. Roots, 11. Stem, 12. Leaves, 13. Petals, 14. Anthers, 15. Pollen. H is for hours after pollination. RLK specific primers were designed from the 3'UTR of each RLKs.


ORK1
LRR III
0.3 0.3 80.2 56.8 1231 1146 653 318 69.9 51.5 67.3 967 5.4 0.3 1.0
ORK2
LRR IV
48.1 13.2 452 733 718 551 512 332 40.5 32.9 212 182 1.2 1.0
ORK3
LRR III
5.2 1.1 38.0 17.6 42.8 28.4 20.8 12.6 3.9 5.7 27.5 5.4 1.0 1.0
ORK4
LRR XIII
5 1.6 203 254 2372 649 53.2 2.8 6.3 634 11.1 0.9 1.0
ORK5
LRR V
16.6 25.5 102 150 319 362 355 185 3.9 4.8 13.6 4.6 2.5 1.0
ORK6
LRR I
1.0 0.5 49.8 40.9 703 367 413 158 52.7 61.5 272 0.7 0.2 1.0
ORK7
LRR III
7 2.5 197 125 773 1067 706 248 37.8 35.3 215 17.7 0.4 1.0
ORK8
LRR XI
6.9 0.3 50.0 40.4 197 166 117 201 74.9 36.4 32.3 66.3 1.6 0.6 1.0
ORK9
SD-1
5.8 6.3 18.4 11.7 18.0 10.5 7.7 1.5 14.6 7.8 11.4 5.0 0.9 1.0
ORK10
C-Lectin
12.4 10.8 52.8 53.9 82.5 139 73.4 38.6 25.8 10.9 14.4 67.9 26 4.3 1.0
ORK11
LRR III
30.9 93.9 1665 1174 1921 2012 3407 1355 1012 159 92.4 1070 151 2.6 1.0
ORK12
LRR XII
2.8 2.0 7.2 8.9 8.1 21.7 7.8 6.9 2.1 0.6 0.6 3.5 2.5 1.1 1.0
ORK13
LRR V
4.1 18.9 20.2 12.8 29.4 43.5 23.6 14.7 4.5 3.6 5.3 12.6 7.6 2.0 1.0
ORK14
LRR III
0.5 0 9.1 7.0 56.2 138 135 30.6 6.8 4.7 3.2 46.3 0.4 0.3 1.0
ORK15
LRR XIII
0.2 0 15.8 33.6 325 271 229 23.7 5.5 2.0 4.7 66.9 2.6 0.9 1.0


ORK16
LRR X
1536 2143 9462 4248 3052 4771 4924 5033 1985 3650 7786 7258 1090 891 1.0
ORK17
LRR VI
19.5 3.5 909 579 2589 2684 3799 557 164 17.2 4.0 368 20 3.3 1.0
ORK18
LRR VI
84.7 20.5 4756 6470 13600 16400 8904 6829 5385 662 532 759 188 26.4 1.0
ORK19
LRR IX
778 1111 4069 3728 12800 7952 9620 4728 2826 710 1719 6794 589 801 1.0
ORK20
LRR VIII-2
1.0 1.9 2.2 1.6 4.8 3.9 4.7 9.0 6.1 1.7 3.5 3.2 0.9 1.1 1.0
ORK21
Thaumatin
1.0 0.2 20 5.0 5.6 4.8 10.6 51.4 28.6 10.6 4.7 22.3 0 0 0
ORK22
LRR XI
4239 2726 14900 9324 24400 17400 21400 5419 2411 2007 3758 11400 89 113 1.0
ORK23
L-Lectin
1.0 0.4 2.1 3.2 1.6 4.2 2.1 1.7 0.9 0.7 0.3 0.9 0.6 1.0 6.6
ORK24
LRR VI
91.4 174 237 101 174 111 154 295 122 68.8 289 176 182 64.6 1.0
ORK25
LRR V
4.3 1.6 49.0 41.0 59.2 167 110 67.0 26.1 19.0 16.7 53.6 12.5 1.1 1.0
ORK26
LRR XI
16.3 13.3 383 282 440 895 359 664 269 38.1 16.4 236 28.0 4.7 1.0
ORK27
LRR III
25.0 1.5 480 319 778 1965 958 1489 1117 30.9 24.0 609 33.0 2.7 1.0
ORK28
LRR V
0.6 0.5 2.8 2.1 10.7 8.3 20.5 9.8 1.7 0.5 3.1 1.2 1.3 0.9 1.0
ORK29
LRR VII
5.0 5.7 53.8 40.1 139 344 99.6 16.9 9.2 37.7 31.9 291 9.8 1.7 1.0
ORK30
L-Lectin
10.7 0.3 35.5 35.6 27.4 44.5 19.6 10.4 5.7 17.0 8.6 530 10.2 1.9 1.0

Transcriptional regulation of the ORKs following pollination and fertilization

Since most of the isolated RLKs were predominantly expressed in female reproductive tissues, we determined if these RLKs were regulated at the transcriptional level following pollination, fertilization, and later in ovules during early seed development. An arbitrary threshold value of 1.5 fold variation was set and genes were scored as induced after pollination or reduced after pollination by comparing expression levels at 0 h (unpollinated) and 24 h after pollination. Similarly, genes were scored as induced after fertilization or reduced after fertilization by comparing expression profiles between 24 h and 48 h after pollination (fertilization takes place from 36 to 42 h post-pollination in *S. chacoense*), and later during early embryogenesis by comparing the average expression level between 8 DAP and 12 DAP with control ovules (0 h). These later stages (8 DAP and 12 DAP) correspond to ovules bearing embryos from the 8 to 16 cells stages to globular embryos respectively. Furthermore, pollination effects can be detected both locally (in the style) or at a distance in the ovary before the pollen tubes reach the ovary (Lantin et al., 1999). These two possibilities were thus treated separately (Table 6). Figure 7 shows a Venn diagram representation showing the relationship between the 30 *S. chacoense* ORKs and their expression profile during pollination and fertilization. The most striking results were observed in the induced after fertilization class. Twenty-three out of thirty (77%) ORK genes were induced following fertilization and, of these, 22 belonged to the large LRR-RLK class. Only ORK10 belonged to the C-lectin class and its induction level was at the selected threshold value only. Of these, seven ORKs showed a greater than 5-fold increase in mRNA levels: ORK28 (5,2-fold), ORK7 (6,2-fold), ORK14 (8,1-fold), ORK15 (9,7-fold), ORK6 (17,2-fold), ORK1 (21,7-fold), and ORK4 (22,7-fold). Conversely, 16 ORKs (53%) showed a significant decrease in expression following pollination in the style. Five of these showed a greater than 5-fold decrease in mRNA levels: ORK17 (5,5-fold), ORK-14 (12,0-fold), ORK27 (16,2-fold), ORK8 (19,8-fold), and ORK30 (31,5-fold). Out

of those that were induced after fertilization in the ovary, 48% (11/23) had their expression concomitantly reduced in the style. Also noteworthy, while there are only five out the thirty genes that are not part of the LRR family, four of these five genes were among the non-induced or repressed genes.

Figure 7. Venn diagram showing the relationship between the 30 *S. chacoense* ORKs and their expression profile during pollination and fertilization.

Individual ORKs were placed in their respective expression categories based on the fold-induction or fold-repression threshold value of 1.5 as determined in Table 6. Numbers correspond to their respective ORK. *Although ORK12 did not show any immediate mRNA increase 48 h after pollination, a 3-fold transient increase was observed 72 h after pollination.

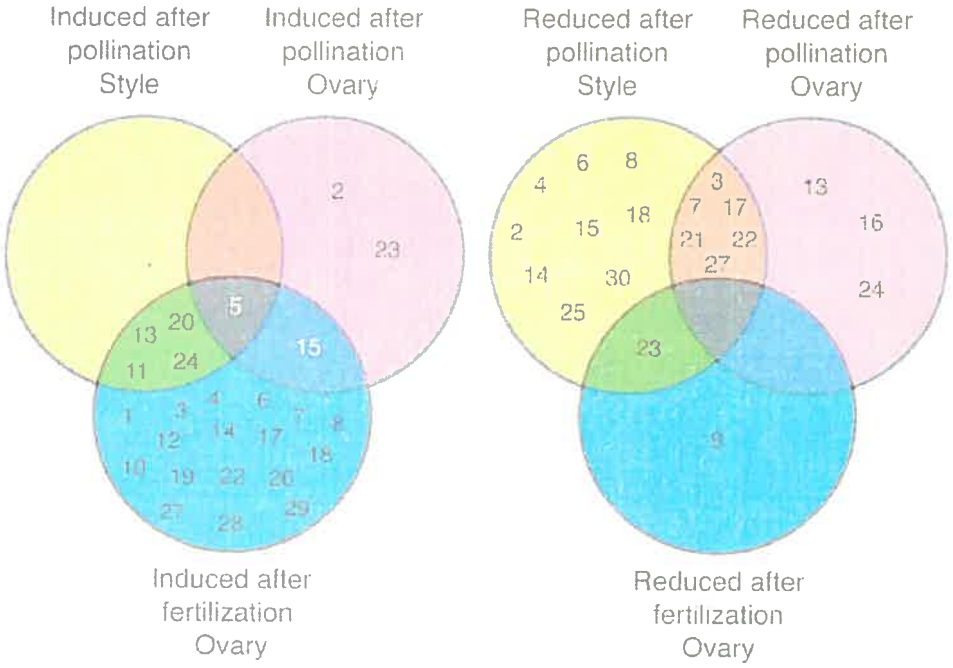


Tableau 6. Comparison of observed versus expected frequency of each functional categories using a table frequency analysis with linear-log between the *S. chacoense* ESTs and the tomato VIRSC EST set.

Significant differences: a statistical analysis (χ2- test) was performed to determine if the differences were significant and to determine if the subtraction strategy had introduced a bias into the gene pool isolated. A χ2 value of 78.3322 ($p < 1E-5$) confirmed that there were significant differences between the subtracted set versus the non subtracted set for seven of the functional categories.

Class	**Observed numbers**	**%**	**Expected numbers**	**%**	**Significant differences**
Metabolism	829	13,0	644	16,1	-
Energy	319	5,0	196	4,9	
Cell Growth, Cell Division and DNA Synthesis	433	6,8	236	5,9	+
Transcription	599	9,4	368	9,2	
Protein Synthesis	395	6,2	172	4,3	+
Protein Destination	542	8,5	284	7,1	+
Transport Facilitation	147	2,3	112	2,8	
Intracellular Transport	344	5,4	200	5	
Cellular Biogenesis	351	5,5	228	5,7	
Cellular Communication/Signal Transduction	357	5,6	308	7,7	-
Cell Rescue, Defense, Cell Death and Ageing	542	8,5	372	9,3	
Cellular Organization	1460	22,9	820	20,5	+
Development	57	0,9	60	1,5	-
	6375		4000		

Tableau 7. Induction and repression level after pollination (24 h/0 h), fertilization (48 h/24 h) and later during embryogenesis (average of 8 days and 12 days/0 h).

ORK	Class	Pollination				Fertilization		Later during embryogenesis	
		Induction		Repression					
		Style	Ovary	Style	Ovary	Induction	Repression	Induction	Repression
ORK1	III					21,7		2,4	
ORK2	VI		1,6	3,6					
ORK3	III			4,6	2,2	2,4			2,3
ORK4	XIII			3,1		22,7		1,7	
ORK5	V	1,5	1,5			2,1		2,6	
ORK6	I			2,0		17,2		9,9	
ORK7	III			2,8	1,6	6,2		2,4	
ORK8	XI			19,8		4,9		2,8	
ORK9	SD-1						1,6		5,7
ORK10	C-Lectin					1,5			1,6
ORK11	III	3,0				1,6			
ORK12	XII								1,6
ORK13	V	4,5			1,6	2,3			2,1
ORK14	III			12,0		8,1		2,0	
ORK15	XIII		2,1	3,5		9,7			
ORK16	X				2,2				2,7
ORK17	VI			5,5	1,6	4,5			2,5
ORK18	VI			4,1		2,1			
ORK19	IX					3,3			
ORK20	VIII-2	1,9				3,0		3,4	
ORK21	Thaumatin			4,2	4,0			2,0	
ORK22	XI			1,6	1,6	2,6			3,8
ORK23	L-Lectin		1,5	2,4			1,9		1,6
ORK24	VI	1,9			2,3	1,7			
ORK25	V			2,8					
ORK26	XI					1,6			
ORK27	III			16,2	1,5	2,4		2,7	
ORK28	V					5,2		2,0	
ORK29	VII					3,5			4,1
ORK30	L-Lectin			31,5					4,4

In the 'later during embryogenesis' class eleven ORKs were induced while another eleven ORKs had their expression level reduced above the threshold level. Again, in the genes had their expression level reduced, of the five ORKs that did not belong to the LRR-RLK class, four were repressed. Another interesting aspect is revealed by the presence of a group of thirteen ORKs that are modulated by pollination, but at a distance in the ovary, since their expression level can be increased (ORK2, ORK5, ORK15, and ORK23) or decreased (ORK3, ORK7, ORK13, ORK16, ORK17, ORK21, ORK22, ORK24, and ORK27) before the pollen tubes reach the ovary. Thus, expression of these RLKs is either modulated developmentally, or by long-distance signaling cues as determined before for other genes (Lantin et al., 1999).

Discussion

During plant sexual reproduction, pollination and fusion of the gametes are expected to trigger a vast array of signal transduction events that will lead to a reorganization of the expressed gene set, which will in turn direct growth and development of the embryo and give rise to a mature fruit. In the present work, which aimed at gaining a better knowledge of the transcriptome induced following fertilization and early embryogenesis, we report the generation of more than 7700 EST sequences from *Solanum chacoense* ovules taken at diverse stages of embryo development, with a particular focus on genes involved in signal transduction events. The presence of a self-incompatibility system in *S. chacoense* not only enables a total control over pollination and fertilization timing, but also makes it possible to address issues related to genes differentially regulated by compatible and incompatible pollinations as well as genes regulated at a distance in the ovary following pollination (Lantin et al., 1999).

Despite the fact that the fruits of several domesticated Solanaceous species (e.g. tomato, eggplant, tomatillo and tamarindo) are very important from an economic point of view, gene expression studies in the female gametophyte have not been extensively

characterized until recently (Drews and Yadegari, 2002; Van der Hoeven et al., 2002; Hu et al., 2003; de Folter et al., 2004; Hennig et al., 2004; Sprunck et al., 2005). Several key aspects of plant development are known and are suspected to be controlled by protein kinases. Because of their capacity to amplify very subtle stimulus, components of signaling modules, such as receptor kinases, mitogen-activated protein kinases (MAPKs), and small signaling ligands, are generally expressed at low levels and are weakly represented in most EST sequencing projects (Hu et al., 2003). In order to increase the number of unigenes found, recent large scale EST sequencing projects in Solanaceous species (Van der Hoeven et al., 2002; Ronning et al., 2003) have relied on the use of a large number of diversified libraries, while other methods used to decrease gene redundancy (based on library normalization) have also been developed. Despite all the disadvantages linked to sequence redundancy, EST sequencing remains a method of choice for gene discovery as well as a cost-effective method for tagging thousands of genes from a given organism. This is reflected by the fact that nearly 6 millions (5 817 901) viridiplantae ESTs had been deposited at NCBI (Jan 01, 2005). Yet only 53 *S. chacoense* cDNA sequences have been deposited at NCBI and no single ESTs have been deposited so far. Thus, the *S. chacoense* data presented here is therefore the first large scale sequencing effort addressing this organism that has been widely used for introgression of valuable traits in the cultivated potato (Hawkes, 1990), and for gametophytic self-incompatibility studies (Matton et al., 1999; O'Brien et al., 2002). In order to analyze the transcriptional program induced during embryo development, libraries from mixed-stages fertilized ovules were selected for a medium scale EST sequencing project. To increase the isolation of weakly expressed mRNAs; reduce the redundancy of the EST dataset; and decrease the cost related to DNA sequencing, the selection of the clones to be sequenced was preceded by a virtual subtraction (negative selection) of the target tissue cDNAs. The virtual subtraction screen was first mentioned by Li and Thomas (1998) and was developed to isolate tissue-specific genes, and genes represented by low-abundance mRNAs in plant embryos. It uses random-primed PCR cDNA probes derived from non-target tissue mRNAs converted to double-stranded cDNAs. A similar strategy termed negative subtraction hybridization was used by

Nelson et al. (1999) to isolate weakly expressed transcripts from human prostate tissue. It used high-density cDNA clone arrays that were hybridized to a first-strand cDNA probe derived from mRNAs isolated from the same tissue used to build the library to be screened. The authors selected the cDNA clones representing the lowest quartile of the hybridization intensities for sequencing. The methodology resulted in a significant decrease in redundancy from the negative selection libraries compared to non-normalized library, but only limited EST numbers were used, thus introducing a bias due to the limited sampling depth of the libraries (average of less than a 1000 clones sampled for each library). Nonetheless, from this limited data set, a three-fold redundancy in ESTs was observed, from 33% in a non-normalized library to 11% in the negative selection library. In two recent studies, two research groups also used a negative selection screen based on the hybridization of the target tissue library with a first-strand reverse transcribed mRNAs isolated either from the same tissues (Vodkin et al., 2004) or from fungal cells subjected to different physiological conditions (Ray et al., 2004). In the study involving plant tissue, Vodkin et al. (2004) used the negative selection screen to isolate weakly expressed genes from an immature soybean cotyledon library. This strategy increased the number of new genes identified from the ESTs sequenced from the negatively subtracted pool when compared to the non-normalized library. Again, the EST numbers used were too low (only 931 sequences came from the non-subtracted library and 1528 were sequenced from the negatively subtracted pool), thus introducing a bias due to the limited sampling depth of the libraries. For the current study, we used a self-probe strategy that used doubled-stranded cDNAs labeled to high specific activity. Since the probe used to screen the library corresponds exactly to the RNA profile of the tissue used for the library, this results in a non-biased sensitive probe, that represents the actual transcriptome in a given tissue. In our hands, the use of a radioactively-labeled first-strand cDNA probe derived from reverse-transcribed mRNAs showed a lower sensitivity than the use of a double-stranded cDNA pool as the starting material (Chantha, S.-C. and Matton, D. P., unpublished results). With a first-strand cDNA probe, almost half the colonies would have been scored as negatives (close to background level), compared to approximatively 20% with the double-stranded

cDNA probe procedure used in this study. Furthermore, when we used a first-strand cDNA probe strategy, the same profiles were obtained, irrespective of the mRNA source used. For example, no differences were observed between a probe made from whole flowers compared to ovaries, or from a mixture of root, stem, leaf and anther tissues, or from unpollinated pistils, suggesting a lesser degree of discrimination with the use of such probes (Chantha, S.-C. and Matton, D. P., unpublished results), and indicating that the same highly expressed genes are found in all tissues examined. Without library normalization, recent EST sequencing efforts of the same magnitude as the one presented here, produced a much lower percentage of unigenes, from 43.1% for 7106 ESTs sequenced from sunflower embryonic libraries (Ben et al., 2005), to 47,7% for 11 954 ESTs from various tissues in cassava (Lopez et al., 2004), and to 50,5% for 15,781 ESTs from common bean libraries also covering various tissues (Ramirez et al., 2005). These numbers are quite similar with the one obtained from our VIRSC non-normalized control set (54%, table 3). Thus the virtual subtraction used in the present study, considerably increased the number of unigenes isolated from each library screened (91% from the SV5 and 82% from the SV6 library respectively), for a total of 82% considering the two libraries together (6374 unigenes out of 7741 ESTs). Furthermore, comparison of the two libraries that correspond to different developmental stages indicates only limited redundancy at the gene expression level since only 304 unigenes were shared between the two libraries. A total of 6678 unigenes (3113 and 3565 unigenes for SV5 and SV6 respectively) was obtained from the sum of the two libraries, but this was only reduced to 6374 unigenes when the two libraries were merged for analysis (Table 3). Thus the contribution of each library is quite distinct and complementary, and suggests a complex transcriptional shift between the various stages of the ovules and the embryo it bears during early stages of embryo development. This was also observed using stage specific libraries in the sunflower embryo EST project (Ben et al., 2005).

Another interesting observation is that a total of 387 (5%) *S. chacoense* sequences corresponded to unique matches (genes that are new to the database but that have strong

coding potential)(Table 4). A similar figure had been observed for the tomato EST collection when compared to available plant sequences (Van der Hoeven et al., 2002). Thus the subtraction not only led to the identification of ovule-expressed genes but to the finding of new genes, corresponding to roughly 1 % of the predicted *Solanum lycopersicon* genome with only a limited sequencing effort, and although already more than 400,000 sequences from *Solanum tuberosum* and *Solanum lycopersicon* ESTs have been deposited in Genbank (see the taxonomy browser at http://www.ncbi.nlm.nih.gov/Taxonomy/Browser for details about the number of deposited sequences).

The functional annotation of our 7741 ESTs set revealed that, although the genes sequenced were selected on the basis of their weak expression level, no major difference was observed between the subtracted set and a non-normalized EST set (the VIRSC set of tomato ovary ESTs), as far as percentage of genes found per category (Figure 5). This implies that lowly expressed genes are expressed throughout all functional categories and are not limited to specific categories like Development or Cellular communication and signal transduction. Interestingly, in a study on the transcriptional program during early reproductive stages in *Arabidopsis thaliana*, Hennig et al. (2004), came to the conclusion that the transcriptional reprogramming during reproductive development involved genes from all major functional classes, although the authors found a significant enrichment for genes encoding proteins involved in metabolism, transcription, and cellular organization. Similarly, these three functional categories were the most represented in our gene pool (Figure 5).

Our specific interest in signal transduction events involved in ovule and embryo development led us to analyze all the receptor-like kinases found in our gene pool by quantitative RT-PCR analyses over a wide range of tissues and developmental stages (Figure 6). Of the 30 RLKs isolated twenty-five corresponded to the LRR-RLK group (Table 5), the largest RLK group in Arabidopsis. Surprisingly, twenty-eight of the analyzed RLK showed their strongest expression level in ovary or young fruit tissues (Figure 6). The

only two RLK analyzed that did not show their strongest expression level in the female reproductive tissues were members of the L-lectin family. This strongly suggests that the subtraction procedure did not introduce any bias toward genes weakly expressed in the target tissue while being strongly expressed elsewhere in the plant. Despite the fact that several of the twenty-five LRR-RLK found in this EST project fall within subfamilies containing well characterized members, such as CLAVATA1 (Clark et al., 1997) and HAESA (Jinn et al., 2000) from subfamily LRR XI, BRI1 (Li and Chory, 1997) in subfamily LRR X, ERECTA (Torii et al., 1996) in subfamily LRR XIII, and FLS2 (Asai et al., 2002) in subfamily LRR XII, none of the thirty RLK found was an ortholog of a RLK with an already known function, suggesting that many more RLKs are to be found having important roles during reproductive development. To our knowledge only one of our ORK ortholog has been previously characterized in *Arabidopsis* (Tarutani et al., 2004). The authors have analyzed T-DNA lines for RLK902 (At3g17840 corresponding to ORK11) and generated a double-mutant containing T-DNA insertions in RLK902 and RLK1 (a very similar RLK) and could not observed any phenotype. When the RLK902 promoter was fused to the GUS reporter gene, expression was observed in the root tips, the lateral root primordia, the stipules, and the floral abscission zones. RLK902 and RLK1 are the two closest relative of a large family but display very different GUS staining, suggesting a different function.

Twenty-three out of thirty (77%) ORK genes were induced following fertilization and, of these, 22 belonged to the large LRR-RLK class. Interestingly, a coordinated regulation was found between half the ORKs that were being up-regulated in the ovary after fertilization and down-regulated in the style after pollination (Figure 7). This could suggest that, among the RLKs characterized, some might be involved in pollen-pistil interactions, their expression following the progression of the pollen tube in the style and ovary, while others are clearly fertilization-related only, showing a very low level of expression in style and ovary before pollination and fertilization, and a steep increase in mRNA expression levels following fertilization, as determined for ORK28, ORK7,

ORK14, ORK15, ORK6, ORK1, and ORK4 (Figure 6 and 7). On the other hand, genes belonging to the SD-1, L/C-lectin and thaumatin family may not play an active role during fertilization and fruit growth since none are induced at fertilization time and/or are repressed later during embryogenesis above the selected threshold level. The fact that no single ORK was found to be solely pollination-induced might be the consequence of the tissue sampled for the EST project since from the gynoecium, only depericarped ovaries and isolated ovules were used to make the libraries. Apart from responses resulting from a direct interaction between pollen and pistil, another interesting class of ORKs is exemplified by ORKs that are transcriptionally induced or repressed at a distance, before the pollen tube reach the ovary (Figure 7). Pollen tube growth in the style is known to cause major cellular deterioration and ultimately death of the transmitting tissue (Cheung, 1996). Among the cues that have been shown to modulate gene expression at a distance in the ovary, wounding and jasmonates have been shown to be potent inducers (Lantin et al., 1999). Ethylene production, has also been shown to be highly regulated during pollination and would also be a good candidate in modulating responses at a distance in pistil tissues (Singh et al., 1992).

It is noteworthy that, among the 30 ORKs analyzed, 28 showed their peak expression in ovule or ovary tissues, and for 23 of these, leaf was their second highest expressing tissue. One L-lectin RLK (ORK30) had its highest expression level in leaf and then in gynoecium tissue. In a transcriptional profiling experiment using the Arabidopsis ATH1 microarray, Hennig et al. (2004), also noted that flowers expressed mainly the same genes as leaves, but that there were many more flower-specific than leaf-specific genes. These two results are in agreement with Goethe's hypothesis proposed more than 200 years ago that flower organs represent modified leaves (Goethe, 1831). We also observed that genes belonging to the LRR family nearly all responded positively (induced) following fertilization in the ovary while the opposite pattern was observed for all other families analyzed. In line with these results we observed that some families, namely LRR III, LRR V, LRR VI, and LRR XIII appear to be over-represented in the gynoecium, from two to ten

fold, when compared to their respective representation among the Arabidopsis RLKs (Table S2, see annexe). These findings suggest that the female reproductive tissue uses a rather unique set of receptor kinases, which is not expressed at a high level in other tissue and that families of RLK might have evolved to co-regulated precise developmental processes such as embryogenesis.

Earlier estimates from reassociation kinetics had proposed that the number of genes expressed in the male-gametophyte was as high as 20 000 genes (Goldberg et al., 1989), and since the female gametophyte is an actively growing tissue, there is no reason to believe that this figure would be lower in the later. More recent analyses based on microarray hybridization suggests that around 15,500 genes are expressed in flowers (Hennig et al., 2004), although the detection limit of the microarrays compared to quantitative PCR might underestimate the actual number of genes expressed in reproductive tissues. Nonetheless, out of 7741 ESTs sequenced from a normalized (virtually subtracted) library, 6374 unigenes were isolated, thus representing a large fraction of the total number of genes expressed in female reproductive tissues. This new EST dataset combined with the production of cDNA arrays derived from it should be an invaluable tool in addressing these issues as well as the transcriptional profiling of genes during fertilization and early embryogenesis in a solanaceous species.

Acknowledgements

We thank Gabriel Téodorescu for plant care and maintenance. We also thank Jean-Nicolas Sylvestro for helping with the colony plating and hybridization; Kateryna Vyetrogon, Daphné Dupéré-Richer, Thomas Chevalier-Laliberté, and Harmanjit Singh for helping out with EST annotations; Homère Lamarre-Ménard for his work on our database web interface, and Stéphane Daigle for statistical analyses. Special thanks also to Professor Terry L. Thomas (Texas A&M University) for his guidance and support at the beginning of this project. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Canada Research Chair program. H. Germain, M. O'Brien, and S.-C. Chantha are the recipient of Ph. D. fellowships from NSERC and from Le Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT, Québec). D. P. Matton holds a Canada Research Chair in Functional Genomics and Plant Signal Transduction.

References

1. Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G.,Gocayne, J.D., White, O. 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377: 3-174.

2. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J Mol Biol 215: 403-410.

3. Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.L., Gomez-Gomez, L., Boller, T., Ausubel, F.M. and Sheen, J. 2002. MAP kinase signalling cascade in Arabidopsis innate immunity. Nature 415: 977-983.

4. Becraft, P.W. 2002. Receptor kinase signaling in plant development. Annu Rev Cell Dev Biol 18: 163-192.

5. Ben, C., Hewezi, T., Jardinaud, M.F., Bena, F., Ladouce, N., Moretti, S., Tamborindeguy, C., Liboz, T., Petitprez, M. and Gentzbittel, L. 2005. Comparative analysis of early embryonic sunflower cDNA libraries. Plant Mol Biol 57: 255-270.

6. Bonaldo, M.F., Lennon, G., and Soares, M.B. 1996. Normalization and subtraction: Two approaches to facilitate gene discovery. Genome Res 6: 791-806.

7. Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. 2000. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res 10: 1617-1630.

8. Cheung, A.Y. 1996. The pollen tupe growth pathway: its molecular and biochemical contributions and responses to pollination. Sexual Plant Reproduction 9: 330.

9. Clark, M., Hennig S, Herwig R, Clifton SW, Marra MA, Lehrach H, Johnson SL, Group tW; WU-GSCnEST Group. 2001. An Oligonucleotide Fingerprint Normalized and Expressed Sequence Tag Characterized Zebrafish cDNA Library. Genome Res 11: 1594-1602.

10. Clark, S.E., Williams, R.W. and Meyerowitz, E.M. 1997. The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. Cell 89: 575-585.

11. de Folter, S., Busscher, J., Colombo, L., Losa, A. and Angenent, G.C. 2004. Transcript profiling of transcription factor genes during silique development in Arabidopsis. Plant Mol Biol 56: 351-366.

12. Drews, G.N. andYadegari, R. 2002. Development and function of the angiosperm female gametophyte. Annu Rev Genet 36: 99-124.

13. Fei, Z., Tang, X., Alba, R.M., White, J.A., Ronning, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.J. 2004. Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. Plant J 40: 47-59.

14. Goethe, J.W.v. (1831) Essai sur la métamorphose des plantes. Translation of Versuch die Metamorphose der Pflanzen zu erklären. J. G. Cotta, Stuttgart

15. Goldberg, R.B., Barker, S.J. and Perez-Grau, L. 1989. Regulation of gene expression during plant embryogenesis. Cell 56: 149-160.

16. Gong, Z., Yan, T., Liao, J., Lee, S.E., He, J. and Hew, C.L. 1997. Rapid identification and isolation of zebrafish cDNA clones. Gene 201: 87-98.

17. Hawkes, J.G. (1990) The potato: evolution, biodiversity, and genetic resources. Smithsonian Institution Press, Washington, D. C.

18. Hennig, L., Gruissem, W., Grossniklaus, U. and Kohler, C. 2004. Transcriptional programs of early reproductive stages in Arabidopsis. Plant Physiol 135: 1765-1775.

19. Hu, W., Wang, Y., Bowers, C. and Ma, H. 2003. Isolation, sequence analysis, and expression studies of florally expressed cDNAs in Arabidopsis. Plant Mol Biol 53: 545-563.

20. Initiative, A.G. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796-815.

21. Jinn, T.L., Stone, J.M. and Walker, J.C. 2000. HAESA, an Arabidopsis leucine-rich repeat receptor kinase, controls floral organ abscission. Genes Dev 14: 108-117.

22. Jones, J.D.G., Dunsmuir, P. and Bedbrook, J. 1985. High level expression of introduced chimeric genes in regenerated transformed plants. Embo J 4: 2411-2418.

23. Lantin, S., O'Brien, M. and 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: 371-386.

24. Li, J. andChory, J. 1997. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell 90: 929-938.

25. Li, Z. andThomas, T.L. 1998. PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in Arabidopsis. Plant Cell 10: 383-398.

26. Lopez, C., Jorge, V., Piegu, B., Mba, C., Cortes, D., Restrepo, S., Soto, M., Laudie, M., Berger, C., Cooke, R., Delseny, M., Tohme, J. and Verdier, V. 2004. A unigene catalogue of 5700 expressed genes in cassava. Plant Mol Biol 56: 541-554.

27. Matton, D.P., Luu, D.T., Xike, Q., Laublin, G., O'Brien, M., Maes, O., Morse, D. and Cappadocia, M. 1999. Production of an S RNase with dual specificity suggests a novel hypothesis for the generation of new S alleles. Plant Cell 11: 2087-2097.

28. Nelson, P.S., Hawkins, V., Schummer, M., Bumgarner, R., Ng, W.L., Ideker, T., Ferguson, C. and Hood, L. 1999. Negative selection: a method for obtaining low-abundance cDNAs using high-density cDNA clone arrays. Genet Anal 15: 209-215.

29. O'Brien, M., Kapfer, C., Major, G., Laurin, M., Bertrand, C., Kondo, K., Kowyama, Y. and Matton, D.P. 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: 985-996.

30. Radelof, U., Hennig, S., Seranski, P., Steinfath, M., Ramser, J., Reinhardt, R., Poustka, A., Francis, F. and Lehrach, H. 1998. Preselection of shotgun clones by oligonucleotide fingerprinting: an efficient and high throughput strategy to reduce redundancy in large-scale sequencing projects. Nucleic Acids Res 26: 5358-5364.

31. Ramirez, M., Graham, M.A., Blanco-Lopez, L., Silvente, S., Medrano-Soto, A., Blair, M.W., Hernandez, G., Vance, C.P. and Lara, M. 2005. Sequencing and Analysis of Common Bean ESTs. Building a Foundation for Functional Genomics. Plant Physiol 137: 1211-1227.

32. Ray, A., Macwana, S., Ayoubi, P., Hall, L.T., Prade, R. and Mort, A.J. 2004. Negative subtraction hybridization: an efficient method to isolate large numbers of condition-specific cDNAs. BMC Genomics 5: 22.

33. Ronning, C.M., Stegalkina, S.S., Ascenzi, R.A., Bougri, O., Hart, A.L., Utterbach, T.R., Vanaken, S.E., Riedmuller, S.B., White, J.A., Cho, J., Pertea, G.M., Lee, Y., Karamycheva, S., Sultana, R., Tsai, J., Quackenbush, J., Griffiths, H.M., Restrepo, S., Smart, C.D., Fry, W.E., Van Der Hoeven, R., Tanksley, S., Zhang, P., Jin, H., Yamamoto, M.L., Baker, B.J. and Buell, C.R. 2003. Comparative analyses of potato expressed sequence tag libraries. Plant Physiol 131: 419-429.

34. Rudd, S., Mewes, H.W. and Mayer, K.F. 2003. Sputnik: a database platform for comparative plant genomics. Nucleic Acids Res 31: 128-132.

35. Shimada, T., Kita, M., Endo, T., Fujii, H., Ueda, T., Moriguchi, T. and Omura, M. 2003. Expressed sequence tags of ovary tissue cDNA library in Citrus unshiu Marc. Plant Science 165: 167-168.

36. Shiu, S.H. andBleecker, 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: 10763-10768.

37. Singh, A., Evensen, K.B. and Kao, T.-H. 1992. Ethylene synthesis and floral senescence following compatible and incompatible pollinations in *Petunia inflata*. Plant Physiol 99: 38-45.

38. Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. and Efstratiadis, A. 1994. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci U S A 91: 9228-9232.

39. Sprunck, S., Baumann, U., Edwards, K., Langridge, P. and Dresselhaus, T. 2005. The transcript composition of egg cells changes significantly following fertilization in wheat (Triticum aestivum L.). Plant J 41: 660-672.

40. Tarutani, Y., Morimoto, T., Sasaki, A., Yasuda, M., Nakashita, H., Yoshida, S., Yamaguchi, I. and Suzuki, Y. 2004. Molecular characterization of two highly homologous receptor-like kinase genes, RLK902 and RKL1, in Arabidopsis thaliana. Biosci Biotechnol Biochem 68: 1935-1941.

41. Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F. and Komeda, Y. 1996. The Arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. Plant Cell 8: 735-746.

42. Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G. and Tanksley, S. 2002. Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell 14: 1441-1456.

43. Vodkin, L.O., Khanna, A., Shealy, R., Clough, S.J., Gonzalez, D.O., Philip, R., Zabala, G., Thibaud-Nissen, F., Sidarous, M., Stromvik, M.V., Shoop, E., Schmidt, C., Retzel, E., Erpelding, J., Shoemaker, R.C., Rodriguez-Huete, A.M., Polacco, J.C., Coryell, V., Keim, P., Gong, G., Liu, L., Pardinas, J. and Schweitzer, P. 2004. Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. BMC Genomics 5: 73

Chapitre 4

The ScORK17 receptor kinase is predominantly expressed in ovules and is involved in seed development

Hugo Germain, Madoka Gray-Mitsumune, Daniel, P. Matton

Contribution des co-auteurs

Madoka Gray-Mitsumune a réalisé l'essai de déglycosylation, a effectué les expériences. J'ai effectué les analyses de Northern, les transgéniques chez *S. chacoense,* j'ai caractérisé les transgéniques chez *A. thaliana* et la rédaction du papier sauf la section sur l'*in situ* laquelle a été rédigée par Madoka Gray et Daniel Matton.

The ScORK17 receptor kinase is predominantly expressed in ovules and is involved in seed development

Hugo Germain, Madoka Gray-Mitsumune, Daniel P. Matton

Institut de Recherche en Biologie Végétale (IRBV), Département de Sciences Biologiques, Université de Montréal, 4101 Sherbrooke Est, Montreal , QC, H1X 2B2 , Canada

Corresponding author :

Daniel P. Matton

[redacted]

Phone: +1-514-8723967

Fax: +1-514-8729406

Source de financement

This work was supported by NSERC, FQRNT and by the Canada Research Chair program.

Keywords :

Keywords: plant receptor kinase, ovule, heterologous transformation, seed set.

Abstract

Solanum chacoense Ovule Receptor Kinase 17 is a receptor kinase of the LRR-VI subfamily and its expression is highly specific to the female reproductive tissue. There are eleven members of the LRR-VI subfamily in *Arabidopsis thaliana* and, to this day, no members has been attributed a function. We observed that a phylogenetic tree inferred using the kinase domain of LRR-VI subfamily members separates the family into two clades ; one containing an average of 8,2 LRR per protein and a second clade containing an average of 2,7. Additionnally no LRR-VI members are found in tandem repeat on the *Arabidopsis* chromosome, thus suggesting a developmental role for *ScORK17*. *In situ* hybridization was analyzed for *ScORK17* and was shown to be mainly present in the single ovule integument that is developing into the seed coat and in the endothelium. Transient expression also revealed that *ScORK17* is N-glycosylated *in planta*. One T-DNA insertional line for the closest *ScORK17* putative ortholog in *Arabidopsis* was rendered homozygous for phenotypical analyses but no abnormalities were observed in the homozygous population. Overexpression of *ScORK17* in *S. chacoense* did not produce plants with an altered phenotype. However, when a truncated *S. chacoense ScORK17* clone was transformed in *A. thaliana*, the resulting transgenic plants showed reduced seed set, supporting a developmental role for *ScORK17*.

Introduction

Complex organisms, such as plants, use an elaborate communication scheme between cells in order to ensure optimal growth and development. Receptor kinases are well suited as mediators of this intercellular communication mechanism. Their extracellular domain can perceive the information (ligand) coming from the environment or from neighboring cells and relay this information to the nucleus via the phosphorylation of intracellular protein targets through the action of their intracellular kinase domain. During the last decade receptor-like kinases (RLKs) have been associated with nearly all aspects of plant growth and development: pathogen perception (Gomez-Gomez and Boller 2000), hormone signaling (Li and Chory 1997; Matsubayashi et al. 2002), anther and embryo development (Canales et al. 2002), organ initiation (Torii et al. 1996), meristem maintenance (Clark et al. 1993; Clark et al. 1997) and others mechanisms.

Despite all the functions accomplished by receptor kinases, very few RLK have been associated with seed development. Extra sporogenous cells (EXS) and CRINKLY4 are two of the very few RLK that have been found to impair seed development. EXS mutant plants accumulate a larger number of sporogenous cells in the anther and have smaller embryonic cells, delayed embryo development, and smaller mature embryos (Canales et al. 2002). CRINKLY4 was discovered in maize (Becraft et al. 1996), where loss of CRINKLY4 function inhibit aleurone formation in the seed and affects leaf epidermal cell differentiation (Becraft et al. 1996). CRINKLY4 was also characterized in *Arabidopsis* (ACR4) where it was shown that overexpression or antisense suppression of ACR4 lead to embryo malformation (Tanaka et al. 2002) and also disrupted cell layer organization of epidermis-related tissues in leaf and during ovule integument development (Gifford et al. 2003; Watanabe et al. 2004).

Apart from gain of function and loss of function mutants, heterologous transformation can also be highly informative when trying to assign a function to a gene. Heterologous transformation has generally been used to evaluate the functional equivalence or relationship between related genes of different species (Chantha et al. 2006). Several studies in plants have reported the use of heterologous transformation in order evaluate the role of gene in other species when no phenotype were observed in the host specie or when functional redundancy of members of gene families hampers the use of loss-of-function or knock-out lines (Fu et al. 2006; Hirschi 1999; Hirschi et al. 2000). For example, membrane transporter protein originating from multigene family, such as permease, H+-ATPase and cyclic nucleotide-gated cation channel, have been transformed in the yeast in order to study the effect and specificity of individual genes (Alemzadeh et al. 2006; Ali et al. 2006; Weig and Jakob 2000). Heterologous transformation has also been used from one plant specie to another. For example, the CAX1 and CAX2 (calcium exchanger 1 and 2) genes are functionally redundant in Arabidopsis and loss-of-function mutant of either gene do not display a detectable phenotype in Arabidopsis (Hirschi et al. 2000). Additionally, plants overexpressing the CAX2 gene in Arabidopsis do not show an altered protein level of CAX2, but when the AtCAX2 is overexpressed in tobacco, the transgenic plants accumulate higher level of Ca^{2+}, Cd^{2+}, and Mn^{2+} and were more tolerant to elevated Mn^{2+} levels (Hirschi et al. 2000). Other examples of plant/plant heterologous transformation revealing gene function is the expression of a *Solanum tuberosum* MADS-box gene in tobacco (Garcia-Maroto et al. 2000) and the successful expression of the barley HVA1 gene in the perennial grass Agrostis stolonifera that resulted the attenuation of the water-deficit in this specie (Fu et al. 2006).

From a negative screen that focused on weakly expressed transcripts in ovule and ovary tissues following fertilization, we have isolated 30 receptor-like kinases named ScORK1 to 30 for *Solanum chacoense* Ovule Receptor Kinase (Germain et al. 2005). Fine expression analysis through quantitative RT-PCR showed that amongst the 30 RLKs, 28 were predominantly expressed in female reproductive tissues and 23 were transcriptionally

up-regulated following fertilization. Of these, four (*ScORK2, ScORK17, ScORK18*, and *ScORK24)* were members of the LRR-VI family. In this paper we have used expression analysis and heterologous transformation in order to decipher the possible roles of *ScORK17*. Aberrant expression of a truncated version (dominant negative construct) of *ScORK17* in *Arabidopsis thaliana* lead to incomplete seed set in *A. thaliana siliques*. To our knowledge this is the first report of a function for a member of the LRR-VI subfamily of receptor-like kinases.

Material and methods

Sequence analysis

Raw DNA sequence was *in silico* translated using MacVector 8.1.2 (Accelrys, San Diego, CA, USA). The presence of a signal peptide was assessed using SignalP 3.0 (Bendtsen et al. 2004) and the location of the transmembrane domain was positioned using TMHMM 2.0 (Krogh et al. 2001). Putative N-glycosylation sites were analyzed using NetGly 1.0 (Gupta and Brunak 2002) and the structure of the leucine-rich repeat domain was analyzed using ScanSite 2.0 (Obenauer et al. 2003). The phylogenic tree was inferred using a neighbor-joining algorithm (Saitou and Nei 1987) with 10000 bootstraps either on the full length sequences or on the trimmed catalytic domains.

Transgenic plants production and analysis

S. chacoense transgenics were either overexpression lines (OX) with the full length *ScORK17* cDNA cloned downstream of the CaMV 35S promoter or truncated *ScORK17* expressing only the signal peptide, extracellular domain and transmembrane domain (dominant negative constructs) cloned downstream of the CaMV 35S promoter. Prior to transformation all constructs were fully sequenced. *S. chacoense* transformation with *Agrobacterium tumefaciens* strain LBA4404 was carried out as described previously

(Matton et al. 1997). Transgenic *Arabidopsis* plants consisted either of the GABI-Kat 055D10 insertional mutant line for At3g03770 (Rosso et al. 2003) or lines that have been agroinfiltrated with the above mentioned *S. chacoense* constructs. *A. thaliana* Col. plants were grown under long day conditions (16 h light/8 h dark) at 25°C in a growth chamber in 1:1:1 mixture of perlite, vermiculite and topsoil. *S. chacoense* plants were greenhouse grown in topsoil.

Seed count and clearing

Siliques were harvested and placed in deionized water. Dehydration was performed by transferring the siliques in increasing concentration solutions of ethanol (25%, 50%, 75%) for one hour each and left in 100% ethanol overnight. For clearing with methylsalicylate, the siliques were transferred in increasing ratio of methylsalicylate/ethanol solutions (25:75, 50:50, 75:25) for one hour each and left overnight in 100% methylsalicylate. The number of seeds per siliques was counted using a binocular and observed at a magnification of 10X.

RNA expression analyses

Total RNA was isolated as described previously (Jones et al. 1985). The RNA concentration was determined by measuring its absorbance at 260 nm and verified by agarose gel electrophoresis and ethidium bromide staining. Equal loading of total RNA on RNA gel blots was verified with a *S. chacoense* 18S RNA probe. RNA gel blot analyses were performed as described previously (O'Brien et al. 2002). Probe for RNA blot were synthesized using a random-labeled PCR kit (Ambion, Austin, TX, USA). Membranes were exposed at room temperature on an europium screen and scanned on a Typhoon 9200 Phosphorimager (GE Healthcare, Baie dUrfé, QC, Canada). *In situ* hybridizations were performed as described previously (Gray-Mitsumune et al. 2006; O'Brien et al. 2005). Microscopic observations were taken on a Zeiss Axioimager M1 microscope and pictures were taken with a Zeiss Axiocam HRc camera.

Deglycosylation assay

For N-deglycosylation assay, the ScORK17 complete protein sequence was cloned into the tandem affinity purification (TAP) vector pC-TAPa (Rubio et al. 2005) and expressed under the control of a double enhancer CaMV 35S promoter *in planta* through agroinfiltration of *Nicotiana benthamiana* leaves co-infected with the p19 silencing suppressor gene as previously described (Voinnet et al. 2003). Crude extracts from infected leaves collected 3 days after infection were treated with N-glycosidase F (PNGaseF) following the manufacturer instructions (New England Bio-lab, Pickering, ON, Canada). Agroinfection of *Nicotiana benthamiana* leaves was done as previously described with *Agrobacterium tumefaciens* strain C58C1 (Voinnet et al. 2003). The extract was separated on acrylamide gel, transferred on PVDF membrane and detected using the anti-HIS antibody (Sigma, Oakville, ON, Canada).

Immunoblotting procedure

Protein sample were run on a 8% polyacrylamide gel at 150 V for 1 hour in running buffer (3,5 mM SDS, 25 mM Tris and 192 mM glycine in one liter) and then transferred to PVDF membrane in transfer buffer (25 mM Tris and 192 mM glycine and 20% methanol in one liter). Immunoblotting was done as follows: PVDF membranes were hydrated in methanol for 10 seconds, equilibrated in TBS (20 mM Tris pH7,5, 150 mM NaCl) for 15 minutes, blocked for 1 hour in TBST+Milk (TBS + 0,02% Tween 20 and 2,5% defatted dry-milk), incubated for 1 hour with the anti-His antibody in TBST+Milk, washed three times 15 minutes in TBST, incubated with the monoclonal anti-rabbit IgG γ-chain specific conjugated with the peroxidase (Sigma) for one hour and washed three times 15 minutes in TBST. Visualization was done using ECL Plus (GE Healthcare, Baie d'Urfé, QC, Canada) and exposed on a Bioflex Econo Film (Interscience, Laval, Qc, Canada).

Results

ScORK17 sequence analysis and phylogeny

The EST corresponding to ScORK17 (DN980445) was fully sequenced and most probably contained the full-length or near full-length ScORK17 cDNA (2665 bp, Genbank accession number EF517229) since it corresponded to the length obtained by RNA gel blot analysis (~2600 nt, see Fig. 9). The longest open reading frame deduced from the cDNA sequence coded for a 85,8 kD protein of 778 amino acids with a predicted pI of 8.42. The first methionine found is located at position 53 of the longest translated product (nucleotide position 159-161). Although no in frame stop codon was found upstream of the first methionine, we believe that the first 158 nt are part of the 5' untranslated region (UTR) for the following two reasons. Firstly, although the first 52 amino acids are in frame with the first methionine, BLAST analyses with this region did not retrieved any similar sequences from publicly available databases. Secondly, using the *ScORK17* nucleotide sequence, a seed derived tomato (*Solanum lycopersicum*) EST (SGN-U342343) was found in the SOL Genomics Network database (http://www.sgn.cornell.edu/). This single EST aligned with the *ScORK17* 5' end with 94% nucleotide sequence identity. Alignment of the translated product of SGN-U342343 with ScORK17 showed 95% amino acid sequence identity (98% similarity) starting from the first methionine found in each cDNA. Furthermore, in tomato, in frame stop codons were found upstream of the first methionine and the length of the 5'UTRs were comparable, 158 nt for *ScORK17* and 130 nt for SGN-U342343. This data strongly suggests that methionine 53 starts the open reading frame of *ScORK17*. More in depth sequence analysis revealed that *ScORK17* encodes a receptor kinase of the LRR-VI family (Shiu and Bleecker 2003) and exhibits the typical structure of LRR-RLK: a predicted signal peptide (1-27), a transmembrane domain (394-416), a kinase domain (483-757), and a short C-terminal tail (758-778) (Fig. 8a). A more exhaustive analysis of the extracellular domain revealed that it contained ten predicted leucine-rich repeats (LRR)

located between amino acids 110 and 341 (Fig. 8a , b) (Obenauer et al. 2003). These LRRs are flanked by two cysteine pairs. The N-terminal cysteine pair (CNSEPNTALTLMC) is located 60 amino acids from the start codon and is very similar to the cysteine pair (CNSEPSPSLTVVC) in the closest *Arabidopsis* ortholog (At3g03770, see also below for phylogenetic analysis). The C-terminal cysteine pair (CLSNKEQWQHPYSFC) is located 14 amino acids downstream of the last LRR (LRR10). Seven N-glycosylation sites are also predicted from the ectodomain sequence (Gupta and Brunak 2002). Analysis of the kinase catalytic domain revealed that ScORK17 could be classified as an ACF kinase (alternative catalytic function). In *Arabidopsis*, these kinases comprises 20% of the *Arabidopsis* kinome and their catalytic domain lack one or more of three highly conserved residues (K/D/D) thought to be required for catalytic activity in subdomains II (K), VI (D), and VII (D) (Dardick and Ronald 2006). Analysis of all LRR-VI kinases in both *Arabidopsis* and rice revealed that they all belong to the ACF kinase group (Shiu et al. 2004).

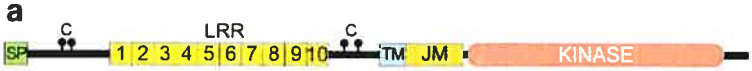
BLASTP analysis revealed that ScORK17 shared 72% similarity (54% identity) with a *Medicago truncatula* RLK (ABD28527 or AC148965), 71% similarity (52% identity) with the At3g03770 *Arabidopsis* LRR-VI family RLK (Shiu and Bleecker 2003), and 66% similarity (46% identity) with a rice (*Oryza sativa*) LRR-VI family RLK (Osi021700.1) (Shiu et al. 2004). The second closest hit in *Arabidopsis*, At2g02780, shares strikingly less sequence similarity (54%), leaving At3g03770 as the most likely *Arabidopsis* ortholog for ScORK17. To better determine the closest orthologs in *Arabidopsis* and rice for ScORK17 we performed a phylogenetic analysis with all the LRR-VI family members and with the Clavata1 RLK from LRR-XI family. Since these two organisms have been fully sequenced, LRR-VI family members have already been annotated (Shiu et al. 2004). Because in this family the LRRs are quite variable in numbers (from 3 to 10 in *Arabidopsis*, and from 0 to 10 in rice as predicted by Scansite 2.0) the phylogenetic analysis was performed with the trimmed kinase domains only (without juxtamembrane and C-terminal domain), as used previously to determine their family grouping (Shiu et al. 2004). This explains why some LRR-VI family members in rice have

no predicted LRR. The phylogenetic analysis is shown in Fig. 8c. *ScORK17* clusters with a highly supported group of four RLKs, including the ones retrieved from the BLASTP as being the most similar. Using a motif prediction algorithm (Scansite 2.0) we also determined the number of LRRs present in each RLK under high stringency conditions (Fig. 8c). Although the phylogeny was based on the kinase domain only, *ScORK17* clustered with RLKs also having the most numerous LRRs in their ectodomains, suggesting a tight functional constrain and possibly that this subgroup was already present before the divergence between eudicots and monocots. Interestingly, phylogenetic associations between the kinase domain mostly reflected the LRR organization with RLKs possessing numerous LRRs (8-10) clustering together, while RLKs possessing fewer LRRs ($\leq$6) also clustering together in two separate highly supported branches (branches A and B on Fig. 8c). LRR-VI RLKs from branch A (8.2 $\pm$ 2.3 LRRs) have three-fold more LRRs than the ones on the B branch (2.7 $\pm$ 2.4 LRRs).

Figure 8. Illustration *ScORK17* primary structure, aligment of the LRR and phylogenetic tree of the LRR-VI members.

A. Illustration *ScORK17* primary structure. The dots correspond the cystein pairs. SP, signal peptide; LRR, leucine-rich repeat; TM, transmembrane domain; JM, intracellular juxtamembrane domain; KINASE, kinase catalytic domain. **B** The ten LRRs are aligned to emphasize most frequent amino acids found at regularly spaced positions. Dark grey boxes represent conserved residues in more than half the sequences while light gray boxes represents conservative substitutions. **C** A neighbour-joining clustering from the trimmed kinase domain of the ScORK17 receptor kinase and all the LRR-VI subfamily RLK members from *A. thaliana* and rice, as well as a highly similar RLK from *Medicago truncatula*. The kinase domain from the CLAVATA1 RLK from the LRR-XI subfamily was chosen as the out group. Bootstrap values were calculated from 10000 replicas. The branch corresponding to the ScORK17 most probable orthologs is highlighted in bold. Numbers of LRR predicted for each receptor kinase appear on the right of each accession number or gene name.

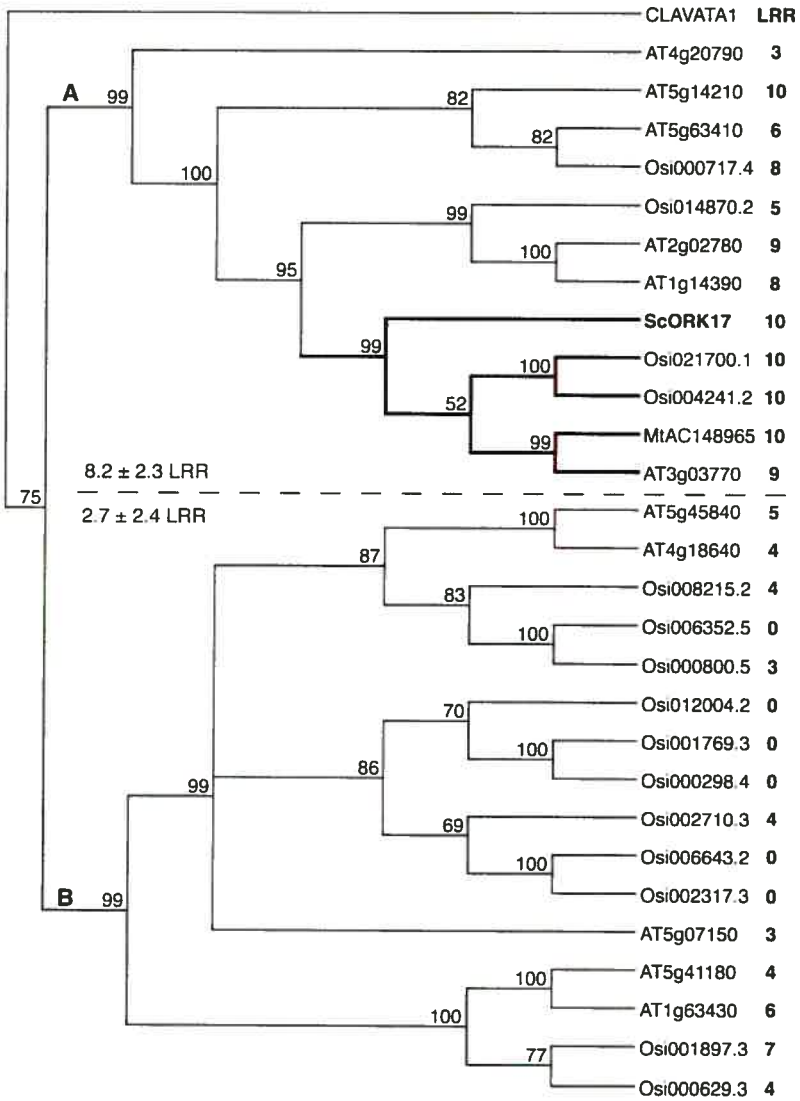
a



b

LRR1	110	**L**	K	V	**L**	S	**L**	V	S	L	G	L	R	**G**	T	L	**P**	K	K	I	G	F	**L**	S	132
LRR2	134	**L**	E	I	I	N	I	S	S	**N**	F	F	Y	**G**	E	I	**P**	G	E	I	S	Y	**L**	K	156
LRR3	158	**L**	Q	T	**L**	I	**L**	D	D	**N**	K	F	T	**G**	Q	V	**P**	E	G	V	D	L	**L**	H	180
LRR4	182	**L**	S	V	**L**	S	F	K	N	**N**	S	F	S	**G**	S	L	**P**	N	S	L	S	N	**L**	Q	204
LRR5	206	**L**	R	I	**L**	S	I	S	G	**N**	N	F	S	**G**	V	V	**P**	N	L	H	N	L	S		227
LRR6	229	**L**	Q	V	**L**	D	L	E	S	**N**	-	N	L	**G**	P	N	F	P	N	I					246
LRR7	250	**L**	V	S	**L**	V	L	R	K	**N**	K	F	S	L	G	V	**P**	K	E	L	S	S	C	Y	272
LRR8	274	**L**	K	K	**L**	D	I	S	S	**N**	E	L	V	**G**	P	F	S	P	T	V	L	S	**L**	P	296
LRR9	298	**L**	S	Y	**L**	D	I	S	G	**N**	K	L	T	**G**	K	L	L	K	N	V	T	C	S	Q	320
LRR10	322	**L**	S	F	V	N	L	S	S	**N**	Y	L	T	**G**	E	L	**P**	D	C	L	K				341

c



ScORK17 expression analysis and deglycosylation assay

ScORK17 mRNA expression was first assessed by RNA gel blot analysis using a variety of vegetative and reproductive tissues (Fig 9a). In order to get a more detailed picture of *ScORK17* expression in the developing fruit, the structure was dissected in ovules plus plancenta and pericarp separately. *ScORK17* mRNA expression is absent from most non-female reproductive tissues, except for a very weak expression in pollen, although this faint band was not reproducibly observed in other RNA gel blot analysis experiments (data not shown), confirming previous results obtained by quantitative real-time PCR (Germain et al. 2005). In ovary tissues, *ScORK17* expression is also barely detectable before fertilization but increases strongly following fertilization, in full concordance with previous quantitative RT-PCR analyses. Peak accumulation occurs four days after pollination (96 HAP) which corresponds to ovules bearing mostly 4-celled proembryos (Fig. 9a, b and Lafleur, E. and Matton, D.P., unpublished observations). *ScORK17* mRNA accumulation decreases thereafter and reaches pre-fertilization levels by 16 DAP. *ScORK17* mRNA signal is detected both in ovules and placenta as well as in the fruit pericarp.

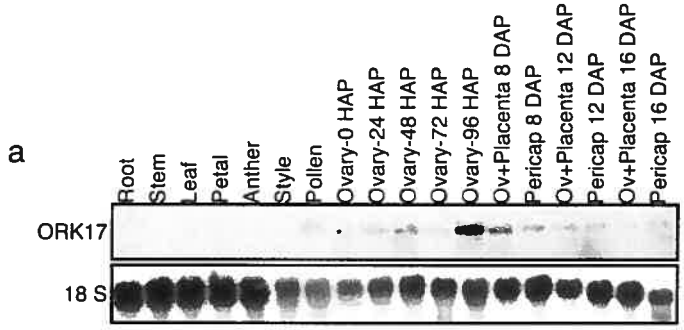
In order to determine more precisely the spatial expression pattern of *ScORK17*, *in situ* RNA hybridizations were performed using ovaries taken 4 days after pollination, when peak accumulation occurs (Fig. 9a). *Solanum chacoense* produces unitegmic-tenuinucellate ovules, a trait that occurs almost universally in the asterid clade (Albach et al. 2001). As for other solanaceous species, the nucellus is completely utilized during megagametophyte development (Dnyansagar and Cooper 1960; Lee and Cooper 1958; Souèges 1907). The embryo sac is thus in direct contact with the endothelium, also called the integumentary tapetum, a layer that is formed as the result of the transformation of the integumental epidermis into specific secretory cells on the nucellus side of the megagametophyte and that retains meristematic properties (Batygina 2002). The endothelium layer is easily recognized with its large isodiametric cells with clearly defined nucleus and dense cytoplasm (Fig. 9b, panel i and ii, red arrowheads). The remaining layers of the integument

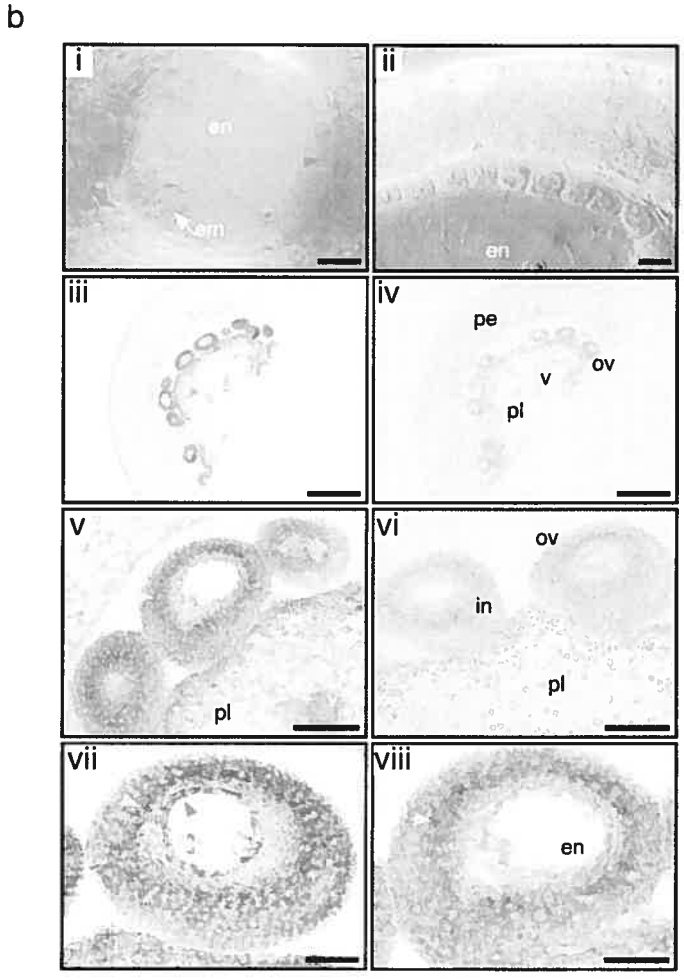
are mostly made of parenchymous cells. Immediately adjacent to the endothelium, the inner integumentary cells appear crushed, probably due to the expansion of the developing endosperm (Fig, 9b, panel ii, green arrowhead). This is consistent with what has been described in other solanaceous species (Dnyansagar and Cooper 1960; Lee and Cooper 1958; Sin et al. 2006).

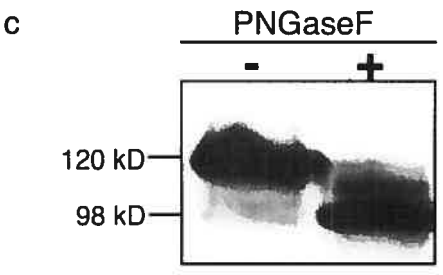
As seen in Fig. 9b panels iii, v, and vii, strong *ScORK17* mRNA signal was detected in the single ovule integument that is developing into the seed coat. By this stage, *Solanum* ovule integument would have undergone a few rounds of cell division (Olson 1988; Sin et al. 2006). Also, ovules contain cellularized endosperm that can be seen as a large vacuolated area in the center (Fig. 9b panels i, ii, v, vii, viii). As observed through DIC microscopy (Fig. 9b panel ii), the cell layers just outside of the endothelium also appear crushed, and this is more prominently visible at the ovule's chalazal end (Fig. 9b, panel viii). Interestingly, the strongest *ScORK17* mRNA signal was detected in the cell layer next to the crushed cells, forming a more densely stained ring (Fig. 9b, panels v, vii, and viii, yellow arrowhead), while the crushed cell layer was much less stained and appeared as a light blue ring around the endothelium (Fig. 9b, panels vii and viii, green arrowhead). There was also an intense signal in the endothelium (Fig. 9 panel vii, read arrowhead). *ScORK17* mRNA signal was also present in the endosperm; however, due to large vacuoles in the cellular endosperm, it was not as strong as the one observed in the integument (Fig. 9b panels vii and viii). Occasionally, an intense signal was observed around the endosperm nuclei (data not shown). In paraffin embedded sections where embryos could be identified, no clear *ScORK17* mRNA signal could be found (Fig. 9b panel viii, white arrow). *ScORK17* mRNA signals were also detected in the placenta, mainly in the placental epidermis and in the placenta vasculature (Fig. 9b panels iii and v), and light staining was consistently observed in the ovary pericarp. *ScORK17* sense probe hybridization were performed as negative controls gave no signal in pericarp, ovule, or placenta (Fig. 9b panels iv and vi).

Figure 9. Expression profile of ScORK17 and protein deglycosylation assay.

A RNA gel blot analysis from vegetative and reproductive tissues. Equal loading of each sample was verified by ethidium bromide staining prior to gel transfer (not shown) and by 18S rRNA hybridization. **B** *In situ* mRNA localization of *ScORK17* transcripts in fertilized ovaries. Ovaries were harvested 4 days after pollination and *in situ* hybridization was performed using either antisense (iii, v, vii, and viii) or sense (iv and vi) probes from the *ScORK17* cDNA clone. Digoxigenin labeling is seen as bluish purple staining. Hybridizations were performed under identical conditions for both antisense and sense probes. Scale bars: 25 μm (i); 50 μm (vii and viii); 100 μm (ii, v, and vi) and 500 μm (iii and iv). Abbreviations: em, embryo; en, endosperm; in, integument; ov, ovule; pe, pericrap; pl, placenta; v, vasculature. Arrowhead indicates the endothelium layer. Red arrowheads indicate the endothelium layer surrounding the embryo sac; green arrowheads indicate the crushed layer of integument cells immediately surrounding the endothelium; yellow arrowheads indicate the hybridization dense cell layer surrounding the crushed cells layer; white arrow indicates the 4-celled proembryo 4 DAP. **C** Deglycosylation assay from transiently expressed ScORK17 protein in *Nicotiana benthamiana* leaves through agroinfiltration. A ~22 kD gel shift is observed after treatment (+) with PNGase F (N-glycosidase F). (-) lane, sample without PNGase F treatment.


a
Root
Stem
Leaf
Petal
Anther
Style
Pollen
Ovary-0 HAP
Ovary-24 HAP
Ovary-48 HAP
Ovary-72 HAP
Ovary-96 HAP
Ov+Placenta 8 DAP
Pericap 8 DAP
Ov+Placenta 12 DAP
Pericap 12 DAP
Ov+Placenta 16 DAP
Pericap 16 DAP
ORK17
18 S


b
i
en
em
ii
en
iii
iv
pe
ov
v
pl
v
pl
vi
ov
in
pl
vii
viii
en


c
PNGaseF
-
+
120 kD
98 kD

Since several N-glycosylation sites were predicted in ScORK17 from the primary amino acid sequence, a N-deglycosylation assay was performed to determine if ScORK17 is a N-glycosylated receptor kinase. The *ScORK17* open reading frame (ORF) fused to a 9xMyc, a hexahistidine (6xHis) and a protein A binding domains tags was transiently expressed through agroinfiltration in *Nicotiana benthamiana* leaves and a crude protein extract was submitted to deglycosylation with N-glycosidase F (PNGaseF). PNGaseF cleaves between the innermost GlcNAc and asparagine residues of high mannose, hybrid, and complex oligosaccharides from N-linked glycoproteins. A mobility shift of ~22 kDa (120 kDa to 98 kDa) was detected on the membrane with an anti-His antibody (Fig. 10c), suggesting that the protein is N-glycosylated. Such a difference in molecular weight is consistent with the presence of several glycosylation sites, as predicted from the primary amino acid sequence. The theoretical molecular weight of the tagged ectodomain of ScORK17 with the signal peptide removed is 76 kDa while the observed molecular weight of the deglycosylated ScORK17 protein is 98 kDa. This difference between theoretical and observed molecular weight (22 kDa) could be caused by additional post-translational modifications or N-glycosylations resistant to PNGaseF (Altmann et al. 1995).

Transgenic plant analysis.

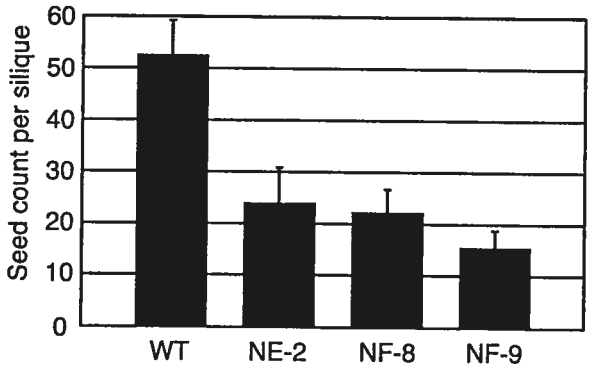
In order to obtain information with regards to the function of *ScORK17* we generated transgenic plants using either the full length *ScORK17* clone or a truncated construct in which the intracellular domain was deleted. This later construct was designed to act as a dominant negative form of *ScORK17*. We obtained plants that strongly overexpressed the transgene from both constructs (data not shown) and these plants were analyzed for phenotypical abnormalities. No defects were observed when the *S. chacoense* plants were transformed with the *ScORK17 S. chacoense* clones. Given the fact that our construct was under the control of the 35S promoter we speculated that, either the transgenes were undergoing a mechanism of post-translational gene-silencing to alleviate an unfavorable effect of the high *ScORK17* transcript levels, or that phenotypical defects

were either absent or undetected. We had previously shown that the 35S promoter was highly active in leaf, anther and ovary tissues (Gray-Mitsumune et al. 2006; O'Brien et al. 2007).

However, when we transformed the *ScORK17* truncated construct in *A. thaliana*, we observed that siliques contained significantly less seeds. The mutant plants obtained from independent transformed lines had an average of 15.7 to 24.2 seeds per siliques compared to an average of 52.3 seeds per siliques in the wild-type (Fig. 10). We confirmed expression of the transgene in the mutant plants using semi-quantitative RT-PCR (data not shown). The average number of seeds per siliques in the mutant lines was 22.2 (SD=6.2) compared to 52.3 (SD=8.0) in the wild-type. We carried a careful observation to detect whether the mutant siliques also contained aborted seeds but none was observed, suggesting that the defect appeared earlier during ovule development. We hypothesized, as mentioned in Diener (2000), that the differences in DNA sequence between the foreign gene and the host might have circumvented the host gene-silencing mechanism, enabling the observation of a mutant phenotype (Hirschi et al. 2000). No phenotype other than wild-type was observed in the only insertional T-DNA mutant line available at the time, possibly due to the fact that the RLK family LRR-VI contains several members (11) and a related member could take over the function of the *Arabidopsis* gene corresponding to *ScORK17*.

Figure 10. Seed count in wild-type siliques versus mutant siliques overexpressing the truncated ScORK17

Seed count in wild-type siliques versus siliques from three independent transformed lines (NE-2, NF-8, NF-9) overexpressing the truncated *ScORK17* construct comprising only the signal peptide, extracellular and transmembrane domains. The number of siliques analyzed was n = 54 for *A. thaliana* WT Columbia ecotype. For the dominant negative transgenic *A. thaliana* lines, n= 30 for the NE-2 line, n=25 for the NF-8 line, and n = 10 for the NF-9 line.



Discussion

Following fertilization, the embryo (and the fruit) starts to grow at a very rapid pace. This extremely active growth necessitates communication between the different structures of the fruit, including the embryo, the placenta, and the endosperm in order to maintain a synchronized and coordinated development. Biomolecules, most likely peptides, could mediate this communication through receptor kinases. To investigate the signal transduction events that take place during early embryogenesis in plants, we constructed a cDNA library enriched for weakly abundant RNA expressed in developing ovules (Germain et al. 2005). Despite the fact that receptor kinases have been associated with nearly all aspect of plant development very few RLK have been associated with embryo development. We investigated the function of *ScORK17* since it presented an expression pattern highly specific to the early developing fruit following fertilization and thus was likely to be involved in fruit or embryo development.

Sequence analysis revealed that *ScORK17* is a member of the LRR-VI subfamily and displayed all the characteristics and domains found in plant receptor kinases. The phylogenetic analysis of all the LRR-VI members in rice and Arabidopsis highlighted the fact that two major clades are observed in the LRR-VI subfamily. This separation corresponds to the presence of a small or a large number of leucine-rich repeats although the phylogeny was performed with the kinase domain only: one clade had an average of 8.2 LRRs per protein while the other one had an average of 2.7 LRRs per protein (Fig. 8c). Shiu et al. (2001) had previously observed that when the kinase domain of all the *Arabidopsis* receptor kinases was used to build a phylogenetic tree, members that had a similar extracellular domain type (LRR, S-domain, thaumatine, etc) tend to fall within the same clade. Grouping based on the structural arrangement of LRRs or on the location of introns in the extracellular domains of the individual RLKs was also observed (Shiu and

Bleecker 2001). According to Shiu et al. (2004) the LRR-VI subfamily forms an ancestral unit, representing a clade (group) present before the monocot–dicot split (Shiu et al. 2004). Their analysis also led to the observation that genes involved in developmental processes were never found as tandem repeat in the genome whereas genes involved in resistance or defense response arose from duplication events which can be measured from a deviation to the 1:1 *Arabidopsis*–rice gene ratios (11 *Arabidopsis* : 15 rice in LRR-VI). Although nearly half of the LRR-VI *Arabidopsis* members are located on chromosome five none of them are found in tamdem repeat (data not shown), hence supporting a developmental role for LRR-VI members and ScORK17.

RNA gel blot analysis (Fig. 9a) and quantitative RT-PCR results (Germain et al. 2005) have shown that *ScORK17* is exclusively expressed in female reproductive tissues, and is strongly up-regulated following fertilization. A finer analysis using RNA in situ hybridization revealed that the *ScORK17* hybridization signal was strong in the ovule's integument and in the endothelium, a layer that is formed as the result of the transformation of the integumental epidermis into specific secretory cells on the nucellus side of the megagametophyte (Batygina 2002). The endothelium has been proposed to serve multiple functions, mainly nutrient transport to the embryo sac and, in species where an endothelium persists in maturing seeds, like numerous species in the Solanaceae, a protective function toward the embryo and endosperm has also been proposed (Batygina 2002). Although no clear specialized cellular structures can be observed in the ovule's single integument, strongest *ScORK17* mRNA signal was detected in the cell layer next to the crushed cells surrounding the endothelium, forming a densely stained ring (Fig. 9b, panels v, vii, and viii), while the crushed cell layer was much less stained and appeared as a light blue ring around the endothelium (Fig. 9b, panels vii and viii). In *Solanum americanum*, Sin et al. (2006) described this layer as being gradually crushed by the pressure exerted by expanding endosperm, while Lee and Cooper (1958) considered that the inner integumentary cells

immediately adjacent to the endothelium disintegrate. From our DIC pictures of cleared ovules, these cells appeared thin and elongated, as if they had been flattened, and large intercellular spaces between these cells were created, either the consequence of the cell's flattening or due to a decreased number of cells, possibly explaining the weaker staining observed by in situ hybridization analysis. Weak *ScORK17* expression was also observed throughout the pericarp and in the placenta (Fig. 9b, panel iii). In the latter tissue, *ScORK17* expression was weak everywhere except in the vasculature and was as strong in the placental epidermis as in the ovule's integument (Fig. 9b, panel v). Overall, the strong and very specific staining in ovules suggest that *ScORK17* plays a role in post-pollination ovule development, mainly in integument growth, but most probably not directly in embryo or endosperm development since these tissues showed no or weak staining. Contrary to the ACR4 receptor kinase gene involved in epidermis tissues in leaf and during ovule integument development (Gifford et al. 2003; Watanabe et al. 2004) and that is ubiquitously expressed in *Arabidopsis* (Tanaka et al. 2002), *ScORK17* displayed a more restricted expression pattern, mainly in the ovule integument and in the epidermal layer of the placenta. Strong mRNA expression in the placenta's epidermal layer that is contiguous with the ovule's integument, and expression in the ovule in layers derived from epidermal cells might suggest that the ScORK17 kinase is expressed in cells that express an epidermal cell fate but restricted to the female reproductive tissues.

Deglycosylation assay using transient expression in *Nicotiana benthamiana* leaves and PNGaseF treatement revealed that the predicted putative N-glycosylation sites were glycosylated in vivo. Yet, following the 22 kDa mobility shift observed after PNGaseF deglycosylation, the ScORK17 protein still remained 22 kDa larger than predicted. In plants and insects N-glycans (susceptible to PNGaseF) can be further modified with fucose in an α1,3-linkage to the asparagine which are then resistant to PNGaseF degradation (Altmann et al. 1995; Kubelka et al. 1993; Tretter et al. 1991). Such modifications were also noted when the ectodomain of the phytosulfokine receptor was expressed in tobacco BY-2 cells (Shinohara et al. 2007). The presence of extensive glycosylation could also

explain that our attempt to obtain a polyclonal antibody from rabbit immunized with an *in vitro* expressed extracellular domain (devoid of post-translational modifications) could not recognize the plant produced ScORK17 protein (data not shown).

In order to establish whether we could demonstrate the function of *ScORK17 in planta* we generated transgenic plants in *S. chacoense* and *A. thaliana* expressing either the full length *ScORK17* or a truncated *ScORK17* construct that contained only the signal peptide followed by the extracellular and transmembrane domains. Although we obtained plants that overexpressed the transcript in *S. chacoense*, no phenotype other than wild-type could be observed. However, when we transformed *A. thaliana* with the truncated *ScORK17* construct, plants with a reduced seed set were obtained, confirming that *ScORK17* has a role in fruit and/or seed development. It is likely that we did not observed a phenotype in *Solanum chacoense* because the ScORK17 protein level could be strictly regulated and degradation of the supernumerary transcript and/or protein, or truncated protein in the case of the dominant negative construct would occur, through a silencing mechanism. This mechanism would have been avoided in *A. thaliana* due the difference in sequence homology between *A. thaliana* and *S. chacoense* (Diener and Hirschi 2000). Heterologous transformation from plant to yeast (Alemzadeh et al. 2006; Ali et al. 2006; Weig and Jakob 2000) or from one plant species to another is of common use in order to uncover phenotype when silencing mechanism precludes protein accumulation in the host species either in a transformed population (Fu et al. 2006; Hirschi 1999; Hirschi et al. 2000) or when silencing affects subsequent generation (Liu and Zhang 2004).

We presented evidence that *ScORK17*, a plant receptor kinase of the LRR-VI subfamily, is involved in plant development. *ScORK17*, whose expression is highly specific to the fruit and more precisely to the ovule integument, partially block seed formation when expressed in *Arabidopsis*. The results presented in this paper represent to our knowledge, the first functional study for a RLK member of the LRR-VI subfamily, but does not allow at this stage any speculation with regards to the precise mechanism in which ScORK17 acts

during seed formation and ovule development. Additional experiment involving detailed microscopic analysis will be required to pinpoint the precise time at which de developmental defect takes place and in which signaling pathway ScORK17 functions.

Acknowledgements

We thank Edith Lafleur for providing DIC images of fertilized ovules and for in situ hybridization pictures. We also thank Gabriel Téodorescu for plant care and maintenance. H. Germain was the recipient of Ph. D. fellowships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Fonds Québécois de la Recherche en Nature et Technologie (FQRNT). This work was also supported by the Canadian Foundation for Innovation and by the Canada Research Chair program (D. P. Matton).

Bibliography

1. Albach DC, Soltis PS, Soltis DE (2001) Patterns of embryological and biochemical evolution in the asterids [Review]. Syst. Bot 26: 242-262

2. Alemzadeh A, Fujie M, Usami S, Yoshizaki T, Oyama K, Kawabata T, Yamada T (2006) ZMVHA-B1, the gene for subunit B of vacuolar H+-ATPase from the eelgrass Zostera marina L. Is able to replace vma2 in a yeast null mutant. Journal of bioscience and bioengineering 102: 390-395

3. Ali R, Zielinski RE, Berkowitz GA (2006) Expression of plant cyclic nucleotide-gated cation channels in yeast. Journal of experimental botany 57: 125-138

4. Altmann F, Schweiszer S, Weber C (1995) Kinetic comparison of peptide: N-glycosidases F and A reveals several differences in substrate specificity. Glycoconjugate journal 12: 84-93

5. Batygina TB (ed) (2002) Embryology of flowering plants: Terminology and concepts. . Science Publishers, Inc., Endfield, NH, USA

6. Becraft PW, Stinard PS, McCarty DR (1996) CRINKLY4: A TNFR-like receptor kinase involved in maize epidermal differentiation. Science 273: 1406-1409

7. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783-795

8. Canales C, Bhatt AM, Scott R, Dickinson H (2002) EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in Arabidopsis. Curr Biol 12: 1718-1727

9. Chantha SC, Emerald BS, Matton DP (2006) Characterization of the plant Notchless homolog, a WD repeat protein involved in seed development. Plant Mol Biol 62: 897-912

10. Clark SE, Running MP, Meyerowitz EM (1993) CLAVATA1, a regulator of meristem and flower development in Arabidopsis. Development (Cambridge, England) 119: 397-418

11. Clark SE, Williams RW, Meyerowitz EM (1997) The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. Cell 89: 575-585

12. Dardick C, Ronald P (2006) Plant and animal pathogen recognition receptors signal through non-RD kinases. PLoS Pathog 2: e2

13. Diener A, Hirschi H (2000) Heterologous expression for dominant-like gene activity. Trends in plant science 5: 10-11

14. Dnyansagar VR, Cooper DC (1960) Development of the seed of Solanum Phureja. Am. J. Bot. 47: 176-186

15. Fu D, Huang B, Xiao Y, Muthukrishnan S, Liang GH (2006) Overexpression of barley hva1 gene in creeping bentgrass for improving drought tolerance. Plant Cell Rep

16. Garcia-Maroto F, Ortega N, Lozano R, Carmona MJ (2000) Characterization of the potato MADS-box gene STMADS16 and expression analysis in tobacco transgenic plants. Plant Mol Biol 42: 499-513

17. Germain H, Rudd S, Zotti C, Caron S, O'Brien M, Chantha SC, Lagace M, Major F, Matton DP (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: 515-532

18. Gifford ML, Dean S, Ingram GC (2003) The Arabidopsis ACR4 gene plays a role in cell layer organisation during ovule integument and sepal margin development. Development (Cambridge, England) 130: 4249-4258

19. Gomez-Gomez L, Boller T (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Molecular cell 5: 1003-1011

20. Gray-Mitsumune M, O'Brien M, Bertrand C, Tebbji F, Nantel A, Matton DP (2006) Loss of ovule identity induced by overexpression of the fertilization-related kinase 2 (ScFRK2), a MAPKKK from Solanum chacoense. Journal of experimental botany 57: 4171-4187

21. Gupta R, Brunak S (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. Pacific Symposium on Biocomputing, pp 310-322

22. Hirschi KD (1999) Expression of Arabidopsis CAX1 in tobacco: altered calcium homeostasis and increased stress sensitivity. The Plant cell 11: 2113-2122

23. Hirschi KD, Korenkov VD, Wilganowski NL, Wagner GJ (2000) Expression of arabidopsis CAX2 in tobacco. Altered metal accumulation and increased manganese tolerance. Plant Physiol 124: 125-133

24. Jones JDG, Dunsmuir P, Bedbrook J (1985) High level expression of introduced chimeric genes in regenerated transformed plants. EMBO Journal 4: 2411-2418

25. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305: 567-580

26. Kubelka V, Altmann F, Staudacher E, Tretter V, Marz L, Hard K, Kamerling JP, Vliegenthart JF (1993) Primary structures of the N-linked carbohydrate chains from honeybee venom phospholipase A2. European journal of biochemistry / FEBS 213: 1193-1204

27. Lee JH, Cooper DC (1958) Seed development following hybridization between diploid Solanum species from Mexico, Central and South America. Am. J. Bot. 45: 104-110

28. Li J, Chory J (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell 90: 929-938

29. Liu Y, Zhang S (2004) Phosphorylation of 1-aminocyclopropane-1-carboxylic acid synthase by MPK6, a stress-responsive mitogen-activated protein kinase, induces ethylene biosynthesis in Arabidopsis. The Plant cell 16: 3386-3399

30. Matsubayashi Y, Ogawa M, Morita A, Sakagami Y (2002) An LRR receptor kinase involved in perception of a peptide plant hormone, phytosulfokine. Science 296: 1470-1472

31. Matton DP, Maes O, Laublin G, Xike Q, Bertrand C, Morse D, Cappadocia M (1997) Hypervariable Domains of Self-Incompatibility RNases Mediate Allele-Specific Pollen Recognition. The Plant cell 9: 1757-1766

32. O'Brien M, Bertrand C, Matton DP (2002) Characterization of a fertilization-induced and developmentally regulated plasma-membrane aquaporin expressed in

reproductive tissues, in the wild potato Solanum chacoense Bitt. Planta 215: 485-493

33. O'Brien M, Chantha SC, Rahier A, Matton DP (2005) Lipid signaling in plants. Cloning and expression analysis of the obtusifoliol 14alpha-demethylase from Solanum chacoense Bitt., a pollination- and fertilization-induced gene with both obtusifoliol and lanosterol demethylase activity. Plant Physiol 139: 734-749

34. O'Brien M, Gray-Mitsumune M, Kapfer C, Bertrand C, Matton DP (2007) The ScFRK2 MAP kinase kinase kinase from Solanum chacoense affects pollen development and viability. Planta 225: 1221-1231

35. Obenauer JC, Cantley LC, Yaffe MB (2003) Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic acids research 31: 3635-3641

36. Olson AR (1988) Postpollination placental development of a diploid Solanum tuberosum. Can. J. Bot. 66: 1813-1817

37. Rosso MG, Li Y, Strizhov N, Reiss B, Dekker K, Weisshaar B (2003) An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol 53: 247-259

38. Rubio V, Shen Y, Saijo Y, Liu Y, Gusmaroli G, Dinesh-Kumar SP, Deng XW (2005) An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. Plant J 41: 767-778

39. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4: 406-425

40. Shinohara H, Ogawa M, Sakagami Y, Matsubayashi Y (2007) Identification of ligand binding site of phytosulfokine receptor by on-column photoaffinity labeling. The Journal of biological chemistry 282: 124-131

41. Shiu SH, Bleecker AB (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proceedings of the National Academy of Sciences of the United States of America 98: 10763-10768

42. Shiu SH, Bleecker AB (2003) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. Plant Physiol 132: 530-543

43. Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH (2004) Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. The Plant cell 16: 1220-1234

44. Sin SF, Yeung EC, Chye ML (2006) Downregulation of Solanum americanum genes encoding proteinase inhibitor II causes defective seed development. Plant J 45: 58-70

45. Souèges R (1907) Développement et structure du tégument séminal chez les Solanacées. Ann. Sci. Nat. Bot. XII 6: 1-124

46. Tanaka H, Watanabe M, Watanabe D, Tanaka T, Machida C, Machida Y (2002) ACR4, a putative receptor kinase gene of Arabidopsis thaliana, that is expressed in the outer cell layers of embryos and plants, is involved in proper embryogenesis. Plant & cell physiology 43: 419-428

47. Torii KU, Mitsukawa N, Oosumi T, Matsuura Y, Yokoyama R, Whittier RF, Komeda Y (1996) The Arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. The Plant cell 8: 735-746

48. Tretter V, Altmann F, Marz L (1991) Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha 1----3 to the asparagine-linked N-acetylglucosamine residue. European journal of biochemistry / FEBS 199: 647-652

49. Voinnet O, Rivas S, Mestre P, Baulcombe D (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J 33: 949-956

50. Watanabe M, Tanaka H, Watanabe D, Machida C, Machida Y (2004) The ACR4 receptor-like kinase is required for surface formation of epidermis-related tissues in Arabidopsis thaliana. Plant J 39: 298-308

51. Weig AR, Jakob C (2000) Functional identification of the glycerol permease activity of Arabidopsis thaliana NLM1 and NLM2 proteins by heterologous expression in Saccharomyces cerevisiae. FEBS letters 481: 293-298

Chapitre 5

Characterization of ScORK28, a transmembrane functional receptor kinase predominantly expressed in fertilized ovaries

Hugo Germain, Josée Houde, Madoka Gray-Mitsumune, Tatsuya Sawasaki, Yaeta Endo, Daniel, P. Matton

Contribution des co-auteurs

Madoka Gray-Mitsumune a réalisé essais de déglycosylation, a effectué avec moi le criblage de lignées t-DNA. Les Dr. Endo et Sawasaki ont supervisé mon stage au Japon, m'ont fourni de judicieux conseils scientifiques, m'ont accueillis dans leur laboratoire pour trois mois et m'ont donné accès à leur matériel sans restriction aucune ce qui m'a permis d'exprimer ma protéine; ScORK28. Josée Houde a fait les transformations en cellules de tabac. J'ai effectué les analyses de Northern, les criblages de t-DNA, les transgéniques chez *S.chacoense* et les criblages nécessaires, les essais kinases et la rédaction du papier sauf le paragraphe sur l'*in situ.*

Characterization of ScORK28, a transmembrane functional receptor kinase predominantly expressed in fertilized ovaries

Hugo Germain[1], Josée Houde[1], Madoka Gray-Mitsumune[1], Tatsuya Sawasaki[2], Yaeta Endo[2], Daniel, P. Matton[1]

(1) Institut de Recherche en Biologie Végétale (IRBV), Département de Sciences Biologiques, Université de Montréal, 4101 Sherbrooke Est, Montreal , QC, H1X 2B2 , Canada

(2) Department of Applied Chemistry, Faculty of Engineering, and Venture Business Laboratory, Ehime University, Matsuyama 790-8577, Japan

Corresponding author :

Daniel P. Matton

[redacted]

Phone: +1-514-8723967

Fax: +1-514-8729406

Source de financement

This work was supported by NSERC, FQRNT and by the Canada Research Chair program.

Keywords :

Receptor-like kinase, ovule, fertilization, *Solanum chacoense*

Abstract

Solanum chacoense Ovule Receptor Kinase 28 (*ScORK28*) was found among thirty receptor kinase originating from an ovaries cDNA library enriched for weakly expressed mRNA. This leucine-rich repeat receptor kinase (LRR-V subfamily) displayed high level of tissue specificity at the RNA and protein levels and was predominantly expressed in female reproductive tissues. Protein expression analyses *in planta* revealed that ScORK28 was N-glycosylated and ScORK28:GFP fusion analyses showed that it was localized at the plasma membrane. Bacterial expression of ScORK28 kinase catalytic domain revealed that ScORK28 is an active Mg^{2+}-dependent kinase that can either auto or transphosphorylate. It also showed that the intracellular juxtamembrane domain is necessary for kinase activity and is most probably a major phosphorylation site. Transgenic overexpression lines with full-length or dominant negative constructs showed that ScORK28 protein accumulation was tightly controlled as no plants with significantly higher ScORK28 protein levels could be obtained even in plants strongly overexpressing *ScORK28* mRNAs, suggesting a role for this receptor kinase in plant reproductive development.

Introduction

One key aspect of all developing organisms is their ability to sense their extracellular environment. This function is mediated, in part, by receptor kinases. These proteins are composed of three different subdomains; an extracellular domain that perceives the signal or ligand, a hydrophobic transmembrane domain that anchors the protein to the phospholipids bi-layer, and a cytoplasmic kinase domain that transduces the signal to intracellular target(s). Although most of the knowledge regarding the *modus operandi* of these signal transducer was gained in animal cell, numerous evidences indicate that it will be similar in plants (Shah et al., 2002). Upon ligand binding the receptor kinase homo- or heterodimerize thus bringing the kinase domains in close proximity to allow transphosphorylation (activation), enabling the recruitment and/or phosphorylation of downstream effectors (Nam and Li, 2004). The receptor-like kinase (RLK) gene family represents one of the largest gene family in *Arabidopsis* with 417 members (Shiu and Bleecker, 2001b). The family can be further subdivided in 21 subclasses based on the phylogeny of their kinase domain. Of these 21 subclasses, the largest by far is the leucine-rich repeat family with 216 members (Shiu and Bleecker, 2001b). Mutant analysis have revealed that RLKs are involved in diverse biological processes including development (Clark et al., 1993b; Torii et al., 1996; Schmidt et al., 1997), defense (Song et al., 1995; Scheer et Ryan, 1999; Gomez-Gomez and Boller, 2000), hormone sensing (Kinoshita et al., 2005) and reproduction (Nasrallah et al., 1994; Muschietti et al., 1998). In contrast to this, biochemical characterization of the intracellular domain of plant receptor kinase has only been performed for a few receptor kinase (Williams et al., 1997; Oh et al., 2000; Shah et al., 2001). For example, the phosphorylation activity of receptor kinases is regulated by divalent cations such as Mn^{2+} or Mg^{2+} but, in general, plant receptor kinases have been shown to have a preference for Mn^{2+} although some receptor kinase are active in the

presence of only Mg^{2+} (Schaller and Bleecker, 1993; Reuveni and DuPont, 2001). The phosphorylated residues are generally located outside of the kinase domain, either in the juxtamembrane domain or the C-terminal domain. A detailed analysis of the *in vivo* phosphoprylation sites of several plant receptor kinases revealed that 75% of the phosphorylation sites are located either in the juxtamembrane domain or in the C-terminal domain (Nuhse et al., 2004). As for animal receptor kinases, plant RLKs can be separated in RD and non-RD kinases. Non-RD kinase are defined by the absence of an arginine residue in kinase subdomain VI and exhibit lower kinase activity than RD kinases (Johnson et al., 1996a).

Using a negative selection screen we have previously identified 30 receptor kinases expressed in ovules and ovaries from *Solanum chacoense* (Germain et al., 2005). One of these, Ovule Receptor Kinase 28 (*ScORK28*) displayed an expression pattern highly specific to ovaries and was therefore chosen for further biochemical, cellular and molecular characterization to determine its involvement in fertilization and early fruit development.

Materials and method

Sequence analysis

Raw DNA sequence was *in silico* translated using the MacVector 8.1.2 (Accelrys, San Diego, CA, USA). The presence of a signal peptide was assessed using SignalP 3.0 (Bendtsen et al., 2004) and the location of the transmembrane domain was positioned using TMHMM 2.0 (Krogh et al., 2001). Putative N-glycosylation sites were analysed using NetGly 1.0 (Gupta and Brunak, 2002) and the structure of the leucine-rich repeat domain was analyzed using ScanSite 2.0 (Obenauer et al., 2003). The phylogenic tree was inferred using a neighbor-joining algorithm (MacVector 8.1.2) with 1000 bootstraps either on the full length sequences or on the trimmed catalytic domains.

Transgenic plants

Arabidopsis thaliana Col. plants were grown under long day condition (16 h light/8 h dark) at 25°C in a growth chamber in 1:1:1 mixture of perlite, vermiculite and topsoil. *Solanum chacoense* plants were grown in topsoil in the greenhouse. *Arabidopsis* plants consisted of T-DNA line obtained from the SALK institute (At1g53730 SALK line number: 116320, 116318, 062310, 054337, 077702 and At3g14350 SALK line number: 039120, 068033, 110007. The *S. chacoense* transgenic was an overexpression lines (OX) with the full length *ScORK28* cDNA cloned downstream of the CaMV 35S promoter. Prior to transformation the construct was fully sequenced.

Protein extraction, western blot analysis and deglycosylation assay

Plant tissues were hand dissected, ground in liquid nitrogen and stored at -80°C until protein extraction. Protein were extracted in a buffer containing 0.1M Tris-HCl pH 8, 0.1% SDS, 2% β-mercaptoethanol, and 1X of protease cocktail inhibitor (Roche Diagnostic, Laval, Qc, Canada). Protein concentration was evaluated in triplicate using the (Bradford) BIO-RAD Protein Assay (Bio-Rad, Hercules, CA, USA) and 10 μg of protein were mixed with Laemmli buffer, denatured at 95°C for 5 minutes and loaded on a 8% acrylamide gel (Sigma, Oakville, ON, Canada). The gel was run at 150 V for 1 hour in running buffer (3,25 mM SDS, 25 mM Tris and 192 mM glycine in one liter) and then transferred to PVDF membrane in transfer buffer (25 mM Tris and 192 mM glycine, and 20% methanol in one liter). To generate specific polyclonal antibodies the extracellular domain of ScORK28 was expressed *in vitro* as previously described (Sawasaki et al., 2002). The protein was then separated on 2-D gel, electroeluted and injected into NZ white rabbits.

Immunoblotting was done as follows: PVDF membranes were hydrated in methanol for 10 seconds, equilibrated in TBS (20 mM Tris pH7.5, 150 mM NaCl) for 15 minutes, blocked for 1 hour in TBST+milk (TBS + 0.02% Tween 20 (v/v) and 2.5% defatted dry-milk (w/v)), incubated for 1 hour with the polyclonal anti-ScORK28 in TBST+milk, washed three times 15 minutes in TBST, incubated with the monoclonal anti-rabbit IgG γ-chain specific conjugated with the peroxidase enzyme (Sigma) for one hour and washed three times 15 minutes in TBST. Detection was done using ECL Plus (GE Healthcare) and exposed on a Bioflex Econo Film (Interscience, Markham, ON, Canada).

For N-deglycosylation assay, the ScORK28 complete protein sequence was cloned into the tandem affinity purification (TAP) vector pC-TAPa (Rubio et al., 2005) and expressed under the control of a double enhancer CaMV 35S promoter *in planta* through agroinfiltration of *Nicotiana benthamiana* leaves co-infected with the p19 silencing

suppressor gene as previously described (Voinnet et al., 2003). Crude extracts from infected leaves collected 3 days after infection were treated with N-glycosidase F (PNGase F) following the manufacturer instructions (New England Bio-lab, Pickering, ON, Canada). The extract was separated on acrylamide gel, transferred on PVDF membrane and detected as previously described using the anti-HIS antibody (Sigma).

Transient expression for protein localization

ScORK28 was PCR amplified from the cDNA clone (forward primer GGGGACAAGTTTGTACAAAAAAGCAGGCTCGAGGAGCCTTTCTTCTGTTCCC and reverse primer GGGGACCACTTTGTACAAGAAAGCTGGGTCTGGTTCATAGTCCGGGCCATC) and cloned using the Gateway recombination vector pZeo (Invitrogen, Burlington, ON, Canada). From the donor vector, the *ScORK28* open reading frame was transferred by recombination in the pMDC83 vector upstream and in frame with the GFP reporter protein (Curtis et Grossniklaus, 2003). The fusion protein was verified by sequencing. Onion cells were transiently transformed by particle bombardement using this *ScORK28-GFP* construct. The epidermis of onion bulb scales was prepared by peeling the inner epidermis from fresh onion bulbs. Five µg of DNA was mixed with 10 µl of 0.1 M spermidine, 25 µl of 2.5 M $CaCl_2$ and 25 µl of tungstene microcarrier (60 µg/ml in 50% glycerol), vortexed vigorously for 5 minutes and centrifuged at 800 *g* in a microcentrifuge for 10 sec. The pellet was washed once with 70% ethanol and the pellet was resuspended in 5µl 100% ethanol. The DNA-tungstene particles were bombarded into cells at a pressure of 1200 psi and target distance of 10 cm, with a homemade biolistic /He particle delivery system under a vacuum chamber at 24 mm Hg. The transformed cells were kept alive by placing the onion peels onto 0.5X Murashige-Skoog medium overnight.

Protein kinase activity assay

The kinase domain of *ScORK28* was cloned in fusion with a N-terminal His-tag in the pQE30 vector (Qiagen, Mississauga, ON, Canada) and expressed in the Rosetta-gami bacterial strain (Novagen, Mississauga, ON, Canada) at 16°C. Two versions of the kinase domain of ScORK28-6XHis were expressed. The first one with the full length intracellular domain (encompassing the internal cytoplasmic juxtamembrane domain, the kinase domain and the C-terminal domain) and named $ScORK28_{308\text{-}711}$ (forward primer GAGAGAGGATCCAGATTGAAGCGGTCATCGACGG; reverse primer GAGAGTCGACAAACTATGGTTCATAGTCCGGGCC). The second construct was made without the internal juxtamembrane domain and named $ScORK28_{409\text{-}711}$ (forward primer GAGAGAGGATCCAGCTTCAATGTTGATAACCTTATTGGC and same reverse primer as $ScORK28_{308\text{-}711}$). Rosetta-gami cells were lyzed using a French Press set at 900 psi and the recombinant protein was purified on nickel Sepharose (GE Healthcare) following the manufacturer instructions with minor modifications. Binding of the cleared lysate was done in 30 mM imidazole, with washes in 60 mM imidazole and elution in 500 mM imidazole. The eluate was then dialyzed twice against 1250 volumes of kinase buffer (50 mM Tris-HCl pH 7.6). For the kinase activity assay $MgCl_2$ and/or $MnCl_2$ was added to a final concentration of 10 mM and DTT was added to a concentration of 0.1 mM and 0.1 µL ^{32}P γ-ATP (370 MBq/ml) was added to a final reaction volume of 20 µL. The reaction was incubated at 25°C for 30 minutes and the product was separated on a 15% acrylamide gel, transferred to a PVDF membrane, exposed on a europium screen and scanned on a Typhoon 9200 Phosphorimager (GE healthcare).

Results

ScORK28 sequence analysis and phylogeny

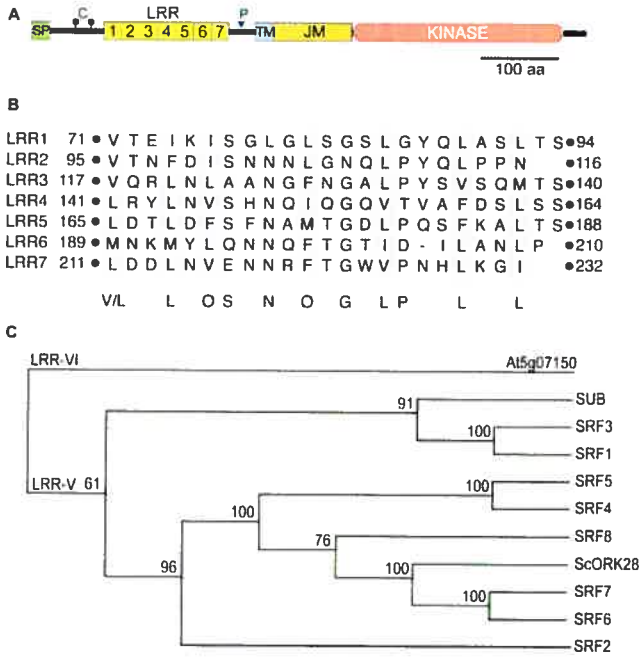
We have previously isolated 30 receptor-like kinases from a negative screen that focused on weakly expressed transcripts in ovule and ovary tissues following fertilization (Germain et al., 2005). Fine expression analysis through quantitative RT-PCR showed that amongst the 30 RLKs, 28 were predominantly expressed in female reproductive tissues and 23 were transcriptionally up-regulated following fertilization. Twenty-five RLKs belonged to the LRR-RLK family, the largest RLK family in *Arabidopsis* (Shiu et Bleecker, 2003). A few RLKs were chosen for further analysis, based on their high ovary-specific expression and low expression in other tissues. One of these RLK is *ScORK28* and is described hereafter.

The EST corresponding to *ScORK28* (DN98256) was fully sequenced (Genbank number EF447217) and most probably contained the full-length or near full-length ScORK28 cDNA (2329 bp) since its 5'-UTR sequence (124 bp long) contained three in frame stop codons upstream of the first methionine, The longest open reading frame deduced from the cDNA sequence coded for a 76.7 kD protein of 711 amino acids with a predicted pI of 5.87. A signal peptide sequence targeting the protein to the secretory pathway is also present in the first 23 amino acids of the deduced amino acid sequence, as well as a single transmembrane domain predicted from amino acid position 285 to 307 (Fig. 11A) (Bendtsen et al., 2004) immediately followed by positively charged residues (KKR) acting as a stop transfer sequence. The signal peptide is followed by seven imperfect leucine-rich repeat (Fig. 11B) from amino acid position 71 to 232, as predicted from Scansite. The seven LRR are flanked by a cysteine pair on their N-terminal side that has a near perfect match (V to I mutation) with the plant specific consensus sequence for N-terminal cysteine pair (CxnWxGVt/sC) usually observed in RLK-LRR family II, III, VI, IX, X, XI, XII, XIII (Dievart and Clark, 2003). It should be noted that the occasionally

observed cysteine pair in C-terminal of the LRR domains is absent from ScORK28 as for all the members of the *Arabidopsis* LRR-V family. A small island of prolines (8 out of 13 residues) is also found after the LRRs between amino acids 246 and 258. The transmembrane domain is followed by a short juxtamembrane domain, a kinase domain and a short C-terminal domain (Fig. 11A). A total of eight predicted phosphorylation sites were predicted by ScanSite in the intracellular domain (Obenauer et al., 2003). Four are located in the juxtamembrane domain (amino acid 308 to 408) and the other four are located in the kinase domain (409-679). A putative N-glycosylation site is also present on the asparagine located at position 145 of the immature protein as determined from predictive algorithms (Gupta and Brunak, 2002).

Figure 11. A. Illustration ScORK28 primary structure. B. The seven LRR are aligned to emphasize most frequent amino acids found at regularly spaced positions. C. A neighbour-joining clustering from the complete amino acid sequences of the ScORK28 receptor kinase and the most closely related RLKs from *A. thaliana* (LRR-V subfamily).

A.The dots represents the cystein pair. SP, signal peptide; LRR, leucine-rich repeat; TM, transmembrane domain; JM, intracellular juxtamembrane domain; KINASE, kinase catalytic domain. B. Grey boxes represent conserved hydrophobic residues. A consensus sequence is presented underneath the alignment and represents amino acids conserved in >50% of the LRR. ∅, aliphatic amino acid; C, conserved cystein pair; P, proline-rich island. C. A neighbour-joining clustering from the complete amino acid sequences of the ScORK28 receptor kinase and the most closely related RLKs from *A. thaliana* (LRR-V subfamily). A LRR-VI RLK was chosen as the out group. Bootstrap values were calculated from 1000 replicas.



Similarity searches using the Blast algorithm retrieved mainly receptor-like kinases from various plant species. A global phylogenetic analysis against all *Arabidopsis* RLKs (Shiu et Bleecker, 2003), and using only the trimmed kinase domain, confirmed that ScORK28 belonged to the small RLK subfamily LRR-V (data not shown). This family has nine members in *Arabidopsis* that are dispersed on all chromosomes. The best known member of the family is *STRUBBELIG*, a RLK that affects the formation and shape of several organs, the orientation of the division plane, and cell proliferation (Chevalier et al., 2005). Figure 11C shows a phylogenetic analysis using the complete amino acid sequence of the nine RLK-LRRV subfamily members (also known as the STRUBBELIG receptor family or SRF family) from *Arabidopsis* (SUB and SRF1- to 8) and a RLK of the LRR-VI subfamily (At5g07150) as the outgroup. This analysis shows that ScORK28 is more closely related to the SRF6 (At1g53730) and SRF7 (At3g14350) members of the SRF family. Close association between ScORK28 and SRF6 and SRF7 is obtained whether using the full protein sequences or the kinase domain only (data not shown).

ScORK28 is predominantly expressed in young fruits and is glycosylated

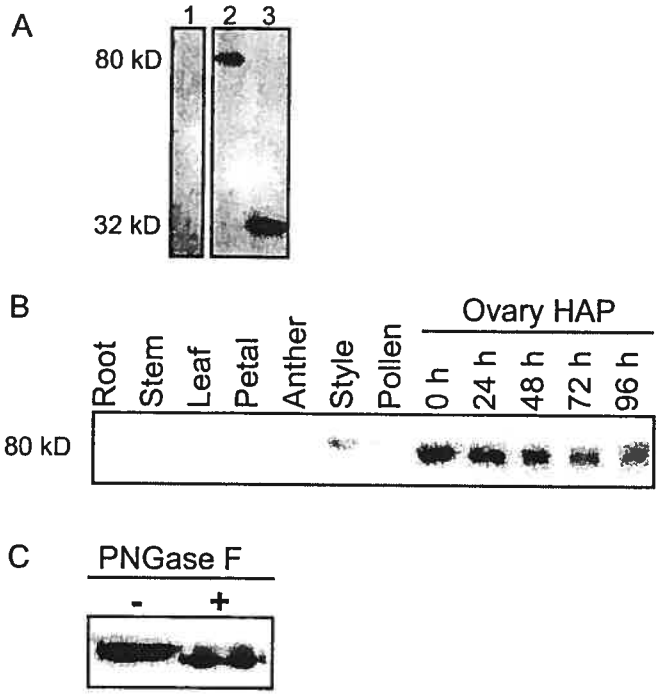
To determine if ScORK28 protein expression pattern followed *ScORK28* mRNA expression (Fig 6), a specific antibody directed against the extracellular domain (ectodomain) of ScORK28 was produced. The ScORK28 ectodomain was synthesized in a cell-free protein system from wheat embryos as described previously (Sawasaki et al., 2002). Targeting the ectodomain for antibody production minimizes the possibility of cross-reaction with other protein kinases since ectodomains are highly divergent among RLKs in sharp contrast to the much more conserved kinase catalytic domain. Specificity of

the antibody is shown in Fig 12A. In lane 1, pre-immune sera was used to immunodetect the ScORK28 ectodomain encoding protein used to inject the rabbit. In lane two, sera from final bleed was used to detect ScORK28 from whole flower or recombinant ScORK28 lane 3. Only one band was observed against the flower tissue showing the specificity of the antibody. Immunoblotting using an anti-ScORK28 antibody (Fig. 12B) clearly shows a predominant ScORK28 accumulation in ovaries. Contrary to its fertilization-induced transcription (Fig. 6), the ScORK28 protein accumulated to higher levels in unfertilized ovaries and decreased slightly afterward in ovaries following fertilization. Furthermore, although *ScORK28* mRNAs were detected in root, stem, leaf, petal, anther, and pollen (Fig. 6), no translation product of *ScFRK28* could be detected in these tissues. These results suggests that there is a strict translational control of the *ScORK28* mRNAs.

The ScORK28 molecular weight on gel was estimated at around 83 kD, compared to a predicted 76.7 kD MW as deduced from the primary amino acid sequence. Since ScORK28 possesses one predicted N-glycosylation site in the ectodomain (position 145), a N-deglycosylation assay was performed to determine if ScORK28 is glycosylated. The ScORK28 open reading frame (ORF) fused to a hexahistidine (6His) tag was transiently expressed through agroinfiltration in *Nicotiana benthamiana* leaves and a crude protein extract was submitted to deglycosylation with N-glycosidase F (PNGase F). PNGase F cleaves between the innermost GlcNAc and asparagine residues of high mannose, hybrid, and complex oligosaccharides from N-linked glycoproteins. A small mobility shift of ~5 kDa was detected on the membrane with an anti-His antibody (Fig. 12C), suggesting that the protein is N-glycosylated. Such a small difference in molecular weight is consistent with the presence of only one glycosylation site, as predicted from the primary amino acid sequence.

Figure 12. Expression profile of ScORK28 and deglycosylation assay.

A. Specificity test of the anti-SCORK28 antibody, lane 1 pre-immune sera, lane 2 whole flower protein extract, lane 3 recombinant ScORK28 ectodomain. B. Western blot analysis C. Deglycosylation assay from transiently expressed ScORK28 protein in *Nicotiana benthamiana* leaves. A 5 kD gel shift is observed after treatment (+) with PNGase F (N-glycosidase F).

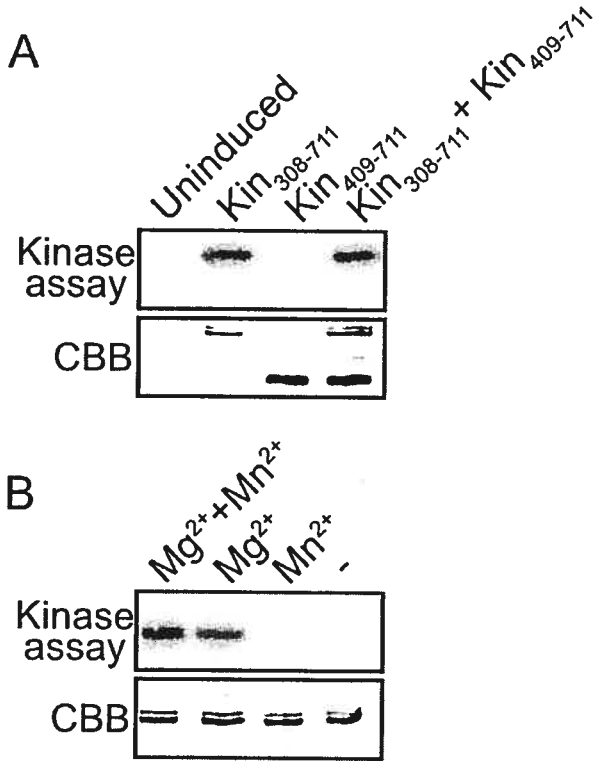


ScORK28 is a functional protein kinase

In order to determine if ScORK28 was a functional protein kinase, hexahistidine-tagged truncated constructs were expressed in *E. coli*. Only the cytoplasmic part of the RLK was used in order to maximize the likelihood of proper expression in bacteria. Two constructs were tested. The first one, $ScORK28_{308-711}$, contained the full cytoplasmic part of the protein, including the juxtamembrane, kinase, and C-terminal domains. The second construct, $ScORK28_{409-711}$, had the juxtamembrane domain further deleted. Figure 13 A clearly shows that $ScORK28_{308-711}$ can autophosphorylate (or transphosphorylate) while $ScORK28_{409-711}$ cannot (equal loading is shown on figure 13A, bottom panel). When both bacterially expressed proteins are mixed together, phosphorylation is only detected on the higher molecular weight $ScORK28_{308-711}$ protein construct (Fig. 13A, $KIN_{308-711}$ + $KIN_{409-711}$). This suggests that $ScORK28_{308-711}$ cannot phosphorylate the shorter $ScORK28_{409-711}$, indicating that the juxtamembrane domain is either necessary for the association of the two proteins or it encompasses the targeted phosphorylation site(s). Generic kinase substrates like histone and myelin basic protein were also tested with both expression constructs and no phosphorylation could be detected on these substrates (data not shown). Bivalent metal ion requirements for ScORK28 kinase activity were also assessed using the bacterially expressed $ScORK28_{308-711}$ construct (Fig 13B). The enzymatic assay was performed with or without either Mg^{2+} or Mn^{2+}, or with both ions. Bivalent metal ion are clearly necessary for ScORK28 phosphorylation activity since no activity can be detected without Mg^{2+} while Mn^{2+} alone cannot stimulate its phosphorylation activity. Presence of both Mg^{2+} and Mn^{2+} seems to slightly increase the kinase activity over Mg^{2+} alone.

Figure 13. Kinase activity assay of ScORK28.

A. ScORK28 kinase activity from bacterially expressed protein. Two histidine-tagged kinase constructs were used for the phosphosphylation assay. The $Kin_{308-711}$ construct corresponds to the complete cytoplasmic part of the ScORK28 protein, comprising the juxtamembrane and the kinase catalytic domains as well as the C-terminal tail (Fig. 11A). The $Kin_{409-711}$ construct has the JM domain (308-408) deleted. B. Divalent cation requirement for ScORK28 activity ($Kin_{308-711}$ construct). Lower panels, Coomassie Brilliant Blue (CBB) staining of equivalent gel indicating equal loading of the expressed proteins.

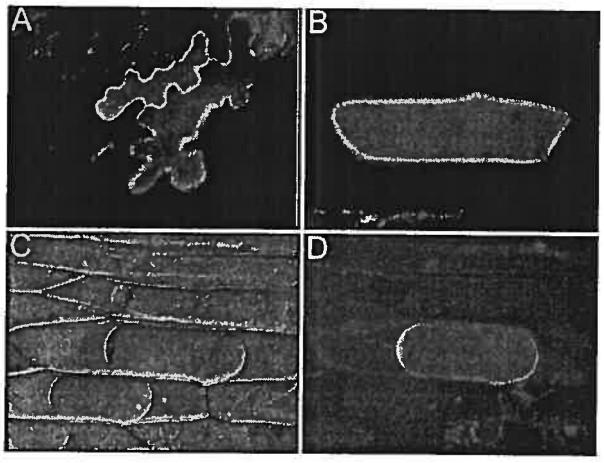


ScORK28 is targeted to the plasma membrane

Although LRR-RLKs possess a transmembrane domain and a putative signal peptide, the subcellular localization of very few LRR-RLKs has been investigated. To determine the subcellular localization of ScORK28, a fusion protein harboring a green fluorescent protein (GFP) downstream (C-terminal) of the ScORK28 ORF was constructed. Transient expression through *Agrobacterium* infiltration of *N. benthamiana* leaves showed that the tagged protein is located at the cell periphery but does not allow a clear discrimination between a cell wall or plasma membrane localization (Fig. 14A). To better localize the ScORK28 RLK, transient expression through particle bombardment of onion epidermal cells with *ScORK28-GFP* DNA coated tungsten microbeads was performed. Again, ScORK28 could be unambiguously located to the cell periphery in non-plasmolyzed cells (Fig. 14B). Plasmolysis of this epidermal cell layer revealed that ScORK28 was associated with the cell plasma membrane and not the cell wall (Fig. 14D). A DIC image of the same cell layer shows the cell structures after plasmolysis, clearly showing the cell wall and the plasma membrane of the shrunken protoplasm (Fig. 14C). We therefore conclude that the signal peptide targets ScORK28 to the plasma membrane where it remains.

Figure 14. Plasma membrane localization of ScORK28.

A. The ScORK28 construct fused to the GFP (ScORK28-GFP) was transiently expressed in *N. benthamiana* epidermal cells through agroinfiltration along with the p19 silencing suppression construct to maximize expression and observed under fluorescent illumination Two cells strongly expressing ScORK28-GFP are visible. B. Epidermal onion cells expressing the ScORK28-GFP construct after micro particles bombardment and observed under fluorescent illumination. C. Differential interference contrast (DIC) microscopy of the same cells shown in B after plasmolysis showing the shrinkage of the protoplast (white arrow). D. Same cell as in B but after plasmolysis showing the retention of the fluorescent signal around the plasma membrane.



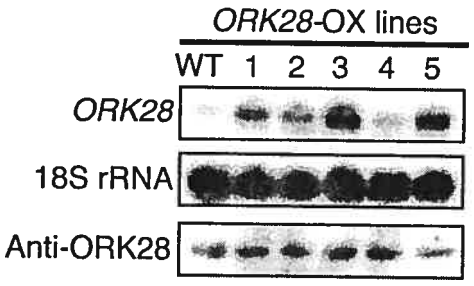
Transgenic plants analysis

In order to uncover the function of ScORK28 *in planta,* transgenics plants overexpressing the full length *ScORK28* cDNA was made. *S. chacoense* plants strongly overexpressing the mRNA corresponding to the full length transgene were obtained (Fig. 15A top panel) but a Western blot analysis of these plants revealed that the ScORK28 protein level was not affected and remained steady (Fig. 15A bottom panel). Expression level of the ScORK28 protein in these plants was not significantly higher than in the wild-type plants despite the fact that the expression of the transgene is under the control of the strong CaMV 35S promoter. This result indicates that the protein level is tightly regulated post-transcriptionally, as observed previously with the tissue-specific distribution of the protein (Fig. 15A).

Arabidopsis t-DNA mutants corresponding to the two closest putative orthologs (At1g53730 and At3g14350, see material and methods for mutant line accessions) were analyzed for altered phenotypes but only wild-type looking plants were observed in single mutants indicating a possible redundancy effect in this family of eight genes (data not shown).

Figure 15. Analysis of transgenic lines expressing full length or dominant negative *ScORK28* constructs.

A. Selection of transgenic lines overexpressing *ScORK28* under the control of the CaMV 35S promoter. Upper panel. RNA gel blot analysis of *ScORK28*-OX lines probed with the *ScORK28* full length ^{32}P labeled insert. Middle panel. RNA gel blot analysis with *ScORK28*-OX lines with a 18S ribosomal probe. Lower panel. Western blot analysis of the same *ScORK28*-OX lines with an anti-ScORK28 antibody directed against the ScORK28 ectodomain.



Discussion

The importance of receptor kinases in plants has emerged in the early 1990's with the discovery of the first plant receptor-like kinase (Walker and Zhang, 1990). Since then, hundreds of plant receptor kinases have been found, mainly through the analysis of completely sequenced genomes (Shiu et al., 2004). It is now well recognized that receptor-like kinases have functions in all aspects of plant growth and development (Morillo et Tax, 2006), including hormone perception (He et al., 2000), meristem maintenance (Clark et al., 1993b), defense responses (Scheer and Ryan, 2002), and symbiotic interaction (Stracke et al., 2002). Receptor kinases are generally classified in several subgroups (up to almost 80 subgroups) based on their phylogenic relationships and their specific ectodomains (Shiu et Bleecker, 2001b; Shiu et al., 2004). Protein kinases can also be further divided into RD or non-RD kinase (Johnson et al., 1996a). RD kinases are regulated by activation loop phosphorylation and typically carry a conserved arginine (R) residue immediately preceding the invariant aspartate in the kinase subdomain VI required for catalytic activity. Most RD kinases are much more active than non-RD kinases, although there are exceptions (Johnson et al., 1996a). Recently, Dardick and Ronald (2006) made a thorough classification of all protein kinases in rice and *Arabidopsis* as well as in humans, *Drosophila*, *C. elegans* and yeast. RD kinases comprises 70% of *Arabidopsis* kinome and include well known RLKs like ERECTA, BRI1, CLAVATA1 and HAESA (Dardick and Ronald, 2006). Apart from having the conserved R, they also have an aspartate (D) residue immediately following the R in subdomain VI (HRD motif), a conserved lysine (K) in subdomain II, as well as a conserved D in subdomain VII (DFG motif). Non RD-kinases lack this conserved R in subdomain VI but have the other conserved amino acids (K/D/D). They comprises 10% of the *Arabidopsis* kinases, show very little kinase activity and have been associated in plants and animals with innate immunity and pathogen perception, like Xa21 and FLS2 in plants, and IRAK in human. A third class, termed ACF for alternative catalytic function comprises 20% of the *Arabidopsis* kinome and these protein kinases lack

one or more of three highly conserved residues (K/D/D) thought to be required for catalytic activity in subdomains II (K), VI (D), and VII (D) (Dardick and Ronald, 2006).

We identified ovule receptor kinase 28 (*ScORK28*) from an EST library enriched in cDNAs corresponding to weakly expressed mRNA species (Germain et al., 2005). *ScORK28* expression at the protein and RNA level was mainly restricted to the developing female reproductive tissues and the gene was found to be transiently up-regulated five-fold following fertilization (Germain et al., 2005) and therefore a role in early embryo and seed development was suspected. Amino acid sequence analysis and phylogenetic analysis revealed that the ScORK28 kinase belonged to the LRR-V subfamily. In *Arabidopsis*, this family comprises 9 members, including *STRUBBELIG* (*SUB*), a RLK that affects the formation and shape of several organs, the orientation of the division plane, and cell proliferation (Chevalier et al., 2005). SUB is also identical to SCRAMBLED, a RLK involved in position-dependant specification of root epidermal cells in *Arabidopsis* (Kwak et al., 2005b). The other eight members have been named SRF for STRUBBELIG receptor family, and all are classified as ACF kinases. All members of this family have from six to seven predicted LRRs that can be easily aligned. The SUB RLK has been described as having six LRR but an imperfect LRR preceding these six can also be found in what was called the SUB domain.

Although ScORK28 is an ACF kinase, lacking many key amino acid residues thought to be involved in catalytic activity, *in vitro* autophosphorylation activity could be detected in the presence of magnesium ions but not in the presence of manganese ions (Fig. 13). Most kinase assays are generally performed with both Mg^{2+} and Mn^{2+}, and plant receptor kinases tend to have a preference for Mn^{2+} (Schaller et Bleecker, 1993), to specifically require Mn^{2+} (Charpenteau et al., 2004), or to display activity with either manganese or magnesium (Nirmala et al., 2006). The only other RLK in the LRR-V family tested for kinase activity, the SUB RLK, was shown not to have any kinase activity, at least from bacterially expressed constructs with or without the JM domain, and in the presence

of myelin basic protein (MBP). SUB is also an ACF kinase, as for all LRR-V kinases, but also has other mutated residues in highly conserved regions (Chevalier et al., 2005). The phosphorylation sites of plant receptor kinase was thoroughly studied by Peck's group using mass spectrometry (Nuhse et al., 2004). They showed that most phosphorylation events take place either within the juxtamembrane domain or in the C-terminal domain and only a minor fraction takes place in the kinase domains itself. These results are supported by the fact that BRI1 has four confirmed phosphorylation sites (out of five) in its juxtamembrane domain (Wang et al., 2005b). For ScORK28, four predicted phosphorylation sites (out of eight) are located in the juxtamembrane domain. A bacterially expressed construct encompassing only the kinase domain (Kin $_{409\text{-}711}$) was not capable of auto-phosphorylation (Fig. 13) nor had kinase activity toward generic substrates like histone or MBP (data not shown). Furthermore, in kinases assays with both the short (Kin $_{409\text{-}711}$) and long version with the JM domain (Kin $_{308\text{-}711}$), phosphorylation could only be detected on the JM containing expressed protein, suggesting that phosphorylated residues are located in this JM domain, in agreement with the results of Nuhse et al. (2004). Kinase activity for other RLK including CLV1, BRI1, Xa21, and CR4 has also been obtained in constructs that included the JM domain (Williams et al., 1997; Friedrichsen et al., 2000; He et al., 2000; Jin et al., 2000). MARK, a maize RLK expressed during embryogenesis, has all the features of ACF kinases, like ScORK28, and was considered dead (or inactive) kinase based on kinases assays with expressed protein that did not included the complete JM domain (Llompart et al., 2003). It would be of interest to determine if addition of the JM domain in the MARK kinase assay would lead to an active kinase. These results taken together strongly suggest that ScORK28 phosphorylation occurs in the JM domain although we cannot exclude the possibility that the juxtamembrane domain is necessary for protein dimerization and transphosphorylation. In the latter case, absence of dimerization with the ScORK28$_{409\text{-}711}$ construct would preclude phosphorylation.

It is generally assumed that receptor kinase proteins localize to the plasma membrane. This assumption may prove wrong as for the ethylene histidine-kinase receptor

that localizes to the ER (Chen et al., 2002). Evidences also exist showing that some receptor kinase may be retained in Golgi vesicles (Goto et al., 2005; Niwa et al., 2006). Our results both in bombarded onion cells or through agroinfiltration of *N. benthamiana* leaves (Fig. 14) clearly demonstrated that ScORK28 localizes to the plasma membrane, as would be expected for a recptor-kinases involved in intercellular communication.

Although we have obtained valuable and novel information regarding the *modus operandi* of ScORK28 as a receptor kinase, including the requirement of the juxtamembrane domain and of Mg^{2+} for its phosphorylation activity, its plasma membrane localization, and its glycosylation status, functional analyses in transgenic plants could not yet reveal a precise role for this RLK. Analysis of eight T-DNA lines in *Arabidopsis* (see Material and Methods for the line accessions) corresponding to putative *S. chacoense* orthologs (Fig. 10) and *S. chacoense* transgenic plants overexpressing the full length *ScORK28* clone did not yield plants showing obvious developmental defects. Therefore, we cannot speculate about the function of *ScORK28* in fertilization and early embryogenesis. Gene redundancy in *A. thaliana* insertional mutant lines might preclude the observation of a mutant phenotype and double mutants will be necessary to uncover the role of *ScORK28* orthologs. Similarly, in *S. chacoense*, a related LRR-V RLK (ScORK25) also shares a highly similar expression pattern, also suggesting gene redundancy (Germain et al., 2005). To circumvent this, OX-lines expressing either a full length or dominant negative constructs were produced. Surprisingly, although overexpression could be found at the mRNA level, tight control at the post-transcriptional level precluded accumulation of ScORK28 protein at levels significantly higher than WT, emphasizing the importance proper expression of this receptor kinase in plant growth and development. Such a mechanism is not uncommon (Hirschi et al., 2000) and has been postulated to be caused by the host silencing mechanism (Diener and Hirschi, 2000).

Acknowledgements

H. Germain was the recipient of Ph. D. fellowships from the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Fonds Québécois de la Recherche en Nature et Technologie (FQRNT). This work was also supported by the Canadian Foundation for Innovation and by the Canada Research Chair program (D. P. Matton). We also thank Gabriel Téodorescu for plant care and maintenance, Alan O'Brien for technical assistance, and Dr. Jean Rivoal and Sonia Dorion for their assistance with protein purification and insightful discussion.

References

1. **Bendtsen, J. D., Nielsen, H., von Heijne, G. and Brunak, S.** (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783-95.

2. **Charpenteau, M., Jaworski, K., Ramirez, B. C., Tretyn, A., Ranjeva, R. and Ranty, B.** (2004). A receptor-like kinase from Arabidopsis thaliana is a calmodulin-binding protein. *Biochem J* **379**, 841-8.

3. **Chen, Y. F., Randlett, M. D., Findell, J. L. and Schaller, G. E.** (2002). Localization of the ethylene receptor ETR1 to the endoplasmic reticulum of Arabidopsis. *J Biol Chem* **277**, 19861-6.

4. **Chevalier, D., Batoux, M., Fulton, L., Pfister, K., Yadav, R. K., Schellenberg, M. and Schneitz, K.** (2005). STRUBBELIG defines a receptor kinase-mediated signaling pathway regulating organ development in Arabidopsis. *Proc Natl Acad Sci U S A* **102**, 9074-9.

5. **Clark, S. E., Running, M. P. and Meyerowitz, E. M.** (1993). *Clavata1*, a regulator of meristem and flower development in *Arabidopsis*. *Development* **119**, 397-418.

6. **Curtis, M. D. and Grossniklaus, U.** (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* **133**, 462-9.

7. **Dardick, C. and Ronald, P.** (2006). Plant and animal pathogen recognition receptors signal through non-RD kinases. *PLoS Pathog* **2**, e2.

8. **Diener, A. and Hirschi, H.** (2000). Heterologous expression for dominant-like gene activity. *Trends Plant Sci* **5**, 10-1.

9. **Dievart, A. and Clark, S. E.** (2003). Using mutant alleles to determine the structure and function of leucine-rich repeat receptor-like kinases. *Curr Opin Plant Biol* **6**, 507-16.

10. **Friedrichsen, D. M., Joazeiro, C. A., Li, J., Hunter, T. and Chory, J.** (2000). Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. *Plant Physiol* **123**, 1247-56.

11. **Germain, H., Rudd, S., Zotti, C., Caron, S., O'Brien, M., Chantha, S. C., Lagace, M., Major, F. and Matton, D. P.** (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**, 515-32.

12. **Gomez-Gomez, L. and Boller, T.** (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Mol Cell* **5**, 1003-11.

13. **Goto, S., Niwa, Y. and Kobayashi, H.** (2005). A cytoplasmic signal of receptor-like kinase in golgi vesicles to chloroplast biogenesis. In *Protein Phosphorylation in Plant Signaling*. Tsukuba, Japan.

14. **Gupta, R. and Brunak, S.** (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. In *Pacific Symposium on Biocomputing*, pp. 310-322.

15. **He, Z., Wang, Z. Y., Li, J., Zhu, Q., Lamb, C., Ronald, P. and Chory, J.** (2000). Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science* **288**, 2360-3.

16. **Hirschi, K. D., Korenkov, V. D., Wilganowski, N. L. and Wagner, G. J.** (2000). Expression of arabidopsis CAX2 in tobacco. Altered metal accumulation and increased manganese tolerance. *Plant Physiol* **124**, 125-33.

17. **Jin, P., Guo, T. and Becraft, P. W.** (2000). The maize CR4 receptor-like kinase mediates a growth factor-like differentiation response. *Genesis* **27**, 104-16.

18. **Johnson, L. N., Nobel, M. E. M. and Owen, D. J.** (1996). Active and inactive protein kinases: structural basis for regulation. *Cell* **85**, 149-158.

19. **Jones, J. D. G., Dunsmuir, P. and Bedbrook, J.** (1985). High level expression of introduced chimeric genes in regenerated transformed plants. *Embo J* **4**, 2411-2418.

20. **Kinoshita, T., Cano-Delgado, A., Seto, H., Hiranuma, S., Fujioka, S., Yoshida, S. and Chory, J.** (2005). Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**, 167-71.

21. **Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L.** (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-80.

22. **Kwak, S. H., Shen, R. and Schiefelbein, J.** (2005). Positional signaling mediated by a receptor-like kinase in Arabidopsis. *Science* **307**, 1111-1113.

23. **Llompart, B., Castells, E., Rio, A., Roca, R., Ferrando, A., Stiefel, V., Puigdomenech, P. and Casacuberta, J. M.** (2003). The direct activation of MIK, a germinal center kinase (GCK)-like kinase, by MARK, a maize atypical receptor kinase, suggests a new mechanism for signaling through kinase-dead receptors. *J Biol Chem* **278**, 48105-11.

24. **Morillo, S. A. and Tax, F. E.** (2006). Functional analysis of receptor-like kinases in monocots and dicots. *Curr Opin Plant Biol* **9**, 460-9.

25. **Muschietti, J., Eyal, Y. and McCormick, S.** (1998). Pollen tube localization implies a role in pollen-pistil interactions for the tomato receptor-like protein kinases LePRK1 and LePRK2. *Plant Cell* **10**, 319-30.

26. **Nam, K. H. and Li, J.** (2004). The Arabidopsis transthyretin-like protein is a potential substrate of BRASSINOSTEROID-INSENSITIVE 1. *Plant Cell* **16**, 2406-17.

27. **Nasrallah, J. B., S.J., R. and Nasrallah, M. E.** (1994). Genetic evidence for the requirement of the Brassica S-locus receptor kinase gene in the self0incompatibility *Plant Journal* **5**, 373-384.

28. **Nirmala, J., Brueggeman, R., Maier, C., Clay, C., Rostoks, N., Kannangara, C. G., von Wettstein, D., Steffenson, B. J. and Kleinhofs, A.** (2006). Subcellular localization and functions of the barley stem rust resistance receptor-like serine/threonine-specific protein kinase Rpg1. *Proc Natl Acad Sci U S A* 103, 7518-23.

29. **Niwa, Y., Goto, S., Nakano, T., Sakaiya, M., Hirano, T., Tsukaya, H., Komeda, Y. and Kobayashi, H.** (2006). Arabidopsis mutants by activation tagging in which photosynthesis genes are expressed in dedifferentiated calli. *Plant Cell Physiol* **47**, 319-31.

30. **Nuhse, T. S., Stensballe, A., Jensen, O. N. and Peck, S. C.** (2004). Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. *Plant Cell* **16**, 2394-405.

31. **Obenauer, J. C., Cantley, L. C. and Yaffe, M. B.** (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* **31**, 3635-41.

32. **Oh, M. H., Ray, W. K., Huber, S. C., Asara, J. M., Gage, D. A. and Clouse, S. D.** (2000). Recombinant brassinosteroid insensitive 1 receptor-like kinase autophosphorylates on serine and threonine residues and phosphorylates a conserved peptide motif in vitro. *Plant Physiol* **124**, 751-66.

33. **Reuveni, M. and DuPont, F. M.** (2001). Manganese enhances the phosphorylation of membrane-associated proteins isolated from barley roots. *Journal of Plant Physiology* **158**, 699-708.

34. **Rubio, V., Shen, Y., Saijo, Y., Liu, Y., Gusmaroli, G., Dinesh-Kumar, S. P. and Deng, X. W.** (2005). An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. *Plant J* **41**, 767-78.

35. **Sawasaki, T., Ogasawara, T., Morishita, R. and Endo, Y.** (2002). A cell-free protein synthesis system for high-throughput proteomics. *Proc Natl Acad Sci U S A* **99**, 14652-7.

36. **Schaller, G. E. and Bleecker, A. B.** (1993). Receptor-like kinase activity in membranes of Arabidopsis thaliana. *FEBS Lett* **333**, 306-10.

37. **Scheer, J. M. and Ryan, C. A.** (1999). A 160-kD systemin receptor on the surface of lycopersicon peruvianum suspension-cultured cells. *Plant Cell* **11**, 1525-36.

38. **Scheer, J. M. and Ryan, C. A., Jr.** (2002). The systemin receptor SR160 from Lycopersicon peruvianum is a member of the LRR receptor kinase family. *Proc Natl Acad Sci U S A* **99**, 9585-90.

39. **Schmidt, E. D. L., Guzzo, F. and Vries, M. A. J. T. S. C. d.** (1997). A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent tp form embryos. *Development* **124**, 2049-2062.

40. **Shah, K., Russinova, E., Gadella, T. W., Jr., Willemse, J. and De Vries, S. C.** (2002). The Arabidopsis kinase-associated protein phosphatase controls internalization of the somatic embryogenesis receptor kinase 1. *Genes Dev* **16**, 1707-20.

41. **Shah, K., Vervoort, J. and de Vries, S. C.** (2001). Role of threonines in the Arabidopsis thaliana somatic embryogenesis receptor kinase 1 activation loop in phosphorylation. *J Biol Chem* **276**, 41263-9.

42. **Shiu, S. H. and 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**, 10763-8.

43. **Shiu, S. H. and Bleecker, A. B.** (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. *Plant Physiol* **132**, 530-43.

44. **Shiu, S. H., Karlowski, W. M., Pan, R., Tzeng, Y. H., Mayer, K. F. and Li, W. H.** (2004). Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. *Plant Cell* **16**, 1220-34.

45. **Song, W. Y., Wang, G. L., Chen, L. L., Kim, H. S., Pi, L. Y., Holsten, T., Gardner, J., Wang, B., Zhai, W. X., Zhu, L. H. et al.** (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270, 1804-6.

46. **Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K. et al.** (2002). A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* 417, 959-62.

47. **Torii, K. U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R. F. and Komeda, Y.** (1996). The Arabidopsis ERECTA gene encodes

a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell* **8**, 735-46.

48. **Voinnet, O., Rivas, S., Mestre, P. and Baulcombe, D.** (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J* **33**, 949-56.

49. **Walker, J. C. and Zhang, R.** (1990). Relationship of a putative receptor protein kinase from maize to the S-locus glycoproteins of Brassica. *Nature* **345**, 743-6.

50. **Wang, X., Goshe, M. B., Soderblom, E. J., Phinney, B. S., Kuchar, J. A., Li, J., Asami, T., Yoshida, S., Huber, S. C. and Clouse, S. D.** (2005). Identification and functional analysis of in vivo phosphorylation sites of the Arabidopsis BRASSINOSTEROID-INSENSITIVE1 receptor kinase. *Plant Cell* 17, 1685-703.

51. **Williams, R. W., Wilson, J. M. and Meyerowitz, E. M.** (1997). A possible role for kinase-associated protein phosphatase in the Arabidopsis CLAVATA1 signaling pathway. *Proc Natl Acad Sci U S A* **94**, 10467-72.

Discussion et perspectives

L'article présenté au chapitre 2 : 'Plant bioactive peptides: an expanding class of signaling molecules' fait état des connaissances actuels sur les petits peptides de signalisation connus chez les plantes. Nous avons délibérément choisi de discuter de ces peptides moins connus car les papiers de synthèse s'attardent généralement aux mêmes peptides et plusieurs papiers de synthèse semblables avaient déjà été publiés sur ces peptides relativement bien caractérisés (Franssen et Bisseling, 2001; Ryan et Pearce, 2001; Ryan et al., 2002; Chilley, 2003; Boller, 2005). La découverte de nouveaux peptides de signalisation potentiels (et orphelins) s'est accrue de façon exponentielle au cours des cinq dernières années principalement à cause des nouveaux outils de bioinformatiques et de la disponibilité de séquences complètes de génomes.

La principale nouveauté dans le domaine des petits peptides de signalisation est la découverte de famille de peptides comme les RALF, CLE, IDA, etc (Cock et McCormick, 2001; Olsen et al., 2002; Butenko et al., 2003). Sur le plan biologique, la découverte de ces familles, dont les membres présentent souvent un fort taux d'identité de séquence entre eux est extrêmement intéressante. Bien que dans certains cas une redondance fonctionnelle ait été observée (Fiers et al., 2005), il semble que l'expression spatio-temporelle soit fortement régulé de façon à restreindre les domaines d'expressions de chacun des membres (Butenko et al., 2003).

Autre élément de compréhension relativement nouveau dans ce domaine est l'abondance de modifications post-traductionnelles chez ces petits peptides. Il y a déjà sept ans que des additions de groupements comme la sulfatation des phytosulfokines avait été démontrés (Yang et al., 1999). Plus récemment, l'hydroxyprolination des systémines de tomate a été mise en évidence. Cependant, le clivage de peptide signal semble être la

modification la plus fréquemment rencontrée. Plus récemment, la présence de sites de clivages en N-terminal des peptides matures a été démontré chez les RALF (Pearce et al., 2001), mais aussi en C-terminal des peptides matures comme chez CLAVATA3 (Pearce et al., 2001) semble aussi être une caractéristique fréquente chez ce type de protéines. Bien que très peu de résultats (séquençage protéique, MALDI-TOF MS) aient démontré la présence de site de clivage comme pour RALF et CLAVATA3, un grand nombre de petits peptides de signalisation potentiels présentent des caractéristiques de séquences particulières laissant présager que les sites de clivages internes seraient une caractéristique répandue. Ces caractéristiques sont la présence d'une séquence très conservée en C-terminal de la protéine et la présence de résidus dibasiques en amont et en aval de ce bloc de résidus conservés. Nous avions postulé dans notre article qu'il ne serait guère surprenant d'observer des sites de clivages en C-terminal des petits peptides, bien que jamais démontré cette hypothèse tenait du fait que nous avions observé la présence de résidus dibasiques conservés en C-terminal des petits peptides. Des résultats très récents ont démontrés que notre hypothèse s'avérait juste. Deux groupes indépendants ont publiés en même temps l'observation de clivage en C-terminal du domaine CLE de CLAVATA3 (Fiers et al., 2006; Kondo et al., 2006).

Dans le troisième chapitre, nous rapportons l'utilisation d'une méthode d'enrichissement de banque des ADNc en ARN faiblement exprimé. Nous avons souhaité enrichir notre population d'ARN messager en ARN rare afin d'obtenir de l'information sur des gènes qui sont soit : 1) exprimés dans un petit nombre de cellules, ou 2) exprimés dans un bon nombre de cellule mais à un niveau très faible. Puisque notre laboratoire se concentre sur les événements de transduction de signaux se produisant pendant le développement embryonnaire, la petite taille des structures en cause; l'embryon lui-même, les synergides, le placenta, etc, nécessitait une méthode d'enrichissement sans quoi ces ARN aurait été 'noyés' parmi les clones correspondant à des ARN très abondants. Afin d'obtenir de l'information sur les gènes faiblement exprimés, les projets de séquençages d'EST ne font habituellement qu'augmenter le nombre de clones séquencés (Van der

Hoeven et al., 2002; Ronning et al., 2003), tout en sachant très bien que le nombre d'unigènes n'augmentent pas de façon proportionnelle avec le nombre de séquences. Cette approche a pour effet de rendre les projets de séquençage d'EST très couteux, en plus de diminuer la quantité d'information obtenue par séquence. Notre méthode a permis de diminuer de façon substantielle la redondance des clones séquencés, donc d'augmenter le nombre d'unigènes et, par le fait même, de diminuer les coûts liés au séquençage.

Nous avons séquencé plus de 7 700 EST d'ovules de *Solanum chacoense* récoltés à différents stade du développement. De ces 7 700 clones, nous avons obtenus 6 374 unigènes ce qui est très largement supérieur aux autres projet de séquençage d'EST de taille similaires. Par exemple, un projet réalisé chez le tournesol visant à séquencer des gènes de banques aussi construites à partir d'embryon a permis d'isoler 43,1% d'unigènes à partir de 7 106 EST (Ben et al., 2005). Un autre projet de 11 954 EST chez le *Cassava* a permis de trouver 47,7% d'unigènes (Lopez et al., 2004). La grande force de la méthode que nous avons utilisée est donc la surreprésentation en unigènes suite à la soustraction par rapport à des projets ne faisant pas appel à cette méthode. De plus, nos résultats démontrent clairement que les deux banques que nous avons utilisés c'est-à-dire ovaires de 2 à 6 jours (SV5) et ovaires dépéricarpés de 7 à 17 jours (SV6) contiennent des populations d'ARN messagers complètement différentes. SV5 nous a permis de trouver 3113 unigènes alors que SV6 nous a permis de repêcher 3565 unigènes pour un total de 6678 unigènes. Seuls 304 unigènes étaient communs entre les deux banques, confirmant l'importante différence entre le programme transcriptionnel du développement de l'embryon précoce (2-6 jours) et celui de l'embryon tardif (7-17 jours).

Puisque notre approche permet d'aller chercher des gènes faiblement exprimés, donc rarement séquencés par les approches traditionnelles, nous avons pu *découvrir* des gènes qui n'avaient jamais été séquencés auparavant. En effet ≈5% de nos séquences, soit 387 unigènes correspondent à des gènes qui ne se trouvaient pas dans les bases de données publiques avant que nous les y déposions. Cela est vraiment considérable si nous prenons

en considération qu'il y avait déjà lorsque nous avions soumis notre article plus de 400 000 EST de *Solanum lycopersicum* et *Solanum tuberosum* (soit plus de 30 000 unigènes du premier et 17 000 unigènes du second).

L'analyse du niveau d'expression par RT-PCR de trente récepteurs kinase a ensuite révélé que 28 des 30 récepteurs analysés présentaient leur maximum d'expression dans les fruits suite à la fécondation. De ces 28 PRK, 25 étaient de la famille des LRR-PRK. De plus, les deux PRK dont le niveau d'expression maximal n'était pas dans les fruits n'étaient pas des LRR-PRK mais de la famille des L-lectines. Des trente PRK, aucun n'avaient d'ortologues ayant une fonction connue chez *A. thaliana*. Il peut paraître un peu surprenant que les récepteurs dont nous avons analysé le niveau d'expression dans les différents tissus par RT-PCR aient, tous ou presque, leur maximum d'expression dans les fruits en développement (28/30). Ce résultat peut sembler paradoxal puisque ces gènes avaient été sélectionnés sur la base du fait qu'ils étaient faiblement exprimés. Une hypothèse plausible (mais peu réjouissante) pouvant expliquer ce résultat serait la possibilité que tous les gènes (et pas seulement ces PRK) soient fortement exprimés à ce moment dans la cellule. Ce n'est pas improbable puisque, suite à la fécondation, le fruit est en croissance extrêmement active et sa masse sera multipliée plusieurs centaines (voire des milliers) de fois dans les jours suivants. Si cette hypothèse est vraie, le fait que les trente récepteurs kinases démontrent leur maximum d'expression à ce stade ne serait pas vraiment informatif au niveau développemental, mais serait plutôt le reflet d'un accroissement d'activité transcriptionelle générale. Heureusement, nous avons maintenant des données de micropuces d'ADN qui démontre clairement que plusieurs patrons d'expressions sont observés parmi nos gènes, et que tous ne sont pas activés transcriptionnellement suite à la fécondation ou la pollinisation. Par exemple, 114 gènes sont induits suite à la pollinisation alors que 186 gènes sont réprimés lors de la pollinisation. Donc nos récepteurs avaient le 'choix' d'être induits ou réprimés. Ces nouveaux résultats de micropuces supportent donc notre conclusion initiale qui suggérait que le haut niveau d'expression de ces gènes dans les tissus du gamétophyte femelle soit vraiment le reflet de leur rôle au niveau de ce tissu. Donc cette nouvelle

méthode d'enrichissement de banque permet d'obtenir beaucoup plus d'unigène que les méthodes conventionnelles, mais en plus elle permet de trouver des gènes spécifiques à certains tissus, cellules ou fonctions.

Il est important de noter que les projets de micropuces d'ADN ne détectent généralement pas les variations d'expression des récepteurs kinases à cause de leur très faible niveau d'expression. Par exemple notre projet de micropuces a détecté des variations de plus de 2-fold pour seulement trois (sur trente) des récepteurs kinases de notre liste alors que nous avons observés des variations dans plusieurs cas de plus de 1000-fold. Le faible niveau d'expression des récepteurs kinases et la faible sensibilité des micropuces d'ADN fait en sorte que peu d'informations peuvent être obtenues pour les récepteurs kinases avec cette méthode (observation personnelle). Notre méthode d'enrichissement de gènes rares combiné à une analyse de RT-PCR en temps réel constitue donc une approche particulièrement efficace pour analyser les modulations d'expression de ce type de gènes.

ORK17 a donc été isolé de la banque précédemment mentionnée. *ORK17* code pour un récepteur kinase de la sous-famille LRR-VI. Son domaine d'expression évalué par Northern et par RT-PCR est très spécifique aux fruits. Une analyse d'expression plus fine intratissulaire par hybridation *in situ* d'ARN nous a permis de déterminer que bien qu'*ORK17* soit exprimé partout dans le fruit, son expression est maximale dans le tégument et dans l'endothélium de l'ovule, suggérant qu'il pourrait avoir un rôle au niveau plus spécifique de l'ovule.

À ce jour très peu de récepteurs kinases ont démontré un rôle dans l'ovule ou dans le fruit. Le gène *CRINKLY4* (CR4), qui code pour un récepteur kinase de la famille TNFR, a été découvert par insertion de transposon chez le maïs (Becraft et al., 1996). Les mutations dans le gène *CRINKLY4* cause un problème de différentiation au niveau de l'épiderme et inhibe aussi la formation de l'aleurone (Becraft et al., 1996), mais n'altère pas directement le développement de l'embryon. De son côté, le gène *EXTRA SPOROGENOUS CELLS* (*EXS*) est impliqué dans la formation des gamètes mâles et

femelles. En effet, les mutants *exs* ont un nombre accru de cellules sporogènes et ne forment pas de cellules du tapetum ou de la lamelle médiane (structure de l'anthère). Au niveau du développement des gamètes femelles, les mutants *exs* ont des embryons qui se développent plus lentement, ont des cellules plus petites et développent des graines plus petites que les plantes de type sauvage (Canales et al., 2002). Finalement, les gènes *TOADSTOOL1* et *TOADSTOOL2* découvert chez *Arabidopsis*, par le laboratoire de Frans Tax, codent pour deux récepteurs kinases dont les fonctions se chevauchent. Les doubles-mutants *toad1/toad2*, ont un patron de division anormal dans le domaine central et basal de l'embryon ce qui mène éventuellement à la mort de l'embryon.

Afin de déterminer la fonction d'ORK17 nous avons créé des plantes transgéniques sur-exprimants ORK17 ou une version tronquée d'ORK17 n'ayant pas de domaine intracellulaire. Nous avons utilisé ces constructions complètes et tronquées pour transformer *Solanum chacoense* et *A. thaliana*. Nous avons obtenu un phénotype pour *ORK17* à l'aide d'une transformation hétérologue dans *Arabidopsis*. Généralement les transformations hétérologues sont utilisées pour vérifier si deux protéines de deux organismes différents peuvent effectuer la même fonction. Par exemple, nous avons déjà transformé une drosophile mutante pour le gène *Notchless* avec le gène *Notchless* de *S. chacoense* pour voir si celle-ci pouvait restaurer le phénotype sauvage (Chantha et al., 2006). Dans notre cas, les plantes transgéniques d'*Arabidopsis* que nous avons généré suite à une transformation avec un clone de *S. chacoense* ont développé des siliques contenant un nombre de graines inférieurs par rapport au type sauvage. Nous avions préalablement généré des plantes transgéniques chez *S. chacoense* avec cette même construction (sans domaine intracellulaire) mais nous n'avions pas obtenu de différence par rapport au phénotype sauvage et ce malgré une forte expression du transgène. Nous avons émis certaines hypothèses pour expliquer ce phénomène; dans le cas des constructions sans domaine intracellulaire, il est possible que la protéine tronquée soit traduite mais qu'un mécanisme de contrôle de qualité fasse en sorte que cette protéine 'anormale' soit dégradée (Pedrazzini et al., 1997). Dans le cas des plantes qui devraient sur-exprimer le clone pleine

longueur, mais qui finalement sous-expriment (plantes co-supprimées), il est possible qu'un phénomène de redondance, utilisant un récepteur semblable, soit utilisé pour effectuer la signalisation (n'oublions pas qu'il y a plus de 200 LRR-RLK). Finalement, il est problable que nous obtenions un phénotype chez *A. thaliana* avec notre clone de *S. chacoense* car la séquence est suffisamment différente de son ortologue et ce qui lui permet d'échapper au mécanisme de 'silencing' de *A. thaliana* (Hirschi et al., 2000).

ORK28 code lui aussi pour un récepteur kinase de type LRR mais appartient à la sous-famille LRR-V. La sous-famille LRR-V contient un total de 9 membres chez *A. thaliana* dont un membre connu appelé *STRUBBELIG* (*SUB*) (Chevalier et al., 2005). Il s'agit aussi du même gène que *SCRAMBLED* (Kwak et al., 2005a). Le gène *SCRAMBLED* a été décrit comme un gène requis par la cellule pour interpréter sa position dans l'épiderme racinaire en développement et pour réguler l'expression des gènes *GLABRA2, CAPRICE, WEREWOLF,* et *ENHANCER OF GLABRA3* tous des facteurs de transcription impliqués dans le destin cellulaire des cellules racinaires (Kwak et al., 2005a). De son côté, *STRUBBELIG* a été décrit comme ayant un rôle plus général, c'est-à-dire que *SUB* influencerait la formation et la forme de plusieurs organes en influençant la morphogenèse cellulaire, l'orientation des divisions cellulaires, et la prolifération cellulaire (Chevalier et al., 2005). Des analyses de mutagenèses dirigées suggèrent que le domaine kinase soit nécessaire à la fonction de *SUB* quoique celui-ci n'ait pas d'activité kinase *in vitro* (Chevalier et al., 2005). Les autres membres de la famille sont appelés SRF (*STRUBBELIG-related family*) mais n'ont pas de phénotype publié et tous se classe dans le groupe des kinases ACF tout comme ORK28.

Tout comme *ORK17*, et plusieurs autres récepteurs que nous avons identifiés, ORK28 présente un profil d'expression d'ARN messagers par RT-PCR et par Northern relativement spécifique aux tissus reproducteurs femelles. Nous avons donc suspecté qu'il pouvait jouer un rôle à ce niveau et donc poussé l'investigation un peu plus à fond.

Nous avons d'abord cherché à déterminer si ORK28 était vraiment localisé au niveau de la membrane plasmique. Pour ce faire, nous avons fusionné ORK28 à la *green fluorescent protein*. Nous savons clairement pu démontrer qu'ORK28 est localisé à la membrane plasmique dans les cellules d'oignons et dans les cellules de tabac. Bien que ça puisse sembler une évidence, puisqu'il s'agit d'un récepteur, des exemples de récepteurs non-localisés à la membrane plasmique ont déjà été démontrés. Par exemple, la communauté scientifique a crû pendant des années, à tort, que le récepteur de l'éthylène était localisé à la membrane plasmique alors qu'il était en fait à la membrane du réticulum endoplasmique (Chen et al., 2002). Plus récemment un groupe japonais a démontré qu'un récepteur kinase de la famille des LRR-PRK était localisé à la membrane des vésicules de Golgi (Goto et al., 2005). Bien qu'il soit probable que la majorité des récepteurs kinases soient localisés à la membrane plasmique, je crois qu'il est prudent de confirmer cette hypothèse, ce que nous avons fait.

Ensuite nous avons tenté de déterminer si le domaine kinase d'OR28 pouvait s'auto/transphosphoryler. ORK28 fait partie de la famille des kinases ACF, laquelle est définie par l'absence de certains résidus que l'on croit important pour l'activité catalytique. Malgré l'absence de certains résidus important l'activité de récepteur kinase à domaine ACF a déjà été démontrée (Mu et al., 1994; Li et al., 2005). Nous avons pu déterminer qu'ORK28 est une kinase fonctionnelle mais que l'amputation du domaine juxtamembranaire la rend inactive. De plus, il n'a pas été possible de phosphoryler des substrats génériques tels que l'histone H3, la protéine basique de la myéline (MBP) ou la version tronqué d'ORK28 sans domaine juxtamembranaire. Les essais kinases *in vitro* réalisé par Chevalier et al. (2005) avec SUB avait démontrés qu'elle était inactive avec ou sans domaine juxtamembranaire. Il faut noter cependant qu'il manque à SUB plusieurs résidus dans des régions fortement conservées (Chevalier et al., 2005).

Les études précédentes portant sur d'autres récepteurs ne s'attardaient pas à la présence ou l'absence du domaine juxtamembranaire dans leurs constructions et ce malgré

les papiers de Pamela Ronald et ceux de Scott Peck qui démontraient respectivement que le domaine juxtamembranaire est nécessaire pour la perception (He et al., 2000) et qu'il peut être abondamment phosphorylé (Nuhse et al., 2004). Par exemple des essais kinase réalisés avec une construction comprenant seulement le domaine kinase et C-terminal du récepteur MARK se sont révélés infructueux et ont mené les auteurs à conclure que leur kinase était inactive (Llompart et al., 2003). Par ailleurs, des essais kinases réalisés avec le récepteur CLAVATA1, à partir d'une construction appelée 'kinase domain' ont fonctionné mais comprenaient le domaine juxtamembranaire (Williams et al., 1997). De la même façon, les essais kinases réalisés avec le domaine intracellulaire du récepteur CRINKLY4 (chez le maïs) ont aussi fonctionnés et comprenaient le domaine juxtmembranaire (Jin et al., 2000). Bref, plusieurs exemples démontrent que des kinases de récepteurs sont actives lorsqu'elles possèdent leur domaine juxtamembranaire (CLAVATA1, CRINKLY4, BRI1) et l'affirmation que la PRK MARK était inactive est sujette à caution puisque les constructions utilisées n'incluaient pas le domaine JM. Nos résultats démontrent hors de tout doute que la PRK ScORK28 ne peut être active en absence de son domaine juxtamembranaire. De plus, nous avons démontré que la kinase active ne peut phosphoryler la kinase tronquée. Ces résultats peuvent être expliqué de différentes façons; soit il s'agit de phosphorylation ayant lieu dans le domaine juxtamembranaire; soit il peut s'agir de transphosphorylation nécessitant la présence du domaine juxtamembranaire pour l'interaction entre les deux protéines.

Afin de tenter de déterminer le rôle d'ORK28 nous avons fait des constructions sur-exprimant soit, ORK28 pleine longueur ou une version dominante négative, c'est-à-dire amputée de son domaine intracellulaire. Bien que nous ayons obtenus des plantes sur-exprimants l'ARN messagers et exprimants la protéine, aucun défaut développemental apparent n'a été observé. Encore une fois, le fait qu'il s'agit d'une famille de gène pourrait expliquer le fait que nous n'observions pas de phénotype. Il est intéressant de noter que seul deux plantes sur-exprimants la protéines tronquées ont été obtenu à partir de 20 plantes et elles exprimaient la protéine à un niveau très faible (considérant qu'elles étaient sous le

contrôle du promoteur fort 35S) indiquant un mécanisme de contrôle très stricte de l'expression de cette protéine. De tels mécanismes sont communs (Hirschi et al., 2000) et pourraient possiblement être éviter si nous utilisions un système de transformation hétérologue comme nous avons fait pour ORK17 afin de déjouer le mécanisme de *silencing* de la plante.

Directions futures

Nous avons découvert plusieurs récepteurs kinase, mais il nous manque encore beaucoup d'information concernant leur fonctionnement ou le fonctionnement des récepteurs de plante en général. Je vais donc énumérer ci-après certaines expériences qui pourraient être réalisés par quelqu'un qui souhaiterait poursuivre mon projet ou entreprendre des projets reliés.

Tel que mentionné dans le résumé, nous aurions bien souhaité pouvoir découvrir un ligand pour nos récepteurs, cependant nous n'avons pas eu le temps de nous attaquer à cette tâche. Une technique classique pour déterminer une interaction entre deux protéine est la chromatograhie d'affinité. Cependant, l'importante différence de taille entre le récepteur et le ligand pose un problème au niveau de la détection. En effet, si on suppose que tous les récepteurs, d'une taille hypothétique, mais réaliste de 100 kilodalton, que nous aurions fixés à une colonne puissent lier chacun un ligand (taille entre 1 et 5 kilodalton) que nous récupérons à 100% lors de l'élution, celui-ci sera récupéré en quantité entre 20 et 100 fois plus faible que le récepteur et sera donc difficile à détecter. Afin de résoudre ce problème, il faudrait se servir du ligand comme appât, mais cela est impossible dans notre cas puisque nous ignorons son identité. De plus, le très faible niveau d'expression de ces gènes, qui sont souvent exprimés dans un domaine très restreint ou dans un très petit nombre de cellules, rend l'approche par chromatographie d'affinité pratiquement impossible dans notre cas. Pour s'attaquer à la découverte d'un ligand je priviligierais donc plutôt une approche par 'phage display'. Il nous est possible d'exprimer ORK17 de façon transitoire *in planta* et de

le purifier en quantité qui serait suffisante pour ce type d'approche. La protéine purifiée pourrait servir d'appât pour des phages exprimant à leur surface des épitopes dont il nous est possible d'obtenir la séquence et de la comparer par la suite aux séquences de nos banques ou du génome d'*Arabidopsis*. Je ne privilégierais certainement pas une approche par deux-hybrides parce que nous avons démontré que nos récepteurs sont glycosylés et donc se retrouveraient dans une conformation non 'naturelle' dans la cellule. De plus, les banques contiennent rarement les petits ARNm, codants pour les petits peptides de signalisation qui subissent généralement eux aussi des modifications post-traductionnelles importantes.

Il serait aussi intéressant de déterminer si nos récepteurs possèdent la propriété de s'autodimériser *in vivo*. À ces fins, nous pourrions utiliser la méthode de BiFC (*bimolecular fluorecence complementation*) (Bracha-Drori et al., 2004). Cette méthode consiste à fusionner une protéine d'intérêt à la portion N-terminale de la protéine fluorescente jaune (YFP) et une autre protéine d'intérêt avec la portion C-terminale de la YFP. Si les deux protéines d'intérêts interagissent physiquement entre elles, l'émission de fluorescence serait observée en microscopie à fluorescence. Nous avons déjà utilisé cette méthode au laboratoire en exprimant les constructions de façon transitoire dans des cellules d'oignons et elle fonctionne bien.

Un autre aspect qui mérite que l'on s'y attarde est le recyclage des récepteurs. Tel que mentionné dans l'introduction, nous savons que chez les mammifères, le récepteur est internalisé dans une vésicule suite à son activation (Dautry-Varsat et al., 1983). Cette internalisation permet de rendre inactif le récepteur. Le ligand et le récepteur se dissocient à l'intérieur de la vésicule et le récepteur est retourné à la surface de la cellule si cela s'avère nécessaire. Chez les plantes, nous savons que le récepteur peut être internalisé suite à son activation (Shah et al., 2002; Russinova et al., 2004) mais nous ignorons son destin par la suite. Je crois que l'utilisation d'un récepteur très bien caractérisé comme BRI1 et dont le l'activateur peut s'acheter serait idéal. Il serait possible d'exprimer de façon transitoire en

protoplaste ou de faire un transgénique stable sur-exprimant une fusion BRI1/GFP, d'activer le récepteur avec une solution de brassinolide et de suivre le destin du récepteur en microscopie. Je crois qu'il s'agirait d'une première si le recyclage pouvait être démontré chez les plantes et je crois qu'un système BRI1/brassinolide serait optimal.

Conclusion

Nous avons présenté au chapitre 2 une publication synthèse portant sur de petits peptides de signalisation potentiels généralement peu connus. Cette classe de protéines découvertes depuis peu s'est élargie à une vitesse fulgurante depuis quelques années. Malgré le grand nombre de protéines faisant partie de cette classe, la très grande majorité d'entre elles demeurent orphelines de fonction. Le défi des prochaines années sera de réussir à trouver des fonctions à ces peptides malgré tous les obstacles techniques mentionnés au chapitre 2.

Le chapitre 3 présente les résultats d'un projet de séquençage d'EST. Ce projet est novateur car il permet d'introduire un biais vers les gènes qui sont faiblement exprimés. De cette façon, il est possible d'obtenir plus d'unigènes par réaction de séquence qu'en faisant un projet de séquençage traditionnel. Avoir plus d'unigènes permet d'avoir plus d'information sur les gènes qui sont exprimés dans un tissu à un moment donné, soit son transcriptome. De plus, dans cet article, nous avons analysé les niveaux d'expression dans différents tissus de trente récepteurs kinases différents. Il s'agit du plus grand projet d'analyse d'expression de récepteurs kinases de plantes par RT-PCR quantitatif et de façon surprenante presque tous ces récepteurs ont montré un maximum d'expression dans les stades précoces des fruits en développement.

Les chapitres quatre et cinq présentent une caractérisation de deux récepteurs kinases que nous avons isolés de notre banque. Nous avons choisi ces deux récepteurs car ils s'expriment de façon relativement spécifique dans les fruits après la pollinisation, suggérant qu'ils pouvaient avoir une fonction dans ce tissu. Bien qu'ORK28 soit exprimé très spécifiquement dans le fruit, nous n'avons pu observer de phénotype dans nos lignées transgéniques sur-exprimantes ou dominantes négatives de même que chez les mutants insertionels d'*Arabidopsis*. Nous avons obtenus des données biologiques permettant de dire qu'ORK28 code pour un récepteur kinase fonctionnel, glycosylé et lié à la membrane

plasmique. Pour ORK17 nous savons qu'il est aussi glycosylé et que le clone provenant de *S. chacoense* donne un phénotype de courtes siliques et un nombre de graines réduits lorsque transformé dans *A. thaliana en version* dominant négatif. Des expériences additionnelles seraient intéressantes sur ces récepteurs kinases afin de comprendre leur mécanisme d'action.

Tel que mentionné dans l'introduction l'état des connaissances chez les récepteurs kinases animaux se trouve largement en avance par rapport à l'état des connaissances chez les récepteurs de plantes. Les domaines souffrant le plus de retard sont les caractérisations biochimiques d'interactions entre un récepteur et son ligand, les mécanismes de dimérisation entre un récepteur et un autre récepteur partenaire, la quasi-totale absence de voie de signalisation bien caractérisées en aval des récepteurs et les mécanismes d'inactivations et de recyclages des récepteurs. Afin de vraiment faire progresser les connaissances sur les récepteurs kinases de plantes je crois qu'en plus de continuer à chercher les fonctions d'un nombre sans cesse croissant de récepteurs il faut avant tout approfondir les connaissances sur un petits nombre de récepteurs comme BRI1, CLV1, SRK ou SERK comme il a été fait chez les animaux avec les ERBB afin de vraiment comprendre les mécanismes sous-jacents à l'activation des situés en aval.

Bibliographie

Aderem, A., and Ulevitch, R.J. (2000). Toll-like receptors in the induction of the innate immune response. Nature **406,** 782-787.

Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.L., Gomez-Gomez, L., Boller, T., Ausubel, F.M., and Sheen, J. (2002). MAP kinase signalling cascade in Arabidopsis innate immunity. Nature **415,** 977-983.

Becraft, P.W., Stinard, P.S., and McCarty, D.R. (1996). CRINKLY4: A TNFR-like receptor kinase involved in maize epidermal differentiation. Science **273,** 1406-1409.

Ben, C., Hewezi, T., Jardinaud, M.F., Bena, F., Ladouce, N., Moretti, S., Tamborindeguy, C., Liboz, T., Petitprez, M., and Gentzbittel, L. (2005). Comparative analysis of early embryonic sunflower cDNA libraries. Plant molecular biology **57,** 255-270.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J Mol Biol **340,** 783-795.

Boller, T. (2005). Peptide signalling in plant development and self/non-self perception. Current opinion in cell biology **17,** 116-122.

Bower, M.S., Matias, D.D., Fernandes-Carvalho, E., Mazzurco, M., Gu, T., Rothstein, S.J., and Goring, D.R. (1996). Two members of the thioredoxin-h family interact with the kinase domain of a Brassica S locus receptor kinase. The Plant cell **8,** 1641-1650.

Bracha-Drori, K., Shichrur, K., Katz, A., Oliva, M., Angelovici, R., Yalovsky, S., and Ohad, N. (2004). Detection of protein-protein interactions in plants using bimolecular fluorescence complementation. Plant J **40,** 419-427.

Braun, D.M., Stone, J.M., and Walker, J.C. (1997). Interaction of the maize and Arabidopsis kinase interaction domains with a subset of receptor-like protein kinases: implications for transmembrane signaling in plants. Plant J **12,** 83-95.

Burgess, A.W., Cho, H.S., Eigenbrot, C., Ferguson, K.M., Garrett, T.P., Leahy, D.J., Lemmon, M.A., Sliwkowski, M.X., Ward, C.W., and Yokoyama, S. (2003). An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. Molecular cell **12,** 541-552.

Butenko, M.A., Patterson, S.E., Grini, P.E., Stenvik, G.E., Amundsen, S.S., Mandal, A., and Aalen, R.B. (2003). Inflorescence deficient in abscission controls floral organ abscission in Arabidopsis and identifies a novel family of putative ligands in plants. The Plant cell **15,** 2296-2307.

Cabrillac, D., Cock, J.M., Dumas, C., and Gaude, T. (2001). The S-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins. Nature **410,** 220-223.

Canales, C., Bhatt, A.M., Scott, R., and Dickinson, H. (2002). EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in Arabidopsis. Curr Biol **12,** 1718-1727.

Chantha, S.C., Emerald, B.S., and Matton, D.P. (2006). Characterization of the plant Notchless homolog, a WD repeat protein involved in seed development. Plant molecular biology **62,** 897-912.

Charpenteau, M., Jaworski, K., Ramirez, B.C., Tretyn, A., Ranjeva, R., and Ranty, B. (2004). A receptor-like kinase from Arabidopsis thaliana is a calmodulin-binding protein. Biochem J **379,** 841-848.

Chen, Y.F., Randlett, M.D., Findell, J.L., and Schaller, G.E. (2002). Localization of the ethylene receptor ETR1 to the endoplasmic reticulum of Arabidopsis. The Journal of biological chemistry **277,** 19861-19866.

Chevalier, D., Batoux, M., Fulton, L., Pfister, K., Yadav, R.K., Schellenberg, M., and Schneitz, K. (2005). STRUBBELIG defines a receptor kinase-mediated signaling pathway regulating organ development in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **102,** 9074-9079.

Chilley, P. (2003). Polypeptide hormones: signaling molecules in plants. Vitam Horm **66,** 317-344.

Clark, S.E., Running, M.P., and Meyerowitz, E.M. (1993a). CLAVATA1, a regulator of meristem and flower development in Arabidopsis. Development (Cambridge, England) **119,** 397-418.

Clark, S.E., Running, M.P., and Meyerowitz, E.M. (1993b). *Clavata1*, a regulator of meristem and flower development in *Arabidopsis*. Development **119,** 397-418.

Clouse, S.D., Langford, M., and McMorris, T.C. (1996). A brassinosteroid-insensitive mutant in Arabidopsis thaliana exhibits multiple defects in growth and development. Plant physiology **111,** 671-678.

Cock, J.M., and McCormick, S. (2001). A large family of genes that share homology with CLAVATA3. Plant physiology **126,** 939-942.

Cock, J.M., Vanoosthuyse, V., and Gaude, T. (2002). Receptor kinase signalling in plants and animals: distinct molecular systems with mechanistic similarities. Current opinion in cell biology **14,** 230-236.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol **133,** 462-469.

Dardick, C., and Ronald, P. (2006). Plant and animal pathogen recognition receptors signal through non-RD kinases. PLoS Pathog **2,** e2.

Dautry-Varsat, A., Ciechanover, A., and Lodish, H.F. (1983). pH and the recycling of transferrin during receptor-mediated endocytosis. Proceedings of the National Academy of Sciences of the United States of America **80,** 2258-2262.

Derynck, R., and Zhang, Y.E. (2003). Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature **425,** 577-584.

Dievart, A., and Clark, S.E. (2003). Using mutant alleles to determine the structure and function of leucine-rich repeat receptor-like kinases. Curr Opin Plant Biol **6,** 507-516.

Ehsan, H., Ray, W.K., Phinney, B., Wang, X., Huber, S.C., and Clouse, S.D. (2005). Interaction of Arabidopsis BRASSINOSTEROID-INSENSITIVE 1 receptor kinase with a homolog of mammalian TGF-beta receptor interacting protein. Plant J **43,** 251-261.

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kalo, P., and Kiss, G.B. (2002). A receptor kinase gene regulating symbiotic nodule development. Nature **417,** 962-966.

Fiers, M., Golemiec, E., Xu, J., van der Geest, L., Heidstra, R., Stiekema, W., and Liu, C.M. (2005). The 14-Amino Acid CLV3, CLE19, and CLE40 Peptides Trigger Consumption of the Root Meristem in Arabidopsis through a CLAVATA2-Dependent Pathway. The Plant cell **Online first**.

Fiers, M., Golemiec, E., van der Schors, R., van der Geest, L., Li, K.W., Stiekema, W.J., and Liu, C.M. (2006). The CLAVATA3/ESR motif of CLAVATA3 is functionally independent from the nonconserved flanking sequences. Plant physiology **141,** 1284-1292.

Franssen, H.J., and Bisseling, T. (2001). Peptide signaling in plants. Proceedings of the National Academy of Sciences of the United States of America **98,** 12855-12856.

Friedrichsen, D.M., Joazeiro, C.A., Li, J., Hunter, T., and Chory, J. (2000). Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. Plant Physiol **123,** 1247-1256.

Germain, H., Rudd, S., Zotti, C., Caron, S., O'Brien, M., Chantha, S.C., Lagace, M., Major, F., and Matton, D.P. (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,** 515-532.

Giranton, J.L., Dumas, C., Cock, J.M., and Gaude, T. (2000). The integral membrane S-locus receptor kinase of Brassica has serine/threonine kinase activity in a membranous environment and spontaneously forms oligomers in planta. Proceedings of the National Academy of Sciences of the United States of America **97,** 3759-3764.

Gomez-Gomez, L., and Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Mol Cell **5,** 1003-1011.

Gomez-Gomez, L., Bauer, Z., and Boller, T. (2001). Both the extracellular leucine-rich repeat domain and the kinase activity of FSL2 are required for flagellin binding and signaling in Arabidopsis. The Plant cell **13,** 1155-1163.

Goto, S., Niwa, Y., and Kobayashi, H. (2005). A cytoplasmic signal of receptor-like kinase in golgi vesicles to chloroplast biogenesis. In Protein Phosphorylation in Plant Signaling (Tsukuba, Japan).

Gupta, R., and Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. In Pacific Symposium on Biocomputing, pp. 310-322.

Hashimoto, S., Matsumoto, K., Gon, Y., Furuichi, S., Maruoka, S., Takeshita, I., Hirota, K., Yodoi, J., and Horie, T. (1999). Thioredoxin negatively regulates p38 MAP kinase activation and IL-6 production by tumor necrosis factor-alpha. Biochemical and biophysical research communications **258,** 443-447.

He, Z., Wang, Z.Y., Li, J., Zhu, Q., Lamb, C., Ronald, P., and Chory, J. (2000). Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. Science **288,** 2360-2363.

Hirschi, K.D., Korenkov, V.D., Wilganowski, N.L., and Wagner, G.J. (2000). Expression of arabidopsis CAX2 in tobacco. Altered metal accumulation and increased manganese tolerance. Plant physiology **124,** 125-133.

Holland, S.J., Gale, N.W., Gish, G.D., Roth, R.A., Songyang, Z., Cantley, L.C., Henkemeyer, M., Yancopoulos, G.D., and Pawson, T. (1997). Juxtamembrane tyrosine residues couple the Eph family receptor EphB2/Nuk to specific SH2 domain proteins in neuronal cells. The EMBO journal **16,** 3877-3888.

Honegger, A., Dull, T.J., Szapary, D., Komoriya, A., Kris, R., Ullrich, A., and Schlessinger, J. (1988). Kinetic parameters of the protein tyrosine kinase activity of EGF-receptor mutants with individually altered autophosphorylation sites. The EMBO journal **7,** 3053-3060.

Honegger, A.M., Schmidt, A., Ullrich, A., and Schlessinger, J. (1990). Evidence for epidermal growth factor (EGF)-induced intermolecular autophosphorylation of the EGF receptors in living cells. Molecular and cellular biology **10,** 4035-4044.

Hunter, T., and Plowman, G.D. (1997). The protein kinases of budding yeast: six score and more. Trends in biochemical sciences **22,** 18-22.

Huse, M., Muir, T.W., Xu, L., Chen, Y.G., Kuriyan, J., and Massague, J. (2001). The TGF beta receptor activation process: an inhibitor- to substrate-binding switch. Molecular cell **8,** 671-682.

Jin, P., Guo, T., and Becraft, P.W. (2000). The maize CR4 receptor-like kinase mediates a growth factor-like differentiation response. Genesis **27,** 104-116.

Johnson, L.N., Nobel, M.E.M., and Owen, D.J. (1996a). Active and inactive protein kinases: structural basis for regulation. Cell **85,** 149-158.

Johnson, L.N., Noble, M.E., and Owen, D.J. (1996b). Active and inactive protein kinases: structural basis for regulation. Cell **85,** 149-158.

Kayes, J.M., and Clark, S.E. (1998). CLAVATA2, a regulator of meristem and organ development in Arabidopsis. Development (Cambridge, England) **125,** 3843-3851.

King, N., Hittinger, C.T., and Carroll, S.B. (2003). Evolution of key cell signaling and adhesion protein families predates animal origins. Science **301,** 361-363.

Kinoshita, T., Cano-Delgado, A., Seto, H., Hiranuma, S., Fujioka, S., Yoshida, S., and Chory, J. (2005). Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. Nature **433,** 167-171.

Kondo, T., Sawa, S., Kinoshita, A., Mizuno, S., Kakimoto, T., Fukuda, H., and Sakagami, Y. (2006). A plant peptide encoded by CLV3 identified by in situ MALDI-TOF MS analysis. Science **313,** 845-848.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305,** 567-580.

Kwak, S.H., Shen, R., and Schiefelbein, J. (2005a). Positional signaling mediated by a receptor-like kinase in Arabidopsis. Science **307,** 1111-1113.

Kwak, S.H., Shen, R., and Schiefelbein, J. (2005b). Positional signaling mediated by a receptor-like kinase in Arabidopsis. Science **307,** 1111-1113.

Lee, H.S., Chung, Y.Y., Das, C., Karunanandaa, B., van-Went, J.L., Mariani, C., and Kao, T.H. (1997). Embryo sac development is affected in Petunia inflata plants transformed with an antisense gene encoding the extracellular domain of receptor kinase PRK1. Sexual Plant Reproduction **10,** 341-350.

Li, J., and Chory, J. (1997). A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell **90,** 929-938.

Li, J., and Jin, H. (2007). Regulation of brassinosteroid signaling. Trends in plant science **12,** 37-41.

Li, J., Wen, J., Lease, K.A., Doke, J.T., Tax, F.E., and Walker, J.C. (2002). BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell **110,** 213-222.

Li, Y.L., Sun, J., and Xia, G.X. (2005). Cloning and characterization of a gene for an LRR receptor-like protein kinase associated with cotton fiber development. Mol Genet Genomics **273,** 217-224.

Llompart, B., Castells, E., Rio, A., Roca, R., Ferrando, A., Stiefel, V., Puigdomenech, P., and Casacuberta, J.M. (2003). The direct activation of MIK, a germinal center kinase (GCK)-like kinase, by MARK, a maize atypical receptor kinase, suggests a new mechanism for signaling through kinase-dead receptors. The Journal of biological chemistry **278,** 48105-48111.

Lopez, C., Jorge, V., Piegu, B., Mba, C., Cortes, D., Restrepo, S., Soto, M., Laudie, M., Berger, C., Cooke, R., Delseny, M., Tohme, J., and Verdier, V. (2004). A unigene catalogue of 5700 expressed genes in cassava. Plant molecular biology **56,** 541-554.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science **298,** 1912-1934.

Matsubayashi, Y., Ogawa, M., Morita, A., and Sakagami, Y. (2002). An LRR receptor kinase involved in perception of a peptide plant hormone, phytosulfokine. Science **296,** 1470-1472.

Morillo, S.A., and Tax, F.E. (2006). Functional analysis of receptor-like kinases in monocots and dicots. Curr Opin Plant Biol **9,** 460-469.

Mu, J.H., Lee, H.S., and Kao, T.H. (1994). Characterization of a pollen-expressed receptor-like kinase gene of Petunia inflata and the activity of its encoded kinase. The Plant cell **6,** 709-721.

Muschietti, J., Eyal, Y., and McCormick, S. (1998). Pollen tube localization implies a role in pollen-pistil interactions for the tomato receptor-like protein kinases LePRK1 and LePRK2. Plant Cell **10,** 319-330.

Nam, K.H., and Li, J. (2002). BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. Cell **110,** 203-212.

Nam, K.H., and Li, J. (2004). The Arabidopsis transthyretin-like protein is a potential substrate of BRASSINOSTEROID-INSENSITIVE 1. The Plant cell **16,** 2406-2417.

Nasrallah, J.B., S.J., R., and Nasrallah, M.E. (1994). Genetic evidence for the requirement of the Brassica S-locus receptor kinase gene in the self0incompatibility Plant Journal **5,** 373-384.

Nirmala, J., Brueggeman, R., Maier, C., Clay, C., Rostoks, N., Kannangara, C.G., von Wettstein, D., Steffenson, B.J., and Kleinhofs, A. (2006). Subcellular localization and functions of the barley stem rust resistance receptor-like serine/threonine-specific protein kinase Rpg1. Proc Natl Acad Sci U S A **103,** 7518-7523.

Niwa, Y., Goto, S., Nakano, T., Sakaiya, M., Hirano, T., Tsukaya, H., Komeda, Y., and Kobayashi, H. (2006). Arabidopsis mutants by activation tagging in which photosynthesis genes are expressed in dedifferentiated calli. Plant & cell physiology **47,** 319-331.

Nuhse, T.S., Stensballe, A., Jensen, O.N., and Peck, S.C. (2004). Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. The Plant cell **16,** 2394-2405.

Obenauer, J.C., Cantley, L.C., and Yaffe, M.B. (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic acids research **31,** 3635-3641.

Oh, M.H., Ray, W.K., Huber, S.C., Asara, J.M., Gage, D.A., and Clouse, S.D. (2000). Recombinant brassinosteroid insensitive 1 receptor-like kinase autophosphorylates on serine and threonine residues and phosphorylates a conserved peptide motif in vitro. Plant physiology **124,** 751-766.

Olsen, A.N., Mundy, J., and Skriver, K. (2002). Peptomics, identification of novel cationic Arabidopsis peptides with conserved sequence motifs. In Silico Biol **2,** 441-451.

Opresko, L.K., Chang, C.P., Will, B.H., Burke, P.M., Gill, G.N., and Wiley, H.S. (1995). Endocytosis and lysosomal targeting of epidermal growth factor receptors are mediated by distinct sequences independent of the tyrosine kinase domain. The Journal of biological chemistry **270,** 4325-4333.

Park, A.R., Cho, S.K., Yun, U.J., Jin, M.Y., Lee, S.H., Sachetto-Martins, G., and Park, O.K. (2001). Interaction of the Arabidopsis receptor protein kinase Wak1 with a glycine-rich protein, AtGRP-3. The Journal of biological chemistry **276,** 26688-26693.

Pearce, G., Strydom, D., Johnson, S., and Ryan, C.A. (1991). A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science **253,** 895-898.

Pearce, G., Moura, D.S., Stratmann, J., and Ryan, C.A., Jr. (2001). RALF, a 5-kDa ubiquitous polypeptide in plants, arrests root growth and development. Proceedings of the National Academy of Sciences of the United States of America **98,** 12843-12847.

Pedrazzini, E., Giovinazzo, G., Bielli, A., de Virgilio, M., Frigerio, L., Pesca, M., Faoro, F., Bollini, R., Ceriotti, A., and Vitale, A. (1997). Protein quality control along the route to the plant vacuole. The Plant cell **9,** 1869-1880.

Reuveni, M., and DuPont, F.M. (2001). Manganese enhances the phosphorylation of membrane-associated proteins isolated from barley roots. Journal of Plant Physiology **158,** 699-708.

Ronning, C.M., Stegalkina, S.S., Ascenzi, R.A., Bougri, O., Hart, A.L., Utterbach, T.R., Vanaken, S.E., Riedmuller, S.B., White, J.A., Cho, J., Pertea, G.M., Lee, Y., Karamycheva, S., Sultana, R., Tsai, J., Quackenbush, J., Griffiths, H.M., Restrepo, S., Smart, C.D., Fry, W.E., Van Der Hoeven, R., Tanksley, S., Zhang, P., Jin, H., Yamamoto, M.L., Baker, B.J., and Buell, C.R. (2003). Comparative analyses of potato expressed sequence tag libraries. Plant physiology **131,** 419-429.

Rubio, V., Shen, Y., Saijo, Y., Liu, Y., Gusmaroli, G., Dinesh-Kumar, S.P., and Deng, X.W. (2005). An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. Plant J **41,** 767-778.

Russinova, E., Borst, J.W., Kwaaitaal, M., Cano-Delgado, A., Yin, Y., Chory, J., and de Vries, S.C. (2004). Heterodimerization and endocytosis of Arabidopsis brassinosteroid receptors BRI1 and AtSERK3 (BAK1). The Plant cell **16,** 3216-3229.

Ryan, C.A., and Pearce, G. (2001). Polypeptide hormones. Plant physiology **125,** 65-68.

Ryan, C.A., Pearce, G., Scheer, J., and Moura, D.S. (2002). Polypeptide hormones. The Plant cell **14 Suppl,** S251-264.

Sachsenmaier, C., Sadowski, H.B., and Cooper, J.A. (1999). STAT activation by the PDGF receptor requires juxtamembrane phosphorylation sites but not Src tyrosine kinase activation. Oncogene **18,** 3583-3592.

Saitoh, M., Nishitoh, H., Fujii, M., Takeda, K., Tobiume, K., Sawada, Y., Kawabata, M., Miyazono, K., and Ichijo, H. (1998). Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. The EMBO journal **17,** 2596-2606.

Sawasaki, T., Ogasawara, T., Morishita, R., and Endo, Y. (2002). A cell-free protein synthesis system for high-throughput proteomics. Proc Natl Acad Sci U S A **99,** 14652-14657.

Schaller, G.E., and Bleecker, A.B. (1993). Receptor-like kinase activity in membranes of Arabidopsis thaliana. FEBS Lett **333,** 306-310.

Scheer, J.M., and Ryan, C.A. (1999). A 160-kD systemin receptor on the surface of lycopersicon peruvianum suspension-cultured cells. Plant Cell **11,** 1525-1536.

Scheer, J.M., and Ryan, C.A., Jr. (2002). The systemin receptor SR160 from Lycopersicon peruvianum is a member of the LRR receptor kinase family. Proc Natl Acad Sci U S A **99,** 9585-9590.

Schmidt, E.D.L., Guzzo, F., and Vries, M.A.J.T.S.C.d. (1997). A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent tp form embryos. Development **124,** 2049-2062.

Shah, K., Vervoort, J., and de Vries, S.C. (2001). Role of threonines in the Arabidopsis thaliana somatic embryogenesis receptor kinase 1 activation loop in phosphorylation. J Biol Chem **276,** 41263-41269.

Shah, K., Russinova, E., Gadella, T.W., Jr., Willemse, J., and De Vries, S.C. (2002). The Arabidopsis kinase-associated protein phosphatase controls internalization of the somatic embryogenesis receptor kinase 1. Genes & development **16,** 1707-1720.

Shiu, S.H., and Bleecker, A.B. (2001a). Plant receptor-like kinase gene family: diversity, function, and signaling. Sci STKE **2001,** RE22.

Shiu, S.H., and Bleecker, A.B. (2001b). Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc Natl Acad Sci U S A **98,** 10763-10768.

Shiu, S.H., and Bleecker, A.B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. Plant physiology **132,** 530-543.

Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F., and Li, W.H. (2004). Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. Plant Cell **16,** 1220-1234.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., Fauquet, C., and Ronald, P. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science **270,** 1804-1806.

Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K., and Parniske, M. (2002). A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature **417,** 959-962.

Szekeres, M. (2003). Brassinosteroid and systemin: two hormones perceived by the same receptor. Trends in plant science **8,** 102-104.

Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A., and Hinata, K. (2000). The S receptor kinase determines self-incompatibility in Brassica stigma. Nature **403,** 913-916.

Teis, D., Wunderlich, W., and Huber, L.A. (2002). Localization of the MP1-MAPK scaffold complex to endosomes is mediated by p14 and required for signal transduction. Developmental cell **3,** 803-814.

Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F., and Komeda, Y. (1996). The Arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. Plant Cell **8,** 735-746.

Trotochaud, A.E., Hao, T., Wu, G., Yang, Z., and Clark, S.E. (1999). The CLAVATA1 receptor-like kinase requires CLAVATA3 for its assembly into a signaling complex that includes KAPP and a Rho-related protein. The Plant cell **11,** 393-406.

Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G., and Tanksley, S. (2002). Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. The Plant cell **14,** 1441-1456.

van der Knaap, E., Song, W.Y., Ruan, D.L., Sauter, M., Ronald, P.C., and Kende, H. (1999). Expression of a gibberellin-induced leucine-rich repeat receptor-like protein kinase in deepwater rice and its interaction with kinase-associated protein phosphatase. Plant physiology **120,** 559-570.

Vanoosthuyse, V., Tichtinsky, G., Dumas, C., Gaude, T., and Cock, J.M. (2003). Interaction of calmodulin, a sorting nexin and kinase-associated protein phosphatase with the Brassica oleracea S locus receptor kinase. Plant physiology **133,** 919-929.

Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J **33,** 949-956.

Walker, J.C., and Zhang, R. (1990). Relationship of a putative receptor protein kinase from maize to the S-locus glycoproteins of Brassica. Nature **345,** 743-746.

Wang, X., and Chory, J. (2006). Brassinosteroids regulate dissociation of BKI1, a negative regulator of BRI1 signaling, from the plasma membrane. Science **313,** 1118-1122.

Wang, X., Li, X., Meisenhelder, J., Hunter, T., Yoshida, S., Asami, T., and Chory, J. (2005a). Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. Developmental cell **8,** 855-865.

Wang, X., Goshe, M.B., Soderblom, E.J., Phinney, B.S., Kuchar, J.A., Li, J., Asami, T., Yoshida, S., Huber, S.C., and Clouse, S.D. (2005b). Identification and functional analysis of in vivo phosphorylation sites of the Arabidopsis BRASSINOSTEROID-INSENSITIVE1 receptor kinase. Plant Cell **17,** 1685-1703.

Wang, Y., Pennock, S., Chen, X., and Wang, Z. (2002). Endosomal signaling of epidermal growth factor receptor stimulates signal transduction pathways leading to cell survival. Molecular and cellular biology **22,** 7279-7290.

Warren, C.M., and Landgraf, R. (2006). Signaling through ERBB receptors: multiple layers of diversity and control. Cellular signalling **18,** 923-933.

Wengier, D., Valsecchi, I., Cabanas, M.L., Tang, W.H., McCormick, S., and Muschietti, J. (2003). The receptor kinases LePRK1 and LePRK2 associate in pollen and when expressed in yeast, but dissociate in the presence of style extract. Proceedings of the National Academy of Sciences of the United States of America **100,** 6860-6865.

Wilde, A., Beattie, E.C., Lem, L., Riethof, D.A., Liu, S.H., Mobley, W.C., Soriano, P., and Brodsky, F.M. (1999). EGF receptor signaling stimulates SRC kinase phosphorylation of clathrin, influencing clathrin redistribution and EGF uptake. Cell **96,** 677-687.

Williams, R.W., Wilson, J.M., and Meyerowitz, E.M. (1997). A possible role for kinase-associated protein phosphatase in the Arabidopsis CLAVATA1 signaling pathway. Proc Natl Acad Sci U S A **94,** 10467-10472.

Yang, H., Matsubayashi, Y., Nakamura, K., and Sakagami, Y. (1999). Oryza sativa PSK gene encodes a precursor of phytosulfokine-alpha, a sulfated peptide growth factor found in plants. Proceedings of the National Academy of Sciences of the United States of America **96,** 13560-13565.

Yarden, Y., and Schlessinger, J. (1987). Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor. Biochemistry **26,** 1443-1451.

Annexe

Table S1. Ovule Receptor Kinases Real-Time RT-PCR primer list

RLK	PRIMER NAME	PRIMER SEQUENCE
ORK1	SV5_08G05f	AAAGGCCTACAATGAAAGATGCCCTTC
ORK1	SV5_08G05r	CAACGGTTCTTCTCATCTTTATGGAGCA
ORK2	SV5_09B07f	GTGGAGACTGCCTTAGCGGAATCAC
ORK2	SV5_09B07r	CCACTGAGGTTTCCACTGTTGGTTG
ORK3	SV5_9F11-f	ATCCCAAACCTTGCCGGACT
ORK3	SV5_9F11-r	TTCCGGCGAGAACCACAACT
ORK4	SV5_10G11f	AAAACTCCCTTTGACTTGGGTTCA
ORK4	SV5_10G11r	TCATCCACCGTAAAACATTCCATC
ORK5	SV5_40C08-f	CGCGCATACTGGGAATTCAGC
ORK5	SV5_40C08-r	TCGCGCTTCTTTGAGGTGCAT
ORK6	SV5_46H11f	GGGGTTTCCAGATTCCTTCAATTT
ORK6	SV5_46H11r	CATCCCTCTCCACTGACCTTGAG
ORK7	SV5_52D12f	TGCTGAAACTAGGAGTGGAAGTGG
ORK7	SV5_52D12r	CAACAATGCAAAGCCAAAGGTAAA
ORK8	SV5_55A01f	TTGCTGCTATTCCTTTGGCTCAT
ORK8	SV5_55A01r	CCCATTTCATCACAACATTGGAA
ORK9	SV6_10H08f	TGGATAGAAGAAAGGGCAATGCAA
ORK9	SV6_10H08r	GATGGTCTGCTAGATGCGTGTGTTC
ORK10	SV6_17D05f	CCATCAATGCGTCCGAGGAT
ORK10	SV6_17D05r	TCATGGACTGGTTCAAGCCTCA
ORK11	SV6_21C05-f	GCTGCCTCAAAGGATGGCAAGT
ORK11	SV6_21C05-r	CCTTCTTTCCCATCAACCCAAACA
ORK12	SV6_21E09f	CCAGGTTTTGATCATCTTGATGGAG

ORK12	SV6_21E09r	CCAGACGAGCAACAACTTCTTTCA
ORK13	SV6_21H05f	GCAGATTCAGGCATGGAGTAACC
ORK13	SV6_21H05r	CTGGCAGGGCACTGCATCTA
ORK14	SV6_23B07f	TCTCTTTCCCCCTCCCTTGC
ORK14	SV6_23B07r	GGATCATCAGATTTTCCAAAAGTGC
ORK15	SV6_23E05f	TCATGTCACCATGCCCGAGT
ORK15	SV6_23E05r	GGCTCCAAAAGCAAGGAGCA
ORK16	SV6_24B06f	CTGAGAAGCACGGCAAGTGG
ORK16	SV6_24B06r	TTGCAGCTCCTCCCCAAAAG
ORK17	SV6_26A08-f	GCAACCAAGAGTCCCCTGGTCA
ORK17	SV6_26A08-r	CTGAAACCTCATATAGCCAAAAGTGTTGC
ORK18	SV6_29A11f	TGGCCTCCCACATTCCATTT
ORK18	SV6_29A11r	CCAATGTGTTGAGTGCAGAGAGC
ORK19	SV6_30C11f	GGGAACCCCATCAGAGACCA
ORK19	SV6_30C11r	GCCTGCCACTTCTTAACTGCTTG
ORK20	SV6_30G09f	GGACATGGCAAGGGCAGAAG
ORK20	SV6_30G09r	TGGTGGTGTGAGTTCTTTCTCGTT
ORK21	SV6_32E08f	AAAACCGCGCCCATTTTCAT
ORK21	SV6_32E08r	CCTCACCGGAAATCAAATCCA
ORK22	SV6_34A04f	AGCAGGCAGCAAAGGTGACG
ORK22	SV6_34A04r	TTGCCTGTCACCAGCTCCAA
ORK23	SV6_36B12f	CCAGAGGCCAACCATGAAGG
ORK23	SV6_36B12r	TGCATACAAGTAAAAGCAGCTAAGAGG
ORK24	SV6_37C04f	GGGAAAGATGGTGATGGAGTGG
ORK24	SV6_37C04r	CCATAGCGCGTGAGATCAGGA
ORK25	SV6_39C12f	CCAGCCTGAGCCTGAGTTCC
ORK25	SV6_39C12r	CAACGCAATGCAACCAACAA
ORK26	SV6_39E01f	TGGGTTGATGGGAAAGAAGGAA

ORK26	SV6_39E01r	TGCAATAACTTTGGGTAGTTGGATGA
ORK27	SV6_48C03f	AGCACCATCACCAATTTACAGAACC
ORK27	SV6_48C03r	CCCCCACACAAAGGAAAATGA
ORK28	SV6_49A12-f	CTCGAGGGGCTCTATCCTGTTAAA
ORK28	SV6_49A12-r	GAGTACTCAACGTCGCATGGAAA
ORK29	SV6_49F04f	CCCGCGAAACGACCAAGTAT
ORK29	SV6_49F04r	TGCATTGTTGCCAATGCCTTT
ORK30	SV6_49G04f	GCCAACAATGCGACAAAACG
ORK30	SV6_49G04r	AAACCTCCGCGATGGTCAAA
Ubiquitin	UBQ-f	GCTGGCAAGCAGTTGGAAGAT
Ubiquitin	UBQ-r	TGGATGTTGTAGTCCGCCAGA

Table S2. Numbers and percentage of *Solanum chacoense* ORK per RLK classes compared with *Arabidopsis thaliana* RLKs.

	S. chacoense Ovule ESTs		*Arabidopsis* RLKs[1]	
RLK Class	Total number	%	Total number	%
LRR I	1	3,3	45	10,8
LRRIII	6	20,0	43	10,3
LRRV	4	13,3	11	2,6
LRRVI	4	13,3	11	2,6
LRRVII	1	3,3	10	2,4
LRRVIII-2	1	3,3	12	2,9
LRRIX	1	3,3	4	1,0
LRRX	1	3,3	15	3,6
LRRXI	3	10,0	28	6,7
LRRXII	1	3,3	9	2,2
LRRXIII	2	6,7	3	0,7
L-lectin	2	6,7	44	10,6
C-Lectin	1	3,3	1	0,2
Thaumatin	1	3,3	3	0,7
SD-1	1	3,3	32	7,7
Total	30	100	271	100

[1] Data taken from (Shiu et Bleecker, 2001b)

Table S3. Ovule Receptor Kinases EST accession numbers and their closest *Arabidopsis thaliana* match

Internal database identifier	db EST	ORK #	RLK family	*Arabidopsis* closest match
SV5_08G05	DN979658	ORK1	III	At5g67280
SV5_09B07	DN979684	ORK2	VI	At1g14390
SV5_9F11	DN979716	ORK3	III	At1g60630
SV5_10G11	DN979771	ORK4	XIII	At5g62230
SV5_40C08	DN977318	ORK5	V	At4g03390
SV5_46H11	DN977960	ORK6	I	At5g48740
SV5_52D12	DN978586	ORK7	III	At3g24660
SV5_55A01	DN978827	ORK8	XI	At5g61480
SV6_10H08	DN983490	ORK9	SD-1	At4g03230
SV6_17D05	DN983985	ORK10	C-Lectin	At1g52310
SV6_21C05	DN979962	ORK11	III	At3g17840
SV6_21E09	DN979990	ORK12	XII	At3g47110
SV6_21H05	DN980024	ORK13	V	At4g03390
SV6_23B07	DN980155	ORK14	III	At2g36570
SV6_23E05	DN980191	ORK15	XIII	At2g35620
SV6_24B06	DN980260	ORK16	X	At1g34420
SV6_26A08	DN980445	ORK17	VI	At3g03770
SV6_29A11	DN980737	ORK18	VI	At3g14350
SV6_30C11	DN980857	ORK19	IX	At2g01820
SV6_30G09	DN980903	ORK20	VIII-2	At3g09010
SV6_32E08	DR398320	ORK21	Thaumatin	At5g38280
SV6_34A04	DN981113	ORK22	XI	At4g28490
SV6_36B12	DN981325	ORK23	L-Lectin	At3g53380

SV6_37C04	DN981425	ORK24	VI	At4g18640
SV6_39C12	DN981625	ORK25	V	At1g53730
SV6_39E01	DN981638	ORK26	XI	At1g28440
SV6_48C03	DN982480	ORK27	III	At5g58300
SV6_49A12	DN982561	ORK28	V	At1g53730
SV6_49F04	DN982613	ORK29	VII	At2g24230
SV6_49G04	DN982625	ORK30	L-Lectin	At2g37710

QK
3
U54
2007
v.008