

Université de Montréal

Impact des changements climatiques et de la variabilité génétique sur le développement et la virulence du nématode à kyste du soya (*Heterodera glycines*)

Par Anne-Frédérique Gendron St-Marseille

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

Thèse présentée à la faculté des Arts et des Sciences
en vue de l'obtention du grade de doctorat
en sciences biologiques

Mai 2019

© Anne-Frédérique Gendron St-Marseille, 2019

Page d'identification du jury

Université de Montréal

Faculté des Études Supérieures et Postdoctorales

Cette thèse intitulée

Impact des changements climatiques et de la variabilité génétique sur le développement du
nématode à kyste du soya (*Heterodera glycines*)

Présentée par

Anne-Frédérique Gendron St-Marseille

A été évaluée par un jury formé des personnes suivantes

Marc St-Arnaud, Ph. D.
Président rapporteur

Jacques Brodeur, Ph. D.
Directeur de recherche

Benjamin Mimee, Ph. D.
Co-directeur

Simon Joly, Ph. D.
Membre du jury

David Joly, Ph. D.
Examineur externe

Christopher Bryant
Représentant du doyen

Résumé

Les invasions biologiques dans les agroécosystèmes engendrent de lourdes pertes économiques. Parmi les nombreuses espèces en cause, on retrouve les nématodes phytoparasites, vers microscopiques s'attaquant principalement aux racines. Présent dans tous les principaux pays producteurs de soya, le nématode à kyste du soya (NKS), *Heterodera glycines*, serait à lui seul responsable annuellement de plusieurs milliards de dollars de pertes. La rotation avec des cultivars résistants est le moyen le plus efficace de contrôler les populations de NKS, mais la surutilisation des mêmes lignées a conduit à la sélection d'individus virulents et mené à leur inefficacité. À ce jour, les mécanismes ainsi que les gènes de virulence associés au contournement de la résistance continuent de mystifier les scientifiques. Dans cette thèse, les effets des changements climatiques sur la reproduction et l'établissement du NKS ainsi que sur la phénologie de son hôte, le soya, ont été étudiés. Le premier modèle bioclimatique simulant le cycle de vie du NKS et du soya a été développé. Il a démontré que le nématode peut déjà se reproduire dans toutes les régions du Québec et que la hausse attendue des températures dans le futur proche (2041-2070) permettrait au NKS de pratiquement doubler le nombre de générations produites par saison de croissance dans toutes les régions. De plus, la production de soya issu du groupe de maturité I pourrait s'étendre à toutes les régions du Québec d'ici 2070. Une étude sur la distribution de la variabilité génétique entre 64 populations américaines et ontariennes et les gènes associés à diverses composantes bioclimatiques et leur rôle dans l'adaptation a également été réalisée. Celle-ci a révélé que la diversité génétique était très élevée entre les populations et qu'un flux de gène continu aurait facilité l'adaptation du NKS à diverses conditions bioclimatiques et son établissement dans toutes les régions nord-américaines où l'on produit du soya. Finalement, cette thèse présente l'analyse des génotypes du NKS et des gènes différenciellement exprimés sur des plants de soya résistant (Peking et PI88788) et sensible (Essex). En plus d'identifier plusieurs protéines liées à la virulence, cette étude a permis de mettre en évidence une région génomique sous forte pression évolutive. Cet îlot génique contient plusieurs répétitions en tandem qui ont divergé et dont certaines sont maintenant utilisées de façon sélective pour le contournement de différents types de résistance.

Mots clés : Nématode à kyste du soya, soya, modélisation bioclimatique, variabilité génétique, isolation par la distance, adaptation, sélection, séquençage de génome complet, transcriptome, gènes différemment exprimés.

Abstract

Biological invasions in agroecosystems are a major cause of economic losses. Plant parasitic nematodes are among the many species causing significant crop damages. The soybean cyst nematode (SCN) is causing billions of dollars of losses in all areas where soybean is produced. Rotation with resistant cultivars is the most effective mean of controlling SCN populations, but the overuse of the same lines has led to the selection of virulent individuals and the ineffectiveness of resistance. To this day, the virulence genes and mechanisms associated with the circumvention of resistance continue to mystify scientists. In this thesis, I explored the effects of climate change on the reproduction and establishment of SCN as well as on the phenology of its host, soybean. I have demonstrated that the nematode can already reproduce in all regions of Québec and that the expected rise in temperatures in the near future (2041-2070) will allow the development of more generations per growing season in all regions. In addition, I have demonstrated that the area suitable for the production of soybean from maturity group I will expand toward the north by 2070, further facilitating the expansion of SCN. I have also explored the genetic variability among more than 64 SCN populations from North America and analyzed the genes associated with various bioclimatic components and their role in adaptation. These analyses revealed that the genetic diversity was very high among SCN populations. This diversity associated with a continuous gene flow between populations has facilitated the adaptation of SCN to various bioclimatic conditions and its establishment in all US and Canadian soybean producing regions. Finally, this thesis presents an analysis of the SCN genotypes and the differentially expressed genes associated with virulence in two resistant soybean lines (Peking and PI88788) and susceptible Essex. This work has identified several proteins associated with virulence and allowed the discovery of a genomic region under strong evolutionary pressure. This island contains several genes in tandem duplications that have diverged and are now used selectively for overcoming different sources of resistance.

Keywords : Soybean cyst nematode, soybean, bioclimatic modeling, genetic variability, isolation by distance, adaptation, selection, whole genome sequencing, transcriptome, differentially expressed genes.

Table des matières

Page d'identification du jury	i
Résumé	ii
Abstract	iv
Liste des tableaux	vii
Liste des figures	viii
Liste des sigles et des abréviations.....	xi
Remerciements	xiv
Corps de l'ouvrage	xvi
Chapitre 1 Introduction générale.....	1
Chapitre 2 Revue de littérature	3
2.1 La plante hôte, le soya (<i>Glycine max</i>)	3
2.2 Le nématode à kyste du soya (<i>Heterodera glycines</i>)	7
2.3 Notions de génétique des populations	17
2.4 Signatures génétiques de l'adaptation	21
2.5 Hypothèses et objectifs.....	27
Chapitre 3 Simulating the impacts of climate change on soybean cyst nematode and the distribution of soybean	29
3.1 Abstract	30
3.2 Introduction	31
3.3 Material and Methods.....	32
3.4 Results	39
3.5 Discussion	44
3.6 Acknowledgements	49
3.7 Supplementary data	49
3.8 References	49
Chapitre 4 Genome Scans Reveal Homogenization and Local Adaptations in Populations of the Soybean Cyst Nematode	57
4.1 Abstract	58
4.2 Introduction	59
4.3 Materials and methods	62

4.4	Results	66
4.5	Discussion	81
4.6	Acknowledgements	86
4.7	Data Availability	87
4.8	Supplementary Materials.....	87
4.9	References	87
Chapitre 5 Genomic profiling of virulence determinants in the soybean cyst nematode (<i>Heterodera glycines</i>).....		96
5.1	Abstract	97
5.2	Introduction	97
5.3	Material and Methods.....	101
5.4	Results	106
5.5	Discussion	123
5.6	Conclusion.....	127
5.7	References	128
5.8	Supplementary material.....	135
Chapitre 6 Discussion générale.....		182
Références		193
Annexe 1 - Autres activités et accomplissements durant le doctorat		207
Annexe 2 - Autre article publié et soumis à un comité de lecture durant le doctorat		210

Liste des tableaux

Table 2.1 Résumé de la production et des superficies cultivées de soya des dix des principaux producteurs mondiaux en 2017 (FAOSTAT, 2019).....	4
Table 2.2 Numéro et nom des lignées utilisées pour effectuer les tests de HG type.....	13
Table 2.3 Classification du degré de différenciation génétique en fonction de la valeur de l'indice de fixation (<i>Fst</i>) selon Wright, 1978.	24
Table 3.1 List of ten climate models from the CMIP5 (2015) used in the construction of the RCP4.5 and RCP8.5 climate change simulations.	34
Table 4.1 Association of outlier loci (single-nucleotide polymorphisms [SNPs]) identified by at least two Bayesian genome scan methods in 64 populations of the soybean cyst nematode from North America with geographic or climatic co-variables.	72
Table 4.2 Identification and effect of outlier loci (single-nucleotide polymorphisms [SNPs]) identified by at least two Bayesian genome scan methods in 64 populations of the soybean cyst nematode from North America.	78
Table 4.3 Putative hitchhiking gene variants with high-impact neighbouring loci under selection and exhibiting a similar allele frequency in <i>Heterodera glycines</i>	79
Table 5.1 Name, origin and population virulence phenotype (HG type) of the populations used in the different experiments of this study.	101
Table 5.2 List of SNPs correlated with virulence on Peking and PI88788.	108
Table 5.3 Variants associated with important allelic frequencies changes or with a change toward homozygous variant associated with females selected on Peking.	110
Table 5.4 Variants associated with important allelic frequencies changes or with a change toward homozygous variant associated with females selected on PI88788.	112
Table 5.5 Gene uniquely absent/present on either cultivar using single-cell PCR.....	118

Liste des figures

Figure 2.1 Production (million de tonnes métriques) et superficies (millions d’hectares) de soya cultivé dans les provinces canadiennes de 1985 à 2017.....	5
Figure 2.2 Croissance relative de la superficie dédiée à la culture du soya au Canada et dans le monde depuis 1985.....	5
Figure 2.3 Représentation du cycle de vie complet du NKS.....	9
Figure 2.4 Différentes utilisations des indices de fixation (<i>Fst</i>) dans l’identification de marqueurs sous sélection et de gènes candidats.....	25
Figure 2.5 Représentations graphiques de quatre cas illustrant l’action des forces évolutives sur la différenciation génétique de sous-population séparées géographiquement.....	26
Figure 3.1 Study zone in red of Québec province representing the extent of the agricultural land.	33
Figure 3.2 Differences between future and reference planting dates for soybean MG-I in Québec, based on greenhouse gases concentration pathway simulations RCP 4.5 and 8.5.....	39
Figure 3.3 Soybean (MG-I) maturity dates (mean values) under the reference period and the 2050 horizon in Québec based on greenhouse gases concentration pathway RCP 4.5 and 8.5.....	40
Figure 3.4 Number of soybean cyst nematode generations completed at 20 cm depth during a single growing season in Québec for the reference period and the 2050 horizon using two RCP simulations.	41
Figure 3.5 Duration (days) and dates associated with the beginning and the end of each generation of soybean cyst nematode completed during a growing season for the reference period and the 2050 horizon at L’Acadie experimental station, Québec.	43
Figure 4.1 Genetic relationship among North American populations of the soybean cyst nematode.	68
Figure 4.2 Distribution of soybean cyst nematode populations used in this study (small dots) and identification of the “oldest” sample in the dataset (MO1, large dot).....	70
Figure 4.3 Number of outlier loci identified by three different Bayesian genome scan methods in 64 populations of the soybean cyst nematode from North America.	71

Figure 4.4 Allele frequencies of 15 outlier loci identified in North American populations of the soybean cyst nematode and associations of the loci with geographic or climatic co-variables. Black squares represent a significant association. 76

Figure 5.1 PCA of all 31408 DEG for all samples representing single female selected on cultivars with a resistant genotypes Peking and PI88788 and on a susceptible cultivar, Essex. .117

Figure 5.2 Highlight of *H. glycines* scaffold 589..... 119

Figure 5.3 Heat map depicting hierarchical clustering of the differentially expressed genes between selected female on Peking and on PI88788 in comparison with female selected on Essex 122

Figure 5.4 Differentially expressed genes between single *Heterodera glycines* females developing on PI88788 and Peking at J4 stage. 123

Liste des équations

Equation 1 $p^2 + q^2 + 2pq = 1$ 18
Equation 2 $Fst = (Ht - Hs)Ht$ 24
Equation 3 $IBD = Fst / (1 - Fst)$ 26
Equation 4 $Y = 0.0041x + 0.0087$ 37
Equation 5 $Y = 0.0008x^2 + 0.04x - 0.4106$ 37

Liste des sigles et des abréviations

AAFC – *Agriculture and Agri-Food Canada*

ADN – Acide desoxyribonucléique

AF – Allele frequency

AFLP – Polymorphismes de longueur des fragments amplifiés

ARN – Acide ribonucléique

BF - *Bayes factor*

bp – Paire de bases, *base pair*

CC – Changement climatique

CMIP5 - *Coupled Model Intercomparison Project Phase 5*

CNV – *Copy number variation*

CO₂ – Dioxyde de carbone, *carbon dioxide*

DEG – *Differentially expressed gene*, gène différentiellement exprimé

DNA – *Deoxyribonucleic acid*

ePBis - *Empirical Bayesian p-value*

FA – Fréquence allélique

FDR – *False discovery rate*

F_{ST} – Indice de fixation, *fixation index*

GBS – Génotypage par séquençage, *genotyping by sequencing*

GIEC - Groupe d'experts intergouvernemental sur l'évolution du climat

GIS – Système d'information géographique, *Geographic information systems*

GWAS - *genome wide association studies*

HG Type – Phénotype d'*Heterodera glycines*, *Heterodera glycines phenotype*

IAS - *Invasive alien species*

IBD – Isolation par la distance, *Isolation by geographical distance*

IPCC - *Intergovernmental Panel on Climate Change*

MCMC - *Markov chain Monte Carlo*

MG – Groupe de maturité, *maturity group*

Mha – Million d’hectare

MNV - *Multi-nucleotide variants*

Mt – Million de tonne

NCBI - *National Center for Biotechnology Information*

NGS - *Next-generation sequencing*

NKS – Nématode à kyste du soya

PCA - *Principal component analysis*

PCR – *Polymerase chain reaction*

PO - *Posterior probabilities*

Pool-seq - *Sequencing of pooled DNA samples*

ppm – Parties par million, *parts per million*

QTL – Locus de caractère quantitatif, *Quantitative Trait Loci*

RCP – *Representative concentration pathway*

RFLP – Polymorphismes de longueur des fragments de restrictions, *restriction fragment length polymorphism*

RNA – *Ribonucleic acid*

SMS – Syndrome de la mort subite

SCN – *Soybean Cyst Nematode*

SNP – Polymorphisme nucléotidique simple, *Single nucleotide polymorphism*

UNEAK - *Universal Network-Enabled Analysis Kit*

US / USA – *United States / United States of America*

Je dédie cette thèse

*À mon mari
pour son support tout au long de mes travaux*

*À mon fils
pour son amour inconditionnel et ses rires contagieux*



Remerciements

Les cinq (six) dernières années ont été incroyables. J'ai eu la chance de travailler dans un Centre de Recherche à Saint-Jean-sur-Richelieu vraiment accueillant et où tous les équipements requis à l'avancement de mes travaux étaient à portée de main. De plus, tous les gens que j'ai eu la chance de côtoyer au Centre de Recherche m'ont considérablement marquée et ont contribué dans un sens ou dans l'autre à la réussite de cette thèse.

D'abord, je tiens à remercier de tout cœur mes deux codirecteurs, Jacques Brodeur et Benjamin Mimee. Tous les deux m'ont aidé à me former en tant que future « super » docteure en biologie. Vous m'avez aidée à évoluer dans le monde de la recherche et vous m'avez permis de découvrir de nouveaux aspects de la biologie.

Aussi, je tiens particulièrement à remercier Benjamin Mimee pour m'avoir donné ma première chance en recherche déjà quand je terminais mes deux maîtrises et de m'avoir consacré beaucoup de temps afin que je relève de nombreux ... parfois trop nombreux défis. J'ai adoré évoluer à tes côtés, tu es quelqu'un d'inspirant, j'espère endosser ce rôle à mon tour un jour ... qui sait ?

Ensuite, je tiens à remercier tous les membres actuels et anciens du laboratoire de nématologie de Saint-Jean-sur-Richelieu pour leur aide et leur support : le trop hot Pierre-Yves Véronneau (merci pour la photo sexy), Nathalie Dauphinais (merci pour les discussions du lab 237), Maxime Gauthier (hein !) et bien sûr, ma belle Éléonore Tremblay (je t'aigine). Je tiens aussi à remercier l'actuel et les anciens stagiaires postdoctoraux : Anatoly Belov, Étienne Lord (of the ring), Joël Lafond-Lapalme et Marc-Olivier Duceppe. Bien entendu, merci à mes collègues et amis Michael (Mic) Sabeh, Mona (ma belle princesse perse) Parizadeh, Lydia (Lydz) Maheux, Sandra (Sandry) Poirier, Guillaume (Guillôme de poulet / Ti-Brin) Trépanier, Carolane (Caporale Carole) Audette et Dave (Dave) Thibouthot Ste-Croix !!! Un merci spécial à tous les défunts stagiaires, en particulier à Yuri Meys l'ensemble de son œuvre !!! Un merci à la gang de pichenotte ... Vous savez qui vous êtes !

Également, merci à Gaétan Bourgeois et à son équipe du laboratoire de Bioclimatologie et modélisation pour leur participation à mon article de modélisation. Finalement, merci aux gens des serres qui ont veillé sur mes plants et mes nématodes !!!

Durant mes études doctorales, j'ai reçu des bourses de plusieurs institutions qui m'ont permis, entre autres, de voyager. Je tiens à remercier : l'Université de Montréal (Bourses de la réussite étudiante, Bourse de Maternité, Bourse de fin d'études), Les Amis du Jardin Botanique (Bourse Pehr-Kalm, 1^{er} prix), L'Institut de Recherche en Biologie Végétale (Bourse Jacques Rousseau), le Centre Sève (deux Bourses (2015 et 2016) pour assister à un congrès), l'European Society of Nematology (ESN Student Bursary Award) et la Société de Protection des Plantes du Québec (Bourse W. E. Sackston).

Enfin, j'aimerais témoigner de ma profonde reconnaissance à ma famille, mes amis et mes amours. Merci à mon mari Marco, mon fils Émile, à ma mère Isabelle, à mon père Louis, à ma sœur Catherine, à ma sœur Marie-Claude, à mes nématodes (Éléonore, Lydia et Mona) pour m'avoir toujours encouragée, aimée et pour avoir égayé mes moments les plus sombres. Merci Jean-François, pour les lunchs gratuits à ton restaurant. Merci à mon grand-père et Huguette de m'avoir préparé des petits plats et m'avoir accueilli chez vous à de nombreuses reprises afin que je me concentre sur mes travaux d'écriture. Merci à mes beaux-parents Diane et Charles pour avoir préparé des petits plats à mon mari, pour m'avoir reçu chez vous afin que je puisse avancer mes projets et d'avoir gardé Émile. Merci à ma grand-mère et à Normand pour tous les mots d'encouragements qui m'ont fait du bien tout au long de mon doctorat. Merci aussi au reste de ma famille (Gilbert, Thérèse, Alex et Vincent) de m'avoir supporté de différentes manières. Si aujourd'hui je termine ce doctorat ... c'est bien grâce à vous tous !!!

Ce projet de recherche a été réalisé à l'aide du programme d'adjoint de recherche (PAR) du gouvernement du Canada. Également, ce projet a pu voir le jour grâce à l'allocation de budgets dédiés à la recherche sur le nématode à kystes du soya par le gouvernement du Canada. Plusieurs expériences ont également été rendues possible grâce au projet SoyaGen (soyagen.ca) financé par Génome Canada, Génome Québec, le Gouvernement du Canada, le Ministère de l'Économie, Science et Innovation du Québec, Semences Prograin Inc., Syngenta Canada Inc., Sevita Genetics, La Coop Fédérée, Grain Farmers of Ontario, Saskatchewan Pulse Growers, Manitoba Pulse & Soybean Growers, the Canadian Field Crop Research Alliance et Les Producteurs de grains du Québec. Finalement, tous les travaux ont été exécutés au centre de Recherche et de Développement de Saint-Jean-sur-Richelieu appartenant à Agriculture et Agroalimentaire Canada.

Corps de l'ouvrage

Chapitre 1. L'introduction générale illustre le contexte général de la thèse et présente les objectifs principaux.

Chapitre 2. Ce chapitre sert à orienter le lecteur sur les concepts abordés dans cette thèse par le biais d'une revue de littérature couvrant les principaux thèmes.

Chapitre 3. Le premier article scientifique intitulé : « Simulating the impacts of climate change on a root endoparasite and on its host » est présenté dans ce chapitre. Dans cette étude, le cycle de vie du nématode à kyste et la croissance de son hôte sont simulés dans un contexte de changements climatiques au Québec.

Chapitre 4. Le second article scientifique intitulé : « Genome scans reveal local adaptations and a continental homogenisation process in North American populations of the soybean cyst nematode *Heterodera glycines* » est présenté dans ce chapitre. Dans cette étude, des analyses génétiques de type génotypage par séquençage ont été effectuées afin d'évaluer la variabilité génétique retrouvée au sein des populations nord-américaines du nématode à kyste du soya. Également, les effets des variations climatiques locales, régionales et interrégionales sur la variabilité génétique intraspécifique et sur la sélection de gènes ont été quantifiés.

Chapitre 5. Le troisième article scientifique non soumis intitulé : « Genomic profiling of the virulence determinants of the soybean cyst nematode, *Heterodera glycines* » est présenté dans ce chapitre. Différentes approches génomiques et transcriptomique visant à identifier des gènes associés à la virulence y sont présentés.

Chapitre 6. Ce chapitre résume les principaux résultats de la thèse, commente leurs implications générales, présente mes principales contributions et suggère des orientations pour les recherches futures.

Annexe 1. Les autres activités menées durant le doctorat y sont décrites.

Annexe 2. Un article de type « review » publié dans la revue *Phytoprotection* y est présenté.

Chapitre 1 Introduction générale

Les invasions biologiques dans les agroécosystèmes sont les principales causes de pertes économiques (Paini *et al.*, 2016). De nombreuses espèces causent d'importants dommages aux cultures et à la production alimentaire parmi celles-ci on retrouve principalement des champignons, des bactéries, des virus, des insectes, de même que des nématodes phytoparasites (Hartman *et al.*, 2011). Les nématodes causant les plus importantes pertes en agriculture sont les endoparasites racinaires qui comprennent *Meloidogyne* spp., *Pratylenchus* spp., *Globodera* spp. et *Heterodera* spp. (Bernard *et al.*, 2017).

Le nématode à kyste du soya (NKS, *Heterodera glycines*) est un endoparasite sédentaire des racines et est l'une des principales causes de pertes économiques dans la production de soya dans tous les pays producteurs (Wrather *et al.*, 2010; Allen *et al.*, 2016, Allen *et al.*, 2017). Depuis la première identification du NKS en Caroline du Nord, aux États-Unis en 1954, sa présence a été notifiée dans presque tous les états producteurs ainsi qu'en Ontario et au Québec (Tylka & Marrett, 2017). Étant donné la popularité de cette culture et l'accroissement continu des superficies cultivées, sa progression dans de nouveaux lieux semble inévitable (United States Department of Agriculture (USDA), 2019). Jusqu'à présent, l'utilisation de cultivars possédants des génotypes résistants en rotation avec des cultures non-hôtes s'est avérée être le meilleur moyen de lutte limitant la reproduction de ce nématode (Niblack *et al.*, 2006; Liu *et al.*, 2017). Toutefois, on constate que les cultivars résistants disponibles sur le marché manquent cruellement de diversité. En effet, selon Mitchum (2016) près de 95 % des cultivars disponibles partagent les mêmes gènes descendant d'une source unique, PI88788. Par conséquent, leur utilisation en continu et s'étalant sur plusieurs décennies a favorisé l'émergence d'individus capables de contourner les mécanismes de défense donc, de populations toujours plus virulentes et incontrôlables (Tylka & Mullaney, 2015; McCarville *et al.*, 2017; Rincker *et al.*, 2017). À ce jour, les mécanismes expliquant le contournement de la résistance ainsi que les gènes qui expliquent la virulence sont toujours inconnus malgré les nombreuses études publiées (voir par exemple Bekal *et al.*, 2015; Gardner *et al.*, 2017; Masonbrinck *et al.*, 2019)

Dans cette thèse, je vais présenter de nouveaux aspects de la biologie du NKS ainsi que celle de son hôte économique principal, le soya. Également, je vais présenter de nouveaux éléments quant à sa diversité génétique et sa virulence. Je vais mettre l'accent sur sa reproduction et son

établissement dans un contexte de changements climatiques, mais également sur la diversité génétique retrouvée au sein des populations, son adaptation et les différences génétiques entre différentes populations virulentes. Pour y parvenir, j'ai adopté différentes approches telles que la modélisation bioclimatique, les statistiques associées à la génétique des populations de même que différentes plateformes de séquençage de nouvelle génération pour effectuer des études d'association. Les objectifs généraux étaient de :

- (1) Développer un modèle bioclimatique permettant de simuler les effets des changements climatiques sur la reproduction et l'établissement du NKS ainsi que sur la phénologie du soya au Québec.
- (2) Évaluer la diversité génétique retrouvée à travers une centaine de populations originaires des États-Unis et du Canada ainsi que trouver des gènes associés à son adaptation dans les différentes régions en fonction de facteurs bioclimatiques.
- (3) Effectuer le profilage génétique de populations et d'individus sélectionnés sur différents génotypes de plantes résistantes afin d'identifier des gènes associés à la virulence.

Chapitre 2 Revue de littérature

2.1 La plante hôte, le soya (*Glycine max*)

La culture du soya gagne en popularité à travers le monde notamment grâce à ses qualités nutritives. Cependant, son importation depuis l'Asie a favorisé l'introduction d'espèces invasives et souvent nuisibles. Dans ce chapitre, la biologie du soya, son importance économique et ses relations avec les parasites seront abordées.

La biologie de la plante

Le soya (*G. max* (L.) Merr.), une plante herbacée annuelle tétraploïde diploïdisée ($2n=40$) appartient à l'ordre des Fabales, famille des Légumineuses, sous-famille est Faboideae (ou Papilionoideae), tribu des Phaseoleae (Giller & Dashiell, 2007). Le soya appartient au genre *Glycine* qui est subdivisé en deux sous-genres, *Glycine* (espèces vivaces) et *Soja* (espèces annuelles) (Giller & Dashiell, 2007). Cette plante a un mode de reproduction autogame, donc l'anthère pollinise directement le stigmate au cours de sa croissance. Le soya est une plante nyctipériodique, donc qui croit de façon optimale dans les régions ayant des jours courts. Les écarts de températures optimales durant l'émergence, la floraison et l'atteinte de la maturité sont respectivement de 15-22°C, 20-25°C et de 15-22°C (Liu *et al.*, 2008).

Une des caractéristiques agronomiques d'importance pour la production du soya réfère au choix des différents groupes de maturité (MG pour *maturity group*) qui classifie les différents cultivars en fonction du différentiel de sensibilité à la photopériode (Cianzio, 2008). De façon générale, huit à dix jours séparent les cultivars les plus hâtifs des plus tardifs à l'intérieur d'un même groupe de maturité donc, environ 10 jours séparent chaque groupe (Cianzio, 2008; Mourtzinis & Conley, 2017). En Amérique, on classifie les cultivars de soya en treize groupes. Les groupes MG000 à MGII tolèrent les longues photopériodes caractéristiques des régions du nord des États-Unis et du sud du Canada (Burton, 1997; Hartman *et al.*, 2011). Les groupes MGIII à MGX évoluent mieux dans les régions ayant des photopériodes courtes, caractéristiques du sud des États-Unis et des régions tropicales (Burton, 1997).

La production mondiale et canadienne du soya

Les superficies de soya sont en forte augmentation à travers le monde, entre 2008 et 2018, les superficies cultivées ont augmenté de 32,7% et la production de 59,7% (United States Department of Agriculture (USDA), 2019). Cette hausse se justifie par les multiples utilisations des huiles et des farines riches en protéines issues de cette oléoprotéagineuse dans l'industrie alimentaire humaine et animale (Ali, 2010; Singh & Shivakumar, 2010). Le soya est l'une des dix cultures les plus importantes au monde en termes de production et de superficies cultivées et se situe au premier rang des cultures oléagineuses (FAOSTAT, 2019). Les principaux pays producteurs sont les États-Unis, le Brésil, l'Argentine, la Chine et l'Inde (FAOSTAT, 2019) (Table 2.1).

Table 2.1 Résumé de la production et des superficies cultivées de soya des dix des principaux producteurs mondiaux en 2017 (FAOSTAT, 2019).

Rang	Pays	Production (MT)*	Superficies cultivées (Mha)**
1	États-Unis	119,5	36,2
2	Brésil	114,6	33,9
3	Argentine	55,0	17,3
4	Chine	13,1	7,3
5	Inde	11,0	10,6
6	Paraguay	10,5	3,4
7	Canada	7,7	2,6
8	Ukraine	3,9	2,0
9	Russie	3,6	2,6
10	Bolivie	3,0	1,3

* Million de tonnes.

** Million d'hectares.

En 2017, le Canada occupait respectivement la huitième et la cinquième position mondiale au niveau de la production et de l'exportation (FAOSTAT, 2019). Au Canada, en 2017, la production de soya se concentrait dans les provinces de l'Ontario (52,2 %), du Manitoba (27,4 %), du Québec (16,1 %), de la Saskatchewan (3,1 %) et des Maritimes (1,2 %) (Statistiques Canada, 2019). D'ailleurs, depuis le début des années 2000, la production au Manitoba a plus que triplé et ne cesse de s'accroître grâce au développement de semences aux qualités agronomiques adaptées à différentes photopériodes (Cober & Morrison, 2010; StatCan, 2019) (figure 2.1).

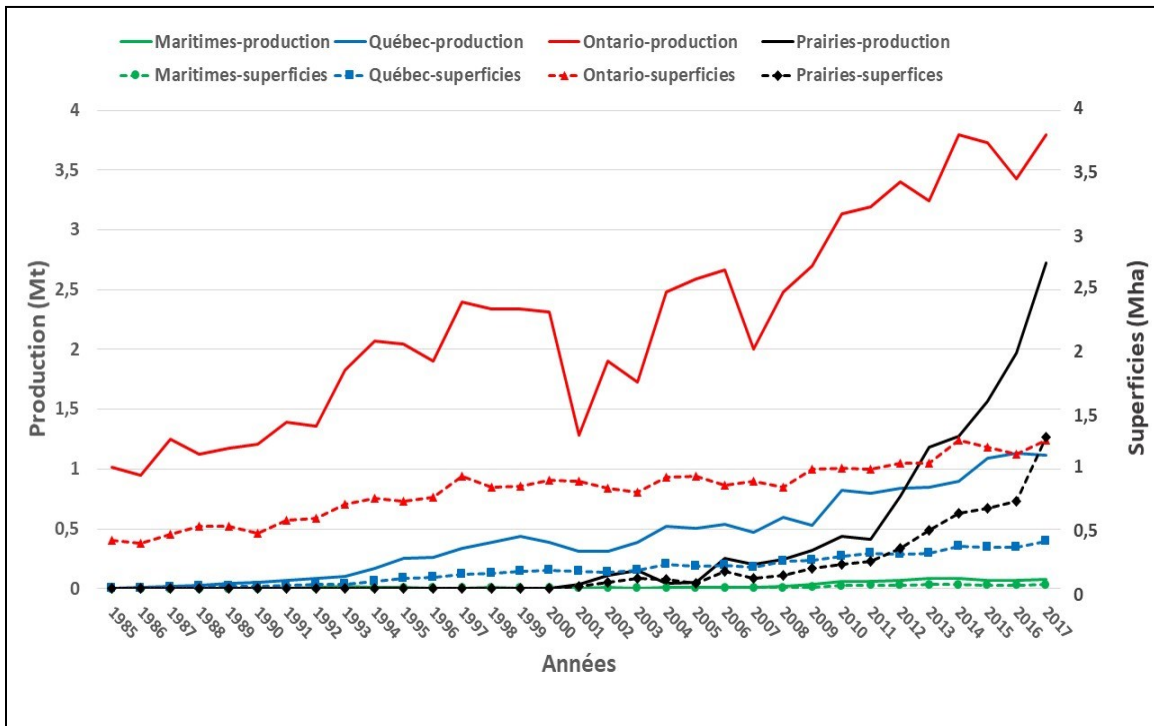


Figure 2.1 Production (million de tonnes métriques) et superficies (millions d’hectares) de soya cultivé dans les provinces canadiennes de 1985 à 2017.

En l’espace de trente ans, les superficies dédiées à la production de soya au Manitoba ont rejoint celles de l’Ontario, et ce, alors que cette province ne s’est réellement intéressée à la culture de cette plante qu’au début des années 2000 (figure 2.1). En outre, tel qu’illustré dans la figure 2.2, la superficie dédiée à la culture du soya a augmenté en l’espace de trente ans beaucoup plus rapidement au Canada que dans le reste du monde (FAOSTAT, 2019; StatCan, 2019).

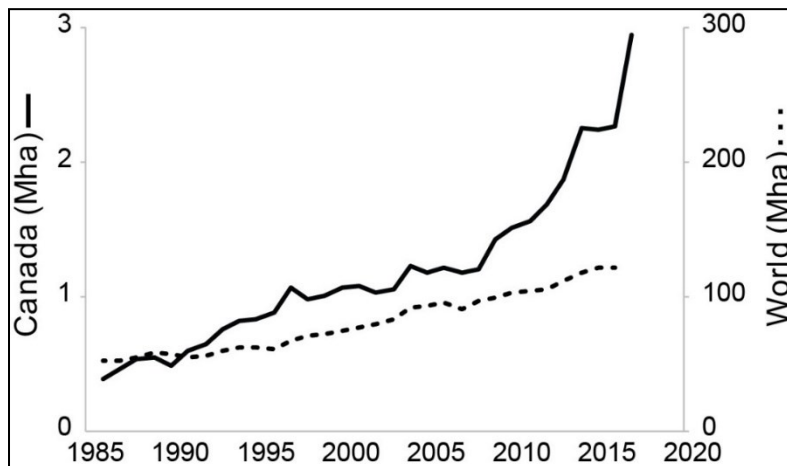


Figure 2.2 Croissance relative de la superficie dédiée à la culture du soya au Canada et dans le monde depuis 1985.

En effet, durant cette période de temps, les superficies cultivées de soya ont sextuplé et sont passées de 400 000 hectares à 2,8 Mha comparativement aux superficies mondiales qui ont pratiquement doublé. Suivant cette tendance, cette culture continuera à gagner en importance dans les prochaines années, d'où la nécessité de réduire les éléments limitant la production (Pagano & Miransani, 2016).

Principaux ravageurs du soya au Canada et aux États-Unis

L'accroissement des échanges commerciaux et l'adoption de nouvelles cultures favorisent l'introduction, l'établissement et la reproduction de certains ravageurs et maladies exotiques et indigènes (Hartman *et al.*, 2011). Dans les principaux pays producteurs de soya, les pertes économiques les plus importantes sont associées à la présence des champignons causant la rouille *Phakospora pachyrhizi* et *P. meibomia*, suivi par le nématode à kyste du soya (NKS) *Heterodera glycines* (Anderson *et al.*, 2004; Hulme, 2009; Wrather *et al.*, 2010).

En 2015, aux États-Unis, les pertes dues aux ravageurs et maladies s'élevaient à 12,5 Mt de soya sur une production totale de 106,9 Mt alors qu'en Ontario ces pertes s'élevaient à 624 000 tonnes de soya sur les 6 Mt produits (Allen *et al.*, 2017; FAOSTAT, 2019). C'est dans les états centraux et du nord des États-Unis où est produit la majorité du soya (78 %) que l'on enregistre les plus importantes pertes de rendements (9,5 % des pertes totales) attribuables aux ravageurs et aux maladies (Bradley *et al.*, 2017). Dans cette partie des États-Unis, les cinq principaux ravageurs et maladies sont: le NKS, les maladies des semis, le syndrome de la mort subite (SMS, *Fusarium virguliforme* anciennement *F. solani* f. sp. *glycines*), la sclérotiniose (*Sclerotinia sclerotiorum*) et la tache brune (*Septoria glycines*) (Bradley *et al.*, 2017). Dans les états du sud des États-Unis, les pertes étaient principalement imputables au NKS, à la cercosporose (*Cercospora sojina*), aux nématodes réniformes, aux maladies de semis et à la pourriture charbonneuse (*Macrophomina phaseolina*) (Bradley *et al.*, 2017). Finalement, en Ontario, les principales maladies et ravageurs sont le NKS, les maladies de semis, la sclérotiniose, le SMS et la pourriture phytophthoréenne (*Phytophthora sojae*) (Bradley *et al.*, 2017). Dans tous les cas, le NKS constitue la plus importante source de perte de rendement dans le soya, 3 Mt de pertes aux États-Unis et au Canada (Ontario).

2.2 Le nématode à kyste du soya (*Heterodera glycines*)

Tel que vu précédemment, plusieurs espèces invasives limitent la production de soya, dont le NKS. Afin de bien cerner la relation hôte-parasite entre le soya et le NKS, ce chapitre décrit l'historique, la biologie et les éléments inhérents à la virulence de ce nématode.

Origine et distribution

Le NKS est un endoparasite obligatoire des racines originaire d'Asie qui a été introduit dans la majorité des pays producteurs de soya (Wrather *et al.*, 2010). Cette espèce aurait coévolué avec son principal hôte économique, le soya, en Chine (Yu, 2011). Le nématode est très répandu en Asie, en effet en plus de la Chine on le retrouve à Taïwan, en Corée du Nord et du Sud, au Japon, en Mongolie, en Inde, en Iran, en Indonésie (Lal & Lal, 2005; Subbotin *et al.*, 2010; CABI, 2019). En Europe, on le retrouve en Russie et en Italie et en Afrique, en Égypte (Riggs, 1977; CABI, 2019).

En Amérique du Nord, il a été détecté pour la première fois en Caroline du Nord en 1954 dans l'Est des États-Unis et en 1987 au Canada, dans le comté de Kent situé au sud-est de l'Ontario (Riggs, 1977; Anderson & Welacky, 1988; Yu, 2011; Tylka & Marett, 2014). Ce n'est qu'en 2013 que le NKS a été détecté au Québec (Mimee *et al.*, 2014b). En Amérique du Sud, on aurait identifié le NKS en Colombie en 1982 (Gómez-Tovar & Medina, 1983). Par la suite, il a été identifié à l'Équateur, au Paraguay, en Bolivie, au Brésil et en Argentine (CABI, 2019).

Principaux symptômes et impact économique du NKS

Les dommages causés aux plants de soja par le NKS sont imputables à la création de sites de nutrition (syncytium) au niveau des racines qui réduisent le transport des éléments nutritifs et de l'eau vers les parties aériennes (Agrios, 2005) (figure 2.3, c).

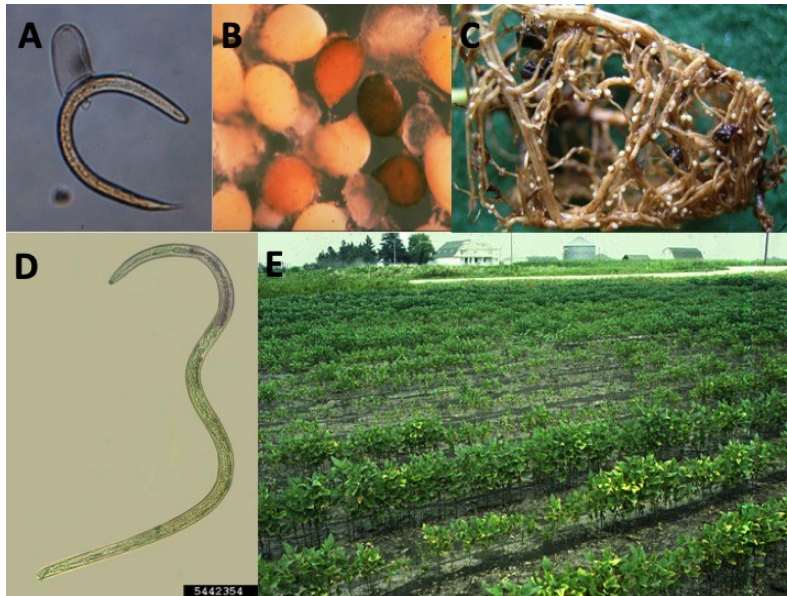


Figure 2.3 Photographies des stades du NKS et des symptômes en champ.
 A) Juvénile (J2); B) Femelles adultes blanches (J4) et kystes bruns; C) Kystes sur des racines; D) Mâle adulte (J4); E) Symptômes en champ.
 Sources : A) et B) Nelson & Bradley, 2003; C) Faghihi & Ferris, 2017; D) Eisenback, 2018; E) Davis & Tylka, 2000.

En général, les plants sévèrement infectés présentent des feuilles chlorotiques et jaunâtres, sont rabougris avec canopée ouverte entre les rangs, ce qui entraîne une production (taille des gousses) fortement limitée (Wang *et al.*, 2003; Niblack, 2005) (figure 2.3, e). Ce parasite peut réduire de 15 à 85 % le rendement des cultures (Niblack, 2005). À l'échelle mondiale, les pertes économiques associées à ce nématode sont estimées à plus d'un milliard de dollars américains annuellement (Allen *et al.*, 2017).

Biologie et cycle de vie de l'espèce

Du point de vue morphologique, une des caractéristiques qui permet de facilement différencier le mâle de la femelle après l'atteinte du stade adulte est le dimorphisme sexuel. Suivant la sortie de l'œuf, l'apparence du nématode reste indifférenciée et il est vermiforme (Figure 2.3, a). C'est à l'atteinte du stade du stade adulte que l'on différencie le mâle qui reste vermiforme et la femelle qui prend une forme de citron (Figure 2.3, b et d).

Plusieurs aspects de la biologie du NKS lui permettent d'être un parasite efficace et de causer de lourds dommages à son hôte. D'une part, le spectre d'hôtes est important et inclut plus d'une

centaine d'hôtes alternatifs tels que des mauvaises herbes pérennes ou annuelles (par ex. *Lamium purpureum* ou *Stellaria media*) qui permet aux populations de se reproduire et de s'établir même en absence de son hôte économique, le soya (Venkatesh *et al.*, 2000; Creech *et al.*, 2007; Yu, 2011). D'autre part, même sous conditions climatiques adverses la survie est élevée même en absence d'hôtes (Perry & Moens, 2011). Plusieurs facteurs expliquent la résilience du NKS, notons sa capacité à entrer en quiescence ou en diapause et la protection conférée par la cuticule du kyste qui renferme des centaines d'œufs, eux-mêmes protégés par plusieurs membranes (Schmitt, 2004; Perry & Moens, 2011).

Le cycle de vie du nématode débute à l'intérieur de l'œuf, où le premier stade larvaire (J1, absent de la figure) subit une première mue pour atteindre le second stade juvénile (J2) qui est la forme infectieuse (Koenning, 2004) (figure 2.4).

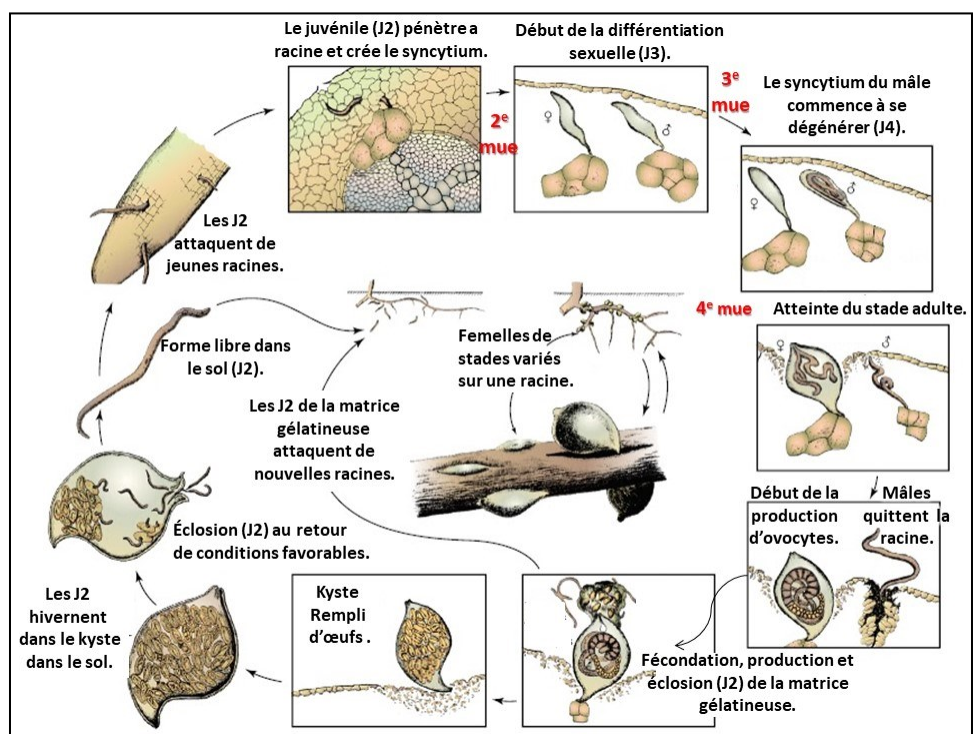


Figure 2.4 Représentation du cycle de vie complet du NKS. Traduit et modifié d'Agrios (2005), p. 846.

Cette larve quitte l'œuf suite à l'écllosion puis, à l'aide d'un stylet, s'introduit dans la racine de l'hôte, choisit une cellule des tissus vasculaires et crée un site d'alimentation appelé syncytium (Niblack, 2005; Niblack *et al.*, 2006). Cette structure produira les ressources essentielles à son

développement au détriment de la plante hôte (Noel, 2004). Par la suite, trois autres mues se succèdent à l'intérieur de la racine jusqu'à l'atteinte de la maturité sexuelle (Noel, 2004). Le mâle, qui demeure vermiforme, arrête de s'alimenter rapidement et sort de la racine pour féconder la femelle. Plusieurs mâles peuvent féconder une femelle donc la variabilité génétique au sein d'une fratrie peut être extrêmement élevée (Lauritis *et al.*, 1983). Celle-ci prend graduellement une forme de citron dont la majeure partie est exposée à l'extérieur de la racine (Noel, 2004). Une fois fécondée, la femelle expulse une partie des œufs dans une matrice gélatineuse (Noel, 2004) où ils sont prêts à éclore et à infecter la racine. Une autre partie des œufs restent enkystés et éclore lorsque des conditions spécifiques seront atteintes (Noel, 2004). Ils peuvent survivre ainsi une dizaine d'années dans le sol (Slack *et al.*, 1972). D'ailleurs, comme les œufs d'un kyste peuvent éclore de façon différentielle dans le temps, il est possible qu'un « ancêtre » féconde une génération actuelle ce qui favorise hautement la préservation d'une importante diversité génétique (Koenning, 2004).

Établissement et adaptation

C'est sous forme de kyste que le NKS est dispersé par divers vecteurs biotiques (humains, animaux) ou abiotiques (vent, eau), et ce, autant à l'échelle régionale qu'à celle d'un champ (Dong & Opperman, 1997; Koenning, 2004; Niblack, 2005). L'établissement du NKS dans pratiquement tous les pays producteurs de soya, peu importe leurs latitudes, démontre que cet organisme s'adapte à une large gamme d'environnements et de conditions climatiques. La température a un effet important sur le cycle de vie de cet organisme poïkilotherme. Il peut compléter un cycle en trois semaines entre 22 et 30°C et en quatre semaines entre 16 et 22°C (Niblack, 2005; Oyekanmi & Fawole, 2010). Aux États-Unis et au Canada, les températures permettent au nématode de compléter de deux à cinq générations durant une saison de croissance (Lauritis *et al.*, 1983). Dans l'hémisphère sud tel qu'en Argentine, on a dénombré jusqu'à sept générations durant une saison de croissance (Burrows & Stone, 1985). Toujours en Argentine, on a constaté durant une saison plus chaude qu'à la normale que la croissance spontanée de plant de soya a permis la production d'une génération supplémentaire de NKS, et ce, même si en juillet les températures ont été situées entre 6-10°C (Doucet *et al.*, 2008). Ainsi, les conditions climatiques influencent directement la dynamique des populations de NKS et, par conséquent, le rendement de la culture (Grulke, 2011; Kakaïre, 2012; Pautasso, 2012; Ahanger, 2013). Donc, il

s'avère très pertinent de se questionner sur l'impact qu'auront les fluctuations de températures attendues à la faveur des changements climatiques (CC) sur la biologie du soya, du NKS et de leur interaction. Aussi, il demeure important de s'interroger sur l'impact de l'environnement sur son adaptation à divers milieux naturels suivant leur introduction.

Impacts des changements climatiques sur la distribution et la biologie du NKS et du soya

L'analyse de la distribution du NKS entre 1954 et 2017 aux États-Unis a démontré une dispersion rapide et un taux d'établissement très élevé dans les régions productrices du centre et du nord des États-Unis (Chakraborty & Datta, 2003; Tylka & Marrett, 2017). Ce phénomène a probablement été accentué par l'adoption de nouvelles zones de production à la faveur des fluctuations du climat (Somasekhar & Prasad, 2012). De même, la modélisation de la phénologie du soya à l'intérieur de la zone de potentiel agricole du Québec, reposant sur les paramètres climatiques du passé récent (1971-2000) et du futur proche (2041-2070), a démontré que des latitudes plus nordiques offrent de nouvelles zones de culture du soya (Gendron St-Marseille, 2013). Également, la modélisation du cycle de vie du NKS en fonction de paramètres climatiques a démontré qu'il pouvait dès lors s'établir à l'intérieur des régions agricoles où le soya se cultive actuellement (Dong & Opperman, 1997; Koenning, 2004; Niblack, 2005) et qu'il possède la capacité de suivre le déplacement de son hôte vers le nord (Gendron St-Marseille, 2013).

Également, la hausse prévue des températures favorisera le taux de développement du NKS, ce qui se traduirait par l'accumulation d'une à deux générations supplémentaires pour atteindre jusqu'à deux générations par saison de croissance dans la zone la plus septentrionale du Québec et jusqu'à six générations dans celle située à l'extrême sud (Gendron St-Marseille, 2013). Ce scénario aura pour effet d'affaiblir davantage le plant de soya, car les surinfections augmentent d'une part la sensibilité des racines aux autres pathogènes et d'autre part, limite l'acheminement des nutriments et de l'eau vers les parties aériennes (Gendron St-Marseille, 2013).

En contrepartie, des modèles prédisent que la hausse des températures, couplée à celle des niveaux de CO₂ atmosphériques auront un effet positif sur la croissance, la densité et la longueur des racines du soya (Lawn & Noel, 1986; Niblack *et al.*, 2006). Toutefois, le réchauffement des températures pourrait affecter de façon plus prononcée les impacts négatifs des premières générations de NKS sur le rendement des cultures puisque l'augmentation des surfaces racinaires

permettra à un plus grand nombre d'individus de se développer (Rogers *et al.*, 1992; Ziska, 1998; Wang *et al.*, 2003). Également, les événements climatiques extrêmes, dont la sécheresse, semblent réduire les rendements du soya de manière plus prononcée dans les champs infestés par le NKS (Tylka, 2012). Ce stress entraîne le nématode à créer son site de nutrition plus profondément dans les tissus vasculaires de la stèle et non dans la zone du cortex des racines, ce qui lui permettrait de continuer d'exploiter les ressources de la plante (Bonner & Schmitt, 1985; Tylka, 2012). Par conséquent, des fréquences plus élevées de sécheresse devraient avoir d'importantes répercussions sur la production de soya.

Le phénotypage des populations de nématode en fonction de leur virulence

Actuellement, le contrôle des populations de nématodes repose sur l'utilisation de cultivars de soya possédant des gènes de résistance. Cependant, aucun cultivar n'est connu pour être capable de limiter la reproduction de toutes les populations de NKS (Elad & Pertot, 2014). En se basant sur la capacité des différentes populations de NKS à se reproduire sur différents génotypes de résistance de cultivars de soya, Niblack *et al* (2002) ont créé un test, *HG Type Test* (pour *Heterodera glycines* phénotype). Ce test permet d'approximer sur la base d'échantillons de sol le ou les types de virulence retrouvés dans les populations. Pour ce faire, on inocule sept lignées indicatrices résistantes de soya avec les kystes d'une population donnée et un plant témoin sensible à tous les HG types (table 2.2).

Table 2.2 Numéro et nom des lignées utilisées pour effectuer les tests de HG type

Type	Numéro de la lignée	Nom de la lignée
	1	PI548402 (Peking)
	2	PI88788
	3	PI90763
Standards résistants	4	PI437654
	5	PI209332
	6	PI89772
	7	PI548316 (Cloud)
Standards sensibles	-	Essex (PI548667)
	-	Lee 74 (PI548658)

Après 30 jours, on compte le nombre de femelles produites sur les racines des plants résistants et on divise par le nombre produit sur le plant sensible. Si la proportion de kystes produits sur des cultivars résistants est au moins supérieur de 10 % en comparaison avec le nombre produit sur les cultivars sensibles (témoins), cette population sera considérée comme virulente. Par exemple, une population peut être classifiée comme 1.2.5.7 si les individus sont capables de se reproduire sur les lignées résistantes #1, #2, #5 et #7.

La distribution des hg types (anciennement connu pour être des races classées de 1 à 16, mais qui ne sont plus utilisées de nos jours) est assez différente entre le Sud et la Nord du 37^{ième} parallèle aux États-Unis. En effet, dans une étude de 1997 (Kim *et al.*, 1997), on retrouvait une prédominance des races 2, 3, 4, 5, 6, 9 et 14 (équivalent des hg types 1.2- ; 0- ; 1.2.3- ; 1- ; 1.3-) alors qu'au nord du 37^{ième} parallèle on retrouvait uniquement les races 3 et 4 (équivalent des hg types 0- et 1.2.3-). Après plus d'une décennie, on retrouvait au sud du 37^{ième} une prédominance des races 1, 2, 4, 5, 10, 14, 15, 16 (ou HG types 1.2.3.5.6.7; 1.2.3.5.7 ; 1.2.5.7 ; 1.3.5.6.7 ; 1.2.5.6.7 ; 2.5.7 ; 7) et au Nord celle des races 1 et 3 (ou HG types 2 ; 2.5 ; 2.5.7 ; 2.7 ; 5 ; 5.7 ; 7 ; 0) (Faghihi *et al.*, 2010). L'étude de Faghihi *et al.* (2010) incluait une province canadienne, l'Ontario et a démontré qu'il y avait une incroyable diversité et qui semble correspondre aux races/HG types retrouvés autant au Nord qu'au sud des États-Unis en plus d'inclure nombreux HG types (12 hg types). En effet, on retrouve toutes les races (en considérant que le cultivar

Pickett et Peking sont identiques tel qu'indiqué par Niblack *et al.*, 2002) soit plus de 24 HG types différents (Faghihi *et al.*, 2010). Une aussi grande diversité de HG types indique que l'adaptation permettant de contourner les mécanismes de résistance de PI 88788 est peut-être en cours, mais pas encore aussi importante qu'au sud du 37^{ième} parallèle où cette lignée résistante est utilisée depuis les années 1960 (Hartwig *et al.*, 1971 ; Faghihi *et al.*, 2010). À l'heure actuelle, plus de 90 % des semences commerciales résistantes sont issues de la même lignée parentale, soit PI88788 (lignée #2) et dans une plus faible proportion, Peking (Concibido *et al.*, 2004). Aux vues du nombre limité de sources de résistance utilisées dans les cultivars commerciaux disponibles sur le marché, il est primordial de comprendre les mécanismes permettant au nématode de contourner les gènes de résistance de la plante.

Les gènes de résistance au nématode à kyste du soya

La variabilité phénotypique du caractère de résistance s'expliquerait à la fois par la présence de différents gènes du parasitisme produisant des protéines associées à la virulence, ou effecteurs, chez le nématode et de gènes de résistance chez la plante (Hogenhout *et al.*, 2009). La résistance du soya est polygénique et près de 70 loci de caractères quantitatifs (QTL) associés à la résistance ont été identifiés chez le soya, les plus étudiés étant Rhg1, Rhg2 et Rhg3 (leur gène majeur est récessif) et Rhg4 et Rhg5 (leur gène majeur est dominant) (Caldwell *et al.*, 1960; Matson et Williams, 1965; Rao-Arelli, 1994; Klink *et al.*, 2009). Deux QTL sont plus importants dans la réaction de résistance, Rhg1 (chromosome 18), qui compte environ neuf gènes et Rhg4 (chromosome 8) qui comptent au moins sept gènes (Klink *et al.*, 2013). On a identifié trois gènes du QTL Rhg1 de PI88788 qui encodent une protéine α -SNAP, un potentiel transporteur d'acides aminés (Cook *et al.*, 2012). Également, un des gènes du QTL Rhg4 encode une enzyme (serine hydroxymethyltransferase) impliquée dans la régulation du métabolisme du carbone et des folates (Liu *et al.*, 2012). Les QTL Rhg1 et Rhg4, bien connus chez PI 88788 et Peking respectivement, provoquent des réponses cytologiques différentes et permettent de distinguer deux groupes de liaison génétique (*linkage group*) (Klink *et al.*, 2013). Le type de résistance appelé « PI88788 » repose principalement sur les gènes situés sur l'allèle rhg1-b du QTL Rhg1, qui ont un comportement dominant chez les cultivars de ce groupe (PI88788, PI209332 et PI549316) (Kim *et al.*, 2010; Cook *et al.*, 2012; Liu *et al.*, 2012; Yuan *et al.*, 2012). La réponse cytologique du groupe PI88788 (regroupant le cultivar du même nom) se caractérise par une nécrose assez lente

(entre 8 à 10 jours) du syncytium et par l'absence d'accroissement de la paroi cellulaire (cell wall apposition) (Liu *et al.*, 2012, Klink *et al.*, 2013). La résistance de type « Peking » requiert la présence simultanée de gènes co-dominants situés sur les QTL Rhg1 et Rhg4 (Kim *et al.*, 2010, Liu *et al.*, 2012, Yuan *et al.*, 2012). Typiquement, la réponse cytologique associée au groupe Peking (Peking, PI89772 et PI90763) se caractérise par une nécrose rapide (environ 48 heures) de la cellule initiale et par l'accroissement de la paroi cellulaire (Cook *et al.*, 2012; Cook *et al.*, 2014). Il a été suggéré que la résistance des plantes est conférée par le nombre de copies de l'allèle rhg1-b associé au QTL Rhg1. En effet, les lignées et cultivars associés au type PI88788 possèdent de sept à dix copies des gènes situés sur le QTL Rhg1, ceux situés sur le même QTL au sein du groupe de type Peking en possèdent moins de trois copies (Brucker *et al.*, 2005; Liu *et al.*, 2012). La lignée indicatrice PI437654 ne se retrouve dans aucun des deux types de résistance puisqu'une majorité de NKS est incapable de se reproduire sur celle-ci et pourrait être une bonne alternative pour le contrôle des nématodes. Cependant, comme les individus capables de se reproduire sur cette lignée parviennent aussi à se reproduire sur les types PI88788 et Peking, l'utilisation de ce cultivar n'est pas recommandée dans un programme de rotation utilisant ces sources de résistance (Gardner *et al.*, 2018).

Les protéines de virulence

Une très grande variabilité génétique se rencontre à l'intérieur des différentes populations de NKS et plusieurs individus seraient porteurs d'allèles « *ror* » (reproduction on a resistant host) favorisant le contournement de la résistance de l'hôte et la capacité à se reproduire du parasite (Zheng *et al.*, 2006; Cook *et al.*, 2012; Wang *et al.*, 2014). Les mécanismes adaptatifs qui mènent au contournement de la résistance demeurent inconnus, mais avec différentes stratégies de croisement génétique, on a identifié deux gènes (non caractérisés) qui pourraient être associés à la virulence (Dong & Opperman, 1997) : *ror-1* serait dominant et permettrait la reproduction sur PI 88788 et *ror-2* serait récessif et permettrait la reproduction sur PI90763 (Dong & Opperman, 1997).

À ce jour, on a identifié plus de 80 effecteurs candidats sécrétés par les glandes œsophagiennes (deux sous-ventrales et une dorsale) et injectées à l'aide du stylet, qui permettent au NKS d'altérer la structure et les fonctions de la cellule hôte (Li *et al.*, 1996; Gao *et al.*, 2001, 2003; Wang *et al.*, 2001; Bekal *et al.*, 2008; Klink *et al.*, 2009; Haegeman *et al.*, 2012; Wang *et al.*,

2014; Noon *et al.*, 2015). En outre, on a aussi identifié chez d'autres nématodes des protéines sécrétées au niveau de l'hypoderme (glutathion peroxydase qui contre les dérivés réactifs de l'oxygène) et de la cuticule (Gp-FAR-1 qui pourrait inhiber l'acide jasmonique) qui permettent au nématode de se camoufler et inhibent certains mécanismes de défense systémique (Prior *et al.*, 2001; Jones *et al.*, 2004). Différents effecteurs sont utilisés selon le stade de développement du nématode. Durant la phase migratoire du J2, la plupart des protéines sécrétées visent à la fois à dissoudre les cellules de la racine et à esquiver la détection par l'hôte en inhibant ou supprimant les hormones associées aux voies métaboliques et de signalisation (Haegeman *et al.*, 2012). Durant la phase migratoire, les glandes subventrales synthétisent différentes enzymes et protéines (cellulases, Hg19C07, β -1,4-endoglucanases, pectate lyases, endo-1,4- β -galactosidases, expansines, etc.) qui une fois injectées par le stylet favorisent l'invasion (de Boers *et al.*, 1999; Wang *et al.*, 1999; de Boers *et al.*, 2002; Qin *et al.*, 2004; Vanholme *et al.*, 2006; Lee *et al.*, 2011).

Suite à la sélection d'une cellule initiale destinée à servir de syncytium, le J2 injecte différents effecteurs qui favorisent son développement jusqu'à l'atteinte du stade adulte. Ces effecteurs (annexines, chorismate mutases, peptides CLE, protéines *ubiquitine extension*, protéines SKP1, Hg30C02, Hg19A06, Hg-VAP-1, Hg-VAP-2 et des protéines *RING-like*, etc.) servent à modifier la physiologie de la cellule, à contourner les mécanismes de défense, à interférer avec les signaux hormonaux et à affaiblir le système de défense de l'hôte (Gao *et al.*, 2001, 2003, 2004; Tytgat *et al.*, 2004; Ithal *et al.*, 2007; Bekal *et al.*, 2008; Patel *et al.*, 2010; Wang, 2010; Haegeman *et al.*, 2010; Hammamouch *et al.*, 2012; Chen *et al.*, 2015; Li *et al.*, 2015; Guo *et al.*, 2017). En outre, on a découvert que durant le maintien du syncytium, des enzymes non sécrétées pourraient jouer un rôle important dans la survie du NKS lorsque le syncytium est dégradé par les mécanismes de défense de la plante. Ces enzymes joueraient un rôle dans la biosynthèse de métabolites de la vitamine B telles que les vitamines B5, B7, B1 et B6 et permettre au nématode de survivre (Craig *et al.*, 2008, 2009).

Chez les nématodes à kyste, peu de gènes d'avirulence reconnus par la plante ont été identifiés jusqu'à maintenant. Chez *H. glycines* on a proposé que le gène HgSLP-1 (domaine SNARE) pourrait interagir avec le gène de résistance Rhg1- α -snap et être reconnu par la plante (Bekal *et al.*, 2015). Cette hypothèse n'a pu être testée *in vivo*. Chez le nématode à kyste pâle (pomme de

terre), *Globodera pallida*, le gène Gp-RBP-1 code pour un effecteur de type SPRYSEC qui génère une réaction d'hypersensibilité lorsque co-exprimé avec le gène de résistance Gpa-2 et a donc été proposé comme gène d'avirulence (Sacco *et al.*, 2009; Mei, 2015, 2018). À l'inverse, chez *Globodera rostochiensis*, le produit du gène SPRYSEC-19 interagirait avec le gène de résistance (CC-NB-LRR) de la plante et bloquerait la réaction de mort cellulaire, ce serait donc un gène de virulence (Rehman *et al.*, 2009; Postma *et al.*, 2012). Si la théorie gène-pour-gène telle que décrite par Vanderplank (1978) semble s'appliquer dans le cas des deux espèces de *Globodera* spp. citées plus haut, il semble que le mécanisme soit plus complexe chez *H. glycines*. La résistance et les interactions entre le NKS et le soya sont très complexes et semblent être contrôlées par plusieurs gènes situés aux locus Rhg1 et Rhg4. L'expression des gènes situés sur ces locus dépend entièrement du nombre de copies, des variations nucléotidiques (la version de l'allèle) et de la méthylation et de l'épistasie existant entre tous ces facteurs voire avec de potentiels gènes mineurs (Yan & Baidoo, 2018; Patil *et al.*, 2019). À l'heure actuelle, ces relations n'ont toujours pas été percées à jour (Yan & Baidoo, 2018; Patil *et al.*, 2019). La découverte de gènes d'avirulence et de virulence chez *H. glycines* permettrait d'identifier les locus associés aux différentes formes de virulence, de développer des outils moléculaires pour associer les génotypes aux phénotypes de virulence et d'orienter les programmes d'amélioration génétique vers des lignées résistantes plus durables.

2.3 Notions de génétique des populations

Tel que mentionné précédemment, bien qu'originaires de l'Asie, il semble qu'aucune barrière environnementale n'a limité la dispersion et l'établissement du NKS à travers les États-Unis et le Canada. En fait, une étude portant sur la modélisation bioclimatique de la phénologie du soya et du cycle du NKS a démontré que les changements climatiques auront une incidence importante tant sur l'ouverture de nouvelles aires disponibles pour la culture du soya plus au nord que sur l'augmentation du nombre de générations de NKS produites durant une saison de croissance (Gendron St-Marseille *et al.*, 2013). Afin de comprendre le mécanisme sous-jacent à l'établissement du NKS, il est nécessaire de détailler les éléments inhérents à son adaptation et à la structure génétique de ses populations.

Variabilité génétique et principe d'Hardy-Weinberg

Plusieurs mécanismes tels que les mutations, la migration, les recombinaisons, la dérive génétique et la sélection naturelle influencent la variabilité génétique et l'évolution de la fréquence des génotypes retrouvés dans une population. En génétique des populations, généralement, on estime la fréquence des génotypes à l'aide du principe d'Hardy-Weinberg (Hartl & Clark, 2007).

Selon le principe d'Hardy-Weinberg, si l'on assume que les plus importantes forces évolutives sont absentes (mutation, migration et sélection naturelle), la fréquence allélique (FA) retrouvée dans les populations (panmictiques) gouvernées par une reproduction sexuée et aléatoire permettrait de déduire la fréquence des génotypes (Ewens, 2004; Hartl & Clark, 2007). Afin que le principe s'applique, d'autres conditions doivent être respectées : l'organisme est diploïde, les générations ne se chevauchent pas, le gène considéré possède deux allèles, les FA entre mâles et femelles sont équivalents et la taille de la population est très grande (Hartl & Clark, 2007). Lorsque ces conditions sont respectées, on considère dès lors que :

Equation 1
$$p^2 + q^2 + 2pq = 1$$

ici p^2 représente le génotype **AA**, q^2 représente le génotype **aa** et $2pq$ représente le génotype **Aa** (Ewens, 2004 et Hartl & Clark, 2007). Cependant, ce principe est loin d'être universel, notamment à cause de la présence des forces évolutives. Chez le NKS l'application stricte de ce principe s'avère impossible déjà parce que les générations sont chevauchantes et parce que plusieurs forces évolutives influencent et maintiennent la diversité génétique.

Forces évolutives

La mutation est l'une des forces évolutives favorisant la variabilité génétique intrapopulationnelle; elle est généralement attribuable à des erreurs dans la réplication de l'ADN ou à des dommages causés par une radiation (Loewe & Hill, 2010). Même si elles sont récurrentes, elles se produisent habituellement à des taux très faibles : de 10^{-4} à 10^{-6} (Loewe & Hill, 2010). La recombinaison favorise la diversité génétique en créant de nouvelles combinaisons alléliques par brassage intra- ou interchromosomique du matériel génétique (Spencer *et al.*, 2006). Les mutations et les recombinaisons créent de la diversité en générant de nouveaux allèles qui seront alors soumis aux trois autres forces évolutives soit la dérive

génétique, la migration et la sélection naturelle (Ewens, 2004). Chez le NKS, le mode de reproduction qui implique une femelle et plusieurs mâles, la dormance ainsi que le chevauchement des générations permettent à des mâles ou des femelles issus de générations précédentes de s'accoupler. Cela favorise le maintien d'allèles et une très grande variabilité génétique dans les populations. En outre, la surutilisation des mêmes génotypes résistants de soya occasionne une très forte pression de sélection pour des individus virulents. Par conséquent, les mutations et les recombinaisons n'ont que très peu d'effets en comparaison.

La dérive génétique est un principe selon lequel des phénomènes aléatoires font fluctuer les FA de génération en génération et est décrite par le modèle mathématique Wright-Fisher (Wright, 1931, Hartl & Clark, 2007). Les effets de la dérive dépendent de la taille de la population : dans une grande population, les FA restent plutôt stables alors qu'ils varient beaucoup dans une petite population (Choudhuri, 2014). Avec le temps, il peut se créer un goulot d'étranglement (*bottleneck effect*) qui favorisera soit la fixation ou à la perte d'allèles dans une population menant à une perte progressive du polymorphisme génétique (diversité) (Chakraborty & Nei, 1978; Choudhuri, 2014). L'effet fondateur (*founder effect*) accompagne parfois le mécanisme de dérive génétique et peut exacerber la réduction de la diversité génétique en changeant radicalement les profils alléliques, génétiques et phénotypiques (Chakraborty & Nei, 1978; Choudhuri, 2014). Cet effet se décrit par la création d'une sous-population caractérisée par un faible nombre de représentants qui ne possèdent qu'un « échantillon » d'allèles de la population d'origine, ce qui limite la variabilité dans la population (Chakraborty & Nei, 1978; Choudhuri, 2014). Même lorsque les populations sont divisées en sous-populations, elles sont rarement complètement isolées et c'est dans ce contexte que la migration (ou la dispersion dans le cas du NKS) intervient pour contrebalancer les forces qui réduisent la diversité des allèles (Nei, 1973; Hartl & Clark, 2004). La migration se décrit comme un échange d'individus (ou de gamètes) entre des sous-populations comprenant des différences génétiques et qui favorisent les flux de gènes entre elles (Nei, 1973; Hartl & Clark, 2007). Une migration limitée augmente le niveau de différenciation génétique entre les populations et mène à une spéciation alors que de fortes activités migratoires augmentent l'homogénéisation de la diversité génétique des sous-populations (Nei, 1973).

Les mutations, les recombinaisons, la dérive génétique et la migration ont toutes pour effet de modifier les génomes dans leur ensemble, la sélection naturelle diffère puisque son influence cible plutôt certaines parties plus spécifiques du génome (Andrews, 2010). Selon Darwin (1859), la sélection naturelle se présente comme étant la capacité d'un groupe d'individus porteurs d'allèles leur conférant une meilleure adaptation à survivre et à se reproduire (*fitness*) donc, à passer leur bagage génétique à leurs descendants. De génération en génération, les allèles conférant une meilleure *fitness* seront de plus en plus présents dans une population, ce qui résultera en une adaptation évolutive (Andrews, 2010). Il existe trois grandes formes de sélection naturelle. La sélection purificatrice ou négative qui est la plus fréquente dans le génome et entraîne la disparition d'allèles délétères à l'échelle du génome entier jusqu'à leur élimination complète dans les populations (Charlesworth *et al.*, 1993 Andrews, 2010). La sélection positive ou directionnelle entraîne la transmission d'allèles conférant un avantage sélectif (*fitness*) aux individus porteurs jusqu'à leur fixation dans la population (Andrews, 2010). Le processus sous-jacent se nomme le balayage sélectif (*selective sweep*) qui décrit l'augmentation de la fréquence des allèles situés à proximité d'un gène qui est positivement sélectionné par auto-stop génétique (*hitchhiking*) (Hartl & Clark, 2007). La taille de la région génomique dans laquelle les FA augmentent dépend du taux de recombinaison : plus le taux est faible, plus la région touchée sera grande (Hartl & Clark, 2007). Finalement, la sélection balancée se caractérise par la conservation de la fréquence d'apparition de plusieurs allèles à un même locus par divers mécanismes (Andrews, 2010). Ce type de sélection assure l'augmentation de la variabilité génétique de ce locus dans la population et la réduction de la différenciation génétique entre les sous-populations (Bamshad & Wooding, 2003; Andrews, 2010).

Tel que décrit précédemment, les forces évolutives façonnent la diversité génétique à des locus bien précis. L'observation de ces locus soumis aux forces évolutives et associés à un environnement, à un hôte voire à un phénotype précis permettrait d'identifier des marqueurs d'intérêt ou des gènes candidats. Dans l'optique d'identifier ces locus, il est possible d'utiliser des scans génomiques. D'ailleurs, c'est la stratégie privilégiée dans la détection de locus outliers chez de nombreuses espèces (par ex. insectes, poissons, chiens, etc.) et qui pourrait faciliter l'étude du potentiel d'adaptation du NKS à de nouveaux environnements et aux cultivars résistants (Feng *et al.*, 2015, Pilot *et al.*, 2016; Liu *et al.*, 2017).

2.4 Signatures génétiques de l'adaptation

Les marqueurs moléculaires

Les marqueurs moléculaires se définissent comme étant des parties du génome qui peuvent porter des variations nucléotidiques ou structurales qui sont utiles pour effectuer des analyses d'association entre le génotype et des traits particuliers (Hayward *et al.*, 2015; Karlik & Tombuloğlu, 2016). Les variations nucléotidiques peuvent comprendre un changement nucléotidique d'une paire de bases (polymorphismes nucléotiques simples (SNPs)) ou de plusieurs (*multi-nucleotide variants* (MNVs)) (Schiessl *et al.*, 2019). Les variations structurales impliquent des changements sur des régions plus importantes (> 50 pb) causés par des insertions, des délétions, des duplications, des translocations, des inversions et des variations du nombre de copies de gènes (copy number variation (CNVs)) (Tattini *et al.*, 2015; Schiessl *et al.*, 2019). Ces variations représentant la variabilité génétique aux échelles individuelles et populationnelles sont imputables aux diverses forces évolutives et événements fondateurs vus précédemment. Quatre qualités sont essentielles pour être un bon marqueur soit être neutre, qu'ils aient une expression codominante, qu'ils soient peu coûteux, qu'ils aient un degré élevé de polymorphisme et qu'ils soient fréquents dans le génome (Duthmal *et al.*, 2018). On peut classer les marqueurs moléculaires en fonction de la méthode de détection (par ex. hybridation, PCR, séquençage, etc.) et du mode de transmission (Nadeem *et al.*, 2018; Teama, 2018).

Diverses techniques utilisées en biologie moléculaire permettent de révéler des marqueurs par exemple, l'hybridation et coupure par enzymes de restriction qui permet de générer des polymorphismes de longueur des fragments de restrictions, les RFLP (Nadeem *et al.*, 2018). Aussi diverses techniques PCR permettent de générer des marqueurs dont les plus connus sont les amplifications aléatoires d'ADN polymorphe ou RAPD (*random amplified polymorphism DNA*), les polymorphismes de longueur des fragments amplifiés (AFLP), des microsatellites ou SSR (parfois référés comme *short tandem repeat* et *sequence length polymorphism*), l'amplification aléatoire des microsatellites polymorphes ou *randomly amplified microsatellite polymorphisms* (RAMP), des *inter simple sequence repeat* (ISSR), les rétrotransposons (longs -LINE pour *long interspersed repetitive element* et courts – SINE pour *small interspersed repetitive element*) et bien d'autres (Guidez, 2014; Nadeem *et al.*, 2018). Les avancées quant aux technologies de séquençage ont menées au développement de plusieurs méthodes de séquençage permettant de

générer des marqueurs moléculaires, parmi celles-ci les plus notables sont : le séquençage Sanger, le pyroséquençage, le séquençage de nouvelle génération (next-generation sequencing, NGS), le génotypage par séquençage (GBS), le séquençage de type *diversity array technology* (DArT Seq) et les SNPs (Heather & Chain, 2016; Nadeem *et al.*, 2018).

Dans le cadre de cette thèse, les marqueurs de types SNP générés à l'aide de différentes méthodes de séquençage (NGS et GBS) ont été privilégiés. En effet, les SNPs sont des marqueurs très fréquents dans les génomes, ils sont relativement stables à l'échelle de l'évolution et ont un faible taux de mutation (Patel *et al.*, 2015). On définit un SNP comme étant une différence de base dans un locus entre deux ou plusieurs individus (Edwards *et al.*, 2007). Il existe trois formes de SNPs : les transitions (C/T, pyrimidines, G/A, purines), les transversions (C/G, A/T, T/G et C/A) et les indels (Edwards *et al.*, 2007). Vu leurs nombreuses qualités, les SNPs peuvent servir à différentes applications telles que l'étude de la génétique des populations (patrons de dispersion, l'évolution, l'adaptation et la génétique comparative) ainsi que pour la construction d'une cartographie physique du génome voire celle des traits (Patel *et al.*, 2015).

Scan génomique

La réduction des coûts associés au séquençage des génomes a facilité l'utilisation de marqueurs génétiques pour les études pangénomiques. En appliquant des scans génomiques sur ces marqueurs il est possible de détecter des locus portant des signes de la sélection qui pourraient contenir ou être à proximité de gènes d'intérêt (Helyar *et al.*, 2011; Haasl & Payseur, 2016). Ces gènes candidats permettraient de relier le génotype au phénotype même sans avoir de connaissances a priori (Stinchcombe & Hoekstra 2008; Gautier *et al.*, 2010; Rellstab *et al.*, 2015). La majorité des scans génomiques reposent sur la théorie neutraliste de l'évolution moléculaire telle que proposée par Kimura (1986) et qui postule que la plupart des modifications génomiques imputables aux forces évolutives démographiques (migration, mode de reproduction, histoire démographique) sont neutres par rapport à la sélection naturelle. Peu importe la forme de la sélection (positive, négative, balancée) seules des régions précises du génome sont affectées contrairement aux forces évolutives qui affectent le génome en entier de façon similaire entre les populations ce qui permet de différencier les effets (Cavalli-Sforza, 1966).

Tests basés sur la corrélation entre la fréquence allélique et des variables environnementales

Un des buts en génétique des populations est d'identifier des gènes ou des mutations responsables des réponses adaptatives. Plusieurs approches existent (gènes candidats, génétique classique (*forward genetic*), génétique inverse (*reverse genetic*)), mais peuvent être limitantes dans les cas où certaines espèces ne possèdent pas de génomes de référence, d'informations sur la fonction des gènes, d'individus isolables, etc. (Li *et al.*, 2008). Dans ces cas précis, une approche d'écologie inverse (*reverse ecology*) peut à tout le moins permettre de trouver des locus (régions) sous sélection menant à l'identification de traits phénotypiques ciblés par la sélection (Li *et al.*, 2008). Plusieurs études ont démontré l'utilité de l'écologie inverse dans l'identification de locus sous sélection par un «agent de sélection» quand les gènes ou phénotypes d'intérêts sont inconnus. Par exemple, elles ont démontré que la domestication (Vigouroux *et al.*, 2003), l'utilisation de pesticides ou d'antibiotiques (Kohn *et al.*, 2002; Wootton *et al.*, 2003) ou des facteurs de l'environnement (Storz *et al.*, 2005; Bonin *et al.*, 2006; Fisher *et al.*, 2013) favorisaient la sélection de locus *outliers*. Ainsi, l'identification de ces *outliers* rend possible l'identification de gènes candidats par d'autres études de validation.

Différentes méthodes de scans génomiques ont été développées pour analyser des données de marqueurs moléculaires et les associer à des données environnementales. Ces scans cherchent à détecter un changement important dans une région du génome tel que l'augmentation de la fréquence d'un allèle ou de plusieurs allèles par effet d'auto-stop génétique (déséquilibre de liaison) attribuable à la sélection (Luikart *et al.*, 2003; Stinchcombe & Hoekstra, 2008; Rellstab *et al.*, 2015). Ainsi, en cas d'adaptation locale, les FA des locus sélectionnés et des marqueurs environnants seront plus élevés et différés entre les populations et les différents environnements (Rellstab *et al.*, 2015). Plusieurs méthodes d'analyse de scans génomiques basées sur une approche d'écologie inverse reposent sur des statistiques bayésiennes pour identifier des locus sous sélection et quantifier le degré de corrélation à une variable environnementale ou une autre covariable spécifique à une sous-population, parmi celle-ci notons Bayescan (Foll & Gaggiotti 2008), BayeScEnv (de Villemereuil & Gaggiotti 2015), et Baypass (Gautier, 2015).

Analyse de la diversité reposant sur la différenciation génétique

En 1951, Wright a défini un indice de variance standardisée (F_{st}) qui quantifie l'hétérogénéité des FA entre les différentes subdivisions d'une population. Précisément, cet indice représente la corrélation entre allèles à l'intérieur d'une sous-population par rapport à l'ensemble des sous-populations et se calcule ainsi :

Equation 2
$$F_{st} = \frac{(H_t - H_s)}{H_t}$$

où H_t est basé sur l'hétérozygotie attendue pour la population totale représentée ($H_t = 2p_iq_i$) et où H_s est basé sur l'hétérozygotie attendue dans les sous-populations ($H_s = 2\bar{p}\bar{q}$). Les valeurs peuvent varier entre 0 (pas de différenciation) et 1 (différenciation totale - les populations ont fixé différents allèles). Cependant, la valeur de 1 étant rarement atteinte il est possible de classer les F_{st} selon quatre classes (Table 2.3)

Table 2.3 Classification du degré de différenciation génétique en fonction de la valeur de l'indice de fixation (F_{st}) selon Wright, 1978.

Valeur du F_{st}	Classification
0,00 - 0,05	Faible
0,05 - 0,15	Modérée
0,15 - 0,25	Importante
0,25 - 1,00	Très importante

L'indice de fixation est fréquemment utilisé dans les scans génomiques, car il peut rapidement permettre de comparer des populations entre elles en effectuant des moyennes de valeurs F_{st} pour chaque locus présent dans un ensemble de populations avec des programmes tels que Popoolation 2 (Kofler *et al.*, 2011). Il est possible de détecter des signatures de sélection en observant des valeurs de FA dépassant les seuils calculés sur la base de simulations de coalescence neutres; par exemple, calculés sur la base d'approches bayésiennes telles que Bayescan (Figure 2.5).

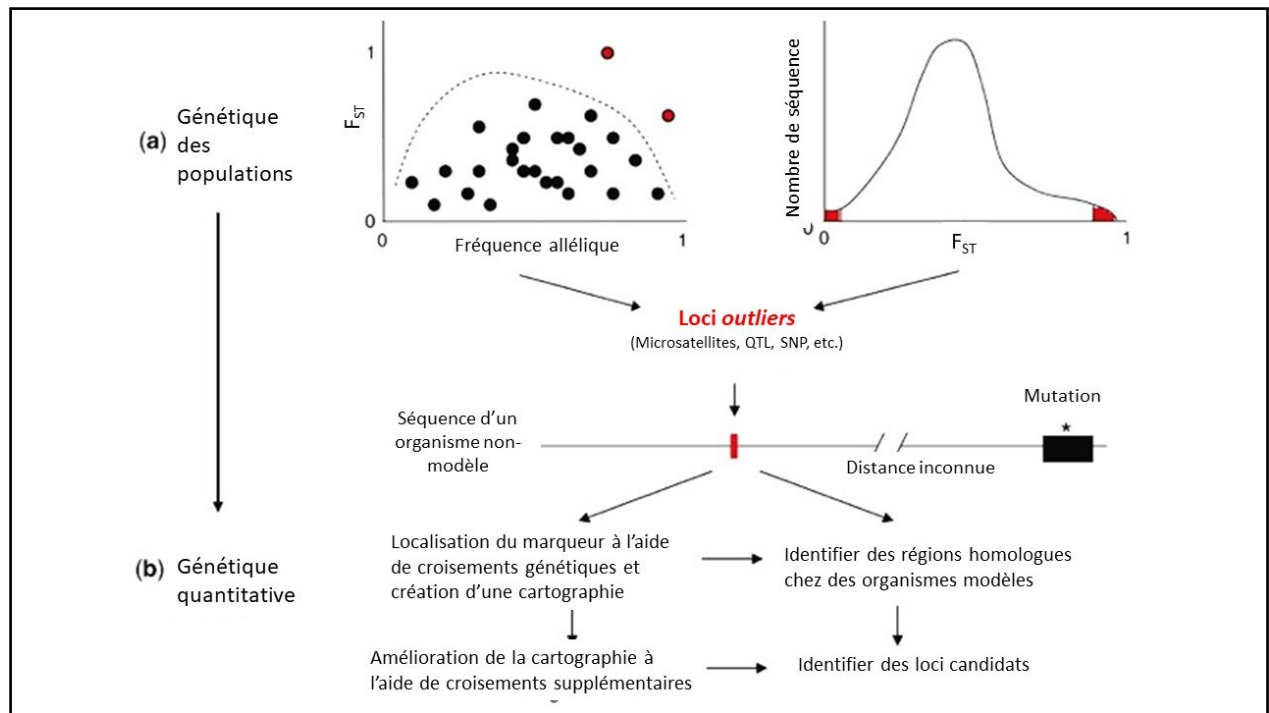


Figure 2.5 Différentes utilisations des indices de fixation (F_{st}) dans l'identification de marqueurs sous sélection et de gènes candidats. A) Identification de locus outliers à l'aide d'une enveloppe de neutralité (pointillé) outliers ou en observant les queues de la distribution des locus. B) Une fois les locus sous sélection identifiés, à l'aide de la génétique quantitative il devient possible d'identifier des gènes candidats soit par croisement génétique ou par la comparaison avec des organismes homologues (traduit de Stinchcombe & Hoekstra, 2008, p.159).

Un autre mode d'identification très simple des gènes candidats est d'observer la distribution des valeurs F_{st} et de trouver des locus sous sélection en observant les valeurs extrêmes (Figure 2.5).

Isolation par la distance

Tel que vu précédemment, les différentes forces évolutives (mutation, dérive génétique, sélection naturelle) vont favoriser l'adaptation aux conditions environnementales et augmenter la différenciation génétique entre les populations (Slatkin, 1987). À l'inverse, la migration va contrebalancer ces forces en augmentant l'homogénéité génétique (Slatkin, 1987). À l'échelle géographique, il est possible d'évaluer l'impact de ces forces sur la différenciation génétique des populations en quantifiant l'isolation par la distance (Rousset, 1997). En utilisant les F_{st} entre paires de sous-populations et la distance géographique les séparant, il est possible d'évaluer l'effet de l'isolation sur la variabilité génétique (Rousset, 1997). On quantifie cette variabilité par le calcul de l'isolation par la distance (IBD, *isolation by distance*) (Equation 3)

Equation 3

$$IBD = F_{st} / (1 - F_{st})$$

En utilisant cette relation entre la distance séparant les populations et la valeur du F_{st} il est possible d'inférer la prépondérance d'une force évolutive par rapport à l'autre pouvant expliquer leur rôle dans le degré de spécialisation d'une sous-population (Figure 2.6)

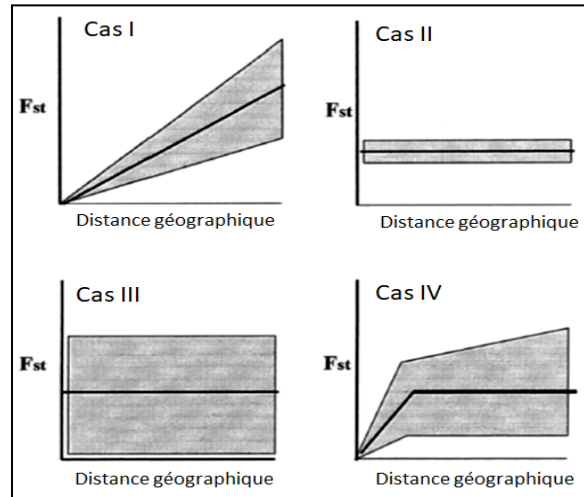


Figure 2.6 Représentations graphiques de quatre cas illustrant l'action des forces évolutives sur la différenciation génétique de sous-population séparées géographiquement. **Cas I**, il y a un équilibre entre la dérive génétique et la migration. **Cas II**, déséquilibre régional, la migration a une plus forte influence que la dérive génétique sur la structure de la population. **Cas III**, en déséquilibre, car la dérive génétique a un effet plus important sur la structure de la population que la migration. **Cas IV**, l'effet de la migration est plus important à petite échelle, alors que c'est la dérive génétique qui influence le plus la structure des populations à une plus grande échelle. (Traduit de Hutchison, D. & Templeton, A. 1999, p.1899).

Tel qu'illustré à la figure 2.6, à petite échelle, la proximité des sous-populations limite l'effet goulot d'étranglement associé à la dérive génétique (perte d'allèles) par l'augmentation de la probabilité d'échange de gènes. Une forte migration à petite échelle géographique réduit la diversité génétique locale, mais augmente la différenciation génétique entre les sous-populations plus éloignées. Ainsi, le rôle de d'une forte migration peut limiter l'adaptation locale par la sélection est critique puisqu'elle peut tout autant favoriser la variabilité génétique et l'augmentation de la taille de la population que diluer les combinaisons génétiques spécialisées et réduire la taille de la population (Slatkin, 1987). Par conséquent, il est primordial d'être en mesure de prédire la capacité des individus formant une population soumise à des changements environnementaux rapides à s'adapter et à s'établir, notamment dans les cas d'introduction d'espèces invasives comme c'est le cas avec le NKS.

2.5 Hypothèses et objectifs

Les hypothèses et objectifs de la thèse sont les suivants :

Chapitre 3

- Hypothèse : Les changements climatiques vont affecter la distribution ainsi que le cycle de vie de l'endoparasite *Heterodera glycines* (NKS) et la phénologie de son hôte *Glycine max* (soya). Cette relation hôte-parasite pourra servir de modèle pour des études similaires.
- Objectifs :
 - 1) utiliser la modélisation bioclimatique couplée à l'utilisation d'un système d'information géographique (SIG) afin de simuler et de visualiser les effets des CC sur la phénologie de l'hôte (soya) et sur le cycle de vie du parasite (NKS);
 - 2) utiliser la modélisation bioclimatique couplée à l'utilisation d'un SIG afin de simuler et de visualiser la dynamique saisonnière des populations de NKS.

Chapitre 4

- Hypothèse : Les gradients climatiques et géographiques influenceront la structure des populations de NKS américaines et canadiennes et auront un effet sur l'adaptation.
- Objectifs :
 - 1) étudier les relations génétiques entre les populations de NKS des États-Unis et du Canada;
 - 2) détecter des populations soumises au processus d'isolation par la distance;
 - 3) détecter des locus sous sélection par la géographie ou par le climat;
 - 4) identifier les fonctions des gènes soumis à la sélection par la géographie ou par le climat.

Chapitre 5

- Hypothèse : La virulence du NKS peut être corrélée à la fréquence allélique à des locus précis.
- Objectifs:

- 1) Trouver une corrélation entre la fréquence allélique (FA) des SNPs trouvés à l'aide d'une approche de génotypage par séquençage (GBS) de pool de populations (57 populations et pools) et le HG type (female index, FI).
- 2) Identifier des changements importants de la FA des variants identifiés à l'aide de la technique de séquençage de génome entier (WGS) entre les pools d'individus de la condition initiale et ceux des individus sélectionnés sur des génotypes de plants de soya résistants.
- 3) Identifier des gènes différenciellement exprimés (RNA-Seq) et uniques à ceux sélectionnés par l'un des génotypes résistants.

Chapitre 3 Simulating the impacts of climate change on soybean cyst nematode and the distribution of soybean

Anne-Frédérique Gendron St-Marseille, Gaétan Bourgeois, Jacques Brodeur et Benjamin Mimee.

Publié tel que : Gendron St-Marseille, A.-F. *et al.* (2019) *Agricultural and Forest Meteorology* **264** : 178–187.

3.1 Abstract

Climate change will undoubtedly impact air and soil temperature in the future. For agriculture, climate warming could be beneficial by opening new lands to profitable crops at higher latitudes. In Canada, soybean production has had a spectacular growth over the last five years with an increase of 27% in the acreage devoted to its cultivation. However, the anticipated climatic changes, coupled with global trading intensification could also favor the introduction and establishment of invasive alien soybean pests. For example, the soybean cyst nematode (SCN) has recently been discovered in the province of Québec, Canada and its presence raised several questions on its reproductive potential, spatial distribution and possible impacts on long term soybean cultivation. To investigate the consequences of climate warming on soybean and SCN development and distribution, and their interactions, two phenology models have been developed and used with the Representative Concentration Pathways (RCP) 4.5 and 8.5, associated with climate change scenarios. These pathways describe two possible future climates based on the level of greenhouse gases concentration in air. Using temperature optimum for soybean maturity group I, our results showed that soybean could be cultivated over all cropland of Québec by the 2050 horizon (2041-2070). Based on phenology models, SCN can currently complete from one to three generations in the different regions of Québec. In the future, the nematode could produce up to four or five generations in an optimistic (RCP 4.5) or pessimistic (RCP 8.5) scenario, respectively. Climate warming will promote the expansion of soybean production to northern areas and expected conditions will be more favourable to SCN development. Accordingly, we should develop more soybean lines with an early maturity and containing other sources of resistance than PI88788 to reduce its reproduction capacity under a more favourable climate in the future.

Keywords

Modelling; Invasive alien species; Soybean; *Glycine max*; Soybean cyst nematode; *Heterodera glycines*.

3.2 Introduction

In an era of major changes, the sustainability of food production is under threat (Sundström *et al.*, 2014; Food and Agriculture Organization of the United Nations (FAO), 2017). Although trade globalization has been beneficial on several aspects, the introduction of many invasive alien species (IAS), has altered agricultural ecosystems and caused losses to the economy (FAO, 2017). In addition, even if empirical evidence is still limited, numerous model studies have predicted that the increasing temperature and atmospheric CO₂ concentration accompanying climate change will have a high impact on agriculture (Savary *et al.*, 2012; Sundström *et al.*, 2014; Bhattarai *et al.*, 2017; Zhao *et al.*, 2017; Deutsch *et al.*, 2018). Based on Representative Concentration Pathways (RCP), which describe greenhouse gas concentration trajectories, the latest report on climate change by the Intergovernmental Panel on Climate Change (IPCC 2014) projects a rise in global mean temperature of 0.3 to 4.8 °C and in CO₂ atmospheric concentration from about 420 ppm to 935 ppm by the late-21st century.

The worldwide production of basic food commodities largely depends on 15 crops (Strange & Scott, 2005), including soybean (*Glycine max* L. Merr.), which ranks sixth in production and fourth in cultivated area (FAOSTAT, 2017ab). Soybean ranks high in protein and other essential nutrients and plays a major role in meeting global food security objectives (Strange & Scott, 2005; Savary *et al.*, 2012). One of the main pathogens of soybean is the soybean cyst nematode (SCN) (*Heterodera glycines* Ichinohe), a below-ground obligate endoparasite that has been accidentally introduced in North America from Asia (Wrather *et al.*, 2010). In the USA, this species causes over \$1 billion in annual losses (Wrather *et al.*, 2009, 2010). Since its first detection in 1954 in North Carolina, SCN has quickly spread across North America via biotic (humans, animals) and abiotic (wind, water) vectors (Riggs, 2004; Tylka & Marett, 2014). In Canada, SCN was first detected in 1987 in Kent County, Ontario (Anderson & Welacky, 1988; Riggs, 2004) and in Québec in 2013 (Mimee *et al.*, 2014). This recent invasion provides a unique opportunity to study the biology of SCN during its phase of establishment. At northern latitudes, the potential is high for expansion of soybean production, as well as for naturalization and spread of SCN (Bootsma *et al.*, 2005; Osborne *et al.*, 2012).

Climate change will impact the interaction between soybean and SCN. For instance, climate warming will favour the expansion of cultivated areas for soybean by increasing the length of the

growing season which, in return, could open new establishment areas for SCN as seen for others species like the brown marmorated stink bug which is actually spreading in eastern Canada (Bootsma *et al.*, 2005; Bebber *et al.*, 2013; Kistner, 2017). Furthermore, environmental changes caused by climate warming can influence the capacity of plants to cope with biotic stresses, possibly increasing their susceptibility to parasitic infections (Anderson *et al.*, 2004; Grulke, 2011; Newton *et al.*, 2012). As for all poikilotherm organisms, high temperature reduces the developmental time of SCN (Lauritis *et al.*, 1983; Bonner & Schmitt, 1985) thereby potentially increasing the number of generations per growing season and result in increased damage to soybean production.

Climate change will modify local environmental conditions in most cropping systems but the impacts on host-parasite interactions and yield remain unknown (Donnatelli *et al.*, 2017). The soybean – nematode relationship could represent a good model to study the impact of climate change on a below-ground endoparasite establishment capacities, reproduction and time of infection in function of rising soil temperature. In this study, we use bioclimatic models to predict the impacts of climate warming on soybean and SCN phenology in Québec, by comparing a reference climate period (1981-2010) to two climatic scenarios, RCP 4.5 and RCP 8.5, of CO₂ concentrations associated with the 2041-2070 period (referred to as 2050 horizon) (IPCC, 2014). Furthermore, a case study is presented to illustrate direct impacts of climate warming during a growing season at field scale. We also propose the use of geographic information systems (GIS) to visualize the effects of climate change on expected (i) soybean growing area, (ii) SCN establishment area, and (iii) SCN reproductive capacity and seasonal population dynamics.

3.3 Material and Methods

Study area

The study area included all agricultural land in Québec as determined by Agriculture and Agri-Food Canada, AAFC (2015) (Figure 3.1). We used the Québec Lambert projection for all maps and the North American Datum of 1983 (NAD83) as the geodesic reference.

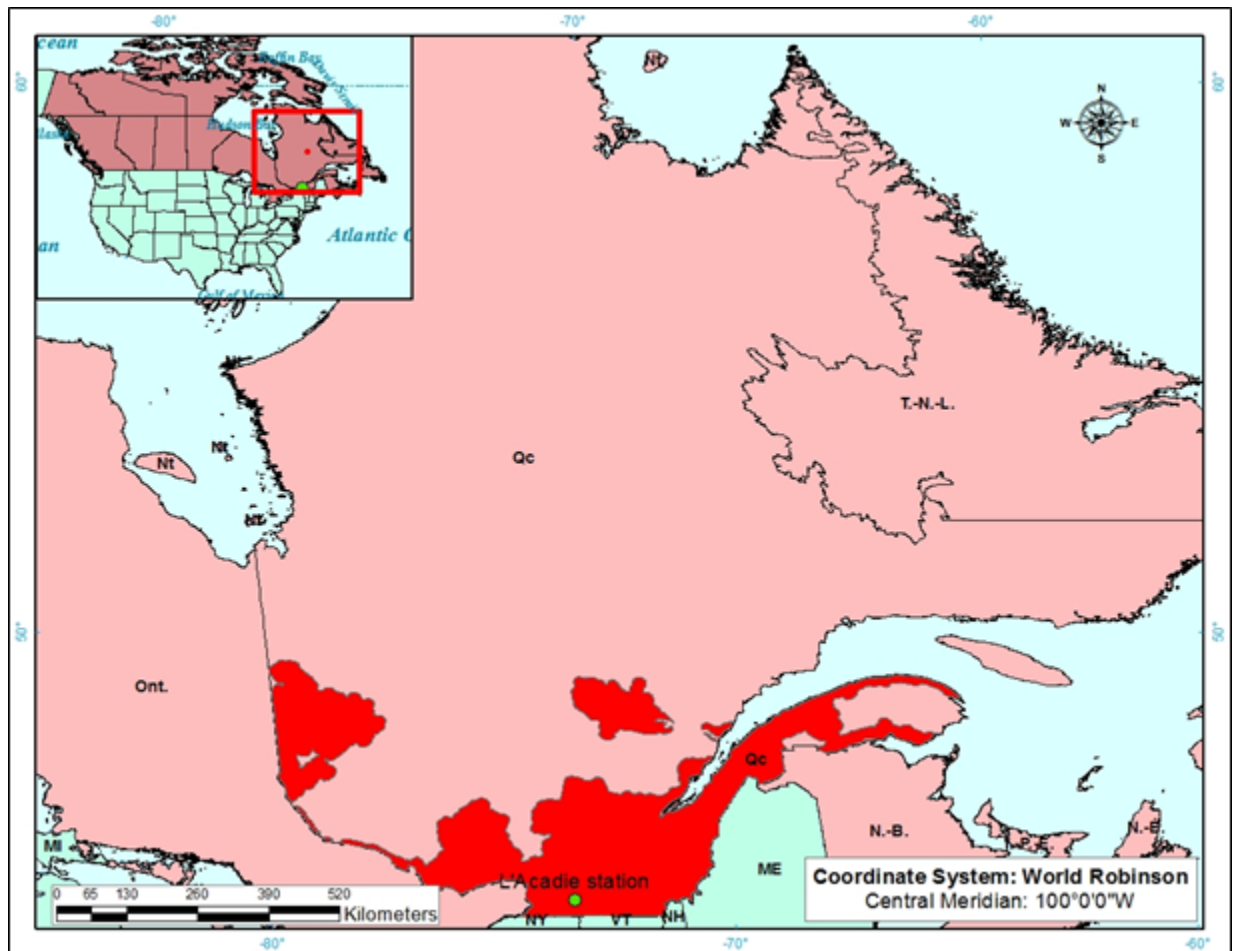


Figure 3.1 Study zone in red of Québec province representing the extent of the agricultural land. The L'Acadie research station of Agriculture and Agri-Food Canada is identified by a green dot (Basemap belong to ArcGIS® software by ESRI).

Climate data of recent years (1981-2010) and climate scenarios of the future (2041-2070)

The air temperatures for the reference years (1981-2010) were selected from the *Daily 10 km Raster-Gridded Climate Dataset for Canada 1950-2012* available on the website of AAFC's National Land and Water Information Service (NLWIS). This dataset includes daily maximum and minimum air temperatures, and precipitations recorded from 1950 to 2012 and interpolated with ANUSPLIN 4.3 software (Hutchinson, 2004) on a 10 km X 10 km grid (Hutchinson *et al.*, 2009) covering approximately 7500 weather stations across Canada.

The future simulations and projections for 2050 horizon (2041-2070) are based on two CO₂ RCPs developed by IPCC (2014). For projections, we selected an intermediate CO₂ emission pathway

(RCP 4.5) and a pessimistic CO₂ emission pathway (RCP 8.5) to foresee the effects of temperature variations on the seasonal phenology of soybean and SCN. Under the intermediate projection (RCP 4.5), global mean surface temperatures are projected to increase from 1.1°C to 2.6°C until the end of the century (2100) accompanied by an average atmospheric CO₂ concentrations from 580 ppm to 720 ppm (\approx 650 ppm CO₂) (Van Vuuren *et al.*, 2011). Under the most pessimistic projection (RCP 8.5), global mean surface temperatures are projected to increase from 2.6°C to 4.°C accompanied by an average atmospheric CO₂ concentration near 1390 ppm CO₂ (Van Vuuren *et al.*, 2011).

Ten global climate models from the Coupled Model Intercomparison Project Phase 5 (CMIP5) (Table 3.1) were used for each RCP CO₂ emission trajectories to simulate the projected climate for the 2050 horizon (Flato *et al.*, 2013; CMIP5, 2015).

Table 3.1 List of ten climate models from the CMIP5 (2015) used in the construction of the RCP4.5 and RCP8.5 climate change simulations.

Modelling Center	Model	Institution	Reference
CSIRO-BOM	ACCESS1-0	CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia), and BOM (Bureau of Meteorology, Australia)	Collier et Uhe, 2012.
CNRM-CERFACS	CNRM-CM5	Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancées en Calcul Scientifique	Voltaire <i>et al.</i> , 2011.
CSIRO-QCCCE	CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organisation in collaboration with the Queensland Climate Change Centre of Excellence	Gordon <i>et al.</i> , 2010.
NOAA GFDL	GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory	Dunne <i>et al.</i> , 2012, 2013.

IPSL	IPSL-CM5A-MR	Institut Pierre-Simon Laplace	Dufresne <i>et al.</i> , 2013.
MIROC	MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	Watanabe <i>et al.</i> , 2010.
MPI-M	MPI-ESM-MR	Max Planck Institute for Meteorology (MPI-M)	Giorgetta <i>et al.</i> , 2012.
NCC	NorESM1-M	Norwegian Climate Centre	Bentsen <i>et al.</i> , 2013.
INM	INMCM4	Institute for Numerical Mathematics	Volodin <i>et al.</i> , 2010.
CCCma	CanESM2	Canadian Centre for Climate Modelling and Analysis	Chylek <i>et al.</i> , 2011.

A total of 20 simulations were performed, 10 for each of the two CO₂ RCPs. The temperature forecasts associated with both RCP emission pathways are based on the average of the data obtained from these simulations (Table 3.1).

The climate models simulating the 2050 horizon were developed by the OURANOS climate scenarios team (Roy, P., unpublished). Climate simulations were adjusted by a method of “quantile - quantile mapping” (ThiemeBl *et al.*, 2011) that allows to correct the bias of a simulation. This method finds the quantiles of model variables and substitutes the simulated values with those of the identical quantile from an observational dataset. To obtain a good resolution, AAFC’s 10 km x 10 km grid was used for the development of soybean and SCN phenology models (Hutchinson, 2004; Hutchinson *et al.*, 2009).

Daily air temperatures recorded over a 30-year period (1981-2010) in Québec and those simulated for the 2050 horizon have been integrated in a software developed by AAFC’s bioclimatology and modelling research team (Plouffe *et al.*, 2014). Daily soil temperatures were

estimated using a number of equations available in the STICS (“Simulateur multIdisciplinaire pour les Cultures Standard”) crop growth and development simulator (Brisson *et al.*, 2008). Air temperature data were used for the soybean phenology model and soil temperature at 20-cm depth for the SCN phenology model since nematodes are more abundant in the first layer of soil between 0 and 20 cm (Alston and Schmitt, 1987).

Optimistic (cooler) and pessimistic (warmer) scenarios have been developed to illustrate the inter-annual variation between simulations within each RCP. For each variable, the ten climate models were run. Only the values obtained by at least 9/10 models in at least 24/30 years were kept. The 10th and 90th percentile values were identified in each climate model output. The average of all the 10th percentile values was reported as the “warmer scenario” and represents the higher temperature while the average of the 90th percentile values was the “cooler scenario” and represents the lower temperature of the distribution. This allows the quantification of the uncertainty associated with the projections for the 2050 horizon.

Bioclimatic models simulating soybean and SCN phenology as a function of photoperiod and air and soil temperature

The model simulating soybean phenology as a function of air temperature and photoperiod was provided by AAFC’s bioclimatology and modelling research team in Saint-Jean-sur-Richelieu, Québec (Plouffe *et al.*, 2014). In summary, thirty data sets obtained over four years, two sites, and multiple planting dates, were used for model conceptualization and calibration. A series of equations are associated to three specific phenological phases of the plant, namely emergence, vegetative and reproductive phases. Equations within each phenological phase include effects of both temperature $f(T)$ and photoperiod $f(P)$. The temperature and photoperiod functions originate from Yan & Hunt (1999) and Setiyono *et al.* (2010), respectively (see Supplementary Material S1 for more details). This phenology model was incorporated in CIPRA (Computer centre for agricultural pest forecasting) software (Plouffe *et al.*, 2014). The map describing potential growing area for soybean in Québec was generated using the soybean phenology model for the reference period and the 2050 horizon. Soybean maturity group I (MG-I), currently sowed in Québec, was used in the model. Maps were produced using ArcMap 10.3.

The development of SCN phenology model was based on a mechanistic approach that uses species-specific physiological tolerance factors such as temperature to predict its developmental rate (Jeschke & Strayer, 2008). Developmental data measurements such as stage duration for temperatures between 16 °C and 32 °C were obtained from the literature (Ross, 1964; Lauritis *et al.*, 1983; Melton *et al.*, 1986). Because growth rates differed between eggs (in the soil) and other developmental stages (in soybean roots), two equations were developed. A linear regression was used to describe the growth of SCN inside the egg, from embryogenesis to hatching of the second juvenile stage (J2), for temperatures between 16 °C and 30 °C.

Equation 4
$$Y = 0.0041x + 0.0087$$

Y represents the number of days to reach the J2 stage and hatch and x is the temperature in Celsius degrees. Then, a polynomial regression described the growth of SCN following penetration of the root (from J2 to adult) for temperatures between 16 °C and 32 °C.

Equation 5
$$Y = 0.0008x^2 + 0.04x - 0.4106$$

Y represents the number of days to reach the adult stage and x is the temperature in Celsius degrees.

In order to simplify the model, we made three important assumptions (reviewed in Jeschke & Strayer, 2008). First, biotic interactions do not influence positively or negatively SCN geographical range over space and time. Second, the genetic and phenotypic compositions of SCN populations are constant over space and time. Third, SCN can disperse wherever its hosts are present, i.e. in this study, where soybean is grown. Of course, biotic interactions can influence SCN dispersal and host susceptibility, this will need to be addressed in future studies.

The SCN phenology model starts with SCN egg hatching which is set to the same day than the predicted soybean sowing date; this in order to simplify the model but also to synchronize SCN life cycle with the presence of the host. Since the soybean planting date varies with years, it is estimated using an algorithm. The algorithm first identifies the median date of the last spring frost for each geographic node (10 km x 10 km) and related depth (20 cm) using reference climate datasets (1981-2010) or datasets generated by the two RCP scenarios. Then, for each year, starting on the median date of the last spring frost, the software identifies a moving 5-day mean temperature greater than or equal to 12.8 °C. Under this temperature, the soybean seeds

develop very slowly and can enter dormancy (Ontario Ministry of Agriculture Food and Rural Affairs (OMAFRA), 2009). The fifth day corresponds to the planting date used in the model.

As parameters associated with SCN reproduction, including migration of the males out of the root to mate with the females, are complex and difficult to include in the model, three days were added to the model after the adult stage is reached and prior to embryogenesis based on migration approximation made by Lauritis *et al.* (1983). The model provides the dates on which the embryonic (four cells to first juvenile stage (J1)), infectious (J2) and adult stages are reached, the duration of each stage, and the number of generations per growing season. The last seasonal SCN generation was considered incomplete if the adult stage is not reached in at least 24 years out of 30 years with a predicted value in at least nine of the ten climate models.

Calculation of key variables

Using the parameters described above, the soybean planting date is calculated for each geographic node (10 km x 10 km) using the ten climate models. Then, each simulation predicts the date when each soybean phenological stage is reached using degree-day accumulation. The average of the 10th and 90th percentile values obtained using each of the ten climate models was reported as the “cooler” and “warmer” scenarios for planting and maturity dates at each of the geographic node. In the same manner, the completion of each SCN developmental stage was calculated using the equations described above starting at the planting date predicted by each of the ten climate models. The number of full generations by growing season is reported for future and reference climate for each of the geographical node. The duration (days) and dates associated with the beginning and the end of each generation was also outputted for the geographical node corresponding to L’Acadie experimental station. The inter-annual variation was also evaluated by averaging the 10th and 90th percentile values obtained by each of the ten climate models. ArcMap 10.3 was used to produce maps illustrating the number of generations potentially produced in the climatic conditions of the reference period (1981-2010) and of the 2050 horizon. All these variables were calculated under both RCP simulations.

3.4 Results

Soybean phenology under different greenhouse gases concentration trajectories

The differences between soybean maturity group I planting dates in Québec depending on RCP simulations and climate change scenarios are shown in figure 3.2.

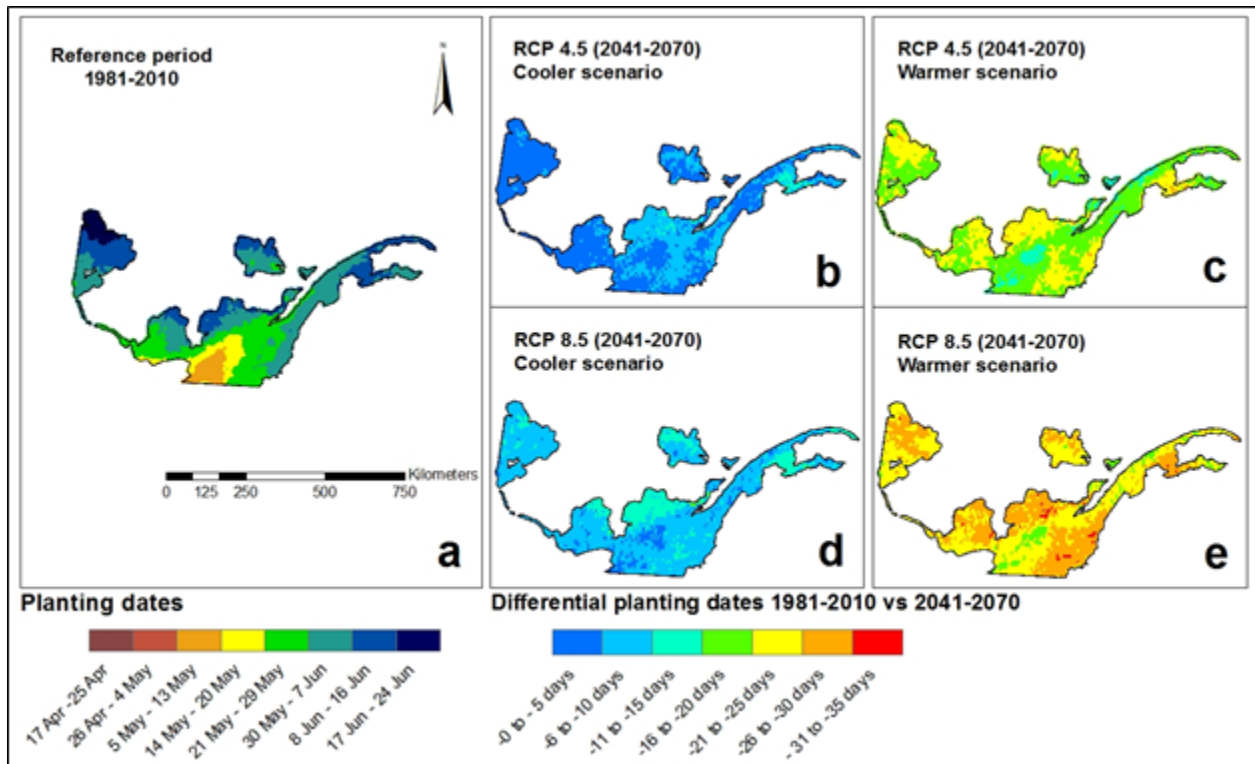


Figure 3.2 Differences between future and reference planting dates for soybean MG-I in Québec, based on greenhouse gases concentration pathway simulations RCP 4.5 and 8.5. Cooler and warmer scenarios represent the inter-annual variation obtained by 10 different climate simulations within each RCP. The cooler scenario is the average of the 10th percentile values obtained using each of the models and represents the lower temperatures of the distribution while the warmer scenario is the mean of the 90th percentile values, representing the higher temperature obtained across the 30 years. Negative values indicate planting dates earlier than the reference period (1981-2010).

Under the reference scenario (1981-2010), soybean can be planted from May 5th in southern Québec to June 24th in the northern area (Figure 3.2a). During the 2050 horizon, models predicted that under the greenhouses concentration pathway of RCP 4.5, it will be possible to plant soybean up to 15 days earlier under a cooler scenario (Figure 3.2b) and up to 25 days earlier under a warmer scenario (Figure 3.2c). For RCP 8.5, models suggested that soybean could be sowed up to 15 days earlier under a cooler scenario (Figure 3.2d) and up to 35 days earlier under a warmer scenario (Figure 3.2e).

Figure 3.3 shows the potential for MG-I soybean to reach maturity currently (Reference period) and under RCP simulations.

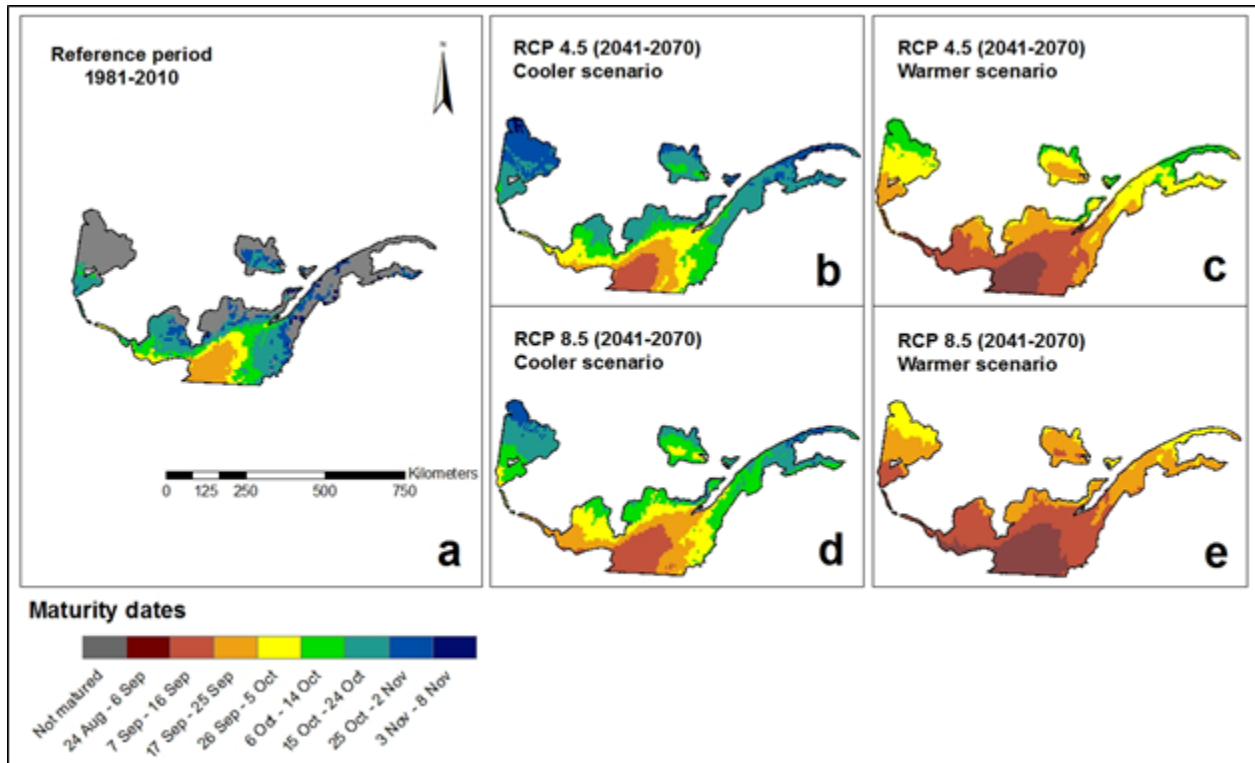


Figure 3.3 Soybean (MG-I) maturity dates (mean values) under the reference period and the 2050 horizon in Québec based on greenhouse gases concentration pathway RCP 4.5 and 8.5. Cooler and warmer scenarios represent the inter-annual variation obtained by 10 different climate simulations within each RCP. The cooler scenario is the average of the 10th percentile values obtained using each of the models and represents the lower temperatures of the distribution while the warmer scenario is the mean of the 90th percentile values, representing the higher temperature obtained across the 30 years. Maturity is not reached in grey areas.

Soybean (MG-I) maturity during the 1981-2010 period can be reached from late September in south-western Québec to late October in other areas (Figure 3.3a). The models predicted that soybean could be harvested from September 7th (south) to November 9th (north) under the cooler scenario of RCP 4.5 simulations (Figure 3.3b). Harvesting will start sooner, from August 24th (south) to mid-October (north), under a warmer scenario (Figure 3.3c). Thus, it will be possible to harvest 14 days earlier in southern Québec under a warmer scenario and 25 days earlier in the north. Under the cooler scenario of RCP 8.5 simulations (Figure 3.3d), the model predicted that soybean could be harvested from September 7th to November 2nd from south to north, respectively, whereas under a warmer scenario (Figure 3.3e) this period will start as soon as August 24th to October 5th from south to north, respectively. When comparing the warmer and the

cooler scenario, the models predicted that in the south it will be possible to harvest up to 14 days earlier while in the north harvest could occur up to 28 days earlier.

Predicted number of SCN generations during a growing season over Québec cropland

In order to figure out SCN reproductive potential at 20 cm depth in Québec's cropland, the total number of SCN generations produced during a single season was mapped for the reference period and RCP's simulations of future climate (Figure 3.4).

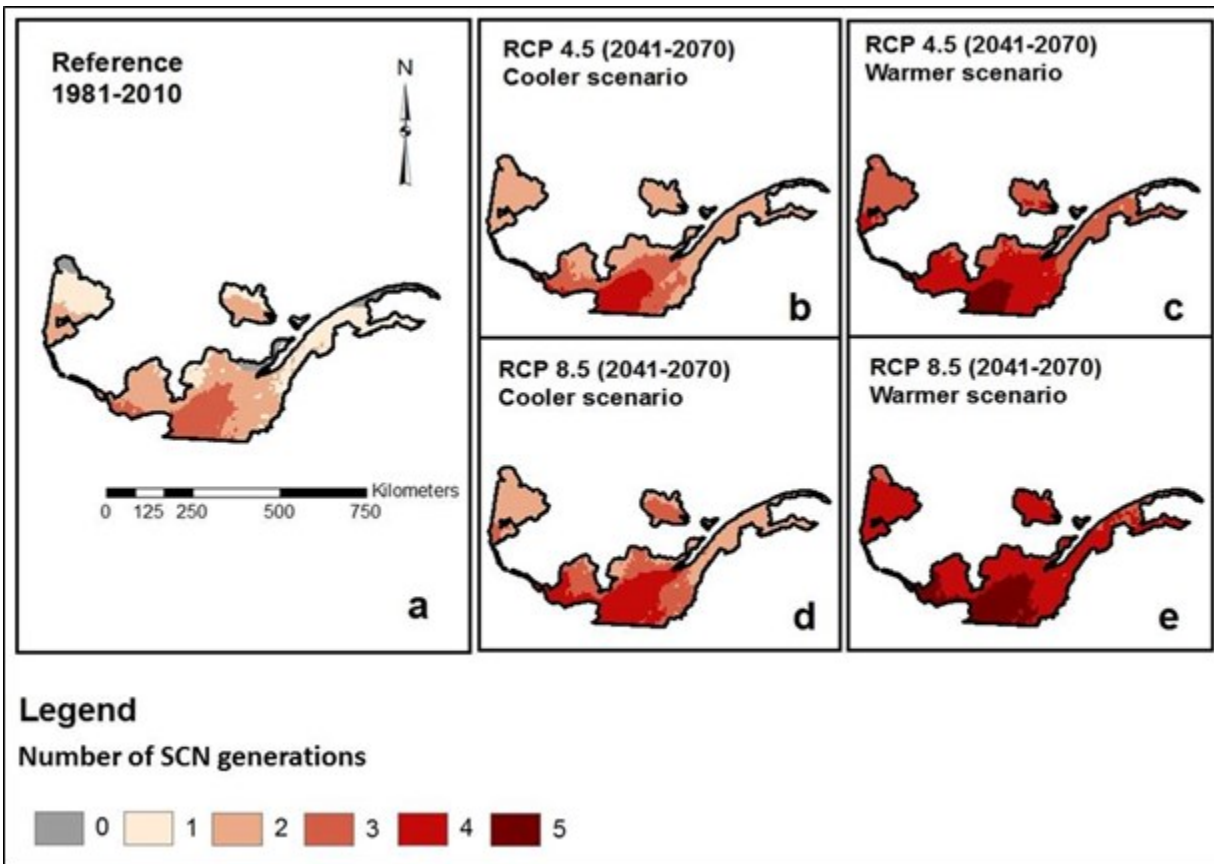


Figure 3.4 Number of soybean cyst nematode generations completed at 20 cm depth during a single growing season in Québec for the reference period and the 2050 horizon using two RCP simulations. Cooler and warmer scenarios represent the variation between simulations within each RCP. For the 2050 horizon, generations had to be completed for at least 24 years over 30 to be recorded.

From the reference period to the 2050 horizon, there will be a marked increase in the number of SCN generations produced over a growing season. Currently, up to three generations can be obtained in the area where soybean is grown (Figure 4.4a) and SCN cannot survive or attain sexual maturity in the northern limits of Québec's cropland. In contrast, both RCP simulations

indicate that future climatic conditions could allow the development of additional SCN generations. Modelling of SCN phenology using RCP 4.5 simulations under the cooler scenario (Figure 4.4b) predicted the completion from four (south) to two (north) generations while under a warmer scenario it predicted the completion from five (south) to two (north) generations. Under the cooler scenario simulated with the RCP 8.5 data (Figure 3.4d), the number of expected full generations is similar to RCP 4.5 results, with the exception that the area where five and four generations could be completed is much larger. Finally, the warmer scenario simulated with RCP 8.5 (Figure 3.4e), predicted the completion of five generations over a larger area in southern Québec, four generations almost everywhere else, except at the northern limits where three generations could be completed.

SCN seasonal population dynamics: a case scenario at L'Acadie station

To illustrate the application of the model on expected seasonal population dynamics of SCN, we selected an experimental site located at L'Acadie (45°17'38" N, 73°20'58" W, 43,80 m), in the middle of the main soybean production area in Québec, and where weather data are available for the past 30 years. Figure 3.5 illustrates the predicted number and duration (from embryogenesis to adult) of SCN generations based on the average of the 10 climate models for the reference period and 2050 horizon and classified into optimistic (cooler) and pessimistic (warmer) scenarios.

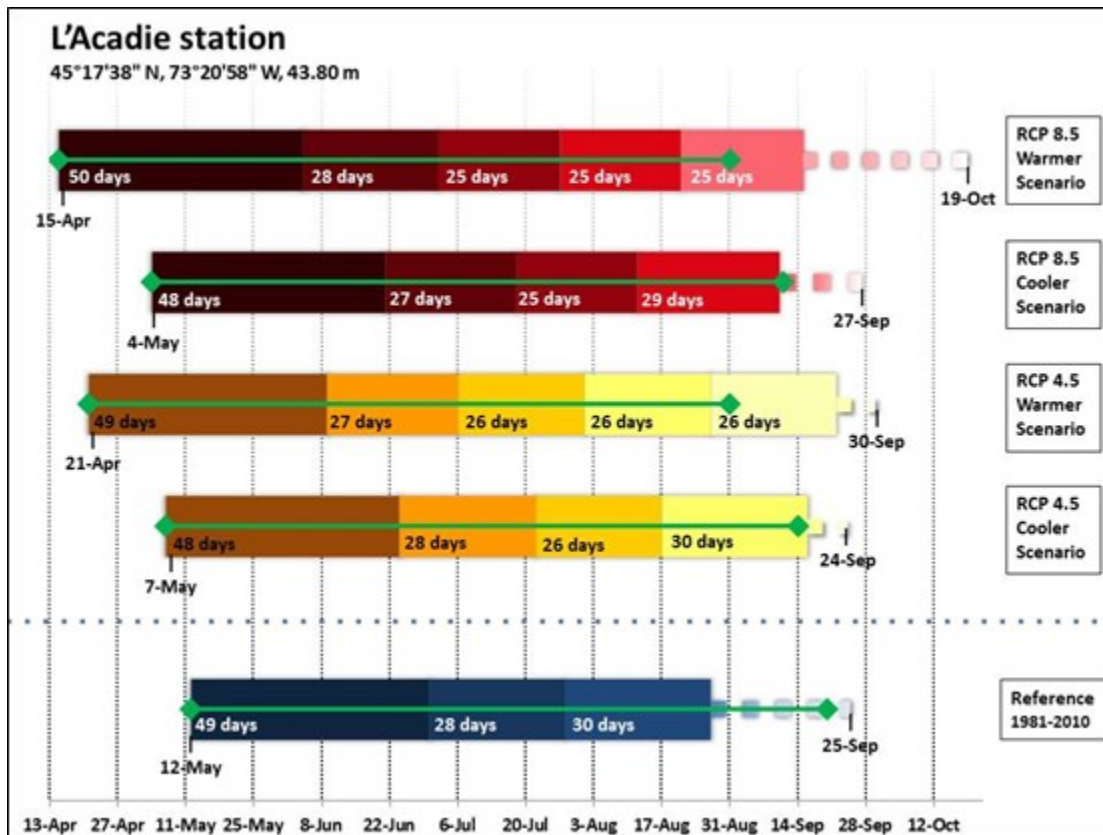


Figure 3.5 Duration (days) and dates associated with the beginning and the end of each generation of soybean cyst nematode completed during a growing season for the reference period and the 2050 horizon at L'Acadie experimental station, Québec. Dotted lines to the right of each bar represent incomplete generation at the end of the season. Blue, yellow and red gradient bar represent the reference period, the RCP 4.5 and the RCP 8.5, respectively. For each RCP, the bottom and upper bars represent the cooler (90th percentile) and warmer (10th percentile) scenarios, respectively. Green lines represent soybean MG-I cropping cycle based on predicted sowing and maturity dates obtained by the 10 climatic simulations.

Modelling suggested that soybean can currently be sowed around May 12th and that SCN can complete three generations from mid-May to the end of August in southern Québec (Figure 3.5, blue gradient). Nematodes most likely enter into winter dormancy between the end of August and September 25th without completing a fourth generation 20 years out of 30.

With the expected increase in temperatures in the 2050 horizon, under the RCP 4.5 planting dates of soybean MG-I (Figure 3.5) under a warmer scenario could be advanced by 16 days compared to the cooler scenario. The duration of interaction (synchronism) between soybean MG-I and SCN populations is similar in both scenarios, but under a warmer climate nematodes could pursue their development even in absence of its host for about a month. SCN developmental and

reproductive period could last 140 and 162 days under the cooler and warmer scenario, respectively. RCP 4.5 simulations under the cooler (Figure 3.5, lower yellow gradient) and warmer (figure 3.5 upper yellow gradient) scenario predicted four and five complete generations, respectively.

Under the RCP 8.5, the planting date under a warmer scenario will be advanced by 19 days compare to the cooler scenario. SCN potential reproductive period varied from 146 days under a cooler scenario to 187 days under a warmer scenario. The RCP 8.5 simulations under the cooler scenario (Figure 3.5, lower red gradient) predicted four complete generations and the initiation of a fifth incomplete generation. Under the warmer scenario (Figure 3.5, upper red gradient), five complete generations are predicted with the completion of a sixth generation 13 years out of 30.

During the 2050 horizon, soybean planting date will be between 5 to 21 days (RCP 4.5) and 8 to 27 days (RCP 8.5) earlier. The duration of interaction between soybean and SCN (from 128 to 138 days) is similar for each scenarios and periods, but under a warmer climate the nematode could pursue its development from about 50 more days. SCN potential reproductive period will be between 4 to 26 days (RCP 4.5) and 10 and 51 day (RCP 8.5) longer during the 2050 horizon. Finally, temperature change during the 2050 horizon could lead to the production of one to two additional generations of SCN.

3.5 Discussion

This study showed that climate change would modify host and parasite phenology and population dynamics. Climate warming will change soybean production in northern latitude by lengthening the growth season and opening new areas to this crop. Concurrently, higher temperatures will also increase the rate of development of SCN, a poikilothermic species.

Growing soybean under climate change in the province of Québec

Phenology modelling for the 2050 horizon predicted a significant increase in the potential soybean growing area (Figures 3.2 and 3.3) and earlier planting dates in all regions where soybean is currently grown. Predictions showed that soybean MG-I could currently be grown all over the existing cultivable land in Québec. Depending on the RCP simulation and latitude, temperature increase associated with climate change could extend the soybean growing season by

one to three weeks for the 2050 horizon as Bootsma *et al.* (2005) demonstrated in their study. According to our results, a shift in MG will likely occur in all regions. For example, for regions in southern parts of the cultivable land, where MG-I soybean is currently grown, the model predicted an earlier planting date. Because soybean growth and development will benefit of longer days (daylight reach its maximum at end of June then start decreasing), planting MG-II soybean could be successful. As demonstrated in other studies, early planting generates higher yields because soybean reproduction starts sooner when days are longer and light is more intense (Zhang *et al.*, 2000; De Bruin & Pederson, 2008; Nafzinger, 2009; Robinson *et al.*, 2009). Therefore, a favourable temperature scenario such as RCP 4.5 or RCP 8.5 where soybean could be planted in southern area in early April rather than in early May would provide growers more seeds per plant. However, the consequences of early planting on pest population dynamics also need to be considered.

Possible effects of climate change on SCN seasonal cycle

The genetic variability and adaptation potential towards climate change of SCN populations established in Québec remain to be explored. Understanding the capacity of the species to adapt to current and future climate is critical to forecast colonization of new areas, as it was shown in North America for other pathogens, insects and weeds (Boland, 2004; Dukes *et al.*, 2009, 2011; Urbanski *et al.*, 2012; Baker *et al.*, 2015). All RCP simulations predicted an increase of the number of generations of SCN during the growing season from currently three to potentially four or five. As for its plant host, warming alone will accelerate SCN development. This species should therefore become more abundant in Québec, as for the United States of America and South America where warmer climate and longer presence of the host provide ideal conditions for its reproduction (Hunt, 2008; Tylka & Marett, 2014). Under the different latitudinal climate and field conditions from Florida to North Dakota, SCN can produce from six to two generations by growing season (University of Illinois at Urbana-Champaign, 1991; Nelson & Bradley, 2003; Iowa State University, 2004; Chen, 2011). These results are consistent with SCN life cycle models run with future temperature data that are predicting four to five (RCP 4.5) or five to six generations (RCP 8.5). These comparisons showed that the model seems consistent with the situation currently observed in southern areas and could help Québec growers to anticipate and manage this pest species under future climate.

Together, the rise of temperature and the expected increase of the soybean growing area could influence and enlarge the establishment area and the distribution of SCN across the province of Québec. Similar situation has occurred in Minnesota where SCN has spread at the same rate as soybean production increased from south to north since the first discovery of SCN in 1978 (Chen, 2011; Kandel, 2013). Simulation of SCN development showed that SCN could currently complete its life cycle in all the cultivable lands in Québec if a susceptible host is available. For now, SCN has been found in different region of southern Québec where MG-000 to MG-I soybean are grown (Mimee *et al.*, 2016). RCP simulations suggest that SCN infestations at higher latitudes, on MG-000 to MG-0, would considerably complicate its management as most of the available resistant cultivars have been developed for higher maturity groups (Orf, 2013; Rincker *et al.*, 2017). Thus, it is very important to anticipate the establishment of SCN populations at higher latitudes by increasing the breeding efforts for SCN resistance in short season soybean and diversifying the sources of resistance.

For a better understanding of the effects of climate changes on SCN life cycle at a farm scale level, a case study was run at a specific location in Québec's main soybean growing area. Simulations predicted significant effects of climate change on soybean phenology and SCN development at L'Acadie, with earlier planting dates (between 5 to 27 days) in the near future. This could allow a shift in maturity groups from MG-I to MG-II with longer vegetative stage duration. Despite the earlier arrival of higher temperatures, the duration of SCN first generations will not be shorter because the cycle will begin sooner, the temperatures will still be low and SCN developmental rate should remain the same during the first generation. On the other hand, summer temperatures will be near optimal and should accelerate the rate of development. However, a shift to MG-II would extend the presence of soybean vegetative stage in the field by 5-20 days. This could allow secondary infections during this stage and limit root growth (Egli, 1995; Pederson *et al.*, 2008). The resulting damages would significantly impair subsequent flowering and pod filling which are critical stages for yield (Wrather & Anand, 1986; Todd, 1993). Currently, soybean plants from MG-I grown at L'Acadie station take about 40 to 43 days to reach flowering and the duration of the first SCN generation is 49 days. Hence, only the initial generation (spring inoculum) invades the roots and limit water and nutrient intakes during this stage.

After flowering, the modelling of soybean MG-I planted at L'Acadie station predicted that the reproductive stage will last for 68-72 days. This stage is critical for yield determination in terms of seed number and pod size. Moreover, SCN could still impair directly and indirectly production by weakening the plant and increasing damages by other diseases like the fungi causing the Sudden Death Syndrome (SDS) and others such as members of the oomycetes (Todd *et al.*, 1987; Wang *et al.*, 2003). Furthermore, RCP simulations predict more generations from mid-summer to fall which could lead to increases in SCN population density and cyst number that will constitute the initial inoculum of the next growing season.

As seen previously in RCP simulations, SCN can pursue its development during more than one month after its host had reached maturity, which can be enough to produce a fifth or a sixth complete generation. These ultimate generations will not directly impact the soybean because at this stage pod production is over but could lead to an increased inoculum that will affect soybean production the following year. In addition, the presence of winter annual weeds such as purple deadnettle (*Lamium purpureum*) or common chickweed (*Stellaria media*), both excellent alternative hosts which are present in Québec, could allow a greater production of cysts as seen in the United States and augment the annual inoculum (Johnson *et al.*, 2008). Altogether, the longer growth season and the presence of winter weeds as the potential to impede significantly the soybean production in Québec in a context of climate change.

Bioclimatic modelling and climate change

In this study we developed a process-based SCN phenology model relying only on the literature available describing the relationship between temperature and SCN development to describe the effect of climate warming on a soil temperature dependent nematode species without considering the effect of other bioclimatic and abiotic factors that could also influence its development. Nevertheless, our study demonstrates that the use of daily fluctuating temperature data coupled with a simple mathematical equation were sufficient to simulate the development and reproduction of a soil species from embryogenesis to sexual maturity and egg laying. This model still needs to be validated with field trials under different temperature regimes; however, when compared with different locations across the United States of America, the number of generations produced during one growing season is realistic. The approach also innovates by combining the

phenology of the host to the temperature-based modelling of pest development. This greatly increases the outcomes of modelling by highlighting the synchronism of different stages and predicting if the pest will attack the plant during a more susceptible or yield-limiting stage. It also allows the computation of indirect effects of climate on pest development such as the presence of the host or not. We consider that this type of simplistic model would help anticipate the potential impact of a pest species and to plan future management of crop and pest control based on potential needs. It is also easily adaptable to other plant-parasite systems as it does not rely on complex parameters of crop growth simulations. In this study, the two chosen RCPs all predicted a global rise of temperatures, sea levels, precipitations and others environmental variables so that the biology of all poikilothermic species could dramatically change in the near future.

To our knowledge, this is the first mechanistic bioclimatic and coupled host-pest phenology model for the prediction of current and future SCN distribution and reproduction potential under different climatic scenarios. Soybean acreages have exploded in Canada during the last decade and will continue to increase. The recent introduction of SCN in Québec has generated a lot of concerns for growers and raised important questions about its potential of establishment in the province. With this regard, the development of a simple and realistic model that could be used in any climatic region to quickly enable authorities to determine the reproductive potential of SCN was the first step of an integrated crop-pest management program.

Future perspectives

The perspective of a warming climate can be seen as a good opportunity for northern agriculture. This would open up new lands to lucrative crops like soybean and extend the growing season in the current production area, allowing the planting of late maturity cultivars with higher yields. However, this will indubitably have major consequences on pests and diseases as already suggested in numerous studies (Bebber *et al.*, 2013; Elad & Pertot, 2014; Kistner, 2017). Many organisms are currently unable to overwinter or have very limited impact under the current climate and could become problematic in future. For SCN, this study has also shown that a shift to MG-II would extend the presence of the host vegetative stage and allow more secondary infections that could negatively impact yield. The shorter SCN developmental time under warmer climate will also accelerate the build-up and spread of nematode populations that could also be

facilitated by an increase in soybean cultivated areas and better connectivity between fields. Therefore, it is crucial to develop and deploy new pest management tools to maximize the chances of a positive outcome. We have to develop cultivars integrating novel sources of resistance. We can also imagine the development of a predictive yield model that integrates winter survival, inoculum density and, eventually, interactions with other biotic and abiotic factors in each region. This information would help the growers to make a conscious choice of cultivar and MG to maximize their profits in a changing climate.

3.6 Acknowledgements

The authors would like to thank Nathalie Beaudry for technical assistance, Danielle Choquette and Dominique Plouffe from the AAFC's bioclimatology and modelling research team for their contribution to the project, and the OURANOS team for providing climatic data. This work was supported by Agriculture and Agri-Food Canada

3.7 Supplementary data

A supplemental document is accessible at :

<https://ars.els-cdn.com/content/image/1-s2.0-S0168192318303320-mmcl.docx>

3.8 References

- Agriculture and Agri-Food Canada (AAFC) 2016. Canada: outlook for principal field crops. Retrieved from <http://www.agr.gc.ca/eng/industry-markets-and-trade/statistics-and-market-information/by-product-sector/crops/crops-market-information-canadian-industry/canada-outlook-for-principal-field-crops/canada-outlook-for-principal-field-crops-february-16-2016/?id=1455720699951> [Nov 2017].
- Agriculture and Agri-Food Canada (AAFC) 2015. Quebec Agroclimatic Atlas Data. Retrieved from <http://open.canada.ca/data/en/dataset/03807263-b1ad-44e9-b2a9-b574cd26e992> [Nov 2017].
- Alston, D. G. & Schmitt, D. P. 1987. Population density and spatial pattern of *Heterodera glycines* in relation to soybean phenology. *Journal of Nematology*, **19**, 336-345.
- Anderson, J. P., Badruzaufari, E., Schenk, P. M., Manners, J. M., Desmond, O. J., Ehlert, C., Maclean, D. J., Ebert, P. R. & Kazan, K. 2004. Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in *Arabidopsis*. *Plant Cell*, **16**, 3460-3479.

- Anderson, T. R. & Welacky, T. W. 1988. First report of *Heterodera glycines* on soybeans in Ontario, Canada. *Plant Disease*, **72**, 453.
- Baker, M. B., Venugopal, P. D. & Lamp, W. O. 2015. Climate change and phenology: *Empoasca fabae* (Hemiptera: Cicadellidae) migration and severity of impact. *PLoS One*, **10**, e0124915.
- Bhattarai, M. D., Secchi, S. & Schoof, J. 2017. Projecting corn and soybeans yields under climate change in a Corn Belt watershed. *Agricultural Systems* **152**, 90-99.
- Bebber, D. P., Ramotowsk, M. A.T. & Gurr, S. J. 2013. Crop pests and pathogens move polewards in a warming world. *Nature Climate Change*, **3**, 985-988.
- Bentsen, M., Bethke, I. & Debernard, J. B. 2013. The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate. *Geoscientific Model Development*, **6**, 687-720.
- Boland, G.J., Melzer, M.S., Hopkin, A.A., Higgins, V. & Nassuth, A. 2004. Climate change and plant diseases in Ontario. *Canadian Journal of Plant Pathology*, **26**, 335-350.
- Bonner, M. J. & Schmitt, D. P. 1985. Population dynamics of *Heterodera glycines* life stages on soybean. *Journal of Nematology*, **17**, 153-158.
- Bootsma, A., Gameda, S. & McKenney, D. W. 2005. Potential impacts of climate change on corn, soybeans and barley yields in Atlantic Canada. *Canadian Journal of Plant Science*, **85**, 345-357.
- Brisson, N., Launay, M., Mary, B. & Beaudoin, N. 2008. Conceptual Basis, Formalisations and Parameterization of the STICS crop model. Éditions Quae, INRA, Versailles, France. 297 p.
- Chen, S. 2011. Soybean cyst nematode management guide. University of Minnesota Extension. Retrieved from <http://www1.extension.umn.edu/agriculture/soybean/soybean-cyst-nematode/> [Nov 2017].
- Coupled Model Intercomparison Project Phase 5 (CMIP5) 2015. CMIP5 - Data Access - Availability. Retrieved from <http://cmip-pcmdi.llnl.gov/cmip5/availability.html> [Nov 2017].
- Chylek, P., Li, J., Dubey, M. K., Wang, M. & Lesins, G. 2011. Observed and model simulated 20th century Arctic temperature variability: Canadian Earth System Model CanESM2, *Atmospheric Chemistry & Physics Discussion*, **11**, 22893–22907.
- Collier, M. & Uhe, P. 2012. CMIP5 datasets from the ACCESS1.0 and ACCESS1.3 coupled climate models. CAWCR Technical Report No. 059. The Centre for Australian Weather and Climate Research, Aspendale, Victoria. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.710.4998&rep=rep1&type=pdf>

- De Bruin, J. L. and Pedersen, P. 2008. Soybean seed yield response to planting date and seeding rate in the Upper Midwest. *Agronomy Journal*, **100**, 696–703.
- Dukes, J. S., Chiariello, N. R., Loarie, S. R. & Field, C. B. 2011. Strong response of an invasive plant species (*Centaurea solstitialis* L.) to global environmental changes. *Ecological Applications*, **21** (6), 1887-1894.
- Deutsch, C. A., Tewksbury, J. J., Tigchelaar, M., Battisti, D. S., Merrill, S. C., Huey, R. B. & Naylor, R. L. 2018. Increase in crop losses to insect pests in a warming climate. *Science*, **361** (6405), 916-919.
- Dukes, J. S., Pontius, J., Orwig, D., Garnas, J. R., Rodgers, V. L., Brazee, N. ... Ayres, M. 2009. Responses of insect pests, pathogens, and invasive plant species to climate change in the forests of northeastern North America: What can we predict? *Canadian Journal of Forest Ressource*, **39**, 231–248.
- Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumon, O. ... Vuichard, N. 2013. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dynamics*, **40**, 2123–2165.
- Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M, Hallberg, R. W., Shevliakova, E. ... Zadeh, N. 2012. GFDL’s ESM2 Global coupled climate–carbon earth system models. Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, **25**, 6646-6665.
- Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L. ... Zadeh, N. 2013. GFDL’s ESM2 Global coupled climate–carbon earth system models. Part II: Carbon system formulation and faseline simulation characteristics. *Journal of Climate*, **26**, 2247-2267.
- Elad, Y. & Pertot, I. 2014. Climate change impacts on plant pathogens and plant diseases. *Journal of Crop Improvement*, **28**, 99-139.
- FAO, 2017. The future of food and agriculture – Trends and challenges. Rome. Retrieved from <http://www.fao.org/3/a-i6583e.pdf> [Dec 2017].
- FAOSTAT, 2017a. 2014 World area harvested of primary crops. Retrieved from <http://faostat.fao.org> [Oct 2017].
- FAOSTAT, 2017b. 2014 World production quantity of primary crops. Retrieved from <http://faostat.fao.org> [Oct 2017].
- Flato, G., Marotzke, J., Abiodun, B. P., Braconnot, S.C., Chou, W., Collins, P. ... Rummukainen, M. 2013. Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (Eds. Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. & Midgley, P.

M.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

- Gendron St-Marseille, A.-F. 2013. Le nématode à kyste du soja (*Heterodera glycines*) : Enjeux des changements climatiques sur sa distribution, sa reproduction et sur les probabilités de synchronisme avec le soja (*Glycines max*). Master thesis. University of Sherbrooke, Canada.
- Giorgetta, M. A., Roeckner, E., Mauritsen, T., Bader, J., Crueger, T., Esch, M. ... Stevens, B. 2013. The atmospheric general circulation model ECHAM6: Model description (Report n°135). Max Planck Institute for Meteorology: Hamburg. Retrieved from https://www.mpimet.mpg.de/fileadmin/publikationen/Reports/WEB_BzE_135.pdf [Sep 2016].
- Gordon, H., O'Farrell, S., Collier, M., Dix, M., Rotstayn, L., Kowalczyk, E. ... Watterson, I. 2010. The CSIRO Mk3.5 Climate Model. CAWCR Technical Report No. 021. The Centre for Australian Weather and Climate Research, Melbourne, Victoria. Retrieved from http://www.cawcr.gov.au/technical-reports/CTR_021.pdf [Sep 2016].
- Gulke, N. E. 2011. The nexus of host and pathogen phenology: Understanding the disease triangle with climate change. *New Phytologist*, **189**, 8–11.
- Hunt, D. J., 2017. *Heterodera Glycines* In: Invasive Species Pest (ed CAB International) Retrieved from <http://www.cabi.org/isc/datasheet/27027> [Nov 2017].
- Hutchinson, M. F. 2004. ANUSPLIN version 4.3. Canberra: Australian National University. Retrieved from <https://researchers.anu.edu.au/publications/38018?term=anusplin+4.3> [Nov 2017].
- Hutchinson, M. F., McKenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E. & Papadopol, P. 2009. Development and testing of Canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003. *Journal of Applied Meteorology and Climatology*, **48**, 725-741.
- Interpanel on Climate Change (IPCC) 2014. Climate Change 2014: Synthesis Report. In: Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Eds. Core writing team, Pachauri, R.K. & Meyer, L. A.), Geneva, Switzerland, 151 pp.
- Jeschke, J. M. & Strayer, D. L. 2008. Usefulness of bioclimatic models for studying climate change and invasive species. *Annals of the New York Academy of Sciences*, **1134**, 1-24.
- Johnson, W. G., Creech, J. E. & Mock, V. A. 2008. Role of winter annual weeds as alternative hosts for soybean cyst nematode. *Crop Management*. <https://doi:10.1094/CM2008-0701-01-RV>

- Juroszek, P. & von Tiedemann, A. 2012. Plant pathogens, insect pests and weeds in a changing global climate: A review of approaches, challenges, research gaps, key studies and concepts. *Journal of Agricultural Science*, **151**, 163-188.
- Kakaire, S., Grove, I. G. & Haydock, P. P. J. 2012. Effect of temperature on the life cycle of *Heterodera schachtii* infecting oilseed rape (*Brassica napus* L.). *Nematology*, **14**, 855-867.
- Kendel, H. 2013. Soybean production – Field guide for North Dakota and Northwestern Minnesota (Guide A-1172). North Dakota State University Extension Service: Fargo. Retrieved from <https://www.ag.ndsu.edu/publications/landing-pages/crops/soybean-production-guide-a-1172> [Nov 2017].
- Kistner, E. J. 2017. Climate change impacts on the invasive brown marmorated stink bug. United States Department of Agriculture – Midwest climate hub. Retrieved from http://www.pestrisk.org/wp-content/uploads/2017/10/3_3_IPRRG11_Kistner.pdf [Dec 2017].
- Lauritis, J. A., Rebois, R. V, & Graney, L. S. 1983. Development of *Heterodera glycines* Ichinohe on soybean, *Glycine max* (L.) Merr., under gnotobiotic conditions. *Journal of Nematology*, **15**, 272-281.
- Melton, T. A., Jacobsen, B. J. & Noel, G. R. 1986. Effects of temperature on development of *Heterodera glycines* on *Glycine max* and *Phaseolus vulgaris*. *Journal of Nematology*, **18**, 468-474.
- Mimee, B., Peng, H., Popovic, V., Duceppe, M.-O., Tetreault, M.-P. & Belair, G. 2014. First report of soybean cyst nematode (*Heterodera glycines* Ichinohe) on soybean in the Province of Quebec, Canada. *Plant Disease*, **98**, 429.
- Mimee, B., Gagnon, A.-È., Colton-Gagnon, K. & Tremblay, É. 2016. Portrait de la situation du nématode à kyste du soya (*Heterodera glycines*) au Québec (2013-2015). *Phytoprotection*, **96**, 33-42.
- Nafzinger, E. 2009. Soybean. In: Illinois Agronomy Handbook: 24th Edition. , p.27-36 University of Illinois at Urbana-Champaign, College of Agriculture, Cooperative Extension Service.
- Nelson, B. D. & Bradley, C. A. 2003. Soybean Cyst Nematode. Retrieved from <https://www.ndsu.edu/pubweb/~bernelso/soydiseseases/cyst.shtml> [Nov 2017].
- Newton, A. C., Torrance, L., Holden, N., Toth, I. K., Cooke, D. E., Blok, V. & Gilroy, E. M. 2012. Climate change and defense against pathogens in plants. *Advance of Applied Microbiology*, **81**, 89-132.
- OMAFRA 2009. Soybean - Planting and crop development. In: Agronomy guide for field crop (Ed. Brown, C.), Service Ontario Publications, Toronto. Retrieved from <http://www.omafra.gov.on.ca/english/crops/pub811/pub811ch2.pdf> [Oct 2107]

Orf, J. 2013. Breeding for SCN resistance in the north. Presentation conducted during Soybean breeders' workshop. St. Louis, Missouri. Retrieved from

http://soybase.org/meeting_presentations/soybean_breeders_workshop/SBW_2013/presentations/13SBW_tues_Orf.pdf [Nov 2017].

Osborne, T., Rose, G. & Wheeler, T. 2012. Variation in the global-scale impacts of climate change on crop productivity due to climate model uncertainty and adaptation. *Agricultural and Forest Meteorology*, **170**, 183-194.

Pederson, P., Kumudini, S., Board, J. & Conley, S. P. 2008. Soybean Growth and Development. In: Using Foliar Fungicides to Manage Soybean Rust (Eds. Dorrance, A. E., Draper, M. A. & Hershman, D. E.), pp. 41-47, Media Distribution Communications and Technology, The Ohio State University, Columbus, Ohio.

Plouffe, D., Bourgeois, G., Beaudry, N., Chouinard, G. & Choquette, D. 2014. CIPRA - Computer Centre for Agricultural Pest Forecasting: Crop Guide. Agriculture and Agri-Food Canada. Technical Bulletin, A42-119/2013E-PDF. 136 p.

Riggs, R. D. 2004. Hosts Range. In: Biology and Management of the Soybean Cyst Nematode (Eds. Riggs, R. D. & Wrather, J. A.), pp. 107-114, Robert D. Riggs and J. Allen Wrather Schmitt and associates of Marcelline, Marcelline, MO.

Robinson, A. P., Conley, S. P., Volenec, J. J. & Santini, J. B. 2009. Analysis of high yielding, early-planted soybean in Indiana. *Agronomy Journal*, **101**, 131–139.

Ross, J. P. 1964. Effect of soil temperature on development of *Heterodera glycines* in soybean roots. *Phytopathology*, **54**, 1228-1231.

Savary, S., Ficke, A., Aubertot, J.-N. & Hollier, C. 2012. Crop losses due to diseases and their implications for global food production losses and food security. *Food Security*, **4**, 519-537.

Setiyono, T. D., Cassman, K. G., Specht, J. E., Dobermann, A., Weiss, A., Yang, H. ... De Bruin, J. L. 2010. Simulation of soybean growth and yield in near-optimal growth conditions. *Field Crops Research*, **119**, 161-174.

Strange, R. N. & Scott, P. R. 2005. Plant disease: A threat to global food security. *Annual Review of Phytopathology*, **43**, 83-116.

Sundström, J. F., Albiñ, A., Boqvist, S., Ljungvall, K., Marstorp, H., Martiin, C. ... Magnusson, U. 2014. Future threats to agricultural food production posed by environmental degradation, climate change, and animal and plant diseases – a risk analysis in three economic and climate settings. *Food Security*, **6**, 201-215.

Themeßl, J. M., Gobiet, A. & Leuprecht, A. 2011. Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, **31**, 1530–1544.

- Todd, T. C. 1993. Soybean planting date and maturity effects on *Heterodera glycines* and *Macrophomina phaseolina* in Southeastern Kansas. *Journal of Nematology*, **25**, 731-737.
- Todd, T. C., Pearson, C. A. S. & Schwenk, F. W. 1987. Effect of *Heterodera glycines* on charcoal rot severity in soybean cultivars resistant and susceptible to soybean cyst nematode. *Annals of Applied Nematology*, **1**, 35-40.
- Tylka, G. L. & Marett, C. C. 2014. Distribution of the soybean cyst nematode, *Heterodera glycines*, in the United States and Canada: 1954 to 2014. *Plant Health Progress*, **15**, 13-15.
- University of Illinois at Urbana-Champaign 1991. RPD No. 501 - The Soybean Cyst Nematode Problem. Retrieved from <http://ipm.illinois.edu/diseases/series500/rpd501/> [Nov 2017].
- Urbanski, J., Mogi, M., O'Donnell, D., DeCotiis, M., Toma, T. & Armbruster, P. 2012. Rapid adaptive evolution of photoperiodic response during invasion and range expansion across a climatic gradient. *The American Naturalist*, **179**, 490-500.
- Iowa State University 2004. Soybean cyst nematode. Integrated Crop Management, Iowa State University Extension and Outreach, Ames, Iowa. Retrieved from http://crops.extension.iastate.edu/soybean/diseases_SCNbiology.html [Nov 2017].
- Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, a., Hibbard, K. ... Rose, S. K. 2011. The representative concentration pathways: An overview. *Climate Change*, **109**, 5-31.
- Voltaire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S. 2011. The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dynamics*, **40**, 2091-2121.
- Volodin, E. M., Dianskii, N. A. & Gusev, A. V. 2010. Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations. *Atmospheric and Oceanic Physics*, **46**, 414-431.
- Wang, J., Niblack, T. L., Tremain, J. A., Wiebold, W., Tylka, G. L., Marett, C. C. ... Schmitd, M. E. 2003. Soybean cyst nematode reduces soybean yield without causing obvious aboveground symptoms. *Plant Disease*, **87**, 623-628.
- Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S. ... Kimoto, M. 2010. Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. *Journal of Climate*, **23**, 6312-6335.
- Wrather, J. A. & Koenning, S. R. 2009. Effects of diseases on soybean yields in the United States 1996 to 2007. *Plant Health Progress*. <https://doi:10.1094/PHP-2009-0401-01-RS>

- Wrather, J. A., Shannon, G., Balardin, R., Carregal, M., Escobar, R., Gupta, G. K. ... Tenuta, A. 2010. Effect of diseases on soybean yield in the top eight producing countries in 2006. *Plant Health Progress*. <https://doi:10.1094/PHP-2010-0125-01-RS>
- Yan, W. & Hunt, L. A. 1999. An equation for modelling the temperature response of plants using only the cardinal temperatures. *Annals of Botany* **84**, 607-614.
- Zhang, L., Wang, R. & Hesketh, J. D. 2000. Effects of photoperiod on growth and development of soybean floral bud in different maturity. *Agronomy Journal*, **93**, 944–948.
- Zhao, C., Liu, C., Piao, S., Wang, X., Lobell, D. B., Huang, Y. ... Asseng, S. 2017. Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the national academy of sciences of the united states of America* **114** (35), 9326-9331.

Chapitre 4 Genome Scans Reveal Homogenization and Local Adaptations in Populations of the Soybean Cyst Nematode

Anne-Frédérique Gendron St-Marseille, Étienne Lord, Pierre-Yves Véronneau, Jacques Brodeur et Benjamin Mimee.

publié tel que : Gendron St-Marseille, A.-F. *et al.* (2018) *Frontier in Plant Science* 9 : doi: 10.3389/fpls.2018.00987

4.1 Abstract

Determining the adaptive potential of alien invasive species in a new environment is a key concern for risk assessment. As climate change is affecting local climatic conditions, widespread modifications in species distribution are expected. Therefore, the genetic mechanisms underlying local adaptations must be understood in order to predict future species distribution. The soybean cyst nematode (SCN), *Heterodera glycines* Ichinohe, is a major pathogen of soybean that was accidentally introduced in most soybean-producing countries. In this study, we explored patterns of genetic exchange between North American populations of SCN and the effect of isolation by geographical distance. Genotyping-by-sequencing was used to sequence and compare 64 SCN populations from the United States and Canada. At large scale, only a weak correlation was found between genetic distance (Wright's fixation index, F_{ST}) and geographic distance, but local effects were strong in recently infested states. Our results also showed a high level of genetic differentiation within some populations, allowing them to adapt to new environments and become established in new soybean-producing areas. Bayesian genome scan methods identified 15 loci under selection for climatic or geographic co-variables. Among these loci, two non-synonymous mutations were detected in *SMAD-4* (mothers against decapentaplegic homolog 4) and *DOP-3* (dopamine receptor 3). High-impact variants linked to these loci by genetic hitchhiking were also highlighted as putatively involved in local adaptation of SCN populations to new environments. Overall, it appears that strong selective pressure by resistant cultivars is causing a large scale homogenization with virulent populations.

Key words: Bayesian outlier detection, genetic diversity, genotyping-by-sequencing, *Heterodera glycines*, isolation by distance.

4.2 Introduction

The introduction of an organism into a new environment can have unpredictable detrimental consequences, including public health problems, losses in biodiversity and ecosystem services or crop yield losses due to exotic weeds, insects and pathogens, altogether resulting in significant economic impact (Pimentel *et al.* 2005). Unfortunately, the steady increase in international trade facilitates the movement and introduction of new invasive species (Hulme 2009). In addition, climate change is altering environmental conditions and could change the species' distribution range or favor their establishment in previously unsuitable habitats (Early and Sax 2014). It is therefore imperative to carry out specific risk assessment in order to target species to be controlled. Alien invasive species, by definition, did not evolve in the biogeographic habitat in which they are introduced. Consequently, they are often poorly adapted to their new environment. Generally, the most successful invaders will have a high potential for rapid adaptation through phenotypic plasticity or microevolution (Novak 2007). Understanding the genetic mechanisms of local adaptation is therefore critical to predict future species distribution.

Plant-parasitic nematodes are microscopic worms that reduce global annual food production by 12.3% and cause more than US\$157 billion in economic losses worldwide (Hassan *et al.* 2013). In North America, one of the most damaging species is the soybean cyst nematode (SCN), *Heterodera glycines* Ichinohe. Since its first detection in 1954 in Hanover County, North Carolina, SCN has been reported in almost every soybean-producing county in the United States (Davis and Tylka 2000), as well as in southwestern Ontario, Canada (Anderson *et al.* 1988). During the 2000s, SCN colonized several new northern localities in the United States (North Dakota in 2003; Mathew *et al.* 2015) and Eastern Canada (northeastern Ontario in 2007 and Québec in 2013; Mimee *et al.* 2014). A few studies have investigated SCN dispersal, which seems to follow a northward and eastward pathway in North America (Tylka and Marett 2017). This dispersion pattern correlates with the expansion of soybean cultivation following the introduction of new soybean varieties with shorter maturity periods and improved tolerance to drought and cold (Shurtleff and Aoyagi 2010; Yu 2011).

It is expected that SCN would survive and multiply throughout the current North American soybean-growing area and complete at least two generations at its northern limit (Gendron St-Marseille 2013). Climate warming should also favour the establishment of SCN at higher latitudes and increase the number of generations per year in most regions. However, these predictions are based strictly on temperature requirements and do not account for the intrinsic capacity of SCN to adapt to new environmental conditions. Genetic variations within a population reflect its evolutionary potential and result from four evolutionary forces that affect individual fitness: mutation, gene flow, selection, and genetic drift (Eizaguirre and Baltazar-Soares, 2014). For most organisms, including SCN, the relative weights of these forces can differ significantly. Mutations are rare events that should not contribute significantly to SCN adaptations in a short time frame. Gene flow depends on the dispersal ability of an organism, which for SCN is achieved mainly by means of human activities at the short spatial scale (Kristjansson 2010). In addition, wind and flooding can carry SCN cysts over very long distances, and contribute to its dispersal at the regional and continental scales. Many different selection pressures can shape the genetic structure of SCN populations, but host plant is probably the strongest selection factor. The ability of SCN to reproduce on a given soybean genotype differs greatly depending on its resistance genes and the nematode's virulence profile (HG type) (Colgrove and Niblack 2008; Niblack *et al.* 2008). Thus, management decisions by growers (for example, the systematic use of resistant cultivars) can result in a strong selection pressure. Finally, the influence of genetic drift will also depend on pest control strategies, because they contribute to dictate population size, although it was shown for cyst nematodes that genetic diversity can be very high within a single cyst (Green *et al.* 1970). Each nematode female can mate with several males and lay hundreds of eggs that can survive for at least a decade in the soil (Slack *et al.* 1972). Thus, even if the diversity appears reduced due to genetic drift under strong selection by resistant cultivars, most alleles probably persist for several years in the population in a “dormant” state.

In other cyst nematode species, genetic diversity at the population level has been studied by means of several techniques, including microsatellite markers, ITS-RFLP (internal transcribed spacer–restriction fragment length polymorphism), RAPD (random amplified polymorphic DNA), and 2-DGE (two-dimensional gel electrophoresis) (e.g., Blok *et al.* 1997; Grenier *et al.*

2001; Plantard *et al.* 2008; Boucher *et al.* 2013). However, these methods focus only on specific sections of the genome, yield very few markers, or do not allow precise comparisons among populations. For example, Eves van den Akker *et al.* (2015) recently published a metagenetic approach to genotype populations of the pale cyst nematode, *Globodera pallida*. While this method was shown powerful to rapidly evaluate genetic diversity and distribution of specific mitotypes, it is based on very few neutral markers, which prevent any assessment for selection. Genome scan is an interesting approach for the identification of genetic loci involved in adaptation to specific selection pressure. It was notably used in *G. pallida* to identify genomic regions associated with virulence on resistant potato cultivars (Eoche-Bosy *et al.*, 2017).

Despite rapid advances in next-generation sequencing (NGS) technologies, sequencing a large number of individual genomes at high coverage in order to perform population genetic studies remains very expensive and may require important quantities of DNA from individuals. Elshire *et al.* (2011) developed a genotyping-by-sequencing (GBS) protocol to rapidly identify single-nucleotide polymorphisms (SNPs). The GBS technique and other restriction-site-associated DNA (RAD) sequencing methods produce large quantities of reads that do not cover the entire genome but have higher sequencing depth, thus reducing sequencing errors (Gautier *et al.* 2013; Anand *et al.* 2016). Loci generated by GBS can be present in both coding and non-coding regions and will be shared between all populations owing to the conservation of restriction sites (Cariou *et al.* 2013). Finally, GBS does not require any prior genomic information for the species being studied, which is an important consideration for SCN since there is no reference genome yet. The optimal gene coverage to reduce the amount of missing data depends on the choice of restriction enzyme (Fu *et al.* 2016). Fortunately, optimal gene coverage was already tested for the closely related species *Globodera rostochiensis* (Mimee *et al.* 2015).

The main objectives of this study were to (1) investigate the genetic relationships among SCN populations from United States and Canada, (2) detect isolation by geographical distance (IBD) between SCN populations from United States and Canada, (3) detect genetic loci under selection associated to environmental and climatic parameters, and (4) identify the putative gene functions contributing to the adaptation of SCN populations to specific environmental conditions.

4.3 Materials and methods

Soybean cyst nematode populations sampling and genotyping-by-sequencing

A total of 64 field populations of SCN, representative of the area currently infested in North America, were sampled or provided by collaborators from 11 US states (Delaware, Iowa, Illinois, Indiana, Kansas, Michigan, Minnesota, Missouri, North Dakota, Ohio, South Dakota) and one Canadian province (Ontario) (Fig. 1). For DNA extraction, 40 cysts were randomly chosen from each population and pooled together. Eggs were extracted from each cyst and then washed twice in sterile filtered water. Total genomic DNA of each pool was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Mississauga, ON, Canada) following the manufacturer's instructions. DNA extracts were quantified using Qubit fluorometric quantification (ThermoFisher Scientific, Burlington, ON, Canada) and normalized at 2 ng/ μ L prior to library preparation and sequencing. These steps were performed following standard protocols (Elshire *et al.* 2011; Poland *et al.* 2012) at the Institute for Integrative and Systems Biology (IBIS; Université Laval, Québec City, QC, Canada). Genotyping-by-sequencing was performed using the method described by Mimee *et al.* (2015) with a combination of two restriction enzymes (PstI/MspI) (New England Biolabs, Whitby, ON, Canada). After the restriction enzyme treatment, all samples (one composite per field population) were barcoded and multiplexed to obtain a single library for the 64 samples, which was sequenced on three Ion Proton chips (ThermoFisher Scientific) at IBIS.

Single-nucleotide polymorphism calling

The UNEAK pipeline (Lu *et al.* 2013), which is part of the TASSEL 3.0 bioinformatics analysis package (Bradbury *et al.* 2007), was used to process the raw reads, since no complete reference genome is yet available for *H. glycines*. This pipeline is designed to call SNPs *de novo*, without a reference genome with high stringency. Only sequences containing a single putative SNP (1 bp mismatch) per sequencing read were kept. Before analysis, the SNPs were further filtered with a minimum coverage (minCov) of 20 reads, a maximum coverage (maxCov) of 10,000 reads, and a minimum allele frequency (MAF) of 0.01.

Population genetics

Clustering of SCN populations using PCA was carried out using the *prcomp* function from the *stats* package in R software (R Core Team, 2017). The *poppr* package (v2.4.1) in R (Kamvar *et al.* 2014) was used to investigate the genetic relationship between populations and to build a phylogenetic tree (Provesti's distance, 10,000 bootstraps, neighbour-joining algorithm counting missing data as equivalent in the distance computation). Visualization of sample coordinates and phylogenetic relationship analyses were carried out using the *phytools* 0.6-20 package in R (Revell 2017).

Fixation index (F_{ST}) values (Wright 1943) were calculated using the classical approach (Hartl and Clark 2007) with the PoPoolation2 software (Kofler *et al.* 2011) to evaluate the genetic differentiation between each pair of populations. Furthermore, the effect of isolation by distance on population structure was tested using the correlation between the genetic distance ratio [$F_{ST}/(1-F_{ST})$], as defined by Rousset (1997), and the geographic distance of population pairs in kilometers (km). The geographic distance between each sample location was calculated with the haversine formula using the *geosphere* package in R. To examine the significance of the relationship between the genetic distance ratio and the pairwise geographic distance distances, we performed a Mantel test (Spearman rank correlation) using the *vegan* package in R with 999 permutations. To evaluate the effect of time on population genetic differentiation, linear regression analyses were run in R software using pairwise F_{ST} distances based on a point of origin selected on the basis of the first reports of SCN in North America. In our dataset, the closest sample to the oldest population was located in Clarkton, Missouri (MO1) (Hegge 1957; Tylka and Maret 2017). To access the genetic isolation from the MO1 location, a Spearman rank correlation (r_s) test was performed at a 0.95 confidence level.

Populations from three states (North Dakota, Minnesota, Illinois) were selected as case studies to evaluate the local (short-scale) genetic differentiation. These states were chosen because there were sufficient samples for comparison and because (i) North Dakota shows a recent introduction of SCN and a continuous northward dispersal of the nematode (Nelson and Bradley 2003; Mathew *et al.* 2015); (ii) Minnesota has a longer history of SCN, with the nematode being first

detected in 1978 and the infested area still expanding each year (Zheng *et al.* 2006; Kurle *et al.* 2011); (iii) in Illinois SCN has been well established in every county for many years (Riggs 2004; Tylka and Marett 2014).

Outlier detection and their association with environmental variables

Two geographic and four climatic covariables were investigated as possible factors explaining loci under selection: latitude (LAT), longitude (LONG), annual mean air temperature (BIO1), maximum air temperature of the warmest month (BIO5), annual precipitation (BIO12), and total precipitation of the warmest quarter (BIO18). All climatic variables were retrieved from the WorldClim global climate database, version 1.4 (Fick and Hijmans 2017), corresponding to historic conditions (1960–1990). The spatial resolution used for the bioclimatic analysis was set at 30 seconds or 0.86 km².

To detect correlations between variations in population allele frequencies of SNPs and environmental factors, we used three different Bayesian methods (software programs): BayPass, version 2.1 (Gautier, 2015), BayeScan, version 2.1 (Foll and Gaggiotti 2008), and BayeScEnv, version 1.1 (de Villemereuil and Gaggiotti 2015). For each program, triplicate runs with different random seeds were performed with a pilot run of 10,000 iterations to estimate starting parameters, a burn-in length of 50,000 iterations, and a minimum of 50,000 iterations, accounting for the small number of SNPs being investigated. Post-run diagnostics were carried out using the *coda* package in R software in order to ensure sufficient iterations and normality of the Markov chain (Plummer *et al.* 2006).

The first software, BayPass, identifies genetic markers subject to selection by covariates such as phenotypic or environmental variables associated with the population of interest (Gautier 2015). This application is based on the BAYENV model proposed by Coop *et al.* (2010) and Günther and Coop (2013), but with several modifications detailed in Gautier (2015), including the reprogramming of the Markov chain Monte Carlo (MCMC) algorithm. For BayPass, we used a pool-size file with the `-d0yij` option set at 800 and 20 pilot runs with a length of 10,000 iterations, a burn-in length of 50,000 iterations, and a chain length of 50,000 iterations. For each SNP, BayPass generates a Bayes factor (BF), quantifying evidence against the null hypothesis, and an empirical Bayesian *p*-value (eBPis), a metric measuring the difference between observed

data and a simulated set of data (posterior distribution) (Kass and Raftery 1995; Andraszewicz *et al.* 2015; Gautier 2015). To be considered under selection, a SNP had to meet two criteria: a BF greater than 10 ($BF > 10$) and an eBPis lower than 0.05 ($eBPis < 0.05$).

The second software, BayeScan, uses a Bayesian likelihood method that assumes a Dirichlet distribution of allele frequencies between populations (Foll 2012). This program estimates the probability that each locus is subject to selection by using a logistic regression on the two locus-population F_{ST} coefficients. This Bayesian method uses a reversible-jump MCMC algorithm to calculate a posterior probability that each locus is under selection. The decision criterion to determine whether a locus is likely to be under a strong selection is the q -value (Foll 2012), analogue to a false discovery rate (FDR) p -value, that must be under 0.05. A second decision criterion was applied to further endorse the selected outliers: the ratio of posterior probabilities (PO). The PO threshold to affirm that a locus was under selection, in comparison with a neutral model, was set to 0.91, which corresponds to a strong relationship on the Jeffreys scale (Foll 2012). We used BayeScan with the default parameters, but we set the minimum number of iterations to 50,000, the length of 20 pilot runs to 10,000 iterations, and the burn-in length to 50,000 iterations.

The third software, BayeScEnv, is similar to BayeScan and uses the F_{ST} index to detect loci with a high level of differentiation in comparison with the entire genome. This program allows a normalization vector to be applied to the environmental data instead of only a binary combination, thus generating a lower number of false positives, according to the creators of the software (de Villemereuil and Gaggiotti 2015). We used BayeScEnv with the default parameters, with the number of iterations set to 50,000 and 20 pilot runs with a length of 10,000 iterations. We used the reported q -value, which is related to the FDR, as our decision criterion, considering only SNPs with a q -value less than 0.05.

Localization of outlier loci, gene function, and genic environment

The SNPs identified in short reads were first retrieved from a draft SCN genome available from SCNBase (<https://www.scnbase.org/>) by means of BLASTN with the default parameters, except for a smaller word size of 4, with the Blast2GO application (Conesa *et al.* 2005). Many of the identified SNP-containing fragments matched multiple genes or genome locations (see

supplementary tables 3 to 5). To assign a putative gene function to each SNP, we compared the aligned sequences to the National Center for Biotechnology Information (NCBI) protein database by means of BLASTX and BLASTP (Altschul *et al.* 1990) on a subset of sequences (nematodes, taxid: 6231) or to all of the NCBI non-redundant (nr) sequence database with an *E*-value significance cutoff of $1e^{-5}$.

As our GBS sequencing covers approximately 0.8% of the genome (see Results section), we explored the genomic regions around outlier loci for genetic hitchhiking in whole-genome sequences from four populations of different origin and distinct virulence profile (ON1, ON34, IL4, and KS2). Sequencing libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA) and sequenced on a MiSeq sequencer (Illumina) using the MiSeq Reagent Kit v3 (600 cycle). Reads were demultiplexed using the Sabre software program (<https://github.com/najoshi/sabre>) and processed with Trimmomatic (v0.32) (Bolger *et al.* 2014) to remove adapters and barcodes. Alignment on the draft reference genome (see above) was done using SAMtools (v0.1.19) (Li *et al.* 2009) and BWA (v0.7.12) (Li and Durbin 2009). Only a window of ± 50 kb around the 15 loci under selection was retained for the genic environment analysis. Variants were called with freebayes (v1.0.2) (Garrison and Marth 2012) and snpeff (v4.2) (Cingolani *et al.* 2012). Distribution of allele frequencies in these four populations for each gene variant was compared with the allele frequency of the corresponding loci under selection in the same population in the genotyping-by-sequencing dataset. Only SNPs exhibiting a similar allele frequency and having a high impact on the predicted protein product were retained in our analysis.

4.4 Results

Genotyping-by-sequencing

A total of 192,576,709 short reads were obtained from the sequencing of the DNA from the 64 SCN populations, following digestion by the PstI/MspI restriction enzymes. The *H. glycines* effective genome length was estimated to be 96,752,286 bp, with an average of one PstI restriction site every 12,688 bases (E. Lord, private communication, 2017). On the basis of the sequence size used by the UNEAK (Universal Network-Enabled Analysis Kit) pipeline (first 64

bp of each read), the expected horizontal coverage was approximately 0.8% of the total SCN genome, and the vertical coverage at each locus, considering the number of reads obtained, was 400×. Before filtering, the UNEAK pipeline identified 3,172 variants. After filtering for low coverage and selecting only SNP variants, two datasets were generated. The first dataset contained 245 SNPs without missing data (loci sequenced in all populations), and the second contained 804 SNPs with missing data.

Population structure

Principal component analysis

A principal component analysis (PCA) was performed using the dataset without missing data, which contained 245 SNPs. The PCA plot of the 64 North American SCN populations revealed a geographically ordered pattern (figure 4.1A).

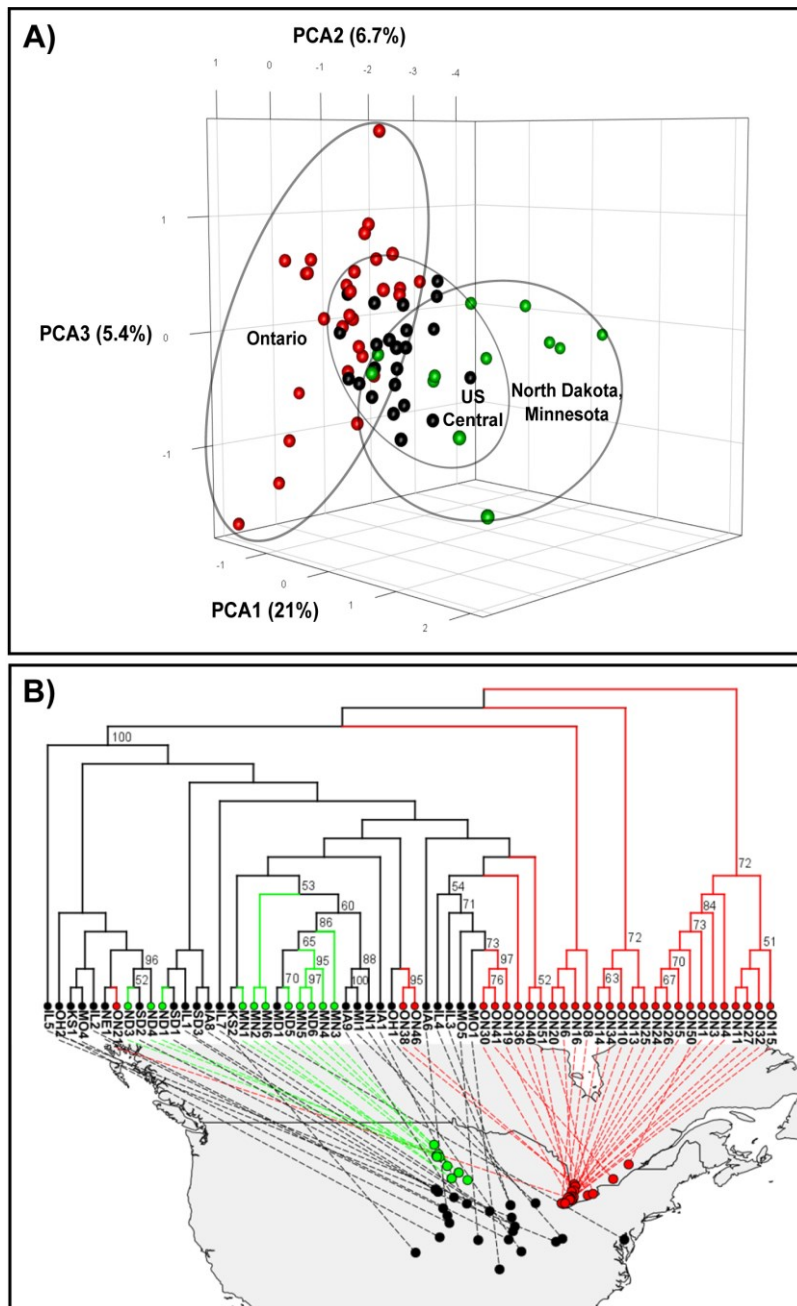


Figure 4.1 Genetic relationship among North American populations of the soybean cyst nematode. (A) Three-dimensional principal component analyses (PCAs). (B) Phylogenetic tree plot based on Provesti's absolute genetic distance. Significant bootstrap values (> 50%) are indicated. Soybean cyst nematode populations from Ontario (Canada), North Dakota and Minnesota, and central US states are indicated in red, green and black, respectively.

Overall, the Ontario populations showed the greatest dispersion, indicating they are more genetically differentiated, while the central regions of the US showed more clustered populations. Also, the group containing populations from Minnesota and North Dakota (in green in figure 4.1A) was clearly different from the Ontario populations by the first axis, which explained 21%

of the total variation. All populations from the central states, although they originated from many states and covered a much wider area, were less diverse and more clustered together.

Phylogenetic tree

To better understand the population structure of each population, we used the 804 SNP dataset to conduct a phylogenetic analysis based on Provesti's distance. The inferred neighbour-joining phylogenetic tree (figure 4.1B) showed that most of the Ontario samples (in red in figure 4.1B) were different from the rest of the North American populations. Some samples from Ontario were found to be different from each others and more similar to those of central states. Such a pattern can also be observed in the PCA analysis (figure 4.1A).

Genetic differentiation

Overall, we observed an average value for Wright's fixation index (F_{ST}) of 0.15 ± 0.07 . However, pairwise genetic distance was highly variable, with a minimum F_{ST} of 0.005 and a maximum F_{ST} of 0.53 (supplementary table 1). The highest F_{ST} value was found between two of the most remote populations, ND6 from North Dakota and ON41 from Ontario, which were separately by approximately 1,720 km. On the other hand, the lowest F_{ST} value was observed between ON3 and ON5, which were separated by only 17.9 km. In general, higher F_{ST} values were found in distant populations, especially between populations from Minnesota (MN4 and MN5), North Dakota (ND3 and ND6), South Dakota (SD2), and Ontario (ON6, ON13, ON25, ON41, and ON46).

Isolation by distance

The analysis of the relationship between genetic distance [$F_{ST}/(1 - F_{ST})$] and geographic distance (km) for all pairs of populations revealed a significant ($p = 0.012$) but weak effect of isolation by distance, with a Mantel correlation (r) of only 0.135. However, when analyzing the effect of isolation by distance at the local scale, a different pattern was found. With respect to US states with recent SCN history, a strong and significant positive correlation was observed between the genetic and geographic distances in North Dakota ($r = 0.430$, $p = 0.013$) and Minnesota ($r = 0.730$, $p = 0.006$). In contrast, no correlation was found among the populations from Illinois ($r = -0.170$, $p = 0.767$).

To further explore the contribution of isolation by distance to the genetic differentiation of United States and Canada populations of SCN, we compared the genetic distance of each population from a theoretical ancestor population (MO1) using the 245 SNP dataset. This analysis included 38 independent populations that were separated by 273 to 1,367 km from MO1 (figure 4.2A).

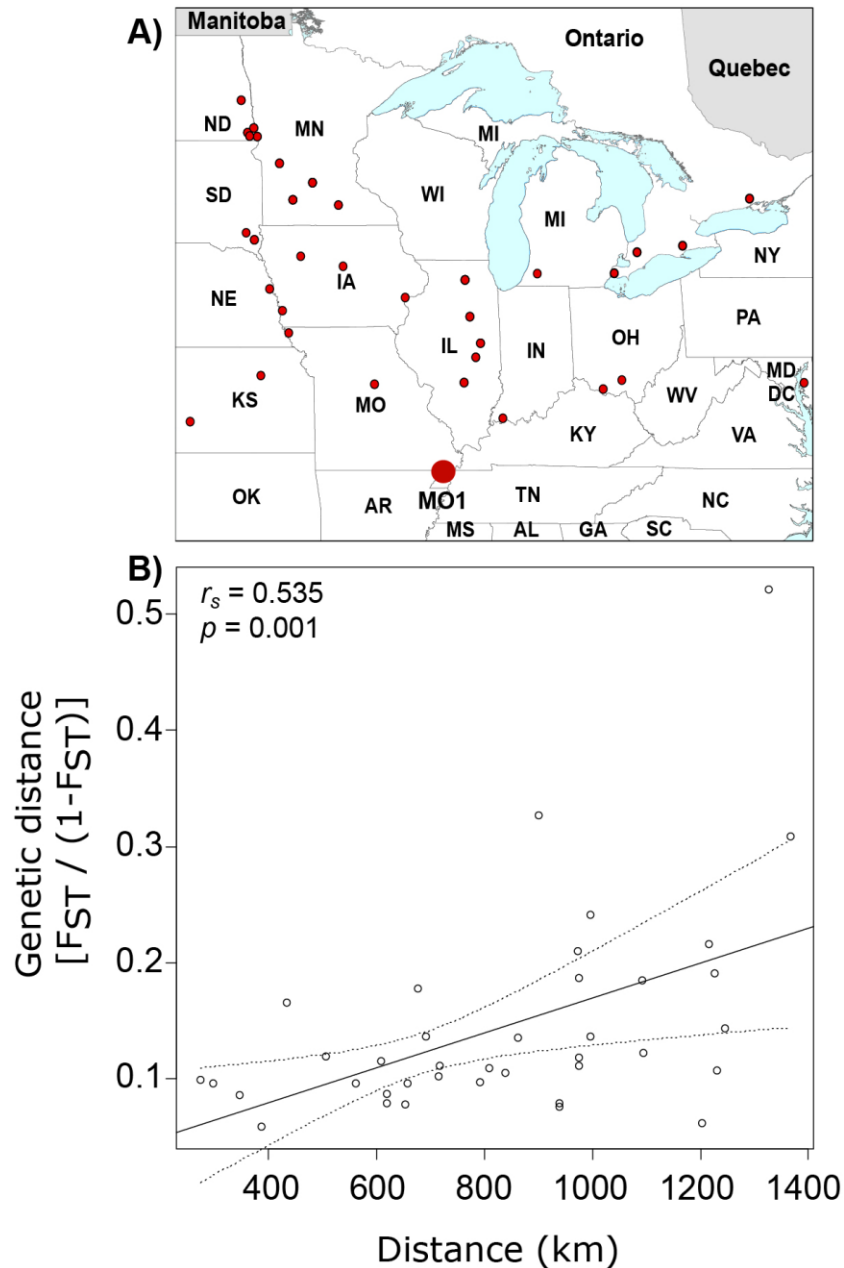


Figure 4.2 Distribution of soybean cyst nematode populations used in this study (small dots) and identification of the “oldest” sample in the dataset (MO1, large dot) (A) as well as the relationship between genetic distance $[F_{ST}/(1 - F_{ST})]$ and geographic distance (B). Linear regression line is shown with 95% confidence intervals estimated using 1000 bootstrap replicates.

The genetic-distance ratio to the “oldest population” varied from 0.06 for IL1 (386 km apart) to 0.52 for ND6 (1,327 km apart) (supplementary table 2) and an overall moderate ($r_s = 0.535$) but significant ($p = 0.001$) relationship was observed between genetic distance and geographic distance (figure 4.2B).

Signature of genetic adaptations

Genome scans

Using three Bayesian inference genome scan approaches, we identified 71 SNPs (out of 804) that were under selection. Among the different computational methods, 48 loci were identified by BayeScan, 23 loci were identified by BayPass, and 16 loci were identified by BayeScEnv (supplementary tables 3 to 5). Out of those 71 SNPs, only one was highlighted by the three methods, while 15 outliers were inferred by at least two of the methods (fig. 4.3).

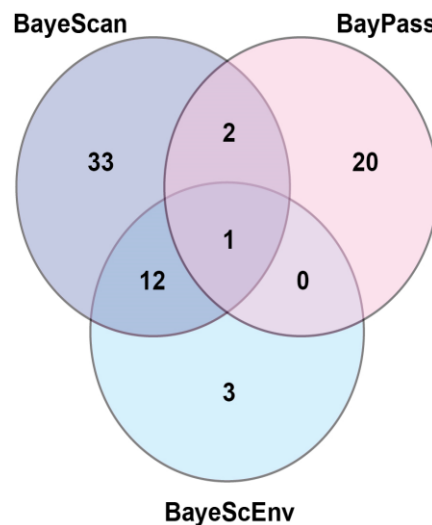


Figure 4.3 Number of outlier loci identified by three different Bayesian genome scan methods in 64 populations of the soybean cyst nematode from North America.

The geographic variables (LAT and LONG) were associated with seven and eight outlier loci, respectively, and the climatic variables (BIO1, BIO5, BIO12, and BIO18) were associated with nine, six, eight, and seven SNPs, respectively (table 4.1).

Table 4.1 Association of outlier loci (single-nucleotide polymorphisms [SNPs]) identified by at least two Bayesian genome scan methods in 64 populations of the soybean cyst nematode from North America with geographic or climatic co-variables. The values presented are the p-value for BayeScan and BayeScEnv and Bayes Factor for BayPass.

SNP ID	Method	Latitude	Longitude	Temperature (annual)	Temperature (maximum)	Precipitation (annual)	Precipitation (warmest quarter)
TP66981	BayeScan		0.998	0.995			
	BayeScEnv		0.017	0.042			
	BayPass						
TP77825	BayeScan				0.999	0.941	
	BayeScEnv				0.029		
	BayPass						
TP146577	BayeScan	0.952	1.000		1.000	0.997	0.997
	BayeScEnv		0.047		0.038		
	BayPass						
TP173057	BayeScan	0.988					
	BayeScEnv						
	BayPass						13.86
TP181514	BayeScan	0.996	0.940	0.992	0.961		

	BayeScEnv		0.050	0.009			
	BayPass						
TP188858	BayeScan					0.999	
	BayeScEnv					0.050	
	BayPass						
TP226227	BayeScan		1.000			1.000	0.956 0.999
	BayeScEnv					0.044	
	BayPass						
TP252086	BayeScan			0.963			
	BayeScEnv			0.050			
	BayPass						
TP324827	BayeScan	0.999	0.999	0.999		1.000	0.999
	BayeScEnv	0.043	0.023	0.006		0.017	
	BayPass						
TP333577	BayeScan	0.984	0.978			0.999	1.000
	BayeScEnv		0.029	0.015			0.040
	BayPass						14.32
TP333913	BayeScan	0.931	0.978	0.976		0.930	

	BayeScEnv		0.029		0.015			
	BayPass							
TP364091	BayeScan				0.976			
	BayeScEnv				0.020			
	BayPass							
TP376687	BayeScan	0.993	0.999		1.000		0.999	0.999
	BayeScEnv				0.038			
	BayPass							
TP380020	BayeScan				0.998		1.000	
	BayeScEnv				0.030		0.027	
	BayPass							
TP380188	BayeScan				0.968			
	BayeScEnv							
	BayPass							13.97

Three SNPs were strictly associated with only one environmental variable: TP188858 was associated with maximum temperature, while TP252086 and TP364091 appeared to be linked to annual temperature. The allele frequencies analysis at these 15 loci revealed SNPs that were specific for a given region, such as TP364091, TP181514, and TP333913 for Minnesota and North Dakota and TP376687, TP146577, and TP226227 for Ontario (figure 4.4).

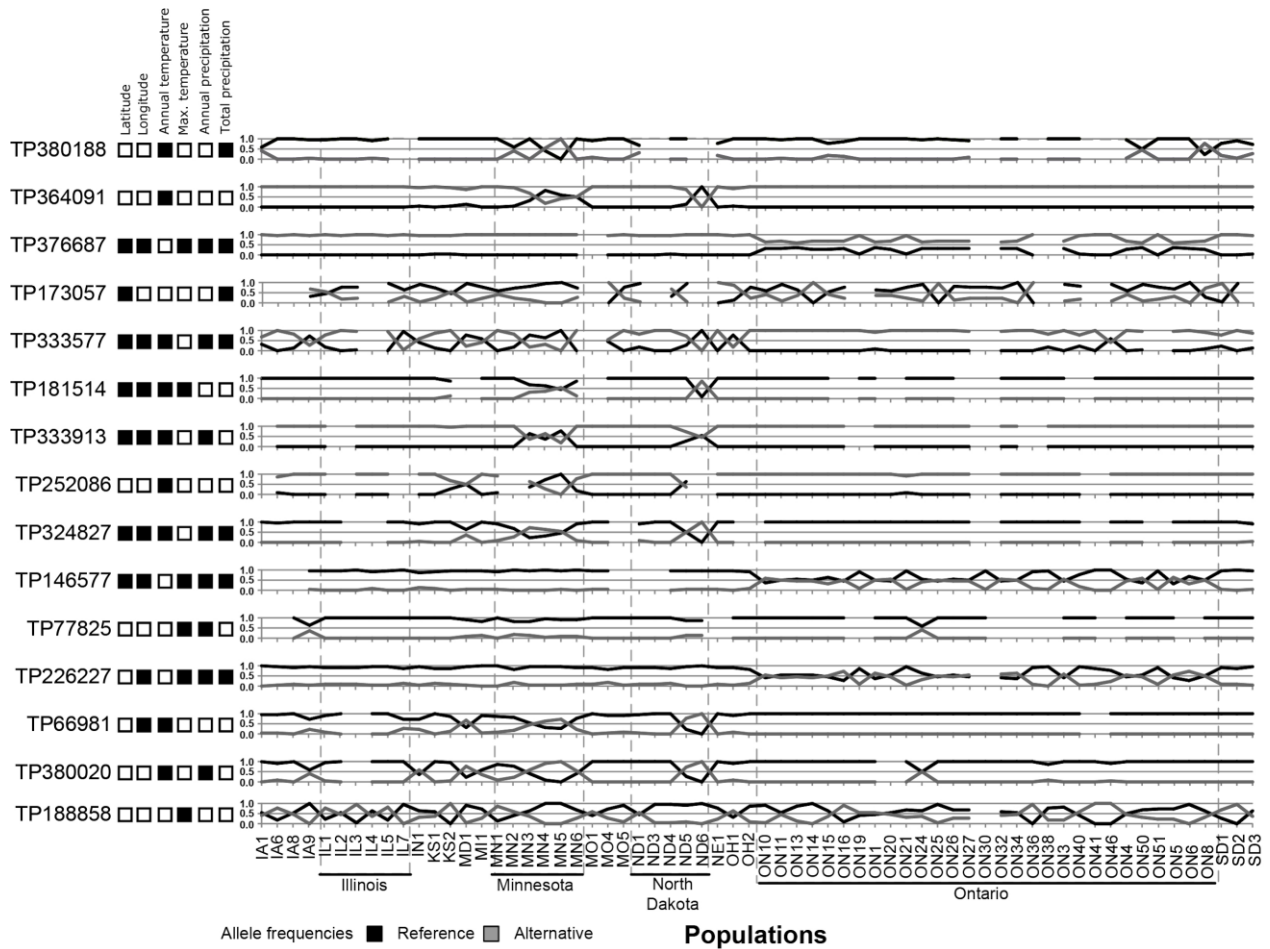


Figure 4.4 Allele frequencies of 15 outlier loci identified in North American populations of the soybean cyst nematode and associations of the loci with geographic or climatic co-variables. Black squares represent a significant association.

Localization, effect, and genic environment of outlier loci

Of the 15 SNPs under selection identified by at least two Bayesian approaches, 10 SNPs were located in predicted genes (seven in exons and three in introns), and five SNPs were located in intergenic regions (table 4.2).

Table 4.2 Identification and effect of outlier loci (single-nucleotide polymorphisms [SNPs]) identified by at least two Bayesian genome scan methods in 64 populations of the soybean cyst nematode from North America.

ID	Location	S or N-S ¹	SNP	a.a. subst. ²	E-value	Best annotation
TP380188	Exon	N-S	C/A	Gln/Lys	2.0e-20	SMAD-4 [<i>Strongyloides ratti</i>]
TP364091	Exon	N-S	G/A	Asp/Asn	2.6e-104	Dopamine receptor 3 [<i>Toxocara canis</i>]
TP376687	Exon	S	T/G	Gly/Gly	4.0e-11	Putative esophageal gland cell protein Hgg-20 [<i>Heterodera glycines</i>]
TP173057	Exon	S	C/T	Ile/Ile	2.0e-11	Leucine-rich repeat-containing domain protein, partial [<i>Necator americanus</i>]
TP333577	Exon	S	C/T	Leu/Leu	1.0e-110	Cadherin-4 [<i>Caenorhabditis elegans</i>]
TP181514	Exon	S	T/C	Gly/Gly	1.6e-100	LET-19 [<i>Caenorhabditis brenneri</i>]
TP333913	Exon	S	T/C	Gly/Gly	1.6e-7	EGL-17 [<i>Brugia malayi</i>]
TP252086	Intron	-	G/A	-	2.0e-64	Esophageal gland-localized secretory protein 11 [<i>Heterodera glycines</i>]
TP324827	Intron	-	C/T	-	3.0e-32	Reverse transcriptase [<i>Ancylostoma ceylanicum</i>]
TP146577	Intron	-	T/A	-	2.0e-11	Hypothetical protein Y032_0209g2110 [<i>Ancylostoma ceylanicum</i>]
TP77825	Intergenic	-	T/A	-	2.0e-57	Nearest gene : Hypothetical protein CRE_19487 [<i>Caenorhabditis remanei</i>]
TP226227	Intergenic	-	T/G	-	3.0e-05	Nearest gene : ATP-dependent DNA helicase PIF1, partial [<i>Toxocara canis</i>]
TP66981	Intergenic	-	T/G	-	-	No significant hit
TP380020	Intergenic	-	G/T	-	-	No significant hit
TP188858	Intergenic	-	G/A	-	-	No significant hit

¹Synonymous (S) or non-synonymous (N-S) mutations in the final protein product.

²a.a., amino acid substitution.

Two SNPs (TP380188 and TP364091) correspond to non-synonymous mutations that induce a change in the amino acid sequence of the resulting protein. The annotations of the genes putatively impacted by these modifications are *SMAD-4* (mothers against decapentaplegic homolog 4) for TP380188 and *DOP-3* (dopamine receptor 3) for TP364091. The sequence corresponding to TP364091 was retrieved in four different genes (g14639.t1, g14642.t1, g14644.t1, and g14654.t1) in the draft reference genome and introduced two different mutations (supplementary tables 3 to 5). All the other SNPs (13/15) were silent mutations (synonymous in exons or located in introns or intergenic regions) and thus probably not the cause of their selection. Among the 42 remaining SNPs that were under selection but identified by only one pipeline, 12 correspond to a non-synonymous modification impacting the predicted protein sequence. These SNPs include the gene *UFMI* (Ubiquitin-fold modifier 1), the gene *PLA2* (85/88 kDa calcium-independent phospholipase A2), and a pyruvate kinase (supplementary tables 3 to 5).

The whole-genome resequencing of four SCN populations for the exploration of the genic environment of outlier loci highlighted 257 genes containing high-impact variants in a ± 50 -kb window around loci under selection. Annotation was available for 148 of these genes, corresponding to 132 different gene functions (supplementary table 6). Of these genes, 25 exhibited allele frequencies similar to their neighbouring outlier loci (table 4.3).

Table 4.3 Putative hitchhiking gene variants with high-impact neighbouring loci under selection and exhibiting a similar allele frequency in *Heterodera glycines*.

ID	Gene	Best annotation	E-value ¹	Type of variants ²
TP146577	g15502.t1	G-protein-coupled receptor, partial [Pristionchus pacificus]	9.10e-57	complex, insertion
TP173057	g18365.t1	85 kda calcium-independent phospholipase a2 [Ascaris suum]	8.50e-46	SNP, insertion
TP173057	g18358.t2	F-box only protein 30 [Ascaris suum]	1.10e-33	4 complexes
TP173057	g18327.t1	Protein CBR-CLEC-223 [Caenorhabditis briggsae]	5.70e-57	SNP

TP173057	g18328.t1	Protein lap1 [Toxocara canis]	1.80e-104	6 SNP, 3 insertions, 2 complexes
TP173057	g18332.t1	Protein scribble-like protein, partial [Stegodyphus mimosarum]	1.00e-33	2 deletions, MNP
TP181514	g3633.t1	Mediator of RNA polymerase II transcription subunit 13-like protein [Ascaris suum]	1.20e-133	Deletion
TP226227	g1982.t1	85/88 kDa calcium-independent phospholipase A2 [Toxocara canis]	8.00e-29	SNP
TP226227	g14545.t1	Glycoside hydrolase domain containing protein [Haemonchus contortus]	1.10e-162	deletion, deletion
TP226227	g12205.t1	Kazal domain and organic anion transporter polypeptide (OATP) family and MFS [Strongyloides ratti]	3.00e-36	deletion, 3 complexes
TP226227	g1979.t1	Serine/threonine-protein kinase haspin [Haemonchus contortus]	1.40e-62	insertion
TP252086	g18812.t1	Bestrophin/UPF0187 family-containing protein [Strongyloides ratti]	2.30e-44	deletion
TP252086	g18825.t1	mtN3/saliva family protein [Oesophagostomum dentatum]	1.30e-55	deletion, SNP, complex
TP333577	g16696.t2	AFG3-like protein 2 [Toxocara canis]	1.30e-286	complex, 2 insertions
TP333913	g14633.t1	Cytohesin-1 [Trichinella zimbabwensis]	1.30e-118	SNP
TP333913	g14626.t1	Fibroblast growth factor 17 [Toxocara canis]		deletion
TP333913	g14617.t1	Immunoglobulin domain containing protein [Haemonchus contortus]	9.60e-42	deletion
TP333913	g14614.t1	Sodium-dependent acetylcholine transporter [Toxocara canis]	1.50e-215	SNP
TP364091	g14633.t1	Cytohesin-1 [Trichinella zimbabwensis]	1.30e-118	SNP
TP364091	g14656.t1	Major facilitator superfamily MFS-1 domain containing protein, partial [Haemonchus contortus]	2.00e-106	deletion
TP380020	g7291.t1	Carboxypeptidase [Eupolyphaga sinensis]	1.00e-60	SNP

TP380020	g8313.t1	Cell Division Cycle related [Caenorhabditis elegans]	1.10e-254	deletion, SNP
TP380020	g8311.t1	GPI ethanolamine phosphate transferase 3, partial [Trichinella papuae]	1.90e-125	deletion
TP380020	g8314.t3	Homeobox protein ceh-6 [Toxocara canis]	4.60e-52	insertion, 2 deletions
TP380020	g8309.t1	Immunoglobulin I-set domain containing protein [Ascaris suum]	5.50e-31	2 deletions
TP380020	g8305.t1	RecQ-mediated genome instability protein 1 [Toxocara canis]	4.60e-28	SNP
TP380188	g2288.t1	Cleavage polyadenylation specificity factor domain containing protein [Haemonchus contortus]	3.90e-209	insertion
TP380188	g2287.t1	Cytochrome b-c1 complex subunit 2, mitochondrial [Strongyloides ratti]	1.20e-63	deletion, SNP

¹Expected value for annotations of each putative hitchhiking gene.

²Variant in bold indicates a stop codon gain. More details on types of variants and position in the gene are provided in Supplementary Table 6. SNP, single-nucleotide polymorphism; MNP, multiple-nucleotide polymorphism.

4.5 Discussion

Since its introduction into North America in 1954, the soybean cyst nematode has spread at a regular pace to most of the soybean-producing areas in the US and reached Canada in 1987. Its northern limit of establishment remains to be determined, as new populations are found each year. Soybean acreage is rapidly increasing in northern latitudes; for example, farmers in the province of Manitoba, Canada, sowed 2.3 million acres with soybean in 2017, a 40% increase in one year in a province where soybean was not grown a decade ago (Statistics Canada 2017). This spectacular change in land use results from a coordinated breeding program for short-season and drought-tolerant soybean cultivars (Tardivel *et al.* 2014). Although soybean is nowadays very profitable in Canada (Manitoba, Northern Ontario, Québec), managing a pest such as SCN could be challenging, as no resistant cultivars are currently adapted to these regions. Simulations based on thermal development have shown that SCN could theoretically survive in these new soybean growing areas (Gendron St-Marseille 2013), but actual establishment is a different matter, as

initial observations in the province of Québec indicated that even though the nematode is detected in many regions, populations weakly reproduced (Mimee *et al.* 2016). This phenomenon probably reflects SCN's very recent history in this part of the world and the poor fitness of the introduced populations but does not exclude possible adaptations of the nematode in the future. To better understand the evolution of SCN population genetics in North America, we compared 64 populations originating from 11 US states and the province of Ontario in Canada.

The dynamics of cyst nematode populations are intrinsically complex. Several thousand cysts from different females can form on a single plant, and each cyst contains hundreds of half-sibling individuals as a result of SCN's polyandrous mode of reproduction. Thus, the number of unique genotypes occurring in a field is massive. Isolating individual nematodes, even hundreds of them, to explore the genetic relationship between populations would be an arduous task and the approach would still be biased owing to limited sub-sampling. For these reasons, we opted for a Pool-Seq (sequencing of pooled DNA samples) approach. Sequencing DNA from pooled samples for each population also has the advantage of keeping the number of redundant sequences low (Schlötterer *et al.*, 2014). However, this approach does not allow the assignment of sequences to individuals and prevents some population genetic analyses such as linkage-disequilibrium estimation (Lynch *et al.*, 2014, Anand *et al.*, 2016). Nevertheless, a number of recent studies have overcome these limitations and have proposed new methods to analyze genetic variations among populations based on Pool-Seq data (Van Tassel *et al.*, 2008; Gautier *et al.*, 2013, Navon *et al.*, 2013; Mimee *et al.*, 2015; Anand *et al.*, 2016).

Both the PCA and phylogenetic analyses showed strong clustering of populations based on their geographic origin, which supports the “nearest-neighbour theory” movement as in the stepping-stone model (Kimura and Weiss 1964; Hutchison and Templeton 1999). We observed a geographic separation dividing SCN populations into two clusters, a northeastern one (Ontario) and a northwestern one (Minnesota, North Dakota) that both share genetic similarities with central populations. In northeast and northwest regions, SCN populations established only recently, and new detections continue to occur in the northern areas. In the central states, SCN populations have been present for a longer time and resistant soybean cultivars are routinely used. We believed that these two conditions contribute to explain the homogeneity of the central

populations. On one hand, the continuous exchange of genetic material would have led to the homogenization of alleles in sympatric populations. On the other hand, the massive use of nematode resistant cultivars that are all derived from a single source of resistance (PI 88788) would have selected for the virulent SCN genotypes, thereby limiting the propagation of some alleles. It is thus surprising to observe a greater genetic differentiation in SCN populations from Ontario since the main assumption is that they originate from central states. These populations should contain less diversity, owing to the founder effect, which would normally result in reduced potential for differentiation. However, the selective pressures mentioned above are not yet present in Ontario, allowing the development of most genotypes. This finding concurs with the results of Faghihi *et al.* (2010), who observed a much greater diversity in SCN population phenotypes (HG types) in Ontario than in Tennessee, Indiana, or Illinois. A similar pattern was obtained in Minnesota by Zheng *et al.* (2006) and in China by (Liu *et al.* 1997) when comparing populations from the north to populations from central regions. However, all these authors suggested that climatic conditions or local environmental factors may also play a role.

According to Hartl and Clark (2007), the overall average F_{ST} value of 0.15 that we obtained by comparing all pairs of SCN populations from North America corresponds to a high level of differentiation. However, when we examined specific pairs of populations, we found that the populations were very similar at the local scale. Also, the differentiation of recent populations (Ontario, Minnesota, North Dakota) from those of the central states was only moderate, whereas comparing these recent populations together generated very high F_{ST} values (up to 0.53). This corroborates results from the PCA analysis and clearly suggests that populations diverged following two distinct routes, one in the northwest toward North Dakota and one in the northeast toward Ontario. This hypothesis is further supported by the isolation-by-distance analysis. Overall, we observed a significant effect of distance on genetic differentiation at the continental scale. The effect of distance was even more noticeable in newly colonized areas like North Dakota and Minnesota or when each population was compared to a common ancestor. On the other hand, the absence of any effect of distance in Illinois and low F_{ST} in the central states clearly suggests that dispersal is frequent enough to weaken signs of isolation by distance after a short period of time. This continuous gene flow combined with the selection of alleles due to the general use of the same resistant soybean lines are thought to cause a rapid homogenization of

SCN populations (Mitchum *et al.* 2007; Niblack *et al.* 2008; Zheng and Chen 2011). If SCN is already adapted to environmental conditions prevailing in northern areas, this homogenization process would accelerate the development of virulent genotypes in these regions (in comparison with the central states) and would be independent from the use of resistant cultivars. This concurs with the finding of Faghihi *et al.* (2010) where 15% of SCN populations in Ontario were able to multiply on PI 548402 and 6% were able to multiply on PI 90763, even though these sources of resistance were not present in commercial cultivars in Ontario.

Although 71 SNPs were found to be under selection among SCN populations, only one SNP was identified by the three Bayesian approaches tested. This SNP (TP333577) was strongly associated with precipitation; but the geographic distribution of allele frequencies (fig. 4.4) at that specific locus did not match with the general pattern of genetic differentiation, instead showing a very local effect (in populations from Minnesota but also in some populations from Iowa, Illinois, North Dakota, and Ohio). Thus, this selection for wetness was either lost in favour of a more stringent pressure or too costly to be maintained in the other regions. Overall, considering the broad range of environmental conditions in which these populations develop, finding only a few genes under selection suggests that SCN possesses an intrinsic capacity to evolve within a large range of temperature and wetness gradients.

Most of the 15 outliers identified by at least two pipelines correlated with temperature, reinforcing that temperature is a key factor that acts on the genetic diversity and natural selection of nematode populations. Nematodes are poikilothermic organisms and depend greatly on temperature to complete their life cycle. Thus, any adaptation that facilitates the development and reproduction of nematodes at lower temperatures would be advantageous. On the basis of the distribution of allele frequencies, however, this adaptation does not seem crucial for survival, as the northern populations did not all exhibit similar patterns. This trait is probably not yet fixed in these populations. Nevertheless, six SNPs under selection and associated with temperature were specific to populations from Ontario (TP376687, TP146577, and TP226227) or only retrieved from Minnesota and North Dakota (TP364091, TP181514, and TP333913), a finding that clearly suggests local adaptations.

The SNP corresponding to TP364091 was retrieved in four out of six populations from Minnesota and in one population from North Dakota. That SNP, which is a non-synonymous mutation in the gene *DOP-3* (dopamine receptor 3) that replaces an aspartic acid with an asparagine, was strongly associated with the annual mean temperature and annual precipitation. This modification was retrieved in two distinct gene sequences (g14642.t1 and g14654.t1), and a synonymous mutation (Pro/Pro) was retrieved in two other sequences (g14639.t1 and g14644.t1) but in a different amino acid, indicating a reading frame shift and a putative upstream indel. These four genes were all annotated as *DOP-3* and were probably the result of gene duplication. Gene duplication has frequently been hypothesized to play an important role in adaptation to the environment (reviewed in Kondrashov 2012). In *Caenorhabditis elegans*, the *DOP-3* protein is well known to modulate chemosensory functions, such as mating and foraging, and to be involved in locomotion (Chase *et al.* 2004; Wood and Ferkey, 2016). The other SNP that introduces a non-synonymous mutation and was highlighted by two genome scan methods, namely, TP380188, was located in the gene *SMAD-4* (mothers against decapentaplegic homolog 4) and was retrieved in only a few populations from Minnesota, North Dakota, and Ontario. The *SMAD-4* protein is a regulator of different cellular processes, including cell differentiation, apoptosis, migration, and proliferation (Nikolic *et al.* 2011). Mutations of *SMAD-2*, *SMAD-3*, *SMAD-4*, or *SMAD-6* were previously shown to result in a 30% reduction in body size in *C. elegans* (Savage-Dunn *et al.* 2003; Watanabe *et al.* 2007). Also of interest is the locus TP122360, identified only by BayPass, that was located in the UFM1 (ubiquitin-fold modifier 1) genes (g7250.t1 and g7228.t1). A deletion in this gene in *C. elegans*, although reducing reproduction rate and life span, increased the survival of this nematode under oxidative or heat stress (Hertel *et al.* 2013).

For the majority of the loci, their selection probably results from genetic hitchhiking rather than a direct contribution to adaptation. To explore that possibility, we analyzed the genic environment in a ± 50 kb window around each SNP in four SCN populations. Genes containing high-impact genetic variations and exhibiting the same allele frequencies as the associated SNP under selection were of particular interest. Two of these genes (g18327.t1 and g15502.t1) coded for G-protein-coupled receptors (GPCR). These proteins are members of a large and very diverse multigene family with hundreds of occurrences in the *C. elegans* genome (Bockaert and Pin

1999; Bargmann 2006). The GCPR are crucial in sensing the local environment and were shown to evolve following alterations in habitat or foraging behaviour (Nei *et al.* 2008). Four genes coding for proteins implicated in metabolite transport into the cell were also identified (g12205.t1, g14656.t1, g18812.t1, and g18825.t1), as was one gene involved in the regulation of those proteins (g18365.t1). All these proteins are required to maintain homeostasis in the cell and to respond to local environmental changes. In a similar study using the tree *Fagus sylvatica*, Pluess *et al.* (2016) found that a version of a potassium transporter was associated with lower precipitation and could contribute to the regulation of growth under dry conditions. Structural changes in the body of an organism as a result of microevolution can also confer a significant advantage in terms of resisting more adverse environmental conditions, such as drought or high temperatures (Hazel and Williams, 1990). In our study, mutations were observed in two genes involved in cell-membrane and cell-wall stability (g18328.t1 and g8311.t1). Lastly, modifications in genes involved in the regulation of transcription (g2288.t1 and g3633.t1) and in the maturation of proteins (g14633.t1 and g18358.t2) have the potential to radically change the proteome of the adapted organism.

Although some SCN genes were found to be under selection and local adaptation was found to be underway in this study, our results also indicate that there is no critical adaptive event required for SCN establishment in northern latitudes. Consequently, all populations should theoretically survive and multiply at high latitudes. The risk is thus real for new soybean areas where cultivars resistant to SCN are not available at this time. Of course, the method we used does not explore the entire genome, and a pan-genomic study of these populations could reveal other loci under selection in an evolutionary process.

4.6 Acknowledgements

The authors thank Drs. Thomas Baum and Andrew Severin for graciously making SCN genome sequences available to us before public release on scnbase.org. The authors also thank Tom Welacky (AAFC) and Terry Niblack (OSU) for providing SCN populations. The authors wish to acknowledge the contribution of Génome Québec, Genome Canada, the Government of Canada, the Ministère de l'Économie, Science et Innovation du Québec, Semences Prograin Inc., Syngenta Canada Inc., Sevita Genetics, La Coop Fédérée, Grain Farmers of Ontario,

Saskatchewan Pulse Growers, Manitoba Pulse & Soybean Growers, the Canadian Field Crop Research Alliance and Les Producteurs de grains du Québec.

4.7 Data Availability

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher. All the SNP sequences and read count used in this study are listed in supplementary table 7.

4.8 Supplementary Materials

The excel data file is accessible at :

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6056837/bin/Data_Sheet_1.XLSX

4.9 References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment searchtool. *J. Mol. Biol.* 215(3):403–410. doi:[10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Anand, S., Mangano, E., Barizzone, N., Bordoni, R., Sorosina, M., Clarelli, F., *et al.* (2016). Next generation sequencing of pooled samples: Guideline for variants’ filtering. *Sci. Rep.* 6:33735. doi:10.1038/srep33735
- Anderson, T.R., Welacky, T.W., Olechowski, H.T., Ablett, G., Ebsary, B.A. (1988). First report of *Heterodera glycines* on soybeans in Ontario, Canada. *Plant Dis.* 72(5):453. doi:10.1094/PD-72-0453C
- Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., Wagenmakers, E-J. (2015). An introduction to Bayesian hypothesis testing for management research. *J. Manag.* 41(2):521–543. doi:[10.1177/0149206314560412](https://doi.org/10.1177/0149206314560412)
- Bargmann, C.I. (2006). Comparative chemosensation from receptors to ecology. *Nature.* 444:295–301. doi:10.1038/nature05402
- Blok, V.C., Phillips, M.S., Harrower, B.E. (1997). Comparison of British populations of potato cyst nematodes with populations from continental Europe and South America using RAPDs. *Genome.* 40(3):286–293.

- Bockaert, J., Pin, J.P. (1999). Molecular tinkering of G protein-coupled receptors: An evolutionary success. *EMBO J.* 18(7):1723–1729. doi:[10.1093/emboj/18.7.1723](https://doi.org/10.1093/emboj/18.7.1723)
- Bolger, A.M., Lohse, M., Usadel B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 30(15):2114–2120. doi:[10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170)
- Boucher, A.C., Mimee, B., Montarry, J., Bardou-Valette, S., Bélair, G., Moffett, P., *et al.* (2013). Genetic diversity of the golden potato cyst nematode *Globodera rostochiensis* and determination of the origin of populations in Quebec, Canada. *Mol. Phylogenet. Evol.* 69(1):75–82. doi:[10.1016/j.ympev.2013.05.020](https://doi.org/10.1016/j.ympev.2013.05.020)
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics.* 23(19):2633–2635. doi:[10.1093/bioinformatics/btm308](https://doi.org/10.1093/bioinformatics/btm308)
- Cariou, M., Duret, L., Charlat, S. (2013). Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecol. Evol.* 3(4):846–852. doi:[10.1002/ece3.512](https://doi.org/10.1002/ece3.512)
- Chase, D.L., Pepper, J.S., Koelle, M.R. (2004). Mechanism of extrasynaptic dopamine signaling in *Caenorhabditis elegans*. *Nat. Neurosci.* 7(1):1096–1103. doi:[10.1038/nn1316](https://doi.org/10.1038/nn1316)
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., *et al.* (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly.* 6(2):80–92. doi:[10.4161/fly.19695](https://doi.org/10.4161/fly.19695)
- Colgrove, A.L., Niblack, T.L. (2008). Correlation of female indices from virulence assays on inbred lines and field populations of *Heterodera glycines*. *J. Nematol.* 40(1):39–45.
- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 21(18):3674–3676. doi:[10.1093/bioinformatics/bti610](https://doi.org/10.1093/bioinformatics/bti610)
- Coop, G., Witonsky, D., Di Rienzo, A., Pritchard, J.K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics.* 185(4):1411–1423. doi:[10.1534/genetics.110.114819](https://doi.org/10.1534/genetics.110.114819)
- Davis, E.L., Tylka, G.L. (2000). Soybean cyst nematode disease. *Plant Health Instr.* doi:10.1094/PHI-I-2000-0725-01. doi:10.1094/PHI-I-2000-0725-01

- de Villemereuil, P., Gaggiotti, O.E. (2015). A new F_{ST} -based method to uncover local adaptation using environmental variables. *Meth. Ecol. Evol.* 6(11):1248–1258. doi:10.1111/2041-210X.12418
- Early, R., Sax, D.F. (2014). Climatic niche shifts between species' native and naturalized ranges raise concern for ecological forecasts during invasions and climate change. *Glob. Ecol. Biogeogr.* 23(12):1356–1365. doi:10.1111/geb.12208
- Eizaguirre, C. and Baltazar-Soares, M. (2014). Evolutionary conservation-evaluating the adaptive potential of species. *Evol. Appl.* 7(9): 963-967. doi:10.1111/eva.12227
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., *et al.* (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 6(5):e19379. doi:10.1371/journal.pone.0019379
- Eoche-Bosy, D., Gautier, M., Esquibet, M., Legeai, F., Bretaudeau, A., Bouchez, O., Fournet, S., Grenier, E., Montarry, J. (2017). Genome scans on experimentally evolved populations reveal candidate regions for adaptation to plant resistance in the potato cyst nematode *Globodera pallida*. *Mol. Ecol.* 26 : 4700–4711. doi: 10.1111/mec.14240
- Eves-van den Akker, S., Lilley, C. J., Reid, A., Pickup, J., Anderson, E., Cock, P. J., Blaxter, M., Urwin, P.E., Jones, J. T., Blok, V. C. (2015). A metagenetic approach to determine the diversity and distribution of cyst nematodes at the level of the country, the field and the individual. *Mol. Ecol.*, 24(23): 5842-5851. doi: 10.1111/mec.13434
- Faghihi, J., Donald, P.A., Noel, G., Welacky, T.W., Ferris, V.R. (2010). Soybean resistance to field populations of *Heterodera glycines* in selected geographic areas. *Plant Health Prog.* doi:10.1094/PHP-2010-0426-01-RS.
- Fick, S.E., Hijmans, R.J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37(12): 4302–4315. doi:10.1002/joc.5086
- Foll, M. (2012). BayeScan v2.1 user manual. Available from: http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.1_manual.pdf.
- Foll, M., Gaggiotti, O.E. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics.* 180(2):977–993. doi:10.1534/genetics.108.092221

- Fu, Y.-B., Peterson, G.W., Dong, Y. (2016). Increasing genome sampling and improving SNP genotyping for genotyping-by-sequencing with new combinations of restriction enzymes. *G3 (Bethesda)*. 6(4):845856. doi:10.1534/g3.115.025775
- Garrison, E., Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.
- Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*. 201(4):1555–1579. doi:10.1534/genetics.115.181453
- Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., *et al.* (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Mol. Ecol.* 22(11):3165–3178. doi:10.1111/mec.12089
- Gendron St-Marseille, A.-F. (2013). Le nématode à kyste du soja (*Heterodera glycines*) : enjeux des changements climatiques sur sa distribution, sa reproduction et sur les probabilités de synchronisme avec le soja (*Glycine max*) au Québec [master's thesis]. [Sherbrooke (QC)]: Université de Sherbrooke. Available from: <http://hdl.handle.net/11143/7214>.
- Green, C.D., Greet, D.N., Jones, F.G.W. (1970). The influence of multiple mating on the reproduction and genetics of *Heterodera rostochiensis* and *H. schachtii*. *Nematologica*. 16(2):309–326. doi:[10.1163/187529270X00333](https://doi.org/10.1163/187529270X00333)
- Grenier, E., Bossis, M., Fouville, D., Renault, L., Mugniéry, D. (2001). Molecular approaches to the taxonomic position of Peruvian potato cyst nematodes and gene pool similarities in indigenous and imported populations of *Globodera*. *Heredity*. 86(3):277–290. doi:10.1046/j.1365-2540.2001.00826.
- Günther, T., Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*. 195(1):205–220. doi:10.1534/genetics.113.152462
- Hartl, D.L., Clark, A.G. (2007). Principles of population genetics. 4th ed. Sunderland. Massachusetts Sinauer Associates.
- Hassan, M.A., Pham, T.H., Shi, H., Zheng, J. (2013). Nematodes threats to global food security. *Acta Agric. Scand. B. Soil Plant Sci.* 63(5):420–425. doi:[10.1080/09064710.2013.794858](https://doi.org/10.1080/09064710.2013.794858)
- Hazel, J.R., Williams, E.E. (1990). The role of alterations in membrane lipid composition in enabling physiological adaptation of organisms to their physical environment. *Prog. Lipid Res.* 29(3):167–227. doi:[10.1016/0163-7827\(90\)90002-3](https://doi.org/10.1016/0163-7827(90)90002-3)

- Hegge, A.H. (1957). Soybean cyst nematode, *Heterodera glycines*, in Missouri. *Plant Dis. Rep.* 41(3):201. doi:[10.1094/PDIS-92-10-1475B](https://doi.org/10.1094/PDIS-92-10-1475B)
- Hertel, P., Daniel, J., Stegehake, D., Vaupel, H., Kailayangiri, S., Gruel, C., *et al.* (2013). The ubiquitin-fold modifier 1 (Ufm1) cascade of *Caenorhabditis elegans*. *J. Biol. Chem.* 288(15):10661–10671. doi:[10.1074/jbc.M113.458000](https://doi.org/10.1074/jbc.M113.458000)
- Hulme, P.E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.* 46(1):10–18. doi:10.1111/j.1365-2664.2008.01600
- Hutchison, D.W., Tempelton, A.R. (1999). Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution.* 53(6):1898–1914. doi:10.1111/j.1558-5646.1999.tb04571
- Kamvar, Z.N., Tabima, J.F., Grünwald, N.J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ.* 2:e281. doi:[10.7717/peerj.281](https://doi.org/10.7717/peerj.281)
- Kass, R.E., Raftery, A.E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90(430):773–795. doi:[10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572)
- Kimura, M., Weiss, G.H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49(4):561–576.
- Kofler, R., Pandey, R.V., Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics.* 27(24):3435–3436. doi:[10.1093/bioinformatics/btr589](https://doi.org/10.1093/bioinformatics/btr589)
- Kondrashov, F.A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. R. Soc. B.* 279(1749):5048–5057. doi:[10.1098/rspb.2012.1108](https://doi.org/10.1098/rspb.2012.1108)
- Kristjansson, G. (2010). Evaluation of import pathways for the soybean cyst nematode (*Heterodera glycines* Ichinohe). Pest Risk Assessment # 09-59. Plant Health Risk Assessment Unit of the Science Advice Division. Canadian Food Inspection Agency: Ottawa. 45p.
- Kurle, J., Malvick, D., Potter, B., Orf, J. (2011). Soybean cyst nematode management guide. Chen S, editor. St. Paul (MN): University of Minnesota Extension. 28 p.

- Li, H., Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler Transform. *Bioinformatics*. 25(14):1754–1760. doi:[10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324)
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25(16):2078–2079. doi:10.1093/bioinformatics/btp352
- Liu, X.H., Li, J.Q., Zhang, D.S. (1997). History and status of soybean cyst nematode in China. *Int. J. Nematol.* 7(1):18–25.
- Lu, F., Lipka, A.E., Glaubitz, J., Elshire, R., Cherney, J.H., Casler, M.D., *et al.* (2013). Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9(1):e1003215. doi:[10.1371/journal.pgen.1003215](https://doi.org/10.1371/journal.pgen.1003215)
- Lynch, M., Bost, D., Wilson, S., Maruki, T., Harrison, S. (2014). Population-genetic inference from pooled-sequencing data. *Genome Biol. Evol.* 6(5):1210–1218. doi:[10.1093/gbe/evu085](https://doi.org/10.1093/gbe/evu085)
- Mathew, F., Markell, S., Jantzi, D., Yan, G., Nelson, B., Helms, T. (2015). Soybean cyst nematode. Plant Disease Management. Fargo (North Dakota): North Dakota State University Extension Service. 4 p.
- Mimee, B., Duceppe, M.-O., Véronneau, P.-Y., Lafond-Lapalme, J., Jean, M., Belzile, F., *et al.* (2015). A new method for studying population genetics of cyst nematodes based on Pool-Seq and genomewide allele frequency analysis. *Mol. Ecol. Resour.* 15(6):1356–1365. doi:10.1111/1755-0998.12412
- Mimee, B., Gagnon, A.-È., Colton-Gagnon, K., Tremblay, É. (2016). Portrait de la situation du nématode à kyste du soja (*Heterodera glycines*) au Québec (2013-2015). *Phytoprotection*. 96(1):33–42. doi:10.7202/1038941ar
- Mimee, B., Peng, H., Popovic, V., Yu, Q., Duceppe, M.-O., Tétréault, M.-P., Bélair, G. (2014). First report of soybean cyst nematode (*Heterodera glycines* Ichinohe) on soybean in the province of Quebec, Canada. *Plant Dis.* 98(3):429. doi:[10.1094/PDIS-07-13-0782-PDN](https://doi.org/10.1094/PDIS-07-13-0782-PDN)
- Mitchum, M.G., Wrather, J.A., Heinz, R.D., Shannon, J.G., Danekas, G. (2007). Variability in distribution and virulence phenotypes of *Heterodera glycines* in Missouri during 2005. *Plant Dis.* 91(11):1473–1476. doi:[10.1094/PDIS-91-11-1473](https://doi.org/10.1094/PDIS-91-11-1473)

- Mitchum, M.G. (2016). Soybean Resistance to the Soybean Cyst Nematode *Heterodera glycines*: An Update. *Phytopathol.*, 106(12), 1444-1450. doi:[10.1094/PHYTO-06-16-0227-RVW](https://doi.org/10.1094/PHYTO-06-16-0227-RVW)
- Navon, O., Sul, J.H., Han, B., Conde, L., Bracci, P.M., Riby, J., *et al.* (2013). Rare variant association testing under low-coverage sequencing. *Genetics*. 194(3):769–779. doi:10.1534/genetics.113.150169
- Nei, M., Niimura, Y., Nozawa, M. (2008). The evolution of animal chemosensory receptor gene repertoires: Roles of chance and necessity. *Nat. Rev. Genet.* 9(12):951–963. doi:10.1038/nrg2480.
- Nelson, B.D., Bradley, C.A. (2003). Soybean cyst nematode (SCN). Fargo (North Dakota): North Dakota State University. Available from: <https://www.ndsu.edu/pubweb/~bernelso/soydiseases/cyst.shtml>
- Niblack, T.L., Colgrove, A.L., Colgrove, K., Bond, J.P. (2008). Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. *Plant Health Prog.* doi:10.1094/PHP-2008-0118-01-RS.
- Nikolic, A., Kojic, S., Knezevic, S., Krivokapic, Z., Ristanovic, M., Radojkovic, D. (2011). Structural and functional analysis of SMAD4 gene promoter in malignant pancreatic and colorectal tissues: Detection of two novel polymorphic nucleotide repeats. *Cancer Epidemiol.* 35(3):265–271. doi:[10.1016/j.canep.2010.10.002](https://doi.org/10.1016/j.canep.2010.10.002)
- Novak, S.J. (2007). The role of evolution in the invasion process. *PNAS*. 104(10):3671–3672.
- Paini, D.R., Sheppard, A.W., Cook, D.C., De Barro, P.J., Worner, S.P., & Thomas, M.B. (2016). Global threat to agriculture from invasive species. *Proc. Nat. Act. Sci.*, 113(27), 7575. doi: [10.1073/pnas.1602205113](https://doi.org/10.1073/pnas.1602205113)
- Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* 52(3):273–288. doi:[10.1073/pnas.0700224104](https://doi.org/10.1073/pnas.0700224104)
- Plantard, O., Picard, D., Valette, S., Scurrah, M., Grenier, E., Mugniéry, D. (2008). Origin and genetic diversity of Western European populations of the potato cyst nematode (*Globodera pallida*) inferred from mitochondrial sequences and microsatellite loci. *Mol. Ecol.* 17(9):2208–2218. doi:10.1111/j.1365-294X.2008.03718

- Pluess, A.R., Frank, A., Heiri, C., Lalagüe, H., Vendramin, G.G., Oddou-Muratorio, S. (2016). Genome–environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytol.* 210(2):589–601. doi:[10.1111/nph.13809](https://doi.org/10.1111/nph.13809)
- Plummer, M., Best, N., Cowles, K., Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News.* 6(1):7–11.
- Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One.* 7(2):e32253. doi:[10.1371/journal.pone.0032253](https://doi.org/10.1371/journal.pone.0032253)
- R Core Team. (2017). R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: <http://www.R-project.org/>.
- Revell, L.J. (2017). Phylogenetic tools for comparative biology (and other things). Available from: <https://cran.r-project.org/web/packages/phytools/phytools.pdf>. doi:10.1111/j.2041-210X.2011.00169
- Riggs, R.D. (2004). History and distribution. In: Schmitt DP, Wrather JA, Riggs RD, editors. *Biology and management of soybean cyst nematode*. pp. 9-40 2nd ed. Marceline (Missouri): Schmitt & Associates of Marceline.
- Rousset, F. (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics.* 145(4):1219–1228.
- Savage-Dunn, C., Maduzia, L.L., Zimmerman, C.M., Roberts, A.F., Cohen, S., Tokarz, R., *et al.* (2003). Genetic screen for small body size mutants in *C. elegans* reveals many TGFβ pathway components. *Genesis.* 35(4):239–247. doi:[10.1002/gene.10184](https://doi.org/10.1002/gene.10184)
- Schlötterer, C., Tobler, R., Kofler, R., Nolte, V. (2014). Sequencing pools of individuals – Mining genome-wide polymorphism data without big funding. *Nat. Rev. Genetics.* 15(11):749–763. doi:[10.1038/nrg3803](https://doi.org/10.1038/nrg3803)
- Shurtleff, W., Aoyagi, A. (2010). *History of soybeans and soyfoods in Canada (1831-2010): Extensively annotated bibliography and sourcebook*. Lafayette (California): Soyinfo Center. 1060 p.
- Slack, D.A., Riggs, R.D., Hamblen, M.L. (1972). The effect of temperature and moisture on the survival of *Heterodera glycines* in the absence of a host. *J. Nematol.* 4(4):263–266.

- Statistics Canada. (2017). Table 001-0017 – Estimated areas, yield, production, average farm price and total farm value of principal field crops, in imperial units, annual, CANSIM (database). [Accessed: November 20, 2017]. Available from: <http://www5.statcan.gc.ca/cansim/a26?lang=eng&id=10017>.
- Tardivel, A., Sonah, H., Belzile, F., O'Donoghue, L.S. (2014). Rapid identification of alleles at the soybean maturity gene E3 using genotyping by sequencing and a haplotype-based approach. *Plant Genome* 7(2). doi:10.3835/plantgenome2013.10.0034
- Tylka, G.L., Marett, C.C. (2017). Known distribution of the soybean cyst nematode, *Heterodera glycines*, in the United States and Canada: 1954 to 2017. *Plant Health Prog.* 18(3):167–168. doi:10.1094/PHP-05-17-0031-BR
- Van Tassel, C., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Taylor Lawley, et al. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods.* 5(3):247–252. doi:[10.1038/nmeth.1185](https://doi.org/10.1038/nmeth.1185)
- Watanabe, N., Ishihara, T., Ohshima, Y. (2007). Mutants carrying two *sma* mutations are super small in the nematode *C. elegans*. *Genes Cells.* 12(5):603–609. doi:10.1111/j.1365-2443.2007.01077
- Wood, J.F., Ferkey, D.M. (2016). GRK roles in *C. elegans*. In: G protein-coupled receptor kinases. Gurevich VV, Gurevich EV, Tesmer JJG, editors. New York (New York): Springer New York: p. 283–299. doi:10.1007/978-1-4939-3798-1_13
- Wright, S. (1943). Isolation by distance. *Genetics.* 28(2):114–138.
- Yu, Q. (2011). Soybean cyst nematode (*Heterodera glycines* Ichinohe) In: Soybean physiology and biochemistry. El-Shemy HA, editor. pp. 461-474. Rijeka (Croatia): InTech Europe. Available from: <http://cdn.intechopen.com/pdfs-wm/22782.pdf>. doi:10.5772/17649
- Zheng, J.Z., Chen, S.Y. (2011). Estimation of virulence type and level of soybean cyst nematode field populations in response to resistant cultivars. *J. Entomol. Nematol.* 3(3):37–43.
- Zheng, J., Li, Y., Chen, S. (2006). Characterization of the virulence phenotypes of *Heterodera glycines* in Minnesota. *J. Nematol.* 38(3):383–390.

Chapitre 5 Genomic profiling of virulence determinants in the soybean cyst nematode (*Heterodera glycines*)

Gendron St-Marseille, A.-F., Véronneau, P.-Y., Lord, E., Belov, A., Brodeur, J. and Mimee, B.

Non soumis, en révision.

5.1 Abstract

Soybean cyst nematode (SCN) is the cause of one of the most important disease of soybean. Currently, the main management strategy relies on planting resistant cultivars but the overuse of a single resistance source (PI88788) has led to the selection of virulent SCN populations. To date, the mechanisms used by the nematode to overcome the resistance genes are still unknown. In this study we used different genomic and transcriptomic approaches to identify the genes involved in either or both Peking or PI88788 resistance breaking mechanisms. A genotyping by sequencing approach highlighted 14 loci significantly associated with virulence across 57 field populations. Then, a genome-wide allelic imbalance analysis between second-stage and fourth-stage females grown on a susceptible line (Essex) or selected on a resistant genotype (Peking or PI88788) highlighted 17 variants associated with virulence on Peking and 109 on PI88788. From those, several were located in genes coding for known effectors involved in nematode parasitism. For example, an allele coding for an amino acid change (different polarity) in the SCN gland protein G20E03 was found to be associated with virulence on PI88788. Using a “single-nematode” sequencing strategy, we established for the first time the full transcriptome of single SCN individuals. With this approach we found an island of tandem duplicates of the gene *hgg-20* that could be directly involved in virulence. One version (T0000282801.1) of the *hgg-20* from this region was only expressed in nematodes developing on Peking and the other version (T000028283.1) was strictly present in nematode overcoming PI88788 resistance. Both genes were expressed in SCN populations before selection by the resistant plants. This finding opens the way for the development of a diagnostic tool for the identification of virulent SCN populations.

Key words: *Heterodera glycines*, pathogenecity, single-nematode sequencing, Whole-genome sequencing, GBS

5.2 Introduction

Soybean cyst nematode (SCN - *Heterodera glycines*) is one of the most important diseases of soybean in all major producing countries (Matsuo *et al.*, 2012; Allen *et al.*,

2016; Allen *et al.*, 2017). In United States and Canada, economic losses caused by SCN have been estimated to be more than a billion US dollars per year (Allen *et al.*, 2017). In the U.S., SCN was first identified in North Carolina in 1954 (Winstead *et al.*, 1955) and has spread to almost every soybean producing states since then (Tylka & Marett, 2017), reaching Canada in 1987, including the province of Québec (Anderson *et al.*, 1988; Mimee *et al.*, 2013; Tylka & Marett, 2017). The high genetic differentiation level and continuous gene flow that characterize SCN populations in North America facilitate their adaptation to diverse bioclimatic conditions and establishment in new areas (Gendron St-Marseille *et al.*, 2018). The ongoing climate warming will also favor SCN development by increasing the number of generations per growing season and by opening new regions to soybean cultivation (Gendron St-Marseille *et al.*, 2019). Therefore, more damages caused by SCN are expected in the future.

Crop rotation and resistant cultivars have been the most effective ways to control SCN (Wang *et al.*, 2003). Breeding efforts have allowed the development of highly resistant varieties that remained effective for many decades with no associated yield drawbacks (Tylka & Mullaney, 2015; Rincker *et al.*, 2017). However, about 95% of these resistant cultivars are derived from a single parental line, PI88788 (Tylka & Mullaney, 2015; Mitchum, 2016). In recent years, the overutilization of PI88788 in the United States had exerted a strong selection pressure that led to the emergence of virulent SCN subpopulations able to overcome this resistance source (Niblack *et al.*, 2008; Zheng & Chen, 2013; McCarville *et al.*, 2017). As an example, a recent study showed that PI88788-type resistance was overcome in all 393 tested fields in Missouri (Howland *et al.*, 2018). Even worse, in almost half of these fields (43%), SCN populations also have the capacity to reproduce on a second source of resistance, PI548436 (Peking), from which most of the other 5% resistant cultivars on the market were developed. It is recommended to rotate the sources of resistance to limit the selection pressure and the development of PI88788-virulent populations (Niblack, 2005). Nevertheless, this situation will remain challenging until the development of new resistant germplasms and a sound understanding of how the nematode circumvents mechanisms of soybean resistance.

Soybean resistance is based on the degeneration of the nematode feeding cell called the syncytium, ending in the necrosis of host cell around the nematode, a mechanism similar to hypersensitive response (HR) (Mitchum, 2016). This phenomenon is polygenic and nearly 70 quantitative trait loci (QTL) have been identified so far (Kim *et al.*, 2010). Two major QTLs, *Rhg1* (about 9 genes) and *Rhg4* (about 7 genes), are closely associated with PI88788 and Peking resistance (Klink *et al.*, 2013). The resistance linked to Peking causes a rapid syncytium degeneration, approximately 48 hrs after infection, and requires the presence of both *Rhg1* and *Rhg4* (Brucker *et al.*, 2005; Liu *et al.*, 2012; Yuan *et al.*, 2012). The resistance associated with PI88788 takes more time to develop, between 8-10 days after infection, and requires only *Rhg1*, but is dependent on the presence of multiple copies of the genes at that locus to be effective (Brucker *et al.*, 2005; Cook *et al.*, 2012; Liu *et al.*, 2012; Cook *et al.*, 2014). The alleles of the α -SNAP (α -soluble N-ethylmaleimide sensitive factor attachment protein) genes present in Peking (*rhg1-a*) and PI88788 (*rhg1-b*) also differ (Bayless *et al.*, 2016). Even if both types of resistance have a similar genetic background, SCN populations have adapted independently to each source (Anand & Brar, 1983).

In SCN, virulence genes were shown to be inherited in a Mendelian mode independently in a dominant (PI 88788) and recessive manner (Peking) using controlled crossing strategy on inbred SCN lines (Dong & Opperman, 1997). Bekal *et al.* (2015) had identified two candidate genes from the comparison of highly inbred SCN populations but they were not validated in other populations. Thus, the identity of the genes underlying virulence remains unknown, probably due to their complex nature (Mitchum, 2016). The recent availability of a reliable assembly and annotation of *H. glycines* genome should facilitate association studies to highlight genes involved in virulence (Masonbrink *et al.*, 2019).

Plant endoparasitic nematodes have acquired the capacity to invade their host without triggering natural defense reactions by secreting a diverse set of effector proteins (Haegeman *et al.*, 2012). In response, the hosts have evolved different resistance mechanisms, typically by recognizing the effectors and initiating a rapid and localized cell death or HR (Williamson & Kumar, 2006). In the potato cyst nematode, *Globodera*

pallida, the protein Gpa2 produced by the NB-LRR (nucleotide-binding and leucine-rich repeat domains) resistance gene recognizes the *Gp-Rbp-1* effector protein secreted by the nematode in incompatible reaction while a single amino-acid polymorphism confers virulence in compatible reaction (Sacco *et al.*, 2009). In SCN, many putative effectors have been discovered, but for most of them their functions remain unknown (Mitchum *et al.*, 2013; Mitchum, 2016). To date, avirulent genes or mechanisms involved in SCN resistance breaking have not been reported. Two main hypotheses can be put forward to explain the overcoming of resistance. First, polymorphism in effector sequences could prevent their recognition by the plant. Second, a difference in gene expression (resulting from regulation, gene copy number or epigenetic control for example) could lead to the absence of a protein involved in HR response or the overproduction of a “virulence factor”.

SCN field populations are characterized by a high genetic diversity and, in a single field, these genotypes may produce very distinct phenotypes (Zheng & Chen, 2013). On the other hand, SCN being an obligatory parasite, obtaining highly inbred strains to performed genetic analysis is cumbersome and only offers a reduced (and dated) representation of the population diversity. Genotyping-by-Sequencing (GBS) is a simple approach allowing the identification of many markers for genome wide association studies (GWAS) without any prior knowledge of the regions of interest or reference genome (Elshire *et al.*, 2011). It is possible to genotype numerous pools of individuals at very low cost to establish links between the observation of a population phenotype with the allele frequency at specific genomic loci. This approach has notably been used to identify genomic regions associated with phenotypes in many species and regions under selection for climatic adaptation in SCN (Sonah *et al.*, 2015; Gendron St-Marseille *et al.*, 2018; Jamil *et al.*, 2018; Rovelli *et al.*, 2018). The objective of this work was to identify genes associated with the virulence on either or both Peking or PI88788. We hypothesized that this approach could be a simple and efficient way to identify SCN genomic regions associated with virulence by screening 57 populations with diverse virulence profiles. To refine this association, we sequenced the whole genome of four of these populations before and after selection of virulent individuals on the resistant soybean genotypes Peking and PI88788. Finally, since virulence could also be associated

with difference in gene expression, RNA-seq was used to establish the transcriptome of pools of larvae at J2 stage and individual female at J4 stage from nine distinct SCN field populations developing on Peking, PI88788 or the susceptible cultivar Essex.

5.3 Material and Methods

SCN Populations

A total of 59 field populations of SCN, representative of the area currently infested in North America, were sampled or supplied by collaborators from 12 US states (Iowa, Illinois, Indiana, Kansas, Maryland, Michigan, Minnesota, Missouri, North Dakota, Nebraska, Ohio, South Dakota) and one Canadian province (Ontario) (Table 5.1).

Table 5.1 Name, origin and population virulence phenotype (HG type) on Peking (line #1) and PI88788 (line #2) of the populations used in the different experiments of this study.

Name	Location	FI* on Peking	FI on PI88788	HG Type	GBS ¹	WGS ²	RNA-Seq ³
IA6	Glenwood, Iowa	4	18	2	x		
IA8	Muscatine, Iowa	0	24	2	x		
IA9	Newell, Iowa	0	13	2	x		
IL1	Arthur, Illinois	2	20	2	x		
IL3	Dekalb, Illinois	45	38	1.2	x		
IL4	Dekalb, Illinois	2	37	2	x	x	
IL5	Pontiac, Illinois	0	25	2	x		
IN1	Evansville, Indiana	7	22	2	x		
KS1	Manhattan, Kansas	0	24	2	x		
KS2	Edwards, Kansas	0	0	0	x		
MD1	Queenstown, Maryland	0	19	2	x		
MI1	Decatur, Michigan	1	7	0	x		
MN1	Danvers, Minnesota	11	12	1.2	x		
MN2	Lamberton, Minnesota	1	11	2	x		
MN3	Stewart, Minnesota	4	40	2	x		
MN4	Stewart, Minnesota	2	25	2	x		
MN5	Waseka, Minnesota	1	11	2	x		
MO1	Clarkton, Missouri	11	35	1.2	x		
ND1	Fairmont, North-Dakota	1	11	2	x		
ND3	Steele, North-Dakota	0	2	0	x		
ND4	Dwight, North-Dakota	0	7	0	x		
ND6	Arthur, North-Dakota	0	0	0	x		
NE1	Herman, Nebraska	0	15	2	x		
OH1	Pike, Ohio	13	10	1.2	x		
OH2	Brown, Ohio	1	19	2	x		
ON1	Harrow, Ontario	2	4	0	x	x	x
ON10	Stoney Point, Ontario	47	1	1	x		x
ON11	Chatham, Ontario	2	6	0	x		
ON12	Blenheim, Ontario	4	20	2	x		
ON13	Wardsville, Ontario	19	18	1.2	x		
ON14	Forest, Ontario	8	17	2	x		

ON15	Kingsville, Ontario	8	18	2	x	
ON16	Amherstburg, Ontario	29	17	1.2	x	x
ON17	Parkhill	5	20	2		x
ON19	Parkill, Ontario	1	2	0	x	
ON2	Maidstone, Ontario	16	1	1	x	
ON20	Dunnville, Ontario	2	10	2	x	
ON21	Kent, Ontario	0	8	0	x	x
ON22	Ontario	1	2	0		x
ON24	Dresden, Ontario	2	24	2	x	x
ON25	Dawn-Euphemia, Ontario	41	14	1.2	x	
ON26	Tupperville, Ontario	0	3	0	x	
ON27	Amherstburg, Ontario	13	6	1	x	
ON3	Grand Point, Ontario	3	2	0	x	
ON32	Chatham, Ontario	4	14	2	x	
ON34	Bothwell, Ontario	23	6	1	x	x
ON35	Pain court, Ontario	38	17	1.2	x	
ON4	Wallaceburg, Ontario	1	5	0	x	
ON47	Elgin, Ontario	4	6	0	x	
ON48	Ottawa, Ontario	0	16	2	x	
ON5	Blenheim, Ontario	16	17	1.2	x	
ON50	Ottawa, Ontario	14	10	1.2	x	
ON51	Chatham, Ontario	0	11	2	x	
ON6	Bothwell, Ontario	11	6	1	x	x
ON7	Rodney, Ontario	2	26	2	x	x
ON8	Stoney Point, Ontario	23	2	1	x	
SD1	Beresford, South-Dakota	0	30	2	x	
SD2	Hurley, South-Dakota	0	1	0	x	
SD3	Bersford, South-Dakota	0	1	0	x	

* FI = Female index: Mean number of females developing on the resistant line divided by the number on the susceptible control multiplied by 100. Population with FI > 10 are considered virulent on this line.

¹ GBS, Genotyping by Sequencing.

² WGS, Whole Genome Sequencing.

³ RNA-Seq, RNA-Sequencing.

The population virulence phenotype (HG type) of each population on resistant lines Peking and PI88788 was confirmed using standard protocol as described in Niblack *et al.* (2002) with only two indicator lines (Peking and PI88788) the cultivar Essex as susceptible control.

Genotyping by sequencing (GBS)

DNA extraction

Forty cysts from each of the 57 populations (from the 59 listed in Table 5.1) used for GBS were gently crushed to extract eggs and washed in sterile water. Total genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Mississauga, ON, Canada) following the manufacturer's instruction. DNA extracts were quantified using

Qubit fluorometric quantification (ThermoFisher Scientific, Burlington, ON, Canada) and normalized at 2 ng/ μ L.

Sequencing and SNP calling

Library preparation and sequencing were performed at the Institute for Integrative and System Biology (IBIS: Université Laval, Québec City, Qc, Canada). The GBS protocol described in Mimee *et al.* (2015) based on the combination of two restriction enzymes (PstI-MspI; New England Biolabs, Whitby, ON, Canada) was used. Samples were barcoded, multiplexed and pooled in a single library that was sequenced on three Ion Proton chips (ThermoFisher Scientific).

The UNEAK pipeline (Lu *et al.*, 2013), from TASSEL 3.0 bioinformatics analysis package was used to call single-nucleotide polymorphism (SNP) with high stringency (Bradbury *et al.*, 2007). A single putative SNP (1-bp mismatch) was tolerated per sequence. The SNPs were further filtered to ensure a minimum coverage (minCov) of five reads, a maximum coverage (maxCov) of 10,000 reads, a minimum allele frequency (MAF) of 0.01 (over all populations) and to keep only those with no more than 20% of missing data.

Genotype-phenotype relationship

The relationship between the virulence (FI on Peking or PI88788) of each population and their allelic frequency (AF) at each locus (SNP) was evaluated using linear regressions in R (R Core Team, 2019). Prior to the correlation analysis, both AF and FI were normalized by dividing their values with the higher value for each variable in a given population. SNPs with a Spearman correlation values (r) greater than ± 0.4 and with a significance level (p-value) lower than 0.01 were kept as good candidates. The position of these SNPs was retrieved from the SCN reference genome (Masonbrink *et al.*, 2019) by aligning the GBS sequences with BWA-MEM (Li & Durbin, 2009). Sequences located in a predicted gene were blasted using Blast2GO (Götz *et al.*, 2008).

Whole genome sequencing

Selection of virulent J4 females on resistant genotypes and DNA extraction

We choose three populations (ON1, ON34 and IL4) on the basis of their HG type (respectively HG type 0, HG type 1 and HG type 2) and virulence capacity on Peking PI88788 to forced selection of virulent individuals on resistant lines. Approximately 500 cysts of the populations ON1, ON34 and IL4 were immersed in water at 25°C in a growth chamber during three weeks before eggs hatched. Then, J2 were harvested and seven-days-old soybean plants were inoculated by pipetting ~200 larvae/plant directly on the root. The population ON1 was inoculated on Peking and PI 88788, ON34 was inoculated on Peking only and IL4 was inoculated on PI88788 only. Twelve plantlets per population were grown in silicone ice cube tray filled with wet sand in a growth chamber at 28°C with 16/8h light/dark photoperiod. After 24 hrs, plants roots were rinsed and plantlets were transferred to plastic cone-tainers (SC10L, Stuewe and sons, inc, Tangent, Oregon, United States) filled with sand and incubated in a growth chamber with the same conditions for 14 days. After 13 days, eight plants were removed and stained with fuch sine as described by Eisenback (2000) to confirm infection and life stage. The next day, the roots were blended as described by Eisenback (2000) and 40 individual J4 females were handpicked from each population, washed three time in sterile water and frozen at -80°C in a microtube. DNA was extracted as described in the GBS section from this material and from eggs from 40 cysts of each population to represent the genetic diversity before the selection on resistant genotypes (founder population).

Library preparation and sequencing

All DNA samples were normalized at 2 ng/μL and 0.5 μl (1 ng) of total DNA was used to prepare paired-end libraries with the Nextera XT Kit library (Illumina, San Diego, CA, United States) following the manufacturer instructions. Size selection and cleanup were done using Agencourt AMPure XP beads (Beckman-Coulter, Indianapolis, IN, United States) with a ratio (bead/sample) of 0.8 to kept long fragments. The quality and concentration of each library were evaluated on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States) with a Agilent DNA 7500 Kit (Agilent Technologies). Sequencing was performed using a 300-cycle (2 × 150-bp paired-end) MiSeq v2 reagent kit (Illumina) on a MiSeq platform according to the manufacturer's instructions.

Bioinformatic processing

Trimmomatic was used to remove the adapters, low quality reads and reads with less than 36 bp long (v. 0.39 - Bolger *et al.*, 2014). Quality trimmed read pairs were mapped to the *Heterodera glycines* genome (Masonbrink *et al.*, 2019) using the BWA package (Li & Durbin, 2009). The alignment statistics were calculated with Samtools tools (v. 1.9 - Li *et al.*, 2009). Variants were called using Freebayes (Garrison & Marth, 2012) and their putative effect analyzed with SnpEff (v. 4.3+ - Cingolani *et al.*, 2012). Variants were screened based on AF to identify those for which an allele was fixed by selection (AF = 0 or 1) and those showing a change in AF greater than $\pm 40\%$ after selection on a resistant genotype compared to the founder population. Variant respecting these conditions and located in a predicted gene were blasted using Blast2GO (Götz *et al.*, 2008). In order to identify variants localized in known effector gene we used the list available on the SCN database (scnbase.org).

RNA-Sequencing

RNA extraction

In this experiment, nine populations with different virulence profiles and able or not to overcome Peking and PI88788 resistance were used.

Larvae (J2) were hatched as detailed in the WGS section. A pool of 30 larvae from each population was used to extract RNA at the pre-parasitic stage. Immature females from each population able to develop on a given resistant plant at the J4 stage were produced as described in the WGS section. The populations ON1, ON21 and ON22 were inoculated on the susceptible line Essex as control, the populations ON7, ON17 and ON24 on PI88788 and ON6, ON10 and ON16 on Peking. Three individual females per population were selected and cut in buffer (see Serra *et al.*, 2018 for recipe) directly after handpicking separately. RNA was extracted from fresh individual females directly after handpicking (14 days after inoculation) using the Smart Seq 2 method (Serra *et al.* (2018). We decided to replace the KAPA HiFi HotStart ReadyMix (2x) (Roche Diagnostic Ltd, Indianapolis, Indiana, USA) with Phusion high fidelity PCR master mix (Thermo Scientific TM, Waltham, Massachusetts, USA) to amplify cDNA. The quality of

the cDNA samples was analyzed with an Agilent DNA 7500 Kit on the Agilent 2100 Bioanalyzer.

Library preparation and sequencing

Sequencing libraries were prepared using the Nextera DNA Flex Library Prep Kit (Illumina) following manufacturer's instructions. Libraries quality was analyzed using the Agilent 2100 Bioanalyzer with an Agilent DNA 7500 Kit (Agilent Technologies). Libraries were normalized and sequenced (paired-end 2 x 150 bp) at McGill and Genome Québec Innovation Center (Montréal, Québec, Canada) on one lane of NovaSeq 6000 (Illumina).

Bioinformatic processing

Sequences were trimmed to keep those with quality score higher than 20 and at least 36 bp long with Trimmomatic (v. 0.39 - Bolger *et al.*, 2014). Abundance of transcripts from single-nematode RNA-Seq samples was determined by pseudo-alignment on the SCN reference genome (Masonbrink *et al.*, 2019) using the Kallisto software with default settings (v. 0.45.0 - Bray *et al.*, 2016). The Edger package (v. 3.8 - Robinson *et al.*, 2010; McCarthy *et al.*, 2012) was used to normalize read counts and to analyze differential expressions between treatments (selected females on Peking with Essex as control and selected females on PI88788 with Essex as control, see table S.5 for the design matrix). A hierarchical clustering of differentially expressed genes (DEG) was performed using R package "stats" (R Core Team, 2019). Clusters were calculated using the average of the euclidian distances between genes. Heatmap was created using "pheatmap" R package (Kolde, 2019). The expression value of the genes of interest was manually checked in the sequences from the J2 pools.

5.4 Results

Genotyping-by-sequencing

Using the GBS approach, we obtained a total of 192,576,709 short reads. Before filtering, the UNEAK pipeline identified a total of 2,389 SNPs. After filtering to remove low coverage sequences and loci with more than 20% of missing data 1,169 SNPs were

retained for further analyses. A linear regression between the AF of these SNPs and the virulence level (FI) of 57 SCN populations identified 14 loci with a $r > \pm 0.4$ and a p-value < 0.01 . Of these, seven were correlated to reproduction on Peking and seven on PI88788. In each group, five out of seven sequences were mapped back to the reference genome (Table 5.2) and located on 9 different scaffolds.

Table 5.2 List of SNPs correlated with virulence on Peking and PI88788.

Scaffold ID	Position	Hetgly gene ID	Spearman correlation	p-value	UniProt ID	Best hit description	Species	Evalue	Location
Peking virulence type									
50	7624	G000005350	0.416	0.0014	O01505	Prion-like-(Q/N-rich)-domain-bearing protein	<i>Caenorhabditis elegans</i>	1.6E-28	Exon
Not mapped		-	0.580	4.6E-5	-	-	-	-	-
12	299065	G000002695	0.568	1.68E-5	KPBA	Probable phosphorylase b kinase regulatory subunit alpha	<i>C. elegans</i>	0.00E+00	Exon
43	161931	G000013619	0.421	0.002	A0A183BZV1	Uncharacterized protein	<i>Globodera pallida</i>	5.1E-24	Exon
Not mapped		-	0.439	0.0020	-	-	-	-	-
124	244979	G000012871	0.407	0.004	N1NVB5	K+/Cl-Cotransporter	<i>C. elegans</i>	0.00E+00	Exon
42	227507	G000011992	0.418	0.0016	A0A183C555	Putative DNA polymerase	<i>Trachymyrmex zeteki</i>	3.5E-69	Exon
PI88788 virulence type									
227	130031	Intergenic	0.444	0.0015	-	-	-	-	-
82	94368	G000013936	0.449	0.0015	A0A090LNI7	Suppressor of Stomatin mutant Uncoordination	<i>Strongyloides ratti</i>	3.6E-110	Exon
148	176776	G000014958	0.414	0.0015	A0A2A6BYM9	G protein-coupled receptor	<i>Pristionchus pacificus</i>	2.00E-110	Exon
601	15013	G000029142	0.428	0.0026	A0A183BXX0	Uncharacterized protein	<i>G. pallida</i>	2.00E-95	Exon
85	129959	G000013853	0.400	0.0043	A0A2L0VDJ0	Putative effector protein	<i>Heterodera avenae</i>	1.2E-173	Intron
85	87041	G000013844	0.410	0.0021	A0A183CK41	Phospholipid-transporting ATPase	<i>G. pallida</i>	0.00E+00	Exon
Not mapped		-	0.479	0.0006	-	-	-	-	-

Interestingly, 10 of these SNPs were located in coding genes and nine of them in exonic sequences. The sequence showing the best correlation for Peking ($r = 0.580$, $p = 4.6E-5$) didn't map to the reference but the SNP with the second best correlation ($r = 0.568$, $p = 1.68E-5$) was located in a probable phosphorylase b kinase regulatory subunit alpha. SNPs correlated with virulence on Peking hits a Prion-like-(Q/N-rich)-domain-bearing protein, a K⁺/Cl⁻-Cotransporter and a putative DNA polymerase. For virulence on PI88788, the best correlation ($r = 0.479$, $p = 0.0006$) was also obtained with an unmapped sequence. The other SNPs showing a significant correlation were located in a suppressor of stomatin mutant uncoordination, a phospholipid-transporting ATPase, a G protein-coupled receptor and a putative effector protein. Two of those SNPs were physically associated on scaffold 85.

Whole Genome Sequencing

The genomic DNA of three populations obtained before (“founder”) and after their selection (“selected”) on a resistant soybean line were sequenced. A total of 1,782,150 variants were identified. The allele frequencies of these variants were analyzed to retain only those with a high variation (40% or more) between founder and selected isolates or selecting a single allele (homozygous) that were consistent in both populations selected on the same plant. After filtering, a total of 784 variant were kept across 589 genes. Of these, 124 variants (found in 86 different genes) appeared to be under selection on Peking (Suppl. table S 5.1 and S 5.2) and 660 (found in 503 genes) on PI88788 (Suppl. table S 5.3 and S 5.4). A total of 12 genes of interest containing variants predicted to have a high or a moderate impact in an annotated protein were associated with virulence on Peking (Table 5.3) and 62 with virulence on PI88788 (Table 5.4).

Of particular interest for virulence on Peking (Table 5.3), two variants were predicted to generate a stop codon in a protein SNF2 (helicase-like protein) family and in a integrase core domain protein.

Table 5.3 Variants associated with important allelic frequency changes or for which a single allele was selected on resistant plant associated with SCN females selected on Peking.

Blast2GO best hit description*	Scaffold ID	Position	gene ID	e-Value	Variant type	Variant type**	Allele frequencies***			
							IL4 Founder /PI88788	ON1 Founder/Peking	ON1 Founder/PI88788	ON34 Peking/Peking
Aspartyl aminopeptidase	111	118624, 118630, 118658	G000014217	6.42E-177	complex & mnp & snp	Mis & Syn	0.7/0.5	0.8/0.3	0.8/0.7	1.0/0.2
Dorsal gland cell-specific expression protein	95	168408	G000016215	1.15E-38	complex & del & snp	Fs & Mis	0.8/0.8	0.8/1.0	0.8/0.8	0.7/1.0
Dorsal gland cell-specific expression protein	159	137849	G000019407	7.92E-12	complex & del & ins	Mis & Clins & Cldel	0.8/0.8	0.8/1.0	0.8/0.7	0.8/1.0
Mucin-2 isoform X1	208	52180	G000021751	9.94E-142	snp	Mis	0.9/0.7	0.6/1.0	0.6/0.8	0.8/1.0
Oxidoreductase. short chain dehydrogenase/reductase family protein	9	777418	G000005953	3.07E-58	snp	Mis	1.0/0.5	0.8/1.0	0.8/0.7	0.7/1.0
Protein Y45F3A.4	120	11736	G000013014	1.22E-23	complex	Mis	0.8/0.6	0.8/1.0	0.8/0.5	0.8/1.0
SNF2 family protein	3	536266	G000002997	0	snp	StG	0.7/0.7	0.7/1.0	0.7/0.8	0.7/1.0
SH3 domain protein	336	4884	G000026873	2.51E-43	complex & snp	Mis & Syn	0.2/0.1	0.4/0.0	0.4/0.4	0.9/0.2
TATA-box binding protein	85	354141	G000013910	7.71E-07	snp	Mis	0.8/0.8	0.8/1.0	0.8/0.7	0.8/1.0

*In bold, protein with putative effector functionality. **Abbreviations: Conservative inframe deletion, Cldel; Conservative inframe insertion, Clins; frameshift, Fs; Missense, Mis; Stop gained, StG; Synonymous, Syn. *** Allelic frequency of the original population / allele frequency after selection for one generation on selected cultivar. *Eg. While the allelic frequency is remaining stable when selected on one of the resistant cultivar (Peking or PI88788), it changes drastically when selected on the other type of resistant cultivar (Peking or PI887788).*

Also, three genes contained at least a variant predicted to generate a frameshift in proteins from which one is known to be a potential effector of the SCN (dorsal gland cell-specific expression protein). Six other genes showed a missense variant, one being a second putative SCN effector (dorsal gland cell-specific expression protein).

Among the genes of interest for virulence on PI88788 (Table 5.4), a variant was predicted to cause a stop codon in a POU domain, class 3, transcription factor 3.

Table 5.4 Variants associated with important allelic frequency changes or for which a single allele was selected on resistant plant associated with SCN females selected on Peking.

Blast2GO best hit description	Scaffold ID	Position	Hetgly gene ID	e-Value	Variant type	Effect type*	Allelic frequencies			
							IL4 Founder/PI88788	ON1 Founder/Peking	ON1 Founder/PI88788	ON34 Founder/Peking
ADP-ribose pyrophosphatase. mitochondrial	137	104461	G000012581	4.37E-09	ins & snp	Fs & Mis	0.7/1.0	0.7/0.6	0.7/1.0	0.3/0.5
Alcohol dehydrogenase 1-like	70	388153	G000010247	5.17E-43	snp	Mis	0.7/0.3	0.8/0.6	0.8/0.3	1.0/1.0
Aldolase and Adducin domain protein	207	47486	G000021128	4.18E-72	snp	Mis	0.8/1.0	0.8/0.8	0.8/1.0	0.9/0.8
AMP deaminase	71	52659	G000010267	0.00E+00	complex & snp	Mis	1.0/0.6	1.0/0.8	1.0/0.6	0.3/0.3
Annexin family	136	63412	G000009935	7.36E-102	snp	Mis	0.1/0.8	0.2/0.2	0.2/0.7	1.0/1.0
Antiviral helicase	34	571740	G000010727	0	snp	Mis	0.8/1.0	0.8/0.6	0.8/1.0	0.3/0.4
Arginine kinase	153	150639	G000017275	6.41E-82	snp	Mis	0.3/0.8	0.4/0.6	0.4/1.0	0.3/0.4
Asparagine synthetase domain-containing protein 1	16	397745	G000005044	1.24E-62	snp	Mis	1.0/0.3	0.6/0.5	0.6/0.1	1.0/1.0
Bm10974. isoform a	20	211654	G000003243	3.09E-15	snp	Mis	0.9/0.4	0.5/0.4	0.5/0.0	0.0/0.0
Cell growth-regulating nucleolar protein-like	80	149975	G000015207	3.49E-33	snp	Mis	0.8/0.4	0.8/0.9	0.8/0.4	1.0/1.0
CRE-ROD-1 protein	114	167374, 167384	G000017737	3.55E-134	snp & complex	Mis	0.6/1.0	0.7/0.8	0.7/1.0	0.1/0.2
Cytosolic endo-beta-N-acetylglucosaminidase	60	297264	G000005564	3.74E-78	snp	Mis	0.1/0.6	0.6/0.4	0.6/1.0	0.0/0.0
DnaJ homolog subfamily C member 10-like	284	38454	G000022305	0.00E+00	snp	Mis	0.0/0.5	0.1/0.0	0.1/0.5	0.0/0.0
DNA polymerase, putative	102	259636	G000014073	4.29E-93	del & snp	Cidel & Dlin & Mis	0.7/0.3	0.8/0.8	0.8/0.3	0.2/0.3
Dorsal gland cell-specific	28	368031	G000010818	1.54E-14	snp	Mis	0.1/0.8	0.4/0.6	0.4/1.0	0.0/0.0

expression protein										
Dorsal gland cell-specific expression protein	28	368040	G000010818	1.54E-14	complex	Mis	0.0/0.7	0.5/0.6	0.5/0.9	0.0/0.0
Dorsal gland cell-specific expression protein	277	36337	G000024468	9.58E-42	complex	Mis	0.5/0.9	0.4/0.3	0.4/1.0	0.0/0.0
Dorsal gland cell-specific expression protein	277	36354	G000024468	9.58E-42	snp	Mis	0.5/1.0	0.4/0.3	0.4/0.8	0.0/0.0
Dorsal gland cell-specific expression protein	277	54464	G000024472	1.49E-28	complex	Mis	1.0/0.5	0.9/0.8	0.9/0.4	1.0/1.0
Dorsal gland cell-specific expression protein	277	54503	G000024472	1.49E-28	snp	Syn	1.0/0.5	1.0/0.9	1.0/0.4	1.0/1.0
Elongation factor Tu GTP-binding domain-containing protein 1	89	214080	G000017040	1.72E-72	snp	Mis	0.8/1.0	0.8/0.5	0.8/1.0	0.0/0.0
Endonuclease/exonuclease/phosphatase family protein putative	92	66725	G000009753	1.99E-66	complex & snp	Mis	0.0/0.7	0.2/0.3	0.2/0.7	0.0/0.0
FLYWCH zinc finger domain protein	217	96836	G000022678	3.66E-17	snp	Mis	0.0/0.7	0.4/0.5	0.4/0.9	0.6/0.7
Gland protein G17G01	188	100704	G000018675	3.51E-33	complex	Mis	0.8/1.0	0.7/0.8	0.7/1.0	0.2/0.6
Gland protein G20E03	34	36124	G000010619	9.82E-121	snp	Mis	0.4/0.0	0.4/0.4	0.4/0.0	0.3/0.2
Glutamyltranspeptidase family-containing protein (gamma)	40	563913	G000004972	1.99E-57	snp	Mis	1.0/0.2	0.9/0.8	0.9/0.4	1.0/1.0
Guanylate cyclase gcy-9	59	32933	G000006195	0.00E+00	snp	Mis	0.3/0.9	0.5/0.5	0.5/1.0	0.0/0.0
Immunoglobulin I-set domain-containing protein	136	116834	G000009955	7.18E-54	snp	Mis	0.9/0.5	0.8/0.7	0.8/0.1	0.0/0.0
Innexin family-containing protein	270	96344	G000024053	2.23E-154	snp	Mis	0.1/0.6	0.3/0.3	0.3/0.7	0.0/0.0
Integrase core domain protein	381	81359	G000018119	2.02E-15	complex & snp	Mis & Syn	0.2/0.8	0.3/0.4	0.3/0.8	0.3/0.1
JmjC domain protein	92	146469	G000009778	9.99E-80	snp	Mis	0.9/0.5	0.8/0.9	0.8/0.2	1.0/1.0

Let-4	308	48679	G000019634	6.67E-56	complex & snp	Mis	0.8/1.0	0.8/0.8	0.7/1.0	0.9/0.5
Let-4	308	48735	G000019634	6.67E-56	complex & snp	Mis	0.7/1.0	0.8/0.8	0.8/1.0	0.9/0.7
Let-4	308	48741	G000019634	6.67E-56	complex & snp	Mis & Syn	0.7/1.0	0.8/0.8	0.8/1.0	0.9/0.7
Let-4	308	48755	G000019634	6.67E-56	complex & snp	Mis & Syn	0.7/1.0	0.8/0.8	0.8/1.0	0.9/0.7
Let-4	308	48766	G000019634	6.67E-56	complex & snp	Mis	0.7/1.0	0.8/0.8	0.8/1.0	0.9/0.6
Lissencephaly-1 -like protein A	463	27304	G000024745	4.92E-27	complex & snp	Mis & Syn	0.2/1.0	0.8/0.6	0.8/1.0	0.0/0.0
LPXTG cell wall anchor domain-containing protein	169	86676	G000023163	7.08E-05	complex	Mis	0.8/0.4	0.8/0.7	0.8/0.2	1.0/1.0
Mannose 4.6 dehydratase 1	182	176886	G000016657	0.00E+00	snp	Mis	0.2/0.8	0.2/0.2	0.2/0.6	0.6/0.6
Methylthioribulose-1-phosphate dehydratase	86	41241	G000010523	6.45E-62	complex & snp	Mis	0.1/0.7	0.2/0.0	0.2/0.6	0.0/0.0
MH1 domain protein	116	265665	G000015913	8.82E-95	snp	Mis	0.9/0.4	0.6/0.5	0.6/0.2	0.8/0.8
Nucleolar protein 10	374	30089	G000027794	1.08E-92	complex & snp	Mis	0.5/1.0	0.6/0.8	0.6/1.0	0.0/1.0
Nuclear hormone receptor family member nhr-48	20	703781	G000003392	1.28E-58	snp	Mis	1.0/0.1	0.8/0.7	0.8/0.4	0.0/0.0
Pectate lyase 1	260	6741, 45222	G000025554	6.40E-78	snp	Mis	0.9/0.0	0.7/0.8	0.7/0.1	0.0/0.2
Pectate lyase 1	260	45232, 45237	G000025559	1.49E-75	snp	Mis	0.7/1.0	0.8/0.8	0.8/1.0	1.0/0.7
Peptidase family M1 containing protein	233	106336	G000024132	1.11E-132	snp	Mis	0.9/0.1	0.9/0.8	0.9/0.5	1.0/1.0
Pentapeptide repeat-containing protein	324	23787	G000024682	7.75E-16	complex & snp	Mis	0.4/0.9	0.3/0.4	0.3/0.7	0.0/0.0
poly(A) polymerase type 3	49	298495	G000011044	1.16E-25	snp	Mis	0.0.8	0.5/0.5	0.5/1.0	0.0/0.0
Poly(A) polymerase gamma-like isoform X1	249	100076	G000020222	3.18E-40	complex & snp	Mis	0.8/1.0	0.8/0.7	0.8/1.0	0.8/0.2

Proline-rich protein 12	1	873420	G000000669	3.47E-55	snp	Mis	0.4/1.0	0.8/0.4	0.8/1.0	1.0/0.7
4-hydroxybenzoate polyprenyltransferase. mitochondrial isoform X1	12	727438	G000002816	1.48E-114	complex & snp	Mis & Syn	0.4/1.0	0.6/0.6	0.6/1.0	0.0/0.4
RNA-directed DNA polymerase (reverse transcriptase) and Integrase domain containing protein	67	266816	G000014324	5.44E-79	snp & snp	Mis	0.9/0.1	0.5/0.4	0.5/0.1	0.1/0.2
RNA-directed DNA polymerase (reverse transcriptase) and Integrase domain containing protein	101	115890	G000016030	3.96E-114	complex & snp	Mis & Syn	0.3/0.0	0.3/0.2	0.3/0.0	0.2/0.3
SecA	151	141595	G000019082	1.10E-90	complex	Mis	0.1/0.5	0.2/0.1	0.2/0.7	0.0/0.0
SecA	151	141615	G000019082	1.10E-90	snp	Mis	0.1/0.6	0.2/0.2	0.2/0.6	0.0/0.0
Small subunit processome component 20 homolog	92	91906	G000009762	0.00E+00	snp	Mis	1.0/0.6	0.8/0.8	0.8/0.0	1.0/1.0
Sodium/hydrogen exchanger	134	228557	G000014176	0.00E+00	ins & snp	Dlin & Dlin & Mis	0.0/0.8	0.3/0.5	0.3/0.8	0.0/0.0
Sugar transporter SWEET	11	319048	G000001344	1.61E-93	snp	Mis & Syn	1.0/0.5	0.8/0.6	0.8/0.3	1.0/1.0
Talin-1	148	207224	G000014967	0	complex	Mis	0.8/1.0	0.8/0.8	0.8/1.0	0.6/0.6
TANC1 isoform X1	101	141620	G000016040	3.81E-19	snp	Mis	0.6/0.0	0.2/0.4	0.2/0.0	0.8/0.6
Transcription factor 3	91	50397	G000018228	1.29E-18	snp	StG & Mis	0.8/1.0	0.8/0.8	0.8/1.0	0.7/0.8
Transthyretin-like family protein	60	87902	G000005507	1.99E-44	snp		0.3/0.7	0.2/0.4	0.2/0.8	0.0/0.0
Transposon-derived protein F54H12.3	343	37814	G000022033	2.81E-60	ins	Ciins	0.0/0.6	0.1/0.1	0.1/0.6	0.1/0.1
Ubiquitin-40S ribosomal protein S27a	246	81564	G000024555	3.75E-27	mnp & snp	SpD & Syn & Int	0.1/0.7	0.2/0.3	0.2/0.6	0.0/0.0

Unknown protein	609	34561	G000027398	2.17E-49	complex & snp	Mis	1.0/0.6	0.8/0.9	0.8/0.3	0.4/0.4
Unknown protein	1	114934 1	G000000758	9.01E-11	snp	Mis	0.7/1.0	0.8/0.8	0.8/1.0	0.7/0.7
Vesicular PolyAmine Transporter	28	210497	G000010791	4.41E-68	snp	Mis	0.4/1.0	0.6/0.6	0.6/1.0	0.3/0.5
Vesicular PolyAmine Transporter	28	215654	G000010791	4.41E-68	snp	Mis	0.7/1.0	0.8/0.7	0.8/1.0	0.6/0.7
Zinc knuckle family protein	387	19251	G000028484	1.33E-56	complex & snp	Mis	0.6/1.0	0.6/0.7	0.6/1.0	0.6/0.4
Zinc transporter 1	26	356372	G000006099	8.28E-139	complex & snp	Mis	0.6/1.0	0.8/0.5	0.8/1.0	0.0/0.0

*In bold, protein with putative effector functionality. **Abbreviations: Conservative inframe deletion, CIdel; Conservative inframe insertion, Clins; frameshift, Fs; Missense, Mis; Stop gained, StG; Synonymous, Syn.

Another variant was predicted to impact a splice donor site in a ubiquitin-40S ribosomal protein S27a. There was one frameshift variant in a ADP-ribose pyrophosphatase, mitochondrial protein, three genes contained an insertion or a deletion and 51 a missense variant. This included several known or putative effectors.

RNA-Seq

It was possible to isolate unmated females from eight of the nine populations (ON10 was dropped) and from 31 of the 32 remaining samples (one missing data in population ON6) A total of 2,891,067,825 paired-end reads were sequenced from these 31 samples. The number of paired reads vary from 64,571,101 to 128,822,060 in all samples. Overall, 31,408 of the 32,271 transcripts present in the *H. glycines* reference transcriptome were detected in this study. A PCA analysis confirmed that gene expression among a treatment (population*plant selection) was consistent between the replicates (Figure 5.1).

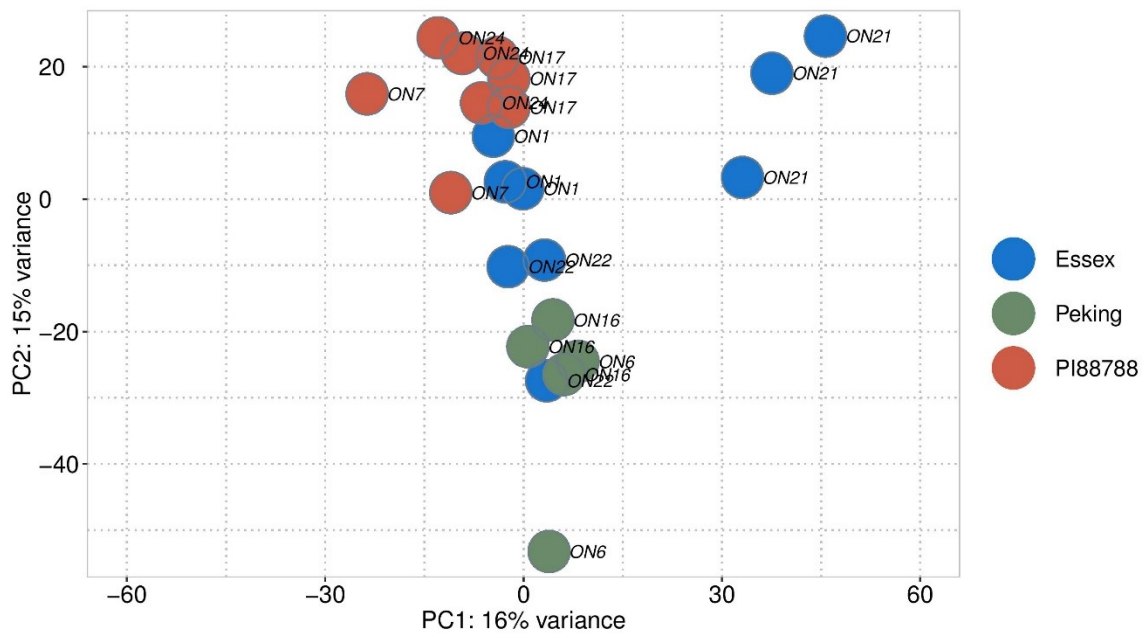


Figure 5.1 PCA of all 31,408 DEG for all samples representing single female selected on resistant (Peking or PI88788) or susceptible (Essex) soybean cultivars.

Four genes were consistently (in all replicates) expressed on Peking and not on PI88788 and eighteen on PI88788 but not on Peking (Table 5.5).

Table 5.5 Genes consistently expressed by virulent SCN individuals on a resistance-type but not the other (Peking or PI88788).

Transcript ID	Best hit description	Species	e-value	% Identity
ON6, ON16 Peking virulence				
T000015132.1	Bm-DAP-1 protein	Loa loa	5,00E-06	40%
T000010480.1	Alcohol dehydrogenase	Anoplophora glabripennis	3,00E-03	36%
T000028280.1	Dorsal gland cell-specific expression protein Hgg-20	Heterodera avenae	2,00E-03	31%
T000028278.2	No significant hit	No significant hit	-	-
ON7, ON17, ON24 on PI88799 virulence				
T000024179.1	CBN-RNR-2 protein	Caenorhabditis brenneri	3,00E-39	26%
T000011288.1	Guanine nucleotide-binding protein subunit beta-2-like 1	Loa loa	0,00E+00	78%
T000019807.2	No significant hit	No significant hit	-	-
T000027269.1	WD domain, G-beta repeat protein	Onchocerca flexuosa	0,00E+00	52%
T000023487.1	Endothelin-converting enzyme 2	Stylophora pistillata	7,00E-20	42%
T000028567.2	KRAB-A domain-containing protein 2-like isoform X1	Myzus persicae	2,00E-04	28%
T000004686.1	PREDICTED: signaling mucin MSB2	Aplysia californica	1,00E-10	36%
T000028283.1	Dorsal gland cell-specific expression protein Hgg-20	Heterodera avenae	4,00E-09	43%
T000017201.1	Single-stranded-DNA-specific exonuclease recj	Devosia submarina	7,90E-01	31%
T000023533.1	Retrotransposon-related protein	Trifolium pratense	2,00E-43	45%
T000010848.1	No significant hit	No significant hit	-	-
T000015581.1	No significant hit	No significant hit	-	-
T000025810.1	MULTISPECIES: ATP-dependent metalloproteinase ftsh/Yme1/Tma family protein	Prevotella	3,00E-03	33%
T000020888.1	No significant hit	No significant hit	-	-
T000013685.1	No significant hit	No significant hit	-	-
T000019144.2	Putative DNA polymerase	Cyphomyrmex costatus	1,00E-83	28%
T000012660.1	Dynein heavy chain, cytoplasmic	Caenorhabditis elegans	2,00E-11	78%
T000026347.1	ADP-dependent glucokinase	Solemya velum gill symbiont	1,00E-06	39%

Interestingly, these genes included different versions of a putative effector (Hgg-20) located on the same scaffold. A closer look at that genomic region revealed the presence of 6 copies of the gene in tandem duplications (Figure 5.2). The distribution of read count for these genes across all populations appear in Table 5.6.

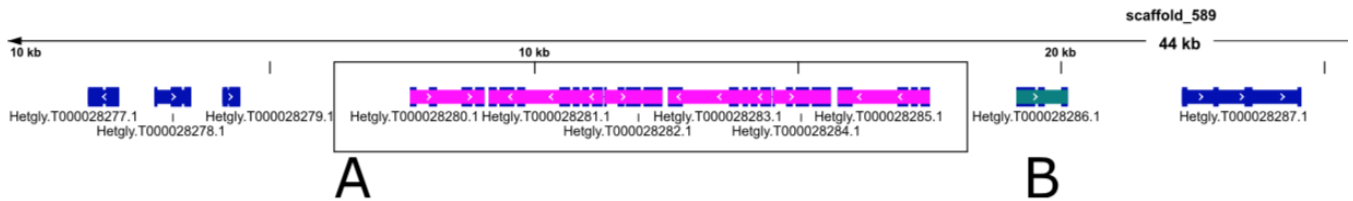


Figure 5.2 Details of gene arrangement on scaffold 589 of the *Heterodera glycines* reference genome. In A, six transcripts with similarity to Hgg-20 proteins. In B, a putative transposase gene.

Very interestingly, two close transcripts (T000028280.1 and T000028283.1) sharing similarities to the putative effector Hgg-20 showed perfect opposite expression in individuals developing on Peking and PI88788. The transcript T000028280.1 was only expressed on Peking while the second transcript T000028283.1 was only expressed on PI88788. Also, T000028283.1 was expressed in almost every females developing on Essex (no selection) and J2 pools. This clearly illustrates that the gene is present in all populations but that only the individuals not expressing the gene (or having a deletion) are able to reproduce on Peking.

A total of 485 significant ($p < 0.01$) differentially expressed genes (DEG) with a fold change ($-2 > \log(\text{FC}) > 2$) were identified when comparing females selected on Peking or PI88788 with the females selected on Essex (control group) (Suppl. table S 5.6). All these genes were differentially expressed on Peking (176 upregulated and 309 downregulated) and only 31 of them on PI88788 (25 upregulated and 6 downregulated). The most differentially expressed genes were coding for Vesicle-associated membrane protein, putative apoptosis inhibitor, several dorsal gland cell-specific expression protein, transcription factors, Ran-binding protein, Glutathione synthetase and several proteins involved in ubiquitination.

A hierarchical clustering of these DEGs identified 10 clusters based on Euclidian distance (Figure 5.3).

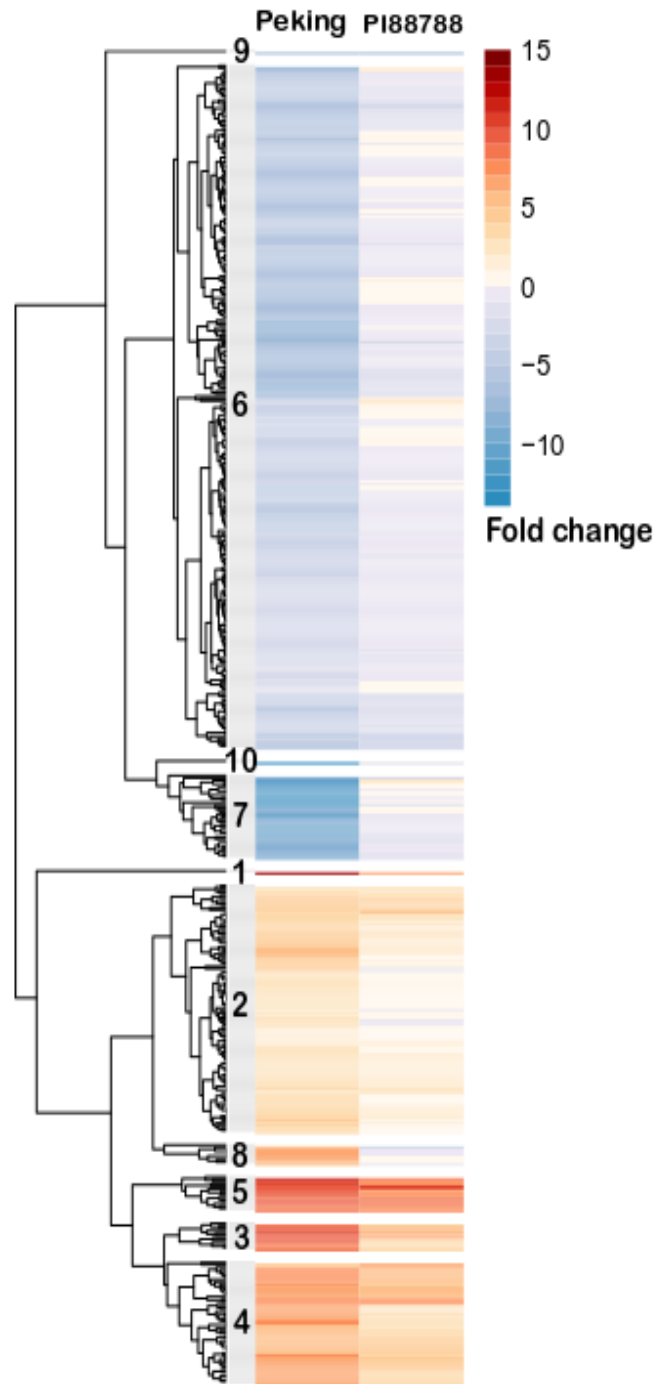


Figure 5.3 Heat map showing hierarchical clustering of the differentially expressed genes between selected female on Peking and on PI88788 in comparison with female selected on Essex. Numbers of each cluster are presented on the figure.

This heatmap showed the massive downregulation of genes (in blue) in females developing on Peking in comparison to those on the control (susceptible soybean) and on

PI88788. The females developing on PI88788 exhibited a similar transcriptome than the control except for a few upregulated genes, mainly in cluster 5. The direct comparison of the genes differentially expressed in females developing on Peking and PI8788 is also presented in Figure 5.4.

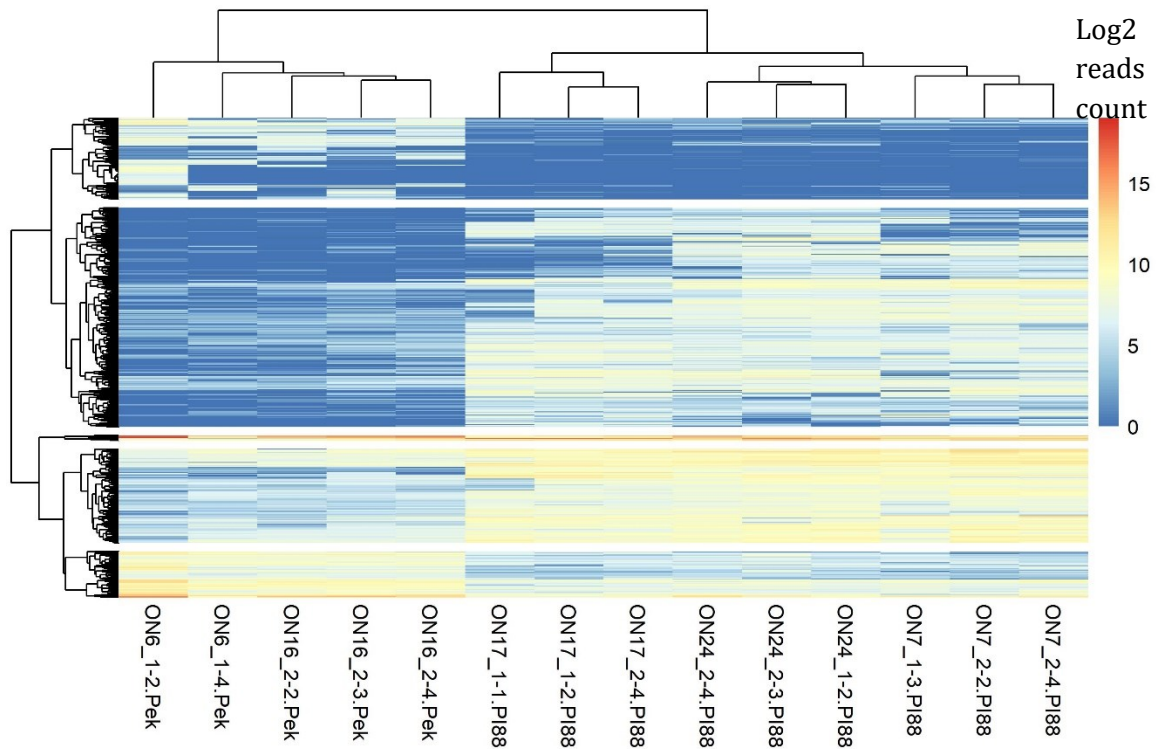


Figure 5.4 Differentially expressed genes between single *Heterodera glycines* females developing on PI88788 and Peking at J4 stage. Each sample represents a single female. ON6 and ON16 samples represent female selected on Peking and ON17, ON24 and ON7 samples represent female selected on PI88788.

5.5 Discussion

In this study, we genotyped 57 populations from across the United States and Canada with the GBS approach to find SNPs markers in regions associated with virulence on Peking or on PI88788 based on allelic imbalance. We found a total of 14 SNPs that were significantly associated with virulence. However, most of these correlations were weak or moderate. This is not surprising considering the high diversity prevailing in field populations of SCN (Faghihi *et al.*, 2010; Gendron St-Marseille *et al.*, 2018). Also, the proportion of the individuals inside a population that are virulent (according to the female

index obtained during the Hg type test) was relatively low, which considerably reduced the chance of finding significant association. This kind of approach was successfully used in different systems but with much more homogenous population (in term of virulence status, e.g. Mimee *et al.* (2015) on *Globodera rostochiensis*). Surprisingly though, the ten markers that were retrieved from the reference genome sequence were all located in genes and nine in exons. This was totally unexpected. The use of field populations is probably not stranger to this. Because of their very high genetic diversity, it is probable that the sequences outside of the coding regions have diverged radically between these populations. One marker was located in a gene identified as a putative effector in *Heterodera avenae* (Chen *et al.*, 2018). Another difficulty faced with this approach is the still fragmented nature of the reference genome that limits the analysis of physical association between this markers and neighboring genes. This will soon be overcome by the upcoming publication of a chromosome level assembly of the SCN genome (R. Masonbrick, personal communication).

In order to directly screen for gene variants associated with virulence on Peking or PI88788, we sequenced the whole genome of females (J4) successfully developing on resistant plants and compared them to a pool of unselected individuals (J2) from the same population. More than 589 genes showed significant change in allele frequency (>40%) or purifying selection (homozygous at the end) after being selected on either Peking or PI88788. This represents a massive change in the allele diversity and illustrates the strong selective pressure that the plant genotype exerts on SCN populations. Of course, most of these variants are not associated with virulence but only carried over in the genotype of virulent individuals. Still, from these variants, 126 (17 in Peking and 109 in PI88788) were predicted to have a moderate or a high impact on the subsequent protein product and justified a closer look at their putative involvement in virulence. The most striking result was the presence of several genes coding for putative effectors identified in other studies (Gems & Maizels, 1996; Davis *et al.*, 2000; Gao *et al.*, 2003; Lilley *et al.*, 2005; Haegeman *et al.*, 2012). The number of putative effectors in our variant table was clearly enriched when compared to their abundance in the reference genome. It was shown that some effectors family are still evolving rapidly (Tomalova *et al.*, 2012; Hu & Hewezi, 2018) and that in the SCN genome, some of these genes were duplicated and under strong

diversifying selection (Masonbrink *et al.*, 2019). For example, in one of these genes, coding for an aspartyl aminopeptidase, we found several variants causing non-synonymous changes in the amino-acid sequence. This protein was previously identified in SCN secretions and involved in protein degradation (Gao *et al.*, 2003; Lilley *et al.*, 2005). We also found many variants in dorsal gland cell-specific expression protein identified in SCN and other cyst nematodes. The dorsal gland proteins are of particular interest because most of the proteins involved in parasitism are produced by these specialized glands in plant parasitic nematodes (see the review of Haegeman *et al.* (2012)). Another protein, isoform of mucin, also contained a variant that was associated with female selected on Peking. This protein is implicated in host defense mechanism and in immune invasion of another nematode (*Toxocara canis*) (Gems & Maizels, 1996; Davis *et al.*, 2000; Loukas *et al.*, 2000). Overall, much more variants were found in females selected on PI88788 (503) than in Peking (86). This again included several putative effectors like annexin, esophageal gland proteins, endonucleases, guanylate cyclase, pectate lyase, peptidase and ubiquitins. From these, the gland protein G20E03 was of particular interest as it has previously been identified in other SCN researches. This protein has shown an increased expression during incompatible interactions on resistant plants (Klink *et al.*, 2013; Gardner *et al.*, 2018; Hua *et al.*, 2018). In our study, only the females homozygous at that locus (position 36124 on scaffold 34) were able to overcome resistance and to develop on PI88788. This SNP in G20E03 is causing a codon modification, changing a cysteine (polar) for a phenylalanine (non-polar) at the position 191. Thus, females overcoming PI88788 are homozygous for a nonpolar phenylalanine. This change in polarity could change the function of this protein or be implicated in host recognition. Further analysis should be performed in order to evaluate the effect of this change in polarity.

Finally, we sequenced the whole transcriptome of single females developing on the resistant soybean line Peking and PI88788 and on the susceptible control Essex. This was done using a protocol designed for single-cell sequencing. Until recently, sequencing the whole transcriptome (or genome) from a single nematode individual was impossible due to the small size of these animals and the low nucleic acid content. Some kits for whole genome amplification existed but were often disappointing by generating highly

biased results with a partial representation of the genome. Here, the sequencing resulted in a 97.3% coverage (31,408 out of 32,271) of the predicted transcripts in the *H. glycines* transcriptome (Masonbrink *et al.*, 2019). In general, comparing virulent females on Peking or PI88788 with the unselected ones on Essex revealed much more differentially expressed genes on Peking than PI88788 and a lower fold change on PI88788 for the same genes. This could be explained by the fact that hypersensitive reaction associated with resistance mechanism in Peking roots are very fast and more aggressive than in PI88788 forcing the nematode to counteract these mechanisms. As shown above, the plant genotype exerts a strong selective pressure that is not limited only to the virulence genes. The internal condition in plant cells of these different genotypes could also differ significantly, explaining the extended change in gene expression on Peking. This included several genes that were downregulated. Beside the previous reasons, since virulence genes are the usual targets for host recognition and the development of plant immunity, their expression could hinder survival (Masonbrink *et al.*, 2019). Therefore, finding effectors with reduced gene expression is not surprising.

The most interesting result of this work was the finding of two genes physically linked and specifically turned off on a given resistant soybean line while equally expressed on the susceptible control or in unselected pools of larvae. The gene transcript T000028280.1 was never expressed on PI88788 and T0000028283.1 never expressed on Peking. We thus propose that the resistance genes could recognize these proteins selectively and that their absence confers virulence on a given resistant genotype. Both transcripts encode a protein similar to a dorsal gland specific expression protein Hgg-20. While there is no specific function for this protein, Hgg-20 was suggested to be involved in plant-nematode interaction (Ithal *et al.*, 2007). A similar observation was made by Klink *et al.* (2013) which identified Hgg-20 as possibly implicated in host defenses evasion in Peking. These genes were located on a scaffold containing several contiguous genes showing very high similarity. This strongly suggests past duplication events and diversifying selection at this locus that has provided this adaptative advantages. Masonbrink *et al.* (2019) recently found, in releasing the first complete assembly of the *H. glycines* genome that there were a very high number of duplication events in this nematode species (34% of the genome showed duplication events). They also found that

effectors were frequently associated with tandem duplications, DNA transposon and LTR-retrotransposon. Interestingly, they identified 144 regions presenting duplications containing Hgg-20-related sequences. In our case, this specific region on scaffold 589 also contains a putative transposase-encoding gene after the tandem duplication. Although region-specific duplication are not always associated with effectors (Masonbrink *et al.*, 2019), this specific region, and transcript levels might be one of the factors explaining the virulence of these SCN populations on Peking and PI88788. Gardner *et al.* (2017) recently generated a lab population of SCN that was virulent on the broad spectrum resistance source PI437654. This hyper-virulent population was able to overcome both PI88788 and Peking resistance. It would be interesting to see if these nematodes lacks both of the transcripts that we identified in the present study. With the numerous and complex mutations and DEGs that we found in this study, it is likely that virulence is linked to an intertwined genetic background of differential gene expression and/or mutant effector genes.

5.6 Conclusion

To our knowledge, this is the first study using “single-cell” sequencing for RNA-Seq study on SCN. Working with individual nematode allowed the isolation of specific effector genes related to plant-nematode interactions that are putatively associated with virulence on resistant soybean genotypes. Both the GBS and WGS approaches highlighted numerous variants with several being statistically associated with virulence. This illustrates again the high genetic diversity of *Heterodera glycines* populations and the complex and active selection forces acting at the population level. This was also shown by various authors and complicated the search of specific effectors responsible for nematode virulence on soybean in the past. However, using this new “single-cell” sequencing now opens new opportunity to study the complete transcriptome (or genome) at the individual level and to pin point specific signal responsible for a given phenotype.

5.7 References

- Allen, T. W., Bradley, C. A., Sisson, A. J., Byamukama, E., Chilvers, M. I., Coker, C. M., . . . Wrather, J. A. (2017). Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2010 to 2014. *Plant Health Progress*, *18*(1), 19-27. doi:10.1094/php-rs-16-0066
- Allen, T. W., Bradley, C. A., Sisson, A. J., Byamukama, E., Chilvers, M. I., Coker, C. M., . . . Wrather, J. A. (2016). First Report of Soybean Cyst Nematode (*Heterodera glycines*) on Soybean from Gansu and Ningxia China. *Plant Disease*, *100*(1), 229-229. doi:10.1094/pdis-04-15-0451-pdn
- Anand, S. C., & Brar, G. S. (1983). Response of Soybean Lines to Differentially Selected Cultures of Soybean Cyst Nematode *Heterodera glycines* Ichinohe. *Journal of nematology*, *15*(1), 120-123.
- Anderson, T. R., Welacky, T. W., Olechowski, H. T., Ablett, G., & Ebsary, B. A. (1988). First report of *Heterodera glycines* on soybeans in Ontario, Canada. *Plant Disease*, *72*, 453.
- Bayless, A. M., Smith, J. M., Song, J., McMinn, P. H., Teillet, A., August, B. K., & Bent, A. F. (2016). Disease resistance through impairment of α -SNAP–NSF interaction and vesicular trafficking by soybean *Rhg1*. *Proceedings of the National Academy of Sciences*, *113*(47), E7375-E7382. doi:10.1073/pnas.1610150113
- Bekal, S., Domier, L. L., Gonfa, B., Lakhssassi, N., Meksem, K., & Lambert, K. N. (2015). A SNARE-Like Protein and Biotin Are Implicated in Soybean Cyst Nematode Virulence. *PLOS ONE*, *10*(12), e0145601. doi:10.1371/journal.pone.0145601
- Bolger, A. M., Usadel, B., & Lohse, M. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633-2635. doi:10.1093/bioinformatics/btm308
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*, 525. doi:10.1038/nbt.3519
- Brucker, E., Carlson, S., Wright, E., Niblack, T., & Diers, B. (2005). *Rhg1* alleles from soybean PI 437654 and PI 88788 respond differentially to isolates of *Heterodera glycines* in the greenhouse. *Theoretical and Applied Genetics*, *111*(1), 44-49. doi:10.1007/s00122-005-1970-3
- Chen, C., Chen, Y., Jian, H., Yang, D., Dai, Y., Pan, L., . . . Liu, Q. (2018). Large-Scale Identification and Characterization of *Heterodera avenae* Putative Effectors

- Suppressing or Inducing Cell Death in *Nicotiana benthamiana*. *Frontiers in Plant Science*, 8(2062). doi:10.3389/fpls.2017.02062
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80-92. doi:10.4161/fly.19695
- Cook, D. E., Bayless, A. M., Wang, K., Guo, X., Song, Q., Jiang, J., & Bent, A. F. (2014). Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie *Rhg1*-Mediated Soybean Resistance to Soybean Cyst Nematode. *Plant Physiology*, 165(2), 630-647. doi:10.1104/pp.114.235952
- Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., . . . Bent, A. F. (2012). Copy Number Variation of Multiple Genes at *Rhg1* Mediates Nematode Resistance in Soybean. *Science*, 338(6111), 1206. doi:10.1126/science.1228746
- Davis, E. L., Hussey, R. S., Baum, T. J., Bakker, J., Schots, A., Rosso, M.-N., & Abad, P. (2000). Nematode Parasitism Genes. *Annual Review of Phytopathology*, 38(1), 365-396. doi:10.1146/annurev.phyto.38.1.365
- Dong, K., & Opperman, C. H. (1997). Genetic Analysis of Parasitism in the Soybean Cyst Nematode *Heterodera glycines*. *Genetics*, 146(4), 1311.
- Eisenback, J. (2000). Techniques for measuring nematode development and egg production. In T. Wheeler (Ed.), *Laboratory Techniques in Nematode Ecology* (pp. 1-4). Hyattsville, MD: Society of Nematologists.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE*, 6(5), e19379. doi:10.1371/journal.pone.0019379
- Faghihi, J., Donald, P. A., Noel, G., Welacky, T. W., & Ferris, V. R. (2010). Soybean Resistance to Field Populations of *Heterodera glycines* in Selected Geographic Areas. *Plant Health Progress*, 11(1), 19. doi:10.1094/php-2010-0426-01-rs
- Gao, B., Allen, R. J., Maier, T., Davis, E. L., Baum, T. J., & Hussey, R. I. (2003). The parasitome of the phytonematode *Heterodera glycines*. *Molecular plant-microbe interactions*, 16(8), 720-726.
- Gardner, M., Dhroso, A., Johnson, N., Davis, E. L., Baum, T. J., Korkein, D., & Mitchum, M. G. (2018). Novel global effector mining from the transcriptome of early life stages of the soybean cyst nematode *Heterodera glycines*. *Scientific Reports*, 8(1), 2505. doi:10.1038/s41598-018-20536-5

- Gardner, M., Heinz, R., Wang, J., & Mitchum, M. G. (2017). Genetics and Adaptation of Soybean Cyst Nematode to Broad Spectrum Soybean Resistance. *G3: Genes|Genomes|Genetics*, 7(3), 835. doi:10.1534/g3.116.035964
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing (Version v.2): arXiv preprint
- Gems, D., & Maizels, R. M. (1996). An abundantly expressed mucin-like protein from *Toxocara canis* infective larvae: the precursor of the larval surface coat glycoproteins. *Proceedings of the National Academy of Sciences*, 93(4), 1665. doi:10.1073/pnas.93.4.1665
- Gendron St-Marseille, A.-F., Bourgeois, G., Brodeur, J., & Mimee, B. (2019). Simulating the impacts of climate change on soybean cyst nematode and the distribution of soybean. *Agricultural and Forest Meteorology*, 264, 178-187. doi:10.1016/j.agrformet.2018.10.008
- Gendron St-Marseille, A. F., Lord, E., Veronneau, P. Y., Brodeur, J., & Mimee, B. (2018). Genome Scans Reveal Homogenization and Local Adaptations in Populations of the Soybean Cyst Nematode. *Front Plant Sci*, 9, 987. doi:10.3389/fpls.2018.00987
- Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., . . . Dopazo, J. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36(10), 3420-3435. doi:10.1093/nar/gkn176
- Haegeman, A., Mantelin, S., Jones, J. T., & Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene*, 492(1), 19-31. doi:<https://doi.org/10.1016/j.gene.2011.10.040>
- Howland, A., Monnig, N., Mathesius, J., Nathan, M., & Mitchum, M. G. (2018). Survey of *Heterodera glycines* Population Densities and Virulence Phenotypes During 2015-2016 in Missouri. *Plant Dis*, 102(12), 2407-2410. doi:10.1094/PDIS-04-18-0650-SR
- Hu, Y., & Hewezi, T. (2018). Nematode-secreted peptides and host factor mimicry. *Journal of Experimental Botany*, 69(12), 2866-2868. doi:10.1093/jxb/ery144
- Hua, C., Li, C., Hu, Y., Mao, Y., You, J., Wang, M., . . . Wang, C. (2018). Identification of HG Types of Soybean Cyst Nematode *Heterodera glycines* and Resistance Screening on Soybean Genotypes in Northeast China. *J Nematol*, 50(1), 41-50. doi:10.21307/jofnem-2018-007
- Ithal, N., Recknor, J., Nettleton, D., Hearne, L., Maier, T., Baum, T. J., & Mitchum, M. G. (2007). Parallel Genome-Wide Expression Profiling of Host and Pathogen During Soybean Cyst Nematode Infection of Soybean. *Molecular Plant-Microbe Interactions*, 20(3), 293-305. doi:10.1094/MPMI-20-3-0293

- Jamil, M., Ali, A., Gul, A., Ghafoor, A., Ibrahim, A. M. H., & Mujeeb-Kazi, A. (2018). Genome-Wide Association Studies for Spot Blotch (*Cochliobolus sativus*) Resistance in Bread Wheat Using Genotyping-by-Sequencing. *Phytopathology*, *108*(11), 1307-1314. doi:10.1094/PHYTO-02-18-0047-R
- Kim, M., Hyten, D. L., Bent, A. F., & Diers, B. (2010). Fine mapping of the SCN resistance locus rhg1-b from PI88788. *The Plant Genome*, *3*(2), 81-89.
- Klink, V. P., Matsye, P. D., Lawrence, K. S., & Lawrence, G. W. (2013). *Engineered soybean cyst nematode resistance*: InTech.
- Kolde, R. (2019). pheatmap: Pretty Heatmaps (Version 1.0.12). Retrieved from <https://CRAN.R-project.org/package=pheatmap>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324
- Li, H., Genome Project Data Processing, S., Wysoker, A., Handsaker, B., Marth, G., Abecasis, G., . . . Fennell, T. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352
- Lilley, C. J., Goodchild, S. A., Atkinson, H. J., & Urwin, P. E. (2005). Cloning and characterisation of a Heterodera glycines aminopeptidase cDNA. *International Journal for Parasitology*, *35*(14), 1577-1585. doi:<https://doi.org/10.1016/j.ijpara.2005.07.017>
- Liu, S., Kandath, P. K., Warren, S. D., Yeckel, G., Heinz, R., Alden, J., . . . Meksem, K. (2012). A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature*, *492*, 256. doi:10.1038/nature11651
- Loukas, A., Hintz, M., Linder, D., Mullin, N. P., Parkinson, J., Tetteh, K. K. A., & Maizels, R. M. (2000). A Family of Secreted Mucins from the Parasitic Nematode *Toxocara canis* Bears Diverse Mucin Domains but Shares Similar Flanking Six-cysteine Repeat Motifs. *Journal of Biological Chemistry*, *275*(50), 39600-39607.
- Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., . . . Costich, D. E. (2013). Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLOS Genetics*, *9*(1), e1003215. doi:10.1371/journal.pgen.1003215
- Masonbrink, R., Maier, T. R., Muppirala, U., Seetharam, A. S., Lord, E., Juvale, P. S., . . . Baum, T. J. (2019). The genome of the soybean cyst nematode (*Heterodera glycines*) reveals complex patterns of duplications involved in the evolution of parasitism genes. *BMC Genomics*, *20*(1), 119. doi:10.1186/s12864-019-5485-8

- Matsuo, É., Sedyama, T., D'Arc de Lima Oliveira, R., Damião Cruz, C., & de Cássia Teixeira Oliveira, R. (2012). Characterization of type and genetic diversity among soybean cyst nematode differentiators. *Ciencia Agricola*, *69*(2), 147-151.
- McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288-4297. doi:10.1093/nar/gks042
- McCarville, M. T., Marett, C. C., Mullaney, M. P., Gebhart, G. D., & Tylka, G. L. (2017). Increase in Soybean Cyst Nematode Virulence and Reproduction on Resistant Soybean Varieties in Iowa From 2001 to 2015 and the Effects on Soybean Yields. *Plant Health Progress*, *18*(3), 146-155. doi:10.1094/php-rs-16-0062
- Mimee, B., Duceppe, M.-O., Véronneau, P.-Y., Lafond-Lapalme, J., Jean, M., Belzile, F., & Bélair, G. (2015). A new method for studying population genetics of cyst nematodes based on Pool-Seq and genomewide allele frequency analysis. *Molecular Ecology Resources*, *15*(6), 1356-1365. doi:10.1111/1755-0998.12412
- Mimee, B., Peng, H., Popovic, V., Yu, Q., Duceppe, M. O., Tétreault, M. P., & Belair, G. (2013). First Report of Soybean Cyst Nematode (*Heterodera glycines* Ichinohe) on Soybean in the Province of Quebec, Canada. *Plant Disease*, *98*(3), 429-429. doi:10.1094/PDIS-07-13-0782-PDN
- Mitchum, M. G. (2016). Soybean Resistance to the Soybean Cyst Nematode *Heterodera glycines*: An Update. *Phytopathology*, *106*(12), 1444-1450. doi:10.1094/PHYTO-06-16-0227-RVW
- Mitchum, M. G., Hussey, R. S., Baum, T. J., Wang, X., Elling, A. A., Wubben, M., & Davis, E. L. (2013). Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist*, *199*(4), 879-894. doi:10.1111/nph.12323
- Niblack, T. L. (2005). Soybean Cyst Nematode Management Reconsidered. *Plant Disease*, *89*(10), 1020-1026. doi:10.1094/PD-89-1020
- Niblack, T. L., Arelli, P. R., Noel, G. R., Opperman, C. H., Orf, J. H., Schmitt, D. P., . . . Tylka, G. L. (2002). A Revised Classification Scheme for Genetically Diverse Populations of *Heterodera glycines*. *Journal of nematology*, *34*(4), 279-288.
- Niblack, T. L., Colgrove, A. L., Colgrove, K., & Bond, J. P. (2008). Shift in Virulence of Soybean Cyst Nematode is Associated with Use of Resistance from PI 88788. *Plant Health Progress*, *9*(1), 29. doi:10.1094/php-2008-0118-01-rs
- R Core Team. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

- Rincker, K., Cary, T., & Diers, B. W. (2017). Impact of Soybean Cyst Nematode Resistance on Soybean Yield. *Crop Science*, 57, 1373-1382. doi:10.2135/cropsci2016.07.0628
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.
- Rovelli, V., Ruiz-González, A., Vignoli, L., Macale, D., Davoli, F., Vieites, D., . . . Buono, V. (2018). *Genotyping-by-Sequencing (GBS) of large amphibian genomes: A comparative study of two non-model species endemic to Italy*.
- Sacco, M. A., Koropacka, K., Grenier, E., Jaubert, M. J., Blanchard, A., Goverse, A., . . . Moffett, P. (2009). The Cyst Nematode SPRYSEC Protein RBP-1 Elicits Gpa2- and RanGAP2-Dependent Plant Cell Death. *PLOS Pathogens*, 5(8), e1000564. doi:10.1371/journal.ppat.1000564
- Serra, L., Chang, D. Z., Macchietto, M., Williams, K., Murad, R., Lu, D., . . . Mortazavi, A. (2018). Adapting the Smart-seq2 Protocol for Robust Single Worm RNA-seq. *Bio-protocol*, 8(4), e2729. doi:10.21769/BioProtoc.2729
- Sonah, H., O'Donoghue, L., Cober, E., Rajcan, I., & Belzile, F. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnology Journal*, 13(2), 211-221. doi:10.1111/pbi.12249
- Tomalova, I., Iachia, C., Mulet, K., & Castagnone-Sereno, P. (2012). The map-1 Gene Family in Root-Knot Nematodes, *Meloidogyne* spp.: A Set of Taxonomically Restricted Genes Specific to Clonal Species. *PLOS ONE*, 7(6), e38656. doi:10.1371/journal.pone.0038656
- Tylka, G. L., & Marett, C. C. (2017). Known Distribution of the Soybean Cyst Nematode, *Heterodera glycines*, in the United States and Canada, 1954 to 2017. *Plant Health Progress*, 18(3), 167-168. doi:10.1094/php-05-17-0031-br
- Tylka, G. L., & Mullaney, M. P. (2015). Soybean cyst nematode-resistant soybean varieties for Iowa. *Coop. Ext. Publ.* Retrieved from <https://store.extension.iastate.edu/Product/Soybean-cyst-nematode-resistant-soybean-varieties-for-Iowa-pdf>
- Wang, J., Niblack, T. L., Tremain, J. A., Wiebold, W. J., Tylka, G. L., Marett, C. C., . . . Schmitt, D. P. (2003). Soybean cyst nematode reduces soybean yield without causing obvious aboveground symptoms. *Plant Disease*, 87, 623-628.
- Williamson, V. M., & Kumar, A. (2006). Nematode resistance in plants: the battle underground. *Trends in Genetics*, 22(7), 396-403. doi:<https://doi.org/10.1016/j.tig.2006.05.003>

- Winstead, N. N., Skotland, C. B., & Sasser, J. N. (1955). Soybean cyst nematode in North Carolina. *Plant Disease Reporter*, 39(1), 9-11.
- Yuan, C.-P., Li, Y.-H., Liu, Z.-X., Guan, R.-X., Chang, R.-Z., & Qiu, L.-J. (2012). DNA sequence polymorphism of the Rhg4 candidate gene conferring resistance to soybean cyst nematode in Chinese domesticated and wild soybeans. *Molecular breeding : new strategies in plant improvement*, 30(2), 1155-1162. doi:10.1007/s11032-012-9703-1
- Zheng, J. W., & Chen, S. Y. (2013). Estimation of virulence type and level of soybean cyst nematode field populations in response to resistant cultivars. *International Journal of Nematology and Entomology*, 1(6), 111-116.

5.8 Supplementary material

Table S. 5.1 Variants associated with a change higher than 40 % between allele frequencies calculated from the founder population (J2) and from females (J4) selected on Peking whereas selection on PI88788 didn't influenced the allelic frequencies between founders and selected females (differences are less than 20%).

Scaffold ID	Position	Hetgly Transcript ID	Variant type ¹	Blast2GO best hit description	e-Value	Variant type ²	Variant impact ³	Allelic frequencies			
								IL4 Founder-PI88788	ON1 Founder-PI88788	ON1 Founder-Peking	ON34 Founder-Peking
665	3200	T000028947.1	complex, ins, snp	S-antigen, putative	5,28E-06	Fs, Fs & Mis, Mis	H, H, M	0.4-0.4	0.7-0.8	0.7-0.3	0.9-0.3
9	173692	T000005787.1	del, snp	NA	NA	Mis, Cldel	M, M	0.6-0.5	0.4-0.3	0.4-0.0	1.0-0.5
111	118624	T000014217.1	complex, snp	aspartyl aminopeptidase	6,42E-177	Mis, Syn	M, L	0.7-0.5	0.8-0.7	0.8-0.3	1.0-0.3
111	118630	T000014217.1	complex, mnp, snp	aspartyl aminopeptidase	6,42E-177	Mis, Mis, Mis	M, M, M	0.7-0.5	0.8-0.7	0.8-0.3	1.0-0.2
336	4884	T000026873.1	complex, snp, snp	SH3 domain protein	2,51E-43	Mis, Syn, Syn	M, L, L	0.2-0.1	0.4-0.4	0.4-0.0	0.9-0.2
111	118658	T000014217.1	complex, complex	aspartyl aminopeptidase	6,42E-177	Syn, Syn	L, L	0.6-0.6	0.8-0.8	0.8-0.3	1.0-0.3
207	125706	T000021145.1	del, del, del	NA	NA	SpR & Int, Int, Int	L, Modif, Modif	0.4-0.3	0.6-0.6	0.6-1.0	0.6-1.0
678	1558	T000029539.1	snp	hypothetical protein Y032_0039g47	2,29E-20	Syn	L	0.5-0.7	0.6-0.5	0.6-1.0	0.2-0.8
1	76812	T000000441.1	snp	NA	NA	Int	Modif	0.9-1.0	1.0-0.8	1.0-0.5	1.0-0.3
23	269825	T000006985.2	ins, snp	Homolog of Odr-2 (Two)	7,69E-44	Int, Int	Modif, Modif	0.6-0.6	0.6-0.5	0.6-0.0	0.8-0.1
58	267287	T000013136.1	snp	kinesin motor domain protein	2,74E-139	Int	Modif	0.6-0.8	0.8-0.8	0.8-0.3	0.8-0.2

61	108021	T000013434.1	complex	titin homolog	8,92E-05	Int	Modif	0.9-0.8	1.0-1.0	1.0-0.5	1.0-0.6
219	22776	T000024004.1	snp	Exocyst complex component 7	1,59E-40	Int, Int	Modif, Modif	0.7-0.9	0.9-0.9	0.9-0.4	0.7-0.3
576	27343	T000028401.1	ins	Ral GDS-Like	3,79E-36	Int	Modif	0.3-0.4	0.7-0.9	0.7-0.3	0.6-0.1
22	377325	NA	ins, mnp, snp	NA	NA	Inter	NA	0.6-0.6	0.8-1.0	0.8-0.4	0.8-0.2
49	263308	NA	complex, snp, snp	NA	NA	Inter	NA	0.6-0.6	0.5-0.6	0.5-0.9	0.5-0.9
51	474889	NA	snp, snp	NA	NA	Inter	NA	0.6-0.6	0.4-0.3	0.4-0.0	1.0-0.6
77	255811	NA	del	NA	NA	Inter	NA	0.7-0.9	0.7-0.8	0.7-0.2	1.0-0.5
89	54457	NA	complex, snp	NA	NA	Inter	NA	1.0-1.0	1.0-0.8	1.0-0.4	0.4-0.0
190	48762	NA	complex, ins, snp	NA	NA	Inter	NA	0.6-0.6	0.7-0.7	0.7-0.0	0.8-0.3
266	13654	NA	del, del, snp	NA	NA	Inter	NA	0.7-0.6	0.4-0.6	0.4-0.9	0.6-1.0

¹ Complex = multiple type of nucleotide changes; del = deletion; ins = insertion; mnp = multiple nucleotide polymorphism; snp = single nucleotide polymorphism.

² Conservative inframe deletion = Cidel; conservative inframe insertion = Clins; frameshift = Fs; missense = Mis; intergenic = Inter; intron = Int; splice region variant = SpR; synonymous = Syn.

³ High = H; Low = L; Moderate = M; Modifier = Mod.

Table S. 5.2 Variants associated with an increasing allele frequency toward homozygous variants (AF < 0.05 or > 0.95) associated with females selected on Peking whereas alleles of females selected on PI88788 remain unstable.

Scaffold ID	Position	Hetgly Transcript ID	Variant type ¹	Blast2GO best hit description	e-Value	Variant type ²	Variant impact ³	Allelic frequencies			
								IL4 Founder-PI88788	ON1 Founder-PI88788	ON1 Founder-Peking	ON34 Founder-Peking
3	536266	T000002997.2	snp	protein, SNF2 family	0	StG, StG	H, H	0.7-0.7	0.7-0.8	0.7-1.0	0.7-1.0
15	371176	T000007782.2	complex, complex, snp	integrase core domain protein	1,54E-178	StG, Syn, Syn	H, L, L	0.7-0.8	0.8-0.8	0.8-1.0	0.8-1.0
69	295950	T000017119.1	complex, snp	NA	NA	Fs & Mis, Mis	H, M	0.6-0.8	0.8-0.8	0.8-1.0	0.6-1.0
72	116110	T000018058.1	complex, complex, snp	BTB/POZ domain-containing protein 2	6,11E-37	Fs & Syn, Syn, Syn	H, L, L	0.6-0.8	0.8-0.6	0.8-1.0	0.7-1.0
95	168408	T000016215.1	complex, del, snp	dorsal gland cell-specific expression protein	1,15E-38	Fs, Mis, Mis	H, M, M	0.8-0.8	0.8-0.8	0.8-1.0	0.7-1.0
3	163715	T000002897.1	snp	hypothetical protein B7P43_G14789	2,43E-124	Mis	M	0.1-0.5	0.2-0.2	0.2-0.0	0.2-0.1
9	777418	T000005953.1	snp	oxidoreductase, short chain dehydrogenase/reductase family protein	3,07E-58	Mis	M	1.0-0.5	0.8-0.7	0.8-1.0	0.7-1.0
15	589003	T000007828.2	complex	NA	NA	Mis, Mis	M, M	0.6-0.6	0.2-0.6	0.2-0.0	0.2-0.0
85	354141	T000013910.1	snp	TATA-box binding protein	7,71E-07	Mis	M	0.8-0.8	0.8-0.7	0.8-1.0	0.8-1.0
120	11736	T000013014.1	complex	Protein Y45F3A.4	1,22E-23	Mis	M	0.8-0.6	0.8-0.5	0.8-1.0	0.8-1.0
159	137849	T000019407.1	complex, del, ins	dorsal gland cell-specific expression protein	7,92E-12	Mis & Clins, Clins, Cidel	M, M, M	0.8-0.8	0.8-0.7	0.8-1.0	0.8-1.0
208	52180	T000021751.1	snp	mucin-2 isoform X1	9,94E-142	Mis	M	0.9-0.7	0.6-0.8	0.6-1.0	0.8-1.0
3	163707	T000002897.1	snp	hypothetical protein B7P43_G14789	2,43E-124	Syn	L	0.1-0.4	0.2-0.2	0.2-0.0	0.3-0.0
124	32650	T000012803.1	snp	Bromodomain protein	4,93E-119	Syn	L	0.8-0.7	0.8-0.6	0.8-1.0	0.7-1.0
207	125706	T000021145.1	del, del, del	NA	NA	SpR & Int, Int, Int	L, Mod, Mod	0.4-0.3	0.6-0.6	0.6-1.0	0.6-1.0
2	545235	T000001988.1	ins, ins, snp	c-terminal tandem repeated domain in type 4 procollagen	0	Int, Int, Int	Mod, Mod,	0.5-0.5	0.8-0.8	0.8-1.0	0.8-1.0

3	1281434	T000003165.1	del	NA	NA	Int	Mod	0.9-0.8	0.8-0.6	0.8-1.0	0.4-1.0
4	1001968	T000000330.1	del, ins, ins	protein canopy homolog 4	6,85E-26	Int, Int, Int	Mod, Mod, Mod	0.4-0.7	0.3-0.5	0.3-0.0	0.6-0.0
15	859770	T000007892.1	snp	hypothetical protein IQ06DRAFT_303516	4,14E-06	Int	Mod	1.0-0.6	0.8-0.6	0.8-1.0	0.7-1.0
24	699431	T000009027.1	snp	NA	NA	Int	Mod	1.0-0.8	0.7-0.7	0.7-1.0	0.6-1.0
31	137246	T000008327.1	snp	rab GTPase-activating protein 1- like isoform X1	0	Int	Mod	0.6-0.6	0.8-0.8	0.8-1.0	0.8-1.0
32	335605	T000013298.1	del	NA	NA	Int	Mod	0.8-0.6	0.8-0.8	0.8-1.0	0.7-1.0
58	15452	T000013099.1	snp	Protein ZK177.4	1,32E-15	Int	Mod	0.5-0.6	0.7-0.5	0.7-1.0	0.7-1.0
86	166909	T000010562.1	complex, ins, snp	leucine Rich Repeat family protein	0	Int, Int, Int	Mod, Mod, Mod	1.0-0.8	0.7-0.7	0.7-1.0	0.8-1.0
153	47105	T000017239.1	snp	Protein CBG01955	1,11E-46	Int	Mod	0.4-0.8	0.6-0.8	0.6-1.0	0.7-1.0
250	68947	T000025432.1	snp	Glycoside hydrolase domain containing protein	5,51E-135	Int	Mod	0.9-0.6	0.7-0.5	0.7-1.0	0.6-1.0
393	78126	T000025477.1	snp	transmembrane amino acid transporter protein	4,12E-85	Int	Mod	0.5-0.4	0.8-0.7	0.8-1.0	0.6-1.0
1	614167	NA	snp	NA	NA	Inter	NA	0.6-0.3	0.8-0.8	0.8-1.0	0.8-1.0
3	207767	NA	del, snp	NA	NA	Inter	NA	0.8-0.8	0.8-0.8	0.8-1.0	0.8-1.0
3	207790	NA	snp	NA	NA	Inter	NA	0.7-0.7	0.8-0.6	0.8-1.0	0.8-1.0
4	415319	NA	snp	NA	NA	Inter	NA	0.0-0.8	0.6-0.7	0.6-0.0	0.7-0.0
9	524834	NA	snp	NA	NA	Inter	NA	0.8-0.8	0.8-0.7	0.8-1.0	0.8-1.0
13	397886	NA	snp	NA	NA	Inter	NA	0.9-0.8	0.8-0.4	0.8-1.0	0.8-1.0
13	397905	NA	snp	NA	NA	Inter	NA	0.9-0.8	0.8-0.3	0.8-1.0	0.8-1.0
18	21094	NA	complex, complex, snp	NA	NA	Inter	NA	0.8-0.8	0.7-0.8	0.7-1.0	0.8-1.0
24	542283	NA	del, snp	NA	NA	Inter	NA	0.6-0.7	0.8-0.6	0.8-1.0	0.7-1.0
27	291911	NA	snp	NA	NA	Inter	NA	0.0-0.4	0.7-0.6	0.7-1.0	0.8-1.0
31	170152	NA	snp	NA	NA	Inter	NA	0.8-0.5	0.5-0.6	0.5-0.0	0.3-0.0
38	78481	NA	snp, snp	NA	NA	Inter	NA	0.7-0.7	0.8-0.7	0.8-1.0	0.7-1.0
48	169327	NA	complex	NA	NA	Inter	NA	0.7-0.6	0.6-0.7	0.6-1.0	0.8-1.0
48	198695	NA	ins	NA	NA	Inter	NA	1.0-0.7	0.8-0.8	0.8-1.0	0.8-1.0
48	198709	NA	snp	NA	NA	Inter	NA	1.0-0.6	0.8-0.8	0.8-1.0	0.8-1.0
62	417488	NA	snp	NA	NA	Inter	NA	0.2-0.2	0.4-0.3	0.4-0.0	0.4-0.0

62	417517	NA	complex, complex	NA	NA	Inter	NA	0.3-0.4	0.5-0.2	0.5-0.0	0.4-0.0
64	415059	NA	snp	NA	NA	Inter	NA	0.9-0.7	0.8-0.4	0.8-1.0	0.6-1.0
70	364792	NA	del, ins	NA	NA	Inter	NA	0.9-0.6	0.7-0.8	0.7-1.0	0.7-1.0
72	130497	NA	snp	NA	NA	Inter	NA	0.8-0.7	0.8-0.5	0.8-1.0	0.8-1.0
103	31465	NA	complex, del	NA	NA	Inter	NA	0.9-0.8	0.8-0.4	0.8-1.0	0.7-1.0
112	99143	NA	ins, ins, ins	NA	NA	Inter	NA	0.7-0.8	0.8-0.6	0.8-1.0	0.7-1.0
156	143924	NA	ins	NA	NA	Inter	NA	0.7-0.7	0.8-0.7	0.8-1.0	0.8-1.0
157	14342	NA	snp	NA	NA	Inter	NA	0.8-0.7	0.7-0.7	0.7-1.0	0.8-1.0
157	14345	NA	complex, complex	NA	NA	Inter	NA	0.7-0.7	0.7-0.7	0.7-1.0	0.8-1.0
157	14358	NA	ins	NA	NA	Inter	NA	0.8-0.7	0.7-0.7	0.7-1.0	0.7-1.0
173	92738	NA	snp	NA	NA	Inter	NA	1.0-0.7	0.8-0.7	0.8-1.0	0.8-1.0
234	120663	NA	snp	NA	NA	Inter	NA	0.8-0.7	0.8-0.7	0.8-1.0	0.3-1.0
234	120727	NA	complex, complex, complex	NA	NA	Inter	NA	0.9-0.7	0.8-0.7	0.8-1.0	0.8-1.0
386	4179	NA	snp	NA	NA	Inter	NA	0.9-0.8	0.8-0.8	0.8-1.0	0.8-1.0
386	4191	NA	complex, complex	NA	NA	Inter	NA	0.9-0.8	0.8-0.8	0.8-1.0	0.6-1.0
386	4215	NA	del	NA	NA	Inter	NA	0.9-0.8	0.7-0.8	0.7-1.0	0.6-1.0
402	63589	NA	complex	NA	NA	Inter	NA	0.8-0.8	0.7-0.7	0.7-1.0	0.8-1.0
420	28453	NA	ins, ins, snp	NA	NA	Inter	NA	0.7-0.4	0.5-0.7	0.5-0.0	0.4-0.0
420	28461	NA	complex, complex, del	NA	NA	Inter	NA	0.3-0.2	0.3-0.4	0.3-0.0	0.2-0.0
425	16095	NA	complex, snp	NA	NA	Inter	NA	0.9-0.7	0.7-0.8	0.7-1.0	0.7-1.0
561	23345	NA	snp	NA	NA	Inter	NA	0.9-0.3	0.8-0.5	0.8-1.0	0.3-1.0
594	7509	NA	snp	NA	NA	Inter	NA	0.3-0.3	0.4-0.7	0.4-0.0	0.3-0.0

¹ Complex = multiple type of nucleotide changes; del = deletion; ins = insertion; mnp = multiple nucleotide polymorphism; snp = single nucleotide polymorphism.

² Conservative inframe deletion = Cidel; conservative inframe insertion = Cliins; frameshift = Fs; missense = Mis; intergenic = Inter; intron = Int; splice region variant = SpR; stop gained = StG; synonymous = Syn.

³ High = H; Low = L; Moderate = M; Modifier = Mod.

Table S. 5.3 Variants associated with a change higher than 40 % between allele frequencies calculated from the founder population (J2) and from females (J4) selected on PI88788 whereas selection on Peking didn't influenced the allelic frequencies between founders and selected females (differences are less than 20%).

Scaffold ID	Position	Hetgly Transcript ID	Variant type ¹	Blast2GO best hit description	e-Value	Effect type ²	Variant impact ³	Allelic frequencies			
								IL4 Founder-PI88788	ON1 Founder-PI88788	ON1 Founder-Peking	ON34 Founder-Peking
1	1816544	T000000890.1	complex, ins	NA	NA	Fs	H	0.5-1.0	0.6-1.0	0.6-0.6	1.0-0.8
233	119219	T000024134.1	complex, ins, ins	NA	NA	Fs	H	1.0-0.3	0.8-0.2	0.8-0.8	1.0-1.0
233	119227	T000024134.1	complex, snp	NA	NA	StG	H	1.0-0.3	0.8-0.2	0.8-0.8	1.0-1.0
233	119249	T000024134.1	complex, ins	NA	NA	Fs, Fs & Mis	H, H	1.0-0.3	0.8-0.1	0.8-0.7	1.0-1.0
246	81564	T000024555.1	mnp, snp, snp	ubiquitin-40S ribosomal protein S27a	3,75E-27	SpD& SpR, SpD, Syn	H, H, L	0.1-0.7	0.2-0.6	0.2-0.3	0.0-0.0
11	319048	T000001344.1	snp, snp	Sugar transporter SWEET	1,61E-93	Mis, Syn	M, L	1.0-0.5	0.8-0.3	0.8-0.6	1.0-1.0
20	52360	T000003200.1	snp	FlaMIngo (cadherin plus 7TM domain) homolog	0,00E+00	Mis, Mis	M, M	0.9-0.1	0.6-0.1	0.6-0.5	0.0-0.1
20	211654	T000003243.2	snp	Bm10974, isoform a	3,09E-15	Mis, Mis	M, M	0.9-0.4	0.5-0.0	0.5-0.4	0.0-0.0
20	418416	T000003304.1	snp	NA	NA	Mis	M	0.9-0.4	0.7-0.2	0.7-0.8	0.0-0.0
20	703781	T000003392.1	snp	Nuclear hormone receptor family member nhr-48	1,28E-58	Mis	M	1.0-0.1	0.8-0.4	0.8-0.7	0.0-0.0
36	113815	T000004395.1	snp	Serine Rich Adhesion Protein-like	5,70E-64	Mis	M	1.0-0.5	1.0-0.6	1.0-0.9	1.0-1.0
40	563913	T000004972.1	snp	Gamma-glutamyltranspeptidase family-containing protein	1,99E-57	Mis	M	1.0-0.2	0.9-0.4	0.9-0.8	1.0-1.0
16	397745	T000005044.1	snp	asparagine synthetase domain-containing	1,24E-62	Mis	M	1.0-0.3	0.6-0.1	0.6-0.5	1.0-1.0

protein 1											
60	87902	T000005507.1	snp	Transthyretin-like family protein	1,99E-44	Mis	M	0.3-0.7	0.2-0.8	0.2-0.4	0.0-0.0
60	297264	T000005564.1	snp	cytosolic endo-beta-N-acetylglucosaminidase	3,74E-78	Mis	M	0.1-0.6	0.6-1.0	0.6-0.4	0.0-0.0
9	444744	T000005856.1	snp, snp	NA	NA	Mis, Mis	M, M	0.8-0.4	0.8-0.3	0.8-0.7	1.0-1.0
59	32933	T000006195.1	snp	Receptor-type guanylate cyclase <i>gcy-9</i>	0,00E+00	Mis	M	0.3-0.9	0.5-1.0	0.5-0.5	0.0-0.0
92	66725	T000009753.1	complex, snp	endonuclease/exonuclease/phosphatase family protein	1,99E-66	Mis, Mis	M, M	0.0-0.7	0.2-0.7	0.2-0.3	0.0-0.0
92	91906	T000009762.1	snp	small subunit processome component 20 homolog	0,00E+00	Mis	M	1.0-0.6	0.8-0.0	0.8-0.8	1.0-1.0
92	146469	T000009778.1	snp	JmjC domain protein	9,99E-80	Mis	M	0.9-0.5	0.8-0.2	0.8-0.9	1.0-1.0
136	63412	T000009935.1	snp	anNEXin family	7,36E-102	Mis	M	0.1-0.8	0.2-0.7	0.2-0.2	1.0-1.0
136	116834	T000009955.1	snp	immunoglobulin I-set domain-containing protein	7,18E-54	Mis	M	0.9-0.5	0.8-0.1	0.8-0.7	0.0-0.0
70	388153	T000010247.1	snp	alcohol dehydrogenase 1-like	5,17E-43	Mis	M	0.7-0.3	0.8-0.3	0.8-0.6	1.0-1.0
71	52659	T000010267.1	complex, snp	AMP deaminase	0,00E+00	Mis, Mis	M, M	1.0-0.6	1.0-0.6	1.0-0.8	0.3-0.3
86	41241	T000010523.1	complex, snp	Methylthioribulose-1-phosphate dehydratase	6,45E-62	Mis, Mis	M, M	0.1-0.7	0.2-0.6	0.2-0.0	0.0-0.0
28	368031	T000010818.1	snp	dorsal gland cell-specific expression protein	1,54E-14	Mis	M	0.0-0.8	0.4-1.0	0.4-0.6	0.0-0.0
28	368040	T000010818.1	complex	dorsal gland cell-specific expression protein	1,54E-14	Mis	M	0.0-0.7	0.5-0.9	0.5-0.6	0.0-0.0
49	298495	T000011044.1	snp	poly(A) polymerase type 3	1,16E-25	Mis	M	0.3-0.8	0.5-1.0	0.5-0.5	0.0-0.0
82	112747	T000013942.1	complex, mnp, snp	NA	NA	Mis, Mis	M, M	0.8-0.2	0.4-0.0	0.4-0.3	0.0-0.0
102	94546	T000014032.1	snp	NA	NA	Mis	M	0.0-0.4	0.2-0.6	0.2-0.1	0.0-0.0
102	259636	T000014073.1	del, del, snp	putative DNA polymerase	4,29E-93	Cidel, Dlins, Mis	M, M, M	0.7-0.3	0.8-0.3	0.8-0.8	0.2-0.3

134	228557	T000014176.1	ins, ins, snp	Sodium/hydrogen exchanger	0,00E+00	Dlins, Dlins, Mis	M, M, M	0.0-0.8	0.3-0.8	0.3-0.5	0.0-0.0
67	266816	T000014324.1	snp, snp, snp	RNA-directed DNA polymerase (reverse transcriptase) and Integrase domain containing protein	5,44E-79	Mis, Mis, Mis	M, M, M	0.9-0.1	0.5-0.1	0.5-0.4	0.1-0.2
80	149975	T000015207.2	snp	cell growth-regulating nucleolar protein-like	3,49E-33	Mis	M	0.8-0.4	0.8-0.4	0.8-0.9	1.0-1.0
116	265665	T000015913.1	snp	MH1 domain protein	8,82E-95	Mis	M	0.9-0.4	0.6-0.2	0.6-0.5	0.8-0.8
182	176886	T000016657.1	snp	GDP-mannose 4,6 dehydratase 1	0,00E+00	Mis	M	0.2-0.8	0.2-0.6	0.2-0.2	0.6-0.6
153	150639	T000017275.1	snp	arginine kinase	6,41E-82	Mis	M	0.3-0.8	0.4-1.0	0.4-0.6	0.3-0.4
381	81359	T000018119.1	complex, complex, snp	integrase core domain protein	2,02E-15	Mis, Mis, Syn	M, M, L	0.2-0.8	0.3-0.8	0.3-0.4	0.3-0.1
151	141595	T000019082.1	complex, snp	preprotein translocase subunit SecA	1,10E-90	Mis, Mis	M, M	0.1-0.5	0.2-0.7	0.2-0.1	0.0-0.0
151	141615	T000019082.1	snp	preprotein translocase subunit SecA	1,10E-90	Mis	M	0.1-0.6	0.2-0.6	0.2-0.2	0.0-0.0
343	37814	T000022033.1	ins	putative uncharacterized transposon-derived protein F54H12.3	2,81E-60	Ciins	M	0.0-0.6	0.1-0.6	0.1-0.1	0.1-0.1
284	38454	T000022305.1	snp	dnaJ homolog subfamily C member 10-like	0,00E+00	Mis	M	0.0-0.5	0.1-0.5	0.1-0.0	0.0-0.0
217	96836	T000022678.1	snp	FLYWCH zinc finger domain protein	3,66E-17	Mis	M	0.0-0.7	0.4-0.9	0.4-0.5	0.6-0.7
466	1643	T000022928.1	snp	NA	NA	Mis	M	1.0-0.6	0.7-0.3	0.7-0.6	1.0-1.0
169	86676	T000023163.1	complex, complex	LPXTG cell wall anchor domain-containing protein	7,08E-05	Mis, Mis	M, M	0.8-0.4	0.8-0.2	0.8-0.7	1.0-1.0
270	96344	T000024053.2	snp	Innexin family-containing protein	2,23E-154	Mis	M	0.1-0.6	0.3-0.7	0.3-0.3	0.0-0.0
233	106336	T000024132.1	snp	peptidase family M1 containing protein	1,11E-132	Mis	M	0.9-0.1	0.9-0.5	0.9-0.8	1.0-1.0
233	119334	T000024134.1	complex, complex, snp	NA	NA	Mis, Mis	M, M	0.8-0.4	0.9-0.3	0.9-0.8	1.0-1.0
277	36337	T000024468.1	complex, snp	dorsal gland cell-specific expression protein	9,58E-42	Mis, Mis	M	0.5-0.9	0.4-1.0	0.4-0.3	0.0-0.0

277	36354	T000024468.1	complex, snp, snp	dorsal gland cell-specific expression protein	9,58E-42	Mis, Mis, Mis	M, M, M	0.5-1.0	0.4-0.8	0.4-0.3	0.0-0.0
277	54464	T000024472.1	complex, complex, snp	dorsal gland cell-specific expression protein	1,49E-28	Mis, Mis, Syn	M, M, L	1.0-0.5	0.9-0.4	0.9-0.8	1.0-1.0
277	54503	T000024472.1	snp	dorsal gland cell-specific expression protein	1,49E-28	Mis	M	1.0-0.5	1.0-0.4	1.0-0.9	1.0-1.0
324	23787	T000024682.1	complex, snp	pentapeptide repeat-containing protein	7,75E-16	Mis, Mis	M, M	0.4-0.9	0.3-0.7	0.3-0.4	0.0-0.0
260	6741	T000025554.1	snp	pectate lyase 1	6,40E-78	Mis	M	0.9-0.0	0.7-0.1	0.7-0.8	0.0-0.2
609	34561	T000027398.1	complex, snp, snp	predicted protein	2,17E-49	Mis, Mis, Mis	M, M, M	1.0-0.6	0.8-0.3	0.8-0.9	0.4-0.4
237	145871	T000028814.1	snp	NA	NA	Mis	M	0.4-0.8	0.5-1.0	0.5-0.6	0.0-0.2
550	12529	T000029200.1	snp	NA	NA	Mis	M	0.2-0.7	0.3-0.8	0.3-0.3	0.3-0.3
4	183618	T000000056.1	snp	Splicing factor 3B subunit 2	0,00E+00	Syn	L	0.2-0.7	0.4-1.0	0.4-0.6	1.0-1.0
4	185559	T000000056.1	snp	Splicing factor 3B subunit 2	0,00E+00	Syn	L	0.2-0.9	0.5-0.9	0.5-0.6	1.0-1.0
4	638116	T000000213.1	snp	Discontinuous actin hexagon	1,23E-82	Syn	L	0.7-0.2	0.6-0.1	0.6-0.6	0.0-0.0
4	962833	T000000317.1	snp	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit epsilon	0,00E+00	Syn	L	0.2-0.8	0.5-1.0	0.5-0.5	1.0-1.0
8	41262	T000001542.1	snp	SERF-like protein [1,66E-11	Syn	L	1.0-0.4	1.0-0.5	1.0-0.8	1.0-1.0
2	620741	T000001995.1	complex, snp	Globin-like domain and Globin, structural domain-containing protein	7,65E-55	Syn, Syn	L, L	0.8-0.1	0.5-0.0	0.5-0.4	0.0-0.0
21	163201	T000002196.2	snp	Transient receptor potential channel	9,70E-156	SpR & Syn	L	0.8-0.3	0.6-0.1	0.6-0.6	0.4-0.5
20	384616	T000003291.1	snp	HMG box	1,67E-28	Syn	L	0.9-0.3	0.7-0.3	0.7-0.7	0.0-0.0
7	701231	T000003603.1	complex, complex, del	Ets-domain protein	1,54E-40	SpR & Int, Int, Int	L, Mod, Mod	0.4-0.9	0.6-1.0	0.6-0.5	0.0-0.0
33	402464	T000003819.1	snp, snp	Homolog of Odr-2 (Two)	3,97E-30	Syn,	L, L	1.0-0.5	0.9-0.4	0.9-0.8	1.0-1.0

											Syn
29	471281	T000004044.1	snp	Transmembrane matrix receptor MUP-4	0,00E+00	Syn	L	0.0-0.5	0.2-0.6	0.2-0.1	0.0-0.0
29	647225	T000004101.1	snp	GTP-ase activating protein for Arf containing protein	1,89E-112	Syn	L	0.3-0.9	0.4-0.8	0.4-0.4	0.0-0.0
40	138909	T000004846.1	ins	cyclin domain protein	1,41E-90	SpR & Int	L	0.8-0.2	0.9-0.4	0.9-0.8	1.0-1.0
50	225859	T000005403.1	snp	Mitochondrial thiamine pyrophosphate carrier	1,36E-99	Syn	L	0.5-1.0	0.3-0.8	0.3-0.5	1.0-1.0
50	495184	T000005474.2	ins	Integrator complex subunit 3	3,13E-104	SpR & Int	L	0.0-0.5	0.1-0.7	0.1-0.2	0.1-0.0
60	15681	T000005484.1	snp	XAP5 protein	1,38E-130	Syn	L	0.1-0.6	0.2-0.7	0.2-0.1	0.0-0.0
59	390120	T000006296.1	complex, ins	NIPA1 (NonImprinted gene in Prader-Willi/Angelman syndrome region 1) homolog	1,89E-179	SpR & Int, SpR & Int	L, L	0.2-0.6	0.5-1.0	0.5-0.6	0.0-0.0
92	66760	T000009753.1	snp	endonuclease/exonuclease/phosphatase family protein	1,99E-66	Syn	L	0.0-0.7	0.2-0.6	0.2-0.3	0.0-0.0
92	90288	T000009762.1	snp	small subunit processome component 20 homolog	0,00E+00	Syn	L	0.9-0.1	0.5-0.0	0.5-0.4	0.0-0.0
28	207093	T000010790.1	snp	NA	NA	SpR & Int	L	0.2-1.0	0.1-0.6	0.1-0.0	0.0-0.1
81	270571	T000011625.1	complex	histidine acid phosphatase	1,83E-132	SpR & Int	L	0.8-0.2	0.5-0.1	0.5-0.6	1.0-1.0
113	276349	T000012382.1	snp	ankyrin repeat domain-containing protein 28	9,27E-71	Syn	L	0.9-0.4	0.7-0.3	0.7-0.7	1.0-1.0
137	32682	T000012558.1	snp	membralin isoform X1	7,79E-135	Syn	L	0.3-0.7	0.2-0.6	0.2-0.3	1.0-1.0
85	217352	T000013870.1	snp	Methionine synthase reductase	1,87E-62	SpR & Int	L	0.9-0.5	0.5-0.0	0.5-0.6	1.0-1.0
102	259494	T000014073.1	snp	putative DNA polymerase	4,29E-93	Syn	L	1.0-0.5	0.8-0.4	0.8-0.9	0.3-0.4
171	182455	T000014555.2	complex, snp	Spectrin beta chain	0,00E+00	SpR & Int,	L, Mod,	0.2-0.7	0.3-0.8	0.3-0.4	1.0-1.0

						Int	Mod				
184	74946	T000016476.1	snp	integrator complex subunit 4	3,45E-105	SpR & Int	L	0.9-0.4	0.9-0.4	0.9-0.8	0.0-0.0
131	31099	T000017827.1	snp	UPF0428 protein CXorf56 -like protein	2,02E-91	Syn	L	1.0-0.1	0.7-0.2	0.7-0.7	1.0-1.0
90	230819	T000018011.1	snp	Spermatogenesis-associated serine-rich protein 2	1,57E-24	Syn	L	0.8-0.3	0.6-0.2	0.6-0.5	0.0-0.0
168	176041	T000018908.1	snp	Amino acid transporter, transmembrane family-containing protein	1,13E-161	Syn	L	0.8-0.1	0.6-0.2	0.6-0.5	0.7-0.6
142	37416	T000019578.1	snp	NA	NA	Syn	L	0.3-0.9	0.4-0.9	0.4-0.4	1.0-1.0
143	83375	T000020343.1	complex, ins, snp	CBR-NURF-1 protein, variant	2,61E-97	SpR & Int, Int, Int	L, Mod, Mod	0.7-0.2	0.5-0.0	0.5-0.4	0.0-0.0
143	124925	T000020351.1	snp, snp	Protein CBG00614	8,85E-37	Syn, Syn	L, L	0.7-0.2	0.9-0.4	0.9-0.9	0.2-0.2
271	83878	T000020960.1	snp	FAD dependent oxidoreductase	0,00E+00	Syn	L	0.0-0.7	0.2-0.7	0.2-0.2	0.0-0.1
239	62626	T000021180.1	snp	DNA topoisomerase 3-beta-1	0,00E+00	Syn	L	0.0-0.8	0.1-0.6	0.1-0.2	1.0-1.0
186	137143	T000021672.1	snp	dorsal gland cell-specific expression protein	8,55E-20	Syn	L	1.0-0.3	0.6-0.2	0.6-0.7	0.5-0.5
413	22257	T000021953.1	snp, snp	cupin family protein	2,13E-109	Syn, Syn	L, L	0.3-0.7	0.3-0.7	0.3-0.5	1.0-0.9
294	64177	T000022428.1	snp	NCA Localization Factor	7,35E-65	Syn	L	0.3-0.7	0.2-0.6	0.2-0.3	0.0-0.0
169	86658	T000023163.1	del	LPXTG cell wall anchor domain-containing protein	7,08E-05	SpR & Int	L	0.9-0.4	0.8-0.2	0.8-0.7	1.0-1.0
216	123735	T000023674.2	snp	ATP-dependent DNA helicase PIF1	1,15E-130	Syn	L	0.1-0.7	0.2-0.6	0.2-0.2	0.4-0.4
270	96349	T000024053.2	snp	Innexin family-containing protein	2,23E-154	Syn	L	0.1-0.6	0.3-0.7	0.3-0.3	0.0-0.0
277	36324	T000024468.1	complex, snp, snp	dorsal gland cell-specific expression protein	9,58E-42	Syn, Syn, Syn	L, L, L	0.4-0.9	0.3-1.0	0.3-0.4	0.0-0.0
246	81556	T000024555.1	del	ubiquitin-40S ribosomal protein S27a	3,75E-27	SpR & Int	L	1.0-0.2	0.8-0.4	0.8-0.9	1.0-1.0

461	62957	T000027710.1	complex, snp	putative DNA polymerase	7,93E-119	Syn, Syn	L, L	0.8-0.2	0.6-0.0	0.6-0.4	0.1-0.2
587	2690	T000029364.1	snp	Hypothetical 27.8 kDa protein T25D10.4 in chromosome II, putative	3,13E-05	SpR & Int	L	1.0-0.5	0.8-0.3	0.8-0.9	1.0-1.0
719	4398	T000029390.1	snp	Ubiquitin thioesterase trabid	1,24E-17	Syn	L	0.7-0.2	0.9-0.4	0.9-1.0	1.0-1.0
4	3245	T000000001.1	snp	vacuolar protein sorting-associated protein 54	1,50E-175	Int	Mod	0.2-0.7	0.3-0.8	0.3-0.4	1.0-1.0
4	533227	T000000177.1	snp	NeuroPeptide Receptor family	3,01E-151	Int	Mod	0.4-1.0	0.6-1.0	0.6-0.5	1.0-1.0
4	1008067	T000000331.1	snp	DNA ligase 1	0,00E+00	Int	Mod	0.9-0.3	0.6-0.1	0.6-0.4	0.0-0.0
3	1353331	T000003178.1	snp	lysine-specific demethylase 5C-like isoform X1	1,93E-103	Int	Mod	0.1-0.5	0.6-1.0	0.6-0.7	0.0-0.0
20	49800	T000003200.1	snp	FlaMIngo (cadherin plus 7TM domain) homolog	0,00E+00	Int	Mod	1.0-0.6	0.8-0.3	0.8-0.7	1.0-1.0
20	190379	T000003240.1	complex, complex, snp	calpain D	7,13E-169	Int, Int, Int	Mod, Mod, Mod	1.0-0.4	0.4-0.0	0.4-0.3	0.0-0.0
20	266083	T000003261.1	snp	UBX domain-containing protein 7	1,62E-77	Int	Mod	0.1-0.6	0.4-0.8	0.4-0.4	1.0-1.0
20	303658	T000003270.1	snp	mediator complex subunit MED14	1,67E-11	Int	Mod	1.0-0.0	0.5-0.0	0.5-0.6	0.0-0.0
20	303691	T000003270.1	ins	mediator complex subunit MED14	1,67E-11	Int	Mod	1.0-0.0	0.5-0.0	0.5-0.5	0.0-0.0
20	304232	T000003270.1	snp	mediator complex subunit MED14	1,67E-11	Int	Mod	1.0-0.5	0.9-0.4	0.9-0.9	1.0-1.0
20	308798	T000003271.1	snp	Ubiquitin Fusion Degradation (yeast UFD homolog)	1,61E-62	Int	Mod	1.0-0.4	0.4-0.0	0.4-0.4	0.0-0.0
20	399099	T000003293.1	ins, snp, snp	nematode cuticle collagen domain protein	5,44E-04	Int, Int, Int	Mod, Mod, Mod	1.0-0.5	0.7-0.2	0.7-0.7	0.0-0.0
20	531183	T000003332.1	snp	7 transmembrane receptor	2,12E-66	Int	Mod	1.0-0.0	0.7-0.2	0.7-0.6	0.0-0.0
20	536194	T000003333.1	complex, del	Nuclear Hormone Receptor family	2,02E-08	Int, Int	Mod, Mod	0.9-0.4	0.6-0.1	0.6-0.8	0.0-0.0
20	736517	T000003399.1	snp	REX1 DNA Repair	7,00E-79	Int	Mod	1.0-0.1	0.8-0.3	0.8-0.6	0.0-0.0
33	402385	T000003819.1	snp	Homolog of Odr-2 (Two)	3,97E-30	Int	Mod	1.0-0.5	0.9-0.4	0.9-0.8	1.0-1.0
6	438094	T000004221.1	snp	hnRNP-R, Q splicing factor family	1,20E-166	Int	Mod	1.0-0.6	0.9-0.5	0.9-1.0	1.0-1.0

36	88312	T000004389.1	snp	Zinc finger protein	0,00E+00	Int	Mod	1.0-0.3	0.9-0.4	0.9-0.7	0.1-0.0
36	153952	T000004406.1	snp	protein Lilipod	1,24E-127	Int	Mod	1.0-0.6	1.0-0.4	1.0-0.8	1.0-1.0
36	525188	T000004507.1	del	NA	NA	Int	Mod	1.0-0.5	0.9-0.4	0.9-0.9	1.0-1.0
10	678679	T000004727.1	snp	differentially expressed in FDCP 8 homolog isoform X1	2,95E-46	Int	Mod	0.0-0.8	0.4-0.8	0.4-0.5	0.8-0.7
16	162759	T000005000.1	snp	NA	NA	Int	Mod	0.9-0.5	0.8-0.4	0.8-0.9	1.0-1.0
16	162792	T000005000.1	snp	NA	NA	Int	Mod	0.9-0.5	0.8-0.4	0.8-0.9	1.0-1.0
16	248607	T000005017.1	ins	putative phospholipase B-like 2	7,14E-86	Int	Mod	1.0-0.4	0.7-0.3	0.7-0.6	1.0-1.0
16	610496	T000005096.1	snp	43 kDa receptor-associated protein of the synapse	2,91E-175	Int	Mod	1.0-0.4	0.9-0.4	0.9-0.8	0.8-1.0
50	220293	T000005399.1	complex, snp	AcetylCholine Receptor	5,33E-169	Int, Int	Mod, Mod	0.4-1.0	0.5-1.0	0.5-0.6	1.0-1.0
59	226869	T000006258.1	ins	pecanex protein	0,00E+00	Int	Mod	0.2-0.7	0.6-1.0	0.6-0.4	0.0-0.0
68	90070	T000006835.1	complex, ins	NA	NA	Int, Int	Mod, Mod	1.0-0.4	0.6-0.2	0.6-0.5	1.0-1.0
68	191502	T000006869.1	del	Pep3/Vps18/deep orange family protein	0,00E+00	Int	Mod	0.0-0.5	0.5-0.9	0.5-0.6	0.0-0.0
17	341733	T000007121.1	complex, mnp	Astacin-like metalloendopeptidase	4,51E-39	Int, Int	Mod, Mod	0.7-0.1	0.8-0.4	0.8-0.7	0.6-0.6
77	129026	T000008086.1	complex, snp	IntraMembrane Protease (IMPAS) family	1,93E-77	Int, Int	Mod, Mod	0.1-0.6	0.5-0.9	0.5-0.3	0.0-0.0
77	233109	T000008116.1	snp	Vacuolar protein sorting-associated protein 41 -like protein	0,00E+00	Int	Mod	1.0-0.5	0.7-0.0	0.7-0.8	1.0-1.0
77	242552	T000008118.1	snp	Protein C39E9.8, isoform d	3,70E-67	Int	Mod	0.0-0.5	0.3-0.8	0.3-0.2	0.2-0.3
46	390950	T000008821.1	snp	Myosin heavy chain	0,00E+00	Int	Mod	0.9-0.2	0.8-0.4	0.8-0.7	0.5-0.5
24	348694	T000008952.1	snp	reverse transcriptase	2,43E-122	Int	Mod	0.8-0.2	0.7-0.3	0.7-0.8	1.0-1.0
37	305249	T000009585.1	ins	Putative tyrosinase-like protein tyr-3	0,00E+00	Int	Mod	0.8-0.3	0.8-0.3	0.8-0.7	1.0-1.0
92	60880	T000009751.1	complex	NA	NA	Int	Mod	0.0-0.4	0.3-0.8	0.3-0.3	0.0-0.0
92	59120	T000009751.1	complex, snp	NA	NA	Int, Int	Mod, Mod	1.0-0.5	0.9-0.5	0.9-0.8	1.0-1.0
136	57883	T000009933.1	snp	CRiM (Cysteine Rich motor neuron protein) homolog	4,43E-97	Int	Mod	0.8-0.4	0.7-0.3	0.7-0.6	0.2-0.0

136	151314	T000009963.1	complex, ins	Hexokinase family and Hexokinase, N-terminal domain and Hexokinase, C-terminal domain-containing protein	3,97E-174	Int, Int	Mod, Mod	0.9-0.4	0.8-0.2	0.8-0.7	0.1-0.0
71	345659	T000010359.2	complex, ins	Potassium voltage-gated channel protein shk-1	7,49E-142	Int, Int	Mod, Mod	1.0-0.4	0.9-0.5	0.9-1.0	1.0-1.0
38	234323	T000010426.1	snp	ADenOsine Receptor homolog	1,50E-96	Int	Mod	0.0-0.7	0.3-0.9	0.3-0.3	0.0-0.0
38	234339	T000010426.1	complex, del	ADenOsine Receptor homolog	1,50E-96	Int, Int	Mod, Mod	0.0-0.6	0.3-0.9	0.3-0.4	0.0-0.0
86	132735	T000010553.1	snp	NA	NA	Int	Mod	0.0-0.8	0.5-1.0	0.5-0.4	1.0-1.0
83	157789	T000011797.1	complex, snp	Transmembrane cell adhesion receptor mua-3	0,00E+00	Int, Int	Mod, Mod	0.3-0.7	0.5-1.0	0.5-0.5	0.1-0.0
109	194070	T000012136.2	complex	Guanylate cyclase	0,00E+00	Int	Mod	0.5-0.9	0.6-1.0	0.6-0.5	1.0-1.0
135	45773	T000012246.1	complex, mnp	Pyruvate dehydrogenase phosphatase regulatory subunit, mitochondrial	1,20E-18	Int, Int	Mod, Mod	0.9-0.4	0.6-0.2	0.6-0.8	1.0-1.0
113	241021	T000012372.1	complex, complex	Putative diacylglycerol kinase K06A1.6	8,68E-158	Int, Int	Mod	0.9-0.4	0.8-0.3	0.8-0.8	1.0-1.0
43	57727	T000013599.1	del	hypothetical protein ANCDUO_02652	5,09E-14	Int	Mod	0.0-0.5	0.1-0.7	0.1-0.0	0.0-0.0
43	57755	T000013599.1	snp	hypothetical protein ANCDUO_02652	5,09E-14	Int	Mod	0.3-0.7	0.3-0.7	0.3-0.4	1.0-1.0
82	89881	T000013935.1	ins	Thioredoxin domain-containing protein 15	2,93E-12	Int	Mod	1.0-0.5	0.7-0.2	0.7-0.7	0.8-1.0
82	97433	T000013936.1	snp	Suppressor of Stomatin mutant Uncoordination	1,10E-107	Int	Mod	1.0-0.6	0.8-0.1	0.8-0.9	1.0-1.0
82	97441	T000013936.1	snp	Suppressor of Stomatin mutant Uncoordination	1,10E-107	Int	Mod	1.0-0.5	0.8-0.1	0.8-0.9	1.0-1.0
82	97448	T000013936.1	del	Suppressor of Stomatin mutant Uncoordination	1,10E-107	Int	Mod	1.0-0.6	0.8-0.2	0.8-0.9	1.0-1.0
82	126360	T000013945.1	ins	NA	NA	Int	Mod	0.0-0.5	0.5-0.9	0.5-0.6	0.0-0.0
80	125568	T000015205.1	snp	DNA-binding protein D-ETS-3 isoform X3	1,53E-61	Int	Mod	0.2-0.7	0.2-0.9	0.2-0.2	0.0-0.0
44	299303	T000015543.1	snp	Sulfhydryl oxidase	9,32E-131	Int	Mod	1.0-0.5	0.4-0.0	0.4-0.5	1.0-1.0
116	269047	T000015913.1	snp	MH1 domain protein	8,82E-95	Int	Mod	1.0-0.6	0.9-0.4	0.9-0.8	1.0-1.0
48	76653	T000016091.1	snp	ankyrin repeat protein	1,51E-19	Int	Mod	0.4-0.9	0.5-1.0	0.5-0.6	1.0-1.0
184	110128	T000016487.1	complex, complex,	E3 ubiquitin-protein ligase hecd-1	0,00E+00	Int, Int,	Mod, Mod,	0.0-0.5	0.3-0.7	0.3-0.2	0.0-0.0

			snp			Int	Mod				
182	142367	T000016649.1	snp	Rap Guanine nucleotide Exchange Factor homolog	1,97E-167	Int	Mod	0.3-0.9	0.3-0.8	0.3-0.4	0.7-0.6
182	153233	T000016651.1	snp	Zinc finger CCCH domain-containing protein 14	2,22E-19	Int	Mod	0.2-0.9	0.3-0.8	0.3-0.2	1.0-0.9
182	153353	T000016651.1	del, snp	Zinc finger CCCH domain-containing protein 14	1,74E-19	Int, Int	Mod, Mod	0.2-0.8	0.2-0.6	0.2-0.3	0.8-0.8
255	90346	T000017164.1	complex, ins, snp	transmembrane protein 50A	1,02E-58	Int, Int, Int	Mod, Mod, Mod	1.0-0.5	0.9-0.4	0.9-0.9	1.0-1.0
153	47891	T000017239.1	del	Protein CBG01955	1,11E-46	Int	Mod	0.3-0.8	0.4-0.9	0.4-0.5	1.0-1.0
153	164714	T000017277.1	snp	Mediator of RNA polymerase II transcription subunit 13	5,69E-113	Int	Mod	0.2-0.7	0.2-0.8	0.2-0.3	0.1-0.2
153	164669	T000017277.1	snp	Mediator of RNA polymerase II transcription subunit 13	5,69E-113	Int	Mod	0.8-0.4	0.8-0.3	0.8-0.8	0.9-0.9
103	123333	T000017525.1	snp	NA	NA	Int	Mod	1.0-0.5	0.5-0.0	0.5-0.3	1.0-1.0
104	184633	T000017604.1	snp	NA	NA	Int	Mod	0.4-0.9	0.6-1.0	0.6-0.4	1.0-1.0
131	50938	T000017830.1	complex, del, snp	NA	NA	Int, Int, Int	Mod, Mod, Mod	0.1-0.7	0.4-0.8	0.4-0.5	0.0-0.0
110	272344	T000019339.1	ins, snp	homeobox domain-containing protein	8,38E-111	Int, Int	Mod, Mod	0.4-1.0	0.6-1.0	0.6-0.4	1.0-1.0
279	8146	T000019529.1	snp	WW domain-containing protein	2,11E-179	Int	Mod	0.4-0.9	0.2-0.6	0.2-0.4	1.0-1.0
142	56606	T000019588.1	ins, snp	transcription factor collier	3,97E-180	Int, Int	Mod, Mod	0.4-0.8	0.4-0.8	0.4-0.5	1.0-1.0
174	20666	T000020024.1	del, del	GSK3-beta interaction protein	1,57E-65	Int, Int	Mod	0.1-0.5	0.4-0.9	0.4-0.3	0.0-0.0
174	38304	T000020027.1	complex, ins	MBOAT family protein	9,73E-124	Int, Int	Mod, Mod	0.0-0.6	0.1-0.8	0.1-0.3	0.0-0.0
174	96334	T000020045.1	snp	CRE-EFN-2 protein	3,37E-45	Int	Mod	1.0-0.6	0.9-0.2	0.9-0.9	1.0-0.9
139	160130	T000020175.1	complex	NA	NA	Int	Mod	0.9-0.3	0.9-0.2	0.9-0.7	1.0-1.0
139	160151	T000020175.1	snp	NA	NA	Int	Mod	0.9-0.4	0.9-0.3	0.9-0.7	1.0-1.0

139	160156	T000020175.1	snp	NA	NA	Int	Mod	0.9-0.4	0.9-0.3	0.9-0.8	1.0-1.0
139	160163	T000020175.1	snp	NA	NA	Int	Mod	0.9-0.4	0.9-0.3	0.9-0.8	1.0-1.0
139	160171	T000020175.1	snp	NA	NA	Int	Mod	1.0-0.4	0.8-0.3	0.8-0.8	1.0-1.0
139	160184	T000020175.1	snp	NA	NA	Int	Mod	1.0-0.4	0.8-0.3	0.8-0.8	1.0-1.0
139	160191	T000020175.1	snp	NA	NA	Int	Mod	1.0-0.4	0.8-0.3	0.8-0.8	1.0-1.0
139	160205	T000020175.1	complex	NA	NA	Int	Mod	1.0-0.4	0.8-0.3	0.8-0.8	1.0-1.0
139	160219	T000020175.1	snp	NA	NA	Int	Mod	0.9-0.4	0.8-0.2	0.8-0.8	1.0-1.0
139	160230	T000020175.1	snp	NA	NA	Int	Mod	0.9-0.4	0.8-0.3	0.8-0.7	1.0-1.0
143	125256	T000020351.1	complex, del	Protein CBG00614	8,85E-37	Int, Int	Mod, Mod	0.7-0.0	0.8-0.4	0.8-0.8	0.0-0.2
313	27136	T000020978.1	snp	regulator of MON1-CCZ1 complex	3,77E-77	Int	Mod	0.0-0.4	0.1-0.5	0.1-0.3	0.0-0.0
200	16182	T000021092.2	complex	ImpB/MucB/SamB family protein	3,51E-74	Int	Mod	0.1-0.6	0.2-0.6	0.2-0.4	0.0-0.0
130	106165	T000021432.1	snp	NA	NA	Int	Mod	0.4-0.9	0.2-0.8	0.2-0.0	1.0-1.0
145	162865	T000022001.1	snp	NA	NA	Int	Mod	1.0-0.6	0.8-0.3	0.8-0.7	0.8-1.0
291	85693	T000022500.1	snp	NA	NA	Int	Mod	0.7-0.1	0.5-0.0	0.5-0.6	0.0-0.0
291	85721	T000022500.1	snp	NA	NA	Int	Mod	0.7-0.1	0.5-0.0	0.5-0.6	0.0-0.0
206	57783	T000023772.1	complex, complex	NADH pyrophosphatase	3,36E-90	Int, Int	Mod	0.0-0.5	0.0-0.5	0.0-0.0	0.0-0.0
206	57793	T000023772.1	snp	NADH pyrophosphatase	3,36E-90	Int	Mod	0.0-0.5	0.0-0.5	0.0-0.0	0.0-0.0
206	84781	T000023777.1	del	Sodium-dependent high-affinity dicarboxylate transporter 3	1,04E-170	Int	Mod	0.0-0.5	0.1-0.6	0.1-0.0	0.0-0.0
206	85037	T000023777.1	snp	Sodium-dependent high-affinity dicarboxylate transporter 3	1,04E-170	Int	Mod	0.0-0.6	0.1-0.8	0.1-0.2	0.0-0.0
206	85088	T000023777.1	ins	Sodium-dependent high-affinity dicarboxylate transporter 3	1,04E-170	Int	Mod	0.0-0.7	0.1-0.6	0.1-0.3	0.0-0.0
270	96152	T000024053.2	complex, snp	Innexin family-containing protein	2,23E-154	Int, Int	Mod, Mod	0.1-0.6	0.3-0.8	0.3-0.2	0.0-0.0
355	74932	T000024083.1	snp	Neuropeptide-Like Protein	9,28E-23	Int	Mod	0.0-0.7	0.1-0.9	0.1-0.0	1.0-1.0
233	119184	T000024134.1	snp	NA	NA	Int	Mod	1.0-0.4	0.9-0.3	0.9-0.8	1.0-1.0
286	8656	T000024383.1	del	NA	NA	Int	Mod	0.4-0.9	0.4-0.9	0.4-0.5	0.0-0.0
286	58404	T000024392.1	complex, del	cytidyltransferase domain protein	1,03E-122	Int, Int	Mod, Mod	0.1-0.9	0.6-1.0	0.6-0.4	0.0-0.0
471	34063	T000024414.1	complex,	replication factor C subunit 3	1,39E-	Int,	Mod,	0.3-1.0	0.3-0.9	0.3-0.3	1.0-1.0

			complex, snp		134	Int, Int	Mod, Mod				
222	152194	T000025147.1	snp	Transposable element Tc3 transposase	5,82E-05	Int	Mod	0.5-0.0	0.7-0.0	0.7-0.8	1.0-1.0
481	29268	T000025326.1	snp	retinoic acid receptor gamma	1,09E-32	Int	Mod	0.2-0.9	0.3-0.8	0.3-0.3	1.0-1.0
481	43145	T000025332.1	complex, del	Intron-binding protein aquarius	0,00E+00	Int, Int	Mod, Mod	0.5-0.9	0.2-0.6	0.2-0.2	1.0-0.9
344	52719	T000026502.1	complex, snp	NA	NA	Int, Int	Mod	1.0-0.5	0.9-0.3	0.9-1.0	1.0-1.0
374	14798	T000027792.2	snp	Polypeptide N-acetylgalactosaminyltransferase 5	0,00E+00	Int	Mod	0.4-0.9	0.6-1.0	0.6-0.6	0.0-0.0
575	19527	T000028173.1	snp	putative uncharacterized transposon-derived protein F54H12.3	2,16E-20	Int	Mod	0.0-0.6	0.1-0.7	0.1-0.0	0.1-0.0
662	24851	T000029255.1	complex, snp	NA	NA	Int, Int	Mod, Mod	0.0-0.4	0.4-0.8	0.4-0.4	0.0-0.0
674	12247	T000029459.1	snp	HNTX-XI-10 precursor	5,69E-10	Int	Mod	0.0-0.6	0.2-0.7	0.2-0.2	0.0-0.0
105	57290		ins, snp, snp	NA	NA	Inter	NA	0.6-1.0	0.2-0.9	0.2-0.3	0.0-0.2
115	131539		complex, snp, snp	NA	NA	Inter	NA	0.6-0.2	0.5-0.0	0.5-0.6	0.0-0.1
115	131557		complex, mnp, snp	NA	NA	Inter	NA	0.7-0.2	0.5-0.0	0.5-0.6	0.0-0.1
116	283596		snp	NA	NA	Inter	NA	0.3-0.7	0.3-0.8	0.3-0.4	0.6-0.5
13	19816		complex, snp	NA	NA	Inter	NA	0.9-0.4	0.7-0.2	0.7-0.5	1.0-1.0
131	77218		del	NA	NA	Inter	NA	0.2-0.7	0.3-1.0	0.3-0.5	0.0-0.0
134	234820		complex	NA	NA	Inter	NA	0.0-0.6	0.5-1.0	0.5-0.3	0.0-0.0
134	234839		snp	NA	NA	Inter	NA	0.0-0.7	0.5-1.0	0.5-0.3	0.0-0.0
135	257796		ins, ins	NA	NA	Inter	NA	1.0-0.6	0.9-0.5	0.9-0.9	1.0-1.0
138	86731		snp	NA	NA	Inter	NA	0.8-0.3	0.8-0.4	0.8-0.8	1.0-1.0
139	100470		snp	NA	NA	Inter	NA	0.1-0.7	0.2-0.6	0.2-0.3	0.3-0.2
14	477480		snp	NA	NA	Inter	NA	0.9-0.5	0.9-0.4	0.9-0.7	1.0-1.0
149	80311		snp	NA	NA	Inter	NA	1.0-0.0	0.7-0.2	0.7-0.6	0.2-0.0

152	85920	snp	NA	NA	Inter	NA	0.5-1.0	0.5-1.0	0.5-0.5	0.0-0.0
152	85927	ins	NA	NA	Inter	NA	0.5-1.0	0.5-1.0	0.5-0.6	0.0-0.0
153	114284	snp	NA	NA	Inter	NA	0.2-0.7	0.2-0.8	0.2-0.3	1.0-1.0
153	114291	complex	NA	NA	Inter	NA	0.2-0.8	0.3-0.8	0.3-0.3	1.0-1.0
16	418105	complex, snp	NA	NA	Inter	NA	1.0-0.5	0.5-0.1	0.5-0.5	1.0-0.9
174	156972	ins	NA	NA	Inter	NA	1.0-0.6	0.8-0.3	0.8-0.7	1.0-1.0
174	174879	complex, del	NA	NA	Inter	NA	0.0-0.6	0.1-0.6	0.1-0.1	0.0-0.0
174	174953	snp	NA	NA	Inter	NA	1.0-0.3	0.8-0.3	0.8-0.8	1.0-1.0
174	185681	snp	NA	NA	Inter	NA	0.0-0.7	0.2-1.0	0.2-0.4	0.0-0.0
18	208380	complex, complex, snp	NA	NA	Inter	NA	0.7-0.3	0.6-0.2	0.6-0.5	0.3-0.4
18	71983	snp	NA	NA	Inter	NA	0.0-0.7	0.4-0.8	0.4-0.3	1.0-1.0
18	72236	snp	NA	NA	Inter	NA	0.0-0.5	0.4-1.0	0.4-0.2	1.0-1.0
18	72290	snp	NA	NA	Inter	NA	0.0-0.5	0.4-0.9	0.4-0.2	1.0-1.0
182	5624	snp	NA	NA	Inter	NA	0.9-0.3	0.6-0.2	0.6-0.6	0.0-0.0
20	175688	complex, ins, snp	NA	NA	Inter	NA	1.0-0.4	0.6-0.1	0.6-0.4	0.0-0.0
206	47299	complex, complex, ins	NA	NA	Inter	NA	1.0-0.6	0.9-0.4	0.9-0.8	1.0-1.0
225	56297	complex, snp, snp	NA	NA	Inter	NA	1.0-0.4	1.0-0.5	1.0-1.0	1.0-1.0
234	54529	snp	NA	NA	Inter	NA	1.0-0.4	1.0-0.4	1.0-1.0	1.0-1.0
235	121421	snp	NA	NA	Inter	NA	0.1-0.7	0.4-0.8	0.4-0.3	1.0-1.0
236	63900	complex, snp	NA	NA	Inter	NA	1.0-0.5	0.6-0.2	0.6-0.8	0.6-0.4
236	72114	complex, complex, snp	NA	NA	Inter	NA	0.5-0.1	0.5-0.0	0.5-0.4	0.0-0.0
248	113975	snp	NA	NA	Inter	NA	0.0-0.6	0.6-1.0	0.6-0.8	1.0-0.9

249	80222	complex, mnp, snp	NA	NA	Inter	NA	0.1-0.6	0.3-0.8	0.3-0.5	0.0-0.0
260	68581	snp	NA	NA	Inter	NA	0.9-0.1	0.9-0.5	0.9-1.0	1.0-1.0
265	4619	complex, snp	NA	NA	Inter	NA	0.3-0.8	0.3-0.8	0.3-0.3	0.0-0.0
265	71807	snp	NA	NA	Inter	NA	0.3-0.8	0.3-0.8	0.3-0.4	1.0-1.0
298	17362	complex	NA	NA	Inter	NA	0.9-0.5	0.8-0.4	0.8-0.7	1.0-1.0
311	21163	snp	NA	NA	Inter	NA	1.0-0.4	0.4-0.0	0.4-0.5	0.0-0.0
314	25557	snp	NA	NA	Inter	NA	0.5-1.0	0.5-1.0	0.5-0.7	1.0-1.0
325	91667	complex	NA	NA	Inter	NA	0.8-0.1	0.7-0.1	0.7-0.7	0.4-0.4
325	91690	snp	NA	NA	Inter	NA	0.8-0.4	0.7-0.1	0.7-0.7	0.4-0.3
327	5769	snp	NA	NA	Inter	NA	1.0-0.6	0.9-0.4	0.9-0.8	1.0-1.0
328	56591	snp	NA	NA	Inter	NA	0.8-0.4	0.8-0.3	0.8-0.7	0.8-0.8
33	540854	snp	NA	NA	Inter	NA	1.0-0.5	0.7-0.3	0.7-0.8	1.0-1.0
355	76933	del	NA	NA	Inter	NA	0.0-1.0	0.2-0.7	0.2-0.3	1.0-1.0
359	73758	snp	NA	NA	Inter	NA	0.1-0.6	0.4-1.0	0.4-0.4	1.0-1.0
36	343692	snp	NA	NA	Inter	NA	1.0-0.4	0.9-0.5	0.9-0.9	1.0-1.0
36	466814	ins	NA	NA	Inter	NA	1.0-0.3	0.8-0.4	0.8-0.9	1.0-1.0
367	79211	snp	NA	NA	Inter	NA	0.4-0.9	0.5-0.9	0.5-0.5	0.4-0.5
38	206953	del, snp	NA	NA	Inter	NA	0.0-0.5	0.2-0.7	0.2-0.3	0.0-0.0
381	42365	snp	NA	NA	Inter	NA	0.5-0.9	0.6-1.0	0.6-0.4	0.5-0.4
4	427239	ins, snp	NA	NA	Inter	NA	0.3-0.9	0.5-1.0	0.5-0.5	1.0-1.0
40	143820	snp	NA	NA	Inter	NA	0.9-0.4	0.9-0.4	0.9-0.8	1.0-1.0
413	16497	complex, complex	NA	NA	Inter	NA	0.8-0.3	0.8-0.4	0.8-0.8	0.0-0.0
45	205723	snp	NA	NA	Inter	NA	0.3-0.8	0.6-1.0	0.6-0.5	1.0-1.0
45	324218	snp	NA	NA	Inter	NA	0.8-0.4	0.6-0.2	0.6-0.7	0.0-0.0
451	27416	snp	NA	NA	Inter	NA	0.5-0.9	0.2-0.7	0.2-0.0	0.8-0.9
463	22128	snp	NA	NA	Inter	NA	0.9-0.5	0.6-0.0	0.6-0.6	1.0-1.0
48	59725	snp, snp	NA	NA	Inter	NA	0.4-0.8	0.3-0.8	0.3-0.4	1.0-1.0
49	196931	snp	NA	NA	Inter	NA	0.1-0.6	0.4-0.8	0.4-0.4	0.0-0.0

49	55825	complex, complex, complex	NA	NA	Inter	NA	0.6-0.1	0.5-0.0	0.5-0.7	0.0-0.0
497	37968	snp	NA	NA	Inter	NA	1.0-0.4	0.9-0.4	0.9-1.0	1.0-1.0
503	30030	ins, snp, snp	NA	NA	Inter	NA	0.0-0.6	0.2-0.6	0.2-0.2	0.0-0.0
503	30035	complex, complex, mnp	NA	NA	Inter	NA	0.0-0.6	0.2-0.6	0.2-0.2	0.0-0.0
52	143677	snp	NA	NA	Inter	NA	0.8-0.4	0.4-0.0	0.4-0.5	1.0-1.0
541	24454	del	NA	NA	Inter	NA	1.0-0.4	0.8-0.4	0.8-0.9	1.0-1.0
541	26069	snp	NA	NA	Inter	NA	1.0-0.6	0.9-0.2	0.9-0.9	1.0-1.0
564	6312	complex, mnp	NA	NA	Inter	NA	0.9-0.3	1.0-0.6	1.0-0.9	0.6-0.7
571	5718	snp, snp	NA	NA	Inter	NA	0.5-1.0	0.5-1.0	0.5-0.5	0.0-0.0
587	201	snp	NA	NA	Inter	NA	1.0-0.4	0.9-0.4	0.9-0.9	1.0-1.0
587	477	snp	NA	NA	Inter	NA	0.1-0.9	0.2-0.8	0.2-0.4	0.0-0.0
609	27668	complex, complex, snp	NA	NA	Inter	NA	0.2-0.7	0.4-0.9	0.4-0.3	0.6-0.5
609	27684	complex, mnp	NA	NA	Inter	NA	0.2-0.7	0.4-0.9	0.4-0.4	0.6-0.5
63	60667	snp	NA	NA	Inter	NA	0.7-0.2	0.5-0.0	0.5-0.4	0.3-0.4
72	211540	complex, del, snp	NA	NA	Inter	NA	0.3-0.8	0.4-0.9	0.4-0.6	0.4-0.4
73	276024	ins, ins	NA	NA	Inter	NA	1.0-0.6	0.8-0.2	0.8-0.7	1.0-1.0
80	202853	snp	NA	NA	Inter	NA	0.9-0.5	0.8-0.4	0.8-1.0	1.0-1.0
82	177862	snp	NA	NA	Inter	NA	0.9-0.4	1.0-0.6	1.0-1.0	1.0-1.0
86	198890	del, ins, snp	NA	NA	Inter	NA	0.0-0.7	0.6-1.0	0.6-0.4	1.0-1.0
9	506010	complex, del, ins	NA	NA	Inter	NA	1.0-0.4	0.8-0.2	0.8-0.9	1.0-1.0
90	300810	snp	NA	NA	Inter	NA	1.0-0.2	0.5-0.0	0.5-0.3	0.0-0.0

90	300832	snp	NA	NA	Inter	NA	1.0-0.6	1.0-0.6	1.0-1.0	1.0-0.8
90	48477	snp	NA	NA	Inter	NA	0.3-0.9	0.6-1.0	0.6-0.7	1.0-1.0
95	156905	complex, complex, ins	NA	NA	Inter	NA	0.8-0.3	0.9-0.1	0.9-0.8	0.8-0.7

¹Complex = multiple type of nucleotide changes; del = deletion; ins = insertion; mnp = multiple nucleotide polymorphism; snp = single nucleotide polymorphism.

²Conservative inframe deletion = Cidel; conservative inframe insertion = Clins; disruptive inframe insertion = Dlins; frameshift = Fs; missense = Mis; intergenic = Inter; intron = Int; SPLICE DONOR REGION = sPd; splice region variant = SpR; stop gained = StG; synonymous = Syn.

³High = H; Low = L; Moderate = M; Modifier = Mod.

Table S. 5.4 Variants associated with an increasing allele frequency toward homozygous variants ($AF < 0.05$ or > 0.95) associated with females selected on PI88788 whereas alleles of females selected on Peking remain unstable.

Scaffold ID	Position	Hetgly Transcript ID	Variant type ¹	Blast2GO best hit description	e-Value	Effect type ²	Variant impact ³	Allelic frequencies			
								IL4 Founder-PI88788	ON1 Founder-PI88788	ON1 Founder-Peking	ON34 Founder-Peking
1	1816544	T000000890.1	complex, ins	NA	NA	Fs, Fs	H, H	0.5-1.0	0.6-1.0	0.6-0.6	1.0-0.8
12	304537	T000002697.1	ins	NA	NA	Fs & StaL	H	0.8-1.0	0.8-1.0	0.8-0.7	0.8-0.5
14	459766	T000007641.1	del, snp	NA	NA	Fs, Fs	H, H	0.3-0.0	0.2-0.0	0.2-0.3	0.5-0.4
14	459892	T000007641.1	snp	NA	NA	SpA & Int	H	0.2-0.0	0.2-0.0	0.2-0.2	0.4-0.4
34	50107	T000010623.1	del	NA	NA	Fs	H	0.8-1.0	0.7-1.0	0.7-0.7	0.0-0.1
71	365565	T000010362.1	ins	hypothetical protein CRE_17270	0	Fs	H	0.8-1.0	0.6-1.0	0.6-0.7	0.8-0.8
91	50397	T000018228.1	snp, snp	POU domain, class 3, transcription factor 3	1,29E-18	StG, Mis	H, M	0.8-1.0	0.8-1.0	0.8-0.8	0.7-0.8
137	104461	T000012581.1	ins, snp	ADP-ribose pyrophosphatase, mitochondrial	4,37E-09	Fs, Mis	H, M	0.7-1.0	0.7-1.0	0.7-0.6	0.3-0.5
149	112949	T000022084.1	ins, ins, ins	NA	NA	Fs, Fs, Fs	H, H, H	0.6-0.0	0.4-0.0	0.4-0.5	0.4-0.7
576	36216	T000028402.1	del	NA	NA	Fs	H	0.5-1.0	0.6-1.0	0.6-0.2	1.0-0.7
1	77317	T000000441.1	complex, complex, complex	NA	NA	Mis, Mis, Mis	M, M, M	0.3-0.0	0.5-0.0	0.5-0.4	0.8-0.9
1	873420	T000000669.1	snp	Proline-rich protein 12	3,47E-55	Mis	M	0.4-1.0	0.8-1.0	0.8-0.4	1.0-0.7
1	1149341	T000000758.1	snp	hypothetical protein Tcan_05939	9,01E-11	Mis	M	0.7-1.0	0.8-1.0	0.8-0.8	0.7-0.7
1	1816522	T000000890.1	complex, snp	NA	NA	Mis, Mis	M, M	0.5-1.0	0.6-1.0	0.6-0.6	1.0-0.8
1	1816539	T000000890.1	snp	NA	NA	Mis	M	0.5-1.0	0.6-1.0	0.6-0.6	1.0-0.8
12	126930	T000002641.1	del	NA	NA	Didel	M	0.7-1.0	0.7-1.0	0.7-0.8	0.5-0.5
12	126945	T000002641.1	snp	NA	NA	Mis	M	0.7-1.0	0.7-1.0	0.7-0.8	0.5-0.4
12	127040	T000002641.1	snp	NA	NA	Mis	M	0.8-1.0	0.7-1.0	0.7-0.8	0.3-0.4
12	727438	T000002816.1	complex, snp	4-hydroxybenzoate polyprenyltransferase, mitochondrial isoform X1	1,48E-114	Mis, Syn	M, L	0.4-1.0	0.6-1.0	0.6-0.6	0.0-0.4
14	459575	T000007641.1	snp	NA	NA	Mis	M	0.2-0.0	0.2-0.0	0.2-0.2	0.4-0.5
14	459803	T000007641.1	snp	NA	NA	Mis	M	0.2-0.0	0.2-0.0	0.2-0.3	0.4-0.3
16	164129	T000005001.1	snp	NA	NA	Mis	M	0.8-1.0	0.7-1.0	0.7-0.6	0.5-0.6

26	356372	T000006099.1	complex, snp	zinc transporter 1	8,28E-139	Mis, Mis	M, M	0.6-1.0	0.8-1.0	0.8-0.5	0.0-0.1
28	210497	T000010791.1	snp	Vesicular PolyAmine Transporter	4,41E-68	Mis	M	0.4-1.0	0.6-1.0	0.6-0.6	0.3-0.5
28	215654	T000010792.1	snp	Vesicular PolyAmine Transporter	4,91E-56	Mis	M	0.7-1.0	0.8-1.0	0.8-0.7	0.6-0.7
34	36124	T000010619.1	snp	putative gland protein G20E03	9,82E-121	Mis	M	0.4-0.0	0.4-0.0	0.4-0.4	0.3-0.2
34	313800	T000010677.1	complex, mnp, snp	NA	NA	Mis, Mis, Mis	M, M, M	0.8-1.0	0.7-1.0	0.7-0.6	1.0-0.4
34	571740	T000010727.1	snp	antiviral helicase	0	Mis	M	0.8-1.0	0.8-1.0	0.8-0.6	0.3-0.4
34	574799	T000010728.1	complex, snp	NA	NA	Mis, Mis	M, M	0.9-1.0	0.7-1.0	0.7-0.6	0.0-0.0
89	214080	T000017040.1	snp	Elongation factor Tu GTP-binding domain-containing protein 1	1,72E-72	Mis	M	0.8-1.0	0.8-1.0	0.8-0.5	0.0-0.0
101	115890	T000016030.1	complex, snp, snp	RNA-directed DNA polymerase (reverse transcriptase) and Integrase domain containing protein	3,96E-114	Mis, Mis, Syn	M, M, L	0.3-0.0	0.3-0.0	0.3-0.2	0.2-0.3
101	141620	T000016040.1	snp	protein TANC1 isoform X1	3,81E-19	Mis	M	0.6-0.0	0.2-0.0	0.2-0.4	0.8-0.6
114	167374	T000017737.1	snp, snp	CRE-ROD-1 protein	3,55E-134	Mis, Mis	M, M	0.6-1.0	0.7-1.0	0.7-0.8	0.1-0.2
114	167384	T000017737.1	snp	CRE-ROD-1 protein	3,55E-134	Mis	M	0.6-1.0	0.7-1.0	0.7-0.8	0.1-0.2
114	167390	T000017737.1	complex	CRE-ROD-1 protein	3,55E-134	Mis	M	0.6-1.0	0.7-1.0	0.7-0.8	0.1-0.2
148	207224	T000014967.1	complex	Talin-1	0	Mis	M	0.8-1.0	0.8-1.0	0.8-0.8	0.6-0.6
188	100704	T000018675.1	complex	putative gland protein G17G01	3,51E-33	Mis	M	0.8-1.0	0.7-1.0	0.7-0.8	0.3-0.6
207	47486	T000021128.1	snp	class II Aldolase and Adducin domain protein	4,18E-72	Mis	M	0.8-1.0	0.8-1.0	0.8-0.8	0.9-0.8
249	100076	T000020222.1	complex, snp	poly(A) polymerase gamma-like isoform X1	3,18E-40	Mis, Mis	M, M	0.8-1.0	0.8-1.0	0.8-0.7	0.8-0.2
260	45222	T000025559.1	snp	pectate lyase 1	1,49E-75	Mis	M	0.7-1.0	0.8-1.0	0.8-0.8	1.0-0.7
260	45232	T000025559.1	snp	pectate lyase 1	1,49E-75	Mis	M	0.7-1.0	0.8-1.0	0.8-0.8	1.0-0.7
260	45237	T000025559.1	snp	pectate lyase 1	1,49E-75	Mis	M	0.7-1.0	0.8-1.0	0.8-0.8	1.0-0.7
296	44441	T000025586.1	del	NA	NA	Didel	M	0.3-0.0	0.3-0.0	0.3-0.2	0.6-0.7
308	48679	T000019634.1	snp	Leucine-rich repeat-containing protein let-4	6,67E-56	Mis	M	0.8-1.0	0.8-1.0	0.8-0.8	0.9-0.5
308	48735	T000019634.1	complex, snp	Leucine-rich repeat-containing protein let-4	6,67E-56	Mis, Mis	M, M	0.7-1.0	0.8-1.0	0.8-0.8	0.9-0.7
308	48741	T000019634.1	complex, snp	Leucine-rich repeat-containing protein let-4	6,67E-56	Mis, Syn	M, L	0.7-1.0	0.8-1.0	0.8-0.8	0.9-0.7
308	48755	T000019634.1	complex,	Leucine-rich repeat-containing	6,67E-56	Mis,	M, L	0.7-1.0	0.8-1.0	0.8-0.8	0.9-0.7

			snp	protein let-4		Syn						
308	48766	T000019634.1	complex, snp	Leucine-rich repeat-containing protein let-4	6,67E-56	Mis, Mis	M, M	0.7-1.0	0.8-1.0	0.8-0.8	0.9-0.6	
374	30089	T000027794.1	complex, snp	Nucleolar protein 10	1,08E-92	Mis, Mis	M, M	0.5-1.0	0.6-1.0	0.6-0.8	0.0-0.0	
387	19251	T000028484.1	complex, complex, snp	Zinc knuckle family protein	1,33E-56	Mis	M	0.6-1.0	0.6-1.0	0.6-0.7	0.6-0.4	
463	27304	T000024745.1	complex, snp	Lissencephaly-1 -like protein A	4,92E-27	Mis, Syn	M, L	0.2-1.0	0.8-1.0	0.8-0.6	0.0-0.0	
517	40746	T000027947.1	snp	NA	NA	Mis	M	0.8-1.0	0.6-1.0	0.6-0.5	0.3-0.7	
1	1149390	T000000758.1	snp	hypothetical protein Tcan_05939	9,01E-11	Syn	L	0.6-1.0	0.8-1.0	0.8-0.7	0.7-0.6	
12	126951	T000002641.1	snp	NA	NA	Syn	L	0.7-1.0	0.7-1.0	0.7-0.8	0.6-0.5	
12	492241	T000002751.1	snp	Sec16 ortholog	5,14E-141	Syn	L	0.6-1.0	0.5-1.0	0.5-0.7	0.6-0.4	
12	720329	T000002814.1	ins	Bm13019, isoform b	2,98E-19	SpR & Int	L	0.5-1.0	0.8-1.0	0.8-0.7	0.5-0.4	
12	734736	T000002818.1	snp	pyridoxal-dependent decarboxylase domain protein	0	Syn	L	0.6-1.0	0.6-1.0	0.6-0.7	0.7-0.5	
14	459563	T000007641.1	complex	NA	NA	SpR & Int	L	0.2-0.0	0.2-0.0	0.2-0.2	0.4-0.6	
14	459795	T000007641.1	snp	NA	NA	Syn	L	0.2-0.0	0.2-0.0	0.2-0.2	0.5-0.3	
17	618859	T000007182.1	snp	CUB domain protein	1,19E-10	Syn	L	0.8-1.0	0.7-1.0	0.7-0.8	0.2-0.8	
34	313714	T000010677.1	snp	NA	NA	Syn	L	0.7-1.0	0.6-1.0	0.6-0.6	1.0-0.7	
44	17323	T000015492.1	snp	AMP-binding enzyme	2,19E-50	Syn	L	0.3-0.0	0.2-0.0	0.2-0.4	0.5-0.8	
44	17337	T000015492.1	complex, snp	AMP-binding enzyme	2,19E-50	SpR & Int	L	0.3-0.0	0.2-0.0	0.2-0.3	0.6-0.8	
52	245764	T000005679.1	snp	poly-gamma-glutamate biosynthesis protein	8,03E-136	Syn	L	0.3-1.0	0.7-1.0	0.7-0.5	1.0-0.7	
65	347682	T000017478.1	snp	rab GTPase-activating protein 1- like isoform X1	1,37E-17	SpR & Int	L	0.8-1.0	0.8-1.0	0.8-0.8	1.0-0.8	
82	46319	T000013927.1	snp	Nck-associated protein 1	0	Syn	L	0.8-1.0	0.5-1.0	0.5-0.7	0.0-0.0	
82	50180	T000013927.1	complex, snp	Nck-associated protein 1	0	Syn	L	0.5-1.0	0.7-1.0	0.7-0.5	0.0-0.0	
82	50198	T000013927.1	snp	Nck-associated protein 1	0	Syn	L	0.5-1.0	0.7-1.0	0.7-0.5	0.0-0.0	
127	130694	T000011707.1	snp	ribose-phosphate diphosphokiNase	0	Syn	L	0.3-0.0	0.2-0.0	0.2-0.2	0.7-0.9	
137	59716	T000012568.1	complex, ins	importin-7 isoform X1	0	SpR & Int, SpR & Int	L, L	0.8-1.0	0.7-1.0	0.7-0.6	0.0-0.0	
188	156402	T000018690.1	snp	putative gland protein G17G01	4,87E-88	Syn	L	0.7-1.0	0.7-1.0	0.7-0.8	0.6-0.6	

193	126724	T000018732.2	snp	Methyltransferase-like protein 13	1,44E-34	Syn	L	0.8-1.0	0.8-1.0	0.8-0.8	0.8-0.5
225	105145	T000027889.1	snp	zinc finger protein 3-like isoform X2	1,97E-09	Syn	L	0.8-1.0	0.8-1.0	0.8-0.8	0.5-0.6
250	48773	T000025426.1	snp	NA	NA	SpR & Int	L	0.8-1.0	0.7-1.0	0.7-0.8	0.4-0.7
250	129401	T000025443.1	snp	dorsal gland cell-specific expression protein	9,21E-13	SpR & Int	L	0.2-0.0	0.4-0.0	0.4-0.7	0.0-0.4
260	45213	T000025559.1	snp	pectate lyase 1	1,49E-75	SpR & Int	L	0.7-1.0	0.8-1.0	0.8-0.8	1.0-0.7
260	45242	T000025559.1	snp	pectate lyase 1	1,49E-75	Syn	L	0.7-1.0	0.8-1.0	0.8-0.8	1.0-0.7
270	76865	T000024046.1	snp	Pao retrotransposon peptidase family protein	3,46E-89	Syn	L	0.7-1.0	0.6-1.0	0.6-0.6	0.0-0.0
275	81552	T000023265.1	snp	retrotransposon gag protein	2,43E-19	Syn	L	0.7-1.0	0.8-1.0	0.8-0.7	0.5-0.7
277	51732	T000024471.1	snp	dorsal gland cell-specific expression protein	1,02E-35	Syn	L	0.5-1.0	0.7-1.0	0.7-0.6	0.8-0.5
285	59025	T000020577.1	snp	Polycystin cation channel, PKD1/PKD2 domain-containing protein	0	Syn	L	0.7-1.0	0.8-1.0	0.8-0.8	1.0-0.6
285	59034	T000020577.1	snp	Polycystin cation channel, PKD1/PKD2 domain-containing protein	0	Syn	L	0.8-1.0	0.8-1.0	0.8-0.8	1.0-0.6
296	44979	T000025587.1	complex, snp	NA	NA	InC, InC	L, L	0.2-0.0	0.2-0.0	0.2-0.2	0.6-0.5
298	28295	T000026899.1	snp	Coatomer subunit gamma	0	Syn	L	0.8-1.0	0.8-1.0	0.8-0.7	0.7-0.3
308	48714	T000019634.1	snp	Leucine-rich repeat-containing protein let-4	6,67E-56	Syn	L	0.6-1.0	0.7-1.0	0.7-0.8	0.9-0.5
469	54842	T000027310.1	snp	DNA helicase Pif1-like	3,29E-120	Syn	L	0.7-1.0	0.7-1.0	0.7-0.7	0.8-0.8
474	60582	T000029311.1	snp	NA	NA	Syn	L	0.6-1.0	0.7-1.0	0.7-0.6	1.0-0.8
1	1324371	T000000808.1	complex, del	protein CutA homolog	1,19E-38	Int, Int	Mod, Mod	0.7-1.0	0.8-1.0	0.8-0.5	0.7-0.7
1	1324405	T000000808.1	del, snp	protein CutA homolog	1,19E-38	Int, Int	Mod, Mod	0.7-1.0	0.8-1.0	0.8-0.6	0.8-0.6
9	250473	T000005803.1	snp	PREDICTED: thrombospondin-3	5,30E-12	Int	Mod	0.6-1.0	0.6-1.0	0.6-0.7	0.7-0.7
11	58327	T000001265.1	snp	Transcription elongation factor B polypeptide 3	2,17E-60	Int	Mod	0.5-1.0	0.6-1.0	0.6-0.6	0.4-0.7
12	128028	T000002641.1	snp	NA	NA	Int	Mod	0.8-1.0	0.7-1.0	0.7-0.2	0.7-0.5
12	129997	T000002641.1	snp	NA	NA	Int	Mod	0.8-1.0	0.6-1.0	0.6-0.8	1.0-0.4
12	803514	T000002837.1	ins, ins	PaTched Related family	2,93E-107	Int, Int	Mod, Mod	0.5-1.0	0.7-1.0	0.7-0.7	0.3-0.8
14	459549	T000007641.1	snp	NA	NA	Int	Mod	0.2-0.0	0.2-0.0	0.2-0.2	0.4-0.6
14	488953	T000007645.2	snp	Protein lap1	1,40E-169	Int	Mod	0.8-1.0	0.8-1.0	0.8-0.6	0.0-0.0

16	490929	T000005066.1	snp	TWiK family of potassium channels	3,37E-124	Int	Mod	0.8-1.0	0.5-1.0	0.5-0.8	1.0-0.7
16	490963	T000005066.1	snp	TWiK family of potassium channels	3,37E-124	Int	Mod	0.8-1.0	0.6-1.0	0.6-0.8	1.0-0.7
16	540299	T000005076.1	snp	Innexin unc-9	1,85E-08	Int	Mod	0.7-1.0	0.8-1.0	0.8-0.8	0.7-0.6
17	643844	T000007186.1	ins	NA	NA	Int	Mod	0.5-1.0	0.7-1.0	0.7-0.6	0.8-0.8
18	617225	T000010092.1	snp	hypothetical protein ZOSMA_5828G00010, partial	9,10E-35	Int	Mod	0.6-1.0	0.8-1.0	0.8-0.7	1.0-0.8
18	817920	T000010142.1	complex, snp	integrase core domain protein	2,27E-46	Int, Int	Mod, Mod	0.7-1.0	0.7-1.0	0.7-0.8	0.8-0.8
23	344325	T000006991.2	complex, del, del	zinc finger, C2H2 type	3,61E-73	Int, Int, Int	Mod, Mod, Mod	0.3-0.0	0.4-0.0	0.4-0.5	0.3-0.4
26	356443	T000006099.1	complex, snp	zinc transporter 1	8,28E-139	Int, Int	Mod, Mod	0.6-1.0	0.8-1.0	0.8-0.5	0.0-0.1
30	187721	T000008195.1	snp	Protein CBG10707	1,25E-21	Int	Mod	0.7-1.0	0.6-1.0	0.6-0.4	0.1-0.2
34	40516	T000010620.1	complex	glutathione synthase	5,86E-165	Int	Mod	0.7-1.0	0.6-1.0	0.6-0.7	0.4-0.7
45	376234	T000016334.1	snp	Zinc metalloproteinase NAs-12	2,92E-72	Int	Mod	0.2-0.0	0.2-0.0	0.2-0.2	0.0-0.2
49	136583	T000011009.1	snp	NA	NA	Int	Mod	0.6-1.0	0.7-1.0	0.7-0.6	0.6-0.6
52	85097	T000005644.1	snp	AFG3-like protein 2	0	Int	Mod	0.8-0.0	0.3-0.0	0.3-0.3	1.0-1.0
52	125909	T000005650.1	snp	NA	NA	Int	Mod	0.2-1.0	0.6-1.0	0.6-0.7	0.1-0.0
52	276998	T000005688.1	complex, ins	WAP-type 'four-disulfide core	1,20E-90	Int, Int	Mod, Mod	0.7-0.0	0.3-0.0	0.3-0.5	0.1-0.2
58	164254	T000013122.1	snp	LIM domain protein	2,15E-116	Int	Mod	0.1-1.0	0.7-1.0	0.7-0.7	0.1-0.1
64	198921	T000014444.1	snp	AF273736_1hypothetical esophageal gland cell secretory protein 9	4,51E-52	Int	Mod	0.7-1.0	0.7-1.0	0.7-0.8	0.7-0.4
70	311073	T000010227.1	snp	60S ribosomal protein L18	1,14E-91	Int	Mod	0.5-1.0	0.8-1.0	0.8-0.6	0.0-0.0
70	346731	T000010238.1	snp	transposase	6,26E-44	Int	Mod	0.5-1.0	0.7-1.0	0.7-0.7	0.0-0.0
70	361339	T000010242.1	ins	SMAD domain, Dwarfina-type and MAD homology 1, Dwarfina-type domain and SMAD/FHA domain and MAD homology, MH1 domain and SMAD domain-like-containing protein	6,95E-22	Int	Mod	0.6-1.0	0.8-1.0	0.8-0.8	0.3-0.3
70	361422	T000010242.1	complex, snp	SMAD domain, Dwarfina-type and MAD homology 1, Dwarfina-type domain and SMAD/FHA domain and MAD homology, MH1 domain and SMAD domain-like-containing protein	6,95E-22	Int, Int	Mod, Mod	0.6-1.0	0.6-1.0	0.6-0.8	0.2-0.3
70	375964	T000010245.1	complex,	SMAD domain, Dwarfina-type	5,73E-58	Int,	Mod,	0.4-1.0	0.8-1.0	0.8-0.5	0.3-0.0

			snp	and MAD homology 1, Dwarfintype domain and SMAD/FHA domain and MAD homology, MH1 domain and SMAD domain-like-containing protein		Int	Mod					
74	147204	T000014997.1	snp	patched family protein	1,40E-150	Int	Mod	0.4-1.0	0.8-1.0	0.8-0.6	0.0-0.0	
82	68152	T000013930.1	ins	NA	NA	Int	Mod	0.7-1.0	0.7-1.0	0.7-0.5	0.0-0.0	
89	204816	T000017038.2	complex, complex, del	Ubiquitin carboxyl-terminal hydrolase 7	9,01E-18	Int, Int, Int	Mod, Mod, Mod	0.7-1.0	0.7-1.0	0.7-0.6	0.0-0.3	
89	214217	T000017040.1	snp	Elongation factor Tu GTP-binding domain-containing protein 1	1,72E-72	Int	Mod	0.8-1.0	0.7-1.0	0.7-0.6	0.0-0.0	
89	241582	T000017042.1	complex, complex, complex	patched family protein	0	Int, Int, Int	Mod, Mod, Mod	0.8-1.0	0.8-1.0	0.8-0.6	0.5-0.2	
114	167409	T000017737.1	complex, snp, snp	CRE-ROD-1 protein	3,55E-134	Int, Int, Int	Mod, Mod, Mod	0.6-1.0	0.7-1.0	0.7-0.8	0.2-0.1	
121	118320	T000013548.1	complex, snp	probable ATP-dependent RNA helicase DHX34	0	Int, Int	Mod, Mod	0.6-1.0	0.8-1.0	0.8-0.8	0.7-0.8	
122	151650	T000013798.1	snp	CRE-SUR-7 protein	1,08E-17	Int	Mod	0.8-1.0	0.8-1.0	0.8-0.4	1.0-0.8	
124	189973	T000012857.1	snp	Uncharacterized protein F10E9.4	1,17E-09	Int	Mod	0.4-1.0	0.7-1.0	0.7-0.8	0.3-0.3	
128	159347	T000019691.1	ins	NA	NA	Int	Mod	0.5-0.0	0.3-0.0	0.3-0.2	1.0-1.0	
137	100073	T000012580.1	complex, complex, complex	Tectonin beta-propeller repeat-containing protein 2	1,02E-51	Int	Mod	0.7-0.0	0.3-0.0	0.3-0.3	0.9-1.0	
140	196326	T000014897.1	complex, complex, del	Ribosomal Proteins L2, RNA binding domain protein	1,23E-104	Int, Int, Int	Mod, Mod, Mod	0.6-1.0	0.8-1.0	0.8-0.8	0.7-0.5	
142	97479	T000019598.1	del, snp	intermediate filament tail domain protein	0	Int, Int	Mod, Mod	0.4-1.0	0.3-1.0	0.3-0.6	0.0-0.1	
154	71520	T000020493.1	snp	zinc finger CCHC domain-containing protein 8	1,61E-15	Int	Mod	0.8-1.0	0.6-1.0	0.6-0.7	0.6-0.7	
154	167017	T000020514.1	snp	NA	NA	Int	Mod	0.7-1.0	0.8-1.0	0.8-0.8	1.0-0.8	
188	102937	T000018676.1	snp	NA	NA	Int	Mod	0.3-0.0	0.2-0.0	0.2-0.2	0.5-0.3	
188	156448	T000018690.1	complex, snp	putative gland protein G17G01	4,87E-88	Int, Int	Mod, Mod	0.7-1.0	0.8-1.0	0.8-0.8	0.7-0.4	
200	71991	T000021100.1	snp	L-density lipoprotein receptor-related protein	0	Int	Mod	0.7-1.0	0.2-1.0	0.2-0.4	0.4-0.1	
209	76825	T000022449.1	complex,	S-antigen	8,35E-13	Int,	Mod,	0.3-1.0	0.8-1.0	0.8-0.4	0.2-0.6	

			snp			Int	Mod				
222	152194	T000025147.1	snp	Transposable element Tc3 transposase	5,82E-05	Int	Mod	0.5-0.0	0.7-0.0	0.7-0.8	1.0-1.0
236	34156	T000026150.1	snp	Astacin-like metalloendopeptidase	1,59E-14	Int	Mod	0.7-1.0	0.8-1.0	0.8-0.8	1.0-0.8
374	29991	T000027794.1	complex, snp	Nucleolar protein 10	1,08E-92	Int, Int	Mod, Mod	0.5-1.0	0.6-1.0	0.6-0.7	0.0-0.0
439	721	T000026754.1	snp	Multiple PDZ domain protein	0	Int	Mod	0.8-1.0	0.6-1.0	0.6-0.4	0.4-0.4
469	30116	T000027308.1	snp	NA	NA	Int	Mod	0.3-0.0	0.3-0.0	0.3-0.4	0.2-0.2
1	62012		snp	NA	NA	Inter	NA	0.7-1.0	0.7-1.0	0.7-0.7	0.5-0.4
1	1964626		complex, snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	0.8-0.6
12	304028		del	NA	NA	Inter	NA	0.7-1.0	0.7-1.0	0.7-0.6	0.6-0.3
15	114048		snp, snp	NA	NA	Inter	NA	0.2-0.0	0.2-0.0	0.2-0.2	1.0-1.0
15	501771		snp	NA	NA	Inter	NA	0.5-0.0	0.3-0.0	0.3-0.3	0.3-0.4
16	203115		complex, complex, complex	NA	NA	Inter	NA	0.5-0.0	0.3-0.0	0.3-0.6	0.4-0.3
21	644562		snp	NA	NA	Inter	NA	0.7-1.0	0.6-1.0	0.6-0.8	0.7-0.8
25	452116		ins	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	0.7-0.5
34	434288		snp	NA	NA	Inter	NA	0.7-1.0	0.7-1.0	0.7-0.8	1.0-0.7
34	580490		snp	NA	NA	Inter	NA	0.7-1.0	0.7-1.0	0.7-0.8	0.0-0.0
38	224030		complex, snp, snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.4	1.0-0.8
39	283193		complex, ins	NA	NA	Inter	NA	0.2-0.0	0.3-0.0	0.3-0.4	0.5-0.2
41	483358		snp, snp	NA	NA	Inter	NA	0.8-1.0	0.7-1.0	0.7-0.7	1.0-0.7
41	483382		complex, snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.7	1.0-0.7
45	113616		snp	NA	NA	Inter	NA	0.7-1.0	0.4-1.0	0.4-0.7	0.4-0.8
45	456118		snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	0.7-0.5
52	415901		snp	NA	NA	Inter	NA	0.3-1.0	0.6-1.0	0.6-0.8	0.8-0.8
55	92265		complex, del, snp	NA	NA	Inter	NA	0.0-1.0	0.8-1.0	0.8-0.8	0.6-0.5
66	140074		snp	NA	NA	Inter	NA	0.8-1.0	0.7-1.0	0.7-0.7	0.8-0.7
66	140079		snp	NA	NA	Inter	NA	0.8-1.0	0.6-1.0	0.6-0.7	0.8-0.6
67	89685		snp	NA	NA	Inter	NA	0.7-1.0	0.8-1.0	0.8-0.8	0.5-0.6
69	37958		snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	0.9-0.7
81	138441		complex	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	1.0-0.6
82	199807		snp	NA	NA	Inter	NA	0.5-1.0	0.8-1.0	0.8-0.6	1.0-0.7
87	73759		snp	NA	NA	Inter	NA	0.6-1.0	0.7-1.0	0.7-0.6	1.0-0.8
89	229909		snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.7	0.3-0.4

89	229936	complex, complex, snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.7	0.4-0.5
94	116655	snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	0.5-0.6
105	149144	ins	NA	NA	Inter	NA	0.8-1.0	0.7-1.0	0.7-0.8	0.0-0.1
139	76487	snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.7	0.4-0.5
152	85920	snp	NA	NA	Inter	NA	0.5-1.0	0.5-1.0	0.5-0.5	0.0-0.0
152	85927	ins	NA	NA	Inter	NA	0.5-1.0	0.5-1.0	0.5-0.6	0.0-0.0
188	43991	complex, snp, snp	NA	NA	Inter	NA	0.7-1.0	0.8-1.0	0.8-0.7	1.0-0.7
188	104536	snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	0.8-0.8
188	104562	ins, snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.7	0.4-0.5
191	120593	complex, snp	NA	NA	Inter	NA	0.3-0.0	0.2-0.0	0.2-0.3	0.7-0.4
197	67948	complex	NA	NA	Inter	NA	0.8-1.0	0.7-1.0	0.7-0.8	1.0-0.7
207	75873	snp	NA	NA	Inter	NA	0.7-1.0	0.5-1.0	0.5-0.7	0.9-0.8
221	163912	complex, complex, snp	NA	NA	Inter	NA	0.6-1.0	0.5-1.0	0.5-0.8	1.0-0.5
236	38705	complex, ins, ins	NA	NA	Inter	NA	0.3-0.0	0.2-0.0	0.2-0.3	0.3-0.8
236	86525	del	NA	NA	Inter	NA	0.7-1.0	0.8-1.0	0.8-0.7	1.0-0.8
250	21995	ins	NA	NA	Inter	NA	0.7-1.0	0.7-1.0	0.7-0.8	0.0-0.4
252	93988	ins	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.5	0.6-0.8
252	93995	snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.5	0.6-0.8
264	86756	snp	NA	NA	Inter	NA	0.7-1.0	0.8-1.0	0.8-0.7	1.0-0.7
268	25854	snp	NA	NA	Inter	NA	0.4-0.0	0.4-0.0	0.4-0.2	0.2-0.3
268	25864	complex, snp	NA	NA	Inter	NA	0.4-0.0	0.4-0.0	0.4-0.2	0.2-0.3
323	34737	snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.7	0.6-0.5
331	33481	complex, complex	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	1.0-0.8
331	33511	complex	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	1.0-0.7
381	86576	snp	NA	NA	Inter	NA	0.6-1.0	0.8-1.0	0.8-0.8	0.8-0.7
388	63320	snp	NA	NA	Inter	NA	0.6-1.0	0.6-1.0	0.6-0.7	0.5-0.6
474	60510	snp	NA	NA	Inter	NA	0.7-1.0	0.7-1.0	0.7-0.6	1.0-0.8
474	60523	snp	NA	NA	Inter	NA	0.7-1.0	0.7-1.0	0.7-0.6	1.0-0.8
571	5718	snp, snp	NA	NA	Inter	NA	0.5-1.0	0.5-1.0	0.5-0.5	0.0-0.0
648	12836	snp	NA	NA	Inter	NA	0.8-1.0	0.8-1.0	0.8-0.8	0.6-0.5
648	12897	snp	NA	NA	Inter	NA	0.8-1.0	0.7-1.0	0.7-0.8	0.7-0.4
648	13092	snp	NA	NA	Inter	NA	0.8-1.0	0.7-1.0	0.7-0.6	0.6-0.3

658	21181	snp	NA	NA	Inter	NA	0.7-1.0	0.7-1.0	0.7-0.7	0.8-0.6
-----	-------	-----	----	----	-------	----	----------------	----------------	---------	---------

¹ Complex = multiple type of nucleotide changes; del = deletion; ins = insertion; mnp = multiple nucleotide polymorphism; snp = single nucleotide polymorphism.

² Conservative inframe deletion = CIdel; conservative inframe insertion = Clins; disruptive inframe insertion = Dlins; frameshift = Fs; initiator codon variant = InC; missense = Mis; intergenic = Inter; intron = Int; splice acceptor variant = SpA; splice donor region = sPd; splice region variant = SpR; start lost = StaL; stop gained = StG; synonymous = Syn.

³ High = H; Low = L; Moderate = M; Modifier = Mod

Table S. 5.5 Design matrix for edgeR contrast.

Sample_id	Peking*	PI88788
kallisto_N701---N503_ON21_1_2_J2	0	0
kallisto_N701---N504_ON1_1-1	1	1
kallisto_N701---S517_ON24_2-3	0	2
kallisto_N702---N502_ON6_1-2	2	0
kallisto_N702---N504_ON1_1-3	1	1
kallisto_N702---S517_ON1_2-3	1	1
kallisto_N703---N502_ON6_1-4	2	0
kallisto_N703---N504_ON21_1-4	1	1
kallisto_N703---S517_ON7_1-3	0	2
kallisto_N704---N502_ON16_2-2	2	0
kallisto_N704---N504_ON21_2-2	1	1
kallisto_N704---S517_ON7_2-2	0	2
kallisto_N705---N502_ON16_2-3	2	0
kallisto_N705---N504_ON1_1-2_J2	0	0
kallisto_N705---S517_ON7_2-4	0	2
kallisto_N706---N502_ON16_2-4	2	0
kallisto_N706---N504_ON7_1-2_J2	0	0
kallisto_N706---S517_ON21_1-3	1	1
kallisto_N709---N502_ON22_1-4	1	1
kallisto_N709---N504_ON6_1-1_J2	0	0
kallisto_N709---S507_ON22_1-2_J2	0	0
kallisto_N709---S517_ON24_2-4	0	2
kallisto_N710---N504_ON16_1-2_J2	0	0
kallisto_N710---S502_ON22_2-2	1	1
kallisto_N710---S517_ON17_1-1	0	2
kallisto_N711---N502_ON22_2-4	1	1
kallisto_N711---N504_ON17_1-1_J2	0	0
kallisto_N711---S517_ON17_1-2	0	2
kallisto_N712---N502_ON24_1-2	0	2
kallisto_N712---N504_ON24_1-1_J2	0	0
kallisto_N712---S517_ON17_2-4	0	2

* (0) means that the dataset is unused; (1) means control samples; (2) means treatment samples.

Table S. 5.6 Results of Edger analysis with Peking as treatment samples and Essex as the control samples. Only coding sequences with log (FC) > 3 and p-value < 0.01 are listed.

Hetgly transcript ID	Peking log(FC)	Peking Log(CPM)	Peking edgeR adj. P-value	PI88788 log(FC)	PI88788 Log(CPM)2	PI88788 edgeR adj. P-value	Clusters	SCN coding sequences file protein description (species)
T000015125.1	13,865	5,367	3,50E-55	6,198	-1,412	0,094	1	60S ribosomal protein L10a (Caenorhabditis elegans)
T000017987.1	-12,503	4,858	0,00091	-1,817	5,036	0,68	10	ATP-dependent DNA helicase PIF1 (Danio rerio)
T000004176.1	4,62	1,227	0,00045	4,517	1,724	0,0073	2	NA
T000013532.1	3,988	0,281	5,00E-04	5,097	1,845	0,00026	2	NA
T000013534.1	4,021	2,467	5,00E-04	4,121	3,137	0,0077	2	NA
T000006629.1	4,537	0,409	1,50E-05	3,48	0,05	0,017	2	Histone-lysine N-methyltransferase SUV39H2 (Xenopus tropicalis)
T000023110.1	4,009	4,525	2,40E-08	3,828	4,953	0,00027	2	NA
T000007950.1	4,278	2,146	0,00055	3,436	1,955	0,057	2	Mitochondrial import receptor subunit TOM40 homolog (Caenorhabditis elegans)
T000007915.1	4,082	3,336	9,20E-05	3,092	3,016	0,014	2	Succinyl-CoA ligase [GDP-forming] subunit beta%2C mitochondrial (Caenorhabditis elegans)
T000013613.1	5,157	0,471	4,20E-06	1,792	-1,511	0,54	2	Uncharacterized protein (Plasmodium falciparum (isolate Dd2))
T000005910.1	4,554	2,587	8,10E-09	2,342	1,219	0,17	2	NA
T000021339.1	5,248	3,08	1,20E-14	1,436	0,432	0,46	2	NA
T000023942.1	4,346	1,094	2,80E-06	2,192	-0,124	0,12	2	NA
T000013675.1	5,108	4,954	3,20E-33	1,337	2,282	0,3	2	Pre-mRNA-splicing factor ATP-dependent RNA helicase ddx-15 (Caenorhabditis elegans)
T000015151.1	4,049	4,296	9,20E-16	2,37	3,404	0,0052	2	Protein polybromo-1 (Homo sapiens)
T000005911.1	4,463	3,878	1,20E-10	1,781	2,132	0,25	2	Rap ran-GAP domain containing protein (Haemonchus)
T000011277.1	3,448	4,753	2,80E-08	2,784	4,74	0,00042	2	Probable ATP-dependent RNA helicase DHX35 homolog (Caenorhabditis elegans)
T000005912.1	4,47	1,338	7,30E-08	1,761	-0,299	0,36	2	NA
T000013533.1	3,15	3,229	0,00099	3,055	3,683	0,0064	2	Ubiquitin-protein ligase E3B (Xenopus tropicalis)
T000005954.1	4,184	0,697	5,70E-05	1,742	-0,649	0,46	2	NA
T000002914.1	4,553	4,915	2,60E-07	1,284	2,725	0,32	2	NA
T000010779.1	3,574	4,092	1,20E-05	2,222	3,494	0,077	2	Uncharacterized protein K02A2.6 (Caenorhabditis elegans)
T000022790.1	3,574	4,092	1,20E-05	2,222	3,494	0,077	2	Uncharacterized protein K02A2.6 (Caenorhabditis elegans)
T000007920.1	3,744	5,975	0,00031	2,03	5,057	0,013	2	Rap1 GTPase-activating protein 1 (Homo sapiens)
T000023460.1	3,541	2,778	0,00012	2,201	2,197	0,16	2	NA
T000013535.1	2,951	2,211	4,40E-06	2,78	2,592	0,0016	2	Histone-lysine N-methyltransferase SUV39H2 (Xenopus tropicalis)

T000023934.1	5,084	5,938	3,20E-27	0,611	2,813	0,74	2	NA
T000005914.1	4,586	1,191	0,00077	1,094	-0,939	0,79	2	Disintegrin and metalloproteinase domain-containing protein 10 (<i>Xenopus laevis</i>)
T000022839.1	3,698	7,151	1,20E-08	1,979	6,244	0,038	2	Uncharacterized abhydrolase domain-containing protein DDB_G0269086 (<i>Dictyostelium discoideum</i>)
T000013666.1	4,228	4,181	2,30E-17	1,379	2,366	0,19	2	Ral GTPase-activating protein subunit alpha-1 (<i>Homo sapiens</i>)
T000028929.1	4,535	3,144	4,90E-08	0,942	0,801	0,7	2	Homeobox protein six1b (<i>Danio rerio</i>)
T000013676.1	4,7	4,255	1,40E-23	0,743	1,583	0,77	2	Dorsal gland cell-specific expression protein (<i>Heterodera avenae</i>)
T000005909.1	3,858	2,693	0,00032	1,495	1,288	0,44	2	NA
T000027507.1	2,708	2,177	0,00031	2,63	2,627	0,0042	2	Uncharacterized protein (<i>Caenorhabditis elegans</i>)
T000023461.1	3,63	5,863	0,00011	1,676	4,764	0,2	2	Zinc metalloproteinase nas-36 (<i>Caenorhabditis elegans</i>)
T000011282.1	3,08	3,803	5,30E-06	2,164	3,567	0,043	2	Probable ATP-dependent RNA helicase DHX35 homolog (<i>Caenorhabditis elegans</i>)
T000005967.1	3,982	1,566	3,30E-06	1,228	-0,045	0,49	2	NA
T000013621.1	4,029	2,924	4,80E-10	0,911	1,007	0,61	2	Casein kinase II subunit alpha (<i>Caenorhabditis elegans</i>)
T000021332.1	3,883	0,399	2,20E-05	1,025	-1,115	0,75	2	Transposon Tf2-8 polyprotein (<i>Schizosaccharomyces pombe</i> (strain 972 / ATCC 24843))
T000013673.1	3,323	2,611	6,60E-05	1,455	1,647	0,29	2	Ovarian tumour%2C otubain domain-containing protein (<i>Strongyloides</i>)
T000004177.1	2,552	1,743	0,00088	2,131	1,897	0,027	2	NA
T000004189.1	2,552	1,743	0,00088	2,131	1,897	0,027	2	NA
T000018637.1	3,53	7,932	4,60E-07	1,028	6,468	0,094	2	SNF-related serine/threonine-protein kinase (<i>Homo sapiens</i>)
T000013665.1	3,859	4,518	3,70E-21	0,687	2,562	0,59	2	60S ribosomal protein L21 (<i>Caenorhabditis elegans</i>)
T000019763.1	2,976	4,936	1,30E-05	1,532	4,284	0,2	2	Kynurenine--oxoglutarate transaminase 3 (<i>Homo sapiens</i>)
T000002632.1	3,225	3,45	7,50E-05	1,234	2,391	0,31	2	Uncharacterized protein (<i>Strongyloides papillosus</i>)
T000007250.1	3,225	3,45	7,50E-05	1,234	2,391	0,31	2	NA
T000017636.1	3,225	3,45	7,50E-05	1,234	2,391	0,31	2	Uncharacterized protein (<i>Strongyloides papillosus</i>)
T000002907.1	2,932	4,478	0,00016	1,51	3,852	0,1	2	Dorsal gland cell-specific expression protein (<i>Heterodera avenae</i>)
T000028923.1	2,938	4,542	1,00E-09	1,479	3,89	0,096	2	Similar to Calponin homolog OV9M (<i>Onchocerca volvulus</i>)
T000028244.1	2,723	3,768	9,30E-05	1,551	3,342	0,11	2	NA
T000024467.1	2,673	4,75	2,00E-06	1,575	4,372	0,071	2	Protein disabled (<i>Drosophila melanogaster</i>)
T000016185.1	4,016	4,774	2,30E-10	0,196	2,386	0,95	2	NA
T000004166.1	2,43	2,725	0,00079	1,764	2,679	0,11	2	Putative gland protein G17G01 (<i>Heterodera glycines</i>)
T000027665.1	3,635	0,513	0,00062	0,546	-1,078	0,86	2	NA
T000011278.1	2,493	5,395	3,00E-05	1,654	5,227	0,0052	2	Probable ATP-dependent RNA helicase DHX35 homolog (<i>Caenorhabditis elegans</i>)
T000028373.1	3,163	1,502	0,00027	0,938	0,352	0,59	2	NA
T000024469.1	2,535	6,388	1,30E-09	1,471	6,047	0,0013	2	Dorsal gland cell-specific expression protein (<i>Heterodera avenae</i>)

T000001930.1	2,49	3,645	0,00033	1,426	3,297	0,15	2	Dynein heavy chain%2C cytoplasmic (Caenorhabditis elegans)
T000021792.1	3,171	2,407	7,30E-05	0,714	1,081	0,81	2	NA
T000024471.1	2,364	6,375	4,60E-06	1,466	6,162	0,02	2	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000010224.1	3,043	3,429	0,00025	0,759	2,217	0,7	2	RRP15-like protein (Caenorhabditis briggsae)
T000013674.1	3,101	4,478	8,10E-12	0,657	3,143	0,58	2	OTU domain-containing protein 7A (Homo sapiens)
T000027660.1	3,389	4,025	9,20E-16	0,357	2,291	0,84	2	NA
T000000480.1	4,201	4,25	1,40E-05	-0,555	1,392	0,7	2	Cathepsin 7 (Mus musculus)
T000025571.1	2,415	4,651	1,60E-05	1,031	4,088	0,29	2	NA
T000026670.1	2,312	6,197	1,00E-05	1,127	5,776	0,11	2	NA
T000025299.1	2,222	3,041	0,00033	1,205	2,752	0,21	2	DC-STAMP domain-containing protein 2 (Homo sapiens)
T000027211.1	2,589	3,478	0,00018	0,765	2,618	0,64	2	RING-H2 finger protein ATL80 (Arabidopsis thaliana)
T000028385.1	2,192	4,306	2,00E-05	1,136	3,982	0,26	2	BTB/POZ domain-containing protein 2 (Homo sapiens)
T000001439.1	2,499	2,575	0,00023	0,822	1,833	0,52	2	NA
T000016963.1	1,964	3,967	0,00031	1,228	3,862	0,081	2	Uncharacterized protein (Fragment) (Myotis brandtii)
T000020732.1	2,602	4,284	5,00E-06	0,538	3,282	0,72	2	NA
T000024721.1	2,897	2,428	0,00029	0,243	1,051	0,94	2	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000024723.1	2,897	2,428	0,00029	0,243	1,051	0,94	2	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000012475.1	2,695	5,576	0,00027	0,421	4,429	0,62	2	Phospholipid phosphatase 2 (Bos taurus)
T000016165.1	2,499	3,674	3,00E-05	0,609	2,784	0,68	2	NA
T000016940.1	2,262	3,819	8,30E-06	0,787	3,219	0,37	2	NA
T000023936.1	2,149	4,087	4,10E-05	0,895	3,642	0,38	2	Reticulocyte-binding protein 2 homolog a (Plasmodium falciparum (isolate 3D7))
T000013672.1	1,843	4,259	0,00018	1,193	4,219	0,21	2	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000021834.1	2,188	5,426	9,40E-05	0,733	4,84	0,44	2	NA
T000013662.1	2,413	6,29	3,90E-07	0,475	5,381	0,61	2	Disintegrin and metalloproteinase domain-containing protein 10 (Mus musculus)
T000023949.1	2,645	2,81	3,30E-05	0,237	1,62	0,91	2	NA
T000023940.1	2,645	2,81	3,30E-05	0,237	1,62	0,91	2	NA
T000029517.1	2,431	3,464	2,00E-04	0,445	2,543	0,78	2	Aldose reductase (Mus musculus)
T000000221.1	2,396	3,13	0,00066	0,468	2,246	0,78	2	Protein MoeA (Enterobacter aerogenes)
T000000228.1	2,396	3,13	0,00066	0,468	2,246	0,78	2	Protein MoeA (Enterobacter aerogenes)
T000000236.1	2,396	3,13	0,00066	0,468	2,246	0,78	2	Protein MoeA (Enterobacter aerogenes)
T000028170.1	3,128	4,623	3,30E-05	-0,382	2,729	0,89	2	NA
T000009757.1	2,311	3,28	0,00015	0,393	2,412	0,83	2	NA
T000015131.1	2,22	3,944	1,10E-05	0,288	3,076	0,86	2	NA
T000015130.1	1,881	3,638	0,00078	0,547	3,154	0,58	2	Probable tyrosine-protein kinase kin-31 (Caenorhabditis elegans)

T000023941.1	1,755	6,446	0,00022	0,666	6,115	0,36	2	NA
T000013654.1	2,317	2,168	0,00062	0,06	1,145	0,98	2	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000024466.1	1,569	5,918	0,00035	0,769	5,767	0,24	2	Dual specificity protein phosphatase 16 (Homo sapiens)
T000015128.1	2,358	2,761	0,00015	-0,056	1,638	0,98	2	NA
T000006002.1	2,067	4,202	0,00012	0,126	3,363	0,94	2	Cre-nsy-4 protein (Ascaris)
T000008129.1	1,743	5,354	0,00079	0,387	4,859	0,74	2	Uncharacterized protein (Acyrtosiphon pisum)
T000013284.1	1,915	5,431	6,50E-07	0,193	4,722	0,87	2	NA
T000013637.1	2,157	3,05	0,00015	-0,096	2,043	0,97	2	Ovarian tumour%2C otubain domain-containing protein (Strongyloides)
T000018385.1	1,578	6,692	0,00067	0,244	6,226	0,81	2	MICOS complex subunit Mic60 (Fragment) (Rattus norvegicus)
T000011651.1	2,142	4,621	4,90E-05	-0,321	3,506	0,82	2	Zinc metalloproteinase nas-5 (Caenorhabditis elegans)
T000013607.1	1,713	4,363	4,00E-04	-0,151	3,621	0,92	2	Cholinesterase (Felis catus)
T000005960.2	10,452	1,983	7,30E-19	6,053	-1,45	0,15	3	Similar to Calponin homolog OV9M (Onchocerca volvulus)
T000025025.1	9,348	0,94	2,00E-09	6,126	-1,366	0,15	3	NA
T000013611.1	9,621	1,202	1,70E-10	5,635	-1,728	0,11	3	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000014113.1	9,46	1,025	2,80E-07	5,153	-2,007	0,2	3	Unconventional myosin heavy chain 6 (Caenorhabditis elegans)
T000015124.1	9,078	1,472	9,80E-10	5,075	-1,475	0,15	3	NA
T000023930.1	8,516	1,542	2,00E-15	4,841	-1,062	0,12	3	NA
T000008007.1	9,313	2,711	8,30E-09	3,982	-1,367	0,33	3	NA
T000013670.1	9,558	5,486	3,70E-32	3,296	0,159	0,31	3	60S ribosomal protein L21 (Caenorhabditis elegans)
T000023929.1	8,546	4,746	1,20E-11	3,179	0,289	0,44	3	NA
T000021369.2	8,405	2,133	5,00E-06	3,128	-1,721	0,47	3	NA
T000013616.1	7,8	1,228	1,10E-12	2,822	-2,054	0,46	3	NA
T000015145.1	7,966	-0,333	3,00E-06	6,552	-1,108	0,02	4	NA
T000019314.1	7,248	-0,876	1,00E-05	7,221	-0,564	0,00094	4	NA
T000018076.1	7,08	-1,094	0,00042	7,041	-0,754	0,038	4	Uncharacterized protein (Pristionchus pacificus)
T000016955.2	7,885	1,563	2,70E-13	5,609	0,016	0,026	4	NA
T000010778.1	7,418	3,102	7,50E-05	5,832	2,193	0,0011	4	NA
T000020627.1	7,479	1,616	1,30E-06	5,506	0,34	0,091	4	Serine/threonine-protein phosphatase PP1 (Zea mays)
T000028924.1	7,282	-0,859	6,10E-06	5,61	-1,77	0,065	4	NA
T000007918.1	7,353	1,879	7,90E-08	5,474	0,786	0,13	4	Inhibitor of growth protein 3 (Homo sapiens)
T000002058.1	6,564	3,597	2,00E-30	5,954	3,646	4,70E-11	4	NA
T000028082.1	5,634	-1,465	0,00028	6,813	-0,131	0,016	4	Regulator of nonsense transcripts 1 homolog (Arabidopsis thaliana)
T000004153.1	6,728	2,5	4,20E-21	5,495	1,978	1,60E-05	4	NA
T000019765.3	6,865	2,945	1,00E-05	5,141	1,892	0,029	4	TBC1 domain family member 5 (Mus musculus)
T000015137.1	7,046	4,002	1,10E-17	4,783	2,509	0,057	4	Protein polybromo-1 (Homo sapiens)
T000023469.2	6,565	2,741	0,00015	5,255	2,137	0,014	4	Fragile X mental retardation syndrome-related protein 1

								homolog (<i>Xenopus tropicalis</i>)
T000007713.2	6,565	2,741	0,00015	5,255	2,137	0,014	4	Fragile X mental retardation syndrome-related protein 1 homolog (<i>Xenopus tropicalis</i>)
T000006631.1	6,681	1,926	4,00E-09	5,135	1,045	0,027	4	NA
T000021335.1	6,525	-0,005	3,60E-09	4,695	-1,027	0,026	4	Uncharacterized protein K02A2.6 (<i>Caenorhabditis elegans</i>)
T000015135.1	6,656	3,542	4,10E-17	4,455	2,051	0,014	4	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13 (<i>Bos taurus</i>)
T000028844.1	6,955	2,682	6,60E-05	4,016	0,556	0,26	4	NA
T000021331.1	7,081	-0,303	2,60E-07	3,871	-2,146	0,23	4	Integrase core domain protein (<i>Ancylostoma duodenale</i>)
T000015764.2	6,49	0,229	0,00013	4,451	-0,914	0,31	4	Zinc finger BED domain-containing protein 4 (<i>Mus musculus</i>)
T000023931.1	6,58	0,013	2,90E-09	4,242	-1,363	0,09	4	NA
T000013516.1	4,404	0,611	0,00018	6,406	3,103	3,00E-05	4	Methyltransferase-like protein 13 (<i>Mus musculus</i>)
T000006633.3	6,572	3,536	0,00017	4,161	1,881	0,096	4	Cathepsin B (<i>Sus scrofa</i>)
T000028246.1	5,846	0,215	8,50E-08	4,648	-0,262	0,015	4	NA
T000004194.1	5,675	0,126	2,00E-07	4,816	-0,077	0,016	4	NA
T000004175.1	5,675	0,126	2,00E-07	4,816	-0,077	0,016	4	NA
T000004191.1	5,675	0,126	2,00E-07	4,816	-0,077	0,016	4	NA
T000023123.1	5,511	4,143	6,90E-10	4,908	4,208	4,20E-06	4	Putative cuticle collagen 145 (<i>Caenorhabditis briggsae</i>)
T000015133.1	6,573	3,279	5,80E-11	3,799	1,27	0,17	4	NA
T000027206.1	5,703	1,879	0,00011	4,358	1,231	0,014	4	Ecdysone receptor (<i>Lucilia cuprina</i>)
T000019315.1	5,744	-0,153	1,30E-06	4,293	-0,807	0,036	4	NA
T000015165.1	6,398	0,362	3,40E-06	3,623	-1,422	0,34	4	NA
T000005973.1	6,827	1,013	9,30E-12	3,127	-1,474	0,27	4	NA
T000015142.1	6,176	0,055	1,40E-06	3,772	-1,267	0,075	4	NA
T000015153.1	6,176	0,055	1,40E-06	3,772	-1,267	0,075	4	NA
T000013680.1	7,069	2,81	1,20E-19	2,874	-0,407	0,32	4	Organic solute transporter alpha-like protein 1 (<i>Caenorhabditis elegans</i>)
T000013620.1	7,184	3,249	4,10E-26	2,623	-0,241	0,29	4	OTU domain-containing protein 7A (<i>Homo sapiens</i>)
T000028927.1	6,35	2,99	7,30E-20	3,137	0,628	0,081	4	NA
T000025390.1	6,75	1,567	1,00E-05	2,582	-1,31	0,62	4	NA
T000013667.1	6,81	3,973	4,20E-10	2,496	0,618	0,52	4	Leukocyte receptor cluster member 1 (<i>Homo sapiens</i>)
T000004152.1	5,769	0,36	1,20E-06	3,478	-0,988	0,094	4	NA
T000006004.1	6,098	-0,295	2,20E-06	3,049	-1,961	0,55	4	NA
T000023932.1	5,477	3,434	3,00E-22	3,623	2,335	0,014	4	Sentrin-specific protease 1 (<i>Mus musculus</i>)
T000012635.1	5,095	-0,173	0,00038	3,527	-0,97	0,15	4	NA
T000013658.1	6,378	3,346	1,30E-18	2,114	0,211	0,53	4	Lon protease homolog%2C mitochondrial (<i>Zea mays</i>)
T000006001.1	6,119	2,369	4,20E-06	2,363	-0,378	0,61	4	Probable E3 ubiquitin-protein ligase XBOS34 (<i>Oryza sativa</i> subsp. <i>japonica</i>)
T000021348.1	6,325	1,272	7,30E-09	1,812	-1,736	0,62	4	NA
T000004172.1	5,371	2,703	0,00026	2,696	0,904	0,33	4	Uncharacterized protein (<i>Pristionchus pacificus</i>)
T000027863.1	5,348	2,227	1,30E-06	2,676	0,426	0,34	4	NA

T000011259.1	9,552	1,12	8,50E-07	10,325	2,276	3,70E-09	5	NA
T000015141.1	11,48	2,993	1,70E-21	8,215	0,309	0,0064	5	Vesicle-associated membrane protein-associated protein B (Rattus norvegicus)
T000015092.1	9,264	0,876	2,70E-06	9,557	1,523	0,0069	5	Trafficking protein particle complex subunit 12 (Homo sapiens)
T000007052.1	10,74	2,273	1,10E-21	7,853	-0,026	0,00051	5	Putative apoptosis inhibitor 021L (Invertebrate iridescent virus 3)
T000008005.1	9,138	2,769	9,60E-16	8,728	2,94	0,00069	5	Transcription factor Adf-1 (Drosophila melanogaster)
T000011075.1	10,644	2,181	8,80E-14	7,156	-0,654	0,067	5	Ribonuclease UK114 (Capra hircus)
T000008246.1	9,085	1,492	1,10E-06	8,484	1,376	0,059	5	NA
T000025738.1	9,693	1,263	5,70E-16	7,819	-0,071	0,004	5	NA
T000007952.1	8,891	0,478	8,90E-08	8,314	0,357	0,0089	5	Uncharacterized protein (Ancylostoma ceylanicum)
T000026005.1	9,55	1,103	2,00E-07	7,552	-0,324	0,01	5	NA
T000004154.1	8,309	-0,009	1,70E-05	8,393	0,44	0,00084	5	Putative histone H1.6 (Caenorhabditis elegans)
T000006633.2	8,383	0,003	3,30E-05	7,862	-0,043	0,012	5	Cathepsin B (Sus scrofa)
T000025300.1	9,055	0,594	0,00028	6,927	-0,814	0,015	5	Histone-lysine N-methyltransferase SUV39H2 (Xenopus tropicalis)
T000019292.1	8,085	2,111	1,10E-16	7,671	2,443	0,0011	5	NA
T000017414.1	-2,476	4,028	0,00029	1,466	5,614	0,04	6	Foot protein 1 variant 1 (Perna viridis)
T000002859.1	-3,498	4,275	0,00011	2,314	6,596	0,05	6	NA
T000006544.1	-1,694	4,779	0,00031	0,205	5,476	0,89	6	Transketolase-like protein 2 (Bos taurus)
T000009196.1	-1,837	5,608	0,00021	0,321	6,389	0,71	6	Retrovirus-related Pol polyprotein from transposon 17.6 (Drosophila melanogaster)
T000027656.1	-1,957	5,541	0,00029	0,211	6,274	0,92	6	Uncharacterized protein (Thelohanelus kitauei)
T000006572.1	-1,619	5,161	0,00054	-0,359	5,569	0,69	6	NA
T000025892.1	-2,056	5,536	0,00013	0,049	6,2	0,98	6	Acyl-CoA synthetase family member 4 (Mus musculus)
T000029269.1	-2,605	3,984	2,80E-05	0,585	4,997	0,42	6	NA
T000009834.1	-2,156	4,478	0,00027	-0,272	4,999	0,86	6	Digestive organ expansion factor homolog (Mus musculus)
T000025904.1	-2,3	4,901	7,70E-06	-0,146	5,497	0,89	6	Transposase (Meloidogyne javanica)
T000008027.1	-1,987	4,889	0,00047	-0,49	5,298	0,59	6	Succinyl-CoA ligase [GDP-forming] subunit beta%2C mitochondrial (Caenorhabditis elegans)
T000028937.1	-2,454	5,242	4,10E-06	-0,088	5,876	0,95	6	NA
T000020514.1	-2,277	4,426	4,80E-06	-0,272	4,956	0,83	6	NA
T000002888.1	-2,387	4,449	0,00097	-0,168	5,045	0,92	6	NA
T000017115.1	-2,019	3,955	0,00044	-0,574	4,33	0,5	6	Ran-binding protein 9 (Xenopus laevis)
T000003462.1	-2,016	4,322	0,00054	-0,578	4,697	0,5	6	NA
T000000068.1	-2,761	4,313	0,00048	0,141	5,092	0,93	6	DNA-directed RNA polymerase II subunit RPB1 (Caenorhabditis briggsae)
T000028831.1	-2,42	2,547	0,00076	-0,255	3,093	0,85	6	Ran-binding protein 10 (Mus musculus)
T000029717.1	-1,964	6,484	0,00054	-0,727	6,794	0,28	6	Transposon Ty3-I Gag-Pol polyprotein (Saccharomyces cerevisiae (strain ATCC 204508 / S288c))

T000027016.1	-2,551	5,727	9,70E-05	-0,155	6,339	0,92	6	BTB/POZ domain-containing protein 2 (Homo sapiens)
T000002908.1	-2,635	4,921	4,40E-05	-0,156	5,54	0,92	6	NA
T000017105.1	-2,55	5,497	5,70E-06	-0,273	6,054	0,82	6	Protein bark beetle (<i>Drosophila melanogaster</i>)
T000008025.1	-2,588	4,772	0,00091	-0,269	5,337	0,86	6	Probable Werner syndrome ATP-dependent helicase homolog 1 (<i>Caenorhabditis elegans</i>)
T000025305.1	-2,037	5,005	5,40E-05	-0,821	5,29	0,5	6	Histone-lysine N-methyltransferase SUV39H2 (Homo sapiens)
T000008552.1	-2,973	3,987	0,00038	0,097	4,757	0,97	6	NA
T000021240.1	-2,664	5,765	1,20E-05	-0,265	6,337	0,86	6	Protein bark beetle (<i>Drosophila melanogaster</i>)
T000029263.1	-3,544	2,134	0,00012	0,6	3,196	0,64	6	NA
T000017109.1	-2,751	4,47	4,80E-06	-0,194	5,079	0,89	6	Ran-binding proteins 9/10 homolog (<i>Drosophila melanogaster</i>)
T000027657.1	-2,808	3,802	0,00014	-0,145	4,437	0,94	6	Ankyrin repeat domain-containing protein 17 (Homo sapiens)
T000023337.1	-3,371	3,897	0,00065	0,416	4,862	0,75	6	Uncharacterized ATP-dependent helicase C05C10.2 (<i>Caenorhabditis elegans</i>)
T000028316.2	-3,408	2,843	9,90E-05	0,436	3,817	0,71	6	BTB/POZ domain-containing protein 6 (Homo sapiens)
T000021237.1	-2,852	2,838	0,00028	-0,125	3,481	0,94	6	Dorsal gland cell-specific expression protein (<i>Heterodera avenae</i>)
T000005877.1	-3,552	1,564	0,00028	0,564	2,592	0,59	6	Similar to Histone H4 (<i>Ascaris suum</i>)
T000023613.1	-2,417	4,036	0,00096	-0,575	4,456	0,57	6	NA
T000028314.1	-3,157	4,221	4,00E-07	0,154	5,031	0,89	6	Retrovirus-related Pol polyprotein from transposon 297 (<i>Drosophila melanogaster</i>)
T000017099.1	-2,532	5,124	0,00058	-0,487	5,59	0,67	6	NA
T000025128.1	-2,727	4,773	1,80E-05	-0,293	5,336	0,82	6	Uncharacterized protein (<i>Thelohanellus kitauei</i>)
T000017100.1	-2,762	3,775	8,00E-05	-0,273	4,35	0,86	6	Uncharacterized protein F23F12.3 (<i>Caenorhabditis elegans</i>)
T000028490.1	-3,079	5,002	0,00041	0,013	5,733	1	6	NA
T000025127.1	-4,27	3,483	0,00085	1,186	4,955	0,48	6	Anaphase-promoting complex subunit 11 (<i>Caenorhabditis elegans</i>)
T000021030.1	-3,375	2,183	0,00025	0,274	3,053	0,86	6	Uncharacterized protein (<i>Haemonchus placei</i>)
T000025031.1	-3,212	5,215	1,40E-05	0,077	5,988	0,97	6	Ecdysone-induced protein 74EF (<i>Drosophila virilis</i>)
T000017076.1	-2,66	6,001	1,90E-05	-0,483	6,477	0,57	6	NA
T000025891.1	-2,786	5,773	7,10E-05	-0,364	6,311	0,69	6	NA
T000017113.1	-2,865	4,709	6,80E-05	-0,287	5,283	0,84	6	Ankyrin repeat domain-containing protein 17 (<i>Mus musculus</i>)
T000028832.1	-2,993	2,963	0,00035	-0,176	3,594	0,92	6	NA
T000008014.1	-2,445	5,054	0,00056	-0,733	5,412	0,35	6	Oligoribonuclease (<i>Bordetella pertussis</i> (strain Tohama I / ATCC BAA-589 / NCTC 13251))
T000013708.1	-3,032	3,537	5,20E-05	-0,157	4,181	0,94	6	Foot protein 1 variant 1 (<i>Perna viridis</i>)
T000007932.1	-2,594	5,458	0,00076	-0,608	5,885	0,44	6	Triple functional domain protein (<i>Danio rerio</i>)
T000017083.1	-3,274	2,942	0,00013	0,057	3,702	0,98	6	NA
T000017634.1	-2,853	4,961	9,10E-06	-0,406	5,484	0,84	6	Protein disabled (<i>Drosophila melanogaster</i>)
T000010267.1	-3,098	5,153	5,30E-13	-0,175	5,796	0,86	6	AMP deaminase 2 (Homo sapiens)
T000020866.1	-2,354	4,691	0,00074	-0,926	4,97	0,2	6	Probable DNA replication complex GINS protein PSF2 (<i>Brugia</i>)

								malayi)
T000003281.1	-2,305	4,499	1,50E-05	-1,001	4,747	0,11	6	Innexin-3 (Caenorhabditis elegans)
T000017092.1	-2,732	4,11	0,00025	-0,581	4,549	0,64	6	NA
T000019782.1	-2,696	6,543	0,00014	-0,622	6,968	0,47	6	NA
T000028634.1	-2,967	4,325	2,10E-07	-0,404	4,856	0,75	6	Dual specificity protein kinase splA (Dictyostelium discoideum)
T000019775.1	-2,343	4,771	0,00067	-1,056	5,009	0,18	6	MAU2 chromatid cohesion factor homolog (Xenopus tropicalis)
T000001318.1	-3,239	4,849	2,10E-08	-0,171	5,505	0,91	6	NA
T000007987.1	-3,538	3,89	8,60E-05	0,117	4,701	0,95	6	39S ribosomal protein L41%2C mitochondrial (Caenorhabditis elegans)
T000026827.1	-2,867	5,04	1,00E-05	-0,562	5,501	0,51	6	NA
T000022625.1	-2,863	4,713	3,40E-06	-0,606	5,153	0,48	6	NA
T000007970.1	-2,432	5,6	4,20E-05	-1,038	5,852	0,31	6	Neuronal acetylcholine receptor subunit alpha-9 (Gallus gallus)
T000021033.1	-3,579	1,283	0,00026	0,109	2,063	0,96	6	NA
T000021034.1	-3,579	1,283	0,00026	0,109	2,063	0,96	6	NA
T000026398.1	-3,293	6,299	0,00022	-0,226	6,933	0,89	6	Pao retrotransposon peptidase family protein (Brugia malayi)
T000029609.1	-3,462	3,909	2,90E-09	-0,066	4,621	0,98	6	Homeobox protein ceh-32 (Caenorhabditis elegans)
T000021417.1	-2,83	3,477	0,00021	-0,718	3,874	0,48	6	NA
T000028946.1	-3,096	3,978	9,20E-05	-0,513	4,471	0,75	6	Ran-binding protein 10 (Danio rerio)
T000025627.1	-3,422	1,893	0,00043	-0,195	2,531	0,92	6	NA
T000028821.1	-2,646	4,148	0,00011	-0,987	4,431	0,27	6	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000017822.1	-3,062	4,38	0,00054	-0,591	4,85	0,76	6	NA
T000013090.1	-3,677	4,743	5,50E-07	-0,005	5,499	1	6	Ecdysone-induced protein 74EF (Drosophila virilis)
T000017082.1	-3,67	3,588	1,30E-05	-0,012	4,331	1	6	Endonuclease III homolog (Caenorhabditis elegans)
T000021231.1	-3,95	4,724	1,30E-11	0,207	5,598	0,9	6	NA
T000025628.1	-3,71	1,278	0,00041	-0,055	1,982	0,99	6	Putative esophageal gland cell protein Hgg-20 (Fragment) (Heterodera glycines)
T000008547.1	-3,382	3,864	3,90E-05	-0,433	4,402	0,75	6	NA
T000013084.1	-3,478	3,107	3,90E-05	-0,346	3,688	0,83	6	NA
T000016147.1	-3,33	3,905	2,10E-05	-0,51	4,414	0,66	6	NA
T000008070.2	-3,564	3,624	0,00025	-0,277	4,241	0,88	6	Solute carrier family 35 member E1 homolog (Drosophila melanogaster)
T000028827.1	-4,535	0,879	0,00011	0,649	1,977	0,7	6	Uncharacterized protein (Oesophagostomum dentatum)
T000012880.1	-2,668	3,981	1,00E-04	-1,243	4,185	0,16	6	NA
T000009095.1	-3,149	4,964	0,00021	-0,77	5,362	0,45	6	NA
T000028386.1	-3,884	3,246	3,60E-05	-0,059	3,975	0,98	6	Heterochromatin-associated protein MENT (Gallus gallus)
T000005955.1	-4,473	2,636	5,20E-07	0,526	3,689	0,57	6	NA
T000008015.1	-3,506	3,788	0,00082	-0,447	4,344	0,82	6	NA

T000012893.1	-2,924	5,136	9,20E-06	-1,091	5,417	0,33	6	Putative uncharacterized transposon-derived protein F54H12.3 (<i>Caenorhabditis elegans</i>)
T000019690.1	-3,611	2,87	0,00019	-0,446	3,408	0,79	6	NA
T000011376.1	-4,102	4,202	8,00E-04	0,026	4,984	1	6	Serine/threonine-protein phosphatase PP1 (<i>Zea mays</i>)
T000006611.1	-3,525	5,848	6,90E-06	-0,553	6,349	0,61	6	Acylamino-acid-releasing enzyme (<i>Mus musculus</i>)
T000022468.1	-4,055	2,399	0,00045	-0,028	3,141	1	6	NA
T000017114.1	-4,042	3,205	1,60E-05	-0,062	3,942	0,98	6	NA
T000006571.1	-4,198	2,093	3,30E-05	0,089	2,898	0,98	6	ABC transporter ATP-binding protein/permease wht-3 (<i>Caenorhabditis elegans</i>)
T000018803.1	-3,934	3,709	0,00018	-0,316	4,327	0,92	6	Glc(2)Man(9)GlcNAc(2)-PP-Dol alpha-1%2C2-glucosyltransferase (<i>Homo sapiens</i>)
T000027645.1	-4,017	6,896	4,90E-11	-0,241	7,551	0,89	6	Ran-binding protein 4 (<i>Heterodera glycines</i>)
T000028825.1	-3,442	5,953	1,50E-05	-0,826	6,345	0,48	6	Ran-binding proteins 9/10 homolog (<i>Drosophila melanogaster</i>)
T000009117.1	-3,754	2,282	0,00054	-0,526	2,796	0,75	6	NA
T000017147.1	-3,641	7,342	2,70E-06	-0,661	7,809	0,72	6	Uncharacterized protein (Fragment) (<i>Picoides pubescens</i>)
T000021581.1	-4,001	2,451	4,40E-06	-0,304	3,067	0,85	6	Eukaryotic translation initiation factor 3 subunit H (<i>Brugia malayi</i>)
T000014910.1	-3,605	3,624	3,10E-05	-0,719	4,062	0,52	6	NA
T000021325.1	-4,173	3,659	0,00046	-0,168	4,349	0,88	6	NA
T000021215.1	-4,169	3,728	1,10E-06	-0,182	4,413	0,93	6	Dual specificity protein kinase splA (<i>Dictyostelium discoideum</i>)
T000017071.1	-3,943	4,502	0,00055	-0,408	5,075	0,83	6	NA
T000010279.1	-4,303	5,047	1,60E-19	-0,067	5,791	0,97	6	Ankyrin repeat domain-containing protein 49 (<i>Mus musculus</i>)
T000027322.1	-3,782	2,73	0,00092	-0,588	3,218	0,71	6	NA
T000018989.1	-4,225	2,612	0,00025	-0,164	3,299	0,95	6	NA
T000021236.1	-4,564	3,376	3,20E-08	0,136	4,227	0,95	6	Dual specificity protein kinase splA (<i>Dictyostelium discoideum</i>)
T000020892.1	-4,853	2,744	8,70E-06	0,425	3,758	0,77	6	NA
T000027334.1	-3,6	3,363	0,00067	-0,834	3,756	0,55	6	NA
T000027333.1	-3,6	3,363	0,00067	-0,834	3,756	0,55	6	NA
T000020506.1	-5,481	2,774	2,50E-05	1,029	4,173	0,59	6	NA
T000008033.1	-3,427	6,243	3,60E-06	-1,038	6,574	0,23	6	Rap1 GTPase-activating protein 1 (<i>Homo sapiens</i>)
T000026178.1	-4,255	3,117	4,30E-05	-0,211	3,797	0,89	6	NA
T000016274.1	-4,16	2,937	0,00074	-0,33	3,548	0,89	6	Uncharacterized protein (<i>Anisakis simplex</i>)
T000027800.1	-4,16	2,937	0,00074	-0,33	3,548	0,89	6	Uncharacterized protein (<i>Anisakis simplex</i>)
T000008030.1	-3,716	4,653	1,30E-05	-0,79	5,07	0,2	6	Inhibitor of growth protein 3 (<i>Homo sapiens</i>)
T000000866.1	-3,946	3,723	1,00E-04	-0,583	4,224	0,59	6	NA
T000028631.1	-4,695	2,974	3,50E-07	0,137	3,822	0,96	6	Dual specificity protein kinase splA (<i>Dictyostelium discoideum</i>)
T000017090.1	-4,223	2,125	1,00E-05	-0,336	2,733	0,85	6	NA

T000025636.1	-3,782	2,151	0,00016	-0,797	2,55	0,57	6	NA
T000026830.2	-4,26	2,559	2,00E-04	-0,33	3,177	0,87	6	NA
T000017635.1	-3,854	5,064	1,20E-06	-0,753	5,503	0,62	6	Dual specificity protein phosphatase 16 (Homo sapiens)
T000021042.1	-5,449	4,724	7,30E-12	0,813	5,984	0,13	6	Similar to Calponin homolog OV9M (Onchocerca volvulus)
T000021220.1	-4,202	3,581	1,60E-07	-0,465	4,14	0,75	6	NA
T000003128.1	-5,542	1,321	0,00041	0,87	2,575	0,58	6	Transposase (Meloiodogyne javanica)
T000016498.1	-4,808	5,746	0,00016	0,132	6,604	0,96	6	Uncharacterized protein (Toxocara canis)
T000028826.1	-5,075	0,733	0,00065	0,398	1,681	0,86	6	Uncharacterized protein (Oesophagostomum dentatum)
T000017061.1	-4,481	3,359	6,50E-07	-0,245	4,02	0,88	6	Uncharacterized protein F23F12.3 (Caenorhabditis elegans)
T000024292.1	-3,891	3,2	0,00088	-0,854	3,597	0,47	6	Ran-binding protein 9 (Homo sapiens)
T000018654.1	-3,781	4,504	8,50E-07	-1,001	4,846	0,42	6	NA
T000020886.1	-4,858	5,05	1,50E-15	0,071	5,875	0,96	6	Protein polybromo-1 (Homo sapiens)
T000004725.1	-4,887	1,716	0,00013	0,091	2,529	0,98	6	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000027329.1	-4,832	2,193	1,60E-05	0,015	2,979	1	6	Ran-binding proteins 9/10 homolog (Drosophila melanogaster)
T000013078.1	-4,79	0,947	5,30E-05	-0,044	1,675	0,99	6	BTB/POZ domain-containing protein 2 (Homo sapiens)
T000025306.1	-3,417	3,491	9,20E-05	-1,473	3,676	0,35	6	DC-STAMP domain-containing protein 2 (Homo sapiens)
T000027006.1	-3,865	4,995	5,00E-07	-1,038	5,331	0,32	6	Uncharacterized protein (Strongylocentrotus purpuratus)
T000005880.2	-5,545	3,948	6,20E-15	0,555	5,049	0,5	6	Similar to Histone H4 (Ascaris suum)
T000020156.1	-4,421	3,738	1,40E-05	-0,597	4,248	0,77	6	NA
T000029262.1	-4,788	2,943	2,50E-07	-0,234	3,619	0,91	6	14-3-3 protein beta/alpha-A (Xenopus laevis)
T000007971.1	-3,531	3,203	0,00049	-1,497	3,394	0,25	6	tRNA dimethylallyltransferase (Lactococcus lactis subsp. lactis (strain IL1403))
T000007975.1	-3,531	3,203	0,00049	-1,497	3,394	0,25	6	tRNA dimethylallyltransferase (Lactococcus lactis subsp. lactis (strain IL1403))
T000011366.1	-5,097	3,836	1,70E-08	0,045	4,655	0,99	6	NA
T000025040.1	-4,841	3,416	6,80E-08	-0,26	4,085	0,89	6	Uncharacterized protein C1259.12c (Schizosaccharomyces pombe (strain 972 / ATCC 24843))
T000020891.1	-5,065	3,601	9,40E-12	-0,074	4,357	0,97	6	NA
T000022335.1	-4,791	1,784	0,00092	-0,35	2,393	0,89	6	Uncharacterized protein C1259.12c (Schizosaccharomyces pombe (strain 972 / ATCC 24843))
T000021413.1	-3,862	4,622	0,00011	-1,319	4,878	0,34	6	Major myo-inositol transporter IolT (Bacillus subtilis (strain 168))
T000025023.1	-4,822	1,504	0,00041	-0,377	2,102	0,85	6	Uncharacterized protein C1259.12c (Schizosaccharomyces pombe (strain 972 / ATCC 24843))
T000027866.1	-5,204	1,209	1,20E-05	-0,004	1,976	1	6	NA
T000024262.1	-3,332	2,951	3,00E-04	-1,889	3,036	0,024	6	Transposon Tf2-8 polyprotein (Schizosaccharomyces pombe (strain 972 / ATCC 24843))
T000005883.1	-5,335	3,653	2,70E-11	0,113	4,501	0,94	6	Disintegrin and metalloproteinase domain-containing protein 10 (Bos taurus)

T00000865.1	-4,64	2,476	0,00048	-0,586	2,988	0,68	6	NA
T000011027.1	-5,251	2,272	0,00021	0,002	3,052	1	6	Uncharacterized protein K02A2.6 (<i>Caenorhabditis elegans</i>)
T000017825.1	-4,082	4,158	0,00012	-1,182	4,466	0,42	6	NA
T000026824.1	-4,51	3,383	6,90E-07	-0,771	3,832	0,47	6	NA
T000003022.1	-4,436	2,723	0,00066	-0,86	3,128	0,51	6	NA
T000019209.1	-3,935	3,354	0,00046	-1,376	3,595	0,26	6	Uncharacterized protein (<i>Oryza sativa</i> subsp. <i>japonica</i>)
T000012905.1	-3,491	5,338	3,00E-05	-1,859	5,44	0,03	6	Glutathione synthetase (<i>Xenopus laevis</i>)
T000000861.1	-4,601	3,56	5,00E-07	-0,766	4,017	0,58	6	NA
T000028635.1	-5,033	1,792	0,00067	-0,337	2,408	0,91	6	Uncharacterized protein K02A2.6 (<i>Caenorhabditis elegans</i>)
T000021222.1	-4,475	2,192	0,00029	-0,902	2,579	0,5	6	NA
T000024479.1	-3,774	3,587	1,20E-05	-1,61	3,75	0,074	6	NA
T000005882.1	-5,514	1,419	3,20E-06	0,122	2,251	0,96	6	Organic solute transporter alpha-like protein 1 (<i>Caenorhabditis elegans</i>)
T000012881.1	-6,014	0,046	0,00021	0,613	1,068	0,87	6	Putative 3OC02 parasitic stage effector (<i>Heterodera</i>)
T000024310.1	-7,022	2,207	0,00041	1,612	4,001	0,32	6	Uncharacterized protein (<i>Lottia gigantea</i>)
T000013689.1	-5,476	4,052	2,70E-11	0,042	4,874	0,98	6	Uncharacterized protein (<i>Nippostrongylus brasiliensis</i>)
T000022339.1	-5,439	1,986	0,00019	-0,028	2,761	1	6	NA
T000029219.1	-4,108	3,373	1,10E-07	-1,369	3,622	0,16	6	449)
T000021031.1	-5,658	3,421	8,30E-14	0,171	4,308	0,9	6	NA
T000020896.1	-4,996	3,161	0,00037	-0,492	3,724	0,88	6	NA
T000025038.2	-5,398	2,207	1,10E-06	-0,095	2,946	0,97	6	NA
T000017428.1	-3,724	4,415	9,20E-05	-1,789	4,545	0,11	6	BTB/POZ domain-containing protein 2 (<i>Homo sapiens</i>)
T000005875.1	-5,482	3,717	9,50E-15	-0,038	4,495	0,98	6	Organic solute transporter alpha-like protein 1 (<i>Caenorhabditis elegans</i>)
T000028630.1	-5,405	3,068	2,00E-05	-0,145	3,795	0,94	6	Dual specificity protein kinase sPlA (<i>Dictyostelium discoideum</i>)
T000025038.1	-6,018	3,654	5,10E-10	0,457	4,695	0,83	6	NA
T000024306.1	-5,43	1,095	0,00036	-0,137	1,787	0,97	6	Uncharacterized protein (<i>Strongylocentrotus purpuratus</i>)
T000007712.1	-3,013	4,82	0,00072	-2,558	4,771	0,02	6	NA
T000012311.1	-5,426	1,892	7,10E-06	-0,147	2,602	0,96	6	Cadherin EGF LAG seven-pass G-type receptor 3 (<i>Mus musculus</i>)
T000029264.1	-4,581	5,3	3,20E-05	-0,992	5,673	0,22	6	NA
T000028456.1	-5,085	2,142	8,00E-08	-0,512	2,693	0,71	6	Dual specificity protein kinase sPlA (<i>Dictyostelium discoideum</i>)
T000005880.1	-6,877	-0,486	0,00091	1,257	0,89	0,48	6	Similar to Histone H4 (<i>Ascaris suum</i>)
T000021244.1	-5,083	1,87	3,50E-06	-0,538	2,404	0,74	6	Ran-binding proteins 9/10 homolog (<i>Drosophila melanogaster</i>)
T000001881.1	-5,657	2,768	0,00052	0,022	3,569	1	6	ATP synthase subunit s%2C mitochondrial (<i>Rattus norvegicus</i>)
T000022627.1	-3,794	8,047	5,10E-05	-1,871	8,158	0,022	6	Pro-Pol polyprotein (<i>Simian foamy virus type 3 (strain LK3)</i>)
T000008966.1	-5,401	2,72	0,00096	-0,268	3,378	0,95	6	NA
T000021251.1	-4,629	1,958	0,00054	-1,05	2,303	0,42	6	ATP-dependent DNA helicase PIF1 (<i>Danio rerio</i>)

T000011398.1	-6,026	1,622	0,00036	0,322	2,576	0,92	6	Uncharacterized oxidoreductase YhdF (Bacillus subtilis (strain 168))
T000026181.1	-4,517	5,575	2,30E-07	-1,205	5,88	0,21	6	NA
T000005234.1	-5,832	1,784	8,20E-05	0,078	2,61	0,99	6	Histone H3.3 (Xenopus tropicalis)
T000012900.1	-3,275	3,54	1,70E-05	-2,483	3,508	0,0033	6	Glutathione synthetase (Dictyostelium discoideum)
T000010277.1	-6,303	3,618	2,70E-16	0,517	4,695	0,57	6	NA
T000013583.1	-4,871	4,357	2,20E-06	-0,935	4,747	0,35	6	NA
T000019410.1	-3,471	4,538	0,00042	-2,346	4,546	0,0029	6	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000026024.1	-3,788	4,905	3,30E-05	-2,031	4,985	0,12	6	Putative neurobeachin homolog (Caenorhabditis briggsae)
T000025890.1	-5,051	1,38	0,00061	-0,774	1,809	0,57	6	NA
T000003451.1	-5,245	3,419	1,60E-08	-0,616	3,934	0,68	6	Pro-Pol polyprotein (Feline foamy virus)
T000012661.1	-5,285	1,26	0,00038	-0,58	1,767	0,76	6	Dynein heavy chain%2C cytoplasmic (Caenorhabditis elegans)
T000028921.1	-6,257	1,813	1,30E-06	0,385	2,807	0,82	6	Similar to Calponin homolog OV9M (Onchocerca volvulus)
T000010324.1	-5,618	2,39	5,90E-11	-0,256	3,061	0,86	6	Lon protease (Methylocella silvestris (strain BL2 / DSM 15510 / NCIMB 13906))
T000010323.1	-5,618	2,39	5,90E-11	-0,256	3,061	0,86	6	Lon protease (Methylocella silvestris (strain BL2 / DSM 15510 / NCIMB 13906))
T000027868.1	-5,364	4,29	1,10E-17	-0,561	4,831	0,42	6	NA
T000022466.1	-5,74	2,52	0,00053	-0,186	3,222	0,96	6	Uncharacterized protein C1259.12c (Schizosaccharomyces pombe (strain 972 / ATCC 24843))
T000028253.1	-4,967	2,092	3,30E-05	-0,962	2,478	0,52	6	Uncharacterized protein (Pristionchus pacificus)
T000026183.1	-5,136	2,924	1,00E-06	-0,872	3,344	0,46	6	NA
T000002028.1	-4,769	3,529	2,20E-05	-1,247	3,828	0,38	6	Similar to Pathogenesis-related protein (Juniperus ashei)
T000028833.1	-5,377	3,358	3,50E-06	-0,65	3,865	0,69	6	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000008293.1	-5,022	3,045	0,00053	-1,079	3,401	0,56	6	NA
T000002009.1	-4,49	2,933	0,00092	-1,611	3,125	0,35	6	NA
T000025035.1	-5,841	2,146	4,60E-08	-0,278	2,802	0,89	6	Uncharacterized protein C1259.12c (Schizosaccharomyces pombe (strain 972 / ATCC 24843))
T000016495.1	-5,749	3,688	3,90E-06	-0,39	4,307	0,9	6	Actin-depolymerizing factor 1%2C isoforms a/b (Caenorhabditis elegans)
T000005361.1	-5,405	4,064	0,00022	-0,743	4,538	0,63	6	Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1A (Caenorhabditis briggsae)
T000016271.1	-4,722	0,732	9,10E-05	-1,429	0,946	0,29	6	M-phase phosphoprotein 8 (Homo sapiens)
T000028836.1	-5,315	2,327	2,40E-05	-0,869	2,748	0,61	6	Uncharacterized protein K02A2.6 (Caenorhabditis elegans)
T000020152.1	-5,33	3,394	4,60E-06	-0,874	3,821	0,57	6	NA
T000027317.1	-5,738	3,651	1,20E-08	-0,481	4,234	0,77	6	Choline kinase A2 (Caenorhabditis elegans)
T000021039.1	-6,276	3,377	7,30E-11	-0,007	4,179	1	6	NA
T000021032.1	-6,228	4,091	7,90E-18	-0,085	4,855	0,96	6	NA
T000004086.1	-6,412	2,952	0,00026	0,09	3,788	0,98	6	Dual specificity protein kinase CLK2 (Homo sapiens)

T000025632.1	-5,464	2,647	6,30E-11	-0,863	3,068	0,5	6	NA
T000025635.1	-5,464	2,647	6,30E-11	-0,863	3,068	0,5	6	NA
T000022610.1	-5,351	1,47	0,00033	-1,029	1,826	0,64	6	Palmitoyltransferase ZDHHC9 (Mus musculus)
T000017086.1	-5,984	3,763	1,50E-07	-0,413	4,372	0,84	6	NA
T000020889.1	-5,925	4,463	1,40E-14	-0,496	5,041	0,6	6	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13 (Bos taurus)
T000005887.1	-6,253	2,422	2,40E-10	-0,22	3,111	0,88	6	NA
T000013585.1	-5,725	3,616	5,40E-06	-0,778	4,074	0,6	6	NA
T000022614.1	-5,168	3,318	1,60E-10	-1,343	3,597	0,22	6	RRP15-like protein (Caenorhabditis briggsae)
T000017088.1	-5,36	3,387	5,20E-05	-1,199	3,705	0,36	6	Ribosome-binding protein 1 (Canis lupus familiaris)
T000029394.1	-6,886	3,349	3,00E-06	0,293	4,31	0,91	6	NA
T000027328.1	-5,707	1,904	0,00056	-0,903	2,316	0,66	6	NA
T000022342.1	-6,637	-0,692	0,00054	-0,054	-0,054	1	6	NA
T000024866.1	-4,511	-0,036	0,00083	-2,27	-0,01	0,15	6	NA
T000028633.1	-6,242	3,338	4,80E-08	-0,543	3,896	0,76	6	Dual specificity protein kinase splA (Dictyostelium discoideum)
T000017641.1	-5,391	3,027	2,10E-11	-1,397	3,286	0,035	6	NA
T000020151.1	-4,27	4,523	2,50E-08	-2,579	4,526	0,0027	6	NA
T000019777.1	-5,267	1,217	0,00015	-1,612	1,414	0,4	6	Polynucleotide 5'-hydroxyl-kinase nol-9 (Caenorhabditis elegans)
T000012660.1	-6,498	-0,8	0,00079	-0,388	-0,299	0,86	6	Dynein heavy chain%2C cytoplasmic (Drosophila melanogaster)
T000007902.1	-5,255	2,528	1,40E-05	-1,632	2,727	0,44	6	NA
T000009173.1	-6,523	2,824	5,60E-05	-0,408	3,438	0,89	6	Retrovirus-related Pol polyprotein from transposon 17.6 (Drosophila melanogaster)
T000007711.1	-3,909	4,142	7,00E-07	-3,026	4,075	0,0063	6	Mitochondrial import receptor subunit TOM22 homolog (Macaca fascicularis)
T000024865.1	-5,077	2,986	1,40E-06	-1,882	3,121	0,23	6	Ubiquitin-conjugating enzyme E2 D2 (Xenopus laevis)
T000016493.1	-6,662	4,007	1,20E-06	-0,305	4,671	0,88	6	Retinol dehydrogenase 13 (Mus musculus)
T000022626.1	-5,013	4,66	7,20E-05	-1,999	4,777	0,052	6	Envelopment polyprotein (Uukuniemi virus (strain S23))
T000028638.2	-6,284	1,122	0,00073	-0,741	1,586	0,8	6	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000012902.1	-5,501	5,135	2,30E-18	-1,613	5,345	0,18	6	Meckelin (Homo sapiens)
T000004125.1	-6,573	1,827	0,00014	-0,617	2,352	0,82	6	Putative ribosome-binding factor A%2C mitochondrial (Homo sapiens)
T000026023.1	-4,594	1,69	2,40E-06	-2,599	1,692	0,042	6	NA
T000006587.2	-6,485	3,069	2,60E-08	-0,717	3,557	0,78	6	Uncharacterized protein (Helobdella robusta)
T000008132.1	-7,06	1,868	2,90E-06	-0,25	2,545	0,95	6	NA
T000001919.1	-6,523	1,736	0,00039	-0,787	2,184	0,7	6	Pachytene checkpoint protein 2 homolog (Oryza sativa subsp. indica)
T000005884.1	-6,06	6,256	3,90E-11	-1,295	6,549	0,2	6	60S ribosomal protein L21 (Caenorhabditis elegans)

T000014076.1	-4,848	12,368	5,00E-04	-2,522	12,387	0,056	6	NA
T000025638.1	-6,174	4,648	7,20E-14	-1,203	4,973	0,26	6	NA
T000004728.1	-6,454	3,15	2,70E-08	-1,003	3,54	0,39	6	NA
T000004731.1	-6,454	3,15	2,70E-08	-1,003	3,54	0,39	6	NA
T000021036.1	-7,385	4,747	3,30E-21	-0,113	5,502	0,93	6	NA
T000023851.1	-6,595	1,7	6,50E-07	-0,97	2,083	0,78	6	NA
T000020157.1	-6,595	1,7	6,50E-07	-0,97	2,083	0,78	6	NA
T000021806.1	-5,808	1,385	2,20E-06	-1,76	1,549	0,42	6	NA
T000003449.1	-7,034	2,134	6,00E-04	-0,572	2,669	0,86	6	NA
T000017416.1	-5,656	6,323	0,00039	-2,043	6,443	0,44	6	Innexin-3 (Caenorhabditis elegans)
T000021221.1	-6,503	3,971	5,20E-05	-1,258	4,282	0,4	6	NA
T000002962.1	-8,053	0,539	0,00091	0,287	1,428	0,95	6	Calponin homology domain-containing protein DDB_G0272472 (Dictyostelium discoideum)
T000013694.1	-7,528	2,21	9,40E-12	-0,254	2,885	0,84	6	Lon protease homolog%2C mitochondrial (Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987))
T000012903.1	-5,611	3,11	7,40E-13	-2,185	3,199	0,089	6	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000013584.1	-6,852	5,403	1,30E-09	-0,948	5,806	0,39	6	2-oxoisovalerate dehydrogenase subunit alpha%2C mitochondrial (Caenorhabditis elegans)
T000022348.1	-6,378	2,865	3,60E-07	-1,427	3,127	0,23	6	NA
T000022441.1	-7,891	2,813	6,20E-08	-0,045	3,591	0,99	6	Uncharacterized protein C1259.12c (Schizosaccharomyces pombe (strain 972 / ATCC 24843))
T000016497.1	-6,704	4,39	9,20E-05	-1,273	4,681	0,49	6	Transthyretin-like protein 46 (Caenorhabditis elegans)
T000023842.1	-6,708	2,819	2,10E-06	-1,406	3,09	0,57	6	NA
T000024333.1	-7,519	2,052	6,30E-06	-0,739	2,535	0,69	6	Dorsal gland cell-specific expression protein (Heterodera avenae)
T000004324.1	-5,019	2,911	0,00053	-3,305	2,849	0,03	6	NA
T000014077.1	-5,116	15,746	0,00049	-3,289	15,681	0,0095	6	NA
T000029217.1	-8,416	2,217	6,10E-08	-0,046	2,996	0,99	6	NA
T000001928.1	-7,443	-0,01	1,90E-05	-1,18	0,275	0,45	6	Acetyl-coenzyme A synthetase (Drosophila melanogaster)
T000027320.1	-7,148	3,091	8,10E-06	-1,538	3,326	0,31	6	NA
T000004199.1	-7,375	1,976	1,70E-10	-1,342	2,262	0,28	6	NA
T000025521.1	-6,38	2,713	1,10E-05	-2,485	2,757	0,082	6	Retrovirus-related Pol polyprotein from transposon 17.6 (Drosophila melanogaster)
T000020158.1	-6,829	3,033	5,40E-10	-2,059	3,153	0,35	6	NA
T000024869.1	-7,662	0,179	1,00E-05	-1,512	0,387	0,45	6	Serine/threonine-protein phosphatase PP1 isozyme 7 (Arabidopsis thaliana)
T000023695.1	-7,387	2,282	1,20E-05	-1,985	2,414	0,43	6	Double homeobox protein A (Homo sapiens)
T000005379.1	-10,032	2,415	0,00096	2,304	4,757	0,25	7	Homeobox-containing protein 1 (Homo sapiens)
T000022440.1	-8,868	1,298	0,00058	0,963	2,613	0,76	7	Brain tumor protein (Drosophila melanogaster)

T000028283.1	-8,779	1,206	0,00022	0,511	2,266	0,82	7	NA
T000004218.1	-9,36	1,763	0,00077	0,709	2,955	0,82	7	Ubiquitin carboxyl-terminal hydrolase 8 (Homo sapiens)
								Dorsal gland cell-specific expression protein (Heterodera avenae)
T000028638.1	-8,049	2,941	8,20E-06	-0,604	3,477	0,81	7	NA
T000001932.1	-8,097	0,578	0,00048	-0,788	1,008	0,83	7	NA
T000021035.1	-8,931	1,356	0,00031	-0,214	2,025	0,96	7	NA
T000019807.2	-10,099	2,48	1,50E-06	0,943	3,81	0,69	7	NA
T000017034.2	-8,449	2,876	6,30E-06	-0,755	3,354	0,83	7	NA
T000020888.1	-8,463	0,911	3,10E-07	-0,748	1,369	0,64	7	NA
T000026182.1	-8,764	1,196	6,50E-06	-0,509	1,754	0,86	7	NA
T000011086.1	-10,215	2,593	2,60E-12	0,854	3,881	0,34	7	Ribonuclease UK114 (Capra hircus)
T000022345.1	-9,14	1,552	1,60E-06	-0,244	2,226	0,94	7	NA
T000017061.2	-8,193	0,661	3,00E-04	-1,199	0,978	0,59	7	Uncharacterized protein F23F12.3 (Caenorhabditis elegans)
T000012907.1	-8,358	2,959	1,20E-14	-1,191	3,288	0,65	7	NA
T000024770.1	-10,79	3,16	7,50E-05	1,194	4,685	0,69	7	Chloride channel protein clh-3 (Caenorhabditis elegans)
T000021696.1	-9,287	1,687	0,00018	-0,46	2,28	0,91	7	Cullin-1 (Pongo abelii)
T000017103.1	-8,986	1,403	1,20E-05	-0,805	1,848	0,72	7	NA
T000027336.2	-8,671	1,108	0,00027	-1,258	1,394	0,6	7	NA
T000013740.1	-8,246	0,712	5,00E-04	-1,779	0,879	0,59	7	Disorganized muscle protein 1 (Caenorhabditis elegans)
T000017089.1	-9,321	1,722	1,70E-08	-0,883	2,152	0,57	7	NA
T000016477.1	-10,407	2,781	6,40E-06	0,138	3,659	0,97	7	Uncharacterized protein K02A2.6 (Caenorhabditis elegans)
T000026253.1	-10,431	2,804	2,20E-05	0,107	3,685	0,98	7	Ran-binding protein 10 (Danio rerio)
T000023360.2	-9,235	1,643	3,30E-07	-1,142	1,997	0,55	7	Uncharacterized protein F23F12.3 (Caenorhabditis elegans)
T000002689.2	-9,224	1,629	0,00013	-1,174	1,954	0,64	7	NA
T000005359.1	-9,69	2,081	7,30E-08	-0,721	2,561	0,67	7	Casein kinase II subunit beta (Danio rerio)
T000020150.1	-8,776	1,207	6,10E-05	-1,978	1,335	0,37	7	NA
T000020895.1	-9,913	5,593	4,20E-21	-0,892	6,029	0,36	7	60S ribosomal protein L10a (Caenorhabditis elegans)
T000020871.1	-10,599	2,969	7,00E-13	-0,292	3,649	0,89	7	NA
T000027335.1	-8,04	3,442	0,00027	-2,898	3,44	0,077	7	NA
								Dorsal gland cell-specific expression protein (Heterodera avenae)
T000001862.1	-11,208	3,573	0,00028	0,067	4,418	0,99	7	NA
T000006526.1	-11,366	3,728	1,10E-08	-0,118	4,487	0,98	7	spidroin-1-like%2C partial (Corvus brachyrhynchos)
T000023443.1	-11,077	3,442	9,30E-05	-0,769	3,934	0,85	7	Uncharacterized protein (Dracunculus medinensis)
T000023166.1	-10,197	2,574	0,00044	-2,338	2,649	0,44	7	NA
T000028968.1	6,213	3,121	7,50E-05	0,424	-0,907	0,92	8	NA
T000016964.1	6,213	3,121	7,50E-05	0,424	-0,907	0,92	8	NA
								Probable tyrosine-protein kinase kin-31 (Caenorhabditis elegans)
T000015127.1	6,991	2,868	5,10E-19	-0,752	-2,011	0,86	8	Probable tyrosine-protein kinase kin-31 (Caenorhabditis elegans)
								Probable tyrosine-protein kinase kin-31 (Caenorhabditis elegans)
T000015126.1	5,62	2,794	2,80E-08	0,001	-0,92	1	8	NA
T000023883.1	6,399	1,777	5,20E-12	-0,855	-2,304	0,85	8	NA

T000017706.1	6,399	1,777	5,20E-12	-0,855	-2,304	0,85	8	NA
T000000193.1	5,103	0,089	0,00031	-0,038	-2,379	1	8	NA
T000028277.1	5,325	0,877	1,00E-05	-3,267	-2,469	0,26	8	NA
T000022645.1	-6,027	3,805	0,00035	-6,291	3,628	7,10E-08	9	NA

Chapitre 6 Discussion générale

L'introduction d'espèces exotiques envahissantes telles que le NKS dans les agroécosystèmes constitue une menace à l'équilibre écosystémique et génère des pertes économiques associées à la baisse de rendement, au contrôle des populations et à la recherche de solutions de gestion durables. Ce sont les différentes composantes du milieu environnemental et climatique, les traits phénotypiques et génomiques de l'hôte et de l'espèce introduite de même que les moyens de lutte et leur durabilité qui vont expliquer la réussite ou l'échec d'une invasion. Dans cette thèse, les différents objectifs visaient à décomposer les éléments inhérents à une espèce exotique invasive en utilisant comme modèle le NKS. Dans cette optique, les objectifs devaient répondre à plusieurs questions. D'abord (1) est-ce que le NKS est susceptible de se propager dans un nouvel environnement tel que le Québec ? Étonnamment, la modélisation bioclimatique réalisée dans le cadre de ce doctorat a démontré qu'actuellement, le NKS a le potentiel de s'établir dans toutes les aires de production du soya puisque la température ne constitue pas une barrière à la reproduction de cette espèce poïkilotherme. En outre, les recherches ont démontré que son expansion pourrait être favorisée par les hausses de températures prévues dans le futur proche puisqu'à la fois de nouvelles aires de production du soya s'ouvriront et que le nombre de générations produites doublera dans certaines régions au cours d'une saison de croissance. Ensuite (2) d'où proviennent les populations de NKS récemment introduites au Canada et ces populations possèdent-elles des gènes favorisant leur adaptation à de nouvelles composantes environnementales et climatiques ? À l'aide d'approches liées à la génétique des populations, les travaux ont démontré que la grande diversité génétique qui caractérise les populations nord-américaines de NKS ainsi que le flux de gènes continu entre les régions ont facilité son adaptation aux diverses conditions bioclimatiques et son établissement dans les nouvelles zones de production de soya. Finalement (3) quels sont les gènes associés au contournement de la résistance chez l'hôte et qui affectent la durabilité de ce moyen de lutte ? La réponse à cette question n'est pas simple, mais cette thèse a permis d'identifier pour la première fois une région génomique qui semble être directement associée à la reconnaissance du NKS par les principaux gènes de résistance. Mis ensemble, tous ces résultats ouvrent plusieurs pistes prometteuses pour les recherches futures et le développement d'outils pour faire face à cette menace de façon durable et protéger une industrie en plein essor.

Évaluer les risques de propagation et d'établissement des espèces récemment introduites dans un contexte de changements climatiques.

Tel que mentionné précédemment, le NKS a été découvert au Québec pour la première fois en 2013. Son identification a soulevé une panoplie de questions quant à la dynamique de cette espèce sur un nouveau territoire et aux risques inhérents à sa présence au Québec. C'est dans cette optique que l'on a effectué une modélisation bioclimatique de son hôte économique principal, le soya et du nématode.

Plusieurs composantes vont expliquer la capacité d'une espèce à s'établir dans un nouvel environnement ainsi qu'à s'y propager. Par exemple, une forte reproduction, une large gamme d'hôtes, une grande plasticité face aux stress environnementaux, l'absence d'ennemis naturels, etc. Parmi celles-ci, la capacité à se reproduire facilement a été retenue comme étant le facteur le plus important dans la réussite de l'invasion du NKS. Étant donné la nature poïkilotherme du NKS et aidée par la littérature, un modèle déterministe reposant sur des régressions logistiques entre la température du sol et le taux de développement du NKS a été développé et présenté dans l'article du chapitre 3. Sachant que la distribution du NKS suit celle de son hôte (Tylka & Marrett, 2017), il serait possible de connaître l'aire de dispersion potentielle du NKS en utilisant un modèle reposant sur la phénologie (déterminée principalement par la température de l'air et la photopériode) de son hôte économique principal, le soya.

L'intégration de données climatiques actuelles (1981-2010) à ces deux modèles (cycle de vie du NKS et phénologie du soya) a permis de révéler le potentiel reproductif du NKS dans un nouvel environnement dépassant son actuelle limite septentrionale de distribution (figure 3.4). Étonnamment, le modèle a prédit que le NKS pouvait déjà se reproduire au Québec (figure 3.4), et que le climat lui permettait, dans les régions où l'on produit les plus grandes quantités de soya, de produire jusqu'à trois générations par saison de croissance. Ainsi, les deux modèles reposant sur l'utilisation de la température pour faire varier les paramètres du cycle de vie du NKS et de la phénologie du soya donnent une indication claire quant aux risques déjà présents pour la production de soya au Québec.

Par la suite, sachant que le climat est en constante évolution et que les modèles généraux prévoient une hausse marquée des températures, la question de recherche a été étendue au futur

proche (2041-2070). Selon le dernier rapport du GIEC (Groupe d'experts Intergouvernemental sur l'Évolution du Climat, 2014), l'augmentation des concentrations de CO₂ atmosphériques devrait influencer à la hausse les températures mondiales (0,3-4,8 °C) d'ici 2100. Parmi les différents scénarios proposés par le GIEC, deux ont été retenus dans le cadre de mes recherches, le premier, plus optimiste (RCP 4.5), prévoit une hausse de près de 1,1 - 2,6°C d'ici 2100 alors que le second, plus pessimiste (RCP 8.5), prévoit plutôt une hausse de près de 2,6 – 4,0 °C. Ainsi, des jeux de données climatiques simulées par Ouranos sur la base de différents scénarios climatiques pour la période 2041-2070 ont été intégrés aux deux modèles. L'utilisation de données climatiques simulées plutôt qu'actuelles a permis non seulement de prédire l'agrandissement de l'aire de production du soya (figure 3.2 et 3.3), mais en plus d'entrevoir la possibilité d'un changement de groupe de maturité en se basant sur les dates d'atteinte de maturité (figure 3.3). Grâce à ces prédictions, il est possible d'inférer non seulement que de nouvelles zones pourront être affectées par le NKS suivant la distribution de son hôte, mais qu'en plus, on peut s'attendre à un changement de groupe de maturité qui lui, influera sur le nombre de générations produites et sur l'accroissement de la densité de population. En effet, Joyce et Thiessen (2018) mentionnent qu'un changement de groupe de maturité modifiera les dates d'atteinte de la maturité et permettra la production d'une génération supplémentaire favorisant une plus forte densité de population. En outre, la simulation des températures attendues a permis de prédire une hausse marquée de la vitesse requise pour accomplir chaque stade du cycle de vie du NKS (figure 3.4). Alors que durant la période actuelle on peut s'attendre à ce que le NKS accomplisse trois générations puisque la durée de chaque cycle complet varie entre 28 et 49 jours, d'ici 50 ans ce nombre pourrait atteindre cinq générations et la durée du cycle pourrait être aussi courte que 25 jours (figure 3.5).

À notre connaissance, nous avons développé le premier modèle bioclimatique simulant tant le cycle de vie d'un nématode endoparasite en fonction des températures du sol et de son hôte en fonction des températures de l'air pour la période actuelle et future. Les résultats présentés dans le cadre de l'article ont démontré qu'aucune barrière climatique ne freinera la progression du NKS vers le nord, au contraire, la hausse de la température prévue pour le futur favorisera l'ouverture de nouvelles aires de production de soya plus au nord et accentuera la vitesse des taux de développement de l'hôte et du parasite. Ainsi comme le Québec est le troisième producteur de soya au Canada et que ce dernier est le septième producteur mondial, il est essentiel de limiter la

propagation de cette espèce ainsi que sa reproduction afin de maintenir la production et rester compétitif sur le marché. Toutefois, jusqu'à présent peu de solutions se sont avérées efficaces pour prévenir de nouvelles introductions et la propagation du NKS à cause des trop nombreuses voies d'entrée et de dissémination (tractus intestinal d'oiseaux, échange de machinerie agricole, forts vents, inondations, souliers souillés, etc.). L'unique solution réside dans le contrôle des populations en utilisant des espèces non hôtes ainsi que des cultivars résistants. Il est donc urgent d'investir dans la recherche afin de déployer des cultivars adaptés au Québec et surtout durables.

Bien que simples, les modèles semblent avoir démontré leur efficacité quant aux prédictions basées sur les températures. Même si le climat est en constante évolution, les modèles sont assez performants pour qu'on puisse facilement intégrer de nouveaux jeux de données climatiques. En outre, plusieurs facteurs limitants le taux de développement ont été exclus et pourraient être intégrés dans le futur.

Détermination du potentiel adaptatif d'une espèce dans un nouvel environnement à l'aide d'une approche génomique

Après avoir confirmé le potentiel d'établissement, de propagation et de reproduction du NKS dans un nouvel environnement, plusieurs éléments quant au processus d'adaptation du NKS sont restés inconnus. En effet, les prédictions précédentes reposaient uniquement sur l'effet des fluctuations de températures sur le développement et aucunement sur la capacité intrinsèque du NKS à s'adapter à un nouvel environnement. De plus, le modèle ne donnait aucune indication quant à la route d'introduction du NKS jusqu'au Québec, qui représente toujours sa limite de distribution la plus septentrionale en Amérique. Ainsi, on a voulu connaître les liens génétiques qui unissent les différentes populations de NKS présentes aux États-Unis et au Canada. Également, on a cherché à établir si certains gènes facilitaient l'adaptation du NKS en identifiant des signatures génétiques associées à des données environnementales et climatiques. Dans l'article présenté au chapitre 4, une approche génomique a été préconisée et repose sur la prémisse que les variations génétiques entre les populations refléteraient leur potentiel évolutif et adaptatif (Andrews, 2010; Eizaguirre & Baltazar-Soares, 2014). Pour être en mesure de séquencer plusieurs dizaines de populations (64 populations) des États-Unis et du Canada, l'approche du génotypage par séquençage (GBS) a été utilisée, car elle permettait, à faible coût, de produire une grande quantité de séquences, donc de marqueurs, avec une bonne profondeur de

séquençage (voir chapitre 4 pour plus de détails). Au moment de préparer cet article, les quantités d'ADN prélevées par individus étaient insuffisantes pour effectuer un séquençage, c'est pourquoi nous avons choisi de «pooler» plusieurs individus, ce qui a complexifié les analyses et leur interprétation. Cependant, l'utilisation d'un pool d'individus a eu pour avantage de capturer la diversité génétique retrouvée au sein de chaque population.

L'analyse de la structure des populations a été nécessaire afin de répondre aux questions associées aux voies d'introduction et aux relations génétiques entre les populations nord-américaines. Afin d'analyser la structure des populations, différentes approches et méthodes ont été préconisées et ont été appliquées aux jeux de données constitués de marqueurs SNP : une analyse en composantes principales (ACP ou PCA en anglais) (figure 4.1a); la création de l'arbre phylogénétique (figure 4.1b); le calcul de l'indice de différenciation génétique pour chaque population; l'analyse de l'isolation par la distance (figure 4.2 ab). Ces analyses ont permis d'identifier deux grands groupes, celui du nord-est (Ontario) et celui du nord-ouest (Dakota du Nord et Minnesota) partageant des similarités génétiques avec les populations « d'origines » (centre des États-Unis) (voir figure 4.1ab) et qui reflétaient également les deux grandes routes d'introduction (nord-est et nord-ouest). Elles ont également montré que dans les groupes du nord-est et du nord-ouest, la diversité génétique était beaucoup plus élevée que dans les états centraux qui sont beaucoup plus homogènes. Ce résultat a surpris, car on s'éloignait de la logique de l'effet fondateur qui stipule que les populations récemment introduites sont moins diverses, dû à un échantillonnage d'une portion des allèles présents dans la population d'origine. Ce résultat a mené à inférer que la pression de sélection causée par la surreprésentation de PI88788 en tant que source de résistance a eu pour effet de limiter la sélection négative en sélectionnant des individus partageant les mêmes gènes dans les états centraux. Cette déduction reposait sur des observations de même nature faites par d'autres chercheurs (voir Liu *et al.*, 1997; Niblack *et al.*, 2008; Fagihi *et al.*, 2010; Zheng & Chen, 2011), mais aussi par l'utilisation moins répandue et sur une moins longue période de temps de la source de résistance PI88788 dans les nouveaux territoires infestés.

En outre, l'analyse de la différenciation génétique et de l'isolation par la distance ont démontré que plus on s'éloigne de la population d'origine « ancestrale » (dans notre jeu de données il s'agissait d'une population du Missouri) plus la différenciation génétique augmente à l'échelle régionale. Cependant cet effet ne fut pas observé à l'échelle locale. En effet, les indices de

différentiations génétiques étaient très faibles dans les populations centrales indiquant un fort flux génique et, mis en relation avec la faible diversité génétique, on obtient une rapide homogénéisation de celles-ci. D'ailleurs, les travaux de Faghihi et al. (2010) confirmaient à la fois la voie d'introduction des états centraux vers l'Ontario (nord-est) et l'homogénéisation des populations centrales puisqu'ils ont trouvé que 15 % des populations ontariennes pouvaient se reproduire sur PI548402 et 6% sur PI 90763. Ces deux derniers cultivars étant absents dans les cultivars commerciaux disponibles sur le marché ontarien, il était surprenant de trouver des individus déjà aptes à contourner leurs mécanismes de défense sans y avoir été exposés au préalable et cela ne pouvait être expliqué que par une introduction de ces états.

Finalement, l'utilisation de scans génomiques a permis de détecter 71 locus sous sélection desquels 15 étaient validés par plus d'une approche. Le faible nombre de locus potentiellement soumis à une sélection par des facteurs environnementaux ou climatiques nous indique que le NKS est probablement déjà adapté à une large gamme d'environnements. Ainsi, si le NKS est déjà adapté aux conditions climatiques qui prévalent à la limite septentrionale de son aire de distribution actuelle, on peut s'attendre à ce que l'utilisation de cultivars commerciaux résistants ait un effet très limité dans le temps pour réduire la reproduction des populations. Compte tenu du fait que les populations ancestrales de l'Ontario proviennent des populations centrales, l'utilisation de la source de résistance PI88788 sera rapidement inefficace pour limiter la reproduction du NKS. En outre, comme les populations du Québec arrivent probablement de l'Ontario, le même problème se posera. De plus, de nouvelles introductions de NKS en provenance des régions où les populations ont été sélectionnées par les cultivars résistants apporteront forcément des individus beaucoup plus virulents. Ainsi, il est urgent de déployer de nouvelles sources de résistance sur le marché, mais aussi d'identifier les gènes responsables de la virulence.

Profilage génomique et transcriptomique de la virulence chez *Heterodera glycines*

Tel que vu précédemment, il existe une incroyable diversité génétique au sein des populations de NKS qui s'explique par les flux migratoires continus, par l'éclosion aléatoire « d'ancêtres » persistant dans les sols en dormance et protégés par les kystes ainsi que le mode reproductif qui implique la polyandrie (plusieurs mâles peuvent se reproduire avec une seule femelle). De plus, à l'heure actuelle, aucun cultivar de soya ne possède une résistance universelle à tous les individus

présents dans une population ce qui limite la durabilité de leur utilisation en particulier puisque près de 95% de ce qui est disponible sur le marché dépend d'une unique source PI88788. C'est pourquoi la dernière partie de cette thèse portait sur l'identification des gènes impliqués dans la virulence.

Comme mentionné plus haut, au début de ce projet il n'était pas possible de séquencer des individus uniques à cause des quantités d'ADN requises pour le séquençage. Dans l'optique de maximiser les chances d'identifier des gènes candidats associés aux HG type, l'approche du GBS de pool d'individus issus de multiples populations permettant d'avoir une foule d'informations génétiques sur une vaste gamme de phénotypes de virulence a été utilisée (voir la section 2.1 pour plus de détails). Les SNP identifiés par le GBS, ont permis d'effectuer des centaines de régressions logistiques entre la fréquence allélique (FA) de tous les SNP (804) et le female index (FI) de chaque population (64) associés au HG type 1 (virulent sur la source de résistance Peking) et au HG type 2 (virulent sur la source de résistance PI88788) qui ont permis d'identifier 14 locus statistiquement associés à la virulence (table 5.2). Étonnamment, la plupart de ces marqueurs étaient localisés directement dans des gènes. Parmi ces gènes candidats, on retrouvait un homologue d'un effecteur connu d'*Heterodera avenae* corrélé avec la capacité à contourner les gènes de résistance de PI88788. L'approche du GBS appliquée sur un pool d'individus composant des populations ayant une forte diversité génétique quoique très utile dans le cadre de la génétique des populations a énormément complexifié l'interprétation de fréquences alléliques dans le cadre d'une étude d'association entre le phénotype et le génotype. En effet, il a été difficile de faire corrélérer les FA avec les FI c'est pourquoi une nouvelle approche a été initiée.

Afin de limiter la variabilité génétique des pools d'individus indifférenciés et non sélectionnés par les plantes résistantes (stades J1-J2) et de faciliter les analyses, 40 femelles du stade J4 ont été sélectionnées sur Peking et PI88788. Aussi, afin d'augmenter la probabilité de trouver un gène candidat, on a procédé au séquençage de génome complet pour chaque pool d'échantillons sélectionnés ainsi que des populations d'origine. L'analyse du déséquilibre allélique (*allelic imbalance*) des variants communs a permis d'identifier près de 784 variants dans 586 gènes parmi lesquels 86 (voir tables supplémentaires S 5.1 et S 5.2) étaient associés à la sélection sur Peking et 503 étaient associés à la sélection sur PI88788 (voir tables supplémentaires S 5.3 et S 5.4). Un total de 17 gènes ont été identifiés et possédaient des variants prédits pour avoir un

impact important sur la séquence d'acides aminés (table 5.3). Parmi ceux-ci, trois gènes étaient connus pour être des effecteurs (*aspartyl aminopeptidase*) ou pour être potentiellement associés au parasitisme (*dorsal gland cell-specific expression protein, mucin*) (Gems & Maizels, 1996; Davis, 2000; Gao *et al.*, 2003; Lilley *et al.*, 2005; Haegerman *et al.*, 2012) (voir la section 2.2. pour informations sur les effecteurs). En outre, on a identifié 109 gènes associés à la virulence sur PI88788 (table 5.4) dans lesquels les variants auraient un fort impact sur la séquence d'acides aminés. Treize de ces gènes ont préalablement été associés au parasitisme du NKS : *annexin, esophageal gland proteins, endonuclease, guanylate cyclase, pectate lyase, peptidase* et *ubiquitin*) (Voir Haegerman *et al.*, 2012 qui ont proposé une belle revue des gènes confirmés pour être impliqués dans le parasitisme). Parmi ces gènes, celui codant pour la *gland protein G20E03* est le plus connu et est fréquemment identifié dans les études pour être fortement exprimé lors d'interactions incompatibles dans les plantes résistantes. Notre analyse a révélé que seuls les individus ne possédant qu'un allèle de ce gène (homozygotes) étaient capables de contourner les mécanismes de résistance de PI88788. L'allèle restant codait pour une phénylalanine (apolaire) alors que l'autre allèle présent dans la population virulente codait pour une cystéine (polaire). Dans le futur, il faudrait identifier l'importance du changement de polarité des acides aminés sur la structure et la fonction de cette protéine. Il faudrait aussi confirmer si l'homozygotie à ce locus est constant dans les populations et s'il est uniquement relié à la virulence sur PI88788. Plusieurs autres variants détectés dans le génome semblent être intimement liés au contournement de la résistance, mais ces derniers demanderont une analyse plus approfondie afin de valider leur impact. L'utilisation de cette approche s'est révélée très prometteuse, car nous avons réussi à identifier des dizaines de gènes porteurs de variants à forts impacts dont certains étaient déjà connus pour être impliqués dans le parasitisme.

Finalement, en 2018, Serra *et al.*, ont publié une méthode permettant d'effectuer un séquençage complet d'ARN (RNA-Seq) à partir d'un seul nématode. Grâce à leurs travaux, nous serons les premiers à analyser l'expression de l'ensemble des gènes du NKS dans des individus uniques. Ainsi, dans cette dernière partie de l'article, neuf populations ontariennes ont été sélectionnées d'une part, afin de limiter la variabilité génétique due à l'origine géographique et d'autre part, pour leur FI sur les lignées Peking et PI88788. Au total, 23 femelles sélectionnées sur des lignées résistantes (Peking ou PI88788) ou sensibles (Essex) ont été isolées des racines afin d'analyser les gènes différentiellement exprimés (DEG) en fonction de la sélection par la plante. Le patron

général du profil d'expression des femelles sélectionnées sur Peking ou sur PI88788 versus celles non sélectionnées sur Essex (cultivar sensible) (figure 5.1) nous confirme que les mécanismes de virulence sont probablement différents entre Peking et PI88788, mais aussi que les gènes surexprimés ou réprimés le sont plus fortement sur Peking. Cela est peut-être lié aux réactions d'hypersensibilité plus accentuées et plus rapidement exprimées dans les racines de Peking forçant le nématode à contrebalancer ces mécanismes (voir la section 2.2 pour plus de détails sur les mécanismes de résistance). Plusieurs des transcrits ayant un patron d'expression inverse entre les femelles sélectionnées sur Peking et sur PI88788 étaient des gènes connus pour être associés au parasitisme et nécessiteront une analyse plus approfondie dans le futur. L'analyse des DEG entre les femelles sélectionnées sur Peking et sur PI88788 (sans Essex) confirme essentiellement la même chose que la figure 5.1, mais permet surtout de mettre en évidence des gènes uniques associés à chaque gène de résistance.

L'analyse de ces gènes uniquement associés aux femelles sélectionnées sur Peking ou sur PI88788 a permis d'identifier un total de 22 transcrits uniques, quatre associés à la virulence sur Peking et 18 à celle sur PI88788. Parmi ceux-ci, deux gènes issus du même scaffold (589) des *dorsal gland cell-specific expression* (hgg-20) ont retenu l'attention, car l'un (T000028080.1) était unique à la virulence sur Peking. L'analyse de cette région génomique nous a permis d'identifier un îlot contenant des duplications en tandem de différentes versions du même gène, hgg-20 (voir figure 5.2). Celui-ci est connu pour être exprimé dans les réactions d'incompatibilité avec des populations virulentes d'*H. glycines* autant durant les premiers stades (J2-J3) que dans les derniers stades (J3-J4) (Ithal *et al.*, 2007; Klink *et al.*, 2008). Les recherches menées par Ital *et al.* (2007), ont même montrées que l'expression d'hgg-20 était plus forte dans les derniers stades lors d'une interaction incompatible avec des individus virulents sur Peking. La découverte des différences d'expression des gènes uniques associées avec les deux réactions incompatibles sur Peking et PI88788 laisse à penser que cet îlot pourrait être impliqué dans la virulence. Cependant, d'autres éléments restent à valider avant la soumission de cet article. Notamment, nous devons déterminer si l'expression sélective de ces deux gènes est due à un contrôle transcriptionnel (répresseur, défaut de promoteur, dégradation de l'ARNm) ou à une large délétion dans l'ADNg. Si la dernière possibilité s'avérait, il serait dès lors possible de développer facilement un outil diagnostique permettant d'anticiper la virulence d'une population sur les cultivars issus de Peking ou PI88788 en mettant en évidence ces délétions.

Principales contributions

- J'ai réalisé la conception du modèle simulant le cycle de vie du NKS en réponse à la température ainsi que les analyses cartographiques associées de l'article présenté au chapitre 3.
- J'ai effectué les manipulations en laboratoire ainsi que les analyses de structures génétiques et les scans génomiques de l'article présenté au chapitre 4.
- J'ai effectué la manipulation ainsi que modifié les protocoles associés aux tâches en laboratoire, j'ai effectué les analyses sur les données de GBS, WGS et RNA-Seq présentés au chapitre 5.

Conclusion et perspective

Caractériser la biologie du NKS depuis son cycle de vie jusqu'à l'identification génétique et transcriptomique des gènes associés à ses traits adaptatifs et phénotypiques a été un défi de taille. Toutefois, il a été relevé dans les chapitres 3, 4 et 5. Dans cette thèse, nous avons modélisé le cycle de vie du NKS en fonction des températures et démontré son potentiel de propagation et d'établissement dans un nouvel environnement ainsi que sa capacité à s'y reproduire sous différents climats (actuel et futur). Ce premier et unique modèle bioclimatique ouvre la porte à sa réutilisation avec des données climatiques mises à jour, à son amélioration en intégrant de nouveaux paramètres limitant le taux de développement et ultimement à sa validation en champs. Également, dans cette thèse il a été démontré que la grande diversité génétique et le flux génétique continu qui caractérisent les populations de NKS en Amérique du Nord facilitent leur adaptation à diverses conditions bioclimatiques et leur établissement dans de nouvelles régions. Malheureusement, nous avons également démontré que cette adaptation rapide était aussi vraie pour les cultivars résistants et que les sources utilisées présentement seraient bientôt très peu utiles. Basé sur ces informations, il est impératif d'inclure de nouvelles sources de résistance dans les programmes d'amélioration génétique et de développer des cultivars résistants qui seront également adaptés à des conditions plus nordiques que celles où l'on cultive le soya en ce moment. Finalement, les analyses GBS, WGS et RNA-seq ont permis d'identifier des centaines de gènes qui pourraient être liés à la virulence. Grâce au séquençage d'individus uniques, nous avons découvert un îlot qui semble assez spécifique pour nous permettre de développer des marqueurs permettant de discriminer les deux types de virulence. Ces outils diagnostiques auront

plusieurs utilités. Ils permettront de mieux conseiller les producteurs sur le choix de cultivars à utiliser dans leur champ afin de maximiser les rendements et la rentabilité de leur ferme. D'autre part, la mise en évidence de populations virulentes dans un champ permettra de retirer la pression de sélection qui favorise son développement et de ralentir sa dispersion. Cela permettra d'étendre la durée de vie des sources de résistances en limitant le flux de gènes de virulence. Dans certains champs, il deviendra sans doute préférable de prolonger les rotations avec des cultures non hôtes au lieu de mettre en péril l'efficacité des cultivars résistants dans les autres champs. Dans de futures études, les gènes d'intérêt identifiés dans cette étude pourront également être inactivés par transformation génétique ou des ARN interférence afin de valider leur rôle dans la virulence. La confirmation de ces cibles permettrait également de développer des plants de soya exprimant des ARNi inhibant directement la production des protéines de virulence (voir l'article de Qi *et al.*, 2019 pour plus de détails). L'utilisation de plusieurs cibles dans des lignées différentes, combinée à un bon suivi de l'évolution des populations au champ, permettrait un contrôle optimal de ce parasite jusqu'ici dévastateur.

Références

- Agrios, GN (2005) Plant diseases caused by nematodes. In Agrios, G. N. (Ed.). Plant pathology. (5th ed., pp. 826-874). Oxford : Elsevier Academic Press.
- Ahanger, R.A., Bhat, H.A., Bhat, T.A., Ganie, S.A., Lone, A.A., Wani, I.A., Ganai, S.A., Haq, S., Khan, O.A., Junaid, J.M. & Bhat, T.A. (2013). Impact of Climate Change on Plant Diseases. *Int. J. Modern Plant & Anim. Sci*, 1, 105-115.
- Allen, T.W., Bradley, C.A., Sisson, A.J., Byamukama, E., Chilvers, M. I., Coker, C. M., ... Wrather, J.A. (2017). Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2010 to 2014. *Plant Health Prog.*, 18, 19-27.
- Ali, N. (2010). Soybean processing and utilization. In G. Singh (Éd.). The soybean: botany, production and uses. (1st ed., pp. 345-374). Oxfordshire : CAB International.
- Alston, D.G. & Schmitt, D.P. (1987). Population density and spatial pattern of *Heterodera glycines* in relation to soybean phenology. *J. Nematol.*, 19, 336-345.
- Anderson, J.P., Badruzaufari, E., Schenk, P.M., Manners, J.M., Desmond, O.J., Ehlert, C., ... Kazan, K. (2004). Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in arabidopsis. *Plant Cell*, 16, 3460-3479.
- Anderson, T.R. & Welacky, T.W. (1988). First report of *Heterodera glycines* on soybeans in Ontario, Canada. *Plant Dis.*, 72, 453.
- Andrews, C.A. (2010). Natural selection, genetic drift, and gene flow do not act in isolation in natural populations. *Nat. Edu. Knowl.* 3(10), 5.
- Bamshad, M., & Wooding, S.P. (2003). Signatures of natural selection in the human genome. *Nat. Rev. Genet.* 4, 99–111.
- Bekal, S., Domier, L.L., Gonfa, B., Lakhssassi, N. & Lambert, K.N. (2015). A SNARE-like protein and biotin are implicated in soybean cyst nematode virulence. *PLoS ONE*, 10, e0145601
- Bekal, S., Craig, J.P. Hudson, M.E. Niblack, T. Domier, L.L. & Lambert, K.N. (2008). Genomic DNA sequence comparison between two inbred soybean cyst nematode biotypes facilitated by massively parallel 454 micro-bead sequencing. *Mol. Genet. Genomics*, 279, 535-543.

- Bernard, G. C., Egnin, M. & Bonsi, C. (2017). The Impact of Plant-Parasitic Nematodes on Agriculture and Methods of Control. In Shaw, M.M. & Mahamood, M. (Eds.). *Nematology - Concepts, Diagnosis and Control*. (pp. 121-151). IntechOpen.
- Bonin, A., Taberlet, P., Miaud, C. & Pompanon, F. (2006). Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol*, *23*, 773–783.
- Bonner, M.J. & Schmitt, D.P. (1985). Population dynamics of *Heterodera glycines* life stages on soybean. *J. Nematol.*, *17*, 153-158.
- Bradley, C., Allen, T., Chilvers, M., Giesler, L., Mehl, K., Mueller, D., ... Wise, K. (2017). Soybean Disease Loss Estimates From the United States and Ontario, Canada — 2015. Retrieved from <http://cropprotectionnetwork.org/wp-content/uploads/2017/04/CPN-1018-15-W-1.pdf> (25-08-17).
- Brucker, E., Carlson, S., Wright, E., Niblack, T. & Diers, B. (2005). Rhg1 alleles from soybean PI 437654 and PI 88788 respond differentially to isolates of *Heterodera glycines* in the greenhouse. *Theor. Appl. Genet.*, *111*, 44-49.
- Burrows, P. R., & Stone, A. R. (1985). *Heterodera glycines*. C.I.H. Descriptions of plant parasitic nematodes, Commonwealth Agricultural Bureaux, Set. 8 No 118, 4 pp.
- Burton, J. W. (1997). Soyabean (*Glycine max* (L.) Merr.). *Field Crop Res.*, *53*, 171-186.
- CABI (2019). *Heterodera glycines* (Soybean cyst nematode). In: Invasive Species Compendium. Wallingford, UK: CAB International. www.cabi.org/isc.
- Caldwell, B.E., Brim, C.A. & Ross, J.P. (1960). Inheritance of resistance of soybeans to the cyst nematode, *Heterodera glycines*. *Agron. J.*, *52*, 635-636.
- Chakraborty, R., & Nei, M. (1977). Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution*, *31*, 347-356.
- Chakraborty, S. & Datta, S. (2003). How will plant pathogens adapt to host plant resistance at elevated CO₂ under a changing climate? *New Phytol.*, *159*, 733-742.
- Charlesworth, B., Morgan, M.T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*, 1289–1303.
- Chen, C., Liu, S., Liu, Q., Niu, J., Liu, P., Zhao, J., & Jian, H. (2015). An ANNEXIN-Like protein from the cereal cyst nematode *Heterodera avenae* suppresses plant defense. *PLoS ONE*, *10*(4), e0122256.
- Choudhuri, S. (2014). Fundamentals of molecular evolution. In Choudhuri, S. (Ed.) *Bioinformatics for Beginners*. (239p). Oxford : Academic Press.

- Cianzio, S.R. (2008). Soybean breeding achievements and challenges. In Kang, M. S. & Priyadarshan, P.M. (Eds). *Breeding Major Food Staples* (pp. 245-273), New York : John Wiley & Sons.
- Concibido, V.C., Diers, B.W. & Arelli, P.R. (2004). A decade of QTL mapping for cyst nematode resistance in soybean. *Crop Sci.*, *44*, 1121–1131.
- Cober, E.R. & Morrison, M. J. (2010). Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. *Theor. Appl. Genet.*, *120*(5), 1005-1012.
- Cook, D.E., Bayless, A.M., Wang, K., Guo, X., Song, Q., Jiang, J., & Bent, A.F. (2014). Distinct copy number, coding sequence, and locus methylation patterns underlie Rhg1 - mediated soybean resistance to soybean cyst nematode. *Plant Physiol.*, *165*(2), 630-647.
- Cook, D.E., Lee, T.G., Guo, X., Melito, S., Wang, K., Bayless, A.M., ... Bent, A.F. (2012). Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science*, *338*, 1206-1209.
- Craig, J. P., Bekal, S., Hudson, M., Domier, L., Niblack, T. & Lambert, K.N. (2008). Analysis of horizontally transferred pathway involved in vitamin B6 biosynthesis from the soybean cyst nematode *Heterodera glycines*. *Mol. Biol. Evol.*, *25*, 2085-2098.
- Craig, J.P., Bekal, S., Niblack, T., Domier, L. & Lambert, K.N. (2009). Evidence for horizontally transferred genes involved in the biosynthesis of vitamin B1, B5, and B7 in *Heterodera glycines*. *J. Nematol.*, *41*, 281-290.
- Creech, J.E., Santini, J.B., Conley, S.P., Westphal, A., & Johnson, W.G. (2007). Purple deadnettle (*Lamium purpureum*) and soybean cyst nematode response to cold temperature regimes. *Weed Sci.*, *55*(6), 592-598.
- Davis, E. L., Hussey, R. S., Baum, T. J., Bakker, J., Schots, A., Rosso, M.-N. & Abad, P. (2000) Nematode Parasitism Genes. *Ann. Rev. Phytopathol.*, *38*, 365-396.
- Davis, E.L. & Tylka, G.L. (2000). Soybean cyst nematode disease. The Plant Health Instructor. Retrieved from <https://www.apsnet.org/edcenter/disandpath/nematode/pdlessons/Pages/SoyCystNema.aspx> (12-08-19).
- de Boers, J.M., McDermott, J.P., Davis, E.L., Hussey, R.S., Smant, G., & Baum, T.J. (2002). Cloning a putative pectate lyase gene expressed in the subventral esophageal glands of *Heterodera glycines*. *J. Nematol.*, *33*, 9-11.
- de Boers, J.M., Yan, Y., Wang, X., Smant, G., Hussey, R.S., Davis, E.L. & Baum, T.J. (1999). Developmental expression of secretory β -1,4-endoglucanases in the subventral esophageal glands of *Heterodera glycines*. *Mol. Plant-Microbe Interact.*, *12*, 663-669.

- Dhutmal, R. R., Mundhe, A.G & More, A.W. (2018). Molecular marker techniques: a review. *Int.J.Curr.Microbiol.App.Sci*, Special Issue-6, 816-825.
- de Villemereuil, P., & Gaggiotti, O.E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods Ecol. Evol.*, 6, 1248-1258.
- Dong, K. & Opperman, C.H. (1997). Genetic analysis of parasitism in the soybean cyst nematode, *Heterodera glycines*. *Genetics*, 146,1311-1318.
- Doucet M.E., Lax P. & Coronel, N. (2008) The Soybean Cyst Nematode *Heterodera Glycines* Ichinohe, 1952 in Argentina. In: Ciancio A., Mukerji K.G. (eds) Integrated Management and Biocontrol of Vegetable and Grain Crops Nematodes. Integrated Management of Plant Pests and Diseases, vol 2. Springer, Dordrecht.
- Edwards D., Forster J.W., Chagné D. & Batley J. (2007). What are SNPs? In Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E., De Silva, H. (Eds). Association mapping in plants. New York : Springer.
- Eisenback, J. D. (2018). Virginia Polytechnic Institute and State University, Bugwood.org. Retrieved from <https://www.invasive.org/browse/detail.cfm?imgnum=5442354> (12-08-19).
- Eizaguirre, C. & Baltazar-Soares, M. (2014) Evolutionary conservation - evaluating the adaptive potential of species. *Evol. Appl.*, 7, 963-967.
- Elad, Y. & Pertot, I. (2014). Climate change impacts on plant pathogens and plant diseases. *J. Crop Improv.*, 28, 99-139.
- Eves-van den Akker, S., Laetsch, D.R., Thorpe, P., Lilley, C.J., Danchin, E.G., Da Rocha, M., ... Jones, J.T. (2016). The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biol.*, 17(1), 124.
- Ewens, W. J. (2004). Mathematical Population Genetics. 2nd. New York : Springer.
- Faghihi, J., Donald, P. A., Noel, G., Welacky, T. W & Ferris, V. R. (2010). Soybean resistance to field populations of *Heterodera glycines* in selected geographic areas. *Plant Health Prog.* doi:10.1094/PHP-2010-0426-01-RS.
- FAOSTAT. 2019. Données - Cultures. FAO database. Retrieved from <http://www.fao.org/faostat/en/#data> (30-01-2019).
- Faghihi, J. & Ferris, V. R. (2017). Soybean cyst nematode. Retrieved from <https://extension.entm.purdue.edu/publications/E-210/E-210.pdf> (12-09-2019).

- Feng, X.-J., Jiang, G.-F., & Fan, Z. (2015). Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, *Ceracris kiangsu*. *Sci. Rep.* 5, 13758.
- Fischer M.C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K.K. ... Widmer, A. (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Mol. Ecol.*, 22, 5594–5607.
- Foll, M. & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180, 977–993.
- Gao, B., Allen, E., Davis, E.L., Baum, T.J. & Hussey, R.S. (2004). Molecular characterisation and developmental expression of cellulose-binding protein gene in soybean cyst nematode *Heterodera glycines*. *Int. J. Parasitol.*, 34,1377-1383.
- Gao, B., Allen, R., Maier, T., Davis, E.L., Baum, T.J. & Hussey, R.S. (2001). Molecular characterization and expression of two venom allergen-like protein genes from *Heterodera glycines*. *Int. J. Parasitol.*, 31, 1617-1625.
- Gao, B., Allen, R., Maier, T., Davis, E.L., Baum, T.J. & Hussey, R.S. (2003). The parasitome of the phytonematode *Heterodera glycines*. *Mol. Plant-Microbe Interact.*, 16,720-726.
- Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, 201(4), 1555-1579.
- Gautier, M., Hocking, T.D. & Foulley, J.L. (2010). A Bayesian Outlier Criterion to Detect SNPs under Selection in Large Data Sets. *PLOS ONE* 5(8): e11913.
- Gardner, M., Heinz, R., Wang, J. & Mitchum, M. (2018). Genetics and Adaptation of Soybean Cyst Nematode to Broad Spectrum Soybean Resistance. *G3 (Bethesda)*, 7(3), 835–841.
- Gems, D. & Maizels, R. M. (1996) An abundantly expressed mucin-like protein from *Toxocara canis* infective larvae: the precursor of the larval surface coat glycoproteins. *Proc.Nat. Aca. Sci.*, 93, 1665.
- Gendron St-Marseille, A.-F. (2013). Le nématode à kyste du soja (*Heterodera glycines*) : enjeux des changements climatiques sur sa distribution, sa reproduction et sur les probabilités de synchronisme avec le soja (*Glycine max*) au Québec. Essai. Université de Sherbrooke, Sherbrooke.
- Giller, K.E. & Dashiell, K.E. (2007). *Glycine max* (L.) Merr. In van der Vossen, H.A.M. et Mkamilo, G.S. (Eds). Wageningen, Netherlands : PROTA (Plant Resources of Tropical Africa / Ressources végétales de l’Afrique tropicale).
- Gomez-Tovar, J. & Medina, C. (1983) *Heterodera glycines* on soybean in Valle Cauca, Colombia. *Nematropica* 13, 229-237 (Spanish).

- Grulke, N.E. (2011). The nexus of host and pathogen phenology : understanding the disease triangle with climate change. *New Phytol.*, 189, 8–11.
- Guidez, F. (2014). Rétrotransposons et cellules somatiques - Parasites ou acteurs actifs du contrôle épigénétique ? *Med. Sci.*,30, 659-664.
- Guo, X., Wang, J., Gardner, M., Fukuda, H., Kondo, Y., Etchells, J. P., ... Mitchum, M. G. (2017). Identification of cyst nematode B-type CLE peptides and modulation of the vascular stem cell pathway for feeding cell formation. *PLoS pathogens*, 13(2), e1006142.
- Haegeman, A., Kyndt, T. & Gheysen, G. (2010). The role of pseudo-endoglucanases in the evolution of nematode cell wall modifying proteins. *J. Mol. Evol.* 16, 720-726.
- Haegeman, A., Mantelin, S. Jones, J.T. & Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene*, 492,19-31.
- Hamamouch, N., Li, C., Davis, E.L., Hewezi, T., Baum, T.J., Mitchum, M.G., ... Vodkin, L.O. (2012). The interaction of the novel 30C02 cyst nematode effector protein with a plant β -1,3-endoglucanase may suppress host defence to promote parasitism, *J. Exp. Bot.*, 63, 3683–3695.
- Hartman, G.L., West, E.D. & Herman, T.K. (2011). Crops that feed the world 2. Soybean – worldwide production, use, and constraints caused by pathogens and pests. *Food Sec.*, 3, 5-17.
- Hartwig, E. E., Epps, J. M., & Edwards, C. J., Jr. (1971). Registration of Pickett 71 soybeans. *Crop Sci.*, 11, 603.
- Hayward, A.C., Tollenaere, R., Dalton-Morgan, J. & Batley, J. (2015). Molecular marker applications in plants. In Batley, J. (ed.), *Plant Genotyping: Methods and Protocols*. New York : Springer.
- Heather JM & Chain B (2016). The sequence of sequencers: the history of sequencing DNA. *Genomics*, 107, 1–8.
- Hogenhout, S.A., Van der Hoorn, R.A., Terauchi, R. & Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant-Microbe Interact.*, 22,115-122.
- Hulme, P.E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.*, 46, 10-18.
- Hutchison, D. & Templeton, A. (1999). Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evol.*, 53(6), 1898-1914.

- Ithal, N., Recknor, J., Nettleton, D., Hearne, L., Maier, T., Baum, T.J. & Mitchum, M.G. (2007). Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. *Plant-Microbe Interact.*, 20, 293-305.
- Jones, J.T., Reavy, B., Smant, G. & Prior, A.E. (2004). Glutathione peroxidases of the potato cyst nematode *Globodera rostochiensis*. *Gene* 324: 47–54.
- Joyce, A. & Thiessen, L. (2018). Soybean Cyst Nematode. Retrieved from <https://content.ces.ncsu.edu/management-of-soybean-cyst-nematode> (08-05-2019).
- Kakaire, S., Grove, I.G. & Haydock, P.P.J. (2012). Effect of temperature on the life cycle of *Heterodera schachtii* infecting oilseed rape (*Brassica napus* L.). *Nematol.*, 14,855-867.
- Karlik, E. & Tombuloğlu, H. (2016). Molecular markers and their applications, In Hakeem, K.R., Tombuloğlu, H., Tombuloğlu, G. (Eds.), *Plant omics: trends and applications*. Springer International Publishing.
- Kim, M., Hyten, D.L., Bent, A.F. & Diers, B.W. (2010). Fine mapping of the SCN resistance locus *rhg1-b* from PI 88788. *Plant Genome*, 3,81-89.
- Kim, D., Riggs, R.D., Robbins, R.T., & Rakes, L.C. (1997). Distribution of Races of *Heterodera glycines* in the Central United States. *J. Nematol.*, 29 2, 173-179.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217, 624–626.
- Klink, V.P., Hosseini, P., MacDonald, M.H., Alkharouf, N.W. & Matthews, B.F. (2009). Population-specific gene expression in the plant pathogenic nematode *Heterodera glycines* exists prior to infection and during the onset of a resistant or susceptible reaction in the roots of the *Glycine max* genotype Peking. *BMC Genomics*, 10,111.
- Klink, V.P., P.D. Matsye, K.S. Lawrence & G.W. Lawrence 2013. Engineered soybean cyst nematode resistance. In El-Shemy, H. (Ed.) *Soybean - Pest Resistance*, InTech.
- Koenning, S.R. 2004. Population biology. In Schmitt, D.P. Wrather, J.A. & Riggs, R.D. (Eds). *Biology and management of soybean cyst nematode*, 2nd ed. Marceline : Schmitt and Associates of Marceline.
- Kofler, R., Pandey, R.V. & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, 27(24), 3435–3436.
- Kohn, M. H., Pelz, H. J. & Wayne, R. K. 2003. Locus-specific genetic differentiation at *Rw* among warfarin-resistant rat (*Rattus norvegicus*) populations. *Genetics*, 164, 1055–1070.

- Lal, R. & Lal, A (2005). Phytosanitary importance of *Heterodera glycines* for South-east Asian. *Bulletin OEPP* 35, 531-536.
- Lawn, D.A. & Noel, G.R. (1986). Field interrelationships among *Heterodera glycines*, *Pratylenchus scribneri*, and three other nematode species associated with soybean. *J. Nematol.*, 18,98-105.
- Lee, C., Chronis, D., Kenning, C., Peret, B., Hewezi, T., Davis, E.L. & Mitchum, M.G. (2011). The novel cyst nematode effector protein 19c07 interacts with the *Arabidopsis* auxin influx transporter lax3 to control feeding site development. *Plant Physiol.*, 155(2), 866-880.
- Li, J., Faghihi, J., Ferris, J.M., & Ferris, V.R. (1996). The use of RAPD amplified DNA as markers for virulence characteristics in soybean cyst nematode. *Fundam. Appl. Nematol.*, 19, 143-150.
- Li, Y., Wang, K., Xie, H., Wang, Y. T., Wang, D. W., Xu, C. L., ... Wang, D. S. (2015). A nematode calreticulin, rs-crt, is a key effector in reproduction and pathogenicity of *Radopholus similis*. *PloS one*, 10(6), e0129351.
- Lilley, C. J., Goodchild, S. A., Atkinson, H. J. & Urwin, P. E. (2005) Cloning and characterisation of a *Heterodera glycines* aminopeptidase cDNA. *Int. J. Parasitol.*, 35, 1577-1585.
- Liu, L., Ang, K.P., Elliott, J.A.K., Kent, M.P., Lien, S., MacDonald, D., & Boulding, E.G. (2017). A genome scan for selection signatures comparing farmed Atlantic salmon with two wild populations: Testing colocalization among outlier markers, candidate genes, and quantitative trait loci for production traits. *Evol Appl.*, 10, 276– 296.
- Liu, X. J. Jian, W. Guanghua, and S.J. Herbert. (2008). Soybean yield physiology and development of high- yielding practices in Northeast China. *Field Crops Res.*, 105, 157-171.
- Liu, S., Kandath, P.K., Warren, S.D., Yeckel, G., Heinz, R., Alden, J., ... Meksem, K. (2012). A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature*, 492, 256-260.
- Liu X.Z., Li J.Q., Zhang D.S. (1997). History and status of soybean cyst nematode in China. *Int. J. Nematol.*,7, 18–25.
- Loewe, L. & Hill, W.G. (2010). The population genetic of mutation : good, bad and indifferent. *Philos Trans R Soc Lond B Biol Sci.*, 365, 1153–1167.
- Lozano-Torres, J.L., Wilbers, R.H.P., Warmerdam, S., Finkers-Tomczak, A., Diaz-Granados, A., van Schaik, C.C, ... Smant, G. (2014). Apoplastic venom allergen-like proteins of cyst nematodes modulate the activation of basal plant innate immunity by cell surface receptors. *PLoS Pathog.*, 10(12), e1004569.

- Luikart, G., England, P.R., Tallmon, D., Jordan, S. & Taberlet, P. (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Rev. Genet.*, 4, 981-994.
- Matson, A.L. & Williams, L.F. (1965). Evidence of a fourth gene for resistance to soybean cyst nematode. *Crop Sci.*, 5, 477.
- Mei, Y., Thorpe, P., Guzha, A., Haegeman, A., Blok, V., Mackenzie, K., ... Mantelin, S. (2015). Only a small subset of the SPRY domain gene family in *Globodera pallida* is likely to encode effectors, two of which suppress host defences induced by the potato resistance gene Gpa2. *Nematol.*, 17, 409–424.
- Mei, Y., Wright, K.M., Haegeman, A., Bauters, L., Diaz-Granados, A., Govere, A., ... Mantelin, S. (2018). The *Globodera pallida* SPRYSEC effector GpSPRY-414-2 that suppresses plant defenses targets a regulatory component of the dynamic microtubule network. *Front. Plant Sci.*, 9, 1019.
- Mimee, B., Brodeur, J., Bourgeois, G., Moiroux, J., Gendron St-Marseille, A.-F. & Gagnon, A.-E. (2014a). Quels enjeux représentent les changements climatiques en lien avec les espèces exotiques envahissantes pour la culture du soya au Québec ? Retrieved from http://www.ouranos.ca/media/publication/332_RapportBrodeur2014.pdf
- Mimee, B., Peng, H., Popovic, V., Duceppe, M.-O., Tetreault, M.-P. & Belair, G. (2014b). First report of soybean cyst nematode (*Heterodera glycines* Ichinohe) on soybean in the Province of Quebec, Canada. . *Plant Dis.*, 98,429.
- Mourtzinis, S. & Conley, S.P. (2017). Delineating soybean maturity groups across the United States. *Agron. J.*, 109(4), 1397-1403.
- Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., ... Baloch, F. S. (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. & Biotech. Eq.*, 32, 261-285.
- Nelson, B. D. & Bradley, C. A. (2003). Soybean cyst nematode (SCN). Retrieved from <https://www.ndsu.edu/pubweb/~bernelso/soydiseases/cyst.shtml> (12-08-19).
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* 70: 3321–3323.
- Niblack, T.L. (2005). Soybean cyst nematode management reconsidered. *Plant Dis.*, 80,1020-1026.
- Niblack, T.L., Arelli, P.R., Noel, G.R., Opperman, C.H., Orf, J.H. Schmitt, D.P. ... Tylka, G.L. (2002). A revised classification scheme for genetically diverse populations of *Heterodera glycines*. *J. Nematol.*, 34,279-288.

- Niblack, T.L., Colgrove, A. L., Colgrove, K. and Bond, J. P. (2008). Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. Online. Plant Health Progress doi:10.1094/PHP-2008-0118-01-RS.
- Niblack, T.L., Lambert, K.N. & Tylka, G.L. (2006). A model plant pathogen from the kingdom animalia : *Heterodera glycines*, the soybean cyst nematode. Annu. Rev. Phytopathol. 44:283-303.
- Noel, G.R. (2004). Soybean response to infection. In Schmitt, D.P. Wrather, J.A. & Riggs, R.D. (Eds.). Biology and management of soybean cyst nematode. (2nd ed. Pp. 131-151). Marceline : Schmitt and Associates of Marceline.
- Noon, J.B., Hewezi, T., Maier, T.R., Simmons, C., Wei, J.-Z., Wu, G., ... Baum, T.J. (2015). Eighteen new candidate effectors of the phytonematode *Heterodera glycines* produced specifically in the secretory esophageal gland cells during parasitism. *Phytopathology*, 105, 1362–1372.
- Oyekanmi, E.O. & Fawole, B. (2010). Nematodes of soybean and their management. In Singh, G. (Ed.). The soybean: botany, production and uses. Oxfordshire : CAB International.
- Patel, N., Hamamouch, N., Li, C. Y., Hewezi, T., Hussey, R. S., Baum, T. J., ... Davis, E. L. 2010. A nematode effector protein similar to annexins in host plants. *J. Exp. Bot.*, 61, 235-248.
- Patel, D.A., Zander, M., Dalton-Morgan, J. & Batley, J. (2015). advances in plant genotyping: where the future will take us. In Batley, J. (Ed.), Plant Genotyping: Methods and Protocols. Springer New York, New York.
- Patil, G. B., Lakhssassi, N., Wan, J., Song, L., Zhou, Z., Klepadlo, M. ... Nguyen, H. T. (2019). Whole-genome re-sequencing reveals the impact of the interaction of copy number variants of the rhg1 and rhg4 genes on broad-based rsistance to soybean cyst nematode. *Plant Biotechnol. J.*, 17, 1595-1611.
- Pautasso, M., Döring, T.F., Garbelotto, M., Pellis, L. & Jeger, M.J. (2012). Impacts of climate change on plant diseases—opinions and trends. *Eur. J. Plant Pathol.*, 133, 295-313.
- Perry, R.N. & Moens, M. (2011). Survival of parasitic nematode outside the host. In Perry, R.N. & Wharton, D.A. (Eds.). Molecular and physiological basis of nematode survival. Wallingford : CABI Publishing.
- Pilot, M., Malewski, T., Moura, A. E., Grzybowski, T., Oleński, K., Kamiński, S., ... Bogdanowicz, W. (2016). Diversifying selection between pure-breed and free-breeding dogs inferred from genome-wide SNP analysis. *G3-Genes Genom. Genet.*, 6(8), 2285-2298.

- Postma, W.J., Sloopweg, E.J., Rehman, S., Finkers-Tomczak, A., Tytgat, T.O., van Gelderen, K., ... Smant, G. 2012. The effector SPRYSEC-19 of *Globodera rostochiensis* suppresses CC-NB-LRR-mediated disease resistance in plants. *Plant Physiol.*, 160(2), 944-54.
- Prior, A., Jones, J.T., Blok, V.C., Beauchamp, J., McDermott, L., Cooper, A. & Kennedy, M.W. (2001). A surface-associated retinol- and fatty acid-binding protein (Gp-FAR-1) from the potato cyst nematode *Globodera pallida*: lipid binding activities, structural analysis and expression pattern. *Biochem. J.*, 356, 387-394.
- Qi, T., Guo, J., Peng, H., Liu, P., Kang, Z., & Guo, J. (2019). Host-induced gene silencing: a powerful strategy to control diseases of wheat and barley. *Int. J. Mol. Sci.*, 20, 206.
- Qin, L., Kudla, U., Roze, E.H., Govere, A., Popeijus, H., Nieuwland, J., ... Helder, J. (2004). A nematode expansin acting on plants. *Nature*, 427, 30.
- Rao-Arelli, A.P. (1994). Inheritance of resistance to *Heterodera glycines* race 3 in soybean accession. *Plant Dis.*, 78, 898-900.
- Rehman, S., Postma, W., Tytgat, T., Prins, P., Qin, L., Overmars, H., ... Smant, G. (2009). A secreted SPRY domain-containing protein (SPRYSEC) from the plant-parasitic nematode *Globodera rostochiensis* interacts with a CC-NB-LRR protein from a susceptible tomato. *Mol Plant Microbe Interact.*, 22(3), 330-40.
- Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M. & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Mol Ecol*, 24, 4348-4370.
- Riggs, R.D. (1977). Worldwide distribution of soybean-cyst nematode and its economic importance. *J. Nematol.* 9, 34-39.
- Rogers, H.H., Peterson, C.M., McCrimmon, J.N. & Cure, J.D. (1992). Response of plant roots to elevated atmospheric carbon dioxide. *Plant Cell Environ.* 15, 749-752.
- Sacco, M.A., Koropacka, K., Grenier, E., Jaubert, M. J., Blanchard, A., Govere, A., ... Moffett, P. (2009). The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death. *PLoS Pathog.* 5(8):e1000564.
- Schiessl, S.-V., Kathe, E., Ihlen, E., Singh Chawla, H. & S. Mason, A. (2019). The role of genomic structural variation in the genetic improvement of polyploid crops, *Crop J.*, 7, 127-140.
- Schmitt, D.P. (2004). Introduction. In Schmitt, D.P., Wrather, J.A. & Riggs, R.D. (Eds.). *Biology and management of soybean cyst nematode*. 2nd ed. Marceline : Schmitt and Associates of Marceline.
- Singh, G. & Shivakumar, B.G. (2010). The role of soybean in agriculture. In Singh, G. (Ed.). *The soybean : botany, production and use*, 1st ed. Oxfordshire : CAB International.

- Slack, D.A., Riggs, R.D et Hamblen, M.L (1972). The effect of temperature and moisture of *Heterodera glycines* in absence of the host. *J. Nematol.*, 4 (4): 263-266.
- Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236, 787-792.
- Somasekhar, N. & Prasad, J.S. (2012). Plant-nematode interactions : consequences of climate change. In Venkateswarlu, B., Shanker, A.K., Shanker, C. & Maheswari, M. (Eds.). Crop stress and its management: perspectives and strategies. Houten : Springer Netherlands.
- Spencer, C.C., Deloukas, P., Hunt, S., Mullikin, J., Myers, S., Silverman, B., ... McVean, G. (2006). The influence of recombination on human genetic diversity. *PLoS genetics*, 2(9), e148.
- Statistiques Canada (2019). Table 32-10-0359-01 - Estimated areas, yield, production, average farm price and total farm value of principal field crops, in metric and imperial units. Retrieved from <https://www150.statcan.gc.ca/t1/tb11/en/cv.action?pid=3210035901> (14-03-2019).
- Stinchcombe J.R. & Hoekstra, H.E. (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, 100, 158-170.
- Subbotin S.A., Mundo-Ocampo M. & Baldwin J.G. (2010). Description and diagnosis of *Heterodera* species - 29. *Heterodera glycines* Ichinohe, 1952. Systematics of Cyst Nematodes (Nematoda: Heteroderinae). *Nematol. Monogr. Perspect.* 8, Part B, 195-204.
- Tattini, L., D'Aurizio, R., & Magi, A. (2015). Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Frontiers in bioengineering and biotechnology*, 3, 92.
- Teama, S. (2018). DNA polymorphisms: DNA-based molecular markers and their application in medicine. In : Liu, Y. (Ed.) Genetic diversity and disease susceptibility, Chap. 2, 25-40. IntechOpen.
- Tylka, G.L. (2012). Soybean cyst nematode field guide – A reference for identifying, scouting for, and managing soybean cyst nematode. Retrieved from <https://store.extension.iastate.edu/Product/Soybean-Cyst-Nematode-Field-Guide> (02-07-2018).
- Tylka, G.L. & Marett, C.C. (2014). Distribution of the soybean cyst nematode, *Heterodera glycines*, in the United States and Canada: 1954 to 2014. *Plant Health Prog.*, 15, 13-15.

- Upchurch, R.G. & Ramirez, M.E. (2011). Effects of temperature during soybean seed development on defense-related gene expression and fungal pathogen accumulation. *Biotechnol. Lett.*, *33*, 2397-2404.
- United States Department of Agriculture, USDA (2019). Production, Supply and distribution online. Retrieved from <https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery> (30-01-2019).
- Vanderplank J.E. (1978) The Gene-for-Gene and the Protein-for-Protein Hypotheses. In: Genetic and Molecular Basis of Plant Pathogenesis. Advanced Series in Agricultural Sciences, vol 6. Springer, Berlin, Heidelberg.
- Vanholme, B., Mitreva, M., Van Criekinge, W., Logghe, M., Bird, D. & McCarter, J. P. (2006). Detection of putative secreted proteins in the plant-parasitic nematode *Heterodera schachtii*. *Parasitol. Res.*, *98*(5), 414-424.
- Venkatesh, R., Harrison, S. K. and Riedel, R. M. (2000). Weed hosts of soybean cyst nematode (*Heterodera glycines*) in Ohio. *Weed Technol.*, *14*, 156–160.
- Vigouroux, Y., McMullen, M., Hittinger, C. T., Houchins, K., Schulz, L., Kresovich, S., ... Doebley, J. (2002). Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc. Natl. Acad. Sci. USA*, *99*, 9650–9655.
- Wang, J., Lee, C., Replogle, A., Joshi, S., Korkin, D., Hussey, R.S., ... Mitchum, M.G. (2010). Dual roles for the variable domain in protein trafficking and host-specific recognition of *Heterodera glycines* CLE Effector proteins. *New Phytol.*, *187*, 1003-1017.
- Wang, X., Meyers, D., Yan, Y., Baum, T., Smant, G., Hussey, R. & Davis, E. (1999). In planta localization of a β -1,4-endoglucanase secreted by *Heterodera glycines*. *Mol. Plant-Microbe Interact.*, *12*, 64-67.
- Wang, J., Niblack, T.L., Tremain, J.A., Wiebold, W.J., Tylka, G.L., Marett, C.C., Noel, G.R., Myers, O. & Schmidt, M.E. (2003). Soybean cyst nematode reduces soybean yield without causing obvious aboveground symptoms. *Plant Dis.*, *87*, 623-628.
- Wang, G., D. Peng, B. Gao, W. Huang, L. Kong, H. Long, P. Peng & H. Jian. (2014). Comparative transcriptome analysis of two races of *Heterodera glycines* at different developmental stages. *PLoS ONE*, *9*.
- Wootton, J. C., Feng, X. R., Ferdig, M. T., Cooper, R. A., Mu, J. B., Baruch, D. I., Magill, A. J. & Su, X. Z. (2002). Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* *418*: 320–323.

- Wrather, J.A., Shannon, G., Balardin, R., Carregal, L., Escobar, R., Gupta, G.K., ... Tenuta, A. (2010). Effect of diseases on soybean yield in the top eight producing countries in 2006. *Plant Health Prog.* doi: 10.1094/PHP-20100125-01-RS.
- Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, 16, 97–159.
- Wright, S. (1951). The genetical structure of populations. *Ann. Eugen.*, 15, 323–354.
- Wright, S. (1978). Evolution and the Genetics of Populations, Vol. IV. Variability Within and Among Natural Populations. Chicago : University of Chicago Press.
- Yan, G. & Baidoo, R. (2018). Current Research Status of *Heterodera glycines* Resistance and Its Implication on Soybean Breeding. *Engineering*, 4, 534-541.
- Yu, Q. (2011). Soybean Cyst Nematode (*Heterodera glycines* Ichinohe). In El-Shemy, H (Éd.) Soybean physiology and biochemistry, (pp. 461-474). Rijeka : InTech.
- Yuan, C.P., Li, Y.H., Liu, Z.X., Guan, R.X., Chang, R.Z. & Qiu, L.J. (2012). DNA sequence polymorphism of the Rhg4 candidate gene conferring resistance to soybean cyst nematode in Chinese domesticated and wild soybeans. *Mol. Breed.*, 30,1155-1162.
- Zheng, J., Li, Y. & Chen, S. (2006). Characterization of the virulence phenotypes of *Heterodera glycines* in Minnesota. *J. Nematol.*, 38:383-390.
- Zheng J. Z., Chen S. Y. (2011). Estimation of virulence type and level of soybean cyst nematode field populations in response to resistant cultivars. *J. Entomol. Nematol.*, 3, 37–43.
- Ziska, L.H. (1998). The influence of root zone temperature on photosynthetic acclimation to elevated carbon dioxide concentrations. *Ann. Bot.*, 81, 717-721.

Annexe 1 - Autres activités et accomplissements durant le doctorat

Articles et rapports publiés

Gendron St-Marseille, A.-F. Un ver pas ordinaire. (2016). Premier prix dans la catégorie doctorat, Bourse Pehr-Kalm 2016 – Les Amis du Jardin botanique de Montréal Retrieved from <https://static.amisjardin.com/data/Originals/5bea410d-5e0b-4def-b825-4437998ae15e.pdf> (28-08-2017)

Gendron St-Marseille, A.-F. (2016) Des partenaires microscopiques hébergés dans les insectes : les endosymbiotes. *Dire* 25 (1), p. 31-36. Retrieved from <https://www.ficsum.com/dire-archives/hiver-2016/des-partenaires-microscopiques-heberges-dans-les-cellules-les-endosymbiotes/> (28-08-2017).

Gendron St-Marseille, A.-F., Bélair, G., Brodeur, J., Bourgeois, G. et Mimee, B. (2015). Le nématode à kyste du soya et les changements climatiques. *Phytoprotection* 95 (1), p. 41–47.

Mimee, B., Brodeur, J., Bourgeois, G., Moiroux, J., Gendron St-Marseille, A.-F. et Gagnon, A.-È. 2013. Quels enjeux représentent les changements climatiques en lien avec les espèces exotiques envahissantes pour la culture du soya au Québec? Rapport 550012. Ouranos. 53 p.

Présentations orales à des congrès scientifiques (= présentateur)*

Gendron St-Marseille, A.-F.*, Brodeur, J. et Mimee, B. (2017). Diversité génétique, structure des populations et adaptation du nématode à kyste du soya en Amérique du Nord, Congrès de la SPPQ 2017 «Phytoprotection: Pilier d'un agro-écosystème durable», Campus MacDonald de l'Université McGill, Sainte-Anne-de-Bellevue, Québec, Canada, 8-9 juin 2017.

2016 Soybean Cyst Nematode Conference of The American Phytopathological Society, Coral Gables, Florida, U.S.A., December 13–15, 2016. (participation)

Gendron St-Marseille, A.-F.*, Véronneau, P.-Y., Lafond-Lapalme, J., Brodeur, J. and Mimee, B. (2016). Intraspecific genetic diversity of the soybean cyst nematode, SOYAGEN meeting, Université Laval, Québec, Québec, Canada, 30 novembre – 2 décembre 2016.

Gendron St-Marseille, A.-F.*, Véronneau, P.-Y., Lafond-Lapalme, J., Brodeur, J. and Mimee, B. (2016). Structure génétique et effet de l'isolement par la distance des populations nord-américaines du nématode à kyste du soya, *Heterodera glycines*, Journée du Centre Sève, Hôtel Château Bromont, Bromont, Québec, Canada, 24-25 novembre 2016.

Gendron St-Marseille, A.-F.*, Véronneau, P.-Y., Lafond-Lapalme, J., Brodeur, J. and Mimee, B. (2016). Using single nucleotide polymorphism to evaluate intraspecific diversity and genetic

determinants of virulence of the soybean cyst nematode, *Heterodera glycines*, European Society of Nematology 2016 conference, University of Minho, Braga, Portugal, august 28th to september 1st 2016.

- Gagnante d'une bourse pour effectuer une présentation étudiante à l'étranger.

Gendron St-Marseille, A.-F.*, Mimee, B., Bourgeois, G. and Brodeur, J. (2015). Impacts of climate change on soybean phenology and on the soybean cyst nematode life cycle. The Annual Meeting of the Society of Nematologists, Michigan State University, East Lansing, United Statesm 19th-24th july 2015.

Gendron St-Marseille, A.-F.*, Mimee, B. et Brodeur, J. (2015). Génétique des populations et utilisation des marqueurs SNP pour différencier les pathotypes du nématode à kyste du soya. Congrès de la SPPQ 2015 «La révolution génomique au service de la phytoprotection», Château Mont St-Anne, Beaupré, Québec, Canada, 16-18 juin 2015.

- Gagnante de la meilleure présentation étudiante.

Gendron St-Marseille, A.-F.*, Mimee, B., Bourgeois, G. et Brodeur, J. (2014). Modélisation bioclimatique du soya et cycle de vie du nématode à kyste du soya (*Heterodera glycines*) dans un contexte de changements climatiques. Congrès de la SPPQ 2014 «Les plantes bio-industrielles – enjeux environnementaux et phytosanitaires», Hôtel les trois tilleuls, St-Marc-sur-Richelieu, Québec, Canada, 10-11 juin 2014.

Prix et bourses

Bourse d'excellence de la FESP 2019 FESP, Université de Montréal	Juin 2019
Bourse de la réussite étudiante 2018-2019 Université de Montréal	Avril 2019
Bourse d'excellence académique Université de Montréal	Octobre 2017
Bourse Pehr-Kalm – 1^{er} prix, rédaction d'un article scientifique Les Amis du Jardin Botanique de Montréal	Janvier 2017
Bourse Jacques-Rousseau pour participer à un congrès Avril 2016 Institut de recherche en biologie végétale – IRBV	
Bourse pour faire une présentation à un congrès Centre Sève	Avril 2016

ESN Student Bursary Award European Society of Nematology	Avril 2016
Bourse W. E. Sackston pour la meilleure présentation étudiante Société de Protection des plantes du Québec – SPPQ	juin 2015
Bourse pour faire une présentation à un congrès Centre Sève	Juin 2015

Annexe 2 - Autre article publié et soumis à un comité de lecture durant le doctorat

Revue de littérature

Impact des changements climatiques sur les interactions moléculaires entre le nématode à kyste du soya (*Heterodera glycines*) et son hôte principal, le soya (*Glycine max*).

Anne-Frédérique Gendron St-Marseille¹, Guy Bélair¹, Jacques Brodeur², Gaétan Bourgeois¹ et Benjamin Mimee^{1*}

publié tel que : Gendron St-Marseille, A.-F., Bélair, G., Brodeur, J., Bourgeois, G. et Mimee, B. (2015). Le nématode à kyste du soya et les changements climatiques. *Phytoprotection* 95 (1), p. 41–47.

RÉSUMÉ

Au Canada, le nématode à kyste du soya, *Heterodera glycines* (NKS), a d'abord été détecté en Ontario en 1988 et plus récemment au Québec en 2013. Il est la principale cause des pertes économiques associées à la production de soya au Canada et aux États-Unis. Les moyens de lutte contre ce ravageur sont limités et reposent essentiellement sur l'exploitation des gènes de résistance naturels du soya. Il est primordial d'étudier la réponse du NKS ainsi que celle de son hôte sous le climat du Québec afin de développer des stratégies de lutte efficaces. De plus, les fluctuations de température associées aux changements climatiques modifieront les traits biologiques du NKS, la phénologie de son hôte et leur interaction. Cela pourrait ainsi affecter l'efficacité et la durabilité des outils de gestion. Dans ce contexte, l'impact des changements climatiques anticipés sur les interactions entre le nématode à kyste du soya et son hôte principal, le soya, sera présenté.

Mots clés : Changements climatiques, Heterodera glycines, interactions moléculaires, soya.

ABSTRACT

[Impact of climate change on molecular interactions between the soybean cyst nematode (*Heterodera glycines*) and its main host, soy (*Glycine max*)]

In Canada, the soybean cyst nematode, *Heterodera glycines* (SCN), was first detected in Ontario in 1988 and more recently in Quebec in 2013. It is the main cause of economic loss associated with soy production in Canada and the United States. Control methods against this pest are limited and mostly rely on soy's natural resistance genes. It is essential to study the response of the SCN and that of its host under Quebec's climate in order to develop effective control strategies. Moreover, the temperature fluctuations associated with climate change will modify SCN's biological characters, its host's phenology, and their interaction. This could affect both the efficacy and durability of management tools. In this context, the impact of the anticipated climate change on the interactions between the soybean cyst nematode and its principal host, soy, will be presented.

Keywords: Climate change, molecular interactions, soy, soybean cyst nematode.

La production du soya (*Glycine max*)

Les superficies de soya (*Glycine max* L. Merr) sont en forte augmentation à travers le monde (Ali, 2010). Cette hausse se justifie par les multiples utilisations des huiles et des farines riches en protéines issus de cette oléo-protéagineuse dans l'industrie alimentaire humaine et animale (Ali, 2010; Singh et Shivakumar, 2010). En 2013, le soya était la septième culture en importance dans le monde, au premier rang des cultures oléagineuses (FAOSTAT, 2014). Les principaux pays producteurs sont les États-Unis suivis par le Brésil, l'Argentine, la Chine et l'Inde (FAOSTAT, 2014) (Tableau 1). En 2013, le Canada occupait respectivement la huitième et la cinquième position mondiale au niveau de la production et de l'exportation (Food and Agriculture Organization (FAO), 2013). La production de soya se concentre dans les provinces de l'Ontario

(59 % de la production totale), du Manitoba (21 %), du Québec (16%), de la Saskatchewan (2 %) et des Maritimes (2 %) (Statitics Canada, 2014).

Tableau 1 : Production mondiale de soya en 2013 (FAOSTAT, 2014).

Ran	Pays	Production (MT)*	Superficies cultivées (Mha)**
1	États-Unis	89,5	30,7
2	Brésil	81,7	27,9
3	Argentine	49,3	19,4
4	Chine	12,5	6,6
5	Inde	11,9	12,2
6	Paraguay	9,1	3,1
7	Ukraine	3,2	1,4
8	Canada	5,2	1,8
9	Uruguay	2,8	1,2
10	Bolivie	2,3	1,2

* Million de tonnes.

** Million d'hectares.

Avec l'intensification de la production et l'augmentation des superficies cultivées en soya, le nombre d'organismes nuisibles pouvant occasionner des pertes économiques a également augmenté (Canadian Soybean Council (CSC), 2012). Ce phénomène s'explique principalement par l'accroissement des échanges commerciaux et par l'adoption de nouvelles cultures qui ont favorisé l'introduction, l'établissement et la reproduction de certains ravageurs exotiques et indigènes (Hartman *et al.*, 2011). Dans les principaux pays producteurs, les pertes économiques les plus importantes sont associées à la présence des champignons de la rouille *Phakospora pachyrhizi* et *P. meibomia* suivi par le nématode à kyste du soya (NKS) *Heterodera glycines* (Anderson *et al.*, 2004; Hulme, 2009; Wrather *et al.*, 2010). Aux États-Unis, les états centraux qui produisent la majorité du soya sont déjà aux prises avec le NKS, mais en plus, ils doivent maintenant affronter la venue du puceron du soya en provenance du nord depuis 2000 et celle de

la rouille en provenance du sud depuis 2004 (Lee *et al.*, 2006). Au Canada, le NKS est présent principalement dans le sud de l'Ontario depuis 1987, et ce n'est qu'en 2004 que le puceron du soya a été détecté dans trois provinces canadiennes (Manitoba, Ontario et le Québec) (Ragsdale *et al.*, 2011; Tilmon *et al.*, 2011). Bien qu'à ce jour, la rouille du soya n'a pas encore traversé la frontière canadienne, une étude menée sur les espèces invasives démontre qu'à la faveur des changements climatiques, ce champignon pourrait atteindre le Canada prochainement (Ariatti, 2014; Mimee *et al.*, 2014a).

Le nématode à kyste du soya (*Heterodera glycines*)

Le NKS est un endoparasite obligatoire des racines originaire d'Asie qui a été introduit dans la plupart des pays producteurs de soya (Wrather *et al.*, 2010). Ce parasite peut réduire de 15 à 85 % le rendement des cultures (CAB International (CABI), 2013). À l'échelle mondiale, les pertes économiques associées à ce nématode sont estimées à 1,6 G\$ US annuellement (Wang *et al.*, 2003). Au Canada, de 1994 à 2006, les pertes économiques ont augmenté de 4 M\$ US/an à 22 M\$ US/an (Wrather *et al.*, 1997; Wrather *et al.*, 2001; Wrather *et al.*, 2010).

Plusieurs aspects de la biologie du NKS lui permettent d'être un parasite efficace et de causer de lourds dommages à son hôte. En effet, le NKS présente une large gamme d'hôtes, une diversité de formes de virulence, une survie élevée dans des conditions climatiques adverses et la capacité de persister longtemps dans l'environnement en absence d'hôtes (Wrather *et al.*, 1997; Wrather *et al.*, 2010). Cette résilience du NKS réside dans sa capacité à entrer en quiescence ou en diapause et la protection conférée par la cuticule du kyste qui renferme des centaines d'œufs, eux-mêmes protégés par les couches externes formées de trois membranes (Schmitt, 2004; Perry et Moens, 2011). Le cycle de vie débute à l'intérieur de l'œuf, où le nématode subit une première mue pour atteindre le second stade juvénile (J2) qui est la forme infectieuse (Koenning, 2004). Cette larve quitte l'œuf suite à l'éclosion puis, à l'aide d'un stylet, le NKS s'introduit dans la racine de l'hôte, choisit une cellule initiale des tissus vasculaires et crée un site d'alimentation appelé syncytium (Niblack, 2005; Niblack *et al.*, 2006). Cette structure produira les ressources essentielles à son développement au détriment de la croissance de la plante hôte (Noel, 2004). Par la suite, le NKS subit trois autres mues à l'intérieur de la racine pour finalement atteindre la maturité sexuelle (Noel, 2004). Le mâle sort ensuite de la racine pour féconder la femelle dont une partie du corps est maintenant exposée à la surface de la racine. La femelle fécondée expulse

une partie des œufs à l'intérieur d'une masse gélatineuse, lesquels sont prêts à éclore, et garde les autres dans son corps (Noel, 2004). Ce dernier formera un kyste contenant les œufs lors de la mort de la femelle (Niblack, 2005).

C'est d'ailleurs sous forme de kyste que le NKS est dispersé par divers vecteurs biotiques (humains, animaux) ou abiotiques (vent, eau) et ce, autant à l'échelle de pays qu'à celle d'un champ (Dong et Opperman, 1997; Koenning, 2004; Niblack, 2005). La première détection du NKS en Amérique a eu lieu en 1954 en Caroline du Nord et par la suite, il a été retrouvé dans tous les états centraux (du sud au nord) des États-Unis (Niblack, 2005). Au Canada, ce nématode a d'abord été observé en 1987 dans le comté de Kent en Ontario puis, au Québec, durant l'été 2013 dans un champ de soya en Montérégie (Anderson et Welacky, 1988; Mimee *et al.*, 2014b; Tylka et Marett, 2014).

L'établissement du NKS dans pratiquement tous les pays producteurs de soya, peu importe leurs latitudes, démontre que cet organisme s'adapte à une large gamme d'environnements et de conditions climatiques. La température a un effet important sur son cycle de vie de cet organisme poïkilotherme. Il peut compléter un cycle en trois semaines entre 22 et 30°C et en quatre semaines entre 16 et 22°C (Niblack, 2005; Oyekanmi et Fawole, 2010). Ainsi, les conditions climatiques influencent directement la dynamique des populations de NKS et, par conséquent, le rendement de la culture (Gulke, 2011; Kakaire, 2012; Pautasso, 2012; Ahanger, 2013). Donc, il s'avère très pertinent de se questionner sur l'impact qu'auront les fluctuations de températures attendues à la faveur des changements climatiques (CC) sur la biologie du soya, du NKS et de leur interaction. L'augmentation des températures et du nombre d'évènements météorologiques extrêmes (sécheresse, canicule, précipitation) associées aux changements climatiques, pourrait exacerber les dommages occasionnés par le NKS de plusieurs façons. Par exemple, en facilitant sa dispersion, en augmentant le nombre de générations par saison, en modifiant certaines caractéristiques physiologiques des racines du soya qui favoriseraient la reproduction du NKS ou en modifiant les conditions physico-chimiques du milieu qui pourrait interférer dans la relation hôte-parasite en modulant leurs interactions moléculaires (Alston et Schmitt, 1987).

Impacts des changements climatiques sur la distribution et la biologie du NKS et du soya

L'analyse de la distribution du NKS entre 1971 et 1998 aux États-Unis a démontré une dispersion rapide et un taux d'établissement très élevé dans les régions productrices du centre et du nord des États-Unis (Chakraborty et Datta, 2003). Ce phénomène a probablement été accentué par l'adoption de nouvelles zones de production à la faveur des fluctuations du climat (Somasekhar et Prasad, 2012). De même, la modélisation de la phénologie du soya à l'intérieur de la zone de potentiel agricole du Québec, reposant sur les paramètres climatiques du passé récent (1971-2000) et du futur proche (2041-2070), a démontré que des latitudes plus nordiques offrent de nouvelles zones de culture du soya (Gendron St-Marseille, 2013). Également, la modélisation du cycle de vie du NKS en fonction de paramètres climatiques a démontré qu'il pouvait dès lors s'établir à l'intérieur des régions agricoles où le soya se cultive actuellement (Dong et Opperman, 1997; Koening, 2004; Niblack, 2005) et qu'il possède la capacité de suivre le déplacement de son hôte vers le nord (Gendron St-Marseille, 2013).

Également, la hausse prévue des températures favorisera positivement le taux de développement du NKS, ce qui se traduirait par l'accumulation de une à deux générations supplémentaires pour atteindre jusqu'à deux générations par saison de croissance dans la zone la plus septentrionale du Québec et jusqu'à six générations dans celle située à l'extrême sud (Gendron St-Marseille, 2013). Ce scénario aura pour effet d'affaiblir d'avantage le plant de soya, car un nombre de surinfections plus élevé augmente d'une part la sensibilité des racines aux autres pathogènes et d'autre part, limite l'acheminement des nutriments et de l'eau vers les parties aériennes (Gendron St-Marseille, 2013).

En contrepartie, des modèles prédisent que la hausse des températures, couplée à celle des niveaux de CO₂ atmosphériques auront un effet positif sur la croissance, la densité et la longueur des racines du soya (Lawn et Noel, 1986; Niblack *et al.*, 2006). Toutefois, le réchauffement des températures pourrait affecter de façon plus prononcée les impacts négatifs des premières générations de NKS sur le rendement des cultures puisque l'augmentation des surfaces racinaires permettra à un plus grand nombre d'individus de se développer (Rogers *et al.*, 1992; Ziska, 1998; Wang *et al.*, 2003). Également, les événements climatiques extrêmes, dont la sécheresse, semblent réduire les rendements du soya de manière plus prononcée dans les champs de soya infestés par le NKS (Tylka, 2012). Ce stress entraîne le nématode à créer son site de nutrition plus profondément dans les tissus vasculaires de la stèle et non dans la zone du cortex des

racines, ce qui lui permettrait de continuer d'exploiter les ressources de la plante (Bonner et Schmitt, 1985; Tylka, 2012). Par conséquent, des fréquences plus élevées de sécheresses devraient avoir d'importantes répercussions sur la production de soya.

Changements climatiques et mécanismes de défense de l'hôte

De plus, les CC pourraient modifier les interactions moléculaires entre l'hôte et l'endoparasite en interférant avec les mécanismes de défense de la plante (Tylka, 2012). L'action simultanée de stress de nature abiotiques (hydriques, nutritionnels, thermiques, etc.) et biotiques (nématodes, pucerons, champignons, etc.) peut avoir un effet additif, synergique ou antagoniste sur les diverses voies métaboliques des hormones qui confèrent la résistance aux stress chez les plantes (Atkinson et Urwin, 2012; Somasekhar et Prasad, 2012). La principale voie métabolique impliquée dans la défense contre les stress abiotiques est celle de l'acide abscissique (ABA) (Atkinson et Urwin, 2012; Studham et MacIntosh, 2012; Atkinson *et al.*, 2013; Suzuki *et al.*, 2014), laquelle altère la signalisation des principales hormones (acide jasmonique (JA), acide salicylique (AS), éthylène (ET)) associées aux stress biotiques causés par les nématodes phytoparasites (Rosso *et al.*). Différents stress peuvent influencer les interactions plantes-organismes nuisible tels que la sécheresse (Atkinson et Urwin, 2012; Nahar *et al.*, 2012; Studham et MacIntosh, 2012; Atkinson *et al.*, 2013; Suzuki *et al.*, 2014), la hausse de la salinité (Anderson *et al.*, 2004; Atkinson *et al.*, 2013; Prasch et Sonnewald, 2013), les hautes températures (Syvertsen et Levy, 2005; Mittler et Blumwald, 2010), les carences en éléments nutritifs essentiels (potassium) (Dropkin, 1969; Prasch et Sonnewald, 2013; Verdejo-Lucas *et al.*, 2013) et la hausse du niveau de CO₂ atmosphérique (Amtmann *et al.*, 2008). Précisément, il a été démontré que ces stress environnementaux, qui entraînent l'activation des signaux chimiques ABA, influencent négativement la résistance, la croissance et le rendement des cultures à cause des réactions antagonistes associées à l'activation des voies de signalement des AJ, AS initiée par la présence du nématode (Sun *et al.*, 2010; Sun *et al.*, 2011). En revanche, il semblerait que des basses températures favorisent la résistance de certaines espèces de plantes, dont la pomme de terre (*Solanum spp*), par l'entremise d'une adaptation croisée hôte-parasite (Thaler et Bostock, 2004; Suzuki *et al.*, 2014). Compte tenu de la hausse attendue de la prévalence et de l'ampleur de l'ensemble de ces stress à la faveur des CC, il importe de se questionner sur l'évolution de la résistance aux nématodes des diverses plantes cultivées. Néanmoins, un nombre restreint de

recherches porte sur l'impact des stress associés aux CC sur les gènes associés respectivement à la virulence des nématodes phytoparasites et à la résistance de l'hôte (Sysoeva *et al.*, 2011; Sysoeva *et al.*, 2012).

Impacts des changements climatiques sur les gènes de résistance du soya et de virulence du NKS

Actuellement, les sources de résistance chez le soya ne sont efficaces que contre certaines populations de NKS (Elad et Pertot, 2014). La capacité relative de différentes populations de NKS à se reproduire sur sept lignées de soya de référence a mené à l'identification de types HG (*Heterodera glycines*) quantifiant la virulence (Niblack *et al.*, 2002). Cette variabilité phénotypique du caractère de résistance s'expliquerait à la fois par la présence de différents gènes du parasitisme produisant des protéines associées à la virulence, ou effecteurs, chez le nématode et de gènes de résistance chez la plante (Hogenhout *et al.*, 2009). Depuis les années 60, près de 70 loci de caractères quantitatifs (QTL) associés à la résistance ont été identifiés chez le soya, les plus étudiés étant *Rhg1*, *Rhg2* et *Rhg3*, dont le gène majeur est récessif et *Rhg4* et *Rhg5*, dont le gène majeur a un comportement dominant (Caldwell *et al.*, 1960; Matson et Williams, 1965; Rao-Arelli, 1994; Klink *et al.*, 2009). Deux de ces QTL seraient plus importants dans la réaction de résistance, soit *Rhg1* (chromosome 18), qui compte environ neuf gènes et *Rhg4* (chromosome 8) qui compte au moins sept gènes (Klink *et al.*, 2013). On les retrouve chez les cultivars PI 88788 (*Rhg1*) et Peking (*Rhg1* et *Rhg4*), qui ont des réponses cytologiques divergentes et qui forment deux groupes de liaison génétique (*linkage group*) (Klink *et al.*, 2013). Le type de résistance appelé « PI 88788 » repose principalement sur les gènes situés sur l'allèle *rhg1-b* du QTL *Rhg1*, qui a un comportement dominant chez les cultivars de ce groupe (PI 88788, PI 209332 et PI 549316) (Kim *et al.*, 2010; Cook *et al.*, 2012; Liu *et al.*, 2012; Yuan *et al.*, 2012). La réponse cytologique du groupe PI 88788 (regroupant le cultivar du même nom) se caractérise par une nécrose assez lente (entre 8 à 10 jours) du syncytium et par l'absence d'accroissement de la paroi cellulaire (*cell wall apposition*) (Liu *et al.*, 2012, Klink *et al.*, 2013). La résistance de type dit Peking requiert la présence simultanée de gènes co-dominants situés sur les QTL *Rhg1* et *Rhg4* (Kim *et al.*, 2010, Liu *et al.*, 2012, Yuan *et al.*, 2012). Typiquement, la réponse cytologique associée au groupe Peking (Peking, cv. " Forrest ", PI 437654, PI 89772, PI 90763) se caractérise par une nécrose rapide (environ 48 heures) de la cellule initiale et par l'accroissement de la paroi

cellulaire (Cook *et al.*, 2014, Cook *et al.*, 2012). Il a été suggéré que la résistance des plantes est conférée par le nombre de copies de l'allèle *rhg1-b* associé au QTL *Rhg1*. En effet, les lignées et cultivars associés au type PI 88788 posséderaient de sept à dix copies des gènes situés sur le QTL *Rhg1*, alors que ceux situés sur le même QTL au sein du groupe de type Peking en posséderaient moins de trois copies (Brucker *et al.*, 2005; Liu *et al.*, 2012).

Une très grande variabilité génétique se rencontre à l'intérieur des différentes populations de NKS et plusieurs individus seraient porteurs d'allèles « *ror* » (*reproduction on a resistant host alleles*) favorisant le contournement de la résistance de l'hôte et la capacité à se reproduire du parasite (Zheng *et al.*, 2006; Cook *et al.*, 2012; Wang *et al.*, 2014). Les mécanismes qui mènent au contournement de la résistance demeurent inconnus. Cependant, grâce aux multiples recherches conduites sur les effecteurs produits et sécrétés par le NKS à l'intérieur de l'hôte, on comprend mieux son succès parasitaire et l'origine de sa virulence (Liu *et al.*, 2012). À ce jour, on a identifié différentes protéines (effecteurs), sécrétées par les glandes œsophagiennes (deux sous-ventrales et une dorsale) et injectées à l'aide du stylet, qui permettent au NKS d'altérer la structure et les fonctions de la cellule hôte (Li *et al.*, 1996; Bekal *et al.*, 2008; Klink *et al.*, 2009; Haegeman *et al.*, 2012; Wang *et al.*, 2014). Ces altérations diffèrent selon le stade de développement du nématode. Durant la phase migratoire, le stade infectieux, juvénile de stade deux (J2), sécrète des protéines visant à dissoudre les cellules de la racine et à esquiver sa détection par l'hôte en inhibant ou supprimant les hormones associées aux voies métaboliques et de signalament propres aux différents mécanismes de défense de la plante (Haegeman *et al.*, 2012). À la suite de la sélection par le stade J2 d'une cellule initiale servant au développement du syncytium, le NKS produit et injecte d'autres effecteurs qui lui permet de compléter son cycle de vie. Ces effecteurs servent à (i) dissoudre les parois des cellules adjacentes afin d'optimiser le détournement des ressources nutritives de la plante et (ii) favoriser l'ubiquitination (dégradation) de certaines protéines produites par la plante afin d'interférer avec les diverses voies métaboliques et affaiblir le système de défense de l'hôte (Gao *et al.*, 2003; Gao *et al.*, 2004; Ithal *et al.*, 2007; Haegeman *et al.*, 2010; Wang *et al.*, 2010; Haegeman *et al.*, 2012).

Les CC auront probablement des impacts significatifs sur l'intensité de la maladie associés aux interactions entre l'environnement (microclimat), le NKS (biologie, virulence) et le soya (phénologie, résistance) en influençant les diverses interactions moléculaires (Boland *et al.*, 2004;

Parmesan, 2006; Gregory *et al.*, 2009; Elad et Pertot, 2014). En effet, il a été démontré que les CC modifient les interactions parasitaires chez d'autres espèces de nématodes phytoparasites (Boland *et al.*, 2004; Gregory *et al.*, 2009; Atkinson et Urwin, 2012; Newton *et al.*, 2012; Elad et Pertot, 2014). Les hausses des concentrations de CO₂ atmosphérique et de température ont un effet sur les gènes de résistance de plusieurs plantes cultivées. Par exemple, l'élévation des températures limitent, voire annihilent l'efficacité des gènes de résistance de type « R » de différents cultivars de tomate, poivre, luzerne, patate douce, coton, raisin connus pour être résistants à différentes espèces de nématodes à galles du genre *Meloidogyne* (Jablonska *et al.*, 2007; Newton *et al.*, 2012; Ferris *et al.*, 2013). De plus, la hausse du CO₂ atmosphérique aurait également un effet inhibiteur sur les niveaux de production d'hormones (AJ et AS) chez la tomate (*Solanum spp*) qui confèrent une résistance à différents nématodes à galles (Cooper *et al.*, 2005; Sun *et al.*, 2010; Sun *et al.*, 2011). Au niveau du soya, les températures élevées ont un effet négatif sur plusieurs gènes de défense face à différents champignons pathogènes (Upchurch et Ramirez, 2011).

Perspectives

À l'heure actuelle, les effets des CC sur la production d'effecteurs par les nématodes phytoparasites, dont le NKS, sont inconnus. Il s'avère également impossible de déterminer si la perte de la résistance chez la plante est liée à une réaction hypersensible ou à la disparition de la réponse immune suite à la construction d'un syncytium ou d'une cellule géante (nématode à galle) (Ferris *et al.*, 2013). Ainsi, les effets des CC sur le pathosystème soya-NKS devraient être étudiés à l'aide de différentes approches méthodologiques afin d'approfondir les connaissances actuelles et mieux comprendre/anticiper les effets à terme de la hausse des températures et de la concentration de CO₂. Dans un premier temps, il serait pertinent d'étudier la phénologie du soya et le cycle de vie du NKS sous différents scénarios climatiques prévus à l'horizon 2050 (2041-2070) dans les zones cultivables du Québec afin de valider si le climat est propice au développement du nématode et d'identifier les zones à haut risque phytosanitaire. Ensuite, comme l'utilisation de cultivar résistant de soya est le principal moyen de lutte utilisé, il serait primordial de mesurer la performance de cultivars ayant des groupes de maturité adaptés au Québec et résistants au NKS en les exposant à différentes températures et concentrations de CO₂

afin d'anticiper leur performance face aux CC. En outre, comme certaines populations pourraient inclure des individus capables de contourner la résistance, il serait utile d'identifier des marqueurs moléculaires, par exemple, des polymorphismes nucléotidiques simples (SNP) associés aux génotypes de différentes populations de NKS. La découverte de marqueurs capables de différencier les populations de NKS et de prédire leur capacité à se reproduire sur des cultivars résistants permettrait le développement d'outils de diagnostic rapides afin d'optimiser les méthodes de lutte en choisissant des cultivars adaptés. Enfin, l'étude des gènes impliqués dans le contrôle des mécanismes biologiques (par ex. le cycle cellulaire) de l'hôte par le parasite pourrait être réalisée grâce aux nouvelles méthodes d'analyse du transcriptome (RNA-Seq). Cela pourrait éventuellement permettre d'identifier de nouveaux effecteurs chez le NKS impliqués dans le parasitisme mais aussi de mettre en évidence des gènes de défense de la plante réprimant l'expression de gènes de virulence chez le nématode.

RÉFÉRENCES

- Ahanger, R.A., Bhat, H.A., Bhat, T.A., Ganie, S.A., Lone, A.A., Wani, I.A., Ganai, S.A., Haq, S., Khan, O.A., Junaid, J.M. et Bhat, T.A. 2013.** Impact of Climate Change on Plant Diseases. *Int. J. Modern Plant & Anim. Sci.* 1:105-115.
- Ali, N. 2010.** Soybean processing and utilization. Pages 345-374. *In* G. Singh (Éd.). *The soybean: botany, production and uses* First ed. CAB International, Oxfordshire.
- Alston, D.G. et D.P. Schmitt. 1987.** Population density and spatial pattern of *Heterodera glycines* in relation to soybean phenology. *J. Nematol.* 19:336-345.
- Amtmann, A., S. Troufflard et P. Armengaud. 2008.** The effect of potassium nutrition on pest and disease resistance in plants. *Physiol Plant* 133:682-691.
- Anderson, J.P., E. Badruzsafari, P.M. Schenk, J.M. Manners, O.J. Desmond, C. Ehlert, D.J. Maclean, P.R. Ebert et K. Kazan. 2004.** Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in arabidopsis. *Plant Cell* 16:3460-3479.
- Anderson, T.R. et T.W. Welacky. 1988.** First report of *Heterodera glycines* on soybeans in Ontario, Canada. *Plant Dis.* 72:453.
- Ariatti, A. 2014.** Comparison of SBR spread in the US 2008-2011. College of Agricultural Sciences, Penn State. Available at <http://plantpath.psu.edu/research/centers-and-institutes/ceal/research/soybean-rust/historical-risk-assessment/comparison-of-sbr-spread-in-the-us-2008-2011> (accessed on 21/12).
- Atkinson, N.J., C.J. Lilley et P.E. Urwin. 2013.** Identification of genes involved in the response of *Arabidopsis* to simultaneous biotic and abiotic stresses. *Plant Physiol.* 162:2028-2041.
- Atkinson, N.J. et P.E. Urwin. 2012.** The interaction of plant biotic and abiotic stresses: from genes to the field. *J. Exp. Bot.* 63:3523-3543.
- Bekal, S., J.P. Craig, M.E. Hudson, T. Niblack, L.L. Domier et K.N. Lambert. 2008.** Genomic DNA sequence comparison between two inbred soybean cyst nematode biotypes facilitated by massively parallel 454 micro-bead sequencing. *Mol. Genet. Genomics* 279:535-543.
- Boland, G.J., M.S. Melzer, A. Hopkin, V. Higgins et A. Nassuth. 2004.** Climate change and plant diseases in Ontario. *Can. J. Plant Pathol.* 26:335-350.
- Bonner, M.J. et D.P. Schmitt. 1985.** Population dynamics of *Heterodera glycines* life stages on soybean. *J. Nematol.* 17:153-158.

Brucker, E., S. Carlson, E. Wright, T. Niblack et B. Diers. 2005. Rhg1 alleles from soybean PI 437654 and PI 88788 respond differentially to isolates of *Heterodera glycines* in the greenhouse. *Theor Appl Genet* 111:44-49.

CAB International (CABI). 2013. *Heterodera glycines* (soybean cyst nematode). Available at <http://www.cabi.org/isc/?compid=5&dsid=27027&loadmodule=datasheet&page=481&site=144> (accessed on 6/03).

Caldwell, B.E., C.A. Brim et J.P. Ross. 1960. Inheritance of resistance of soybeans to the cyst nematode, *Heterodera glycines*. *Agron. J.* 52:635-636.

Canadian Soybean Council. 2012. Industry statistics. Available at <http://www.soybeancouncil.ca/IndustryStatistics/tabid/200/language/en-US/Default.aspx> (accessed on June 24, 2014).

Chakraborty, S. et S. Datta. 2003. How will plant pathogens adapt to host plant resistance at elevated CO₂ under a changing climate? *New Phytol.* 159:733-742.

Cook, D.E., T.G. Lee, X. Guo, S. Melito, K. Wang, A.M. Bayless, J. Wang, T.J. Hughes, D.K. Willis, T.E. Clemente, B.W. Diers, J. Jiang, M.E. Hudson et A.F. Bent. 2012. Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* 338:1206-1209.

Cooper, W.R., L. Jia et L. Goggin. 2005. Effects of jasmonate-induced defenses on root-knot nematodes infection of resistant and susceptible tomato cultivars. *J. Chem. Ecol.* 31:1953-1967.

Dong, K. et C.H. Opperman. 1997. Genetic analysis of parasitism in the soybean cyst nematode, *Heterodera glycines*. *Genetics* 146:1311-1318.

Dropkin, V.H. 1969. Necrotic reaction of tomatoes and other hosts resistant to *Meloidogyne* : reversal by temperature. *Phytopathology* 59:1632-1637.

Elad, Y. et I. Pertot. 2014. Climate change impacts on plant pathogens and plant diseases. *Journal of Crop Improvement* 28:99-139.

FAOSTAT. 2014. Commodities by country – 2012. FAO. Available at <http://faostat.fao.org/site/339/default.aspx> (accessed on 24/06).

Ferris, H., L. Zheng et M.A. Walker. 2013. Soil temperature effects on the interaction of grape rootstocks and plant parasitic nematodes. *J. Nematol.* 45:49-57.

Food and Agriculture Organization (FAO). 2013. Food outlook – Oilseeds market summary. FAO. Available at http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Oilcrops/Documents/Food_outlook_oilseeds/Food_Outlook_June_13.pdf (accessed on 24/06).

- Gao, B., E. Allen, E.L. Davis, T.J. Baum et R.S. Hussey. 2004.** Molecular characterisation and developmental expression of cellulose-binding protein gene in soybean cyst nematode *Heterodera glycines*. *Int. J. Parasitol.* 34:1377-1383.
- Gao, B., R. Allen, T. Maier, E.L. Davis, T.J. J. Baum et R.S. Hussey. 2003.** The parasitome of the phytonematode *Heterodera glycines*. *Mol. Plant-Microbe Interact.* 16:720-726.
- Gendron St-Marseille, A.-F. 2013.** Le nématode à kyste du soja (*Heterodera glycines*) : enjeux des changements climatiques sur sa distribution, sa reproduction et sur les probabilités de synchronisme avec le soja (*Glycines max*) au Québec. Essai. Université de Sherbrooke, Sherbrooke.
- Gregory, P.J., S.N. Johnson, A.C. Newton et J.S. Ingram. 2009.** Integrating pests and pathogens into the climate change/food security debate. *J. Exp. Bot.* 60:2827-2838.
- Grulke, N.E. 2011.** The nexus of host and pathogen phenology : understanding the disease triangle with climate change. *New Phytol.* 189: 8–11.
- Haegeman, A., T. Kyndt et G. Gheysen. 2010.** The role of pseudo-endoglucanases in the evolution of nematode cell wall modifying proteins. *J. Mol. Evol.* 16:720-726.
- Haegeman, A., S. Mantelin, J.T. Jones et G. Gheysen. 2012.** Functional roles of effectors of plant-parasitic nematodes. *Gene* 492:19-31.
- Hartman, G.L., E.D. West et T.K. Herman. 2011.** Crops that feed the world 2. Soybean – worldwide production, use, and constraints caused by pathogens and pests. *Food Security* 3:5-17.
- Hogenhout, S.A., R.A. Van der Hoorn, R. Terauchi et S. Kamoun. 2009.** Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant-Microbe Interact.* 22:115-122.
- Hulme, P.E. 2009.** Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.* 46:10-18.
- Ithal, N., J. Recknor, D. Nettleton, L. Hearne, T. Maier, T.J. Baum et M.G. Mitchum. 2007.** Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. *Plant-Microbe Interact.* 20:293-305.
- Jablonska, B., J.S.S. Ammiraju, K.K. Bhattarai, S. Mantelin, O. Martinez de Ilarduya, P.A. Roberts et I. Kaloshian. 2007.** The Mi-9 gene from *Solanum arcanum* conferring heat-stable resistance to root-knot nematodes is a homolog of Mi-1 *Plant Physiol.* 143:1044-1054.
- Kakaire, S., Grove, I.G. et Haydock, P.P.J. 2012.** Effect of temperature on the life cycle of *Heterodera schachtii* infecting oilseed rape (*Brassica napus* L.). *Nematology* 14:855-867.

Kim, M., D.L. Hyten, A.F. Bent et B.W. Diers. 2010. Fine mapping of the SCN resistance locus rhg1-b from PI 88788. *the Plant Genome* 3:81-89.

Klink, V.P., P. Hosseini, M.H. MacDonald, N.W. Alkharouf et B.F. Matthews. 2009. Population-specific gene expression in the plant pathogenic nematode *Heterodera glycines* exists prior to infection and during the onset of a resistant or susceptible reaction in the roots of the *Glycine max* genotype Peking. *BMC Genomics* 10:111.

Klink, V.P., P.D. Matsye, K.S. Lawrence et G.W. Lawrence 2013. Engineered soybean cyst nematode resistance, *in* Soybean - Pest Resistance, edited by H. El-Shemy, InTech, doi: 10.5772/54514.

Koenning, S.R. 2004. Population biology. Pages 73-88. *In* D.P. Schmitt, J.A. Wrather et R.D. Riggs (Éds.). *Biology and management of soybean cyst nematode*, Second ed. Schmitt and Associates of Marceline, Marceline.

Lawn, D.A. et G.R. Noel. 1986. Field interrelationships among *Heterodera glycines*, *Pratylenchus scribneri*, and three other nematode species associated with soybean. *J. Nematol.* 18:98-105.

Lee, D.J., C.S. Kim et G. Schaible. 2006. Estimating the Cost of Invasive Species on U.S. Agriculture: The U.S. Soybean Market, p. 1-24. *In* U.A.R. Service (ed.), vol. 243. USDA-ARS / UNL Faculty, Lincoln, Nebraska.

Li, J., J. Faghihi, J.M. Ferris et V.R. Ferris. 1996. The use of RAPD amplified DNA as markers for virulence characteristics in soybean cyst nematode. *Fundam. Appl. Nematol.* 19:143-150.

Liu, S., P.K. Kandoth, S.D. Warren, G. Yeckel, R. Heinz, J. Alden, C. Yang, A. Jamai, T. El-Mellouki, P.S. Juvale, J. Hill, T.J. Baum, S. Cianzio, S.A. Whitham, D. Korkin, M.G. Mitchum et K. Meksem. 2012. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* 492:256-260.

Matson, A.L. et L.F. Williams. 1965. Evidence of a fourth gene for resistance to soybean cyst nematode. *Crop Sci.* 5:477.

Mimee, B., J. Brodeur, G. Bourgeois, J. Moiroux, A.-F. Gendron St-Marseille et A.-E. Gagnon. 2014a. Quels enjeux représentent les changements climatiques en lien avec les espèces exotiques envahissantes pour la culture du soya au Québec ? , p. 51
http://www.ouranos.ca/media/publication/332_RapportBrodeur2014.pdf. Ouranos Québec, Canada.

- Mimee, B., H. Peng, V. Popovic, M.-O. Duceppe, M.-P. Tetreault et G. Belair. 2014b.** First report of soybean cyst nematode (*Heterodera glycines* Ichinohe) on soybean in the Province of Quebec, Canada. . Plant Dis. 98:429.
- Mittler, R. et E. Blumwald. 2010.** Genetic engineering for modern agriculture: challenges and perspectives. Annu. Rev. Plant Biol. 61:443-462.
- Nahar, K., T. Kyndt, Y.B. Nzogela et G. Gheysen. 2012.** Abscisic acid interacts antagonistically with classical defense pathways in rice-migratory nematode interaction. New Phytol. 196:901-913.
- Newton, A.C., L. Torrance, N. Holden, I.K. Toth, D.E. Cooke, V. Blok et E.M. Gilroy. 2012.** Climate change and defense against pathogens in plants. Advance in Applied Microbiology 81:89-132.
- Niblack, T.L. 2005.** Soybean cyst nematode management reconsidered. Plant Dis. 80:1020-1026.
- Niblack, T.L., P.R. Arelli, G.R. Noel, C.H. Opperman, J.H. Orf, D.P. Schmitt, J.G. Shannon et G.L. Tylka. 2002.** A revised classification scheme for genetically diverse populations of *Heterodera glycines*. J. Nematol. 34:279-288.
- Niblack, T.L., K.N. Lambert et G.L. Tylka. 2006.** A model plant pathogen from the kingdom animalia : *Heterodera glycines*, the soybean cyst nematode. . Annu. Rev. Phytopathol. 44:283-303.
- Noel, G.R. 2004.** Soybean response to infection. Pages 131-151. In D.P. Schmitt, J.A. Wrather et R.D. Riggs (Éds.). Biology and management of soybean cyst nematode Second ed. Schmitt and Associates of Marceline, Marceline.
- Oyekanmi, E.O. et B. Fawole. 2010.** Nematodes of soybean and their management. Pages 325-343. In G. Singh (Éd.). The soybean: botany, production and uses. CAB International, Oxfordshire.
- Parmesan, C. 2006.** Ecological and evolutionary responses to recent climate change. Annual Review of Ecology, Evolution, and Systematics 37:637-669.
- Pautasso, M., Döring, T. F., Garbelotto, M., Pellis, L. et Jeger, M.J. 2012.** Impacts of climate change on plant diseases—opinions and trends. Eur. J. Plant Pathol. 133:295-313.
- Perry, R.N. et M. Moens. 2011.** Survival of parasitic nematode outside the host. Pages 1-27. In R.N. Perry et D.A. Wharton (Éds.). Molecular and Physiological Basis of Nematode Survival CABI Publishing, Wallingford.
- Prasch, C.M. et U. Sonnewald. 2013.** Simultaneous application of heat, drought, and virus to *Arabidopsis* plants reveals significant shifts in signaling networks. Plant Physiol. 162:1849-1866.

- Ragsdale, D.W., D.A. Landis, J. Brodeur, G.E. Heimpel et N. Desneux. 2011.** Ecology and management of the soybean aphid in North America. *Annu. Rev. Entomol.* 56.
- Rao-Arelli, A.P. 1994.** Inheritance of resistance to *Heterodera glycines* race 3 in soybean accession. *Plant Dis.* 78:898-900.
- Rogers, H.H., C.M. Peterson, J.N. McCrimmon et J.D. Cure. 1992.** Response of plant roots to elevated atmospheric carbon dioxide. *Plant, Cell Environ.* 15:749-752.
- Rosso, M.-N., R.S. Hussey, E.L. Davis, G. Smant, T.J. Baum, P. Abad et M.G. Mitchum. 2012.** Nematode effector proteins: targets and functions in plant parasitism. Pages 327-354. *In* F. Martin et S. Kamoun (Éds.). *Effectors in Plant–Microbe Interactions*. JohnWiley & Sons, Inc.
- Schmitt, D.P. 2004.** Introduction. Pages 1-18. *In* D.P. Schmitt, J.A. Wrather et R.D. Riggs (Éds.). *Biology and management of soybean cyst nematode* Second ed. Schmitt and Associates of Marceline, Marceline.
- Singh, G. et B.G. Shivakumar. 2010.** The role of soybean in agriculture. Pages 22-47. *In* G. Singh (Éd.). *The soybean : botany, production and use*, First ed. CAB International, Oxfordshire.
- Somasekhar, N. et J.S. Prasad. 2012.** Plant-nematode interactions : consequences of climate change. Pages 547-564. *In* B. Venkateswarlu, A.K. Shanker, C. Shanker et M. Maheswari (Éds.). *Crop Stress and its Management: Perspectives and Strategies* Springer Netherlands, Houten.
- Statitics Canada. 2014.** Table 001-0010 - Estimated areas, yield, production and average farm price of principal field crops, in metric units, annual [Online].
<http://www5.statcan.gc.ca/cansim/a26?id=10010&retrLang=eng&lang=eng> (Accessed on Aug. 20, 2014).
- Studham, M.E. et G.C. MacIntosh. 2012.** Phytohormone signaling pathway analysis method for comparing hormone responses in plant-pest interactions. *BMC Research Notes* 5.
- Sun, Y., H. Cao, J. Yin, L. Kang et F. Ge. 2010.** Elevated CO₂ changes the interactions between nematode and tomato genotypes differing in the JA pathway. *Plant, Cell Environ.* 33:729-739.
- Sun, Y., J. Yin, H. Cao, C. Li, L. Kang et F. Ge. 2011.** Elevated CO₂ influences nematode-induced defense responses of tomato genotypes differing in the JA pathway. *PLoS ONE* 6.
- Suzuki, N., R.M. Rivero, V. Shulaev, E. Blumwald et R. Mittler. 2014.** Abiotic and biotic stress combinations. *New Phytol.* 203:32-43.
- Sysoeva, M.I., V.V. Lavrova, E.M. Matveeva, E.G. Sherudilo et L.V. Topchieva. 2011.** Cross adaptation of potato plants to low temperatures and potato cyst nematode infestation. *Russ. J. Plant Physiol.* 58:999-1004.

Sysoeva, M.I., E.M. Matveeva, V.V. Lavrova et E.G. Sherudilo. 2012. Potato plant responses to temperature drop and phytonematode infestation under continuous lighting. *Acta Horticulturae* 956:621-625.

Syvertsen, J. et Y. Levy. 2005. Salinity interactions with other abiotic and biotic stresses in citrus. *HortTechnology* 15:100-103.

Thaler, J.S. et R.M. Bostock. 2004. Interactions between abscisic-acid-mediated responses and plant resistance to pathogens and insects. *Ecology* 85:48-58.

Tilmon, K.J., E.W. Hodgson, M.E. O'Nealm et D.W. Ragsdale. 2011. Biology of the Soybean Aphid, *Aphis glycines* (Hemiptera: Aphididae) in the United States. *Journal of Integrated Pest Managment* 2:1-7.

Tylka, G.L. 2012. Soybean cyst nematode field guide – A reference for identifying, scouting for, and managing soybean cyst nematode. Available at <https://store.extension.iastate.edu/Product/Soybean-Cyst-Nematode-Field-Guide> (accessed on 02/07).

Tylka, G.L. et C.C. Marett. 2014. Distribution of the soybean cyst nematode, *Heterodera glycines*, in the United States and Canada: 1954 to 2014. *Plant Health Prog.* 15:13-15.

Upchurch, R.G. et M.E. Ramirez. 2011. Effects of temperature during soybean seed development on defense-related gene expression and fungal pathogen accumulation. *Biotechnol. Lett.* 33:2397-2404.

Verdejo-Lucas, S., M. Blanco, L. Cortada et F.J. Sorribas. 2013. Resistance of tomato rootstocks to *Meloidogyne arenaria* and *Meloidogyne javanica* under intermittent elevated soil temperatures above 28 °C. *Crop Protect.* 46:57-62.

Wang, G., D. Peng, B. Gao, W. Huang, L. Kong, H. Long, P. Peng et H. Jian. 2014. Comparative transcriptome analysis of two races of *Heterodera glycines* at different developmental stages. *PLoS ONE* [Online], 9. doi: 10.1371/journal.pone.0091634.

Wang, J., C. Lee, A. Replogle, S. Joshi, D. Korkin, R.S. Hussey, T.J. Baum, E.L. Davis, X.H. Wang et M.G. Mitchum. 2010. Dual roles for the variable domain in protein trafficking and host-specific recognition of *Heterodera glycines* CLE B effector proteins. *New Phytol.* 187:1003-1017.

Wang, J., T.L. Niblack, J.A. Tremain, W.J. Wiebold, G.L. Tylka, C.C. Marett, G.R. Noel, O. Myers et M.E. Schmidt. 2003. Soybean cyst nematode reduces soybean yield without causing obvious aboveground symptoms. *Plant Dis.* 87:623-628.

Wrather, J.A., T.R. Anderson, D.M. Arsyad, J. Gai, L.D. Ploper, A. Porta-Puglia, H.H. Ram et J.T. Yorinori. 1997. Soybean disease loss estimates for the top 10 soybean producing countries in 1994. *Plant Dis.* 81:107-110.

Wrather, J.A., T.R. Anderson, D.M. Arsyad, Y. Tan, L.D. Ploper, A. Porta-Puglia, H.H. Ram et J.T. Yorinori. 2001. Soybean disease loss estimates for the top ten soybean-producing countries in 1998. *Can. J. Plant Pathol.* 23:115-121.

Wrather, J.A., G. Shannon, R. Balardin, L. Carregal, R. Escobar, G.K. Gupta, Z. Ma, W. Morel, T. Ploper et A. Tenuta. 2010. Effect of diseases on soybean yield in the top eight producing countries in 2006. *Plant Health Prog.*, 10.1094/PHP-20100125-01-RS.

Yuan, C.P., Y.H. Li, Z.X. Liu, R.X. Guan, R.Z. Chang et L.J. Qiu. 2012. DNA sequence polymorphism of the Rhg4 candidate gene conferring resistance to soybean cyst nematode in Chinese domesticated and wild soybeans. *Mol. Breed.* 30:1155-1162.

Zheng, J., Y. Li et S. Chen. 2006. Characterization of the virulence phenotypes of *Heterodera glycines* in Minnesota. *Journal of Nematology* 38:383-390.

Ziska, L.H. 1998. The influence of root zone temperature on photosynthetic acclimation to elevated carbon dioxide concentrations. *Ann. Bot.* 81:717-721.

