

Génomique de l'adaptation de Globodera pallida aux résistances de la pomme de terre et conséquences sur les traits d'histoire de vie du nématode

Delphine Eoche-Bosy

► To cite this version:

Delphine Eoche-Bosy. Génomique de l'adaptation de Globodera pallida aux résistances de la pomme de terre et conséquences sur les traits d'histoire de vie du nématode. Amélioration des plantes. Agrocampus Ouest, 2016. Français. NNT: 2016NSARA078. tel-01638318

HAL Id: tel-01638318 https://theses.hal.science/tel-01638318

Submitted on 20 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



AGRO CAMPUS

OUEST

Delphine EOCHE-BOSY • 23 novembre 2016

Thèse AGROCAMPUS OUEST sous le label de l'Université Bretagne Loire pour obtenir le grade de **DOCTEUR D'AGROCAMPUS OUEST** Spécialité Biologie et Agronomie ÉCOLE DOCTORALE • Vie - Agro - Santé (VAS) LABORATOIRE D'ACCUEIL • Institut de génétique, environnement et protection des plantes (IGEPP)

Génomique de l'adaptation de Globodera pallida aux résistances de la pomme de terre et conséquences sur les traits d'histoire de vie du nématode Maria MANZANARES-DAULEUX Professeure, AGROCAMPUS OUEST, UMR INRA -AO-UR1 IGEPP / présidente Elisabeth FOURNIER

Directrice de recherche, INRA Montpellier/ rapporteure Olivier PLANTARD Directeur de recherche, INRA Nantes / rapporteur Philippe CASTAGNONE-SERENO Directeur de recherche, INRA Sophia / examinateur Benoit MOURY Directeur de recherche, INRA Avignon/ examinateur Eric GRENIER

Chargé de recherche, INRA Rennes / directeur de thèse

Josselin MONTARRY

Chargé de recherche, INRA Rennes / co-encadrant de thèse

Génomique de l'adaptation de Globodera pallida aux résistances de la pomme de terre et conséquences sur les traits d'histoire de vie du nématode

Delphine EOCHE-BOSY

23 novembre 2016

Thèse AGROCAMPUS OUEST sous le label de l'Université Bretagne Loire pour obtenir le grade de DOCTEUR D'AGROCAMPUS OUEST Spécialité Biologie et Agronomie

> ÉCOLE DOCTORALE Vie - Agro - Santé (VAS)

LABORATOIRE D'ACCUEIL UMR INRA - AGROCAMPUS OUEST - Université de Rennes 1 Institut de Génétique, Environnement et Protection des Plantes (IGEPP)

COMPOSITION DU JURY

Maria MANZANARES-DAULEUX	Professeure, Agrocampus Ouest, UMR INRA-AO-UR1 IGEPP	Présidente
Elisabeth FOURNIER	Directrice de Recherche, INRA Montpellier	Rapporteure
Olivier PLANTARD	Directeur de Recherche, INRA Nantes	Rapporteur
Philippe CASTAGNONE-SERENO	Directeur de Recherche, INRA Sophia	Examinateur
Benoit MOURY	Directeur de Recherche, INRA Avignon	Examinateur
Eric GRENIER	Chargé de Recherche, INRA Rennes	Directeur de thèse
Josselin MONTARRY	Chargé de Recherche, INRA Rennes	Co-encadrant de thèse

REMERCIEMENTS

Cette belle et riche aventure de trois ans s'achève, et je tiens à remercier toutes les personnes qui l'ont rendue possible et qui ont participé à son bon déroulement.

Tout d'abord, merci à mes deux encadrants de m'avoir fait confiance et d'avoir accepté de me confier ce projet, malgré mon inexpérience dans à peu près tous les champs disciplinaires qui le composent. Josselin, un grand merci pour ton implication, ta disponibilité, ton dynamisme et tes qualités humaines. Éric, merci aussi pour ton encadrement et ton implication dans ce projet, et merci en particulier pour ton accueil, ainsi que celui d'Anne-Sophie et de vos filles, pendant mon séjour au Québec.

Un énorme merci général à tous les membres de cette super équipe Némato - so Royal Jelly - pour cette bonne humeur partagée au cours de ces trois ans, pour l'aide, les connaissances et les conseils apportés tout au long de la thèse, et en particulier au cours des derniers jours de rédaction, pendant lesquels votre soutien a été très précieux. Sylvain, merci pour ton intérêt pour mon projet de thèse et tout ce que tu as puy apporter en connaissances, compétences et réflexions, ainsi que pour ta grande implication dans les relectures et corrections ces dernières semaines, mais aussi merci pour tous les moments « moins scientifiques », pour ces albums of the week, cette liste de films « approuvée par Montebourg » et ces nombreux conseils séries, que je vais maintenant prendre le temps d'écouter et de regarder, pour tes discussions engagées du midi, et pour ton caractère... « authentique » ! Lionel, merci d'avoir pris le temps de m'expliquer tous les secrets et étapes du phénotypage, et merci pour ces discussions pendant mes périodes passées au L2. Magali, merci pour ton importante implication technique (même si tu ne t'en rends pas compte !) à différentes étapes de cette thèse, notamment dans la recherche de gènes, et merci pour ta gentillesse. Didier, merci pour toutes les idées et l'aide que tu as apportées au cours de la thèse, et merci en particulier pour ton implication dans les manips de dernières minutes. Sylvie, merci du temps que tu m'as accordé pour me former à la BM et au génotypage à mon arrivée en thèse. Catherine, Marie-Christine et Christophe, merci pour l'aide ponctuelle apportée et votre sympathie au quotidien.

Je remercie également Denis Tagu, Directeur de l'UMR IGEPP, et Régine Delourme, Responsable de l'équipe Résistance et Adaptation, pour m'avoir accueillie au cours de ces trois années et permis de mener ce projet dans de bonnes conditions de travail. Merci à tous les membres de l'IGEPP, pour les discussions, conseils et questions pertinentes au sujet de ma thèse, qui m'ont également aidée à avancer dans ce projet, et pour l'atmosphère conviviale que vous faite régner dans l'Unité. Merci en particulier aux membres de l'équipe Résistance et Adaptation et du groupe Bactérie-Mildiou-Nématode.

Je remercie les membres de mon jury de thèse, Elisabeth Fournier, Olivier Plantard, Benoit Moury, Philippe Castagnone-Sereno et Maria Manzanares-Dauleux, pour avoir accepté d'évaluer ce travail.

Je remercie également les membres de mon comité de thèse, Marie-Claire Kerlan, Julie Jaquiéry, Mathieu Gautier, Fabrice Legeai, Fabien Halkett et Philippe Vandenkoornhuyse, pour tous les conseils, idées, remarques et orientations apportés, qui ont permis d'enrichir ce travail.

Merci également à Amandine Juhel, stagiaire de Master 2, pour le travail fourni durant ces six mois sur la partie scan génomique avec marqueurs microsatellites, et pour notre collaboration efficace et amicale. Merci également à Jérémy Gauthier pour avoir initié ce travail pendant son stage de Master 2, avant mon arrivée en thèse, et pour les discussions que nous avons pu avoir lorsque nous nous sommes recroisés au cours de ces trois ans.

Merci aux copains du labo, pour tous les bons moments passés en votre compagnie. Merci à mes deux colocs de bureau successifs, compagnons de thèse et désormais amis, Cécile (dont l'accueil chaleureux et les nombreux conseils m'ont bien aidée à amorcer cette thèse) et Nathan (dont la gentillesse et la bonne humeur - et la passion des Pokémons - ont accompagné cette fin de thèse). Merci aussi aux autres membres de l'ancien groupe des stagiaires M2 devenus également des amis, Pierre-Loup, Thibaut, Gaëtan et Elsa, notamment pour ces quelques soirées où j'ai lâché la thèse (...quand je n'étais pas à Saint-Malo, n'est-ce pas Pierre-Loup ?), cette furtive période de badminton et notre excursion au Petit Pois (et *cette* inoubliable boulette). Merci également aux conjoints, Kévin, Cindy et Marine. Merci à Stefaniya, docteur ès potins avant tout, pour cette joie de vivre, ces rires et ces moments partagés quand tu étais au labo, et merci aussi pour ce weekend passé à Ottawa. Merci à Nico, Romain, Lucile et Camille, pour votre soutien, les discussions plus ou moins sérieuses et tous les moments que j'ai pu partager avec vous.

Je remercie aussi les nombreuses personnes de l'Unité qui m'ont apporté un soutien technique et logistique au cours de ces trois années. Merci aux pros de la BM, notamment Fred, Stéphanie et Nathalie, qui ont répondu à toutes mes sollicitations en début de thèse. Merci à Fabrice, Anthony, Jean-Pierre et Cervin pour leur aide en bioinfo. Merci à Pierre, Julie et Lucie pour l'aide et les conseils apportés en génomique des populations. Merci à toute l'équipe AdSL, et notamment Anne-Sophie, Géraldine, Martine, Pascale, Régis et Anne.

Je remercie également toutes les personnes qui ont permis ma participation à plusieurs activités annexes à ma thèse, rendant ces trois années encore plus intéressantes et enrichissantes (et donc en particulier merci à mes encadrants d'avoir accepté que je grignote un peu (beaucoup ?) de temps sur ma thèse pour cela), ainsi que toutes les personnes qui ont partagé ces moments avec moi. J'ai notamment pu acquérir une expérience d'enseignement et je remercie donc Denis Poinsot et Liliane Krespi d'avoir accepté que j'enseigne dans les UEs dont ils sont responsables, un clin d'œil également à Aziz, Thomas, Fabrice et Denis que j'ai eu l'occasion de croiser régulièrement dans ces salles de TP du bâtiment 29 de Beaulieu. Merci aux membres du groupe animation scientifique et technique pour m'avoir intégrée pendant un an au groupe et merci à Yoann d'avoir organisé avec moi la journée des doctorants de l'IGEPP. J'ai également eu l'opportunité de participer à l'organisation d'un colloque d'envergure nationale, les journées des doctorants du département Santé des Plantes et Environnement de l'INRA, et je remercie les autres membres du comité d'organisation, Nico, Alexia, Nathan, Mélanie, Laurence et Yoann, ainsi que toutes les personnes qui ont collaboré avec nous et permis l'organisation de cet évènement. Merci également aux membres de la commission Professionnalisation et Relations Internationales de l'école doctorale VAS dont j'ai fait partie pendant ces trois ans, merci en particulier à Joan et Franck qui président cette commission, ainsi qu'aux deux autres doctorantes de la commission, Hélène et Lélia, avec qui j'ai organisé la journée de l'après-thèse. Merci enfin aux membres de l'Unité avec qui j'ai pu participer à la préparation et à l'animation du stand de l'IGEPP durant la Fête de la Science.

Cette thèse a également été l'occasion pour moi de beaucoup voyager, en France et à l'étranger, que ce soit pour participer à des congrès, à des formations, ou pour travailler avec d'autres équipes de recherche. Je remercie donc mes encadrants pour m'avoir permis d'effectuer ces nombreux déplacements et également les personnes qui m'ont accueillie durant ces séjours. Merci à Olivier Bouchez et Marie Vidal qui m'ont accueillie à la plateforme Génomique de GénoToul pour que je puisse participer à la préparation des librairies et visiter la plateforme de séquençage. Merci à Mathieu Gautier qui m'a accueillie à Montpellier pour me former à l'utilisation de BayPass, et merci également aux membres de son équipe pour leur sympathie. Merci à l'équipe du CRDH pour son accueil chaleureux lors de mon séjour québécois, merci en particulier à Benjamin, Pierre-Yves (merci pour tous ces conseils de visites !), Joël, Éléonore, Nathalie, Mona, Anne-Frédérique et Michael.

Merci aussi aux nombreuses personnes qui ont eu un jour (ou plus) « pitié » de moi et qui m'ont pris « en stop » sur le chemin du labo, merci en particulier à Bruno et sa petite voiture rouge que j'étais bien contente d'apercevoir les jours de pluie.

Merci à mes encadrants de stage de Master 1 et 2, Joan, Yannick, Thiago et Liliane, qui m'ont permis d'ouvrir les portes du monde de la Recherche et m'ont aidée à préparer l'entrée en thèse.

Un grand merci à ma famille et ma famille « étendue », Maman, Papa, Lisa, Thomas, Émilie, Christine, Jean-Paul et Morgane, pour tout votre soutien. Merci en particulier à ma Maman, qui a toujours fait tout son possible pour que je puisse arriver là où j'en suis aujourd'hui.

Enfin, merci au plus grand des soutiens depuis ces quatre dernières années, Franck, pour tes encouragements, ton réconfort, ton amour, ton implication (à tel point que tu vulgarises mieux mon sujet de thèse que moi...) et pour avoir réussi à me supporter ces dernières semaines (et... *Paye ton SNP, thésarde !* ^(C)). Merci d'avoir partagé cette aventure avec moi et vivement les nouvelles qui nous attendent !

TABLE DES MATIÈRES

Remerciements	i
Table des matières	iv
Valorisation du travail de thèse	vii
Liste des tableaux	ix
Liste des figures	x
Avant-Propos	xii

1.1 Identification des loci responsables de l'adaptation	4
1.1.1 L'adaptation : concepts et mécanismes	4
1.1.2 Détecter l'adaptation dans les génomes	8
1.1.3 L'approche d'évolution expérimentale pour étudier l'adaptation	17
1.2 Interactions plante-pathogène et durabilité des résistances	18
1.2.1 La coévolution plante-pathogène	18
1.2.2 Les résistances des plantes	19
1.2.3 Contournement des résistances	23
1.2.4 Gestion durable des résistances	26
1.3 Modèle d'étude : le pathosystème Globodera pallida / pomme de terre	28
1.3.1 Le nématode à kyste Globodera pallida	28
1.3.2 La pomme de terre : hôte principal de Globodera pallida	46
1.3.3 Interactions plante-nématode	50
1.3.4 Évolution expérimentale et résultats phénotypiques	55
1.4 Objectifs et plan de la thèse	59
1.5 Ressources génomiques acquises	60

ARTICLE 1: ADAPTATION TO RESISTANT HOSTS INCREASES FITNESS ON SUSCEPTIBLE HOSTS IN THE PLANT PARAS	SITIC
NEMATODE GLOBODERA PALLIDA	65
2.1 Introduction	66
2.2 Methods	67
2.2.1 Biology of Globodera pallida	67
2.2.2 Selection of virulent and avirulent Globodera pallida lineages	67
2.2.3 Virulence assessment	67
2.2.4 Life-history traits assessment	68
2.2.5 Statistical analysis	69
2.3 Results	69
2.3.1 Virulence level of the different Globodera pallida lineages	69
2.3.2 Life-history traits measurements in virulent and avirulent lineages	70
2.4 Discussion	73

2.5 Acknowledgments	76
2.6 Data archiving	76

Article 2: Experimentally evolved populations of the potato cyst nematode $Globodera$ pairs	LLIDA
ALLOW THE TARGETING OF GENOMIC FOOTPRINTS OF SELECTION DUE TO HOST ADAPTATION	80
3.1 Introduction	81
3.2 Materials and methods	82
3.2.1 Selection of virulent and avirulent lineages of G. pallida	82
3.2.2 DNA extraction and microsatellite genotyping	83
3.2.3 Detection of outlier loci	84
3.2.4 Genic environment of outlier loci	85
3.2.5 Genetic diversity and differentiation	85
3.3 Results	86
3.3.1 Detection of outlier loci	86
3.3.2 Genic environment of outlier loci	88
3.3.3 Genetic diversity and differentiation	89
3.4 Discussion	90
3.5 Acknowledgments	93

ARTICLE 3: GENOME SCANS ON EXPERIMENTALLY EVOLVED POPULATIONS REVEAL CANDIDATE REGIONS FOR
ADAPTATION TO PLANT RESISTANCE IN THE POTATO CYST NEMATODE GLOBODERA PALLIDA
4.1 Introduction
4.2 Materials and methods
4.2.1 Study system: Globodera pallida 101
4.2.2 Selection of virulent and avirulent G. pallida lineages
4.2.3 Pool sequencing
4.2.4 Mapping and SNP calling
4.2.5 Genome scan for adaptive divergence and association with the virulent/avirulent
status
4.2.6 Genic environment of outlier loci104
4.3 Results
4.3.1 Sequencing, mapping and SNP calling
4.3.2 Genome scan for adaptive divergence and association with the virulent/avirulent
status
4.3.3 Genic environment of outlier loci
4.4 Discussion
4.5 Acknowledgments
4.6 Data accessibility

RÉSULTATS COMPLÉMENTAIRES : APPROCHE DE GÉNOTYPAGE PAR SÉQUENÇAGE	. 113
4.7 Introduction	. 113
RÉSULTATS COMPLÉMENTAIRES : APPROCHE DE GÉNOTYPAGE PAR SÉQUENÇAGE 1 4.7 Introduction 1 4.8 Protocole expérimental 1 4.8 Protocole expérimental 1 4.8 Protocole expérimental 1 4.8 Protocole expérimental 1 4.8.1 Matériel biologique et extraction d'ADN 1 4.8.2 Préparation des librairies et séquençage 1 4.8.3 Traitement des données et SNP calling 1 4.8.4 Détection des loci outlier 1 4.8.5 Approche alternative pour la détection des SNPs 1 4.9 Premiers résultats et discussion 1 5.1 Synthèse 1 5.2 Intérêt de la stratégie d'étude 1 5.2.1 Intérêt de l'étude des déterminants et des conséquences de l'adaptation pour u gestion durable des résistances 1 5.2.2 Intérêt de l'évolution expérimentale 1 5.3 Principaux résultats et limites soulevées 1 5.3.1 Hypothèse d'un allèle de virulence préexistant dans les populations 1 5.3.2 Conséquences de l'adaptation 1 5.3.3 Génomique de l'adaptation 1 5.4 Perspectives 1 5.4 Perspectives 1 5.4 Conclus	. 115
RÉSULTATS COMPLÉMENTAIRES : APPROCHE DE GÉNOTYPAGE PAR SÉQUENÇAGE 1 4.7 Introduction 1 4.8 Protocole expérimental 1 4.8 Protocole expérimental 1 4.8.1 Matériel biologique et extraction d'ADN 1 4.8.2 Préparation des librairies et séquençage 1 4.8.2 Préparation des librairies et séquençage 1 4.8.3 Traitement des données et SNP calling 1 4.8.4 Détection des loci outlier 1 4.8.5 Approche alternative pour la détection des SNPs 1 4.9 Premiers résultats et discussion 1 5.1 Synthèse 1 5.2 Intérêt de la stratégie d'étude 1 5.2.1 Intérêt de l'étude des déterminants et des conséquences de l'adaptation pour ugestion durable des résistances 1 5.2.2 Intérêt de l'évolution expérimentale 1 5.3 Principaux résultats et limites soulevées 1 5.3.1 Hypothèse d'un allèle de virulence préexistant dans les populations 1 5.3.2 Conséquences de l'adaptation 1 5.3.3 Génomique de l'adaptation 1 5.3.4 Gènes candidats à l'adaptation 1 5.4 Perspectives 1 5.4.1 Validation du lien génotype-phénotype dans d	. 115
4.8.2 Préparation des librairies et séquençage	. 115
4.8.3 Traitement des données et SNP calling	. 116
4.8.4 Détection des loci outlier	. 116
4.8.5 Approche alternative pour la détection des SNPs	116
4.9 Premiers résultats et discussion	. 116
CHAPITRE 5 : DISCUSSION GÉNÉRALE	. 119
5.1 Synthèse	120
5.2 Intérêt de la stratégie d'étude	. 121
5.2.1 Intérêt de l'étude des déterminants et des conséquences de l'adaptation pour	une
gestion durable des résistances	. 121
5.2.2 Intérêt de l'évolution expérimentale	. 124
5.3 Principaux résultats et limites soulevées	. 126
5.3.1 Hypothèse d'un allèle de virulence préexistant dans les populations	. 126
5.3.2 Conséquences de l'adaptation sur la fitness du nématode	. 126
5.3.3 Génomique de l'adaptation	. 127
5.3.4 Gènes candidats à l'adaptation	. 131
5.4 Perspectives	. 132
5.4.1 Validation du lien génotype-phénotype dans des populations naturelles :	une
étude en cours	. 132
5.4.2 À terme : des perspectives fondamentales et appliquées	. 133
5.5 Conclusion	136
RÉFÉRENCES BIBLIOGRAPHIQUES	. 139
ANNEXES	. 165

Annexe 1 : Supporting Information - ARTICLE 1	166
Annexe 2 : Supporting Information - ARTICLE 2	169
Annexe 3 : Supporting Information - ARTICLE 3	176
Annexe 4 : Protocole expérimental de l'AS-PCR	191

VALORISATION DU TRAVAIL DE THÈSE

Publications scientifiques

- **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*. Accepted for publication in *Molecular Ecology*.
- Eoche-Bosy D, Gauthier J, Juhel AS, Esquibet M, Fournet S, Grenier E, Montarry J (2017) Experimentally evolved populations of the potato cyst nematode *Globodera pallida* allow the targeting of genomic footprints of selection due to host adaptation. *Plant Pathology*, 66, 1022-1030.
- Fournet S, **Eoche-Bosy D**, Renault L, Hamelin FM, Montarry J (2016) Adaptation to resistant hosts increases fitness on susceptible hosts in the plant parasitic nematode *Globodera pallida*. *Ecology and Evolution*, **6**, 2559-2568.

Communications scientifiques

Communications orales

- Eoche-Bosy D, Esquibet E, Fournet S, Gautier M, Legeai F, Bretaudeau A, Grenier E, Montarry J (2016) Genome scans on experimentally evolved *Globodera pallida* populations to identify molecular bases of adaptation to *GpaVvrn* potato resistance. 32nd ESN Symposium, Braga, Portugal, 28 Août-1^{er} Septembre 2016.
- <u>Eoche-Bosy D</u>, Esquibet E, Fournet S, Gautier M, Legeai F, Bretaudeau A, Grenier E, Montarry J (2016) Genome scans on experimentally evolved *Globodera pallida* populations to identify molecular bases of adaptation to *GpaV_{vrn}* potato resistance. *SON/ONTA Nematological Joint Meeting of the Americas*, Montréal, Québec, Canada, 17-21 Juillet 2016.

Bourse Agrocampus Ouest pour la participation à un congrès scientifique international

- <u>Eoche-Bosy D</u>, Esquibet E, Fournet S, Gautier M, Legeai F, Bretaudeau A, Grenier E, Montarry J (2016) Genome scans on experimentally evolved *Globodera pallida* populations to identify molecular bases of adaptation to *GpaVvrn* potato resistance. *Colloque LUNEPAR (LUtte contre les NÉmatodes PARasites)*, Tours, France, 12-13 Mai 2016.
- Juhel A, **Eoche-Bosy D**, Fournet S, Kerlan MC, Grenier E, Montarry J (2014) Identification des régions génomiques associées à l'adaptation de *Globodera pallida* à la résistance de la pomme de terre. *Le Petit Pois Déridé*, Orsay, France, 25-28 Août 2014.

• <u>Eoche-Bosy D</u>, Juhel A, Fournet S, Kerlan MC, Grenier E, Montarry J (2014) Phenotypic and genomic modifications associated with *Globodera pallida* adaptation to potato resistances. *Evolutionary Biology of* Caenorhabditis *and other Nematodes*, Hinxton, Cambridge, Royaume-Uni, 14-17 Juin 2014.

Registration bursary from the Wellcome Trust Scientific Conferences department Prix étudiant de la meilleure présentation (Nathan A. Cobb Foundation Graduate Award)

• <u>Grenier E</u>, Fournet S, Kerlan MC, **Eoche-Bosy D**, Montarry J (2014) Direct and indirect consequences of *Globodera pallida* selection by resistant potato genotypes. 6th *International Congress of Nematology*, Cape Town, Afrique du Sud, 4-9 Mai 2014.

Posters

- <u>Eoche-Bosy D</u>, Juhel A, Grenier E, Montarry J (2015) Investigation of the molecular bases of the adaptation of the cyst nematode *Globodera pallida* to potato resistances with genome scan approaches using microsatellites markers and NGS data. *7èmes Journées des Doctorants SPE*, Rennes, France, 1^{er}-3 Juillet 2015.
- <u>Eoche-Bosy D</u>, Juhel A, Grenier E, Montarry J (2015) Investigation of the molecular bases of the adaptation of the cyst nematode *Globodera pallida* to potato resistances with genome scan approaches using microsatellites markers and NGS data. *SMBE Satellite meeting SMBEBA 2015: Investigating biological adaptation with NGS: data and models*, Montpellier (Hameau de l'étoile), France, 26-29 Mai 2015.
- <u>Eoche-Bosy D</u>, Juhel A, Fournet S, Kerlan MC, Grenier E, Montarry J (2014) Phenotypic and genomic modifications associated with *Globodera pallida* adaptation to potato resistances. *Bioinformatics for Environmental Genomics workshop*, Lyon, France, 27-28 Mai 2014.

LISTE DES TABLEAUX

Chapitre 1

Tableau 1.1	Classification des pathotypes de Globodera selon Kort et al. (1977)	38
Tableau 1.2	Ressources génomiques acquises au cours de la thèse	61

Chapitre 2

Table 2.1	Pearson's correlation coefficients (below the diagonal) and associated	
	adjusted P-values (above the diagonal) between the different life-history	
	traits (#Cyst, Size_cysts, #Larvae, Length_larvae, P and HT) for (A) the	
	virulent and (B) the avirulent Globodera pallida lineages	71
Table 2.2	Results from the analyses of variance (ANOVAS) assessing the generation	
	effect (6x and 10x), the vir/avir effect and the corresponding two-way	
	interaction of these factors on the six life-history traits (#Cyst, Size_cysts,	
	#Larvae, Length_larvae, P and HT)	72

Chapitre 3

Table 3.1	Microsatellite markers identified by three outlier detection methods	
	(DETSEL, FDIST and InRH) in comparisons between pairs of potato cyst	
	nematode lineages NI/ND, SMI/SMD and SM360/SMD	87
Table 3.2	Genetic characteristics of the main Globodera pallida lineages studied	90

Chapitre 5

LISTE DES FIGURES

Chapitre 1

Figure 1.1	Les trois types de sélection naturelle	5
Figure 1.2	Illustration du phénomène de dérive génétique	6
Figure 1.3	Illustration de l'effet bottleneck	7
Figure 1.4	Illustration de l'autostop moléculaire ou hitchhiking effect	10
Figure 1.5	Conséquences d'un balayage sélectif sur la diversité nucléotidique aux	
	sites neutres liés	12
Figure 1.6	Course aux armements et guerre des tranchées : les deux scénarios de	
	l'évolution des interactions hôte-pathogène	20
Figure 1.7	Modèle en zig-zag illustrant les différents niveaux du système	
	immunitaire des plantes	21
Figure 1.8	Schéma des différents scénarios possibles du modèle gène-pour-gène	23
Figure 1.9	Changement dans les populations de pathogènes lors du contournement	
	d'une résistance totale (A et B) et lors de l'érosion d'une résistance	
	partielle (C et D)	25
Figure 1.10	Structure interne d'un nématode mâle et d'un nématode femelle	29
Figure 1.11	Coupe transversale de nématode	30
Figure 1.12	Arbre phylogénétique des Métazoaires	30
Figure 1.13	Arbre phylogénétique du phylum Nemata basé sur les séquences de la	
	petite sous-unité de l'ADN ribosomal	32
Figure 1.14	Structure de la partie antérieure des nématodes phytoparasites	34
Figure 1.15	Synthèse schématique des rangs taxonomiques conduisant au genre	
	Globodera et relations phylogénétiques entre certaines espèces du genre	
	Globodera	34
Figure 1.16	Phylogénie de populations péruviennes de G. pallida basée sur la	
	séquence partielle du cytochrome b et sur 8 marqueurs microsatellites	36
Figure 1.17	Carte de répartition du nématode à kyste de la pomme de terre	
	Globodera pallida, au 01/07/2016, d'après l'Organisation Européenne et	
	Méditerranéenne pour la Protection des Plantes (OEPP)	36
Figure 1.18	Cycle biologique de Globodera pallida	40
Figure 1.19	Nématode à kyste Globodera pallida et dégâts occasionnés	40
Figure 1.20	Production mondiale de pomme de terre (données FAOSTAT 2013)	46
Figure 1.21	Schéma de l'évolution expérimentale	57
Figure 1.22	Illustration du phénotypage	57
Figure 1.23	Effet de la sélection et durabilité des résistances	58

Chapitre 2

Figure 2.1	Percentage of females produced for each Globodera pallida lineage on	
	the resistant potato cultivar lledher and for the positive controls on the	
	susceptible potato cultivar Désirée	70

Figure 2.2	Means (± standard error) of (A) the number of cysts (#Cysts), (B) the size	
	of cysts (Size_cysts), (C) the number of larvae (#Larvae), (D) the length of	
	larvae (Length_larvae), (E) the hatching probability (P) and (F) the mean	
	hatching time (HT) for the virulent and the avirulent Globodera pallida	
	lineages at each generation (6x and 10x)	73

Chapitre 3

Figure 3.1	Distribution of genetic differentiation (F_{ST} values) among potato cyst nematode lineage pairs as a function of heterozygosity for each of 53	
	microsatellite loci in (a) comparison of lineages SMI and SMD and (b) comparison of lineages NI and ND	87
Figure 3.2	Standardized values of the natural logarithm of the gene diversity ratio (InRH) for each of 53 microsatellite loci in a comparison between potato	
Figure 2.2	cyst nematode lineages (a) SMI and SMD, and (b) NI and ND	88
Figure 5.5	in one, two or three pairwise comparisons of potato cyst nematode lineages (NI/ND, SMI/SMD and SM360/SMD)	88

Chapitre 4

Figure 4.1	Inferred relationship among the eight <i>G. pallida</i> lineages represented by a correlation plot and a hierarchical clustering tree derived from the	
	matrix Ω estimated under the core model	107
Figure 4.2	SNP XtX as a function of the BF_{mc} for association with the virulence status	
	covariable, estimated in one of the three independent analyses	108
Figure 4.3	Example of increase of genetic differentiation (represented as <i>XtX</i> estimates computed in BayPass across the whole dataset) and decrease of genetic diversity (represented as expected heterozygosity H in each	
	lineages having evolved on Iledher) in a candidate genomic region	108
Figure 4.4	BF _{mc} (association avec la covariable statut virulent/avirulent de la lignée)	
	en fonction de XtX	118

Chapitre 5

Figure 5.1	Influence des pratiques agricoles sur la structure génétique des	
	populations de <i>G. tabacum</i>	L22
Figure 5.2	Assignation des données de séquençage non alignées sur le génome de	
	référence de G. pallida à différents taxons, avec l'outil Kraken	130
Figure 5.3	Nombre moyen de kystes produits par plante (n=4) dans chacune des 20	
	populations de G. pallida représentatives de la diversité mondiale, sur le	
	cultivar Iledher 1	134
Figure 5.4	Schéma bilan du positionnement de la thèse dans le cadre conceptuel du	
	contournement des résistances1	137

AVANT-PROPOS

Cette thèse est le résultat d'un travail de trois années réalisé au sein de l'Institut de Génétique, Environnement et Protection des Plantes (IGEPP, UMR INRA, AGROCAMPUS OUEST, Université de Rennes 1), dans l'équipe Résistance et Adaptation. Ce travail a été encadré par Josselin Montarry et Éric Grenier et financé par l'INRA et la Région Bretagne.

Ce projet a été suivi par un comité de thèse composé de Mathieu Gautier (CBGP, UMR INRA, IRD, CIRAD, Montpellier SupAgro), Fabien Halkett (IAM, UMR INRA, Université de Lorraine), Julie Jaquiéry (ECOBIO, UMR CNRS, Université de Rennes 1), Marie-Claire Kerlan (IGEPP, UMR INRA, AGROCAMPUS OUEST, Université de Rennes 1), Fabrice Legeai (IGEPP, UMR INRA, AGROCAMPUS OUEST, Université de Rennes 1) et Philippe Vandenkoornhuyse (ECOBIO, UMR CNRS, Université de Rennes 1).

Une partie de ce travail de thèse a été effectué en collaboration avec l'équipe de recherche de Benjamin Mimee du Centre de Recherche et de Développement en Horticulture (CRDH, Agriculture et Agroalimentaire Canada) à Saint-Jean-sur Richelieu au Québec, où j'ai effectué une mobilité internationale durant un mois et demi. Ce séjour a été financé par une bourse de mobilité sortante accordée par l'école doctorale de Rennes Vie-Agro-Santé (VAS) ainsi qu'une bourse Jean et Marie-Louise Dufrenoy – Crédit Agricole d'Ile-de-France Mécénat accordée par l'Accadémie d'Agriculture de France. Le séquençage réalisé dans le cadre de cette collaboration a été financé par le projet nord-américain GLOBAL (GLOBodera Alliance : Risk Assessment and Eradication of *Globodera* spp. in U.S. Production of Potato).

J'ai également collaboré avec Mathieu Gautier du Centre de Biologie pour la Gestion des Populations à Montpellier, où j'ai été accueillie pour être formée à l'utilisation du logiciel BayPass.

J'ai aussi été accueillie à la plateforme Génomique du Génopole Toulouse Midi-Pyrénées (GENOTOUL) par Olivier Bouchez et Marie Vidal, qui m'ont présenté et fait participer à la préparation des librairies du premier projet de séquençage de cette thèse.

Enfin, les analyses bioinformatiques de cette thèse ont été réalisées en collaboration avec trois bioinformaticiens : Fabrice Legeai, Anthony Bretaudeau (IGEPP, UMR INRA, AGROCAMPUS OUEST, Université de Rennes 1) et Claire Lemaitre (IRISA, UMR CNRS, INRIA, Université de Rennes 1).



CHAPITRE 1

INTRODUCTION GÉNÉRALE

CHAPITRE 1 INTRODUCTION GÉNÉRALE

Les associations hôte-pathogène constituent un modèle intéressant pour l'étude de l'adaptation, du fait de la dynamique de coévolution (dans les écosystèmes naturels) ou d'évolution séquentielle (dans les agroécosystèmes) qui régit ces systèmes et qui permet de maintenir chaque protagoniste dans l'état le plus adapté à l'autre. L'intérêt qui leur est porté est d'autant plus important lorsque l'adaptation rapide des pathogènes menace la santé humaine ou la sécurité alimentaire mondiale. L'apport de connaissances sur les causes et les conséquences de ces adaptations s'avérera utile au développement de méthodes de gestion durable des populations pathogènes (Aguileta et al. 2009). Les organismes pathogènes de plantes sont en particulier responsables de dommages conséquents en agriculture. Ils génèrent des pertes économiques importantes, estimées à 45% du potentiel mondial de production agricole (Agrios 2005). En effet, les schémas de production agricole intensive développés au cours du XX^{ème} siècle, privilégiant notamment l'augmentation de la taille des parcelles, l'utilisation de plantes génétiquement homogènes et la réduction de la diversité en espèces cultivées, ont favorisé le développement des épidémies (Stukenbrock & McDonald 2008). Si l'utilisation massive de molécules chimiques ces dernières décennies s'est avérée être une solution de lutte efficace contre les agents phytopathogènes (Aktar et al. 2009), leur usage est d'ores et déjà et deviendra de plus en plus limité à l'avenir, pour de multiples raisons. Leur nocivité pour la santé humaine et leurs impacts sur les écosystèmes, par la pollution des sols et des nappes phréatiques, la réduction de la biodiversité et la destruction des ennemis naturels des bioagresseurs, ont été mis en évidence à de nombreuses reprises (Savary & Teng 1994; Colosio et al. 2013). Les pesticides représentent également un poids économique important, lié à leur coût direct mais aussi aux coûts externes de gestion, comme ceux liés à la dépollution par exemple. Enfin, leur efficacité est parfois remise en cause par l'émergence de populations de pathogènes résistantes.

Face à ce constat, les politiques publiques s'orientent vers une limitation de leur utilisation, comme c'est le cas en France avec la mise en place du Plan Écophyto visant à réduire de manière drastique l'usage de produits phytosanitaires (Engagement n°129 du Grenelle de l'environnement, 2007; Plan Écophyto de réduction des usages de pesticides, 2008). La nécessité de développer et de disposer de moyens de protection des cultures plus respectueux des ressources naturelles s'impose donc un peu plus chaque jour. L'utilisation de la résistance variétale se présente comme une alternative efficace, à faible coût et respectueuse de l'environnement. Cependant, l'absence de stratégies de déploiement raisonnées à des échelles spatiales suffisamment larges (correspondant aux capacités de migration du pathogène) et sur des pas de temps suffisamment longs favorise le contournement des résistances par les agents pathogènes (Castagnone-Sereno 2002; McDonald & Linde 2002; Parlevliet 2002; Brown 2015), remettant en cause les efforts substantiels de recherche et d'investissement réalisés par les établissements publics de recherche et les entreprises privées de sélection. Ceci est d'autant plus problématique que le temps nécessaire à la création de nouvelles variétés résistantes est souvent très long, de l'ordre de 10 ans pour les céréales et de 25 ans pour la pomme de terre par exemple (Rousselle-Bourgeois & Rousselle 1996). En conséquence, la mise en œuvre de stratégies de gestion des résistances variétales s'avère indispensable afin d'en optimiser la durabilité (Cheatham *et al.* 2009; Mundt 2014). Une résistance est considérée comme durable si elle a conservé son efficacité après avoir été utilisée de manière prolongée sur de grandes surfaces et en présence de l'agent pathogène ciblé dans un environnement favorable au développement de la maladie (Johnson 1984). La durabilité d'une résistance dépend donc de la vitesse d'adaptation des populations pathogènes à la sélection qu'elle exerce. Plusieurs paramètres peuvent moduler cette vitesse d'adaptation, tels que le potentiel évolutif des populations pathogènes, les conséquences de l'adaptation sur les traits d'histoire de vie du pathogène, ou encore le nombre et la nature des mutations impliquées dans l'adaptation (Leach *et al.* 2001; McDonald & Linde 2002; Fabre *et al.* 2009). Caractériser ces différents paramètres représente une étape clé dans l'anticipation du contournement des résistances.

L'un des bioagresseurs majeurs de la de pomme de terre, cinquième production agricole mondiale, est le nématode à kyste *Globodera pallida*. Classé sur la liste des espèces de quarantaine dans de nombreux pays (Hockland *et al.* 2006), il fait l'objet d'une lutte obligatoire en Europe d'après la directive européenne 2007/33/CE, qui recommande fortement l'utilisation de la résistance variétale dans les zones où ces nématodes de quarantaine sont détectés. Le développement de résistances efficaces contre ce nématode représente donc un enjeu majeur. Un facteur de résistance prometteur a été identifié dans une accession sauvage de *Solanum vernei* et introduit dans plusieurs fonds génétiques de pomme de terre, aboutissant, en France, à l'inscription au catalogue de la première variété de consommation résistante à *G. pallida*, Iledher. *Solanum vernei* est aujourd'hui la source de résistance la plus largement répandue dans les schémas de sélection et, au niveau européen, la seule source exploitée dans les variétés commerciales, de consommation comme féculières. Cependant, une évolution expérimentale menée en laboratoire a montré que des populations françaises de *G. pallida* étaient capables de contourner cette résistance en quelques générations (Fournet *et al.* 2013).

Dans l'objectif d'améliorer la durabilité de cette source de résistance, mes travaux de thèse se sont intéressés à l'étude du déterminisme moléculaire de l'adaptation de *G. pallida* aux résistances de sa plante hôte, via l'utilisation d'approches novatrices de génomique des populations, ainsi qu'aux conséquences de cette adaptation sur les traits d'histoire de vie du nématode.

Ce travail de thèse s'est positionné à l'interface entre pathologie végétale et génomique des populations et a donc mobilisé différents concepts et outils référant à ces domaines de recherche, que nous exposerons dans ce premier chapitre. Celui-ci se divise en trois synthèses bibliographiques. La première aborde l'approche de génomique des populations employée dans le cadre de notre étude pour identifier les déterminants génétiques de l'adaptation, la seconde développe les contextes évolutifs et agronomiques des interactions plante-pathogène et de la durabilité des résistances, tandis que la troisième présente le pathosystème *Globodera pallida* – pomme de terre, modèle de notre étude. Ce premier chapitre introductif se conclut par la présentation des objectifs de ce travail de thèse et présente les ressources génomiques produites au cours de ce projet.

1.1 Identification des loci responsables de l'adaptation

L'un des objectifs majeurs de la biologie évolutive est d'identifier les bases génétiques et moléculaires de l'adaptation, dans le but de mieux appréhender les liens entre génotypes et phénotypes et d'élucider les mécanismes des changements évolutifs. Malgré de considérables avancées pour tenter d'atteindre cet objectif, il existe encore aujourd'hui peu de systèmes chez lesquels les gènes impliqués dans les adaptations ont été identifiés. Jusqu'à récemment, répondre à cet objectif restait méthodologiquement limité pour la majorité des organismes. Cependant, le développement récent d'outils et de méthodes adaptés permet maintenant d'envisager de relever ce défi. Les progrès récents des outils de biologie moléculaire et de génomique permettent aujourd'hui d'avoir accès à la variation génétique à l'échelle du génome, y compris chez des espèces non-modèles. À partir de ces données et grâce à des approches de génomique des populations, un champ disciplinaire émergeant à l'interface entre génétique des populations et génomique, il est possible d'identifier les régions du génome soumises à la sélection dans le cadre d'évènements adaptatifs. Dans cette première partie, nous reviendrons sur le concept d'adaptation, avant de définir les méthodes permettant d'en détecter les bases moléculaires, en développant particulièrement celles que nous avons employées à travers notre étude. Enfin, nous nous intéresserons à l'utilisation d'approches d'évolution expérimentale dans ce contexte, puisque c'est la démarche originale que nous avons choisi d'adopter au cours de cette thèse.

1.1.1 L'adaptation : concepts et mécanismes

Dans un contexte de biologie évolutive, l'adaptation correspond à l'ajustement fonctionnel des organismes d'une population afin de mieux répondre aux contraintes de leur environnement local. Cet ajustement se traduit par l'acquisition par la population d'un ensemble de caractéristiques, déterminées génétiquement, grâce à l'action de la sélection naturelle, qui favorise la transmission des variants avantageux d'une génération à l'autre. Cela a pour conséquence une modification des fréquences alléliques au cours des générations dans la population. On parle alors ici de microévolution, par opposition à la macroévolution qui s'intéresse aux changements dans les fréquences alléliques à des échelles de temps beaucoup plus importantes, pendant lesquelles se manifesteront des successions ou émergences d'espèces liées aux phénomènes de spéciation et d'extinction.

La sélection naturelle permet d'augmenter la représentation des individus porteurs des variations avantageuses au sein d'une population dans un environnement donné, autrement dit leur valeur sélective. La valeur sélective, ou *fitness*, d'un individu se mesure par le nombre de descendants viables et fertiles produits à la génération suivante. Elle est ainsi dépendante de sa survie jusqu'au stade adulte et de sa fertilité. Une définition simple de la *fitness* est donc la capacité des organismes à survivre et à se reproduire. Afin d'augmenter survie et reproduction de l'organisme dans le milieu, l'ajustement fonctionnel évoqué plus haut doit s'effectuer sur des traits liés au succès de ces deux éléments, autrement appelés traits d'histoire de vie. Chez les agents phytopathogènes, plusieurs traits, comme le taux de reproduction, le taux de multiplication ou l'efficacité de l'infection, sont classiquement utilisés pour mesurer la *fitness* du pathogène (Nelson 1979; Leonard & Czochor 1980; Antonivics & Alexander 1989).

On distingue trois types de sélection : directionnelle, divergente et stabilisante (Fig. 1.1). La sélection directionnelle favorise les individus présentant un même phénotype extrême. Elle peut être négative (ou purifiante), quand elle élimine un variant conférant un désavantage sélectif, ou au contraire positive, quand elle augmente la fréquence d'un variant avantageux qui peut à terme se fixer dans la population. La sélection divergente (ou disruptive) est un cas particulier de sélection positive qui favorise les phénotypes des deux extrêmes au détriment des phénotypes intermédiaires. Finalement, la sélection stabilisante maintient le polymorphisme au niveau d'un locus, menant à un excès de variants aux fréquences intermédiaires dans la population et à l'élimination des phénotypes extrêmes. Elle peut être due à un avantage sélectif aux hétérozygotes (on parle dans ce cas de sélection balancée), à de la sélection fréquence-dépendante négative, situation dans laquelle les allèles rares sont avantagés, ou encore à de la sélection dont l'intensité varie dans le temps et/ou l'espace, avantageant alors différents génotypes dans différents environnements.



Figure 1.1 Les trois types de sélection naturelle. Une population de souris présentant une variation quantitative de la coloration du pelage peut être affectée par trois types de sélection : (**A**) directionnelle, (**B**) divergente, (**C**) stabilisante. Les flèches blanches représentent les pressions de sélection exercées contre certains phénotypes.

Quatre autres forces évolutives sont impliquées dans la variation des fréquences alléliques dans les populations.

La mutation correspond à une modification héritable et aléatoire de la molécule d'ADN, qui peut être bénéfique, neutre ou néfaste pour l'organisme porteur. C'est la source ultime de la variation génétique observée dans les génomes, sur laquelle va agir la sélection. Les taux de mutation varient d'une espèce à l'autre et d'un locus à l'autre. Celui du nématode modèle *Caenorhabditis elegans* est par exemple estimé à 3,1 (± 1,9) x 10⁻⁹

mutations/site/génération (Denver *et al.* 2009), ce qui est très faible par rapport aux taux de mutation d'autres agents pathogènes, de l'ordre de 10⁻⁵ par exemple chez les virus à ARN (Drake & Holland 1999). La probabilité d'apparition d'une mutation est donc relativement rare, mais elle peut devenir un évènement important dans de grandes populations.

La recombinaison est une seconde source de variation génétique et est associée à la reproduction sexuée. Elle crée en effet de nouvelles combinaisons d'allèles (plutôt que de nouveaux allèles) qui génèrent à leur tour de nouveaux phénotypes pouvant présenter des valeurs sélectives variables et sur lesquels la sélection peut agir. La probabilité de recombinaison dans un génome est inversement corrélée à la distance séparant les polymorphismes.

La dérive génétique représente la fluctuation aléatoire des fréquences alléliques au sein d'une population, phénomène induit par l'échantillonnage au hasard de gamètes destinés à former la génération suivante. L'importance de la dérive est liée à la taille des populations : plus une population présente un effectif faible, plus la probabilité que la dérive mène à la fixation ou la disparition d'un allèle est élevée (Fig. 1.2). Ce phénomène tend ainsi à éroder la variation génétique, qu'elle soit neutre, favorable ou défavorable, au sein d'une population, et peut ainsi favoriser la divergence génétique entre populations.



Figure 1.2 Illustration du phénomène de dérive génétique. Évolution au cours de 20 générations de 20 allèles à 20 locus indépendants, dont les fréquences initiales sont de 50% dans (**A**) une population de 9 individus et (**B**) une population de 50 individus. La force de la dérive est inversement proportionnelle à la taille de la population (N), ce qui a conduit à la fixation et la disparition de plusieurs allèles dans la plus petite population.

La migration correspond aux flux de gènes réalisés entre populations indépendantes, par déplacement d'individus ou de gamètes. Les migrants peuvent introduire dans les populations des allèles initialement absents, et ainsi y faire augmenter la variation génétique. Ce processus tend à homogénéiser les populations et s'oppose donc à l'effet antagoniste de la dérive génétique. Dans beaucoup de systèmes biologiques, les taux de migration dépassent de plusieurs ordres de grandeur les taux de mutation et de recombinaison, ce qui en fait la source première de modifications de fréquences alléliques sur de courtes échelles de temps, et leur confère une importance prépondérante dans le contournement des résistances. Un cas particulier de migration est la colonisation d'un nouvel habitat. Lorsque des individus migrent pour fonder une nouvelle population à distance de la population d'origine, il est probable qu'ils n'emportent qu'une partie des allèles existant dans leur population d'origine. Cet « effet fondateur » entraîne des conséquences similaires à celles provoquées par un goulot d'étranglement génétique (ou « *bottleneck effect* »), qui survient lorsqu'une proportion substantielle de la population disparaît ou est empêchée de se reproduire, à savoir une diminution de la diversité dans la nouvelle population (**Fig. 1.3**).



Figure 1.3 Illustration de l'effet *bottleneck*. Faire tomber quelques billes à travers le goulot d'une bouteille est analogue à une réduction drastique de la taille de la population. Par le biais du hasard, les allèles bleus sont surreprésentés dans la nouvelle population, tandis que les allèles jaunes ont été perdus.

La génomique des populations permet d'aborder des questions essentielles en biologie évolutive (Orr 2005; Stinchcombe & Hoekstra 2008) : Quelle est l'origine des variants adaptatifs sélectionnés dans un nouvel environnement ? Préexistent-ils dans la population avant le changement d'environnement (*standing variation*) ou émergent-ils après celui-ci (*new mutation*) ? Quels sont le nombre et les effets des mutations impliquées dans l'adaptation ? Celle-ci repose-t-elle sur la sélection d'une seule mutation à large effet ou provient-elle au contraire de nombreuses mutations à effet faible ? L'évolution de traits similaires dans différentes lignées est-elle le produit de la sélection d'une même mutation ? Existe-t-il des régions du génome plus impliquées que d'autres dans le phénomène adaptatif, autrement dit quelle est la nature du mécanisme adaptatif ? Les mutations affectent-elles principalement des régions intragéniques ou d'autres régions, telles que les régions avoisinantes impliquées dans la régulation de la transcription ?

Autant de questions centrales dans le contexte de notre étude, et que les outils de génomique des populations récemment développés et décrits dans la section suivante permettent d'aborder.

1.1.2 Détecter l'adaptation dans les génomes

On distingue globalement trois grands types d'approches pour identifier les loci responsables de l'adaptation (Stinchcombe & Hoekstra 2008; Nadeau & Jiggins 2010; Pardo-Diaz *et al.* 2015) :

• Les approches de *forward genetics* : Ces approches cherchent de manière générale à identifier les gènes contrôlant un trait adaptatif connu, par l'utilisation de méthodes de génétique d'association (*QTL mapping, association mapping* ou *Genome-wide association study*) (Stinchcombe & Hoekstra 2008; Nadeau & Jiggins 2010; Stapley *et al.* 2010). Le principe est ici de génotyper à un grand nombre de marqueurs répartis sur le génome des individus ségrégeant pour le trait d'intérêt, afin d'identifier une régression positive entre les génotypes des individus au niveau d'un marqueur et la valeur de leur phénotype. Il est alors possible de conclure que le marqueur identifié se trouve à proximité du locus ou QTL (*Quantitative Trait Locus*) impliqué dans l'expression du phénotype en question. Ces approches sont parfois nommées scans génomiques d'association.

• Les approches de *reverse genetics* : Elles se réfèrent aux approches de scans génomiques de sélection, que nous nommerons simplement dans la suite de ce manuscrit scans génomiques. Ces approches cherchent à détecter les signatures laissées par la sélection dans le génome. Contrairement aux approches de *forward genetics*, elles ne requièrent pas de connaissance *a priori* sur le phénotype associé (Luikart *et al.* 2003; Storz 2005; Bonin 2008; Ellegren & Sheldon 2008; Stinchcombe & Hoekstra 2008; Stapley *et al.* 2010). Le génotypage d'individus à un grand nombre de marqueurs est ici aussi la règle. Sans données phénotypiques, la détection des loci sous sélection repose uniquement sur des analyses statistiques des fréquences alléliques au niveau des marqueurs. Ces approches sont au cœur d'un champ de recherche ayant émergé récemment : la génomique des populations.

• Les approches gènes candidats : Cette méthode repose sur une connaissance *a priori* de gènes impliqués dans la formation de phénotypes adaptatifs chez des organismes proches de l'organisme d'intérêt, qui seront ainsi explorés chez ce dernier.

Dans notre cas, des limites techniques, concernant le matériel biologique, et théoriques, sur la connaissance de gènes ou régions génomiques pouvant potentiellement intervenir dans l'adaptation, nous permettaient difficilement d'utiliser les approches de génétique d'association et de gènes candidats. Nous avons donc choisi de nous tourner vers une approche sans *a priori* de scan génomique, explicitée plus en détail dans la suite de ce chapitre.

• Génomique des populations et signature moléculaire de la sélection

Les avancées récentes dans les champs de la bioinformatique et des techniques de biologie moléculaire (et en particulier les techniques de séquençage haut-débit, ou NGS pour *Next-Generation Sequencing*) ont conduit à l'émergence de la génomique des populations. Cette discipline consiste à appliquer les concepts développés en génétique des populations à des jeux de données génomiques. En effet, alors que la génétique des populations se base sur

le génotypage d'individus à quelques loci polymorphes, la génomique permet aujourd'hui d'avoir accès à l'information sur la variabilité génétique à l'échelle du génome entier. Ainsi, la génomique des populations peut se définir au sens large comme l'étude simultanée de nombreux loci (de plusieurs dizaines à plusieurs millions) dans le but d'appréhender le rôle des différentes forces évolutives sur les variations observées au sein des génomes et des populations, pour une meilleure compréhension de leur microévolution (Luikart *et al.* 2003). Selon une définition plus restrictive, cette discipline vise à identifier et différencier les effets locus-spécifiques (*i.e.* adaptatifs) des effets impactant l'ensemble du génome (*i.e.* démographiques) (Black *et al.* 2001; Luikart *et al.* 2003). C'est dans ce contexte que la réalisation de scans génomiques prend tout son sens, puisqu'ils permettent de répondre au principal objectif de la génomique des populations, en identifiant les régions du génome impactées par la sélection naturelle.

Cette démarche s'avère cruciale, pour deux raisons. D'une part, comme nous l'avons déjà évoqué, car c'est l'aspect qui nous intéresse ici, elle permet de cibler les gènes impliqués dans les adaptations. D'autre part, elle est aussi d'un intérêt majeur pour l'amélioration des inférences sur l'histoire démographique et évolutive des populations. En effet, ce type d'étude nécessite le retrait des loci sous sélection dans les jeux de données, puisque seuls les processus impactant le génome dans sa globalité informent de manière fiable sur la démographie et l'histoire phylogénétique des populations. La présence de loci sous sélection dans les jeux de données paramètres des populations, tels que les *F*_{ST}, les taux de migration, les tailles des populations et la phylogénie (Luikart *et al.* 2003). Ainsi, historiquement, les recherches de traces de sélection dans les génomes ont avant tout été réalisées dans le but de supprimer les loci non neutres des études (Beaumont 2005). Aujourd'hui, leur intérêt pour l'identification des gènes responsables de l'adaptation est largement reconnu et exploité.

En plus de répondre aux enjeux théoriques de compréhension de l'évolution, cette nouvelle discipline a d'importantes applications dans de nombreux domaines de recherche. Un sujet de prédilection de la génomique des populations est l'espèce humaine et en particulier les recherches en pharmacogénomique, ce qui explique que dans les années 2000, près des trois quarts des études de scan génomique portaient sur l'espèce humaine (Haasl & Payseur 2016). L'un des objectifs de ces études est la recherche de gènes de résistance aux maladies ou associés à la prédisposition à certaines maladies (Kimura *et al.* 2007). La génomique des populations intervient également dans le domaine de la conservation des espèces (Bonin *et al.* 2007; Russelo *et al.* 2012), dans l'étude de l'adaptation des pathogènes aux traitements chimiques (Bonin *et al.* 2009) et dans l'étude de la domestication (Rubin *et al.* 2010, 2012). Les méthodes employées en génomique des populations permettent d'étudier un large panel d'organismes, et les études ne se restreignent aujourd'hui plus uniquement à quelques espèces modèles. Ceci se vérifie par le fait que depuis 2010, ce sont désormais moins de la moitié des études de scans génomiques qui concernent l'Homme (Haasl & Payseur 2016).

La génomique des populations repose sur deux principes fondamentaux. D'une part, nous l'avons vu, le génome est affecté dans son ensemble par la démographie et l'histoire évolutive des populations, tandis que la sélection n'affecte qu'une partie des loci, qui présentent alors des patrons de variations différents du reste du génome. D'autre part, la sélection d'un locus avantageux affecte également les loci avoisinants. L'augmentation en fréquence de l'allèle sélectionné au locus concerné entraîne en effet une augmentation de la fréquence des allèles des loci auxquels il est lié (*i.e.* avec lesquels il est en déséquilibre de liaison), par un effet d'autostop moléculaire, ou *hitchhiking effect* (Maynard-Smith & Haigh 1974) (Fig. 1.4). Cet effet se traduit par une baisse, voire une élimination de la diversité dans ces régions génomiques, un processus appelé balayage sélectif (*selective sweep*). Le déséquilibre de liaison qui existe entre le locus impliqué dans l'adaptation et les loci proches fait que ces derniers renverront un signal similaire au premier. Ce phénomène crée ainsi un îlot de différenciation génétique autour du locus impliqué dans l'adaptation, véritable signature de la sélection dans le génome. Par l'action de la recombinaison génétique, cette signature s'estompe progressivement le long du chromosome et au cours du temps.



Figure 1.4 Illustration de l'autostop moléculaire ou hitchhiking effect (d'après Storz 2005). Les lignes horizontales représentent des chromosomes homologues échantillonnés dans une population et les ronds noirs correspondent à des mutations neutres. Le rond blanc correspond à une mutation avantageuse qui apparaît puis se fixe rapidement dans la population sous l'effet de la sélection positive, entraînant avec elle une portion de l'haplotype ancestral, représenté en rouge. Les variants neutres initialement liés à la mutation avantageuse augmentent alors en fréquence.

La signature de la sélection n'est pas la même si elle concerne une nouvelle mutation ou une mutation préexistante dans la population. Dans le cas d'une nouvelle mutation qui se fixe rapidement dans la population, la diversité sera éliminée ou très fortement réduite sur une large région génomique, car la recombinaison ne peut pas avoir lieu dans un court intervalle de temps. On parle alors de *hard sweep*. Dans le cas d'une adaptation à partir d'un polymorphisme préexistant, ce variant favorable plus ancien a déjà pu recombiner avec les sites proches. La diversité est donc moins érodée dans cette région génomique et le balayage sélectif est localisé sur une région plus restreinte. On parle ici de *soft sweep*.

La détection précise du locus sous sélection par le scan génomique n'est pas obligatoire, puisqu'il ne fait pas forcément partie du panel de marqueurs utilisés. En revanche, il est probable qu'un des marqueurs soit en déséquilibre de liaison avec le locus sous sélection, et qu'il soit donc détecté lors du scan génomique. La détection d'un signal de sélection est ainsi maximisée par l'étendue du déséquilibre de liaison et la densité des marqueurs dans le génome. La signature de la sélection dans le génome se retrouve au niveau de quatre éléments :

• la diversité génétique, qui se trouve diminuée dans la région impactée (Fig. 1.5) ;

• le déséquilibre de liaison, qui augmente comme nous l'avons vu autour du locus sous sélection ;

• le spectre de fréquences de sites polymorphes (SFS, pour *Site Frequency Spectrum*). Celui-ci indique pour chaque fréquence possible la proportion de sites polymorphes où l'allèle est trouvé à cette même fréquence. Cette relation est inversement proportionnelle en absence de sélection. La sélection induit une déviation de ce SFS : lorsqu'elle est négative, le nombre d'allèles présents à faible fréquence augmente, au détriment de ceux présents à des fréquences intermédiaires ou fortes. Un balayage sélectif a globalement les mêmes effets sur le SFS, avec en plus un excès d'allèles à fortes fréquences. Au contraire, en cas de sélection positive, le nombre d'allèles présents en fortes fréquences augmente ;

• la différenciation génétique entre populations, qui augmente dans la région sous sélection (Fig. 1.5). En effet, un locus impacté par la sélection dans une population présente une variation génétique différente du même locus non impacté par la sélection dans une autre population. Les deux populations sont donc génétiquement plus différenciées au niveau de ce locus (et aux polymorphismes liés) qu'au niveau d'autres loci non impactés par la sélection dans les deux populations.

• Les outils de la génomique des populations

Les épisodes adaptatifs laissent des signatures spécifiques dans les génomes. Les stratégies utilisées pour les identifier reposent sur le même point fondamental : la vaste majorité des loci génotypés suit le même modèle de neutralité, où la variation s'équilibre à une valeur moyenne dépendant du taux de mutation et de la taille de la population. Il est ainsi possible de prédire les caractéristiques attendues du patron de polymorphisme sous une hypothèse de neutralité (l'hypothèse nulle au sens statistique), et de vérifier la conformité des données observées le long du génome. Les marqueurs validant l'hypothèse alternative sont appelés « *outliers* » et sont indicatifs de régions génomiques potentiellement impactées par la sélection.

À l'échelle intra-populationnelle, les méthodes utilisées se basent notamment sur la distribution du polymorphisme (SFS) et/ou la mesure de la diversité génétique. Les tests les plus répandus sont le test d'Ewens-Watterson (Ewens 1972; Watterson 1978), le test HKA (Hudson *et al.* 1987), le test du D de Tajima (Tajima 1989) ainsi que ses extensions (Fu & Li 1993; Fay & Wu 2000). D'autres tests de neutralité sont basés sur l'évaluation du déséquilibre de liaison et l'analyse des haplotypes, tels que le test proposé par Kim & Nielsen (2004), une extension du test de Kim & Stephan (2002) initialement basé sur la distribution spatiale des sites polymorphes et le SFS, auquel les auteurs ont incorporé l'information sur la mesure du déséquilibre de liaison. Sabeti *et al.* (2002) ont également proposé un test basé sur la mesure de l'homozygotie haplotypique étendue (EHH, pour *Extended Haplotype Homozygosity*) permettant d'identifier les haplotypes caractérisés par de nombreux sites homozygotes (dus à la perte de diversité génétique) et présents en fortes fréquences dans la population.



Population B

Figure 1.5 Conséquences d'un balayage sélectif sur la diversité nucléotidique aux sites neutres liés (d'après Storz 2005). Le locus 3 a subi un balayage sélectif dans la population B. Par rapport aux autres loci non liés (1, 2 et 4), ce locus est caractérisé par une réduction de diversité nucléotidique intra-population (π_s) et une augmentation de diversité nucléotidique inter-populations (F_{ST}) (*i.e.* augmentation de la différenciation génétique entre les deux populations).

À l'échelle inter-populationnelle, la gamme des méthodes se divise en trois catégories, que nous avons choisi d'exploiter au cours de notre étude.

La première catégorie, et la plus ancienne, est composée de tests basés sur le calcul du F_{ST} (ou de statistiques proches) afin de déterminer si des loci sont sous sélection différentielle entre populations (ou plus précisément, si ces loci sont plus différenciés que ne l'est le reste du génome, supposé neutre). Le F_{ST} , ou indice de fixation de Wright, peut se définir comme le rapport de la variabilité des fréquences alléliques entre sous-populations sur la variabilité des fréquences alléliques au niveau de la population totale. Cet indice varie de zéro à un. Plus sa valeur est proche de zéro, plus les populations sont génétiquement différenciées. En pratique, la différenciation génétique des populations est considérée comme importante dès que F_{ST} >

0.1. L'approche basée sur le calcul du F_{ST} a été initiée par Lewontin & Krakauer dès 1973, sur une idée proposée par Cavalli-Sforza (1966). Le test proposé par Lewontin & Krakauer (1973) vise à comparer les valeurs des F_{ST} estimées à plusieurs loci à une distribution théorique de ces valeurs sous un modèle neutre. Ce modèle a cependant été remis en question (Nei & Maruyama 1975; Robertson 1975), n'étant pas assez robuste et généralisable. La déviation au modèle neutre a de grandes chances de mener à la détection de faux positifs, or le modèle neutre sous-jacent à ce test ne permet pas de modéliser l'ensemble de la diversité des histoires démographiques et des caractéristiques génétiques des populations. Suite à ce discrédit, l'approche engagée par Lewontin & Krakauer (1973) a été abandonnée pendant une longue période. À partir des années 1990, un regain d'intérêt pour cette approche s'est traduit par le développement de nombreuses méthodes basées sur celle-ci. Ces méthodes présentent différentes approches statistiques ou modèles d'estimation de la distribution neutre, mais sont de manière générale basées sur le développement de modèles de génétique des populations suffisamment robustes pour servir de modèle neutre, dans le but de diminuer le nombre de faux positifs. Une des premières méthodes fut ainsi proposée par Beaumont & Nichols (1996) à travers le test FDIST. Parce que de nombreuses espèces montrent une structuration complexe de leurs populations, Excoffier et al. (2009) ont étendu FDIST de manière à prendre en compte la hiérarchie dans les populations. Bonhomme et al. (2010) ont adapté sur la même idée le test de Lewontin & Krakauer (1973). Vitalis et al. (2001) ont également proposé le test DetSel, consistant à restreindre l'analyse à des scénarios simples, mais réalistes, et à des paires de populations, pour ainsi éviter de devoir émettre des hypothèses sur la structure des populations et leur histoire. Plusieurs méthodes basées sur des modèles de vraisemblance, généralement dans un cadre hiérarchique bayésien, ont aussi vu le jour. Une des premières fut proposée par Beaumont & Balding (2004) et a ensuite permis l'émergence de nombreuses autres méthodes (Riebler et al. 2008; Guo et al. 2009; Gompert & Buerkle 2011), notamment BayeScan (Foll & Gaggiotti 2008) ou SelEstim (Vitalis et al. 2014). Dernièrement, certaines méthodes ont abandonné les F_{ST} et leurs préfèrent l'utilisation de l'analyse en composante principale, telles que PCAdapt (Duforet-Frebourg et al. 2014).

La deuxième catégorie de méthodes a été initiée par Schlötterer (2002) et se base cette fois sur la diminution de la variabilité génétique au sein des populations. Au lieu d'utiliser des simulations pour générer la distribution attendue sous l'hypothèse de neutralité comme précédemment, l'approche empirique de Schlötterer vise à comparer la diversité génétique entre deux échantillons en calculant la variance (V) dans la taille des allèles (ce test ayant été initialement développé pour les microsatellites) (Schlötterer 2002), ou l'hétérozygotie attendue (H) (Kauer *et al.* 2003; Schlötterer & Dieringer 2005). Il a en effet été montré que les logarithmes naturels des ratios V et R sont normalement distribués sous différents scénarios démographiques. De la même manière que précédemment, les *outliers* sont les marqueurs sortant des intervalles de confiance.

Le troisième type d'approche repose sur la corrélation entre fréquences alléliques et variables environnementales, et se rapprochent en cela des études de génétique d'association génotype-phénotype, où le phénotype représente ici les variables de l'environnement (Eckert *et al.* 2010; Bradbury *et al.* 2013). Ces approches sont basées sur un modèle développé par Nicholson *et al.* (2002) calculant le patron de covariance des fréquences alléliques entre populations, qui permet *in fine* de tester l'effet linéaire de variables environnementales sur les fréquences alléliques. Une relation significative traduirait alors une sélection différentielle entre populations. Coop *et al.* (2010) ont développé le test BayEnv, basé sur ce modèle. Plusieurs auteurs ont étendu cette approche, la plus récente étant BayPass (Gautier 2015).

Des méthodes telles que LFMM (Frichot *et al.* 2013) et gINLAnd (Guillot *et al.* 2014) ne font pas appel à des modèles d'évolution neutre, mais corrigent statistiquement pour la structure de population.

Trois types de marqueurs moléculaires sont particulièrement employés dans les scans génomiques (Luikart *et al.* 2003) : les AFLPs (*Amplified Fragment-Length Polymorphisms*), les microsatellites et les SNPs (*Single Nucleotide Polymorphisms*). Lors du développement des premières méthodes de scans génomiques, seuls les deux premiers types de marqueurs étaient accessibles pour caractériser la variation génétique dans les génomes. Cependant, leur nombre restreint (quelques centaines) ne permettaient que de couvrir peu densément les génomes, rendant ardue la détection des régions impactées par la sélection. En effet, la précision des analyses génomiques est d'autant plus grande que le nombre de marqueurs est important. Le développement des techniques de séquençage haut-débit, la chute de leur coût et de la durée de la procédure permet aujourd'hui de cribler beaucoup plus densément le génome, et apporte donc des résultats plus précis, en donnant l'accès aux polymorphismes les plus fréquents au sein des génomes, les SNPs, qui sont aujourd'hui des marqueurs de référence.

Limites des approches de scan génomique

L'accessibilité à un nombre croissant de données génomiques et le développement de diverses méthodes d'analyses permettant d'y détecter les signaux de la sélection ont abouti ces dernières années à de nombreuses applications empiriques, pour une large gamme d'organismes, permettant de commencer à lever le voile sur les bases génétiques de l'adaptation. Cependant, ces études ont dans le même temps mis en évidence les limites de ces approches.

Plusieurs effets confondants peuvent produire au sein de régions génomiques un patron de variation semblable à ce qui est observable lors de balayages sélectifs, affectant alors le résultat des scans génomiques en augmentant le nombre de faux positifs. Un des effets à présent bien connu est l'existence possible d'une structuration cachée dans les populations qui, si elle n'est pas prise en compte dans les analyses, du fait de l'utilisation de modèles trop simplifiés, peut biaiser l'estimation des F_{ST} et ainsi aboutir à une augmentation du nombre de faux positifs détectés (Excoffier et al. 2009; Bierne et al. 2013). L'histoire démographique des populations est également un élément à prendre en compte dans des approches de scan génomique, puisque des changements dans les tailles des populations peuvent aboutir à de faux positifs (Teshima et al. 2006; Hermisson 2009). Par exemple, la survenue d'un bottleneck peut se traduire par une augmentation de la variance dans l'estimation des valeurs des FST, mimant alors un évènement de sélection. Pour des populations en expansion, il peut exister une différence de taille efficace entre le cœur de la population et le front de migration, où celle-ci est réduite. Ainsi, par dérive génétique, un allèle rare peut plus facilement augmenter en fréquence et se fixer au niveau du front de migration. Ce processus appelé surfing induit une structuration géographique du polymorphisme génétique, mimant l'effet d'une sélection associée à l'adaptation à de nouvelles conditions environnementales (Klopfstein et al. 2006; Excoffier & Ray 2008). Le mode de reproduction est aussi un paramètre à considérer, puisqu'il détermine l'importance du déséquilibre de liaison dans les génomes. En effet, la reproduction sexuée permet la recombinaison génétique qui réduit le déséquilibre de liaison, restreignant ainsi le signal de sélection dans une portion relativement faible du génome (de plus en plus faible au fil des recombinaisons), tandis que la reproduction asexuée induit un déséquilibre de liaison très fort qui peut se propager au génome entier, et ainsi rendre difficile l'identification de la région génomique impactée par la sélection. D'autre part, la consanguinité et son cas extrême, l'autofécondation, entraînent un déficit en hétérozygotes dans les populations, conduisant à l'augmentation du taux de faux positifs (De Mita *et al.* 2013). Un autre effet confondant peut être l'existence de barrières d'incompatibilité génétique entre populations. En effet, des incompatibilités génétiques intrinsèques, pré- ou post-zygotiques, et donc indépendantes de l'environnement, peuvent créer des barrières endogènes aux flux de gènes qui peuvent colocaliser dans l'espace avec des barrières (Bierne *et al.* 2011) peut aboutir à la détection d'*outliers* qui ne sont pas impliqués dans l'adaptation à l'environnement, mais qui découleraient d'incompatibilités intrinsèques pouvant pré-dater l'apparition de différences écologiques entre populations.

D'autres facteurs moins fréquemment cités dans la littérature peuvent potentiellement réduire la puissance de détection des tests et mener à de faux négatifs (Haasl & Payseur 2016), et doivent donc être considérés. La sélection cible des variants qui apparaissent dans le génome par divers processus mutationnels (substitution, insertion, délétion, transposition, inversion), mais ceux-ci affectent différemment la signature de la sélection. Il est par exemple connu que des scans génomiques employant des marqueurs microsatellites ont une puissance de détection réduite (Haasl et al. 2014). De plus, il existe également une hétérogénéité dans le taux de mutation des différents variants : chez l'homme par exemple, le taux de mutation des SNPs est de l'ordre de 10⁻⁸-10⁻⁹ mutations/site/génération (Nachman & Crowell 2000), tandis que celui des microsatellites 10⁻² à 10-6 mutations/site/génération, varie de et est en moyenne de 10⁻⁴ mutations/site/génération (Selkoe & Toonen 2006). Comme les mutations, les taux de recombinaison sont hétérogènes au sein des génomes. On définit ainsi des régions à fort taux de recombinaison, ou hotspots, et des régions dans lesquelles les taux de recombinaison sont beaucoup plus faibles, les coldspots. L'identification de régions sous sélection reposant le plus souvent sur la détection de loci outlier liés à la mutation sélectionnée et entraînés par hitchhiking, celle-ci sera rendu plus facile si elle a lieu dans une région à faible taux de recombinaison, tandis qu'une région à forte recombinaison effacera plus rapidement le signal de sélection aux loci avoisinant la mutation sélectionnée. Cependant, l'hétérogénéité des taux de recombinaison peut aussi aboutir à de faux positifs, par l'action de la sélection purifiante. En effet, les régions recombinant peu, telles que les régions centromériques, favorisent la sélection purifiante et présentent ainsi une diversité réduite. Ceci peut alors induire l'augmentation en fréquence d'allèles différents dans des populations distinctes, et donc l'augmentation de la différenciation génétique entre les deux populations dans la région génomique concernée, de façon semblable à une signature de sélection divergente (Charlesworth et al. 1997; Noor & Bennett 2009; Turner & Hahn 2010). Aujourd'hui, les études permettant de comprendre comment varie le taux de recombinaison le long du génome et dans les régions candidates à l'adaptation et la spéciation manquent encore. Enfin, l'architecture génétique du trait influe également sur la puissance de détection. Quand la sélection cible des phénotypes complexes, c'est-à-dire dont la variation dépend de l'action de plusieurs loci, les tests sont peu puissants pour les détecter. En effet, plus le nombre de mutations impliquées augmente, plus l'intensité de la sélection appliquée à chacune des mutations diminue, et la signature de sélection s'en retrouve atténuée (Pritchard & Di Rienzo 2010; Le Corre & Kremer 2012).

Les tests de neutralité utilisés peuvent également être responsables de biais dans les résultats. Il a été mis en évidence que l'utilisation de mesures relatives de détection des *outliers*, tels les *F*_{ST}, peut biaiser les résultats des scans génomiques. En effet, le calcul du *F*_{ST} étant basé sur le ratio entre diversité intra-population et diversité globale, la diminution du premier terme via certains mécanismes (différents de l'adaptation) entraîne l'augmentation de la valeur du *F*_{ST} et aboutit à la détection de faux positifs (Noor & Bennet 2009; Nachman & Payseur 2012; Cruickshank & Hahn 2014). Il a donc été suggéré que des mesures absolues de la divergence génétique, indépendantes des niveaux de diversité des populations - comme la distance Dxy - seraient préférables, néanmoins celles-ci ne se montrent globalement pas performantes pour détecter les régions sous sélection (Cruickshank & Hahn 2014).

La puissance des différentes approches de scan génomique dépend également de la force de la sélection. Les méthodes basées sur l'estimation des F_{ST} échouent globalement à détecter des vrais positifs quand la sélection est faible, tandis que les méthodes basées sur la corrélation entre fréquences alléliques et variables environnementales sont robustes aux différentes forces de sélection. En revanche, elles sont sensibles aux corrélations génétiques entre populations si celles-ci ne sont pas corrigées (De Mita et al. 2013). Les résultats obtenus avec les méthodes basées sur l'estimation des F_{ST} sont également très dépendants des modèles démographiques implémentés (De Mita et al. 2013; Lotterhos & Whitlock 2014). Comme aucun des modèles démographiques ne correspond jamais parfaitement à la situation du jeu de donné testé, il est recommandé d'utiliser plusieurs tests de détection d'outliers, qui permettront de donner plus de poids aux outliers détectés dans les différents tests (Luikart et al. 2003; Vasemägi & Primmer 2005; Barret & Hoekstra 2011). Il est ainsi possible de combiner plusieurs tests d'une même « famille », tels que plusieurs tests basés sur des comparaisons inter-populationnelles, mais aussi d'associer des tests inter-populationnels et des tests intrapopulationnels, ou encore d'associer des approches de forward genomics et de reverse genomics (Stinchcombe & Hoekstra 2008; Butlin 2010; Hohenlohe et al. 2010). Cette dernière proposition a par exemple été menée chez des populations d'épinoche Gasterosteus aculeatus, chez lesquelles l'identification du gène Eda comme responsable de la réduction adaptative, chez les formes d'eau douce, des plaques osseuses au niveau de l'armure (une protection contre la prédation par les vertébrés, moindre en eau douce) a été réalisée en combinant notamment une approche de QTL mapping et une étude du déséquilibre de liaison dans la région candidate, avec marqueurs microsatellites (Colosimo et al. 2005).

Des problèmes techniques liés aux NGS peuvent également mener à la détection de faux positifs. L'ensemble des techniques NGS présentent des taux d'erreurs lors de la lecture des bases. Elles peuvent notamment créer des SNPs artificiels en remplaçant une base par une autre, ou effectuer des insertions-délétions. Les approches basées sur le séquençage de représentations réduites du génome, telles que le RADseq (*Restriction site Associated DNA sequencing*) ou le GBS (*Genotyping-By-Sequencing*), présentent également des erreurs propres. L'un des biais majeurs de ces approches est la génération d'allèles nuls, ou *allele dropout* (Arnold *et al.* 2013; Gautier *et al.* 2013), causée par une mutation sur le site de restriction, qui empêche alors l'action de l'enzyme. Si cette mutation sur le site de restriction est spécifiquement associée à un allèle d'un polymorphisme proche, alors ce dernier sera absent du jeu de données final, ce qui biaisera l'estimation des fréquences alléliques en simulant une différenciation génétique virtuellement élevée entre populations. Plusieurs méthodes existent pour corriger les biais liés à ces problèmes techniques.

Une remarque générale souvent soulevée concernant les perspectives des scans génomiques est le fait que malgré les centaines d'études réalisées jusqu'à aujourd'hui et ayant mené à l'identification de régions candidates à l'adaptation, très peu aboutissent à l'établissement d'un lien entre génotype, phénotype et *fitness* de l'individu. La plupart des études produisent de longues listes de gènes candidats qui ne sont pas suivies d'une validation fonctionnelle. Cette dernière étape reste en effet délicate. Les *outliers* mis en évidence dans les études ne sont pas dans la plupart des cas la mutation responsable de l'adaptation, mais simplement des loci liés. L'étendue du déséquilibre de liaison peut rendre difficile la localisation de la mutation responsable. De plus, l'approche de scan génomique étant une approche sans *a priori*, elle est souvent réalisée sans aucune connaissance du phénotype. Il est donc difficile de mettre en évidence les traits phénotypiques qui diffèrent entre populations, qui pourraient en outre donner des indices sur les potentiels gènes sous-jacents.

1.1.3 L'approche d'évolution expérimentale pour étudier l'adaptation

Une majorité des études de génomique des populations est aujourd'hui réalisée sur des populations naturelles. Cependant, les études basées sur des évolutions expérimentales peuvent également être d'un grand intérêt pour élucider les bases moléculaires de l'adaptation, et font de ce fait l'objet d'un intérêt croissant (Bank *et al.* 2014).

Une étude classique d'évolution expérimentale consiste à exposer, pendant un certain nombre de générations, une souche d'organisme (typiquement virus, bactérie et autres microorganismes, ou encore mouche, *i.e.* Drosophile) à un nouvel environnement, potentiellement contraignant, puis d'en évaluer les conséquences sur sa *fitness* (Bank *et al.* 2014). La petite taille de ces organismes permet de réaliser un grand nombre de cultures en parallèle et leur reproduction généralement rapide permet de suivre l'évolution significative des populations sur un grand nombre de générations. Un exemple célèbre est l'expérience lancée par Richard Lenski en 1988 sur une souche d'*Escherichia coli (e.g.* Lenski *et al.* 1991), ayant abouti après plus de 25 ans à plus de 50 000 générations. Les expériences sont cependant généralement loin d'être aussi longues, et des expérimentations à plus courts termes s'intéressent à des périodes d'adaptation plus courtes, dans le but de caractériser les effets de la microévolution.

L'intérêt de cette approche est de pouvoir contrôler expérimentalement l'histoire démographique et les pressions de sélection subies par les populations expérimentales. Elles permettent ainsi de pouvoir tester les théories adaptatives qui auraient pu être difficiles voire impossibles à tester en populations naturelles, pour des raisons logistiques et méthodologiques, notamment à cause d'effets confondants ou de la complexité de potentielles pressions de sélection en jeu.

Appliquées à un contexte de génomique des populations, les évolutions expérimentales présentent de nouveaux intérêts. Elles ne se restreignent plus à évaluer les conséquences de l'adaptation à l'environnement sur la *fitness*, elles permettent également de faciliter l'étude des bases moléculaires (autrement dit, cette fois, des « causes ») de l'adaptation. Elles permettent donc de faire le lien entre la sélection imposée par un environnement donné et les changements génotypiques et phénotypiques observés dans la population (Martins *et al.* 2014). Les intérêts relevés précédemment sont toujours vrais ici, et permettent de diminuer le risque de faux positifs par le contrôle des divers effets confondants.

Les limites globales soulevées à propos de cette approche se réfèrent à l'impossibilité de reproduire l'écologie d'un environnement naturel (Bailey & Bataillon 2016), et également au fait qu'elles ne se restreignent qu'à quelques organismes, majoritairement de petites espèces asexuées qui se multiplient rapidement. Ces espèces sont le plus souvent facilement séquençables et l'accès à la mutation impliquée dans l'adaptation est direct, ne nécessitant pas d'approche de scan génomique. Martins *et al.* (2014) indiquent que peu d'exemples d'études ayant combiné évolution expérimentale et reséquençage complet de génome pour la recherche des bases moléculaires de l'adaptation existent chez des organismes pluricellulaires sexués, chez lesquels les adaptations proviennent plutôt de mutations préexistantes que de nouvelles mutations.

Au cours de ce projet de thèse, nous avons choisi d'appliquer cette approche innovante combinant évolution expérimentale et scan génomique sur des données de génotypage/séquençage. Cette approche représente un challenge car notre modèle d'étude est un organisme pluricellulaire sexué ne produisant qu'une génération par an. De plus, l'évolution expérimentale a duré « seulement » huit ans, soit huit générations, mais a cependant suffi pour mettre en évidence une adaptation phénotypique.

1. 2 Interactions plante-pathogène et durabilité des résistances

Les systèmes plante-pathogène représentent un cadre intéressant pour l'étude de l'adaptation. Dans les agrosystèmes, les résultats de ce type d'étude trouvent en outre des applications concrètes pour une meilleure gestion des résistances. Nous présenterons dans un premier temps la dynamique coévolutive qui régit les interactions plante-pathogène. Nous expliciterons ensuite ces interactions en présentant les stratégies de résistance mises en place dans le règne végétal et le mécanisme de contournement de ces résistances par les pathogènes. Enfin, nous évoquerons les possibles leviers d'actions pour une nécessaire gestion durable des résistances variétales.

1.2.1 La coévolution plante-pathogène

Les plantes subissent les attaques d'une large gamme de pathogènes (virus, viroïdes, oomycètes, champignons, bactéries, insectes, nématodes), qui provoquent parfois le développement de maladies graves pouvant entraîner des pertes économiques considérables. Cependant, malgré une apparente immobilité et passivité face aux agresseurs, les plantes ont développé au cours de l'évolution des systèmes de défense efficaces contre ces agents pathogènes, faisant de la maladie une exception plutôt qu'une règle dans le monde végétal. En réponse, les pathogènes ont évolué pour tenter de contourner les défenses de la plante. Les systèmes hôte-pathogène sont ainsi basés sur la mise en place d'une relation étroite entre les deux protagonistes, où chacun influence l'évolution de l'autre : c'est le phénomène de coévolution, principe explicitement formulé pour la première fois par Ehrlich & Raven en 1964. Le pathogène possède des gènes qui lui permettent de parasiter la plante, tandis que celle-ci possède des gènes de résistance pour lutter contre les attaques d'un spectre large de pathogènes. Ces gènes coévoluent au gré des interactions par les pressions de sélection qu'exerce chaque partenaire sur les gènes de l'autre. Le cadre conceptuel de cette coévolution

a été décrit au travers de modèles théoriques. On se réfère principalement à deux d'entre eux : le modèle de course aux armements et le modèle de guerre des tranchées (Fig. 1.6).

Le modèle de guerre des tranchées repose sur l'existence d'un polymorphisme stable dans les populations hôtes et pathogènes, par l'action de la sélection balancée. Chez l'agent pathogène, l'allèle contribuant le plus à sa *fitness* va augmenter en fréquence dans la population, créant une pression de sélection sur la population hôte. En réponse, l'allèle le plus avantageux de l'hôte va également augmenter en fréquence, tendant à réduire la *fitness* du pathogène et donc la fréquence de l'allèle correspondant. S'en suivra une diminution de fréquence de l'allèle de l'hôte, avant que le cycle ne reprenne. Sur le long terme, ce modèle correspond à une situation de stabilité, où l'évolution repose sur la variation dans le temps des fréquences d'allèles préexistants anciens, plutôt que sur l'apparition de nouveaux allèles (Stahl *et al.* 1999).

Le modèle de course aux armements repose de manière alternative sur l'apparition et la fixation répétée de nouveaux allèles dans les populations hôtes et pathogènes, conduisant dans le même temps au remplacement des allèles alternatifs. L'allèle offrant la meilleure *fitness* au pathogène va augmenter en fréquence dans la population, au détriment des allèles moins adaptés. Face à cette pression de sélection, un nouvel allèle de l'hôte va lui aussi augmenter en fréquence et se fixer. En réponse à la diminution de *fitness* imposée par la pression de l'hôte, un nouvel allèle avantageux apparu dans la population du pathogène sera positivement sélectionné, pour que le cycle se répète à nouveau. Ce modèle instable correspond donc à l'acquisition perpétuelle de nouvelles armes et défenses (Dawkins & Krebs 1979).

Notons qu'un troisième modèle conciliant les deux précédents, le modèle mixte, a également été proposé, permettant notamment d'autoriser l'apparition de nouveaux allèles dans un modèle de type guerre de tranchées.

Dans les écosystèmes naturels, la dynamique de la coévolution plante-pathogène semble suivre largement le modèle de la guerre de tranchées (Brown & Tellier 2011). En revanche, dans les agrosystèmes, le déploiement de gènes de résistance sur de grandes étendues en monoculture entraîne la sélection d'individus pathogènes qui s'y sont adaptés, obligeant l'utilisation d'une nouvelle résistance, qui sélectionnera à son tour de nouveaux individus adaptés. Ce processus est une forme de « course aux armements » (Brown & Tellier 2011), même s'il serait erroné de parler ici de coévolution, l'agriculture moderne intervenant directement sur les caractéristiques de l'hôte, en choisissant et en renouvelant elle-même les génotypes des plantes cultivées. De plus, face à la forte pression de sélection imposée par la résistance de la plante hôte, les populations pathogènes évoluent, mais la population hôte ne peut pas répondre naturellement aux changements de fréquence des allèles de virulence (terme utilisé ici pour désigner la capacité du pathogène à se multiplier sur un hôte résistant), puisqu'elle sera remplacée. On parle ici plutôt d'évolution séquentielle, lorsque l'évolution d'une espèce en influence une autre, sans que la réciproque soit vraie (Ridley 1996).

1.2.2 Les résistances des plantes

Les plantes présentent deux grandes stratégies de défense : la résistance, qui consiste à interdire ou limiter le développement du pathogène et la tolérance, qui consiste à limiter l'impact négatif du développement du pathogène sur les performances de la plante (Schafer 1971; Parlevliet 1979), et qui n'est donc pas une forme de résistance (Simms & Triplett 1994).



Figure 1.6 Course aux armements et guerre des tranchées : les deux scénarios de l'évolution des interactions hôte-pathogène (d'après Terauchi & Yoshida 2010). **A.** Modèles de l'évolution des fréquences alléliques au cours du temps. Ligne continue : fréquence de l'allèle du pathogène. Ligne pointillée : fréquence de l'allèle de l'hôte. **B.** Généalogie d'un locus impliqué dans l'interaction. Ligne continue : généalogie au sein d'une espèce. Ligne pointillée : généalogie chez une espèce proche. Dans le scénario de la course aux armements, un seul allèle (indiqué par « A ») est sélectionné quand il confère une *fitness* plus importante à l'organisme qui le porte. Les autres allèles (indiqués par « x ») sont éliminés de la population. Dans le scénario de la guerre des tranchées, deux allèles (indiqués par « A1 » et « A2 ») ou plus sont maintenus au cours du temps.

Les plantes disposent de deux types de barrières de résistance. En premier lieu, des barrières constitutives préexistent à l'agression. Cette première ligne de défense contre les agents pathogènes est constituée de barrières physiques et/ou chimiques comprenant la cuticule, les parois cellulaires, les enzymes antimicrobiennes ou les métabolites secondaires (Nürnberger & Lipka 2005). Cependant, certains pathogènes peuvent parvenir à passer ces premières défenses en infectant la plante par des ouvertures naturelles telles que les stomates, par des blessures ou encore via l'action d'enzymes hydrolytiques, qui dégradent la cuticule ou la paroi cellulaire ou neutralisent les molécules toxiques en les métabolisant. En cas d'agression intervient alors le deuxième type de barrières végétales, les barrières induites. En effet, les plantes ont développé un système, gouverné par des facteurs génétiques, qui permet à la fois de détecter l'intrusion de pathogènes et d'activer des systèmes de défense faisant intervenir des évènements cellulaires complexes. Ce système est composé de plusieurs niveaux qui reflètent la coévolution des interactions hôte-pathogène, et a été formalisé par Jones & Dangl (2006) dans le modèle dit « en zig-zag » (Fig. 1.7).


Figure 1.7 Modèle en zig-zag illustrant les différents niveaux du système immunitaire des plantes (d'après Jones & Dangl 2006). 1. La plante détecte les molécules caractéristiques des agents pathogènes (PAMPs, losanges rouges) par les récepteurs PRRs, déclenchant la première barrière de l'immunité (PTI). 2. La sélection favorise les individus pathogènes virulents qui produisent des effecteurs (ronds rouges) interférant avec la PTI, rendant la plante sensible (ETS). 3. La sélection favorise les plantes qui possèdent des gènes de résistance dont les produits reconnaissent les effecteurs du pathogène (Avr-R), déclenchant la deuxième barrière de l'immunité (ETI), version amplifiée d'une PTI qui franchit le seuil d'induction de la HR. 4. La sélection favorise les individus pathogènes virulents ayant perdu l'effecteur rouge ou présentant un nouvel effecteur (rond bleu), permettant de supprimer l'ETI. 5. La sélection favorise l'apparition de nouveaux allèles de résistance chez la plante menant à de nouvelles protéines Avr qui reconnaissent les nouveaux effecteurs du pathogène.

Un premier niveau de résistance basale, la résistance non-hôte, est d'abord mis en place par la plante. Elle est déclenchée par la reconnaissance de motifs moléculaires caractéristiques et conservés entre les différents agents pathogènes (PAMPs, pour Pathogen-Associated Molecular Patterns). Ces éléments, aussi nommés éliciteurs de réactions de défense, peuvent être de nature chimique variée : peptides, protéines, composés lipidiques ou encore polymères de sucres (Chisholm et al. 2006; Ingle et al. 2006). La perception de ces motifs par la plante, grâce à des récepteurs membranaires appelés Pattern Recognition Receptors (PRRs), peut initier une réponse immunitaire de base, la PAMP-Triggered Immunity (PTI), qui consiste en une cascade de signaux aboutissant à la mise en place des voies de défense avec production d'espèces réactives de l'oxygène (ROS, pour Reactive Oxygene Species), dépôt de callose au niveau de la zone infectée ou encore production de protéines de type Pathogenesis Related (PR) aux propriétés antimicrobiennes (Jones & Dangl 2006). Malgré leur rôle dans le déclenchement de la PTI, les PAMPs ont été conservés lors de l'évolution des agents pathogènes, ce qui suggère un rôle important dans le parasitisme de la plante (Chisholm et al. 2006). Cette première composante du système immunitaire végétal est peu spécifique, elle peut en effet être efficace contre l'ensemble des membres d'une espèce, d'un genre ou d'une famille d'agents pathogènes, mais elle présente également une résistance modérée, certains pathogènes ayant appris à masquer leur présence en interférant avec les voies de défense de la plante grâce à la sécrétion d'effecteurs, la rendant sensible (Effector*Triggered Susceptiblity*, ETS). Dans ce dernier cas de figure, la plante peut faire intervenir un deuxième type de résistance, la résistance hôte. Celle-ci peut être de deux types, qualitative ou quantitative, avec des déterminismes différents.

La résistance qualitative est une résistance mono- ou oligogénique, induite par un ou quelques gènes de la plante (ou QTL, pour Quantitative Trait Loci, lorsque le gène n'a pas été identifié), dits gènes ou QTL majeurs de résistance, qui codent des protéines R impliquées dans la reconnaissance spécifique des effecteurs sécrétés par le pathogène. Cette interaction spécifique déclenche la deuxième composante du système immunitaire de la plante : l' « Effector-Triggered Immunity » (ETI), qui induit une réponse d'hypersensibilité (HR), se traduisant par une mort programmée des cellules au niveau des tissus infectés pour isoler et inhiber le développement de l'agent pathogène (Jones & Dangl 2006). Cette reconnaissance spécifique des effecteurs par des gènes majeurs de résistance a été mise en évidence pour la première fois par Flor en 1971, suite à ses travaux sur l'interaction entre le lin (Linum usitatissimum) et la rouille du lin (Melampsora lini) (Flor 1971), et formalisée dans le modèle « gène-pour-gène ». Ce modèle prédit une réponse de la plante de type « tout ou rien » : lorsqu'une protéine codée par un gène de résistance R d'une plante interagit avec la protéine codée par le gène d'avirulence Avr correspondant de l'agent pathogène, l'interaction R-Avr est « incompatible » et conduit à une résistance totale (R et Avr sont des traits dominants). En l'absence du gène R et/ou du gène Avr correspondant, l'interaction sera dite « compatible » et mènera à la sensibilité de la plante (Fig. 1.8). La plupart des protéines R identifiées à ce jour appartiennent à la famille des NBS-LRR (Nucleotide Binding Site - Leucine Rich Repeat) (McHale et al. 2006). Les molécules d'avirulence sont généralement impliquées dans une fonction propre de l'agent pathogène (Vera Cruz et al. 2000). Ce modèle sous-entend une interaction directe de type récepteur-ligand entre la protéine R et la protéine Avr, cependant il est à noter que des interactions plus complexes et indirectes peuvent aussi exister. Celles-ci impliquent un cofacteur de la plante qui se positionne comme « intermédiaire » dans l'interaction : l'effecteur agirait sur ce cofacteur, entraînant une modification structurale qui serait ensuite détectée par la protéine de résistance (Jones & Dangl 2006; van der Hoorn & Kamoun 2008). La résistance qualitative a été très étudiée et est la plus utilisée aujourd'hui en lutte variétale. En effet, la réaction d'hypersensibilité induite ne provoque que de faibles dégâts sur les plantes et la forte pression de sélection imposée aux pathogènes rend cette résistance très efficace. Cependant, elle est aussi peu durable, à cause du faible nombre de gènes de résistance impliqués dans l'interaction, et conduit à des contournements, comme celui observé dans les années 1950-1960 aux États-Unis, où chacune des variétés résistantes de blé utilisées contre la rouille a été contournée (Browning & Frey 1969).

La résistance quantitative est en revanche induite par l'action combinée de plusieurs gènes à effets partiels et cumulatifs, appelés gènes ou QTL mineurs de résistance. On parle donc de résistance polygénique (Schiff *et al.* 2001). Elle se traduit par une gamme continue de niveaux de résistance chez la plante, qui fonctionne vis-à-vis de l'ensemble des génotypes de l'agent pathogène. Il n'y a pas d'interaction spécifique entre les génotypes de l'hôte et de l'agent pathogène. Ainsi, cette résistance ne protège pas complétement la plante de l'agent pathogène car elle ne permet pas de bloquer entièrement le cycle de développement du parasite, mais ralentit le développement de l'épidémie dans la culture. On parle donc également de résistance partielle (Vanderplank 1968; Parlevliet 1979). Les mécanismes moléculaires sous-jacents à ces réponses immunitaires restent cependant peu connus. On admet généralement que ce type de résistance est plus durable, de par son déterminisme polygénique, qui serait plus long à contourner, et aux pressions de sélection plus faibles

exercées sur les populations de pathogènes (Parlevliet & Vanommeren 1988; Palloix *et al.* 2009; Delourme *et al.* 2014). Des études ont cependant déjà mis en évidence de tels contournements (Montarry *et al.* 2012; Caffier *et al.* 2014).



Figure 1.8 Schéma des différents scénarios possibles du modèle gène-pour-gène (Flor 1971). L'interaction incompatible, ou résistance de la plante, survient quand le produit de gène de résistance de la plante (R) reconnaît spécifiquement le produit du gène d'avirulence du pathogène (Avr). En l'absence du gène R et/ou du gène Avr correspondant, l'interaction est compatible et mène à la sensibilité de la plante.

1.2.3 Contournement des résistances

La lutte variétale s'appuie sur l'introduction de résistances dans des variétés cultivées par croisements successifs avec des espèces apparentées issues du milieu sauvage. Cependant, plusieurs observations ont déjà démontré l'émergence de nouvelles populations de pathogènes virulentes aux résistances déployées, causant en retour leur perte d'efficacité (McDonald & Linde 2002). Les délais de contournement sont variables d'un pathogène à l'autre et peuvent survenir à l'issue de quelques mois ou de plusieurs décennies (*e.g.* Johnson 1981; Parlevliet 2002; Kang *et al.* 2005).

Les étapes du contournement

On distingue trois étapes lors du contournement des résistances (Gómez *et al.* 2009). La première étape est l'apparition d'individus virulents dans les populations avirulentes. Elle suppose l'apparition d'une variation génétique au niveau du gène d'avirulence concerné, conférant une perte de reconnaissance par la plante hôte et générant donc une interaction compatible. Le mécanisme évolutif principalement en jeu est évidemment la mutation. Outre la taille du génome des organismes, la taille de la population est un facteur central dans la probabilité d'apparition de nouveaux mutants, celle-ci étant plus importante dans de grandes populations, ce qui est souvent le cas chez les agents pathogènes (McDonald & Linde 2002). La recombinaison peut aussi permettre le brassage génétique et mener à la création de combinaisons d'allèles pouvant aboutir à la virulence. C'est par exemple le cas dans

l'interaction melon-MNSV (*Melon Necrotic Spot Virus*), où l'acquisition de la virulence n'est pas le fait de substitutions nucléotidiques mais d'un (ou plusieurs) évènement de recombinaison (Diaz *et al.* 2004). Des individus virulents peuvent néanmoins préexister dans la population au moment du déploiement d'une nouvelle variété résistante, son contournement ne nécessitant alors que le franchissement des deux étapes suivantes.

La deuxième étape est la multiplication des individus virulents dans la population, alors en compétition avec les individus avirulents. La taille de la population a également une importance majeure ici car elle influence la dérive génétique, pouvant mener à la perte aléatoire d'un allèle de virulence dans de petites populations. Des pratiques agricoles comme l'utilisation de pesticides peuvent être considérées comme des événements qui induisent une réduction drastique de la taille des populations. La sélection est la force évolutive la plus importante concernant le contournement des résistances et la principale mise en jeu dans cette seconde étape. La sélection directionnelle exercée par les nouvelles résistances sur les populations d'agents pathogènes est en effet nettement supérieure à toutes les autres forces évolutives. En éliminant les individus avirulents, elles favorisent le développement d'individus virulents dans la population.

L'étape finale du contournement des résistances est la dispersion des individus virulents à l'échelle de la parcelle puis, au-delà, à l'ensemble des plantes hôtes de l'agroécosystème. La principale force à l'œuvre ici est la migration, qui permet les flux de gènes dans la métapopulation. Une fois encore, la taille des populations joue un rôle majeur, puisque l'ampleur du phénomène dépend du nombre de migrants et donc de la taille de la population source - plus une population sera grande, plus la probabilité de migration sera importante - ainsi que de la taille de la population. Les échanges à grande échelle (intercontinentaux) ou à plus petite (pratiques agricoles) provoqués par les activités humaines jouent ici un rôle majeur en accentuant considérablement les flux de gènes (McDonald & Linde 2002).

• Contournement et érosion des résistances

Les processus de contournement des résistances ne sont pas les mêmes s'ils concernent les résistances dues à des gènes majeurs (résistances qualitatives) ou à des gènes mineurs (résistances quantitatives).

Dans le cas de résistances qualitatives, le contournement des résistances suit généralement un cycle dit de « *boom and bust* » (Figs 1.9A et 1.9B). Dans un premier temps, une nouvelle résistance est déployée à large échelle dans l'agrosystème, c'est le *boom*. La résistance provoque une pression de sélection très forte sur la population de pathogènes, qui conduit dans un premier temps à des réductions drastiques des effectifs des populations. S'en suit une augmentation de la fréquence des individus virulents, qui aboutit à l'émergence d'une nouvelle population virulente capable de contourner la résistance devenue alors inutile, c'est le *bust*. Lorsqu'une autre résistance majeure est déployée, le cycle recommence. Ce type de contournement a été mis en évidence à de nombreuses reprises (McDonald & Linde 2002; Rouxel *et al.* 2003). Les populations de pathogènes pouvant conserver longtemps les allèles de virulence adaptés aux gènes R utilisés dans des cultures précédentes, le déploiement ultérieur de nouvelles variétés contenant un ou plusieurs de ces gènes R peut s'en trouver compliqué (Mundt 2014).

Dans le cas de résistances quantitatives, l'étude des contournements s'avère plus difficile à réaliser. La réponse des populations pathogènes est en effet beaucoup plus variable d'une année à l'autre en fonction des conditions environnementales, et le phénomène est également plus lent que lors d'un contournement de résistance majeur. Il a cependant été montré que le contournement de ces résistances était lié à des décalages de phénotypes (Figs 1.9C et 1.9D) souvent associés à une augmentation progressive de l'agressivité des pathogènes. On parle alors plutôt ici d'érosion que de contournement de résistance.



Figure 1.9 Changement dans les populations de pathogènes lors du contournement d'une résistance totale (A et B) et lors de l'érosion d'une résistance partielle (C et D) (d'après McDonald & Linde 2002). Lors du contournement d'un gène majeur de résistance (A et B), le passage de l'avirulence à la virulence est rapide et les différences de pathogénicité entre individus virulents et avirulents sont claires, le contournement est donc facile à détecter. p et q représentent les fréquences des allèles d'avirulence et de virulence, respectivement, dans la population pathogène. Lors de l'érosion d'une résistance quantitative, l'augmentation de la pathogénicité se fait de manière graduelle et les individus présentent une large gamme de pathogénicité, l'érosion est donc plus difficile à détecter.

Coût de l'adaptation aux résistances

Les gènes d'avirulence sont avant tout impliqués dans diverses fonctions chez l'agent pathogène (Vera Cruz *et al.* 2000). Une mutation au niveau de ce gène peut certes rendre l'individu virulent et ainsi lui permettre de contourner la résistance de la plante hôte, mais elle peut se produire aux dépens de la perte ou de la modification de la fonction assurée initialement par ce gène : c'est le coût de la virulence. Celui-ci peut être plus au moins important selon la fonction première du gène d'avirulence, en particulier si celui-ci était

25

impliqué dans la fitness de l'individu (Leach et al. 2001). Bronson & Ellingboe (1986) ont par exemple montré lors d'une expérience de compétition sur hôte sensible qu'une race virulente d'Erysiphe graminis, un champignon du blé, était approximativement 24% moins vigoureuse qu'une race avirulente. D'autres exemples concernent des oomycètes (e.g. Montarry et al. 2010), des virus (e.g. Janzac et al. 2010), des bactéries (e.g. Leach et al. 2001) ou des nématodes (e.g. Castagnone-Sereno et al. 2007). Confrontés à une plante résistante, les individus virulents resteront avantagés face aux individus avirulents, puisqu'ils seront les seuls à voir leur cycle de développement aboutir, si tant est que la diminution de fitness est nulle ou faible. En revanche, si la résistance n'est pas déployée et si le coût est fort, les individus virulents seront contre-sélectionnés et disparaitront de la population (Vanderplank 1968). C'est sur ce principe que se base une nouvelle stratégie de sélection de résistances, explorée notamment par plusieurs équipes françaises de recherche s'intéressant à Venturia inaequalis, responsable la tavelure du pommier (INRA Angers) et à Phytophthora infestans et P. capsici, responsables respectivement du mildiou de la tomate et du piment (INRA Avignon). Plusieurs résistances majeures ciblant ces agents pathogènes ont été déployées, mais ont systématiquement été contournées. La stratégie étudiée repose sur une première étape d'identification chez le pathogène d'une protéine d'avirulence indispensable à sa survie et pour laquelle une mutation lui serait fatale, puis sur une deuxième étape d'identification chez l'hôte du gène de résistance correspondant, qui devrait alors assurer une résistance durable. Ce mécanisme de coût de virulence n'est pas pour autant généralisable, il n'existe pas pour tous les gènes d'avirulence (Leach et al. 2001).

Les coûts de *fitness* associés au contournement des résistances qualitatives ont été beaucoup plus étudiés que ceux associés à l'érosion des résistances quantitatives. Cependant l'adaptation à une résistance quantitative peut également avoir comme conséquence un coût de *fitness* (exemple de l'interaction entre le PVY et le piment, Montarry *et al.* 2012).

1.2.4 Gestion durable des résistances

Une résistance durable se définit comme une résistance qui conserve son efficacité après avoir été utilisée de manière prolongée sur de grandes surfaces et en présence de l'agent pathogène dans un environnement favorable au développement de la maladie (Johnson 1984). La conception de stratégies efficaces de gestion des résistances variétales s'avère indispensable afin d'en optimiser la durabilité. Ceci suppose à la fois d'agir sur les stratégies de déploiement des résistances, mais également d'intégrer des connaissances sur le potentiel évolutif des agents pathogènes (McDonald & Linde 2002).

Du côté de la plante : utiliser des stratégies de déploiement raisonnées

Diverses stratégies de déploiement des résistances peuvent être mises en place afin de les rendre plus durables :

• La diversification dans le temps : Elle peut s'effectuer de plusieurs manières :

o Introgresser différents gènes majeurs de résistance dans différentes variétés et remplacer une variété par une autre à chaque contournement. Cependant, le délai de

contournement d'une résistance, de l'ordre de quelques années, est beaucoup plus rapide que le délai de création de nouvelles variétés résistantes, de l'ordre de la dizaine voire la vingtaine d'années (Rousselle-Bourgeois & Rousselle 1996).

• Il est aussi possible d'alterner d'une saison à l'autre des variétés portant des gènes de résistance différents, permettant ainsi de freiner considérablement l'adaptation des pathogènes (Turner & Fleming 2002).

• Une autre solution est d'alterner l'utilisation de variétés résistantes et sensibles. En effet, nous avons vu que la multiplication des individus virulents dans la population est permise par la pression de sélection imposée par la résistance sur les individus avirulents. L'utilisation alternative de plantes sensibles permettrait ainsi de diminuer cette pression de sélection et limiterait l'augmentation en fréquence des individus virulents. Cette stratégie s'avérerait d'autant plus efficace si la virulence est associée à un coût de *fitness*, aboutissant alors à une contre-sélection des virulents sur les plantes sensibles.

• La diversification dans l'espace : De la même façon, il est possible de déployer des mélanges de variétés sensibles et résistantes, à l'échelle de la parcelle – par rangs alternés ou par mélanges au hasard - ou de l'écosystème. Ceci permet de réduire la dispersion et donc de ralentir l'épidémie (Pink 2002; Brown 2015). L'efficacité de cette technique a déjà été démontrée lors de la réduction de la propagation épidémique de maladies fongiques, en particulier dans le cas des agents de rouilles et d'oïdium des cultures céréalières (Wolfe 1985; Smithson & Lenné 1996). Cependant de tels mélanges variétaux peuvent aboutir à des difficultés techniques lors de la récolte ou de la transformation. Une deuxième solution est alors l'utilisation de multilignées, mélanges de semences de plusieurs lignées qui ne diffèrent que par leurs gènes de résistance (Jensen 1952).

• Le pyramidage de gène : Cette technique consiste à intégrer plusieurs gènes majeurs de résistance dans une même plante. Ceci nécessiterait un plus grand nombre de mutations chez le pathogène, ce qui, statistiquement, aurait peu de chances de survenir (McDonald & Linde 2002). Cette méthode présente toutefois le risque de voir apparaître un individu multivirulent qui rendrait alors tous les gènes de résistance inexploitables (Brown 2015).

• L'association avec d'autres méthodes de lutte : Cette technique est notamment mise en œuvre dans le cas de l'utilisation de résistances quantitatives partielles. Les exemples les plus courants montrent leur association avec des traitements pesticides (Ellissèche *et al.* 1999), qui peuvent également parfois être associés à des résistances majeures (Pink 2002).

• La combinaison des deux types de résistance : Cette technique consiste à introduire dans des variétés présentant un niveau de résistance partielle satisfaisant des gènes majeurs de résistance. Des associations entre QTL et gènes majeurs ont déjà été mises en évidence dans des pathosystèmes sauvages (Ericson & Burdon 2009), et ce type de combinaison est aujourd'hui exploité pour ralentir le contournement des résistances dans des variétés cultivées. Effectivement, la plus grande durabilité de gènes majeurs de résistance lorsqu'ils sont associés à des QTL de résistance partielle a été démontrée pour le virus Y de la pomme de terre (Palloix *et al.* 2009), pour le champignon *Leptospheria maculans* (Brun *et al.* 2010) et

pour deux espèces de nématodes, *Globodera pallida* (Fournet *et al.* 2013) et *Meloidogyne incognita* (Barbary *et al.* 2014).

► Du côté du pathogène : une meilleure connaissance du potentiel évolutif et des déterminants génétiques de la virulence

Nous avons vu précédemment que les différentes forces évolutives intervenaient dans les trois étapes du contournement, et que leur niveau d'action était conditionné par des caractéristiques propres au pathogène, comme la taille de ses populations par exemple. Ces caractéristiques incluent également le spectre d'hôte, les capacités de dispersion active ou passive, le mode de reproduction ou encore l'histoire de l'espèce au cours des derniers millénaires, et définissent ainsi son potentiel évolutif (McDonald & Linde 2002). L'histoire évolutive, qui résulte de l'action combinée des différentes forces évolutives au cours du temps, déterminera notamment la composition génétique de la population au moment du déploiement de la résistance, et ainsi la préexistence éventuelle dans la population d'individus virulents. La composition initiale et le potentiel évolutif de la population du pathogène sont donc des déterminants de la durabilité d'une résistance. Des études sur la biologie et l'écologie de l'espèce et la génétique de ses populations peuvent permettre de décrire les caractéristiques de l'espèce citées ci-dessus, et ainsi d'appréhender son potentiel évolutif. Connaître la composition génétique de la population revient à quantifier la présence d'individus virulents et avirulents dans la population. Pour cela, une étape préliminaire nécessaire est l'identification des bases génétiques et moléculaires de la virulence.

1.3 Modèle d'étude : le pathosystème Globodera pallida / Pomme de terre

Dans cette troisième partie, nous nous attacherons à présenter chacun des membres du pathosystème qui nous a servi de modèle pour étudier par une approche de scan génomique le déterminisme moléculaire de l'adaptation d'un pathogène à sa plante hôte. Nous mettrons ensuite en avant les interactions qui interviennent au cœur du pathosystème. Dans un dernier temps, nous présenterons en détail l'évolution expérimentale et les résultats phénotypiques auxquels elle a permis d'aboutir, qui ont mené à l'émergence de ce projet de thèse et fourni le matériel biologique nécessaire à l'étude.

1.3.1 Le nématode à kyste Globodera pallida

Le phylum Nemata : Généralités

Morphologie et anatomie

Les nématodes (du grec nêmatos, « fil », et eidos, « apparence ») sont des organismes vermiformes ronds, fins et non segmentés, à symétrie bilatérale (Fig. 1.10). Transparents pour la plupart, ils présentent des tailles très variables, pouvant aller de quelques centaines de micromètres à plusieurs mètres (Blumenthal & Davis 2004), bien que la grande majorité des espèces ne dépasse pas 1 à 2 mm. Le corps des nématodes est recouvert d'une cuticule de collagène continue et souple, mais très résistante et percée de pores permettant les transferts

gazeux. La cuticule est sécrétée par l'épiderme syncitial, sous lequel se trouve une épaisse couche de muscles, exclusivement longitudinaux (il n'existe pas de muscles circulaires transversaux), qui entourent le tube digestif – complet et rectiligne, s'ouvrant par la bouche et se terminant par l'anus – et les glandes génitales (Fig. 1.11). Ces muscles n'autorisent ainsi aux nématodes que des mouvements ondulatoires. Les nématodes sont des animaux triploblastiques (trois feuillets embryonnaires) dont la cavité générale est un pseudocœlome. La pression du liquide dans cette cavité joue le rôle de squelette hydrostatique. Les nématodes disposent d'un système nerveux composé d'un anneau entourant le pharynx et de deux nerfs longitudinaux (ventral et dorsal), mais ne présentent pas de système respiratoire ni circulatoire. Les nématodes possèdent des organes sensoriels, les amphides, situées à l'extrémité antérieure, et les phasmides, à l'extrémité postérieure. Les sexes sont en général séparés. Le mâle est souvent plus petit que la femelle et possède des spicules copulateurs, permettant une fécondation interne. L'amphimixie est la stratégie de reproduction la plus répandue, mais d'autres stratégies comme la parthénogénèse et l'hermaphrodisme existent. La capacité de reproduction est importante chez beaucoup d'espèces.



Figure 1.10 Structure interne d'un nématode mâle et d'un nématode femelle.



Figure 1.11 Coupe transversale de nématode. Le pseudocœlome n'est pas tapissé d'épithélium : les organes baignent directement dans la cavité générale.

Les nématodes ont longtemps été considérés comme un groupe « primitif » du fait de leur morphologie très simple. Ils étaient auparavant classés dans le groupe des pseudocœlomates, c'est-à-dire les organismes dont la cavité interne n'est pas tapissée d'épithélium mésodermique (Fig. 1.12A). Ce groupe était lui-même considéré comme plus « primitif » que les véritables cœlomates (*e.g.* arthropodes, chordés...), cette distinction s'avérant en réalité artificielle, le cœlome pouvant évoluer d'un aspect « cœlome vrai » vers un aspect « pseudocœlome » (le terme pseudocœlomate n'a aujourd'hui plus de validité en classification phylogénétique). Sur la base de la phylogénie moléculaire de la séquence 18S de l'ARN ribosomal, les nématodes sont aujourd'hui regroupés avec les arthropodes (dont le plan d'organisation est complexe) au sein des Ecdysozoaires, c'est à dire l'ensemble des Protostomiens possédant un exosquelette et effectuant donc leur croissance par mues (Aguinaldo *et al.* 1997; Adoutte *et al.* 1999) (Fig. 1.12B). Les nématodes passent en effet par quatre mues successives avant d'atteindre la forme adulte, au cours d'un développement direct. Le plan d'organisation serait donc non pas primitivement simple mais simplifié à partir d'un plan plus complexe.



Figure 1.12 Arbre phylogénétique des Métazoaires (Adoutte *et al.* 1999). **A**. Phylogénie « traditionnelle ». **B**. Phylogénie basée sur les séquences 18S de l'ARN ribosomal.

Diversité spécifique et écologique

Le phylum Nemata se caractérise par une grande homogénéité des organismes qui le composent, sur les plans morphologique et anatomique, qui contraste avec leur importante diversité spécifique et écologique. On estime entre 40 000 et 10 millions le nombre total d'espèces dans le phylum Nemata (Blaxter *et al.* 1998; Dorris *et al.* 1999; Blumenthal & Davis 2004), ce qui place ce clade au deuxième rang dans le règne animal, derrière les insectes. Plus de 26 000 espèces de nématodes ont été décrites à ce jour (Hugot *et al.* 2001) et réparties en deux clades et vingt ordres. En termes d'individus, ils sont les organismes du sol les plus abondants parmi les métazoaires, avec une densité d'environ 1 million par m² (Bernard 1992). Leur rôle dans le maintien de l'équilibre naturel des sols est primordial : ils participent à la décomposition de la matière organique, à l'aération des sols et à la circulation des matières minérales et organiques. Cependant, de par leur petite taille, ils ne représentent souvent qu'une faible proportion de la biomasse animale totale des sols, soit moins de 10% (Fierer *et al.* 2009).

Cette diversité d'espèces occupe des niches écologiques variées sur la planète. Restant inféodés aux milieux humides (leur cuticule percée de pores ne leur permettant pas de retenir l'eau), les nématodes exploitent néanmoins une diversité de milieux tels que les océans, les eaux douces, les fluides corporels, les films d'eau dans le sol ou sur les végétaux. Ils se répartissent sur toute la surface du globe, sous tous les climats, à toutes les profondeurs et altitudes. En cas de conditions environnementales défavorables, ils peuvent entrer en état de vie ralentie ou produire des formes de résistance qui peuvent persister plusieurs années. Les nématodes présentent des formes de vie libre (terrestre ou marine) ou parasitaire (d'animaux ou de végétaux). Parmi les espèces décrites, 40% mènent un mode de vie libre, alors que 45% sont zooparasites et 15% sont phytoparasites. Une diversité de régimes alimentaires est également présente au sein de ce clade. On peut en effet observer des espèces bactériophages (e.g. Caenorhabditis elegans), entomopathogènes (e.g. Steinernema spp., Heterorhabditis spp.), zooparasites spécialisées dans divers organes (e.g. Ascaris spp., Brugia spp.), prédatrices (e.g. Mononchus spp.), phytoparasites des tissus foliaires (e.g. Aphelenchoides spp.), des tiges et des bulbes (e.g. Ditylenchus spp.), des graines (e.g. Anguina spp.) ou des racines (e.g. Globodera spp.). Les relations hôte-parasite et les cycles biologiques sont extrêmement diversifiés. On trouve par exemple des juvéniles zooparasites et des adultes libres, des juvéniles phytoparasites et des adultes zooparasites, ou encore des zooparasites à plusieurs hôtes successifs. Une majorité des espèces animales et végétales a généralement une voire plusieurs espèces de nématodes parasites qui leurs sont associées (Gilabert & Wasmuth 2013). Certaines espèces parasitent l'Homme, causant des pathologies parfois très graves, comme l'onchocércose causée par Onchocerca volvulus, ou encore la filariose lymphatique causée notamment par Wuchereria bancrofti.

Histoire évolutive

Les nématodes présentant un corps mou, très peu ont pu être fossilisés, ce qui limite la fiabilité des estimations temporelles et du calibrage d'une horloge moléculaire dans le phylum. On estimerait leur apparition au début du Cambrien, il y a 550 millions d'années (Ayala *et al.* 1998), dans le milieu marin. Les premiers nématodes étaient probablement des organismes libres qui se nourrissaient dans le compartiment microbien, et ils auraient dans un second temps exploité de nouvelles ressources, apparues plus tardivement, comme des hôtes animaux et végétaux (Wang *et al.* 1999; Sanderson 2003). L'apparition du parasitisme reste néanmoins difficile à dater. Des traces de nématodes zooparasites ont été identifiées dans les fèces fossilisées de dinosaures datant de 240 millions d'années (Poinar & Boucot 2006; Hugot *et al.* 2014) et des spécimens de nématodes entomopathogènes ont été découverts piégés dans de l'ambre datant d'au moins 140 millions d'années (Poinar & Poinar 2003). Ce type de fossile n'a cependant pas encore été découvert chez les nématodes phytoparasites.

La majeure partie des connaissances concernant l'évolution des nématodes provient d'études portant sur l'anatomie comparée de nématodes actuels, leurs régimes trophiques et plus récemment la comparaison de séquences ADN (Powers & Harris 1993; Thomas *et al.* 1997). Sur la base des séquences de la petite sous-unité de l'ARN ribosomal, le phylum Nemata a été divisé en cinq clades majeurs (Blaxter *et al.* 1998) (**Fig. 1.13**), contenant tous des espèces aux modes de vies divers. Il apparaît également que la capacité des nématodes à parasiter des animaux et des plantes est apparue plusieurs fois au cours de l'évolution.



Figure 1.13 Arbre phylogénétique du phylum Nemata basé sur les séquences de la petite sous-unité de l'ADN ribosomal (d'après Blaxter *et al.* 1998). L'arbre représente les cinq clades majeurs (numérotés de l à V) ainsi que le groupe paraphylétique Chromadorida (C), dont la relation avec les Secernentea (clades III, IV et V) reste incertaine. L'écologie trophique (indiquée par un symbole) ainsi que des exemples d'espèces sont renseignés pour chaque groupe taxonomique. Photos de haut en bas : *Meloidogyne incognita, Steinernema masoodi, Phasmarhabditis hermaphrodita* et *Dracunculus medinensis*.

• Les nématodes phytoparasites

Le parasitisme des végétaux serait apparu à au moins trois reprises au cours de l'évolution, dans différents ordres : les Dorylaimida, les Triplonchida, les Aphelenchida et les Tylenchida (Blaxter *et al.* 1998) (Fig. 1.13). Bien qu'ayant divergé les uns des autres très précocement au cours de l'évolution du phylum, les nématodes phytoparasites possèdent un caractère dérivé commun, le stylet (Fig. 1.14). Il s'agit d'une aiguille creuse et protractile située dans la partie antérieure des nématodes, leur permettant à la fois de perforer la paroi des cellules végétales de leur hôte, d'injecter des substances, notamment pour induire le développement de structures spécifiques appelées sites nourriciers, et enfin d'aspirer le contenu de ces cellules afin de se nourrir des composés cellulaires des plantes hôtes. Le stylet est ainsi indispensable pour l'établissement de la vie parasitaire. L'œsophage joue également un rôle important dans ce processus. Celui-ci comprend une partie musculaire et une partie glandulaire. La partie musculaire, véritable pompe aspirante et foulante, injecte le produit des glandes salivaires dans les cellules, à travers le stylet (Fig. 1.14).

Du point de vue de leur comportement vis-à-vis des plantes, les nématodes phytoparasites peuvent être séparés en deux catégories d'importance inégale : d'une part, les parasites des parties aériennes, qui s'attaquent aux bulbes, aux tiges et aux feuilles. D'autre part, les nématodes telluriques, qui effectuent tout leur cycle dans le sol et ne s'attaquent qu'aux racines. Ce sont les plus nombreux. Parmi cette catégorie de nématodes, on distingue les ectoparasites, qui ne pénètrent jamais les racines, mais se contentent de les piquer plus ou moins profondément en fonction de la taille de leur stylet, et les endoparasites, qui pénètrent partiellement ou entièrement dans la racine. Dans ce dernier groupe, on différencie les endoparasites migrateurs, qui creusent des cavités à travers les cellules du cortex racinaires, et les endoparasites sédentaires, qui sélectionnent un site nourricier et y restent fixés tout au long de leur développement.

Aujourd'hui, plus de 4 100 espèces de nématodes phytoparasites ont été décrites (Decraemer & Hunt 2006). Les dégâts causés par ces nématodes sont estimés à 80 milliards de dollars par an (Nicol et al. 2011). Les Dorylaimida et les Triplonchida sont plus problématiques par leur rôle de vecteur de virus que pour les dégâts directs qu'ils provoquent sur les plantes. Les Tylenchida constituent l'ordre le plus important des nématodes phytoparasites, tant en termes de nombre d'espèces qu'en termes de dégâts causés aux plantes qu'ils parasitent. Cet ordre regroupe neuf familles de nématodes dont les Heteroderidae (Ferraz & Brown 2002), responsables des dégâts les plus importants sur les plantes. Cette famille comporte une vingtaine de genres dont les nématodes à galles du genre Meloidogyne et les nématodes à kyste des genres Heterodera et Globodera (Fig. 1.15). Les nématodes à galles et les nématodes à kyste ont été respectivement classés numéros un et deux dans le top 10 des nématodes phytoparasites, de par leur importance scientifique et économique (Jones et al. 2013). Ces nématodes ont des caractéristiques biologiques très proches : ce sont des endoparasites sédentaires biotrophes qui induisent la mise en place d'un site nourricier complexe et qui produisent des femelles renflées. Bien que longtemps regroupés dans la classification sur la base de leurs caractères morphologiques, de récentes études de phylogénie moléculaire ont démontré que les nématodes à kyste et les nématodes à galles ne partageaient pas d'ancêtre commun proche (Holterman et al. 2009). Leurs caractéristiques communes ne seraient donc que des convergences évolutives apparues indépendamment au cours de l'évolution.



Figure 1.14 Structure de la partie antérieure des nématodes phytoparasites. A. Schéma de la structure (d'après Hussey 1989). **B.** Photo de la partie antérieure de *Hemicycliophora sp.*, faisant apparaître le stylet (photo : Sylvain Fournet).



Figure 1.15 Synthèse schématique des rangs taxonomiques conduisant au genre *Globodera* et relations phylogénétiques entre certaines espèces du genre *Globodera* (d'après Subbotin *et al.* 2006). Les aires d'origines et les familles des plantes hôtes des espèces sont indiquées.

Le nématode à kyste de la pomme de terre, *Globodera pallida*

Le nématode à kyste sur lequel est centrée cette étude est le nématode à kyste de la pomme de terre *Globodera pallida*. Il a été décrit pour la première fois par Stone en 1973, qui le nomma *Heterodera pallida* (Stone 1973). Cette espèce était auparavant regroupée avec *G. rostochiensis*, qui attaque également la pomme de terre, sous le nom « *Heterodera rostochiensis sensu lato* ». Skarbilovich (1959) a par la suite décrit le sous-genre « *Globodera* » regroupant les nématodes à kyste rond, ce sous-genre ayant ensuite été élevé au rang de genre (Behrens 1975; Mulvey & Stone 1976). La couleur des femelles immatures permet de différencier *G. pallida* et *G. rostochiensis* : à maturité, les deux espèces deviennent marron, mais *G. rostochiensis* passe par un stade doré alors que *G. pallida* passe par un stade blanc, d'où l'autre nom de « nématode blanc de la pomme de terre » donné à *G. pallida* (« nématode doré de la pomme de terre » pour *G. rostochiensis*). D'autres différences reposent sur la taille du stylet et la position des boutons basaux.

Dans le genre *Globodera*, *G. pallida* est génétiquement plus proche de *G. mexicana* (Fig. 1.15), mais les deux espèces se distinguent par leur gamme d'hôte : *G. pallida* se développe sur pomme de terre tandis que *G. mexicana* se développe sur morelle noire, les deux étant également capables de se développer sur tomate. *G. pallida* présente par contre la même gamme d'hôte que *G. rostochiensis*, enfin *G. tabacum* se développe sur morelle noire, tabac et tomate, mais pas sur pomme de terre.

Globodera pallida est un parasite biotrophe obligatoire spécifique des Solanacées. Il parasite en particulier la pomme de terre, ainsi que d'autres espèces de *Solanum* sauvages et leurs hybrides, et également la tomate, l'aubergine et quelques plantes spontanées comme la douce-amère et la jusquiame noire.

Origine, distribution géographique et structure génétique

Le genre Globodera comprend des espèces d'origines géographiques diverses : G. millefolii et G. artemisiae natives de l'est de l'Europe et G. tabacum, G. rostochiensis, G. mexicana et G. pallida originaires des continents nord et sud-américains (Fig. 1.15). La population ancêtre des Globodera sud-américains proviendrait d'une dispersion depuis l'Amérique du Nord à une période (60 à 80 millions d'années) où l'Amérique du Nord et du Sud étaient encore connectées. En se basant sur cette hypothèse, on estime que l'espèce G. pallida a divergé des autres Globodera il y a 15 à 20 millions d'années (Picard et al. 2008). Une étude phylogéographique menée le long de la Cordillère des Andes au Pérou a mis en évidence une structuration des populations de G. pallida de type sud/nord avec cinq clades distincts (Picard et al. 2007) (Fig. 1.16). Cette structuration suggère une évolution des populations de G. pallida depuis le sud Pérou vers le nord Pérou corrélée avec l'orogénèse des Andes : l'évolution de G. pallida a ainsi suivi l'évolution de son hôte principal, la pomme de terre sauvage (voir partie 1.3.2). La colonisation du nord du Pérou par G. pallida est estimée à 3 à 4 millions d'années. Aujourd'hui, G. pallida est présent sur tous les continents, aussi bien en zones tempérées, au niveau de la mer, qu'en zones tropicales ou sud-tropicales, à des altitudes plus élevées (Fig. 1.17).

Globodera pallida aurait été introduit en Europe au milieu du XIXème siècle, probablement par l'import de tubercules de pomme de terre destinés à l'amélioration de la résistance au mildiou. Les individus introduits provenaient de populations de nématodes



Figure 1.16 Phylogénie de populations péruviennes de *G. pallida* basée sur la séquence partielle du cytochrome **b et sur 8 marqueurs microsatellites** (d'après Picard *et al.* 2007). Cette phylogénie met en évidence une structuration de type sud/nord le long de la cordillère des Andes, avec cinq clades distincts (numérotés de 1 à 5).



Figure 1.17 Carte de répartition du nématode à kyste de la pomme de terre *Globodera pallida*, au 01/07/2016, d'après l'Organisation Européenne et Méditerranéenne pour la Protection des Plantes (OEPP).

issues du clade I au sud du Pérou (Plantard *et al.* 2008). Les populations européennes de *G. pallida* ont donc évolué à partir d'une petite fraction des nématodes originaires du Pérou : elles ne représentent qu'une faible partie de la diversité au sein de l'espèce *G. pallida*. En utilisant des marqueurs microsatellites et en comparant les populations de *G. pallida* introduites en Europe et les populations du sud du Pérou, Plantard *et al.* (2008) ont estimé à un tiers la réduction de la diversité allélique en Europe par rapport à la diversité native.

G. rostochiensis a pendant longtemps été le principal nématode ravageur de la pomme de terre en zones tempérées et froides. L'identification d'un gène de résistance actif contre *G. rostochiensis*, et son introgression dans de nombreuses variétés de pommes de terre cultivées, a entraîné une pression de sélection sur le terrain en faveur du remplacement de cette espèce par *G. pallida*.

On distingue au sein de l'espèce *G. pallida*, mais aussi chez *G. rostochiensis*, des « races biologiques », appelées pathotypes, qui regroupent différentes populations présentant les mêmes capacités à se multiplier sur un ensemble de génotypes résistants et sensibles de *Solanum* spp. Deux pathotypes différents se comportent donc différemment face à la gamme d'hôte, du point de vue de leur succès parasitaire (voir **Encadré 1**).

Cycle biologique (Fig. 1.18)

Le cycle biologique de *G. pallida* est un cycle biologique typique de nématode à kyste. Les exsudats radiculaires de la pomme de terre ont la capacité de faire éclore massivement les œufs contenus dans les kystes présents dans le sol. Une humidité suffisante est également un facteur important pour l'éclosion. Celle-ci n'est jamais totale, une partie des œufs restant à l'état de vie latente. Les kystes libèrent des formes juvéniles de stade 2 (J2, Fig. 1.19A), la première mue permettant de passer du stade J1 à J2 ayant lieu directement dans l'œuf. Grâce à leur stylet perforant, les J2 pénètrent par la zone d'élongation racinaire et progressent vers le cylindre central (Fig. 1.19B), détruisant les cellules corticales situées sur leur passage, grâce au stylet et à leurs sécrétions glandulaires. Arrivés à proximité de l'endoderme, ils testent, avec leur stylet, les cellules qui les entourent avant de choisir celle qui servira de point de départ à l'induction d'un site nourricier (Rice et al. 1985). Les critères qui déterminent ces choix ne sont pas encore connus. Le stylet perce la paroi de la cellule choisie et s'enfonce de quelques microns sans abîmer la membrane plasmique. Une sécrétion fluide, provenant des glandes salivaires ou des amphides, est projetée à l'intérieur de la cellule, induisant d'importants changements cellulaires. Les parois latérales s'épaississent alors que la paroi distale s'amincit. Des plasmodesmes se forment entre cette cellule et les cellules adjacentes, situées à l'opposé du stylet du nématode. Celles-ci se modifient également et la lyse des parois situées entre ces cellules entraîne la formation d'une cellule géante multinucléée, à cytoplasme dense, le syncytium. Il représente une source nutritive essentielle pour le nématode qui est dans l'incapacité d'achever son cycle parasitaire s'il n'est pas fonctionnel. L'expansion du syncytium est principalement longitudinale et se poursuit jusqu'au quatrième stade larvaire (Smant et al. 1998). Il peut être constitué de plus de 200 cellules fusionnées (Gheysen & Fenoll 2002). Les nématodes se nourrissent du cytoplasme du syncytium par l'intermédiaire d'un tube alimentaire, ce syncytium restant fonctionnel tout au long du développement du nématode. Des études montrent que les changements cellulaires ayant lieu chez la plante et conduisant à la création du syncytium sont plutôt dus à une manipulation du génome des cellules végétales par le nématode, favorisant la surexpression de certains gènes, via ses sécrétions probablement, plutôt qu'à une action directe de ces sécrétions sur la cellule (Goellner et al. 2001; Vercauteren et al. 2002). D'autres gènes sont au contraire sous-

Encadré 1 Classification des nématodes à kyste en pathotypes.

Kort *et al.* (1977) ont tenté d'établir une classification des populations européennes de *G. pallida* et de *G. rostochiensis*, d'après leurs taux de multiplication sur une gamme de plusieurs génotypes de pommes de terre présentant des résistances contrastées (**Tableau 1.1**). Le taux de multiplication est défini par le ratio entre le nombre final (Pf) et le nombre initial (Pi) de kystes. On détermine le succès à produire une descendance lorsque Pf/Pi > 1.0 alors que l'échec est défini par un ratio Pf/Pi < 1.0.

Tableau 1.1 Classification des pathotypes de Globodera selon Kort et al. (1977) (d'après Nijboer & Parlevliet1990). + : plante sensible (Pf/Pi > 1.0), - : plante résistante (Pf/Pi < 1.0).</td>

Clone	Rol	Ro2	Ro3	Ro4	Ro5	Pal	Pa2	Pas
1. S. tuberosum ssp. tuberosum	+	+	+	+	+	+	+	+
2. S. tuberosum ssp.andigena (H1)	-	+	+	-	+	+	+	+
3. S. kurtzianum, 60.21.19	-	-	+	+ .	+	+	+	+
4. S. vernei, 58.1642.4	-	(m)	-	+	+	+	+	+
5. S. vernei, 65.346.19	-	-	-	-	-	+	+	+
6. S. vernei, 62.33.3	-		-	-	±	_	_	+
7. S. multidissectum, P 55/7(H2)	+	+	+	+	+	-	+	+

Ainsi, on distingue trois pathotypes chez G. pallida (Pa1 à Pa3) et cinq chez G. rostochiensis (Ro1 à Ro5). Toutefois, le concept de pathotype suppose que l'on connaisse les gènes impliqués dans la résistance chez tous les hôtes différentiels, que la résistance soit monogénique, et que tous les individus appartenant à un même pathotype aient tous les mêmes gènes de virulence. Ces conditions ne sont pas toujours remplies ou sont difficiles à vérifier, le schéma proposé par les auteurs reste donc très imprécis. Les seules interactions clairement définies concernent les gènes H1 et H2. Le gène H1 originaire d'un clone de la sous-espèce Solanum tuberosum ssp. andigena permet de différencier les pathotypes Ro1 et Ro4 avirulents, des pathotypes Ro2, Ro3 et Ro5 virulents. De la même façon, les populations de G. pallida qui ne se développent pas sur les plantes possédant le gène H2, originaire d'un clone de S. multidissectum, sont identifiées comme faisant partie du pathotype Pa1 (avirulent). Comme la différenciation entre chacun des autres pathotypes n'est pas nette (les taux de développement représentant en fait un continuum entre Pa2 et Pa3), on parle de plus en plus de groupes de virulence, comme par exemple le groupe de virulence Pa2/3 (virulent sur H2). À ce jour, il n'est pas possible de distinguer les pathotypes sur la base de tests moléculaires.

Les populations concernées par notre étude appartiennent au pathotype (ou groupe de virulence) Pa2/3.

régulés, notamment les gènes impliqués dans les mécanismes de défense des plantes (Gheysen & Fenoll 2002). Le nématode est ainsi capable de manipuler le génome des cellules nourricières de manière à former, mais aussi à maintenir son site nourricier. Les J2 subiront deux mues successives, donnant les juvéniles J3 puis J4, avant de se transformer lors d'une dernière mue en adultes sexués. Le syncytium a un rôle essentiel dans la détermination du sexe, qui est sous influence purement environnementale : la production de mâles est fortement accentuée lorsque le nématode rencontre des conditions adverses (compétition intra- ou interspécifique trop importante, mauvais état physiologique de la plante attaquée, présence de certains gènes de résistance) et que le fonctionnement du syncytium ne permet pas de subvenir aux importants besoins en nutriments du stade femelle (Mugniéry & Fayet 1981, 1984; Grundler et al. 1991). Les femelles grossissent jusqu'à passer à travers la surface de la racine, mais y restent accrochées par la tête. Elles sont fécondées par les mâles mobiles, sortis des racines, qui meurent après la copulation (la reproduction est strictement amphimictique). Les mâles sont attirés par les femelles grâce à des phéromones sexuelles, sécrétées par la cuticule. La copulation déclenche la ponte des ovocytes. La fécondation et l'embryogénèse ont lieu dans les oviductes de la femelle, généralement fécondée par plusieurs mâles (Green et al. 1970; Triantaphyllou & Esbenshade 1990). Les œufs se développent à l'intérieur de la femelle fécondée, provoquant le gonflement de son corps (Fig. **1.19C**). Sa cuticule se transforme en une enveloppe de protection : le kyste. Un kyste contient plusieurs centaines d'œufs, contenant chacun un juvénile. À maturité, la femelle meurt et le kyste se détache de la racine. Il constitue un excellent moyen de dissémination et de conservation. Après la première mue, les J2 entrent en diapause. Les juvéniles peuvent, en zone tempérée, rester viables pendant 10 à 20 ans (Trudgill et al. 2003). Globodera pallida n'effectue normalement qu'une seule génération par an en zone tempérée, car la diapause des J2 n'est levée que par l'effet de températures hivernales. Néanmoins, en France, dans des situations géographiques favorables à la culture précoce de la pomme de terre, G. pallida peut présenter une diapause peu prononcée et un deuxième cycle peut s'initier en automne sur repousses de pomme de terre. Le cycle complet dure de 50 à 70 jours, en fonction de la température du sol. Les œufs éclosent de façon échelonnée dès que la température atteint 12°C. Les premiers kystes apparaissent en petit nombre en juin, puis en très grand nombre en juillet-août. Leur formation s'arrête en octobre. En l'absence de la plante-hôte, la population ne décroit que d'environ 30% par an et, en cas de forte infestation, celle-ci peut mettre 20 à 30 ans à disparaître. Les J2 et les mâles sont vermiformes, d'une longueur d'environ 470 μm et 1200 µm respectivement, tandis que les femelles fécondées se transforment en une poche sphérique d'environ 450 μm.

Moyens de mouvement et dispersion

Globodera pallida n'a pas de moyen actif de dispersion à large échelle, les mouvements se font uniquement sur de courtes distances par les J2 attirées vers les racines de la plante hôte et par les mâles. Cette faible capacité de dispersion provoque un déficit en hétérozygotes dans les populations de *G. pallida* (Picard *et al.* 2004) comme dans les populations d'autres nématodes à kyste, tels *Heterodera schachtii* qui attaque la betterave (Plantard & Porte 2004) ou *G. tabacum* qui attaque le tabac (Alenda *et al.* 2014). Ce déficit en hétérozygotes est directement dû aux faibles capacités de dispersion active des larves aboutissant à une fréquence élevée de croisements entre apparentés (consanguinité) chez les espèces monovoltines comme *G. pallida* et à l'apparition d'une sous-structuration à une échelle



Figure 1.18 Cycle biologique de Globodera pallida (dessin : Pauline Casters-Picard).



Figure 1.19 Nématode à kyste Globodera pallida et dégâts occasionnés. A. Juvénile de stade 2 (photo : University of Nebraska-Lincoln, USA). **B.** Juvénile dans une racine de pomme de terre, coloré à la fuchsine acide (photo : University of Idaho, USA). **C.** Femelles faisant saillie à la surface de racines de pomme de terre (photo : Central Science Laboratory, York, GB). **D.** Champ de pommes de terre infesté de nématodes à kyste de la pomme de terre (photo : Agriculture and Food Science in Northern Ireland).

inférieure à celle de la plante (effet Wahlund) chez les espèces polyvoltines comme *H. schachtii* et *G. tabacum* (Montarry *et al.* 2015).

Globodera pallida est disséminé dans de nouvelles aires sous forme de kystes présents sur différentes sources d'inoculum, telles que (par ordre d'importance) plants de pomme de terre, sol, bulbes de fleurs, pommes de terre destinées à la consommation ou à la transformation. Ces derniers cas ne sont problématiques que s'il existe un risque de plantation ou si aucune précaution n'est prise concernant le traitement des résidus de terre. L'eau et le vent peuvent également propager les kystes sur de longues distances. Picard & Plantard (2006) ont montré qu'une population de G. pallida se définissait à l'échelle d'un bassin de production de pommes de terre, soulignant ainsi l'influence majeure des pratiques agricoles dans la structuration des populations du nématode. Une étude récente montre que les populations de G. tabacum du sud-ouest de la France sont structurées en deux groupes génétiques qui correspondent aux limites de deux coopératives tabacoles, ce qui illustre bien l'impact des activités humaines sur la dispersion des populations de nématodes à kyste (Alenda et al. 2014). Picard et al. (2004) ont en outre montré chez G. pallida que le maximum de variabilité génétique entre individus était retrouvé à l'échelle d'une parcelle et non d'un champ ou d'une région, puis Plantard et al. (2008) ont affiné cette observation en concluant que la variabilité génétique d'une population se retrouvait à l'échelle d'une plante.

Ressources moléculaires

Thiéry et Mugniéry (1996) ont développé un protocole permettant d'identifier l'espèce G. pallida par PCR-RFLP de la séquence ITS (Internal Transcribed Spacers). Un set de 202 marqueurs microsatellites a été développé au sein de notre équipe de recherche et a été utilisé dans le cadre de cette thèse. Le protocole de développement de ces marqueurs est renseigné à la partie **3.4.2**. Le génome de G. pallida, diploïde et réparti sur 2x9 chromosomes, a récemment été séquencé (version Gp.v1.0; Cotton et al. 2014) et est actuellement disponible sous la forme de 6873 scaffolds. La taille de cet assemblage est de 124,7 Mb, tandis que la taille estimée est de 100 Mb. La N50 est de 122 kb et le contenu en GC est de 36,7 %. L'annotation machine de ce génome a abouti à la prédiction de 16 419 gènes, avec un score CEGMA (Core Eukaryotic Genes Mapping Approach) de 81/85 prédictions de gènes identifiées. Les marqueurs microsatellites ont été développés sur une précédente version du génome, supercontigs (version gpal.201011.supercontigs.fasta, composée de 17 095 http://sanger.ac.uk). Enfin, 9 214 séquences ESTs (Expressed Sequence Tags) sont disponibles dans la base de données du NCBI (http://www.ncbi.nlm.nih.gov/nucest/).

Quelques données moléculaires concernant son espèce sœur *G. rostochiensis* peuvent également être exploitées. Une carte génétique a notamment été construite à partir de marqueurs AFLPs (Rouppe van der Voort *et al.* 1999). Elle indique un ratio global distance physique/distance génétique de 120 kb/cM. Le génome de *G. rostochiensis* a également été récemment séquencé, pour lequel une annotation experte est disponible (version nGr.v1.0; Eves-van den Akker *et al.* 2016).

Dégâts, impacts économiques et facteurs influençant la nocivité des nématodes

Globodera pallida cause des dégâts conséquents sur ses hôtes. Le développement et le maintien du syncytium rivalise avec la croissance de la plante, et entrave la circulation de la sève, entraînant une déficience hydrique et un dérèglement du métabolisme nutritionnel. Les racines brunissent, se ramifient et prennent un aspect buissonnant. La croissance de la pomme de terre est ralentie, les feuilles inférieures se fanent et meurent. Les feuilles supérieures se décolorent et présentent des tâches brunes en bordure de folioles. La plante reste chétive et peu productive. Dans les champs, les attaques se manifestent sous forme d'îlots circulaires ou elliptiques, où les plantes apparaissent plus petites que la normale et le feuillage ne se touche pas entre les rangs (Fig. 1.19D). Néanmoins, les lésions causées au système racinaire peuvent aussi ne laisser transparaître aucun symptôme particulier au niveau des parties aériennes en début d'infestation, et il est ainsi difficile de deviner la présence de nématodes sans une analyse précise du sol. De plus, même en cas de forte infestation, les symptômes d'attaques sur les parties aériennes ne sont pas spécifiques. Ils sont par exemple similaires à ceux induits en cas de stress hydrique, de carence minérale ou d'excès de désherbant. La preuve de la présence de *G. pallida* s'obtient donc forcément par une observation directe de kystes dans le sol, ou de femelles ou kystes sur les racines hôtes, visibles à l'œil nu. L'identification spécifique est possible par observation de la couleur de la femelle au stade approprié, mais une identification précise doit être faite au microscope. L'espèce peut être déterminée sur la base de caractères morphologiques, biométriques ou moléculaires.

En termes d'impact économique, on estime que les nématodes à kyste sont responsables de la perte de 9% de la production totale de pomme de terre à travers le monde (Turner & Rowe 2006). Jusqu'à 80% de la culture peut être perdue quand les populations de nématodes atteignent des niveaux importants, par culture répétée de pomme de terre. En conditions de forte infestation, le rendement peut même être inférieur à la quantité de tubercules plantés (Oerke et al. 1994). Aujourd'hui, la situation est particulièrement sérieuse concernant G. pallida, en raison du manque de cultivars résistants disponibles dans le commerce. Le montant des dégâts est lié au nombre de larves de nématodes par unité de sol à la plantation. La détermination de la densité des populations de nématodes peut être réalisée par deux méthodes : l'analyse du sol et l'observation des racines. Globodera pallida provoque des dégâts quantitatifs mais aussi qualitatifs : les piqûres sur tubercule peuvent en effet rendre le produit non commercialisable. Des coûts additionnels engendrés par les attaques de nématodes sont imputables à l'augmentation de l'irrigation nécessaire pour pallier les perturbations subies par le système radiculaire des plantes parasitées, aux interdictions d'exportation du fait du statut de parasite de quarantaine de l'espèce, et aux traitements nématicides coûteux. Les pertes peuvent être aggravées avec d'autres facteurs influençant la plante, tels que la sensibilité variétale, des carences hydriques ou minérales, la compétition avec les mauvaises herbes ou encore la présence d'autres parasites. L'élimination de ces nématodes s'avère difficile en raison de la grande viabilité des J2 dans les kystes.

Les nématodes étant des animaux édaphiques, les principaux facteurs influençant les conditions du sol agissent directement ou indirectement sur leur nocivité. La température est par exemple un facteur important influençant l'éclosion des œufs, la reproduction, la croissance et la survie. Comme déjà évoqué, l'humidité est également un facteur essentiel influençant la dynamique des populations. Un certain taux d'humidité est nécessaire à l'éclosion des œufs et d'autre part, une humidité trop élevée entraînera un manque d'oxygène dans le sol et ainsi une augmentation des toxines produites par des microorganismes anaérobies, alors qu'une humidité trop faible conduira à un dessèchement aboutissant à l'inactivité ou la mort des individus. Les kystes restent cependant résistants à la dessiccation. D'autres facteurs environnementaux rentrent aussi en compte : la texture du sol influence le développement des racines et l'activité et les mouvements des nématodes, une mauvaise aération entraîne une diminution de la survie et de la densité de la population, et les

propriétés chimiques du sol telles que la salinité, le pH, la teneur en matière organique, engrais et pesticides agissent sur l'éclosion et l'activité des nématodes.

Du fait des dégâts considérables qu'il engendre et de son statut non endémique, *G. pallida* fait partie des organismes nuisibles de quarantaine selon la définition de la FAO (1995) : « organisme nuisible qui a une importance potentielle pour l'économie de la zone menacée et qui n'est pas encore présent dans cette zone ou bien qui y est présent mais n'y est pas largement disséminé et fait l'objet d'une lutte officielle ».

Méthodes de lutte

De par son statut d'organisme de quarantaine, des mesures de lutte obligatoire contre *G. pallida* existent à l'échelle internationale, dont l'application est régie par des textes réglementaires. La législation actuellement en vigueur en Europe réfère à la Directive européenne 2007/33/CE du 11 juin 2007, transcrite en droit français dans l'arrêté du 28 juin 2010 relatif à la lutte contre *G. pallida* et *G. rostochiensis* (voir **Encadré 2**). Cet arrêté stipule que toute parcelle détectée comme contaminée est soumise à des mesures de lutte obligatoire et met en particulier en place un dispositif national de surveillance portant sur au moins 0.5% des surfaces de production de pommes de terre (hors production de plants).

Plusieurs stratégies peuvent être employées, de préférence en association, pour lutter contre *G. pallida* :

• La lutte prophylactique : Il convient de prendre toutes les mesures permettant de garantir la culture de tubercules, plants et bulbes non porteurs de kystes dans la parcelle et l'utilisation de matériel agricole également non porteur de kystes.

• La lutte culturale : Une rotation culturale de 5 à 18 ans selon les sols et la quantité initiale de nématodes dans le sol (Trudgill *et al.* 2003) peut être mise en place, afin de maintenir ou réduire les populations de nématodes et à condition de contrôler les repousses de pommes de terre. Les inconvénients de cette méthode reposent sur sa lenteur, basée sur la mortalité naturelle des larves dans les kystes (climat dépendante), et sur le maintien pendant plusieurs années d'un foyer de contamination dans une zone sensible, avec des risques de dispersion (eau, vent, instruments agricoles). Une accélération de la décontamination peut être réalisée par la mise en place de cultures pièges de pomme de terre. Ces cultures sont ensuite détruites mécaniquement ou chimiquement cinq semaines après la plantation. Cette méthode n'entraîne donc pas de formation de tubercules fils ni de repousse estivale, et permet de réduire de 70 à 80% la population. En conditions climatiques favorables à la culture précoce de pomme de terre, il est possible de jouer sur le différentiel entre vitesse de tubérisation et vitesse de développement des nématodes. L'utilisation de cultures pièges de contre saison, détruites quelques semaines plus tard, peut ici aussi être préconisée.

• La lutte physique : Deux moyens sont possibles, la solarisation, processus utilisant la radiation solaire capturée sous film plastique pour chauffer le sol en surface (50-55°C à 5 cm de profondeur et 40-42°C à 20-25 cm de profondeur), et l'inondation, permettant de tuer les nématodes par asphyxie. Cependant, ces techniques présentent des désavantages qui les rendent très peu utilisées: l'efficacité de la solarisation dépend du climat local (Chauvin *et al.* 2008) et son application est possible uniquement sur petites surfaces, en station expérimentale ou en serre, tandis que l'inondation doit être maintenue plus de 12 mois pour

être réellement efficace (National Research Council (U.S.) – Committee on plant and animal pests 1968), et l'exploitation des parcelles après traitement peut être difficile.

• La lutte biologique : L'utilisation d'ennemis naturels de *G. pallida* est possible, tels que les champignons *Arthrobotrys irregularis* et *Paecilomyces lilacinus*, ou les bactéries des genres *Bacillus* et *Pasteuria* (Schenck *et al.* 1977; Sosamma & Koshy 1997). Cependant, ce type de lutte est bien trop aléatoire en termes de résultats pour constituer une alternative économiquement fiable.

• La lutte chimique : La lutte chimique par l'utilisation de nématicides est actuellement la méthode la plus efficace. Son utilisation est cependant progressivement abandonnée en raison des problèmes environnementaux que posent les produits nématicides. En effet, ces produits sont non sélectifs, alors que nous avons vu précédemment qu'une grande partie des nématodes du sol jouent un rôle important dans l'équilibre de celui-ci. Leur coût est également conséquent et certains ont été interdits à cause de leur toxicité pour l'environnement et la santé humaine, notamment les fumigants tels que le 1,3-Dichloropropène. Les non-fumigants nématostatiques (carbamates et organophosphorés) tels que le Vydate[®] ou le Nemathorin[®] restent autorisés, mais présentent une efficacité moindre.

• La lutte variétale : L'utilisation de variétés résistantes est une alternative récente, la plus intéressante et la moins onéreuse. La culture de variétés résistantes conduit à faire éclore les juvéniles dans le sol, puis à les piéger dans les racines, en bloquant leur développement ou leur multiplication par le développement massif de mâles (phénomène de masculinisation). Après culture, le sol se retrouve ainsi assaini dans une proportion équivalente à celle de l'éclosion des kystes, soit environ 80%. Comparée à un traitement nématicide, son efficacité est équivalente, voire supérieure, et elle n'en comporte pas les inconvénients environnementaux. Par conséquent, cette méthode de lutte est aujourd'hui privilégiée et la création de variétés résistantes est davantage prise en compte dans les programmes de sélection. Les résistances génétiques contre les nématodes identifiées à ce jour se sont globalement révélées plus durables que celles utilisées contre les champignons ou les bactéries pathogènes (Castagnone-Sereno 2002; Molinari 2011). Par exemple, le gène H1 chez la pomme de terre a été utilisé pendant plus de 30 ans contre G. rostochiensis (Fleming 1998), tandis que le gène *Mi-1* a contrôlé pendant plus de 20 ans les nématodes à galles de la tomate avant l'apparition de quelques individus virulents en Californie (Kaloshian et al. 1996). Néanmoins, des cas de contournement de gènes majeurs ont déjà été documentés pour différentes espèces de nématodes à kyste telles que Heterodera glycines, H. avenae, H. schachtii, G. rostochiensis (Huijsman & Lamberts 1972; Lange et al. 1993; Lasserre et al. 1996; Dong & Opperman 1997) et récemment pour G. pallida (Mugniéry et al. 2007; Fournet et al. 2013).

Encadré 2 Extrait de la législation européenne relative à la lutte contre les nématodes à kyste de la pomme de terre.

Des mesures législatives sont indiquées dans la Directive européenne 2007/33/CE du Conseil du 11 juin 2007, qui a pour but d'empêcher la dissémination des nématodes à kyste au sein de l'Union européenne, sans vouloir les éradiquer car ils sont présents dans bon nombre de régions. Cette Directive a été transcrite en droit français dans l'arrêté du 28 juin 2010 relatif à la lutte contre *Globodera pallida* (Stone) et *Globodera rostochiensis* (Wollenweber), nématodes à kyste de la pomme de terre (Ministère de l'Alimentation, de l'Agriculture et de la Pêche). Ci-dessous figurent les principaux éléments de cette Directive :

• La lutte contre *Globodera pallida* (Stone) et *Globodera rostochiensis* (Wollenweber) est obligatoire en tous lieux et de façon permanente sur tout le territoire national (article 1).

• Concernant les pommes de terre de semence, l'analyse de sol par le service régional chargé de la protection des végétaux (SRPV) est obligatoire avant plantation (article 7). L'examen de détection est effectué dans la période allant de la récolte de la dernière culture dans le champ à la plantation ou à l'entreposage des végétaux (article 8).

• Concernant les pommes de terre destinées à la consommation ou à la transformation, 0.5% des surfaces sont analysées chaque année par le SRPV (article 9).

• Si une parcelle est déclarée contaminée, la plantation de pomme de terre est interdite, les repousses doivent être éliminées mécaniquement ou chimiquement et le matériel sortant de la parcelle doit être nettoyé (article 11).

• Des dérogations à l'article 11 peuvent cependant être accordées : des pommes de terre (sauf pommes de terre de semence) peuvent être plantées avec l'accord du SRPV et à condition d'utiliser des variétés résistantes dotées du niveau maximal de résistance disponible et à minima d'un degré de résistance de 7, ou de récolter avant l'arrivée à maturité des femelles de nématode, ou d'avoir désinfecté la parcelle sous la supervision du SRPV (article 12).

• Les mesures de l'article 11 sont à appliquer pour une durée initiale de 6 ans. Cette durée est réduite à 3 ans dans le cas de l'utilisation d'une variété résistante ou de la désinfestation du champ. À l'issue de cette période initiale, un examen de détection est à nouveau effectué et les mesures sont levées si l'analyse est négative. L'ensemble de la procédure est reconduit par périodes de trois ans tant que les nématodes à kyste de la pomme de terre sont détectés (article 15).

1.3.2 La pomme de terre : hôte principal de Globodera pallida

Importance socio-économique de la pomme de terre

La pomme de terre (Solanum tuberosum L.) est une plante à multiplication végétative, cultivée pour ses tubercules de manière annuelle. Elle appartient à la famille des Solanacées, qui comprend d'autres espèces légumières d'importance économique comme la tomate (S. lycopersicon), le piment (Capsicum annuum) et l'aubergine (S. melongena). C'est la 5° production agricole mondiale, après la canne à sucre, le maïs, le riz et le blé, et donc la première denrée alimentaire non céréalière, avec près de 374 millions de tonnes produites en 2013 (Fig. 1.20A, données FAOSTAT 2013). L'Union européenne est à la deuxième place des régions productrices derrière la Chine et devant l'Inde et la Russie. Au sein de l'Union européenne, la France arrive au troisième rang des pays producteurs, derrière l'Allemagne et la Pologne (Fig. 1.20B, données FAOSTAT 2013). Cultivée dans plus de 150 pays, la pomme de terre est un aliment de base pour de nombreux habitants de la planète et joue un rôle clé dans le système alimentaire mondial. Elle est facile à cultiver et sa teneur énergétique est élevée. Les tubercules sont une source importante d'amidon, de protéines, d'antioxydants et de vitamines (Burlingame et al. 2009). D'ici à 2020, il est estimé que plus de deux milliards de personnes dans le monde dépendront de la pomme de terre en tant que source d'alimentation humaine, animale ou de revenus.





Taxinomie et origine

L'ancêtre de la pomme de terre sauvage serait apparu à la fin du Crétacé (65 à 80 millions d'années) en Amérique du Nord ou Amérique Centrale, avant de se répandre en Amérique du Sud à l'Éocène (40 millions d'années). Il aurait ensuite colonisé des milieux favorables vers le nord du Pérou au fur et à mesure de la formation des Andes il y a 20 millions d'années, avant de coloniser à nouveau l'Amérique du Nord au Pliocène, il y 5 millions d'années (Hawkes 1990; Rodríguez & Spooner 2009) (voir Encadré 3).

Encadré 3 La petite histoire de la pomme de terre.

• La pomme de terre naît il y a 8 000 ans, dans la Cordillère des Andes, dans la région du lac Titicaca, où elle pousse à l'état sauvage.

• Les Incas commencent à la cultiver vers 900 av. JC, et la nomment « papa ».





• Vers 1570, la pomme de terre traverse l'Atlantique avec les conquistadores espagnols de retour des Amériques. Introduite d'abord en Espagne sous le nom de « *patata* », elle se diffuse timidement vers l'Italie et les états pontificaux qui la prénomment « *taratouffli* » (petite truffe).

• Elle fait une seconde entrée en Europe au milieu du XVI^{ème} siècle, cette fois-ci par l'Angleterre où l'a ramené l'aventurier Raleigh. Et c'est d'Angleterre qu'elle partira coloniser l'Amérique du Nord. Sa conquête du territoire européen s'accélère, poussée dans les campagnes par les famines et les guerres.

• En 1640, elle est enfin introduite en France, mais restera longtemps réservée à l'alimentation animale.

• Au XVIII^{ème} siècle, Antoine Augustin Parmentier, pharmacien aux armées, découvre ses vertus nutritives pendant son emprisonnement en Prusse. À son retour, il la recommande donc pour résoudre le problème des famines endémiques qui ravagent encore la France à cette époque. Afin de la faire adopter, il a l'idée de faire planter des champs de pommes de terre aux alentours de Paris et de les faire garder le jour seulement par des soldats. La nuit, attirés, les habitants dérobent les précieux tubercules et en assurent ainsi la publicité. Le couronnement de l'action promotionnelle de Parmentier est le dîner qu'il offre au roi et à la reine en 1785 et au cours duquel ne sont servis que des plats comportant des pommes de terre.



• Au XIX^{ème} siècle, d'un légume de jardin, la pomme de terre devient une grande culture. Sa production passe de 1,5 million de tonnes en 1803 à 11,8 millions en 1865. Son importance est devenue telle que, vers 1850, en Irlande, une attaque de mildiou (et l'attitude des Anglais) provoqua la mort d'un million de personnes et l'émigration en Amérique d'un million d'autres.

• Aujourd'hui, la pomme de terre est devenue la cinquième production agricole mondiale et c'est un aliment de base pour de nombreux habitants de la planète. Il en existe des milliers de variétés et des spécialités régionales dans le monde entier. Elle est également utilisée pour l'alimentation animale, l'extraction industrielle de la fécule et la production de plants. On peut s'en servir pour la distillation, en cosmétique et même pour la production d'électricité et la patatogravure !



Photos : Wikipédia (1, 2, 3) et International Potato Center (4).

Le genre Solanum comprend environ 1 400 espèces parmi lesquelles 235 sont tubéreuses et considérées comme apparentées à la pomme de terre. Celles-ci sont largement distribuées sur le continent américain, du sud-ouest des États-Unis jusqu'au sud du Chili et de l'Argentine, du niveau de la mer jusqu'aux hauts plateaux andins. Les zones d'origine et de diversification des espèces tubéreuses s'étendent du nord du Chili au nord du Mexique (Rousselle-Bourgeois & Rousselle 1996). Le centre des Andes (Pérou, Bolivie, Équateur) et le centre du Mexique sont les régions les plus riches en espèces (Hijman & Spooner 2001). Au sein du genre Solanum, la pomme de terre appartient à la section Petota, qui regroupe l'ensemble des pommes de terre sauvages et cultivées. L'une des classifications les plus répandues reconnaît sept espèces de pomme de terre cultivées. L'espèce S. tuberosum est divisée en deux sous-espèces : tuberosum et andigena. Parmi ces huit espèces et sousespèces, seule Solanum tuberosum ssp. tuberosum est cultivée dans le monde entier, tandis que les autres sont circonscrites aux pays andins, où ont été trouvés des milliers de cultivars primitifs. Les communautés d'agriculteurs des Andes exploitent plus de 175 variétés de pommes de terre locales. On reconnaît généralement environ 150 espèces de pomme de terre sauvages, toutes circonscrites au continent américain.

La sous-espèce Solanum tuberosum ssp. tuberosum est issue de l'adaptation à la tubérisation en jours longs de la sous-espèce andine Solanum tuberosum ssp. andigena, et sa variabilité génétique est relativement restreinte. Les clones de la sous-espèce *S. tuberosum* ssp. andigena tubérisent en jours courts dans leur zone d'origine, et seuls les clones sélectionnés initialement pour leur aptitude à tubériser en jours longs forment aujourd'hui la base génétique des variétés commerciales européennes et nord-américaines de pommes de terre (Spire & Rousselle 1996). Cette base génétique restreinte ne couvre donc pas tous les besoins des sélectionneurs, surtout en ce qui concerne les gènes de résistance aux ravageurs et aux maladies. En revanche, de telles sources de résistances sont présentes dans de nombreuses espèces apparentées à *S. tuberosum* (Ross 1986; Hawkes 1990). Il est ainsi possible d'exploiter par croisement sexué la variabilité génétique présente chez ces espèces, qui est particulièrement importante pour la résistance à des stress abiotiques et pour la résistance aux bioagresseurs. Des caractères de résistance à de nombreux bioagresseurs ont ainsi été introduits dans le génome de *S. tuberosum* (Caromel 2004).

Principaux bioagresseurs

La pomme de terre est attaquée par un grand nombre de bioagresseurs : des oomycètes, des virus, des champignons, des bactéries, des insectes, des nématodes (voir **Encadré 4**), qui se manifestent pendant la culture ou lors du stockage. Ils sont responsables de 32% des pertes mondiales de rendement en pommes de terre (Oerke *et al.* 1994). L'importance relative des différents bioagresseurs est variable selon les conditions climatiques, les conditions de conservation et la possibilité d'utiliser ou non des produits phytosanitaires efficaces. Pour éviter leur introduction dans une zone géographique qui en est indemne, certains de ces bioagresseurs figurent sur les listes de quarantaine. Comme nous l'avons vu pour le cas des nématodes à kyste, des mesures drastiques de vérification du matériel végétal et de lutte sont mises en place, mais celles-ci peuvent avoir un poids économique important. Les nématodes à kyste ainsi que l'agent du mildiou *Phytophthora infestans* sont les bioagresseurs les plus dommageables pour la pomme de terre, dans les zones de culture à climat tempéré.

Encadré 4 Principaux bioagresseurs de la pomme de terre.

L'agent du mildiou Phytophthora infestans

Considéré comme le parasite majeur de la pomme de terre, l'oomycète *P. infestans* cause des dégâts pouvant atteindre 90 % de la culture en conditions favorables. Il se développe sur les organes aériens de la plante, mais peut aussi contaminer les tubercules et les rendre impropres à la consommation.



Les nématodes

Ils engendrent des pertes estimées à 12,2 % de la production mondiale (Barker 1998). Les endoparasites provoquent les dégâts les plus importants, en particulier ceux des genres *Meloidogyne* (nématodes à galles) et *Globodera* (nématodes à kyste). *Nacobus aberrans*, présent dans les pays andins, *Pratylenchus spp.* aux États-Unis, et *Ditylenchus destructor*, très répandu dans le monde et sur liste de quarantaine en Europe, causent également des dégâts notables. De nombreuses espèces ectoparasites se nourrissent sans causer de dégâts apparents. Seules quelques espèces des genres *Trichodorus* et *Paratrichodorus* sont considérées comme dangereuses de par leur transmission du virus du Rattle.



Les virus

Le virus de l'enroulement de la pomme de terre (PLRV) et le virus Y de la pomme de terre (PVY, photo) sont les virus les plus répandus et les plus dommageables, mais plus de 20 autres virus ont également été décrits (Kerlan 1996). Certains virus peuvent entraîner des pertes de rendement supérieures à 50 %.

Les champignons

Le rhizoctone brun, la gale argentée (photo), la gangrène, les verticillioses et les fusarioses sont les maladies fongiques les plus répandues sur les cultures de pomme de terre. D'autres champignons tels qu'*Alternaria solani* et *Synchytrium endobioticum* peuvent provoquer de graves dégâts, mais ne sont présents ou importants que dans des zones géographiques restreintes.





Les bactéries

Les maladies bactériennes les plus importantes sont causées par des organismes des genres *Erwinia*, qui provoquent la maladie de la jambe noire et entraînent la pourriture molle, *Streptomyces*, qui sont les agents de la gale commune de la pomme de terre (photo), et *Ralstonia*, qui entraînent un flétrissement des plantes en culture.

Les insectes

Les insectes qui causent le plus de dégâts sont les pucerons, le doryphore, la teigne et le taupin. Les pucerons causent principalement des dégâts indirects en tant que vecteurs de nombreux virus, comme *Myzus persicae* qui transmet le plus de virus sur pomme de terre. Les larves et adultes de doryphore peuvent provoquer une défoliation quasi complète des plantes. Les larves de teigne creusent des galeries dans les feuilles, les tiges ou les tubercules (photo).



Photos : Wikipedia (1), Sutton Bridge CSR (2), www.potato-tuber-blemishes.com (3, 4, 5).

Ressources génétiques et génomiques

Depuis les années 1980, les ressources génétiques et génomiques de la pomme de terre ne cessent de s'améliorer. Les cartes génétiques disponibles pour la pomme de terre ont suivi l'amélioration des techniques de cartographie génétique, qui ont permis de passer des premières cartes génétiques de marqueurs RFLP (Bonierbale *et al.* 1988; Gebhardt *et al.* 1989) à des cartes de marqueurs PCR (AFLP, microsatellites, CAPS et SCAR), puis à une carte à Ultra Haute Densité (Isidore *et al.* 2003) pour aboutir récemment à une cartographie physique (de Boer *et al.* 2011). En 2003, Rickert *et al.* ont permis de mettre en évidence 1498 SNPs et 127 polymorphismes d'insertion/délétion (Indels). Plus de 250 000 ESTs sont disponibles dans les bases de données publiques. Le génome de la pomme de terre a été séquencé en 2011 (The Potato Genome Sequencing Consortium 2011), ce qui en fit la première espèce du clade des Astérides à être séquencée. Son génome mesure 840 Mb, ce qui représente une taille moyenne pour un génome de plante, et 39 000 gènes y sont prédits. Le nombre chromosomique de base de l'ensemble des espèces de *Solanum* tubéreuses est 12. La pomme de terre est une espèce tétraploïde.

En outre, une bonne conservation de la synténie entre les différentes espèces de Solanacées, mise en évidence par génétique comparative, permet également d'exploiter les travaux réalisés sur ces espèces (Bonierbale *et al.* 1988; Livingstone *et al.* 1999; Doganlar *et al.* 2002). C'est le cas notamment pour la tomate, espèce modèle des Solanacées, dont le génome a récemment été séquencé (The Tomato Genome Consortium 2012).

1.3.3 Interactions plante-nématode

Le pouvoir pathogène de *G. pallida*

Le pouvoir pathogène d'un nématode phytoparasite se définit par sa capacité à se reproduire sur une plante. La stratégie mise en place par G. pallida pour parvenir à infecter son hôte fait intervenir plusieurs processus tels que la dégradation des parois des cellules végétales, la suppression ou l'évitement des défenses de la plante et la manipulation des voies de signalisation de son hôte pour générer un site nourricier complexe. Ces différentes étapes mobilisent les gènes du pouvoir pathogène, qui correspondent aux gènes permettant au nématode de parasiter la plante, de se développer jusqu'au stade adulte et de se multiplier. Une partie de ces gènes code des protéines produites au niveau de divers organes de sécrétion, tels que les glandes salivaires, les amphides ou la cuticule (Haegeman et al. 2012), et qui seront déversées dans les tissus de l'hôte. Ces protéines sécrétées sont appelées « effecteurs » et possèdent toutes un peptide signal, c'est-à-dire un chainon de 15 à 20 acides aminés situé à l'extrémité N-terminale de la forme traduite de la protéine destinée à être sécrétée. Un effecteur est plus précisément défini comme toute protéine ou petite molécule du pathogène qui modifie la structure et la fonction des cellules de l'hôte, entraînant ainsi le déclenchement des différents niveaux de réponse de défense de la plante (PAMPs, facteurs d'avirulence) ou facilitant l'infection (facteurs de virulence) (Hogenhout et al. 2009). Alors que les effecteurs ont été très étudiés chez les bactéries et les champignons (Pritchard & Birch 2011), ceux des nématodes ne l'ont été que tardivement. Plusieurs études se sont en particulier intéressées aux sécrétions des nématodes produites au niveau des glandes salivaires et sécrétées via le stylet et ont ainsi permis de mieux comprendre leur rôle dans les différentes étapes du parasitisme. *Globodera pallida* et plus largement les Tylenchida et les Aphelenchida possèdent deux glandes salivaires subventrales et une glande salivaire dorsale.

Pénétration, migration et dégradation des tissus végétaux

Les étapes de pénétration dans la plante et de migration jusqu'au cylindre central font intervenir, en plus de l'action mécanique du stylet, des enzymes permettant de dégrader les parois des cellules végétales, telles que des β -1,4-endoglucanases (cellulases) qui dégradent la cellulose, des pectate lyases et polygalacturonases qui dégradent les pectines ou des xylanases qui dégradent l'hémicellulose, dont les transcrits ont tous été localisés dans les glandes ventrales du nématode (Grenier et al. 2002; Jasmer et al. 2003; Jones et al. 2009). D'autres protéines, comme des glycoprotéines (Duncan et al. 1997), des métalloprotéases (Robertson et al. 1999) ou des protéines assimilées à des protéines de venin d'hyménoptère (Gao et al. 2002), ont été identifiées dans les glandes ventrales de nématodes à kyste, suggérant leur implication potentielle dans ces premières phases du parasitisme, même si leur fonction est encore inconnue. Ces protéines montrent de très fortes similarités avec des homologues de bactéries ou de champignons, ce qui suggère que les gènes du nématode codant ces protéines pourraient avoir été acquis par transfert horizontal (Haegeman et al. 2011; Danchin et al. 2010). D'autres protéines, des expansines et des protéines avec domaine de liaison à la cellulose, agissent également sur les parois cellulaires, non pas directement par une action de lyse mais en facilitant l'action des enzymes de dégradation, en rendant les parois plus accessibles (Qin et al. 2004; Kudla et al. 2005; Jones et al. 2009; Danchin et al. 2010; Haegeman et al. 2010, 2011). Enfin, les nématodes sécrètent des protéases qui peuvent dégrader les protéines de l'hôte faisant partie des parois cellulaires, ou les protéines qui agissent dans les défenses de la plante.

Protection du nématode et suppression des défenses de l'hôte

Tout au long du parasitisme, le nématode est exposé aux réactions de défense mises en place par la plante, dont il doit se protéger. La plante sécrète par exemple des toxines à large spectre anti-pathogène, telles que des phytoalexines, des isoflavonoides ou des terpénoides, dont certaines sont nématicides (Jasmer *et al.* 2003; Wuyts *et al.* 2006, 2007). Les nématodes zooparasites utilisent des glutathione-S-transférases (GSTs) pour détoxifier ces composés, et il se pourrait qu'il en soit de même pour les nématodes phytoparasites. Plusieurs GSTs ont en effet été identifiés chez le nématode à galles *Meloidogyne incognita*, dont une au moins est exprimée dans les glandes salivaires. L'autre grande classe de composés de défense produits par les plantes réunit les espèces réactives de l'oxygène (ROS), que les nématodes contournent en produisant une batterie d'antioxydants.

Concernant le contournement des défenses basales de l'hôte, aucun éliciteur de la PTI n'a à ce jour été identifié chez les nématodes à kyste, même si les symptômes typiques de son déclenchement en présence des nématodes sont observés (Waetzig *et al.* 1999). L'importance du rôle de la camalexine et du glucosinolate dans le déclenchement de la PTI a néanmoins été montrée récemment chez les nématodes à galles (Texeira *et al.* 2016). Concernant la résistance hôte spécifique, même si de nombreux gènes de résistance aux nématodes ont été clonés et cartographiés (voir section suivante), la grande majorité des gènes d'avirulence correspondants reste à identifier (Smant & Jones 2011). Le seul effecteur connu comme déclencheur de l'ETI (déclenchement d'une réaction d'hypersensibilité en présence du gène de résistance *Gpa2*) est la protéine Gp-RBP-1 chez *G. pallida* qui appartient à la famille multigénique des SPRYSECs (*secreted SPRY domain-containing proteins*). Chez *G. pallida* et

plus largement chez les nématodes à kyste, des protéines appartenant à la famille multigénique des SPRYSECs sont spécifiquement exprimées au niveau de la glande dorsale du nématode durant les phases précoces du parasitisme (Blanchard *et al.* 2005; Rehman *et al.* 2009). En plus d'un peptide signal caractéristique des protéines sécrétées hors des cellules, ces protéines possèdent un domaine B30.2 connu pour être impliqué dans des interactions de type protéine/protéine (Woo *et al.* 2006). Ces protéines ciblent différents compartiments infracellulaires de l'hôte : une majorité est exprimée au niveau du noyau, mais d'autres sont retrouvées dans le cytoplasme, comme Gp-RBP-1 par exemple (Jones *et al.* 2009). Ceci suggère que les protéines de type SPRYSEC auraient des cibles cellulaires variées et pourraient aussi agir en tant que suppresseurs des défenses de l'hôte (Jones *et al.* 2009; Rehman *et al.* 2009).

En outre, un autre moyen de supprimer les réactions de défense de l'hôte consisterait en l'exploitation du système ubiquitine-protéasome des plantes, ce qui a déjà été observé dans d'autres pathosystèmes (*e.g.* Birch *et al.* 2009). Ce système, découvert récemment, a un rôle central dans la dégradation contrôlée des protéines, et consiste en la fixation d'ubiquitines sur une protéine à dégrader, puis en la dégradation de cette protéine spécifiquement marquée par le protéasome 26S. Les nématodes pourraient interférer avec ce système en sécrétant des effecteurs intervenant dans l'ubiquitination, qui pourraient notamment cibler des protéines responsables du déclenchement de la PTI. De telles protéines ont été identifiées dans les glandes dorsales d'*Heterodera glycines*, *H. schachtii* et *G. pallida* (Gao *et al.* 2003; Tytgat *et al.* 2004 ; Jones *et al.* 2009).

Si les sécrétions produites au niveau des glandes salivaires et injectées par le stylet sont les principaux composés impliqués dans le parasitisme, et sur lesquels se sont donc majoritairement concentrées les recherches, des sécrétions produites au niveau d'autres organes jouent également un rôle dans la protection du nématode et la suppression des défenses de l'hôte. Chez *G. pallida* par exemple, une protéine sécrétée au niveau de la cuticule, Gp-FAR-1, jouerait un rôle dans la suppression de la synthèse des jasmonates, des phytohormones qui interviennent normalement en tant que médiateurs cellulaires de la résistance végétale (Prior *et al.* 2001). Des sécrétions produites par les deux sensilles antérieures, les amphides, mais aussi les postérieures, les phasmides, interviendraient également dans le parasitisme. Ainsi, la protéine MAP-1 produite au niveau des amphides chez *M. incognita* est supposée être impliquée dans la réaction incompatible avec la résistance de la tomate (Semblat *et al.* 2001).

Induction et maintien du site nourricier

Une fois parvenu au cylindre central, le nématode doit induire un site nourricier et le maintenir tout au long de son développement pour assurer sa nutrition. Ces étapes impliquent la sécrétion de protéines qui modifient localement le développement et le métabolisme de l'hôte, et dont les transcrits ont été localisés à la fois dans les glandes ventrales et dans la glande dorsale. L'initiation du site nourricier implique une réactivation du cycle cellulaire suivi d'un processus de différenciation, des mécanismes qui sont habituellement régulés par des phytohormones, en particulier l'auxine. Le nématode doit donc manipuler localement les balances hormonales pour s'établir dans l'hôte (Grunewald *et al.* 2009). Des chorismate mutases ont été identifiées comme des effecteurs du nématode impliqués dans la manipulation de l'auxine. Ils interfèrent localement avec la voie de biosynthèse de cette hormone et en ajuste également la distribution dans les cellules, dans le but d'en augmenter la concentration au voisinage du site nourricier, permettant ainsi son initiation et son développement (Gheysen & Fenoll 2002). D'autres effecteurs tels que les GSTs et les TTLs

(*Transthyretin-like proteins*) pourraient théoriquement jouer un rôle dans la régulation de l'auxine, mais leur implication réelle reste à déterminer.

En plus de cette action indirecte sur la régulation du cycle cellulaire par une manipulation des mécanismes hormonaux, des protéines du nématode pourraient également intervenir de manière directe, selon un fonctionnement intéressant : ces protéines seraient en fait des homologues fonctionnels de protéines de plantes impliquées dans les processus de régulation cellulaire, et pourraient donc agir directement en déclenchant la voie de signalisation correspondante au sein de la plante hôte, afin d'initier le site nourricier. C'est le cas par exemple des peptides CLE (*Clavata-like Element*) mis en évidence dans la glande dorsale de *G. rostochiensis* (Lu *et al.* 2009).

Enfin, l'alimentation du nématode au niveau du syncytium fait intervenir des protéases qui permettent une digestion extracorporelle des composés cellulaires des cellules hôtes, les nématodes ne possédant pas de système de digestion.

Les facteurs de résistance aux nématodes à kyste de la pomme de terre

Seulement deux résistances aux nématodes à kyste ont été identifiées au sein de la sous-espèce *S. tuberosum* ssp. *tuberosum*. Il s'agit des QTL *GpaVIII*_{tbr} et *GpaXI*^I_{tbr}, cartographiés respectivement sur les chromosomes VIII et XI, et actifs contre le pathotype Pa2/3 de *G. pallida* (Caromel 2004). Cependant, leur effet est relativement faible.

En revanche, plusieurs espèces apparentées, telles que Solanum tuberosum ssp. andigena ainsi que des Solanum sauvages, présentent des niveaux de résistance aux nématodes à kyste plus intéressants, ce qui a poussé les sélectionneurs à initier des programmes d'introgression des résistances identifiées chez ces espèces apparentées (gènes majeurs ou QTL) au sein de S. tuberosum ssp. tuberosum (Turner 1989; Rousselle-Bourgeois & Mugniéry 1995). Cependant, ces sources de résistance présentent souvent une grande spécificité et ne sont efficaces que contre une espèce voire une ou quelques populations. Les gènes H2 et Gpa2 sont pour le moment les seuls gènes majeurs de résistance à G. pallida mis en évidence. Le gène H2 est originaire d'un clone de S. multidissectum et est efficace contre le pathotype Pa1 (Dunnett 1961). Le gène Gpa2 a été identifié sur le clone CPC1673 de la sousespèce S. tuberosum ssp. andigena et cartographié sur le chromosome XII (Rouppe van der Voort et al. 1997), néanmoins ce gène n'est actif que contre deux populations hollandaises du nématode (D383 et D372), appartenant au pathotype Pa2/3. Il est également physiquement lié à Rx (Rouppe van der Voort et al. 1999), gène de résistance au virus X de la pomme de terre (Ritter et al. 1991; Bendahmane et al. 1997), et a ainsi été introduit fortuitement dans des variétés commerciales suite à la sélection pour la résistance au virus X. Mis à part H2 et Gpa2, les autres résistances à G. pallida identifiées jusqu'à présent chez des espèces apparentées ont un déterminisme polygénique. Ces QTL sont au nombre de neuf. Deux QTL ont été identifiés sur les chromosomes V et IX chez S. vernei (Rouppe van der Voort et al. 2000), deux QTL sur les chromosomes V et XI chez S. sparsipilum (Caromel et al. 2005), trois QTL sur les chromosomes V, VI et XII chez S. spegazzinii (Caromel et al. 2003), un QTL sur le chromosome XI de S. tarijense (Gpa3, Wolters et al. 1999) et un QTL sur le chromosome IV de S. tuberosum ssp. andigena (Bradshaw et al. 1998). Les trois QTL identifiés sur les chromosomes V de S. vernei, S. sparsipilum et S. spegazzinii sont en position colinéaire. Ces trois QTL ainsi que le QTL présent sur le chromosome XI de S. tarijense montrent un effet fort, tandis que les autres QTL montrent un effet moindre. La sélection de plantes porteuses de ces différents facteurs de résistance suppose le développement de marqueurs permettant de suivre les résistances. Pour le moment, des marqueurs permettant de suivre les facteurs de résistance situés sur le chromosome V des 3 espèces ont été développés, mais ce n'est pas encore le cas pour les autres QTL. Suite à ces travaux de sélection, la variété lledher, portant le QTL majeur de résistance *GpaVvrn*, issu de *S. vernei*, a été inscrite en 2009 au catalogue français comme première variété de pomme de terre de consommation résistante à *G. pallida*. Au niveau européen, d'autres cultivars résistants basés sur ces même QTL, tels qu'Innovator ou Arsenal, ont été développés.

Des résistances à l'espèce sœur de G. pallida, G. rostochiensis, ont également été identifiées. Le gène majeur H1 a été identifié sur le même clone de S. tuberosum ssp. andigena qui a fourni le gène majeur Gpa2 résistant à G. pallida. Le gène H1, cartographié sur le chromosome V, induit une résistance totale contre les pathotypes Ro1 et Ro4 (Gebhardt et al. 1993; Pineda et al. 1993). Il est présent dans la variété Maris Piper inscrite au catalogue anglais en 1967 (Rice et al. 1985; Mugniéry & Balandras 1986) et aujourd'hui également dans de nombreuses variétés européennes. Un gène dominant induisant une résistance partielle aux pathotypes Ro1 et Ro4, GroV¹, a également été cartographié sur le chromosome V de S. vernei (Jacobs et al. 1996). Un QTL de résistance à G. rostochiensis, Grp1, qui confère également une résistance partielle à G. pallida, a aussi été cartographié sur le chromosome V (Rouppe van der Voort et al. 1998). Le clone tétraploïde duquel est issu le parent résistant utilisé par les auteurs a dans son pedigree de nombreux clones d'espèces sauvages, en particulier trois clones de S. vernei, un clone de S. oplocense, et le clone de S. tuberosum ssp. andigena CPC1673. L'espèce S. spegazzinii fournit également plusieurs résistances à G. rostochiensis : le gène majeur de résistance Gro1, actif contre le pathotype Gro1 et identifié sur le chromosome VII (Barone et al. 1990; Paal et al. 2004), ainsi que les QTL Gro1.4, Gro1.2 et Gro1.3 cartographiés respectivement sur les chromosomes III, X et XI (Kreike et al. 1993; Kreike et al. 1996).

L'importance économique des dégâts causés par les nématodes à kyste sur la tomate est moindre, ce qui explique que la résistance à ces nématodes ait été moins étudiée chez cette espèce. Néanmoins, des accessions d'espèces apparentées à la tomate, *Lycopersicum pimpinellifolium* et *L. peruvianum*, présentent de bons niveaux de résistance (Ellis 1968). Un gène majeur, *Hero*, conférant une résistance totale à plusieurs pathotypes de *G. rostochiensis* et de *G. pallida*, cartographié sur le chromosome IV (Ganal *et al.* 1995), a été introgressé à partir de *L. pimpinellifolium* dans l'espèce cultivée *L. esculentum*.

Les résistances présentées précédemment peuvent agir à différentes étapes du cycle de vie du nématode. Certaines agissent à des étapes précoces en bloquant la larve dès sa pénétration ou lors de sa migration dans la racine. C'est notamment ce qu'il peut se produire lorsque les deux QTL originaires de *S. sparsipilum* sont présents ensemble dans un même génotype. Les résistances peuvent aussi agir à des étapes plus tardives. Lors de l'induction du site nourricier, une réaction d'hypersensibilité peut se produire, perturbant voire bloquant le développement du nématode. Cette réaction est notamment observée avec les résistances monogéniques *H1*, *H2* et *Hero*, et se traduit par une réaction rapide de nécrose accompagnée d'un dépôt de lignine autour du syncytium, qui devient défavorable à la formation des femelles (Rice *et al.* 1985; Mugniéry & Balandras 1986). Enfin, les résistances peuvent agir encore plus tardivement, lors du développement larvaire et de la différenciation sexuelle. Le mauvais fonctionnement du site nourricier ou la mauvaise qualité de la plante hôte peuvent entraîner la masculinisation des juvéniles, réduisant ainsi le taux de multiplication du parasite. En effet, la différenciation sexuelle des nématodes à kyste étant strictement épigénétique,

l'expression de la résistance se manifeste fréquemment par une inversion de la sex-ratio habituellement observée sur plantes sensibles. Lorsqu'une résistance se manifeste par une perturbation du développement des syncytia, et donc par une moins bonne alimentation des nématodes, ceux-ci restent bloqués à un stade larvaire ou évoluent préférentiellement en mâles (Mugniéry & Fayet 1981, 1984; Rice *et al.* 1985, 1987). Ceci est notamment observé dans le cas de l'utilisation de résistances polygéniques issues de *S. vernei, S. sparsipilum, S. spegazzinii*, ainsi que pour le gène *Gpa2*, où le développement du syncytium est freiné sans qu'on observe de réaction de nécrose autour de celui-ci. Ce syncytium atrophié ne permet pas aux nématodes de se nourrir suffisamment, ce qui entraîne leur différenciation en mâles.

La structure et la fonction des protéines codées par les QTL sont souvent méconnues. Cependant, les gènes *Hero* (Ernst *et al.* 2002) chez la tomate, et *Gpa2* (Van der Vossen *et al.* 2000) et *Gro1* (Paal *et al.* 2004) chez la pomme de terre ont pu être clonés et leur structure ainsi étudiée. Les trois gènes codent des protéines de type NBS-LRR (Jones 2001), qui possèdent un domaine de liaison aux nucléotides (*Nucleotide Binding Site*, NBS) impliqué dans la transduction du signal, et un domaine riche en répétitions de leucines (*Leucine-Rich Repeats*, LRR), impliqué dans la spécificité de reconnaissance du pathogène. La classification des NBS-LRR se fait sur la présence ou l'absence d'autres domaines. Les protéines de la première sous-famille (TIR-NBS-LRR) possèdent un domaine TIR (*Toll-Interleukin-like Receptor*), et c'est notamment le cas de *Gro1*. La seconde sous famille (non-TIR-NBS-LRR) regroupe l'ensemble des autres NBS-LRR présentant d'autres domaines, dont les CC-NB-LRR qui possèdent un domaine CC (*Coiled Coil* ou faisceau d'hélices), c'est le cas de *Gpa2* et *Hero*.

1.3.4 Évolution expérimentale et résultats phénotypiques

L'ensemble de cette thèse repose sur l'exploitation d'un matériel biologique obtenu au cours d'une évolution expérimentale. Celle-ci a préalablement permis d'obtenir des résultats phénotypiques (Fournet *et al.* 2013) qui ont motivé l'élaboration de ce projet de thèse. Nous allons ici détailler l'évolution expérimentale et revenir sur les principaux résultats qu'elle a permis d'obtenir, afin de mieux appréhender la présente étude.

Nous avons vu précédemment que *S. vernei* était une source de résistance prometteuse face à *G. pallida*. Plusieurs sélectionneurs privés et l'INRA ont réalisé des croisements entre *S. vernei* et des variétés cultivées afin de créer de nouvelles variétés de pomme de terre commercialisables résistantes à *G. pallida*, ce qui a notamment abouti à l'inscription d'Iledher au catalogue français. Le développement d'un matériel végétal résistant doit idéalement passer par l'étude de la durabilité de cette résistance. Fournet *et al.* (2013) ont ainsi suivi la virulence de deux populations de *G. pallida* au cours d'une évolution expérimentale dans le but d'étudier la variabilité de la durabilité de la résistance conférée par le QTL *GpaVvrn* dans différents fonds génétiques.

• Évolution expérimentale

Quatre génotypes résistants ont été utilisés au cours de l'évolution expérimentale, en plus de la variété sensible Désirée servant de témoin à l'étude, et présentent tous dans leur généalogie le parent *S. vernei*. L'un d'eux, 60.96.1, est la variété lledher inscrite au catalogue français. La présence dans les génotypes résistants du QTL majeur GpaV_{vrn}, qui explique 61%

de la résistance, a été vérifiée par méthode CAPS (*Cleaved Amplified Polymorphic Sequence*, décrit dans Caromel *et al.* 2011). En revanche, la présence du QTL mineur expliquant 24% de la résistance n'a pas été vérifiée.

L'évolution expérimentale a été conduite sur deux populations indépendantes de *G. pallida* ($F_{ST} \approx 0.2$, voir **Chapitre 3**), provenant de deux sites distants et hautement infestés : la population SM, provenant d'un champ proche de Saint-Malo comptant 3 J2 par gramme de sol, et la population N, provenant d'une parcelle sur l'île de Noirmoutier et présentant 82 J2 par gramme de sol. Ces deux populations ont été exposées pendant huit générations successives (soit huit années) aux différents génotypes de pomme de terre résistants et à la variété sensible Désirée, selon le schéma suivant :

• Les trois premières générations (1x, 2x, 3x) ont été réalisées directement dans les parcelles de provenance des deux populations. En effet, à cause du fort niveau de résistance des génotypes de pomme de terre utilisés dans l'étude, l'obtention d'un minimum d'individus nécessaires à initier la génération suivante était dépendante d'une exposition à un grand nombre de plants. Chaque champ a été divisé en cinq microparcelles au sein desquelles un des cinq génotypes de pomme de terre a été planté pendant trois années consécutives. Chaque microparcelle correspondait ainsi à une lignée de nématode.

• Les deux générations suivantes de chaque lignée (4x, 5x) ont été obtenues en prélevant dans chaque microparcelle de chaque champ 20 L de sol, placés ensuite dans des bacs sous serre et plantés avec les mêmes génotypes pendant deux années consécutives.

• Après ces cinq premières générations, les kystes du sol de chaque bac (*i.e.* chaque lignée) ont été extraits avec un élutriateur Kort. Les trois dernières générations de chaque lignée (6x, 7x, 8x) ont été réalisées en pot. À chaque génération et pour chaque lignée, 10 kystes (soit environ 1500 J2) de la génération précédente ont été échantillonnés et placés dans un sac de tulle disposé ensuite dans un pot de 13 cm rempli d'un mélange de sol (2:1 sable:terre tamisée) non contaminé en kystes. Les génotypes de pomme de terre correspondant à chaque lignée ont ensuite été plantés et recouverts du mélange de sol. Ce procédé a été répété quatre fois à chaque génération). Les pommes de terre ont ensuite été mises à pousser sous serre, en conditions contrôlées, pendant 120 jours. Les nouveaux kystes formés au sein des quatre réplicas ont été extraits, poolés et conservés à 4 °C pendant un minimum de quatre mois (correspondant au temps de diapause). Les quatre échantillons de 10 kystes requis pour la génération suivante ont ensuite été prélevés aléatoirement au sein des pools.

La **Figure 1.21** récapitulant l'évolution expérimentale indique la dénomination des lignées obtenues, qui sera utilisée tout au long de ce manuscrit (excepté dans le **Chapitre 2**, qui utilise une dénomination similaire à celle retrouvée dans l'étude de Fournet *et al.* (2013)). Celle-ci reprend le nom de la population d'origine suivi du génotype de pomme de terre sur lequel a été réalisée l'évolution expérimentale. Un suffixe peut ensuite compléter ce nom lorsque l'on s'intéresse à une génération précise de la lignée, par exemple « 8x » pour la huitième génération.

Plusieurs lignées issues de la population N ont été perdues au cours de l'évolution expérimentale, l'efficacité de la résistance ne permettant pas de récupérer suffisamment de kystes pour initier la génération suivante.


Figure 1.21 Schéma de l'évolution expérimentale. Les lignées de nématode ont été obtenues à partir de deux populations naturelles, SM et N, provenant respectivement de Saint-Malo et Noirmoutier, exposées expérimentalement pendant huit générations successives à quatre génotypes de pomme de terre résistants, portant tous le QTL de résistance *GpaVvrn* mais différant par leur fond génétique, et à la variété sensible Désirée. Les trois premières générations (1x, 2x, 3x) ont été réalisées au champ, les deux suivantes (4x, 5x) en bacs sous serre et les trois dernières (6x, 7x, 8x) en pots. Les flèches se décrochant à la septième génération correspondent à du matériel biologique utilisé dans le cadre de la thèse mais ne faisant pas partie de l'étude de Fournet *et al.* (2013). Les lignées utilisées dans la thèse sont soulignées.



Femelles

Figure 1.22 Illustration du phénotypage. Les flèches orange indiquent la position des individus sur la racine de pomme de terre (photos : Delphine Eoche-Bosy).

Résultats de l'étude phénotypique

L'évaluation de la résistance d'une plante se basant sur son efficacité à réduire le nombre de femelles produites par rapport à un témoin non résistant, l'étude phénotypique a consisté à évaluer et comparer le nombre de femelles obtenues sur les différents génotypes. Le phénotypage (Fig. 1.22) a été réalisé sur les cinquième et huitième générations de chaque lignée (*i.e.* 5x et 8x), d'après le protocole développé par Mugniéry & Person (1976), consistant à inoculer en boîte de Pétri des larves de stade J2 sur racines de pomme de terre, et à compter 15 jours plus tard le nombre de femelles obtenues. Les résultats ont mis en évidence que toutes les lignées ayant évolué sur génotype résistant ont contourné la résistance de leur plante hôte (Fig. 1.23). Il a également été montré que le fond génétique de la plante hôte a une influence sur la durabilité de la résistance. En effet, alors que le pourcentage de femelles obtenues sur lledher dans les deux populations au bout de huit générations est devenu équivalent à celui obtenu sur hôte sensible, indiquant donc que la résistance a été totalement contournée, ce n'est pas encore le cas des autres lignées qui ne montrent qu'un contournement partiel de la résistance.





Exploitation du matériel biologique issu de l'évolution expérimentale dans le cadre de la thèse

Le matériel biologique produit lors de l'évolution expérimentale représente une véritable ressource pour explorer plus en détails les causes et conséquences de l'adaptation mise en évidence par l'étude phénotypique, que nous avons exploitée lors de ce travail de thèse.

La grande majorité du matériel précédant la huitième génération de l'évolution expérimentale n'est plus disponible, c'est pourquoi la plupart des expériences menées au cours de ce projet portent sur cette huitième génération. Le protocole d'évolution expérimentale prévoyait également de remultiplier une partie du matériel virulent sur la variété sensible Désirée. C'est notamment ce qui a été fait pour les lignées NI et SMI, dont une partie a été multipliée pendant sept générations sur lledher, puis sur Désirée en huitième génération. Ce matériel a également été inclus à notre étude. Enfin, les lignées SM96F et SM94T ne présentaient pas suffisamment de matériel en huitième génération pour pouvoir être exploitées. Ceci s'explique car ces deux lignées ont été confrontées aux deux résistances les plus durables (Fig. 1.23), qui n'ont permis la production que d'un nombre restreint de kystes, qui ont tous été utilisés pour l'étude phénotypique. Cependant, SM94T ayant elle aussi subi une remultiplication sur Désirée, nous avons choisi d'intégrer à notre étude le matériel SM94T-7xD (*i.e.* sept générations sur 94T.146.52 et une huitième génération sur Désirée). La lignée SM96F a quant à elle été supprimée de l'étude. Toutes les lignées utilisées au cours de notre étude sont soulignées dans la Figure 1.21.

Du matériel antérieur à la huitième génération est encore disponible pour les lignées SMD et SMI. De plus, l'évolution expérimentale s'est poursuivie pour ces lignées au-delà de la huitième génération. Ainsi, les sixième et dixième générations de ces lignées sont également considérées dans une partie de ce manuscrit.

1.4 Objectifs et plan de la thèse

L'utilisation de la résistance variétale s'impose aujourd'hui comme une méthode de lutte à développer et à privilégier. Néanmoins, nombre d'études mettent en évidence la capacité des organismes ciblés à contourner ces résistances. La création variétale étant un processus long et les sources de résistances n'étant pas infinies, il est nécessaire de mettre en place des stratégies de gestion durable de ces résistances. Pour être efficaces, celles-ci doivent inclure des éléments prédicteurs de la vitesse d'adaptation des pathogènes. Le potentiel évolutif est un premier estimateur qui renseigne de manière générale sur les capacités d'un organisme à s'adapter a priori. Cependant, il reste inadapté pour travailler à des échelles plus fines : il ne tient pas compte de la diversité génétique qui peut exister entre les populations d'une même espèce, ni de la taille de ces populations. Ces deux variables peuvent influer sur la fréquence initiale des virulents dans les populations, donc sur l'efficacité de la résistance et sur la dynamique de l'adaptation, donc sur la durabilité de l'utilisation de celle-ci. De même, il ne tient pas compte du déterminisme moléculaire de la virulence (nombre et type de mutations nécessaires à l'acquisition de la virulence), différent vis-à-vis de chaque résistance et influençant également fortement la durabilité d'une résistance. Être capable d'identifier les déterminismes de la virulence et *a fortiori* de quantifier les virulents dans chaque population permettrait donc d'évaluer, de manière beaucoup plus fine, l'efficacité initiale de la variété résistante et le potentiel de la population pathogène à s'adapter. Pour gagner en précision dans cette estimation, en particulier sur la dynamique de l'adaptation (*i.e.* la durabilité), la connaissance des coûts et bénéfices associés à la sélection pour la virulence est également indispensable.

Identifier les déterminants génétiques de l'adaptation et les conséquences de cette adaptation sur la *fitness* du pathogène sont des étapes majeures vers une meilleure compréhension et prédiction du phénomène de contournement.

Il n'existe aujourd'hui que peu de variétés de pomme de terre commercialisables présentant un fort niveau de résistance face à *Globodera pallida*. Un des défis actuels est donc de minimiser la perte potentielle d'efficacité due à l'utilisation répétée d'une même variété résistante. Le QTL *GpaV_{vrn}* est un facteur de résistance efficace qui contrôle la quasi-totalité des populations européennes de *G. pallida*. Néanmoins, l'expérience montre que les populations du nématode sont capables de contourner cette résistance. Pour en améliorer la durabilité, il est nécessaire de comprendre et d'anticiper les réponses évolutives des populations de *G. pallida* face à son déploiement.

Dans cet objectif, ce projet de thèse visait à améliorer les connaissances sur les déterminants génétiques et les conséquences phénotypiques de l'adaptation de *G. pallida* aux résistances de sa plante hôte. Ce travail a été réalisé sur des lignées expérimentales de *G. pallida* ayant été confrontées à des variétés sensibles et résistantes de pomme de terre. Il s'est scindé en deux axes principaux : (i) dans le premier axe, nous avons étudié les traits d'histoire de vie de *G. pallida* impactés par l'adaptation à la résistance de son hôte afin de tester l'existence éventuelle d'un coût de virulence ; (ii) dans le second axe, nous avons cherché à identifier les régions génomiques impliquées dans l'adaptation, par une approche de scan génomique en utilisant différentes densités de marqueurs.

Le manuscrit se structure en trois chapitres qui contiennent chacun un article scientifique et qui correspondent aux objectifs spécifiques de la thèse :

- 1. Caractériser les traits d'histoire de vie impactés par l'adaptation à la résistance issue de *S. vernei* et déterminer si celle-ci est associée à un coût de *fitness* pour le nématode.
- 2. Vérifier, en utilisant des marqueurs microsatellites, la faisabilité d'une approche de scan génomique sur un matériel provenant d'une évolution expérimentale et ayant subi un processus court de divergence.
- 3. Identifier les régions génomiques impliquées dans l'adaptation de *G. pallida* à la résistance issue de *S. vernei* en utilisant des marqueurs SNPs issus du reséquençage de génome complet de pools d'individus. Dans ce chapitre, nous avons également testé la possibilité d'utiliser une plus faible densité de SNPs (obtenus par GBS) pour réaliser les scans génomiques.

1.5 Ressources génomiques acquises

Dans l'objectif d'identifier les déterminants génétiques de l'adaptation de *G. pallida* à la résistance de son hôte, différentes ressources génomiques ont été produites à partir du même matériel biologique de départ, issu de l'évolution expérimentale présentée précédemment. Elles ont été reprises dans le tableau ci-dessous (**Tableau 1.2**), dans l'ordre

dans lequel elles ont été acquises. Ce tableau reflète le cheminement de notre raisonnement et les opportunités qui se sont offertes à nous au long du projet. Le contexte d'acquisition de ces données sera rappelé au début de chaque chapitre.

Tableau 1.2 Ressources génomiques acquises au cours de la thèse. Les croix grisées dans la colonne du génotypage avec marqueurs microsatellites correspondent aux lignées qui ont fait partie du matériel de départ mais dont l'analyse des données de génotypage n'a pas pu aboutir.

Lignées	Génotypage avec 202 marqueurs microsatellites	Pool-Seq (génome complet)	GBS			
	≈ 50 individus	Pools de 300 J2 x2	Pools de 300 J2 x2	Pools de 40 kystes		
SMD-8x	х	x	х	Х		
SMI-8x	Х	Х	х	х		
SM360-8x	Х	-	-	х		
SM94T-7xD	Х	-	-	х		
SMI-7xD	Х	-	-	х		
ND-8x	Х	Х	х	х		
NI-8x	х	Х	х	Х		
NI-7xD	Х	-	-	х		
Résultats	Chapitre 3	Chapitre 4	Chap	itre 4		

CHAPITRE 2

CONSÉQUENCES PHÉNOTYPIQUES DE L'ADAPTATION DE *GLOBODERA PALLIDA* **AUX RÉSISTANCES DE SA PLANTE HÔTE**

CHAPITRE 2 CONSÉQUENCES PHÉNOTYPIQUES DE L'ADAPTATION DE *GLOBODERA PALLIDA* AUX RÉSISTANCES DE SA PLANTE HÔTE

Le contournement rapide de la résistance issue de *S. vernei* constaté en évolution expérimentale prédit une faible durabilité de cette résistance au champ. Cependant, certains paramètres peuvent ralentir la dynamique de l'adaptation, comme la taille des populations, les flux de gènes mais aussi l'existence de coûts de *fitness* associés à l'acquisition de la virulence. Leurs actions peuvent allonger substantiellement la durabilité des résistances au champ, comparativement aux conditions utilisées en laboratoire pour l'évaluer. Il est donc important de tenir compte de ces paramètres lors de l'élaboration de stratégies visant à améliorer la durabilité des résistances.

Dans ce chapitre, nous avons ainsi cherché à caractériser les traits d'histoire de vie du nématode impactés par l'adaptation à la résistance issue de *S. vernei* et à déduire l'existence ou non d'un coût de virulence associé au contournement de la résistance. Pour répondre à cet objectif, nous avons mesuré et comparé plusieurs traits d'histoire de vie exprimés sur hôte sensible par des individus virulents et avirulents, obtenus par évolution expérimentale d'une même population naturelle de *G. pallida* sur génotypes de pomme de terre résistant (cv. lledher) et sensible (cv. Désirée). Ces travaux présentent l'originalité de tester, au même moment et donc dans les mêmes conditions, des individus issus de 6 à 10 cycles de sélection sur ce matériel végétal.

Les résultats de ces travaux font l'objet d'un article intitulé « Adaptation to resistant hosts increases fitness on susceptible hosts in the plant parasitic nematode *Globodera pallida* » publié dans la revue *Ecology and Evolution*.

ADAPTATION TO RESISTANT HOSTS INCREASES FITNESS ON SUSCEPTIBLE HOSTS IN THE PLANT PARASITIC NEMATODE *GLOBODERA PALLIDA*

Sylvain FOURNET*, Delphine EOCHE-BOSY, Lionel RENAULT, Frédéric M. HAMELIN and Josselin MONTARRY

INRA, UMR1349 IGEPP (Institute for Genetics, Environment and Plant Protection), F-35653 Le Rheu, France. *Correspondence: Sylvain Fournet, INRA, UMR1349 IGEPP (Institute for Genetics, Environment and Plant Protection), F-35653 Le Rheu, France. Tel: +33 223 485 158; Fax: +33 223 485 150; E-mail: sylvain.fournet@rennes.inra.fr

Running title: A fitness benefit to be virulent in G. pallida

Trade-offs between virulence (defined as the ability to infect a resistant host) and life-history traits are of particular interest in plant pathogens for durable management of plant resistances. Adaptation to plant resistances (*i.e.*, virulence acquisition) is indeed expected to be associated with a fitness cost on susceptible hosts. Here, we investigated whether life-history traits involved in the fitness of the potato cyst nematode *Globodera pallida* are affected in a virulent lineage compared to an avirulent one. Both lineages were obtained from the same natural population through experimental evolution on resistant and susceptible hosts, respectively. Unexpectedly, we found that virulent lineages were more fit than avirulent lineages on susceptible hosts: they produced bigger cysts, containing more larvae and hatching faster. We thus discuss possible reasons explaining why virulence did not spread into natural *G. pallida* populations.

Keywords: experimental evolution, fitness benefit, life-history traits, selection, trade-off, unnecessary virulence.

Author contributions: SF performed the experimental evolution. SF, DEB, LR and JM performed the experiments according to a protocol elaborated jointly between SF and JM. SF, DEB, FMH and JM analyzed the data, wrote the text and prepared the figures.

Published in *Ecology and Evolution* (2016), vol. 6, pp. 2559-2568

2.1 Introduction

The existence of trade-offs between life-history traits is a central concept in evolutionary biology and ecology because they may result in the wide range of species diversity that can be observed in nature. The trade-off concept is based on a differential allocation of a limited resource to two different traits: maximizing one leads to reduced performance in the other (Stearns 1989). This concept provides an intuitive framework to explain why adaptation of populations to a new environment may be costly: selection of a highly beneficial trait in a new environment will be coupled to a cost in other trait(s).

Host-parasite interactions are particularly well suited to address experimentally the research questions relative to adaptation to a new environment, *that is* the host. Indeed, the increasing restrictions on pesticide use lead to the use of crop cultivars resistant to pathogens which represent an efficient, specific and environment friendly alternative strategy to control crop diseases. In this context, the overcoming of host specific resistance and the acquisition by the parasite of virulence, that is the ability to infect and multiply in host genotypes resistant to other parasite genotypes (Vanderplank 1963; Gandon & Michalakis 2002; Tellier & Brown 2007, 2009), represents an interesting case of host adaptation. Developing a sustainable use of plant resistances therefore requires exploring in detail the consequences of this adaptation on life-history traits strongly involved in parasite fitness that can be in trade-off with virulence.

Plant–parasite interactions often fit the widely accepted gene-for-gene model, which predicts that successful disease resistance is triggered only if a resistance gene product in the host recognizes, directly or indirectly, a specific avirulence gene product from the pathogen (Flor 1956, 1971; Jones & Dangl 2006; Brown & Tellier 2011). In such systems, a mutation from avirulence to virulence will increase pathogen fitness if the host resistant gene is present; many recorded examples of such directional selection, leading to the invasion of pathogen populations by virulent isolates, and consequently to resistance breakdown, have been reported (*e.g.*, Vanderplank 1963; Wolfe 1993; Castagnone-Sereno 2002; Molinari 2011; Brown 2015). Conversely, mutation to virulence can also involve strong fitness penalties in the absence of resistant hosts, and theoretical analyses of coevolution in gene-for-gene systems have shown that costs of virulence are necessary to maintain stable resistance-avirulence polymorphisms (Brown & Tellier 2011).

Many efforts have been devoted to the experimental detection and quantification of virulence costs, and results are broadly in accordance with the predictions of a tradeoff model. Definitive evidence for resistance-breaking costs has been reported, either in single infection or in competition, for several plant-virus systems (Ayme et al. 2006; Sacristan & Garcia-Arenal 2008; Fraile & Garcia-Arenal 2010; Janzac et al. 2010; Poulicard et al. 2010, 2012; Fraile et al. 2011; Ishibashi et al. 2012; Montarry et al. 2012), bacteria (Leach et al. 2001) and fungi or oomycetes (Thrall & Burdon 2003; Abang et al. 2006; Huang et al. 2006; Bahri et al. 2009; Montarry et al. 2010). Plant parasitic nematodes are no exception to this rule. A reproductive fitness cost associated with virulence against the tomato *Mi-1* resistance gene and against the pepper Me3 and Me7 resistance genes has been detected in the root-knot nematode Meloidogyne incognita (Castagnone- Sereno et al. 2007; Djian-Caporalino et al. 2011), but there are no data regarding plant parasitic cyst nematodes. However, the overcoming of the resistance of potato cultivar lledher, the first potato cultivar resistant to the cyst nematode Globodera pallida registered in the French catalogue, has been recently observed in an experimental evolution study (Fournet et al. 2013). The aim of the present work was to determine whether fitness costs of virulence are associated with this resistance breakdown,

by studying life-history traits on a susceptible host in a virulent/avirulent pair of lineages from a single *G. pallida* population. Both lineages were obtained through recurrent selection on a susceptible potato (cv. Désirée) for the avirulent one, and on a resistant potato (cv. Iledher) for the virulent one. To account for the adaptation dynamics, life-history traits were studied in both lineages at the same time, from the sixth to the tenth generations.

2.2 Methods

2.2.1 Biology of Globodera pallida

Globodera pallida is a cyst nematode performing one generation for a vegetation season (Subbotin *et al.* 2010), that is only one generation per year in European climatic conditions. As other cyst nematodes, it enters plant roots as second-stage juveniles (J2) and establishes a specialized feeding structure, the syncytium (Jones & Northcote 1972), which is a severe nutrient sink for the plant. Larvae mature and adult males then exit the roots in order to mate with females. After mating, the females continue to feed from the syncytium. When egg development is completed, and after an early moult of the first-stage juveniles (J1) within eggs, females die and form a cyst, which protects the eggs until infective J2 hatch. The cyst, that is the body of the female, is the only structure which was transmitted from one season to the next, males acting only as gamete-like propagules which do not survive.

2.2.2 Selection of virulent and avirulent Globodera pallida lineages

Virulent and avirulent nematode lineages were established, as described by Fournet et al. (2013), starting from cysts of a French natural G. pallida population collected from an infested field (3 J2 per gram of soil) located near Saint-Malo (Brittany, north-western France), which has never been cultivated with any potato cultivar resistant to G. pallida. The lineages were obtained by rearing the nematode population during ten successive generations on either the resistant potato cultivar lledher or the susceptible cultivar Desiree. Briefly, the first three generations were made directly into the field, the next two generations were made into 20 L tanks under greenhouse conditions and the five last generations into pots inoculated with 10 cysts coming from the previous cycle (see Fournet et al. 2013). The resistance of cv. lledher depends on a major QTL, introgressed from Solanum vernei, explaining a large part (at least 61% – Rouppe van der Voort et al. 2000) of the genetic variance of segregating progenies and responsible for the development of most nematodes into adult males. This QTL was mapped on potato chromosome V and is designated as the GpaV_{vrn} locus. The nematode lineages used here were Gp-Désirée-6x, Gp-Désirée-8x, Gp-Désirée-10x and Gp-Iledher-6x, Gp-Iledher-8x, Gp-Iledher-10x (x being the number of successive generations on the corresponding potato cultivar).

2.2.3 Virulence assessment

Response to selection was measured by estimating the virulence level, defined as the probability for a larva to produce a female, in each lineage. The virulence of a given lineage

was estimated as described by Fournet *et al.* (2013). Briefly, pieces of germinated tubers were allowed to emit roots in a petri dish containing 20 g.L⁻¹ water agar, and one root was inoculated with ten newly hatched larvae coming from a pool of ten cysts placed into potato root exudates. After inoculation, petri dishes were kept in total darkness, under controlled conditions ($20 \pm 1^{\circ}$ C), for 15 days. Every inoculated root was then dissected in water under a transillumination unit for brightfield/darkfield contrasts, and the number of developed females was counted. Virulence level was then calculated for each lineage as the ratio between the number of females and inoculated larvae. Virulence level of each nematode lineage was assessed in this way on the resistant potato cultivar lledher and compared for each generation to the corresponding percentage of females produced by the avirulent lineages (Gp-Désirée-6x, Gp-Désirée-8x, Gp-Désirée-10x) on the susceptible potato cultivar Désirée. This comparison allowed us to define the virulence status of each lineage ("vir" for virulent or "avir" for avirulent; see Fig. 2.1). For each lineage, from 12 to 29 roots (*i.e.*, replicates) were inoculated independently.

2.2.4 Life-history traits assessment

Life-history traits were evaluated on new cysts produced on the susceptible potato cultivar Désirée, under controlled conditions, using larvae issued from the lineages Gp-Désirée-6x, Gp-Iledher-6x, Gp-Désirée-10x and Gp-Iledher-10x. To this end, germinated tubers of the cultivar Désirée were buried in a moistened mixture of sand and soil (1:3) contained in a 150-mL clear plastic screw cup (C.E.B., Angers, France). Capped cups were then stored in the dark at 20°C (± 1) and 70% relative humidity in a climatic chamber until roots were clearly visible through the cup wall. Thereafter, for each lineage, six cups (i.e., six independent replicates) were inoculated with 2 mL of a suspension containing around 150 second-stage juveniles.mL⁻¹. This inoculum suspension was obtained by pooling together the larvae collected every 4 days during a maximum of 12 days from 80 different cysts soaked in potato exudates (cv. Désirée). Freshly hatched larvae were stored at 4°C until the proper inoculum concentration was reached. The final concentration was adjusted by adding the corresponding amount of water and controlled five times by counting all the larvae in 1 mL just before inoculation. Cups were then capped again and stored for 80 days in the climatic chamber, under the same climatic conditions, to allow the complete development and maturation of newly formed cysts.

For each replicate, newly formed cysts were manually collected and numbered (#Cyst) after extraction with a Kort elutriator. The size of each collected cyst (Size_cysts) was evaluated through the surface area using a magnifying stereomicroscope coupled with an image analysis software (Microvision Instruments, Histolab v8.1.0, Evry, France). To evaluate the cysts' contents, ten cysts were randomly picked from each replicate, pooled and crushed in water, and the eggs and larvae were counted (#Larvae) using a magnifying stereomicroscope. For each replicate, that is for each pool of ten cysts, the arc length of ten randomly chosen larvae was measured (Length_larvae) using a magnifying stereomicroscope coupled with an image analysis software (Microvision Instruments, Histolab v8.1.0, Evry, France). To estimate the hatching dynamics, ten other cysts, randomly picked from each replicate, were soaked in potato exudates (cultivar Désirée) in a small plastic cup and then stored in the dark at 20°C (± 1). Every 5 days, the potato exudate was removed and the hatched larvae were counted using a magnifying stereomicroscope. Seventy days after the

start of the experiment, cysts were crushed in water to assess the number of unhatched larvae, so as to calculate the final hatching rate (a proportion) of each replicate. Assuming that the time-tohatching is Weibull-distributed, a simple model (**Data S1**) was adjusted to the hatching curves and two parameters, the hatching probability (P) and the mean hatching time (HT), were estimated for each replicate, through a maximum likelihood approach.

2.2.5 Statistical analysis

All statistical analyses were performed using the R software version 3.1.1 (R Core Team, 2014). Normality and homogeneity of variances were checked using the Shapiro–Wilk and the Leven tests, respectively.

The levels of virulence of each lineage (Gp-Désirée-6x, Gp-Désirée-8x, Gp-Désirée-10x and Gp-Iledher-6x, Gp-Iledher-8x, Gp-Iledher-10x) and of the positive controls (corresponding to the percentage of females produced by the three avirulent lineages, Gp-Désirée-6x, Gp-Désirée-8x, Gp-Désirée-10x, on the susceptible potato cultivar Désirée) were compared with a one-way ANOVA on the percentage of females produced and multiple comparisons of means with the Tukey contrasts test ($\alpha = 0.05$).

The pairwise relationships between the different life-history traits measured (#Cyst, Size_cysts, #Larvae, Length_larvae, P and HT) were separately tested on the virulent and avirulent lineages with the Pearson's product— moment correlation. Significance of the correlations was corrected for multiple comparisons using the p.adjust R function and the false discovery rate (FDR) method developed by Benjamini & Hochberg (1995).

For each life-history trait (#Cyst, Size_cysts, #Larvae, Length_larvae, P and HT), the generation effect, the vir/ avir effect (*i.e.*, the virulent/avirulent status of lineages) and the corresponding two-way interaction effect were tested through an ANOVA. Moreover, for each generation (6x and 10x), the vir/avir effect was tested for the six life-history traits using a one-way ANOVA.

2.3 Results

2.3.1 Virulence level of the different Globodera pallida lineages

The ANOVA model showed a significant lineage effect on the percentage of females produced ($F_{8,161} = 69.56$; P < 0.0001). *Globodera pallida* lineages having evolved on the susceptible potato cultivar Désirée, during either six, eight or ten generations, showed a very low virulence level on the resistant potato cultivar lledher (from 0.5% to 3.5% of females produced – Fig. 2.1). That result highlighted the high efficacy of that resistance, conferred to lledher by the $GpaV_{vrn}$ QTL, and confirmed the avirulent status of those lineages, named hereafter avir(6x), avir(8x) and avir(10x). Conversely, *G. pallida* lineages having evolved on cultivar lledher showed a high virulence level on that cultivar, confirming the virulent status of those lineages, named hereafter vir(6x), vir(8x) and vir(10x). While after six generations, the virulence level reached 38.9% of females produced, indicating that the breakdown of resistance was incomplete (*i.e.*, significantly different from the corresponding control – Fig. 2.1), it reached 62.5% and 58.1% of females produced after eight and ten generations, respectively, indicating that the breakdown was total (*i.e.*, not significantly different from their

controls – Fig. 2.1). Because virulence levels from avir(8x) and vir(8x) on the one hand and avir(10x) and vir(10x) on the other hand were not significantly different; the comparisons of life-history traits between virulent and avirulent lineages were made only using 6x and 10x *G. pallida* lineages in order to perform the fitness cost test.



Figure 2.1 Percentage of females produced for each *Globodera pallida* lineage on the resistant potato cultivar **Iledher and for the positive controls on the susceptible potato cultivar Désirée.** Lineages having evolved on Désirée (Gp-Désirée-6x, Gp-Désirée-8x, Gp-Désirée-10x) and on Iledher (Gp-Iledher-6x, Gp-Iledher-8x, Gp-Iledher-10x) are indicated in grey and black, respectively. White bars correspond to the percentage of females produced by the three avirulent lineages (Gp-Désirée-6x, Gp-Désirée-6x, Gp-Désirée-6x, Gp-Désirée-6x, Gp-Désirée-6x, Gp-Désirée-6x, Gp-Désirée-6x, Gp-Désirée-6x, Gp-Désirée-8x and Gp-Désirée-10x) on Désirée. Letters represent the homogenous groups identified by the Tukey contrasts test at the 5% threshold.

2.3.2 Life-history traits measurements in virulent and avirulent lineages

There was no significant correlation between the different life-history traits measured here, that is number of cysts, size of cysts, number of larvae per cyst, length of larvae, hatching probability and mean hatching time, whatever the status of the *G. pallida* lineages (**Table 2.1A** for the virulent and **Table 2.1B** for the avirulent lineages, respectively). Note that there was also no significant correlation between the different life-history traits when the test was done separately for each lineage (data not shown). Consequently, there was no trade-off between the different life-history traits measured here, and because there was no redundant information, comparisons between virulent and avirulent lineages were relevant for all six life-history traits.

The two-way ANOVAs revealed (1) a small generation effect for the number of cysts (with more cysts produced at the sixth generation than at the tenth generation), (2) a vir/avir effect for the size of cysts and the mean hatching time and (3) no significant interaction effect (Table 2.2). The number of cysts produced on the susceptible cultivar potato Desiree did not

differ significantly between the virulent and the avirulent *G. pallida* lineages (Fig. 2.2A). However, cysts were significantly larger for the virulent than for the avirulent lineages, regardless of the generation considered (Fig. 2.2B). The virulent lineage produced more larvae per cyst than the avirulent lineage at the tenth generation, and there was no effect on that lifehistory trait at the sixth generation (Fig. 2.2C). Those larvae were of similar length in the virulent and the avirulent lineages (Fig. 2.2D). Finally, the mean hatching time was significantly longer in the avirulent than in the virulent lineages (Fig. 2.2F and Data S2), whereas there was no significant difference for the hatching probability, whatever the generation considered (Fig. 2.2E and Data S2).

 Table 2.1 Pearson's correlation coefficients (below the diagonal) and associated adjusted P-values (above the diagonal) between the different life-history traits (#Cyst, Size_cysts, #Larvae, Length_larvae, P and HT) for (A) the virulent and (B) the avirulent *Globodera pallida* lineages.

				Length		
	#Cysts	Size_cysts	#Laivae	larvae	P.	HI
(A)						
#Cysts		0.671	0.780	0.671	0.397	0.813
Size_cysts	-0.27		0.671	0.397	0.671	0.671
#Larvae	-0.11	-0.35		0.780	0.397	0.780
Length_Jarvae	0.26	-0.60	0.12		0.671	0.397
P	-0.49	0.34	0.54	-0.23		0.671
HT	-0.08	-0.29	0.13	0.49	-0.22	
(B)						
#Cysts		0,779	0.514	0.606	0.779	0.281
Size_cysts	-0.19		0.779	0.375	0.281	0.841
#Larvae	0.36	-0.14		0.293	0.386	0.873
Length_larvae	-0.30	0.46	0.55		0.281	0.375
P	0.16	0.63	0.44	0.59		0.779
HT	0.62	-0.09	0.05	-0.47	0.20	

Table 2.2 Result: of these factors 0.001.	s from on th	the analy e six life-h	yses of va history tra	iriance aits (#C	: (ANOVAS Cyst, Size_	() assessing th cysts, #Larvae	ie gen e, Leng	eration e gth_larva	ffect (6x e, P and	and 1 HT). ⊤	0x), the v he statist	ir/avir e cically sig	ffect a	ind the co it effects	rrespond are indica	l ing tv ated a	vo-way in s * <i>P</i> < 0.0	teraction 15, *** <i>P</i> <
	#Cys	sts		Size	_cysts		#Larva	ae		Lengt	th_larvae		۵.			보		
Source of variation	df	<i>F</i> value	P > F	df	<i>F</i> value	P > F	Ъ,	<i>F</i> value	P > F	đ	F value	P > F	£	F value	P > F	¥	<i>F</i> value	P > F
Generation effect	-	5.34	0.032*	-	3.32	0.084	-	0.56	0.461	-	1.22	0.284	-	0.01	0.939	-	1.09	
Vir/avir effect	, -	1.16	0.294	-	20.74	0.0002***	, -	2.98	0.100	-	0.45	0.509	, -	2.51	0.129	, -	16.10	0.0007***
Interaction effect	-	0.80	0.382	-	0.35	0.560	, -	2.00	0.173	-	2.15	0.159	-	0.76	0.393	, -	0.31	0.584
Error	20			20			20			19			20			20		

rom the analyses of variance (ANOVAS) assessing the generation effect (6x and 10x), the vir/avir effect and the corresponding two-way intera	n the six life-history traits (#Cyst, Size_cysts, #Larvae, Length_larvae, P and HT). The statistically significant effects are indicated as *P < 0.05, *	
Table 2.2 Results from the analyse	of these factors on the six life-hist	0.001.



Figure 2.2 Means (± standard error) of (A) the number of cysts (#Cysts), (B) the size of cysts (Size_cysts), (C) the number of larvae (#Larvae), (D) the length of larvae (Length_larvae), (E) the hatching probability (P) and (F) the mean hatching time (HT) for the virulent and the avirulent *Globodera pallida* lineages at each generation (6x and 10x). *F* and *P* are indicated for each comparison and significant vir/avir effects are highlighted in bold.

2.4 Discussion

Evaluating life-history trade-offs associated with unnecessary virulence in plant pathogens is the classical method to demonstrate that selection for virulence could be costly. This point is central to understand the ecology and evolution of plant diseases, and thus to better manage plant resistances on the long run. According to the trade-off concept, resistance-breaking isolates are expected to undergo a fitness cost in susceptible hosts, compared to avirulent isolates. This is true for many plant pathogens such as viruses, oomycetes and fungi (see Laine & Barrès 2013 for a review), but also in the case of rootknot nematodes (Castagnone-Sereno *et al.* 2007; Djian-Caporalino *et al.* 2011). Unexpectedly, in our experimental conditions, we found an apparent fitness benefit for virulent lineages in *G*.

pallida, even on the susceptible host: the virulent lineages produced bigger cysts, containing more larvae (at least for the tenth generation) and hatching more quickly. Generalization of that result could be reinforced using other pairs of virulent and avirulent lineages coming from other natural cyst nematode populations.

The lack of apparent cost related to virulence is a result which might bear several explanations. First, the resistance conferred by GpaVvrn basically alters the quality of the feeding site, and thus acts on a late step of the nematode development cycle (Fournet et al. 2013). Therefore, the traits involved in the early stages of parasitism could be less affected by selection against unnecessary virulence. This seems to be true as the same number of cysts for both lineages was observed, suggesting that virulent larvae exhibit the same abilities as avirulent ones to localize and then penetrate inside the roots, to move within the roots, to build their feeding site and to develop into adult males and females. Accordingly, Castagnone-Sereno et al. (2015) recorded no cost related to unnecessary virulence in life-history traits directly linked to the establishment of the parasitic interaction between tomato and M. incognita. Another possible explanation is that virulence costs were compensated before our first measurement of a potential cost, that is before the sixth generation. For example, Schoustra et al. (2006) showed that fludioxonil-resistant isolates of the ascomycete Aspergillus nidulans suffered a severe fitness cost in the absence of the drug, but that this cost was afterwards waived without loss of the resistance. In the contrary, Janzac et al. (2010) performed a serial passage experiment with potato virus Y in pepper and observed no compensatory mutation showing that the only possibility to compensate a high virulence cost was through the reversion from virulence to avirulence. Because no G. pallida cyst was sampled before the sixth generation, we cannot actually test the hypothesis of an early cost compensation.

Finding that a reproductive fitness benefit (*i.e.*, an increase in the ability of virulent individuals to produce a progeny) is associated with unnecessary virulence is much more surprising than the apparent absence of a cost only. To our knowledge, such an unexpected result has been previously recorded only twice, in the wheat pathogen *Puccinia striiformis* f. sp. *tritici*, for which the *vir9* virulence factor conferred a competitive advantage in the absence of the corresponding resistance gene (Bahri *et al.* 2009), and for one life-history trait of the potato late blight pathogen, *Phytopththora infestans*, where virulent isolates had a shorter latency period than avirulent ones (Montarry *et al.* 2010). However, for the former, this was not true for two other virulence factors (*vir4* and *vir6*), and for the latter this advantage was counterbalanced by costs for other life-history traits leading to an overall fitness lower for virulent isolates compared to avirulent ones on susceptible hosts.

The positive relationship between female body size and fecundity is a well-known correlation, particularly in ectotherms: larger females will produce more eggs (*e.g.*, Reiss 1989; Roff 1992), and this has already been highlighted in nematodes (Morand 1996). In our study, virulent females had a larger body size at the sixth and tenth generations (compared to the avirulent lineage). This may be the consequence of an adaptation to the difficulty to build high-quality feeding sites in the resistant cultivar during the experimental evolution. This may correspond to the "thrifty phenotype" hypothesis developed by Hales & Baker (2001) which asserts that nutrient stress during early development leads to an adapted physiological state that increases fitness in conditions of food scarcity. Such theory was recently confirmed in the free-living nematode *Caenorhabditis elegans* where the progeny and grand progeny of starved larvae are more resistant to starvation (Jobson *et al.* 2015). Accordingly, one can imagine that this adaptation may lead to a more efficient use of the amount of available resources in

susceptible host plants by virulent females, leading to larger body sizes and offspring number, which represents an obvious fitness gain.

The offspring of virulent females moreover hatches faster than the offspring of avirulent females. Here, it is less easy to determine whether it can give a fitness advantage or not as *G. pallida* achieves only one generation per year. Nevertheless, one can imagine that hatching more quickly allows the larvae to penetrate in better quality roots (*i.e.*, not parasitized by a large number of competitive larvae), and should enhance their development success. Moreover, in the case of mixed populations where virulent and avirulent larvae compete, such hatching differences could promote mating between virulent females and virulent males.

If the present work attested that the adaptation to the GpaV_{vrn} QTL involves no cost of virulence but rather provides a fitness advantage, it remains unclear why that virulence mutation did not spread into natural G. pallida populations. This question is particularly relevant as the most parsimonious hypothesis is that the mutation has to be already present in the population to allow adaptation to resistance. Indeed, the fact that some nematode lineages were lost during the experimental evolution study performed by Fournet et al. (2013) constitutes an argument that few individuals carrying the virulence alleles are present in natural populations. Even if the virulence seems to be present in some populations and advantageous, the system is probably more complex in natural conditions than in our laboratory experiments. Indeed, gene flow, which was zero during our experimental evolution experiment but often strong in agro-ecosystems due to passive transport of cysts through human activities (Picard et al. 2004; Plantard & Porte 2004; Alenda et al. 2014), and small effective population size (Jan et al. 2016), which favours the random action of genetic drift, can both strongly reduce the frequency of virulence alleles in natural populations. Another possible explanation could be that trade-offs between virulence and fitness traits exist in the field situation but were overshadowed in our experimental situation. Moreover, the competition between virulent and avirulent individuals which takes place in natural populations could lead to the emergence of virulence costs on traits that have not been studied here (e.g., survival abilities when the host is absent). Several experiments on fungi and viruses have studied such a competitiveness cost by performing competition experiments involving infections with mixture of virulent and avirulent isolates (Abang et al. 2006; Bahri et al. 2009; Janzac et al. 2010) and highlighted in most cases a competitiveness cost leading to the counter-selection of virulent individuals on the susceptible host.

To complement our results, further competition experiments should be performed by studying the evolution of frequencies of *G. pallida* virulent and avirulent variants on a susceptible potato cultivar. However, this kind of experiment requires genetic markers to follow the virulent/ avirulent alleles. Such markers are not yet available for *G. pallida*, but studies are currently performed to develop them. Elucidating the genetic determinism of virulence will also allow insight into whether the gene(s) involved is(are) present in multicopy, which could explain the absence of cost, as a mutated copy (virulent) could be associated with a nonmutated copy (avirulent) which could assure the initial function of the gene. Obviously, the use of different *G. pallida* populations, selected on different genotypes carrying the *GpaVvrn* QTL, should allow our conclusions to be extended and is of a crucial importance to address sustainable management of resistance.

Additional Supporting Information may be found in the Annexes section at the end of the thesis manuscript.

2.5 Acknowledgments

Seed tubers of potato cultivar lledher were kindly provided by the potato breeding company Grocep. We gratefully acknowledge Didier Andrivon and Eric Grenier for comments on earlier draft of this manuscript and Samuel Soubeyrand for feedback on modelling the hatching dynamics. We thank the anonymous reviewers for their helpfull suggestions. FMH acknowledges partial funding from the French National Research Agency (ANR) as part of the "Blanc 2013" program (ANR-13-BSV7-0011, Fun- Fit project), and from the French National Institute for Agricultural Research (INRA) "Plant Health and Environment" Division.

2.6 Data archiving

All data used in this article are available at datadryad.org (http://dx.doi.org/10.5061/ dryad.j9j02).

CHAPITRE 3

APPROCHE DE SCAN GÉNOMIQUE SUR DES LIGNÉES DE *GLOBODERA PALLIDA* **ISSUES D'ÉVOLUTION EXPÉRIMENTALE**

CHAPITRE 3 APPROCHE DE SCAN GÉNOMIQUE SUR DES LIGNÉES DE *Globodera pallida* issues d'évolution EXPÉRIMENTALE

L'adaptation de G. pallida à la résistance issue de S. vernei a été caractérisée au niveau phénotypique dans le chapitre précédent. L'évolution expérimentale a permis de mettre en évidence la capacité des populations de G. pallida à contourner rapidement la résistance sans pour autant subir de coût de *fitness*. Au contraire, nous avons montré que le contournement de la résistance était associé à un gain de *fitness* des individus virulents sur hôte sensible. Ces individus produisent des kystes plus gros, contenant plus de larves qui éclosent plus rapidement, que les individus avirulents. Ce constat renforce l'hypothèse d'une faible durabilité de cette source de résistance au champ. Cependant, la fitness a ici été seulement mesurée en inoculation simple. Il n'est pourtant pas improbable qu'un coût de compétitivité puisse s'exprimer chez les individus virulents s'ils sont en compétition avec des individus avirulents sur une plante hôte sensible. Tester l'existence de ce coût de compétitivité nécessite l'utilisation de marqueurs moléculaires permettant de suivre l'évolution des allèles de virulence et d'avirulence dans la population. Ces marqueurs moléculaires ne sont actuellement pas disponibles. Ces résultats appellent donc à franchir un nouveau cap dans l'étude du contournement de la résistance en nous intéressant aux bases génétiques de l'adaptation. Outre une meilleure description du coût potentiel associé au contournement d'une résistance, identifier les déterminants génétiques de la virulence permettrait une meilleure compréhension et anticipation du processus de contournement, en caractérisant le nombre et le type de gènes et mutations impliqués dans l'adaptation, paramètres jouant un rôle clé dans la durabilité d'une résistance, et en accédant à des marqueurs moléculaires permettant d'estimer et de suivre l'évolution des fréquences des allèles de virulence au champ. L'estimation du potentiel de contournement de chaque population vis-à-vis de la source de résistance étudiée serait alors rendue possible par simple génotypage.

Jusqu'à présent, très peu de données concernant les gènes d'avirulence de nématode existent. Dans le but d'identifier les bases génétiques de l'adaptation à la résistance issue de S. vernei, nous avons donc retenu l'utilisation d'une approche sans a priori de scan génomique sur les lignées de G. pallida issues d'évolution expérimentale. Néanmoins, nous n'étions pas en mesure de savoir si le matériel biologique disponible, issu de quelques générations d'évolution expérimentale, avait bénéficié d'un brassage génétique suffisant pour aboutir à un scan génomique précis, c'est-à-dire permettant l'identification de quelques régions génomiques de taille restreinte uniquement. Dans ce chapitre, nous avons donc exploré cette question en réalisant un premier scan génomique qualifié de faible densité, en génotypant à l'aide de 53 marqueurs microsatellites, polymorphes et distribués dans le génome, des individus issus des lignées expérimentales virulentes et avirulentes à la résistance issue de S. vernei, provenant de deux populations indépendantes, et en utilisant trois tests permettant d'identifier des régions génomiques potentiellement soumises à l'action de la sélection. Les données de génotypage nous ont également permis de caractériser notre matériel biologique, par l'étude de la diversité génétique au sein des lignées et de la divergence entre lignées et entre populations, autant d'informations importantes à connaître dans le cadre de la réalisation d'un scan génomique et dans l'objectif d'une interprétation optimale des résultats de ce scan.

Les résultats de ces travaux font l'objet d'un article intitulé « Experimentally evolved populations of the potato cyst nematode *Globodera pallida* allow the targeting of genomic footprints of selection due to host adaptation » publié dans la revue *Plant Pathology*.

EXPERIMENTALLY EVOLVED POPULATIONS OF THE POTATO CYST NEMATODE GLOBODERA PALLIDA ALLOW THE TARGETING OF GENOMIC FOOTPRINTS OF SELECTION DUE TO HOST ADAPTATION

Delphine EOCHE-BOSY, Jérémy GAUTHIER¹, Amandine S. JUHEL², Magali ESQUIBET, Sylvain FOURNET, Eric GRENIER and Josselin MONTARRY*

IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, Le Rheu 35650, France. *Correspondence: Josselin Montarry, IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, Le Rheu 35650, France. Tel: +33 2 23 48 51 73; Fax: +33 2 23 48 51 50; E-mail: josselin.montarry@inra.fr

Running title: Footprints of selection in G. pallida

In the current agronomical context of pesticide use reduction, deciphering the genetic bases of pathogen adaptation to plant defences is of major importance to improve durability of resistance. Indeed, knowledge of virulence gene frequencies in pathogen populations could allow the prediction of resistance durability before deployment. Globodera pallida is a major pest of potato crops for which a promising resistance QTL, GpaVvrn, has been identified in Solanum vernei. An experimental evolution study, in which G. pallida lineages evolved on resistant or susceptible potato genotypes for up to eight generations, previously showed that G. pallida was able to rapidly overcome GpaVvrn resistance. However, it was not known if enough genetic mixing occurred in these lineages to be able to detect islands of differentiation in a genome scan approach. Here, this question was investigated using 53 polymorphic microsatellite markers distributed along the genome and three different tests based on genetic differentiation and heterozygosity. Eight outlier loci were identified, indicative of genomic regions putatively involved in host adaptation. Several loci were identified by multiple detection methods and/or in two independent adapted lineages. Some candidate genomic regions identified also seemed to be involved in overcoming resistance to nematodes in a plant genotype harbouring the same resistance QTL in a different genetic background. These results validate the feasibility of a genome scan approach on biological material coming from short experimental evolution, and encourage the use of a high coverage genome scan by whole genome resequencing.

Keywords: experimental evolution, genome scan, outlier, resistance, virulence.

Author contributions: SF performed the experimental evolution. DEB, JG, ASJ and ME performed the experiments according to a protocol elaborated jointly by JM, SF and EG. DEB, JG, ASJ and JM analyzed the data. DEB and JM wrote the text and prepared the figures.

Published in *Plant Pathology* (2017), vol. 66, pp. 1022-1030

¹present address: Institut de Recherche sur la Biologie de l'Insecte, UMR CNRS 7261, Université François-Rabelais, 37200 Tours, France.

²present address: UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France.

3.1 Introduction

Host-parasite interactions constitute an interesting system in which biological adaptation to a new environment, *i.e.* the host, can be studied. The identification of the molecular basis of host adaptation in parasites has become a central challenge for human and environmental health (Aguileta *et al.* 2009); this is also true for plant parasites, responsible for severe damage in agricultural crops worldwide.

A comprehensive approach to analyse the molecular bases of adaptation encompasses identification of genomic regions showing footprints of selection, followed by more precise investigation of these regions by searching for candidate genes potentially involved in the adaptation. To this end, population genomics has developed methods based on genome-wide variation, called genome scans, to separate locus-specific effects from genomewide effects (Luikart et al. 2003; Storz 2005). Demographic processes affect the genome in its entirety, whereas selection effects are expected to be locus-specific. Therefore, loci potentially involved in selection show a non-neutral pattern of variation, by exhibiting significantly more differentiation among populations and lower diversity within populations (Luikart et al. 2003; Storz 2005). Due to the process of genetic hitchhiking (Maynard-Smith & Haigh 1974), markers linked to the selected loci will also be influenced by selection, leaving a genomic signature that can be detected by genotyping markers distributed along the genome (Luikart et al. 2003). The comparison between the distributions of summary statistics estimated from simulations under the null hypothesis of neutrality and from the dataset allows the detection of outlier loci (i.e. loci departing significantly from the neutral distribution). In natural populations, confounding environmental effects could lead to false positive outliers. Coupling genome scans and experimental evolution is therefore a promising approach, as it allows the control of selection pressure imposed on populations so that a direct link can be established between this selection pressure and the phenotypic and genetic changes observed (Turner *et al.* 2011). Repeating the search for outliers in different populations allows the strengthening of the presumed involvement in adaptation of outliers found several times (Nosil et al. 2008). Compared to the genome scan approach conducted on natural populations, fewer independent replicates should be necessary when using experimental lineages to target the genomic regions impacted by selection pressure; however, sufficient genetic mixing should also occur during the experimental evolution process (Storz 2005).

Plant-parasitic nematodes are major agricultural pathogens causing severe damage to crops leading to considerable economic losses (Nicol *et al.* 2011). The cyst nematode *Globodera pallida* is an important pathogen of potato and is classified as a quarantine pest on both EPPO and USDA quarantine lists. *Globodera pallida* is a diploid organism with obligate sexual reproduction, which achieves one generation per year in European climatic conditions (Jones 1950). The use of nematicides has enabled the control of these pathogens for decades. However, their use is now strictly limited or even prohibited, as they constitute a threat to environmental and human health (Clayton *et al.* 2008). The currently favoured alternative for control of *G. pallida* populations is the use of resistant potato cultivars. Because the breeding of resistant potato genotypes is a long process, only a few registered cultivars show a high level of resistance to *G. pallida*. A promising resistance factor, the QTL *GpaVvrn*, stimulates the development of most nematodes into adult males, which do not cause symptoms by themselves. This QTL has been introgressed from *Solanum vernei* (a wild species close to the cultivated potato *S. tuberosum*) into several potato genotypes with different genetic backgrounds. The resistant parents all harboured QTL *GpaVvrn*, which mapped onto potato

chromosome V and explained 61% of the resistance in biparental segregating populations (Rouppe van der Voort *et al.* 2000); in addition, they possibly carried a minor QTL, mapped onto chromosome IX and explaining 24% of the resistance (Rouppe van der Voort *et al.* 2000) but the presence of this QTL was not confirmed (Fournet *et al.* 2013). One of the resistant genotypes, cultivar lledher, was registered, in 2009 in the French official catalogue (plantdepommedeterre.org), as the first cultivar showing a high level of resistance to *G. pallida*. In an experimental evolution study, Fournet *et al.* (2013) exposed two independent French *G. pallida* populations to several potato genotypes, harbouring *GpaV_{vrn}* but differing by their genetic background, over eight parasite generations. They demonstrated that the resistance from *S. vernei* could be overcome by both *G. pallida* populations, and thus generated lineages adapted to each resistant potato genotype. Moreover, they showed that the plant genetic background impacts the durability of resistance.

In order to design strategies for the sustainable management of plant resistance, it is necessary to study the evolutionary potential of pathogen populations, including the genetic basis of their adaptation to resistant cultivars. However, an important limitation lies in the scarcity of molecular data available about the response of *G. pallida* to selective pressures imposed by host genotypes. Only one virulence/avirulence gene has been described in *G. pallida*, and it is specific to a particular nematode population (Sacco *et al.* 2009). Recently, a study, based on a whole genome allelic imbalance analysis of single nucleotide polymorphisms (SNPs) in populations of *Heterodera glycines* grown on resistant and susceptible soybean plants, revealed two new candidate virulence genes (Bekal *et al.* 2015); this has opened the way for a population genomic approach to analyse cyst nematodes. Bekal *et al.* (2015) carried out controlled matings between *H. glycines* inbred lines that had been maintained in the laboratory for over 30 generations.

The aim of the present study was to validate the feasibility of a genome scan approach on biological material coming from short experimental evolution. Thus, a genome scan was performed on *G. pallida* lineages, virulent and avirulent to Iledher, coming from two different populations, using 53 polymorphic microsatellite markers distributed along the genome. Three different methods were applied to identify outlier loci. In addition, candidate genomic regions identified using lineages adapted to Iledher were compared to another potato genotype harbouring the same resistance QTL in a different genetic background.

3.2 Materials and methods

3.2.1 Selection of virulent and avirulent lineages of G. pallida

Nematode lineages were established from cysts of two natural French populations of *G. pallida*, SM (from near Saint-Malo, Brittany, northwestern France) and N (from the island of Noirmoutier, western France), coming from infested fields (3 and 82 J2 per gram of soil for SM and N, respectively). All details about the experimental evolution protocol are given in Fournet *et al.* (2013). The lineages used here had been obtained in the previous study by rearing both populations for eight successive cycles on the susceptible potato cultivar Désirée (D) and on two resistant cultivars: lledher (I) and the potato genotype 360.96.21 (360), both possessing the resistance QTL *GpaVvrn*. Because of the high level of resistance of the potato genotypes used during the experimental evolution, the first three generations were grown directly in the fields, the next two generations were grown in tanks under greenhouse

conditions, and the three last generations in pots. Unfortunately, the intermediate generations as well as several lineages, including the lineage from the N population evolving on the potato genotype 360.96.21, were no longer available. Thus, the present study was conducted with the eighth generation of the five remaining available lineages, named SMD, SMI, ND, NI and SM360. As a consequence, the comparison for adaptation to the same resistance QTL in different genetic backgrounds (Iledher and genotype 360.96.21) could only be performed with lineages from the SM population.

3.2.2 DNA extraction and microsatellite genotyping

For each lineage, 50 cysts from the eighth generation of the experimental evolution protocol were randomly sampled and individualized. As the amount of DNA in a single larva was not sufficient to perform genotyping with the whole set of microsatellites markers, three J2 (siblings) per cyst were used and each one was genotyped with a subset of markers (*i.e.* the DNA from the three siblings was not pooled). Extraction of genomic DNA of each J2 was performed following Boucher et al. (2013). Presence of DNA and species identity of these samples were checked by PCR-RFLP of the ITS region as previously described (Boucher et al. 2013). Samples were genotyped using 202 microsatellite markers. The majority of these markers were developed directly from one of the first versions of the G. pallida genome, composed of 17 095 supercontigs (version gpal.201011.supercontigs.fasta; Sanger Institute, http://sanger.ac. uk). The QDD software (Meglécz et al. 2010) was used to select sequences and design primer pairs for sequences containing perfect microsatellite motifs with at least four repeats and flanking regions without tandem repetition. Among those markers, small PCR products (<100 nucleotides) were discarded, and only one marker per supercontig was retained. Eight supercontig sequences containing pathogenicity genes, *i.e.* genes involved in the production of molecules interacting with the host plant (PL1, PL2, ExpB1, ExpB2 and four paralogues of RBP1; Haegeman et al. 2012) and 3446 ESTs sequences (available on NEMBASE4: http://www.nematodes.org/nembase4/) were also screened for microsatellite content using the software MSATCOMMANDER v. 0.8.2 (Faircloth 2008), which was also used to design primers. Eleven microsatellite markers developed for G. rostochiensis, which crosshybridizes with G. pallida (Boucher et al. 2013), were also selected. To avoid unexpected hybridization, selected primers were tested on the whole G. pallida genome by e-PCR (Rotmistrovsky et al. 2004), and only primer pairs that cannot hybridize on another part of the genome (allowing up to three mismatches) were retained. Overall, a total of 191 microsatellite markers were selected, to which were added 11 markers previously developed by Montarry et al. (2015).

The 202 microsatellite markers were each amplified in a 5 μ L reaction mix containing 2 μ L of genomic DNA, previously diluted 1 in 3, 0.125 U *Taq* DNA polymerase, PCR buffer at 1 x final concentration, 2 mM MgCl₂, 0.2 mM each dNTP (all products from Promega), 0.25 μ M of forward and reverse primers for each marker, and 0.25 μ M of a universal primer (M13) labelled with either 6-FAM, VIC, NED, or PET fluorescent dyes (5'-dye-GTGCTGCAACATTTTGCTG-3'; Applied Biosystems). Amplification steps were an initial denaturation at 94 °C for 5 min; 20 cycles consisting of a denaturation step at 94 °C for 30 s, an annealing step at 57 °C for 90 s and an elongation step at 72 °C for 30 s; then 20 cycles of 94 °C for 20 s, 53 °C for 30 s and 72 °C for 30 s; and a final elongation at 60 °C for 30 min. Amplifications were conducted on C1000 thermal cyclers (Bio-Rad). After amplification, the PCR products from eight loci were each

added to 5 μ L H₂O (2 μ L of PCR products with 6-FAM, VIC and NED fluorescent dyes and 4 μ L of PCR products with PET fluorescent dyes); four pairs of loci were each labelled with a different dye and loci of the same pair (labelled with the same dye) had nonoverlapping amplicon sizes. Two microlitres of this dilution were then added to 4.95 μ L HiDi formamide (Applied Biosystems) and 0.05 μ L GeneScan 500 LIZ size standard (Applied Biosystems). Electrophoresis of amplified fragments was carried out in a capillary sequencer ABI PRISM 3130xl (Applied Biosystems). For each marker, peaks of the different runs were automatically binned using GENEMAPPER v. 4.1 (Applied Biosystems) and the allele assignation was additionally proofread manually.

After genotyping, all monomorphic microsatellite markers and those with too many missing data were removed. The final dataset consisted of 53 polymorphic markers (13 di-, 14 tri-, 21 tetra-, two penta- and three hexanucleotide motifs; **Table S1**) successfully amplified from 42 'individuals' in SM360, 47 'individuals' in NI and ND and 48 'individuals' in SMI and SMD (as mentioned above, data for each 'individual' actually come from three larvae of a single cyst) and contained, on average, less than 10% missing data. Despite the fact that the markers were identified on a previous version of the genome, the 53 polymorphic markers were distributed on 47 different scaffolds of the last genome version (Cotton *et al.* 2014).

3.2.3 Detection of outlier loci

Two complementary F_{ST} outlier tests of selection were performed: DETSEL v. 1.0.2 (Vitalis 2012) and the FDIST approach (Beaumont & Nichols 1996) implemented in ARLEQUIN v. 3.5.2, which presents a few differences from the original version (Excoffier & Lischer 2010). Both methods rely on the estimation of the joint distribution of summary statistics, by means of stochastic simulations of neutral gene genealogies in a simple population model. FDIST and DETSEL differ by the underlying demographic model: while FDIST considers an island model of population structure, *i.e.* a set of populations with constant and equal deme size that are connected by gene flow, DETSEL considers a pure divergence model, in which an ancestral population splits into two daughter populations. Analyses were run separately for each lineage pair SMI/SMD, NI/ND and SM360/ SMD. In ARLEQUIN, for each analysis, 100 000 coalescent simulations in 100 demes were performed. In DETSEL, 1 000 000 simulations were performed for each pairwise comparison, assuming no bottleneck before the split and using the following parameters: maximum allele frequency = 0.99; mutation rate μ = 10⁻⁴ (corresponding to the average mutation rate for microsatellite markers; Selkoe & Toonen (2006)); time since population split t = 8 generations; ancestral population size $N_e = 10^4$, 10^5 and 10^6 (as recommended by Vitalis 2012; because of the uncertainty in actual N_e , data have to be analysed using a wide range of values for this parameter). The stepwise mutation model (SMM) was used for all simulations. However, additional simulations were performed using the infinite allele model (IAM), a mutation rate $\mu = 10^{-5}$ and/or assuming a bottleneck, and provided the same outlier list (data not shown). For both F_{ST} outlier tests, a locus was classified as a significant outlier if it lay outside the 95% and 99% confidence envelopes and, as the focus was on the loci putatively involved in divergence between adapted and non-adapted lineages, only the loci falling above the upper confidence limits were considered.

In addition, a third test based on the reduction in heterozygosity was used: the InRH test (Kauer *et al.* 2003; Schlötterer & Dieringer 2005), specifically designed for microsatellite loci. This test is based on the assumption that microsatellites linked to loci under selection will

show reduced levels of diversity between populations (Schlötterer & Dieringer 2005). The InRH statistic was computed for each locus in each lineage pair SMI/ SMD, NI/ND and SM360/SMD by calculating the natural logarithm (In) of the gene diversity ratio $[(1/(1 - H_{lineage1}))^2 - 1] / [(1/(1 - H_{lineage2}))2 - 1]$, where H is the expected heterozygosity. For lineages with monomorphic loci, one different allele was added to one individual, as null values of heterozygosities prevent the estimation of InRH due to division by zero (Kauer *et al.* 2003). LnRH estimates were standardized to obtain a mean of 0 and a standard deviation of 1. As the InRH is approximately normally distributed under the null hypothesis of neutrality (Schlötterer & Dieringer 2005), loci with InRH values lower than -1.96 (or greater than 1.96) and -2.58 (or greater than 2.58) were considered outliers at the 0.05 and 0.01 threshold, respectively, indicating reduced variability in the numerator lineage as compared to the denominator one (or in the denominator lineage compared to the numerator one).

3.2.4 Genic environment of outlier loci

There is no available data on the extent of recombination and linkage disequilibrium in the *G. pallida genome*. However, a genetic linkage map constructed with AFLP exists for the species *G. rostochiensis* (Rouppe van der Voort *et al.* 1999) and indicates an overall physical:genetic distance ratio of 120 kb cM⁻¹. Thus, a search was made for predicted genes located in a window of 240 kb around outlier loci (120 kb either side) in the annotated *G. pallida* genome (version nGp.v1.0; Cotton *et al.* 2014), available as an assembly of 125 Mb in 6873 scaffolds with a N50 of 122 kb. Combining transcriptomic data with manual curation, a total of 16 419 genes were predicted. As the majority of virulence factors identified in plant parasitic nematodes are effectors (Haegeman *et al.* 2012), genes coding for secreted proteins were prioritized in the search. Therefore, predicted proteins were examined for the presence of N-terminal signal peptides using SIGNALP v. 4.1 (Petersen *et al.* 2011). A BLAST search was also performed of the predicted genes against the annotated *G. rostochiensis* genome (version nGr.v1.0; Eves-van den Akker *et al.* 2016), for which an expert annotation has been performed (in contrast to the *G. pallida* genome which has only an automated annotation), in order to obtain more precise information on their functional annotation.

3.2.5 Genetic diversity and differentiation

The genetic characteristics of the main lineages of this study (SMD, SMI, NI and ND) were estimated using three datasets: one dataset containing all 53 microsatellite markers retained, one containing the putatively neutral markers (*i.e.* without the outlier loci detected) and one containing only the outlier loci (combining results of SMI/SMD and NI/ND comparisons). Genetic diversity was estimated through allelic richness (A_R) and unbiased gene diversity (H). Allelic richness A_R corresponds to the mean number of alleles per locus, and was estimated on a reduced sample of 20 individuals using the rarefaction method implemented in POPULATIONS v. 1.2.32 (Langella 2000). Gene diversity H corresponds to the average probability across loci to draw at random different alleles in the same population, and was computed using GENETIX v. 4.05.2 (Belkhir *et al.* 1996–2004). The fixation index F_{IS} was also computed using GENETIX, and the statistical significance of F_{IS} values for each lineage was tested using the allelic permutation method (10 000 permutations). Genetic differentiation among

all lineage pairs was estimated by calculating pairwise F_{ST} values using GENEPOP v. 4.0 (Rousset 2008).

3.3 Results

3.3.1 Detection of outlier loci

A genome scan was performed on the eighth generation of five lineages coming from two independent *G. pallida* populations (SM and N); two of them were virulent to lledher (SMI and NI), one was virulent to another resistant genotype harbouring *GpaV_{vrn}* in a different genetic background (SM360) and two were avirulent (SMD and ND). Each lineage was genotyped with 202 microsatellite markers. Fifty-three polymorphic markers were finally kept to perform the genome scan with three different methods: FDIST, DETSEL and InRH. Analyses were run separately for each lineage pair SMI/SMD, NI/ND and SM360/SMD. A locus was considered to be putatively under positive selection if at least two outlier tests were significant for that particular locus or if that locus was found to be an outlier in at least two pairwise comparisons.

Eight outlier loci were found across the three pairwise comparisons and using the three outlier detection tests (**Table 3.1**). Most of them were found with the three methods and at the 0.01 threshold. Gp249 in SMI/SMD and Gp235 in SM360/SMD were only detected with FDIST and DETSEL, and Gp121 in NI/ND and Gp235 in SMI/ SMD, were found only with DETSEL (**Table 3.1**). In order to lower the bias on the mean neutral F_{ST} estimation induced by outlier loci present in the dataset, and subsequently the simulation of the expected distribution under the null hypothesis, it is advocated to run the analysis a second time after removing outlier loci (*e.g.* Antao *et al.* 2008). This procedure is particularly justified for the NI/ ND comparison, where a locus shows a very high F_{ST} value (Gp101 with $F_{ST} = 0.75$; **Fig. 3.1b**). Therefore each analysis was run a second time after removing the outlier loci at the 0.01 threshold and it was found that Gp121 in NI/ND and Gp235 in SMI/SMD were outliers at the 0.05 threshold in FDIST.

In the comparisons between lineages virulent and avirulent to Iledher (*i.e.* SMI/SMD and NI/ND), seven outlier loci were detected (Figs 3.1 & 3.2; Table 3.1); among them, one outlier, Gp121, was shared by both populations (Fig. 3.3). Among the six other outliers, found in only one population, one of them, Gp101 (found with the three outlier tests in NI/ND comparison) was also detected as an outlier at the 0.05 threshold in both F_{ST} -based tests in the SMI/SMD comparison after the second run, and can therefore be considered as a second outlier shared by both populations (Fig. 3.3).

In the comparison SM360/SMD, four outliers were found (**Table 3.1**). Three of them (Gp121, Gp235 and Gr69) were shared with the SMI/SMD pair (**Fig. 3.3**). Gp121 was also shared, as seen above, with the NI/ND pair.

LnRH indicates in which lineage the reduction of variability – and therefore the potential positive selection – happened. In the SMI/SMD comparison, a reduction of variability was observed in the SMD lineage at the Gp121 locus, while this diversity loss happened in the SMI lineage at the Gp124 and Gr69 loci (Fig. 3.2a). In the NI/ND comparison, reduction of variability happened in ND for Gp101 and NI for Gp153 (Fig. 3.2b) and in the SM360/SMD comparison, it happened in SMD for Gp121 and Gp135, and in SM360 for Gr69.

Table 3.1 Microsatellite markers identified by three outlier detection methods (DETSEL, FDIST and InRH) in comparisons between pairs of potato cyst nematode lineages NI/ND, SMI/SMD and SM360/SMD. DETSEL v. 1.0.2, Vitalis (2012); FDIST, Beaumont & Nichols (1996) implemented in ARLEQUIN v. 3.5.2; InRH, Kauer *et al.* (2003), Schlötterer & Dieringer (2005).* P < 0.05. ** P < 0.01.

	Pairwise com	iparisons	
Outlier locus	NI/ND	SMI/SMD	SM360/SMD
Gp121	DETSEL**	DETSEL** + FDIST**	DETSEL** + FDIST**
		+ InRH**	+ InRH**
Gp101	DETSEL**		
	+ FDIST**		
	+ InRH**		
Gp235		DETSEL**	DETSEL** + FDIST**
Gr69		DETSEL** + FDIST**	DETSEL** + FDIST*
		+ InRH**	+ InRH**
Gp124		DETSEL** + FDIST**	
		+ InRH**	
Gp249		DETSEL** + FDIST**	
Gp153	DETSEL**		
	+ FDIST*		
	+ InRH**		
Gp135			DETSEL* + FDIST* + InRH*



Figure 3.1 Distribution of genetic differentiation (*F***sT values) among potato cyst nematode lineage pairs as a function of heterozygosity for each of 53 microsatellite loci in (a) comparison of lineages SMI and SMD and (b) comparison of lineages NI and ND**. Upper and lower dotted lines: 99th and 1st quantiles of the neutral envelope; upper and lower dashed lines: 95th and 5th quantiles of the neutral envelope; black and grey dots: outlier loci at 0.01 and 0.05 thresholds.



Figure 3.2 Standardized values of the natural logarithm of the gene diversity ratio (InRH) for each of 53 microsatellite loci in a comparison between potato cyst nematode lineages (a) SMI and SMD, and (b) NI and ND. Upper and lower dotted lines: 99% confidence interval (-2.58, +2.58); upper and lower dashed lines: 95% confidence interval (-1.96, +1.96); black dots: outlier loci at 0.01 threshold.



Figure 3.3 Venn diagram showing the microsatellite markers identified as outlier loci in one, two or three pairwise comparisons of potato cyst nematode lineages (NI/ND, SMI/SMD and SM360/SMD). Gp101 is on the line because it was detected in the first analysis in the NI/ND comparison and in the second analysis (after removing outlier loci) in the SMI/SMD comparison.

3.3.2 Genic environment of outlier loci

The predicted genes located in a window of 240 kb around outlier loci (120 kb on either side) were investigated. As a priority, genes coding for secreted proteins (*i.e.* harbouring a signal peptide) were searched for. Predicted genes were also checked using BLAST against the expert-annotated *G. rostochiensis* genome.

Microsatellite markers were developed on a previous version of the *G. pallida* genome. However, when a BLAST search of outlier loci against the annotated genome version was performed, Gp249 was not found and must, therefore, have been discarded from the new genome version. The scaffold where locus Gp135 was localized did not contain any predicted genes. Overall, 187 predicted genes were identified around the six remaining outlier loci (Table S2). About 44% of these corresponded to genes with unknown functions in G. pallida, but 10 harboured a signal peptide, indicating that they potentially encode secreted proteins. Nine others were orthologues of predicted genes in G. rostochiensis that harbour a signal peptide. Although their function is unknown, three genes, one localized near Gp153 and two localized near Gr69, were identified as potential novel effectors by Cotton et al. (2014). Among the identified genes with known functions, five harboured a signal peptide. One of them is described as SPRYSEC, which is well known to be involved in the pathogenicity of nematodes of the genus Globodera (e.g. Sacco et al. 2009). Another gene is also described as SPRYSEC but does not harbour a signal peptide. Interestingly, both genes are localized near Gp101, which was identified as an outlier in both NI/ND and SMI/SMD comparisons. Two genes described as calcium-binding proteins, a protein family also potentially involved in pathogenicity, were also found near Gp124 and Gr69.

3.3.3 Genetic diversity and differentiation

Allelic richness (A_R), unbiased gene diversity (H), fixation index (F_{IS}) of the main lineages of this study (SMD, SMI, NI and ND) and genetic differentiation (F_{ST}) among all these lineage pairs were estimated for three datasets: one dataset containing all retained 53 microsatellite markers, one containing the putatively neutral markers and one containing only the outlier loci (combining results of SMI/SMD and NI/ND comparisons).

Genetic diversity was higher for the *G. pallida* lineages coming from the population of Saint-Malo (SMD and SMI) than for the lineages from the population of Noirmoutier (ND and NI), both for A_R and *H* (Table 3.2). The comparison between the three datasets showed clearly that genetic diversity (A_R and *H*) was lower for the seven outliers than for the 46 neutral markers, and intermediate for the combined dataset (Table 3.2).

A highly significant heterozygote deficit relative to Hardy-Weinberg equilibrium was observed in all lineages: in the dataset with the 46 neutral markers, F_{IS} values ranged from 0.085 to 0.165 (Table 3.2). For all lineages, the F_{IS} was higher for the outlier than for the neutral loci, and this was mainly due to the Gp101 locus, as revealed by a jackknife resampling performed on the outlier dataset (data not shown).

The highest F_{ST} values were found between lineages coming from the two distinct *G. pallida* populations (Saint-Malo and Noirmoutier), except for the dataset containing only the outlier loci, and was on average 0.16 and 0.17 for datasets containing only neutral markers and all markers, respectively. Regarding each pair of virulent and avirulent populations, F_{ST} values were higher for the outlier than for the neutral markers, and intermediate for the combined dataset (F_{ST} SMD/SMI = 0.105, 0.002 and 0.014 for outlier, neutral and combined dataset, respectively; F_{ST} ND/NI = 0.308, 0.009 and 0.049 for outlier, neutral and combined dataset, respectively).

Table 3.2 Genetic characteristics of the main *Globodera pallida* lineages studied. *n*, number of sampled individuals; *A*_R, allelic richness; *H*, Nei's (1978) expected unbiased heterozygosity; *F*_{IS}, fixation index (***P* < 0.01).^aMM, estimate with data from 53 microsatellite markers. ^bNM, estimate with data from 46 putatively neutral markers. ^cOM, estimate with data from 7 outlier markers.

				An			H		~	$F_{\rm ff}$	
Lineage	Locality	п	MM*	NM [±]	OM ^e	MM	NM	OM	MM	NM	OM
SMD	Saint-Malo	48	2.55	2.57	2.38	0.386	0.397	0.318	0.125**	0.085**	0.451**
SMI	Saint-Maio	48	2.49	2.54	2.16	0.396	0.412	0.292	0.136**	0.118**	0.308**
ND	Noirmoutier	47	2.14	2.19	1.81	0.271	0.277	0.231	0.169**	0.165**	0.202**
NI	Noirmoutier	47	2.20	2.23	1.99	0.266	0.278	0.188	0.132**	0.120**	0.243**

3.4 Discussion

In this study, a genome scan approach was used on potato cyst nematode, *G. pallida*, after a short experimental evolution to test whether this is an efficient method to target genomic footprints of selection due to host adaptation. The genome scan performed on *G. pallida* lineages virulent and avirulent to potato cultivar Iledher, and coming from two independent populations, highlighted seven different outlier loci, which constitute candidate genomic regions putatively linked to host adaptation.

As previously recorded in natural populations of several plant parasitic nematode species (Montarry et al. 2015), all the lineages showed a heterozygote deficit. It is known that heterozygote deficit is likely to increase the rate of false positive outliers (De Mita et al. 2013) and so it is possible that some of the outlier loci identified in the present study could be false positives. Nevertheless, there are several ways to determine whether outlier loci, and, by extension, the linked genomic regions, are good candidates for being involved in adaptation. First, it has been recommended that false positives can be avoided by performing more than one type of outlier test and only considering those detected by at least two methods as true positives (Luikart et al. 2003). In the present study, most of the outliers were found in two or three of the tests. Secondly, a good candidate should be identified in both independent populations. In the present study, the strong bottleneck imposed on lineages exposed to the resistant potato genotypes could have induced genetic drift and, therefore, a risk of detecting loci fixed only by drift. However, the use of two independent populations as repeated samples provided stronger support to outlier loci found in both populations, as the probability that the same loci were fixed by chance in two independent populations is very small (here P = 1/53 x1/53 = 0.00036). This particularly applies to the Gp121 locus, which was also found in the third comparison, i.e. SM360/SMD, and to some extent, to the Gp101 locus (Fig. 3.3). When studying natural populations, this strategy is also useful to avoid the detection of outlier loci selected by local environmental variables other than adaptation to the host in some populations, such as climate or pedology. Here, this pitfall was avoided by using lineages coming from an experimental evolution, which allowed control of the environment for at least the last five generations that reproduced under glasshouse conditions (Fournet et al. 2013). Thus, any outlier loci identified were likely to be linked to the resistant or susceptible status of the host plants.

Even though 16 genes were found in a 240 kb window around the outlier locus Gp121, none of them was known as an effector and none of them harboured the recently identified DOG-Box that may be a predictor of secretion (Eves-van den Akker *et al.* 2016); nevertheless, four of the genes may be candidates for playing a role in pathogenicity. Among the 21 genes

found close to locus Gp101, five harbour a signal peptide and two are known as SPRYSEC, which are effectors responsible for the suppression of plant defences and for nematode virulence.

Results obtained with InRH indicated that, in the case of loci Gp121 and Gp101, the reduction of variability, and therefore the potential positive selection, could correspond to adaptation to the cultivar Désirée rather than to lledher. This is not surprising as, before the experiment, none of the potato genotypes was cultivated in either of the population source fields. However, InRH also supported the results obtained for three outlier loci, Gp124, Gr69 and Gp153, that the decrease of diversity occurred in the lineages virulent to lledher. The Gp153 locus could be an interesting candidate, as it was identified as an outlier in the NI/ND pair with all three tests, and a gene close to this locus was described as a potential novel effector by Cotton *et al.* (2014). Two genes described as calcium-binding proteins were found near Gp124 and Gr69; these could also represent good candidates for adaptation to plant resistance as they are homologous to annexins and calreticulins, which are effectors for suppression of plant defence, as well as calcium-binding proteins (Lee *et al.* 2004). It should be emphasized that, at this stage and with the genome coverage used in this study, the genomic regions involved in the adaptation to the QTL *GpaVyra* cannot be precisely identified.

The results of the present study have also provided clues for solving the question of whether the genetic architecture of *G. pallida* virulence is the same or differs between nematodes on potatoes of different genetic backgrounds. Indeed, it is now well established that the plant genetic background can influence the durability of plant resistances to virus, fungi and nematodes (Palloix *et al.* 2009; Brun *et al.* 2010; Fournet *et al.* 2013). Here, three outlier loci were found in both SMI/SMD and SM360/SMD pairs. This result supports the hypothesis that the same genomic regions could be involved in adaptation to lledher and to a potato genotype harbouring the same resistance QTL in another genetic background. However, further research is needed to test this hypothesis, as the present study lacks independent replicates regarding the potato genotype showing the same resistant QTL in a different genetic background.

This work shows that genome scans on experimentally evolved lineages are efficient for the detection of genomic regions impacted by selection and putatively involved in host adaptation. It is important that this approach has been confirmed to be feasible in nonmodel organisms with a generation time much longer than those of model organisms such as Caenorhabditis elegans. In fact, because G. pallida completes only one generation per year, the experimental evolution performed here involved a short process of divergence, *i.e.* only eight generations. If adaptation was clearly shown at the phenotypic level (Fournet et al. 2013), there could be a risk that the indicators of selection would be too widespread along the genome, preventing the detection of the genetic bases of this adaptation. Indeed, eight generations may have been insufficient to allow enough recombination events in the genome and could have led to the detection of an excessive number of outlier loci. However, in the present study, only 6% of microsatellite loci on average were detected as outliers, showing that the nematode lineages used had undergone sufficient genetic mixing in only eight generations. This result is consistent with the findings in the majority of genome scan studies, where 2–10% of markers were typically reported as departing from the neutral expectation (Nosil et al. 2008). Moreover, if the mutation(s) preexisted in the populations, i.e. there was standing variation, it could be predicted that the selective sweep would be restricted enough to allow the detection of a small number of outliers. Several facts support the hypothesis that the individuals virulent to GpaV_{vrn} were already present in the wild populations before the start of the experiment. Bakker (2002) observed that the introduction of many resistant cultivars into Europe has been unsuccessful because G. pallida and G. rostochiensis populations show high levels of virulence, without ever being exposed to these resistance genes in Europe; this suggests that the virulence genes were already present in the initial populations introduced from the Andean region of South America. This is in line with the fact that some native G. pallida populations from South America are also virulent to S. vernei (Hockland et al. 2012). The lineages used in the present study support this hypothesis as, in a previous study, both independent populations were shown to overcome several resistant genotypes in less than eight generations (Fournet et al. 2013). It is more probable that such a result would be caused by a shared pre-existing mutation than by the independent appearance of a new mutation in both populations in a small time interval. In a situation where a novel mutation is selected and rapidly fixed in the population, the diversity is strongly reduced over a large region of the genome because recombination cannot occur in a short period. In contrast, if adaptation occurs from a pre-existing polymorphism, as is proposed for the populations used in the present study, the favourable variant is ancient and has been able to recombine, so the selective sweep is located in a restricted region (Przeworski et al. 2005).

The results of this study are encouraging as they have given a first insight into the possibilities that a genome scan approach offers for investigation of the genetic bases of host adaptation in nematodes and other sexually reproducing plant parasites. The coupled approach used in this study is close to the 'evolve and resequencing' (E&R) studies that combine experimental evolution and genome resequencing to identify loci involved in adaptation and analyse the trajectories of these loci during adaptation (*e.g.* Turner *et al.* 2011). However, E&R studies use time series data, which was not possible for the present investigation as the parental material of the lineages was no longer available. Such studies seem to be mostly performed on model species such as *Drosophila melanogaster*. More generally, the methods of population genomics are still underused in phytopathological research for determination of the adaptive potential of pathogen populations.

In the present study, only 53 microsatellite loci were genotyped in a 100 Mb genome. Even if some candidate genes were highlighted, they remain very hypothetical and it is still not possible to identify precisely the molecular bases of the adaptation. It is now necessary to use a more refined approach, using markers more densely distributed along the genome, *i.e.* single nucleotide polymorphisms obtained by whole genome resequencing. With the recent publication of the G. pallida genome (Cotton et al. 2014), a Pool-Seq approach has begun on the lineages studied here. Such an improved coverage will allow the identification of outliers that cluster physically close to each other, leading to more accurate determination of the genomic regions impacted by selection; some of these outliers may be close to those revealed in the present study. These improved results could be used to develop efficient strategies of deployment of plant resistance even before pinpointing the mutation involved in overcoming the resistance. Indeed, a proper characterization of the nematode populations, including their genetic variability and particularly virulence gene frequencies, could enable the prediction of the efficiency and durability of a plant resistance before its deployment. Thus, SNPs linked to virulence genes could be used to follow the evolution of virulence allele frequencies in populations from the field or from experimental evolution studies, to assist different resistance management strategies.

Additional Supporting Information may be found in the Annexes section at the end of the thesis manuscript.
3.5 Acknowledgments

The authors gratefully acknowledge Didier Mugniéry and Marie-Claire Kerlan who assisted in the experimental evolution. They kindly acknowledge Didier Andrivon, Fabien Halkett and two anonymous referees for comments on previous versions of this manuscript and Sylvie Bardou-Valette for her help in the microsatellite genotyping. D.E.B. is supported by the INRA department SPE and the Région Bretagne through a three year PhD grant.

CHAPITRE 4

IDENTIFICATION DES RÉGIONS GÉNOMIQUES IMPLIQUÉES DANS L'ADAPTATION À LA RÉSISTANCE DE LA PLANTE HÔTE

CHAPITRE 4 IDENTIFICATION DES RÉGIONS GÉNOMIQUES IMPLIQUÉES DANS L'ADAPTATION À LA RÉSISTANCE DE LA PLANTE HÔTE

Dans le chapitre précédent, nous avons mis en évidence la faisabilité d'une approche de scan génomique sur des lignées de *G. pallida* issues d'un processus court de divergence en évolution expérimentale. Nous avons en effet identifié un nombre restreint d'*outliers*, dont deux notamment sont communs aux deux populations indépendantes. Cependant, le faible nombre de marqueurs utilisés ne permet pas de couvrir suffisamment le génome pour cibler des régions génomiques précises. La caractérisation du matériel biologique a notamment permis de confirmer la différenciation génétique marquée entre les populations SM et N et donc leur statut de réplicas indépendants, situation idéale pour la réalisation d'un scan génomique performant.

Dans ce chapitre, nous avons pour objectif d'identifier les régions génomiques candidates à l'adaptation à la résistance issue de *S. vernei* en affinant notre approche par la réalisation d'un scan génomique qualifié de haute densité. Nous avons donc à nouveau utilisé les lignées de *G. pallida* issues d'évolution expérimentale, en nous concentrant sur celles provenant des deux populations indépendantes et ayant évolué sur les variétés résistante lledher et sensible Désirée. L'obtention de marqueurs moléculaires répartis plus densément le long du génome a été rendue possible par la réalisation d'un reséquençage de génome complet de pools d'individus ayant permis d'aboutir à l'identification d'1.6 million de SNPs. Les comptages alléliques aux SNPs identifiés ont été utilisés pour réaliser le scan génomique à l'aide d'un modèle récent testant à la fois la différenciation génétique et l'association à une covariable.

Même si les prix de séquençage tendent à diminuer drastiquement, le coût d'un reséquençage complet de génome reste relativement prohibitif pour pouvoir utiliser cette approche en routine dans l'objectif d'une identification rapide des bases génétiques de l'adaptation à de nouvelles résistances. Forts des résultats précédemment obtenus, nous avons souhaité comparer les performances d'un scan génomique haute densité, réalisé à partir de données issues d'un reséquençage complet de génome, à celles d'un scan génomique moyenne densité, réalisé à partir de données provenant d'un séquençage d'une représentation réduite du génome. Pour cela, nous avons mené une approche de génotypage par séquençage (GBS) sur les mêmes lignées que celles utilisées précédemment. Dans le même temps, une telle approche a été réalisée sur l'ensemble du matériel biologique issu de l'évolution expérimentale. En effet, si l'utilisation de données GBS pour la réalisation d'un scan génomique s'avère concluante, l'étude de lignées ayant évolué sur différents génotypes portant GpaV_{vrn} permettrait d'affiner nos conclusions. Enfin, cette étude GBS représentait également une stratégie alternative d'obtention de marqueurs moléculaires et de régions génomiques candidates, pouvant servir à filtrer les candidats identifiés dans la précédente étude.

Les résultats des travaux portant sur le scan génomique haute densité font l'objet d'un article intitulé « Genome scans on experimentally evolved populations reveal candidate regions for adaptation to plant resistance in the potato cyst nematode *Globodera pallida* »

accepté pour publication dans la revue *Molecular Ecology*. Les travaux portant sur le scan génomique moyenne densité feront également l'objet d'une publication, cependant l'état d'avancement des analyses ne permet pas encore d'en préparer l'ébauche. Les premiers résultats sont néanmoins présentés et brièvement discutés.

GENOME SCANS ON EXPERIMENTALLY EVOLVED POPULATIONS REVEAL CANDIDATE REGIONS FOR ADAPTATION TO PLANT RESISTANCE IN THE POTATO CYST NEMATODE *GLOBODERA PALLIDA*

Delphine EOCHE-BOSY¹, Mathieu GAUTIER^{2,3}, Magali ESQUIBET¹, Fabrice LEGEAI^{4,5}, Anthony BRETAUDEAU^{4,6}, Olivier BOUCHEZ^{7,8}, Sylvain FOURNET¹, Eric GRENIER¹ and Josselin MONTARRY^{1,*}

¹IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, 35650 Le Rheu, France; ²CBGP, INRA, IRD, CIRAD, Montpellier SupAgro, 34988, Montferrier-sur-Lez, France; ³IBC, 34095, Montpellier, France; ⁴IGEPP, BIPAA, INRA, Agrocampus Ouest, Université de Rennes 1, 35042 Rennes, France; ⁵IRISA, GenScale, INRIA, 35042 Rennes, France; ⁶IRISA, GenOuest Core Facility, INRIA, 35042 Rennes, France; ⁷GeT-PlaGe Genotoul, INRA, 31326, Castanet-Tolosan, France; ⁸GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326, Castanet-Tolosan, France. *Correspondence: Josselin Montarry, IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, 35650 Le Rheu, France. Tel: +33 2 23 48 51 73; Fax: +33 2 23 48 51 50; E-mail: josselin.montarry@inra.fr

Running title: Molecular signatures of adaptation to resistance

Improving resistance durability involves to be able to predict the adaptation speed of pathogen populations. Identifying the genetic bases of pathogen adaptation to plant resistances is a useful step to better understand and anticipate this phenomenon. Globodera pallida is a major pest of potato crop for which a resistance QTL, GpaVvrn, has been identified in Solanum vernei. However, its durability is threatened as G. pallida populations are able to adapt to the resistance in few generations. The aim of the present study was to investigate the genomic regions involved in the resistance breakdown by coupling experimental evolution and high-density genome scan. We performed a whole genome resequencing of pools of individuals (Pool-Seq) belonging to G. pallida lineages derived from two independent populations having experimentally evolved on susceptible and resistant potato cultivars. About 1.6 million SNPs were used to perform the genome scan using a recent model testing for adaptive differentiation and association to population-specific covariables. We identified 275 outliers and 31 of them, which also showed a significant reduction of diversity in adapted lineages, were investigated for their genic environment. Some candidate genomic regions contained genes putatively encoding effectors and were enriched in SPRYSECs, known in cyst nematodes to be involved in pathogenicity and in (a)virulence. Validated candidate SNPs will provide a useful molecular tool to follow frequencies of virulence alleles in natural G. pallida populations and define efficient strategies of use of potato resistances maximizing their durability.

Keywords: evolve and resequence, experimental evolution, Pool-Seq, population genomics, selection, virulence.

Author contributions: SF performed the experimental evolution. DEB, ME, SF and JM performed the experiments according to a protocol elaborated jointly by DEB, EG and JM. OB performed the Illumina sequencing at the GeT-PlaGe platform (Toulouse, France). DEB, MG, FL and AB analyzed the data. DEB and JM wrote the text and prepared the figures.

Accepted for publication in Molecular Ecology in July 2017

4.1 Introduction

Because of the threat caused by crop pathogens to global food security, control methods to limit yield losses and maintain sustainable productions are needed. If pesticides have been successfully employed for years, the current societal and political demand for a reduction of their use is strong. In most cases, resistant cultivars are now the chosen alternative, as they can be highly effective, organism-specific, economically and environmentally sustainable. However, this control method can be limited by evolutionary capacities of targeted pathogens, which can adapt to plant resistances (McDonald & Linde 2002). Resistance breakdowns have been documented in a wide range of crop pathogens, such as virus, bacteria, nematodes, fungi or oomycetes (e.g., Castagnone-Sereno 2002; McDonald & Linde 2002; Janzac et al. 2009), whereas the development of resistant cultivars is a long process, in the order of decades, *i.e.* usually longer than the time needed for pathogen populations to overcome resistances. Consequently, the design of strategies allowing an increase in resistance durability has been a major goal in plant pathology research over the last 50 years, especially as resistance genes or QTLs remain a scarce resource. To be efficient, such strategies have to integrate knowledge on the adaptation of pathogen populations to the selective pressure imposed by the resistant plant. Consequently, a significant step towards a better management of resistance durability could be taken by identifying the genetic determinants of adaptation to plant resistances, which would represent direct predictors allowing to link dynamics and genetics of resistance breakdown.

Plant-parasitic nematodes are major agricultural pathogens causing severe damages in crops worldwide (Nicol *et al.* 2011), but sources of resistance against them remain scarce. Several studies have been conducted to optimize the management of these resistances (Djian-Caporalino *et al.* 2014; Barbary *et al.* 2015). Cyst nematodes are among the most economically damaging plant-parasitic nematodes. Over the last few years, increased knowledge about characteristics of cyst nematode populations allowed to better predict the global evolutionary potential of these populations. In particular it was showed that cyst nematode populations represent a real threat to the durability of plant resistances due to an important passive dispersion (Picard *et al.* 2004; Plantard & Porte 2004; Alenda *et al.* 2014) and favored expression of recessives virulence genes due to inbreeding (Montarry *et al.* 2015). The adaptation speed of nematode populations would also be modulated by the number and the nature of genetic mutations required to overcome the plant resistance, and by the presence and, if any, the frequency, of virulent individuals in fields.

The cyst nematode *Globodera pallida* (Stone) is a major pest of potato crop (Oerke *et al.* 1994; van Riel & Mulder 1998) and *Solanum vernei*, a wild potato species, is an interesting source of resistance against *G. pallida* as it is highly effective. The resistance, which leads to the development of most nematodes into adult males, is explained at 61% by the major QTL *GpaV*_{vrn} mapped on the potato chromosome V (Rouppe van der Voort *et al.* 2000). This source of resistance is to date the only one exploited in commercial potato cultivars at the European level. The resistant cultivar lledher, carrying *GpaV*_{vrn}, has been registered in 2009 in the French catalogue as the first cultivar showing a high level of resistance to *G. pallida*. However, Fournet *et al.* (2013) highlighted that *G. pallida* populations were able to completely overcome the resistance of lledher in only few generations in experimental evolution. Identifying the determinants of virulence thus appears to be crucial to better anticipate the speed of adaptation of *G. pallida* populations.

In organisms with relatively small genomes, such as RNA viruses with 10-15 kb genomes, it is possible since several years to sequence the entire genome and to directly compare sequences of virulent and avirulent strains in order to find the mutation responsible for resistance adaptation (Meshi *et al.* 1988; Díaz *et al.* 2004; Ayme *et al.* 2006; Mardis 2008; Janzac *et al.* 2010). However, this approach cannot be considered in eukaryotic organisms whose genomes range from one to hundreds of thousands megabases. Genome scans represent an efficient alternative to target genomic regions involved in adaptation in such organisms, as they analyze genome-wide variations at the light of theoretical predictions about the effects of selection, in order to detect locus specific signatures of positive directional selection (Luikart *et al.* 2003; Storz 2005). Despite their strong potential to elucidate the genetic bases of adaptation, genome scans have never been used to identify the determinants of virulence in nematodes. However, Bekal *et al.* (2015) recently opened the way for population genomic approach in cyst nematodes. They performed a whole genome allelic imbalance analysis of SNPs in *Heterodera glycines* inbred lines grown in the laboratory for over 30 generations on resistant and susceptible soybean plants.

The cyst nematode effectors secreted during the interaction allow the maintenance of the biotrophic interaction for several weeks and the induction of a permanent feeding site. Numerous candidate effectors have been isolated from nematodes to date (see Haegeman et al. 2012; Hewezi & Baum 2012 for reviews) that have roles in the invasion of host tissues, evasion of plant defences, suppression of plant defences or manipulation of plant cell fate. Since any compound secreted by the parasite can potentially be perceived as nonself by the plant, nematode effectors are prime candidates for acting also as avirulence factors when recognized by the host guarding or resistance proteins. In *Globodera rostochiensis*, the sister species of G. pallida which also attacks potato, the venom allergen Gr-VAP1 effector was identified as an avirulence gene product triggering a cell death response in tomato (Lycopersicon esculentum) plants containing the Cf-2 and Rcr3pim resistant genes (Lozano-Torres et al. 2012, 2014). In G. pallida, the avirulence gene Gp-RBP-1, coding for a SPRYSEC effector - acting as a versatile protein-binding platform to disrupt plant innate immunity (Diaz-Granados et al. 2016) - interacts with the GPA2 resistant protein in potato (Sacco et al. 2009). In this case, virulence is due to a single amino-acid polymorphism, however this mutation is widely distributed in European populations (Carpentier et al. 2012) and therefore potato carrying Gpa2 resistance gene is of very limited interest for control of G. pallida. On the contrary, the resistant cultivar lledher shows a high level of resistance to a wide range of European G. pallida populations.

The aim of the present study was to investigate the *G. pallida* genomic regions involved in the breakdown of the resistance of potato cultivar lledher. By analyzing microsatellite data, Eoche-Bosy *et al.* (2017) showed that a genome scan on the *G. pallida* lineages coming from the experimental evolution performed by Fournet *et al.* (2013) was feasible. Here, we took advantage of the same biological material to perform a high density genome scan using data from a whole genome resequencing of pools of individuals (Pool-Seq, Futschik & Schlötterer 2010; Zhu *et al.* 2012; Ferretti *et al.* 2013; Gautier *et al.* 2013; Schlötterer *et al.* 2014) and a recent model testing for adaptive differentiation and association to population-specific covariables (BayPass, Gautier 2015).

4.2 Materials and methods

4.2.1 Study system: Globodera pallida

Globodera pallida is a gonochoristic diploid organism with obligate sexual reproduction, which achieves one generation per year in European climatic conditions (Jones 1950). This obligate parasite enters the plant roots as second-stage juveniles (J2) and establishes a specialized feeding structure, the syncytium (Jones & Northcote 1972), which is a severe nutrient sink for the plant. In this species, sex is environmentally determined and depends on the size and efficiency of the syncytium (Sobczak & Golinowski 2011). Adult males leave the root to mate females, which can be fertilized by several males (Green *et al.* 1970; Triantaphyllou & Esbenshade 1990). After mating, the females continue to feed from the syncytium and when egg development is completed, they die and form a cyst, enclosing hundreds of eggs, which can stay viable for several years in soils.

The genome of *G. pallida* has been recently sequenced (Gpal.v1.0, Cotton *et al.* 2014). It is available as an assembly of 124.7 Mb in 6,873 scaffolds, with a N50 of 122 kb and a GC content of 36.7% (Cotton *et al.* 2014). Combining transcriptomic data with manual curation, a total of 16,419 genes were predicted. The genome of *G. rostochiensis* has also been recently sequenced (nGr.v1.0, Eves-van den Akker *et al.* 2016) and could be a more accurate representation of a *Globodera* genome, as suggested by its higher completeness and low level of gene duplication. This genome is available as an assembly of 95.9 Mb in 4,377 scaffolds, with a N50 of 88 kb and a GC content of 38.1%. Annotation of *G. rostochiensis* genome could also be of better quality as a manual annotation phase followed the initial phase of automated annotation, resulting in the prediction of 14,378 genes.

4.2.2 Selection of virulent and avirulent G. pallida lineages

This study relies on the biological material coming from the experimental evolution performed by Fournet *et al.* (2013). Briefly, nematode lineages were established from cysts of two French natural *G. pallida* populations, SM (near Saint-Malo, Brittany, north-western France) and N (from the island of Noirmoutier, western France) coming from infested fields (**Fig. S1**, Supporting information). The lineages used here were obtained by rearing both populations during eight successive cycles (*i.e.* eight generations) on the susceptible potato cultivar Désirée (D) and on the resistant cultivar lledher (I). The present study was conducted with the eighth generation of the four lineages: two lineages adapted to the resistance of lledher (named hereafter SMI and NI), and two lineages which remained unadapted (named hereafter SMD and ND).

4.2.3 Pool sequencing

As the amount of DNA in a single *G. pallida* individual is very low, pooling individuals from a same lineage was an efficient way to increase the amount of DNA while allowing the accurate estimation of population allele frequencies (*e.g.*, Gautier *et al.* 2013). To accurately represent the genetic variability of each lineage, we chose to sample individuals from different cysts, rather than different individuals from a same cyst, to constitute the pools. We therefore

sampled 300 cysts in each lineage, which was the maximum available, and crushed them individually in sterile water. Two J2 (siblings) were sampled from each cyst and attributed to two different pools, resulting in two pools of 300 individuals for each lineage, i.e. two biological replicates. Water in samples was then vacuum-evaporated in a heated Speed Vac Concentrator (MiVac, Genevac Ltd., Ipswich, UK). As the cuticle of the nematode could prevent the lysis of tissues by the proteinase K, samples were stored at -80°C during two hours and then rapidly heated at ambient temperature, just before DNA extraction. DNA was extracted directly from each pool using Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. A pooled DNA extraction has been favored over an individual DNA extraction following by a pooling of the DNA in equimolar proportions because this last strategy can lead to heterogeneity in the amount of each individual DNA due to measurement (poor estimations of individuals DNA concentration) or pipetting errors (Gautier et al. 2013). Extracted genomic DNA was quantified using a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and quality was estimated using a Nanodrop® ND-2000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). The DNA samples conformed to the required purity criteria (A260/A230 and A260/A280 > 1.8) for gDNA library preparation for sequencing, but not to the required concentration criterion, as the amount of DNA obtained from a 300 individuals pool was still very low (*i.e.* 17 ng of DNA per pool on average). Paired-end libraries were therefore constructed from the totality of DNA available in each pool (i.e. each biological replicate in each lineage) contained in 100 µL AE buffer (Qiagen), using the TruSeq Nano DNA Sample Preparation Kits (Illumina, San Diego, CA, USA), according to the manufacturer's instructions. Briefly, DNA was fragmented using a Covaris M220 ultrasonicator (Covaris, Woburn, MA, USA) and, after a purification step, end repaired. A bead-based size selection was performed, then DNA was A-tailed and ligated to indexed sequencing adapters (60 bp on each side), each biological replicate being identified by a different index. After double-purification, libraries were enriched by eight PCR cycles, followed by a final purification. Library profiles were controlled using a DNA High Sensitivity chip on a BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), showing that the average size of the eight libraries was 513 bp (*i.e.* average insert size of 393 bp). Libraries were then quantified by qPCR on ABI7900HT (Applied Biosystems) in order to pool them in equimolar proportions. The final pool was again quantified by qPCR on ABI7900HT in order to load on each lane a volume corresponding to a DNA concentration of 8 pM. In order to sequence the eight libraries at an expected 150X coverage depth, four lanes were necessary. Sequencing was performed on an Illumina HiSeq[™] 2500 platform (Illumina). Three lanes were sequenced with HiSeq v4 reagent kits (HiSeq PE Cluster Kit v4 and HiSeq SBS Kit v4, Illumina) providing 2x125 bp paired-end reads and the fourth lane was sequenced with TruSeq v3 reagents kit (TruSeq PE Cluster Kit v3 and TruSeq SBS Kit v3, Illumina) providing 2x100 bp paired-end reads. Library preparation and sequencing were performed at the GeT-PlaGe France Genomics sequencing platform (Toulouse, France). All the obtained data were submitted to the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra/).

4.2.4 Mapping and SNP calling

Reads from the 64 fastq files (4 lineages x 2 biological replicates x 4 lanes x 2 (pairedend)) were aligned to the *G. pallida* reference genome using Bowtie2 version 2.1.0 (Langmead *et al.* 2009; Langmead & Salzberg 2012) with default parameters. Duplicate reads were removed using Picard version 1.122.0 (http://broadinstitute.github.io/picard/) and remaining reads were realigned around indels using Realigner Target Creator and Indel Realigner from the Genome Analysis Toolkit, GATK version 2.8.0 (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). The 32 BAM files obtained were merged in a single mpileup file using SAMtools MPileup version 0.1.19 (Li et al. 2009) with minimum mapping quality and minimum base quality set to 20. The whole previous workflow was performed on the Galaxy platform (https://galaxyproject.org/, Blankenberg et al. 2010). The mpileup file was further processed to perform SNP calling and derive read counts for each alternative base, after discarding bases with a Base Alignment Quality score below 25. Genomic positions kept in the dataset had to show only two different bases across all the samples, however, the tri-allelic positions for which the third allele was represented by only one read were included in the analysis as bi-allelic SNPs (after filtering the third allele as a sequencing error). To evaluate the reproducibility of the sequencing results across the technical replicates (sequencing lanes), we carried out a Principal Component Analysis (PCA) based on the allele frequency counts (Fig. **S2**, Supporting information). A last filtering step based on the coverage and the minimum allele count (MAC) was performed: a SNP was kept in the dataset if i) it had a coverage of more than 30 reads and less than 300 reads (corresponding approximately to the 95% percentile of the empirical coverage distribution) in each sample and ii) the minor allele was represented by at least two reads in two different pool samples.

4.2.5 Genome scan for adaptive divergence and association with the virulent/avirulent status

To detect genomic signatures of selection, we used the package BayPass version 2.1 (Gautier 2015) that provides a reimplementation of the Bayesian hierarchical model initially proposed by Coop *et al.* (2010) and includes several extensions that improve accuracy and decision criteria. The underlying BayPass framework allows both to identify overly differentiated SNPs (candidate for adaptive selection) in a robust fashion, through the computation of the *XtX* genetic differentiation statistics (Günther & Coop 2013), and to perform genome-wide association with population-specific covariables. The underlying models explicitly account for the covariance structure among the population allele frequencies that originates from the shared history of the populations under study, through the estimation of the population covariance matrix Ω , which renders the identification of SNPs subjected to selection less sensitive to the confounding effect of demography (Bonhomme *et al.* 2010; Günther & Coop 2013). BayPass is generic enough to be suited for the analyses of data from experimental evolution in which the allele frequency covariance structure is simpler, and can handle data derived from Pool-Seq experiments (Gautier 2015).

The allele count dataset was analyzed under all the models implemented in BayPass with all the parameter set to default, considering the virulence/avirulence status (coded as a binary variable with value -1 and 1 respectively) of each sample for the association analysis. First, analysis under the core model allowed to estimate the *XtX* for each SNP that were calibrated after analyzing a pseudo-observed dataset (POD) simulated under the inference model with parameters set equal to those estimated in the real data. Second, we relied on the Bayes Factor (BF) calculated under the auxiliary covariate model to evaluate association of SNPs to the virulent/avirulent status of the lineages. The auxiliary covariate model involves the introduction of a binary auxiliary variable to classify each locus as associated or not. This

allows to easily compute Posterior Inclusion Probability (and BF) for each locus while explicitly accounting for multiple testing issues. For each SNP, the Bayes Factor (denoted BF_{mc} as in Gautier 2015) was expressed in deciban units (dB) via the transformation 10 log₁₀ (BF). As a decision rule, we then followed the popular Jeffreys' rule (Jeffreys 1961) that quantifies the strength of evidence (here in favor of association of the SNP with the covariable) as 'strong' when 10 dB < BF < 15 dB, 'very strong' when 15 dB < BF < 20 dB and 'decisive' when BF > 20 dB.

Three independent BayPass analyses were carried out for each model, giving different values to initial seed of the (pseudo-) Random Number Generator. Under the core model in particular, the posterior estimates of Ω were found almost identical with a FMD distance (Förstner & Moonen 2003) between each pair of matrices always lower than 0.01. For prioritization purposes, only SNPs that were overly differentiated if XtX > 1% POD significance threshold and if BF_{mc} > 20 in the three independent analyses were further considered as candidate.

As we were interested in identifying genomic regions involved in the adaptation to Iledher, we precisely had to consider only SNPs showing footprints of selection in the Iledher lineages (SMI and NI) and not in the Désirée lineages. It is in fact possible to pinpoint genomic regions involved in host adaptation (rather than resistance adaptation), which can occur both in nematode lineages having evolved on Iledher and in the ones having evolved on Désirée, as highlighted in Eoche-Bosy et al. (2017). For this purpose, we used the InRH test (Kauer et al. 2003; Schlötterer & Dieringer 2005). LnRH is traditionally used for microsatellite datasets, but as it is not based on a particular mutation model (Kauer et al. 2003), it can also be applied to SNP datasets (Mc Evoy et al. 2006; Vasemägi et al. 2012). This test, based on the assumption that markers linked to loci under selection will show reduced levels of diversity within populations (Schlötterer & Dieringer 2005), uses the expected heterozygosity H to compute the InRH statistics. We merged allelic counts of biological replicates and derived from it allelic frequencies at each SNP in SMI, SMD, NI and ND. We computed the InRH statistic for each locus in lineage pairs SMI/SMD and NI/ND by calculating the natural logarithm (In) of the gene diversity ratio $\left[\left(\frac{1}{(1-H_{\text{lineage1}})}^{2}-1\right] / \left[\left(\frac{1}{(1-H_{\text{lineage2}})}^{2}-1\right]\right]$. For lineages with monomorphic loci, one different allele was added to one individual, as null values of heterozygosity prevent the estimation of InRH due to division by zero (Kauer et al. 2003). LnRH estimates were standardized to obtain a mean of 0 and a standard deviation of 1. As the InRH is approximately normally distributed under the null hypothesis of neutrality (Schlötterer & Dieringer 2005), loci with InRH values lower than -1.96 were considered outliers at the 0.05 threshold, indicating reduced variability in SMI and NI lineages as compared to SMD and ND, respectively. Only outliers in both InRH pairwise comparisons were considered, and common outliers to BayPass and InRH were retained to further investigation of their genic environment.

4.2.6 Genic environment of outlier loci

We searched for predicted genes located in a window of 120-kb around outlier loci (60-kb either side) in the annotated *G. pallida* genome. The choice of this value was based on the AFLP-based genetic linkage map of *G. rostochiensis* giving a physical:genetic distance ratio of 120 kb.cM⁻¹ (Rouppe van der Voort *et al.* 1999), as no data on the extent of recombination and linkage disequilibrium was available in *G. pallida*. As the majority of (a)virulence factors identified in plant parasitic nematodes are effectors (Haegeman *et al.* 2012; Mitchum *et al.*

2013), we searched as a priority for genes harboring N-terminal signal peptides in the predicted proteins, using SIGNALP v4.1 (Petersen *et al.* 2011). Moreover, the Phobius web server (Käll *et al.* 2007) was run on the predicted proteins showing a signal peptide in order to identify secreted proteins (*i.e.* without any predicted transmembrane domain). We also performed a BLAST search of the predicted genes against the annotated *G. rostochiensis* genome in order to potentially obtain more precise information on their functional annotation. Eves-van den Akker *et al.* (2016) identified in *G. rostochiensis* and subsequently in *G. pallida* genomes a dorsal gland promoter element motif, termed DOG Box, which may be a strong predictor of secretion, and thus likely effector function. We thus searched for the presence of the predicted genes identified around outlier loci in the *G. pallida* putative DOG effectors list established by Eves-van den Akker *et al.* (2016).

4.3 Results

4.3.1 Sequencing, mapping and SNP calling

Sequencing was carried out on duplicates of the lineages SMI, SMD, NI and ND. Despite the low amount of DNA used for each lineage, the generated sequencing data corresponded to the expectations from a quantitative and qualitative point of view. Sequencing produced two billion reads, *i.e.* 62.5 million reads/sample (biological replicate)/lane, and 76% mapped on the *G. pallida* reference genome, but 24% of these mapping reads mapped more than once on the genome. Duplicate filtering resulted in removing 13.5% of mapped reads. Remaining reads covered 95% of the ungapped assembly at an average depth of 34X/sample/lane. Processed mpileup contained 2,383,040 SNPs. PCA based on the allele frequency counts showed that the main axes of variation of genetic variability across the different samples were clearly associated with (i) the population origin and (ii) the virulence status. A very low variation was observed among the sequencing replicates in each pool sample. This prompted us to merge allele count data from the four sequencing replicates (lanes) for each of the eight pool samples (**Fig. S2**, Supporting information). The final dataset filtered on coverage and MAC consisted of read count data for 1,631,158 SNPs in the eight different selected lineages, corresponding to a density of polymorphic sites of 16/kb on average.

4.3.2 Genome scan for adaptive divergence and association with the virulent/avirulent status

Analysis of the dataset under the BayPass core model allowed us to estimate the scaled covariance matrix of population allele frequencies Ω that quantifies the genetic relationship among each pairs of populations. The resulting estimates of Ω accurately reflected the known structure between samples, *i.e.* a clustering at the higher level by population geographical origin, then by the virulence status and finally by biological replicate within each lineage (Fig. 4.1). *XtX* for each SNP were also estimated and calibrated by analyses of a POD containing 1,600,000 SNPs. At the 1% POD significance threshold, about 33,000 SNPs were identified as overly differentiated (Fig. 4.2). However, analysis under the auxiliary covariate model drastically reduced this outlier list, as less than 400 had also a BF_{mc} > 20 (355, 361 and 357 SNPs in each analysis, respectively) (Fig. 4.2). Overall, 275 outliers were shared by the three

analyses (outlier loci found in only one or two analyses actually exhibited XtX and/or BF_{mc} values close to the thresholds, which explains that they were not found in all analyses). Among these 275 SNPs showing an increase of genetic differentiation between Iledher and Désirée lineages, 31 of them also showed a decrease of genetic diversity in both SMI and NI lineages (Fig. 4.3) as identified by InRH.

4.3.3 Genic environment of outlier loci

The 31 selected outlier loci were distributed on 23 different scaffolds. Among these outlier loci for which the genic environment has been investigated, three outliers were located on scaffolds which do not harbor any predicted gene, 16 were located in intergenic area and 12 were directly located into genes (Table S1, Supporting information). Among the latter, five were located in exons, and four of them corresponded to synonymous mutations. The outlier SNP located on the scaffold 1777 corresponded to a non-synonymous mutation that changes an arginine into a histidine in GPLIN 001438500, which is annotated as a transcribed hypothetical protein. Despite the fragmentation of the G. pallida genome in numerous scaffolds that precluded a clear analysis of the physical distances among most of the identified outliers, it appeared that several outlier loci were sometimes found in the same scaffold. The most significant was the scaffold 988 which contains four outlier SNPs close to each other (< 200 bp). Four other scaffolds (44, 85, 182 and 283) harbored outlier SNPs close to each other (< 200 bp) and a fifth scaffold (66) harbored also two outlier SNPs but that are 75 kb apart. The number of outlier SNPs found in a scaffold was not dependent on the size of this scaffold: for instance, scaffold 988, which is one of the smallest scaffold of interest identified, contains four outlier loci. Overall, 258 predicted genes were identified in a 120-kb window around the 31 outlier loci (Table S1, Supporting information). About 47% (121) of them have unknown function in G. pallida, but 15 harbored a signal peptide, indicating that they potentially encode for secreted proteins, and one, GPLIN_000314000, was described similar to an effector of G. rostochiensis (Table S7 in Cotton et al. 2014). Fourteen predicted genes with known functions also harbored a peptide signal. Among the 29 proteins showing a signal peptide, 20 were predicted to be secreted (Table S1, Supporting information). Fifteen genes coding for proteins harboring a SPRY domain were found at the proximity of eight outlier loci, and three of them were paralog of SPRYSECs (Secreted protein with a SPRY domain), which are known in nematodes of the genus Globodera to be involved in pathogenicity but also in (a)virulence (Rehman et al. 2009; Sacco et al. 2009). The fact that a considerable part of the outlier loci are found near genes coding for a SPRY domain containing protein or even in or at the vicinity of SPRYSECs cannot be only explained by the huge size of this family gene or the prevalence of the SPRY domain in the G. pallida genome. In fact, 299 SPRY domain containing proteins are present in the G. pallida genome (which can be moreover cluster into islands), i.e. 1 for 334 kb, and 30 of them are SPRYSECs, i.e. 1 SPRYSEC for 3.33 Mb (Mei et al. 2015; Eves-van den Akker et al. 2016). When looking at the genic environment of outlier loci, we explored a total of 2 Mb of the genome and we identified 15 SPRY domain containing proteins, *i.e.* 1 for 133 kb, and 3 SPRYSECs, i.e. 1 for 660 kb. This is more than expected by chance in both cases, indicating that the highlighted candidate genomic regions were specifically enriched in genes coding for SPRYSEC proteins. Other genes were also interesting as they encode proteins which have functions linked to G. pallida pathogenicity, e.g., cell wall modifying proteins, or they encode proteins which are known to be secreted by the dorsal pharyngeal gland of G. pallida.

Some genes with unknown function in *G. pallida* had homologues with predicted function in *G. rostochiensis* genome, part of them being putatively paralog of SPRYSECs or involved in pathogenicity. Three predicted genes (GPLIN_000832500, GPLIN_001056700 and GPLIN_001258100) harbored DOG Boxes in their promoter region, two of them having unknown function and the third coding for a SPRYSEC protein.



Figure 4.1 Inferred relationship among the eight *G. pallida lineages* represented by a correlation plot and a hierarchical clustering tree derived from the matrix Ω estimated under the core model. Each sample name indicates its geographical origin (SM for Saint-Malo and N for Noirmoutier), the potato cultivar on which it evolved (I for Iledher and D for Désirée) and the replicate number (1 or 2).



Figure 4.2 SNP XtX as a function of the BF_{mc} for association with the virulence status covariable, estimated in one of the three independent analyses. The vertical dotted line represents the 1% POD significance threshold (XtX = 14.0) and the horizontal dotted line represents the 20 dB threshold for BF_{mc}. Black dots represent the 275 outlier loci of interest (*i.e.* outliers for both XtX and BF_{mc} values) and red dots represent the 31 outlier loci of strong interest (*i.e.* outliers also identified with InRH).



Figure 4.3 Example of increase of genetic differentiation (represented as *XtX* estimates computed in BayPass across the whole dataset) and decrease of genetic diversity (represented as expected heterozygosity H in each lineages having evolved on lledher) in a candidate genomic region. The scaffold represented is the scaffold 988 and the grey line indicates the location of the four outlier SNPs on the scaffold.

4.4 Discussion

In this study, we investigated *Globodera pallida* genomic regions involved in the adaptation to the QTL *GpaV*_{vrn}, a resistance factor against this potato parasite, by performing a genome scan on Pool-Seq data derived from experimentally evolved lineages, adapted or not to the resistant potato cultivar lledher.

Elucidating the molecular bases of virulence is a central challenge in nematology. So far, a few plant-nematode interactions have been studied to characterize the molecular determinants of (a)virulence. In G. pallida, only one virulence gene has been described (Sacco et al. 2009; Carpentier et al. 2012). A priori methods based on candidate genes to identify new or unknown virulence genes are therefore not the most adapted ones at the moment. Our study is the first to exploit genome scan to target genetic bases of adaptation to plant resistance in nematode populations. We demonstrated in a previous study the feasibility of a genome scan approach on our specific biological materiel coming from short experimental evolution (Eoche-Bosy et al. 2017), which was achieved here by the identification of 275 outlier loci and 31 candidates of a strong interest among them, distributed on 23 scaffolds. Several genomic regions are therefore putatively linked to adaptation to the resistance of lledher, however this number should be put in perspective with respects to the moderate quality of the G. pallida reference genome. In fact, the genome assembly is still relatively fragmented (6,873 scaffolds), which could suggest that different genomic regions identified as outliers in our study could in fact form only one or a few. Eves-van den Akker et al. (2016) highlighted that some effector islands identified in G. rostochiensis genome were split across different scaffolds in G. pallida genome. This prevents to make realistic assumptions on the number of genes that could be involved in the overcoming of lledher resistance. However, we assume that only a few virulence genes should be involved as few resistance genetic factors are expected in lledher because of the major effect of the QTL GpaV_{vrn} which is present in the potato cultivar lledher and the fast overcoming of this resistance observed in experimental evolution that support rather the view of a mono- or oligogenic resistance.

The moderate quality of the genome involved other limitations in mapping of sequencing reads, leading to the loss of a substantial part of the data. First, 24% of the sequencing reads did not map on the reference genome, while in the same time, the assembly contains 17% of N's and has a CEGMA completeness score of 74% for complete genes and 81% for partial genes (Parra et al. 2007; Eves-van den Akker et al. 2016). This suggests that a large part of unmapped reads could actually belongs to the G. pallida genome, but corresponding sequences would be absent from the assemblage. These missing sequences could putatively comprise candidate genomic regions to adaptation. Future works are now needed in order to analyze our data using a reference-free method to detect SNPs. For instance, the method developed by Uricaru et al. (2014) (i.e. DiscoSNP) could be used on the 24% of unmapped reads in order to identify new candidate SNPs and on the entire dataset in order to compare the results between genome scans performed using or not the reference genome. Second, heterozygosity of the reference genome is suspected, as about 24% of the sequencing reads mapped more than once on the genome, this hypothesis being supported by the difference between the length of the assembly (124.7 Mb) and the estimated genome size (100 Mb). Sequencing data mapping on these genomic regions were further discarded because of poor mapping score due to the multiple positions mapping, and once again, genomic information about these regions and their possible link to resistance adaptation were lost.

Pool-Seq is now commonly adopted in many studies, included population genomics studies (Rubin *et al.* 2010, 2012; Clément *et al.* 2013; Ferretti *et al.* 2013; Fisher *et al.* 2013), because it provides more accurate estimation of allele frequencies, at reduced sequencing and library preparation costs (Futschik & Schlötterer 2010; Zhu *et al.* 2012; Gautier *et al.* 2013; Rellstab *et al.* 2013; Schlötterer *et al.* 2014). Here, we emphasize the interest of using such an approach for population genomic studies of non-clonal microorganisms, as it allowed to increase the amount of DNA while keeping information about allele frequencies. The good quality of our results also relies on library preparation protocol, which fit the recommendations (Rhodes *et al.* 2014; Kofler *et al.* 2016). The TruSeq Nano DNA kit has been shown to be highly accurate and even some duplicates were obtained due to the PCR cycles, Kofler *et al.* (2016) showed that duplicates were also obtained with free-PCR library preparation kits. Nevertheless, Kofler *et al.* (2016) highlighted the superior impact of the genome quality over the library preparation protocol on the mapping quality. Numbers of individuals in pools and depth of coverage used have also been determined to improve sequencing results and allelic frequency estimations (Gautier *et al.* 2013; Kofler *et al.* 2016).

Consideration of a categorical population specific covariable in the genome scan allowed us to refine the list of outlier loci. Analysis of association further conducted by running the auxiliary covariate model allowed to retain only SNPs associated with the virulence status of the lineages, and not those corresponding to other selective pressures (e.g., local adaptation), which helped to reduce the risk of a misleading biological interpretation, a main limitation of genome scans (Pavlidis et al. 2012). Moreover, evaluating the reduction of diversity allowed us to refine the outlier loci list to those linked only to the adaptation to the resistant cultivar lledher. A direct link between those candidate loci and the adaptation to the resistance cannot yet be done at this stage. Indeed, footprints of selection detected in lineages having evolved on Iledher might well be linked to adaptation to the plant resistance or to the plant itself (including all its genetic background). Also, at this stage, we are unable to definitively link the identified SNPs to the resistance factor GpaV_{vrn} or to another resistance factor present in the genetic background of the cultivar lledher. Although these different situations will only be deciphering by functional validation, there is a strong possibility that at least some of the candidate genomic regions identified in this study are involved in adaptation to lledher resistance and in particular to its major resistance QTL GpaV_{vm}.

Sixteen over the 31 outlier loci investigated for their genic environment have at least one gene coding for a secreted protein in their neighboring region (even if all secreted proteins are not effectors, all effectors are secreted protein) and several outliers are at the vicinity of genes coding for proteins specifically secreted in the nematode dorsal pharyngeal gland. The pharyngeal glands are cells in which effectors are produced, and thus can be regarded as a toolbox for infection (Eves-van den Akker & Birch 2016). Cyst nematodes have two sets, subventral and dorsal: the former are primarily active while the nematode migrates through host tissue, while the latter are primarily active during the sedentary parasitic stages (Endo 1987; von Mende 1997; Davis et al. 2000). Some of the identified secreted proteins are SPRYSECs effectors which act as a versatile protein-binding platform for the nematodes to target a wide range of host proteins during parasitism including plant resistance proteins (Rehman et al. 2009; Sacco et al. 2009; Diaz-Granados et al. 2016). Two of them, located on scaffolds 182 and 782, were paralog of RBP-1, the only virulence gene identified in G. pallida (Sacco et al. 2009; Carpentier et al. 2012). It is likely that some of the other genes coding for proteins harboring a SPRY domain could actually be SPRYSECs, the absence of signal peptide being not ascertained as they are 5' truncated (N's region) in the reference genome. The fact that the highlighted candidate genomic regions were enriched in genes coding for a SPRY domain containing protein (1 / 133 kb for the 2 Mb explored versus 1 / 334 kb for the whole *G. pallida* genome) and in gene coding for SPRYSEC proteins (1 / 660 kb versus 1 / 3.33 Mb) supports the view that some SPRYSEC proteins should be well indeed involved in the adaptation of *G. pallida* to lledher.

It can be noticed that the InRH test also identified, among the 275 outlier loci, six SNPs showing a decrease of genetic diversity in both Désirée lineages (Table S2, Supporting information). These loci were distributed on four scaffolds, different from those on which the 31 outliers were located. Although we did not consider them in this study, those results could further be used to study adaptation to the potato cultivar Désirée. Only one of the eight microsatellite markers identified as outliers by Eoche-Bosy *et al.* (2017) was located on the same scaffold as one of the 31 outlier SNPs identified here, and no one was located on the same scaffold as the six outlier SNPs involved in Désirée adaptation. The microsatellite marker Gp235 was indeed located at about 150 bp from the outlier SNP identified here on the scaffold 216. The fact that the other outlier microsatellites were not found on the same scaffold that outlier SNPs is not surprising, due to the small number of microsatellites used and to the fragmentation of the genome.

By identifying genomic regions putatively involved in the adaptation to the resistance from S. vernei, the present study has taken a further step towards understanding and identifying the determinants of virulence. Even if those loci are not the mutation responsible for the adaptation, they can be enough linked to it to be used as molecular tools to determine virulence allele frequencies in field populations before the deployment of resistant cultivars, and to study more accurately the potential fitness costs or benefits linked to the mutation from avirulence to virulence. Indeed, a recent study suggested that adaptation to lledher involves an increase of fitness on a susceptible potato cultivar (Fournet et al. 2016). However, this study did not test for competition between avirulent and virulent individuals, because of the lack of molecular markers allowing to follow (a)virulence alleles. And it was shown in the Potato virus Y that a cost of competitiveness could occur, even in absence of fitness cost in simple inoculation (Janzac et al. 2010). Several approaches can be used to determine and follow virulence allele frequencies in natural populations. Among the simple SNP genotyping methods, the Allele-Specific PCR (AS-PCR) (Gaudet et al. 2009) allows SNPs to be detected in minimally equipped laboratories. This method relies on obtaining a PCR product specific to the SNP polymorphism using a common reverse primer and two forward allele-specific primers with different tails that will amplify two allele-specific PCR products of different lengths. Other methods will detect any base change at the SNP site of interest and do not require design of primers specific to each allele of interest. These usually correspond to highthroughput SNP genotyping technologies which require expensive equipment but have the advantage to provide accurate quantification of allele frequencies, detect any base change at the SNP site of interest and do not require primers specific to each allele. Pyrosequencing that entails sequencing of short stretches of nucleotides surrounding a known polymorphism in a pooled sample containing multiple alleles was mostly used till now. But other sequencing technologies can be used as pooled resequencing approaches were shown to be also well suited to determine allele frequencies (Rellstab et al. 2013). Next Generation Sequencing was used to assess changes in the frequency of different haplotypes in populations of the blood stage of Plasmodium falciparum (Takala et al. 2006) and to assess frequencies of virulent isolates in field populations of a plant pathogenic fungus, Leptosphaeria maculans in order to predict the risk of resistance breakdown in crops (Van de Wouw & Howlett 2012). Another interesting high-throughput SNP genotyping technology is the Sequenom MassARRAY platform that has several attractive features for users desiring an accurate custom SNP genotyping assay with modest multiplexing and minimal assay setup costs due to unmodified oligonucleotide primers. It utilizes a homogeneous reaction format with a single extension primer to generate allele-specific products with distinct masses, multiplexed PCR reactions, a single termination mix and universal reaction conditions for all SNPs (Gabriel *et al.* 2009).

Additional Supporting Information may be found in the Annexes section at the end of the thesis manuscript.

4.5 Acknowledgments

The authors gratefully acknowledge D. Mugniéry and M.C. Kerlan who took part in the experimental evolution, L. Renault, S. Bardou-Valette, C. Porte and D. Fouville who helped for the protocol elaboration and the pooled samples preparation, and M. Vidal for its help in libraries preparation. DEB is supported by the INRA department SPE and the Région Bretagne through a three year PhD grant. We also acknowledge two anonymous reviewers and the subject editor for their useful comments.

4.6 Data accessibility

Sequencing data were submitted to the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra/). BioSample accessions are: SAMN07249861 (SMD replicate 1), SAMN07249862 (SMD replicate 2), SAMN07249863 (SMI replicate 1), SAMN07249864 (SMI replicate 2), SAMN07249865 (NI replicate 1), SAMN07249866 (NI replicate 2), SAMN07249867 (ND replicate 1) and SAMN07249868 (ND replicate 2).

RÉSULTATS COMPLÉMENTAIRES : APPROCHE DE GÉNOTYPAGE PAR SÉQUENÇAGE

4.7 Introduction

Le scan génomique haute densité réalisé précédemment a permis de mettre en évidence les régions génomiques de *G. pallida* impactées par la sélection lors du contournement de la résistance de sa plante hôte. La réalisation d'une telle approche s'avère donc efficace pour répondre à l'objectif d'identification des bases génétiques de l'adaptation aux résistances variétales. Cependant, elle n'en reste pas moins coûteuse de par la nécessité d'un reséquençage de génome complet visant à identifier un large panel de marqueurs moléculaires densément répartis le long du génome. Il est difficilement envisageable pour le moment de réaliser une telle approche en routine pour l'identification systématique des bases génétiques de l'adaptation aux résistances variétales. Une autre stratégie couramment utilisée pour abaisser les coûts liés au séquençage est de séquencer une représentation réduite du génome. Il est par exemple possible de cibler uniquement l'exome, le transcriptome, ou encore une portion réduite du génome grâce à l'utilisation d'enzymes de restriction lors de la préparation des librairies (*reduced representation genomic libraries*, RRL, Van Tassell *et al.* 2008). En associant ces méthodes au *pooling* d'individus, il est alors possible de réduire drastiquement les coûts de séquençage.

Dans l'optique d'identifier une méthode assurant la réalisation d'un scan génomique de qualité à moindre coût, nous nous sommes en particulier intéressés à la méthode utilisant des RRLs. Nous avons ainsi souhaité savoir si un jeu de SNPs répartis de manière moyennement dense sur le génome pourrait être suffisant pour nous permettre de cibler assez précisément les régions génomiques sous sélection. De plus, une des limites inhérente à l'utilisation de RRLs est l'existence possible d'un polymorphisme dans le site de restriction pouvant biaiser l'estimation des fréquences alléliques, un phénomène appelé allele dropout (Arnold et al. 2013; Gautier et al. 2013). En effet, au niveau de tels sites, l'ADN ne sera coupé que chez une fraction des individus de la population, i.e. ceux présentant l'allèle complémentaire à l'enzyme de restriction. Or, si cet allèle est spécifiquement associé, par déséquilibre de liaison, à l'allèle d'un SNP à proximité, l'estimation des fréquences alléliques au niveau de ce SNP s'en retrouvera biaisée, puisque le deuxième allèle, bien qu'existant, ne sera jamais observé dans les séquences. Le phénomène d'allele dropout a ainsi pour conséquence de surestimer la variation génétique dans et entre populations, ce qui peut s'avérer très problématique dans le cas de l'utilisation de données de séquençage issues de RRLs pour la réalisation de scans génomiques, où la performance de l'analyse repose précisément sur la fiabilité de l'estimation des fréquences alléliques. Ce problème se révèle d'autant plus important dans les cas où la préparation des RRLs est réalisée à partir de pools d'individus, car les régions concernées par les biais, i.e. où des reads seront manquants, seront plus difficiles à identifier (Schlötterer et al. 2014). Gautier et al. (2013) ont montré que ce biais restait généralement faible, excepté chez les populations présentant de grandes tailles efficaces. Ce n'est pas le cas des populations naturelles de nématodes à kyste (Jan et al. 2016) où la taille efficace est de l'ordre de 100, cependant celle-ci n'a jamais été évaluée dans les agrosystèmes, qui favorisent par ailleurs les fortes densités d'individus. Nous souhaitions donc également évaluer l'impact de ce possible biais sur les résultats de scans génomiques.

Si l'utilisation de ce type de données pour la réalisation d'un scan génomique s'avère concluante, elle pourrait nous permettre d'étudier l'ensemble du matériel génomique issu de l'évolution expérimentale, afin d'affiner nos conclusions. Ainsi, nous pourrions également cibler les régions génomiques impactées par la sélection lors du contournement de la résistance d'autres génotypes qu'Iledher, afin de déterminer si l'architecture génétique de la virulence est la même ou si, au contraire, différents trajets adaptatifs peuvent exister lors du contournement de génotypes résistants présentant le même QTL de résistance mais des fonds génétiques différents, connus pour influencer la durabilité de la résistance (Fournet et al. 2013). Si les trajets adaptatifs sont les mêmes, il serait ainsi possible de dissocier, parmi les régions génomiques identifiées précédemment, celles réellement impliquées dans l'adaptation à GpaV_{vrn} de celles potentiellement impliquées dans l'adaptation au reste du fond génétique. Les lignées ayant évolué pendant sept générations sur la variété résistante lledher puis pendant une génération sur la variété sensible Désirée pourraient également être exploitées dans l'objectif de tester l'hypothèse de réversion de la virulence lorsque la pression de sélection est levée. Le niveau de virulence de ces lignées est en effet resté inchangé phénotypiquement, mais il est nécessaire de vérifier si les régions identifiées précédemment présentent toujours une différenciation significativement plus élevée que sous l'attendu d'un équilibre entre forces évolutives neutres. Si c'est bien le cas, alors cette approche de multiplication sur Désirée pourrait être utilisée pour identifier les facteurs de virulence impliqués dans des résistances bien plus durables que celle d'Iledher, issues des différentes sources de résistances telles que S. vernei, S. spegazzinii ou S. sparsipilum. En effet, les effectifs des lignées virulentes face aux résistances les plus durables (donc aussi les plus prometteuses) ne permettent actuellement pas de mener un scan génomique sans passer par une étape de multiplication sur une variété sensible.

L'équipe de Benjamin Mimee, du Centre de Recherche et de Développement en Horticulture (Saint-Jean-sur-Richelieu, Québec, Canada), a récemment développé et testé sur l'espèce *G. rostochiensis* une approche associant génotypage par séquençage (*Genotyping-By-Sequencing*, GBS) et Pool-Seq, afin d'identifier des marqueurs SNPs utilisés par la suite dans des études phylogénétiques (Mimee *et al.* 2015). Leur méthode est adaptée de la procédure de GBS décrite par Elshire *et al.* (2011), basée sur un séquençage haut-débit d'une RRL obtenue par digestion du génome par des enzymes insensibles à la méthylation, reconnaissant des séquences plutôt riches en GC, le tout concourant à cibler davantage les régions codantes du génome. Cette stratégie permet en une seule étape à la fois d'identifier de nombreux SNPs et de caractériser leurs fréquences dans différentes populations, le tout rapidement, à un coût restreint et sans nécessiter le recours à un génome de référence. Mimee *et al.* (2015) ont ainsi associé le GBS au Pool-Seq, en séquençant des échantillons préparés directement à partir de kystes de nématode. En effet, la polyandrie étant courante chez les nématodes à kyste, la diversité au sein d'un nombre même restreint de kystes reste élevée, et est considérée comme représentative de la diversité de la population (Éric Grenier, communication personnelle).

Une coopération engagée avec Benjamin Mimee et un séjour effectué dans son laboratoire de recherche ont ainsi permis de réaliser l'approche de GBS sur le matériel biologique issu de l'évolution expérimentale, dans le triple objectif de :

1. comparer les performances d'un scan génomique haute densité réalisé à partir de données issues d'un reséquençage complet de génome et celles d'un scan génomique moyenne densité réalisé à partir de données provenant de séquençage de RRLs ;

- 2. caractériser encore plus précisément les bases génétiques de l'adaptation aux résistances issues de *S. vernei* ;
- 3. aider au crible des régions candidates issues du scan génomique.

4.8 Protocole expérimental

4.8.1 Matériel biologique et extraction d'ADN

En raison des deux objectifs distincts de cette étude énoncés précédemment, deux jeux d'échantillons (JE) différents ont été préparés.

Pour répondre au premier objectif et afin de réaliser une comparaison stricte des méthodes, nous avons utilisé un jeu d'échantillons (JE1) en tout point similaire à celui utilisé dans l'étude précédente, *i.e.* huit pools de 300 individus issus de 300 kystes différents provenant des lignées issues des deux populations SM et N et ayant évolué sur lledher et Désirée, soit deux réplicas par lignée. En effet, même si l'étude précédente évoque la préparation de deux réplicas biologiques par lignée, nous avions en réalité préparé et procédé à l'extraction d'ADN du double d'échantillons, assurant ainsi une parfaite identité du matériel biologique entre les deux études.

Pour répondre au deuxième objectif, nous avons adapté le protocole de Mimee *et al.* (2015) en réalisant des échantillons de pools de kystes (JE2). Quarante kystes ont été échantillonnés dans chacune des lignées issues de l'évolution expérimentale (voir Fig. 1.21). Les kystes ont été éclatés, les débris de cuticules retirés et le pool de larves obtenu pour chaque lignée placé dans un tube d'eau stérile. La suite du protocole pour l'extraction d'ADN suit la même démarche que l'étude précédente. La quantité d'ADN obtenue à partir de ces pools de 40 kystes était en moyenne de 275 ng, ce qui est cohérent avec le rendement obtenu pour le pool de 300 larves (sachant qu'un kyste contient en moyenne 150 larves).

4.8.2 Préparation des librairies et séquençage

Les échantillons du JE2 ont été normalisés à 100 ng d'ADN. Étant donné la faible quantité d'ADN dans les échantillons du JE1, ceux-ci n'ont pas subi de normalisation. Les librairies ont été préparées en utilisant l'automate Ion ChefTM System, un kit de réactifs de séquençage Hi-QTM et des puces PI v3 (Thermo Fisher Scientific, Wilmington, DE, USA), selon la procédure décrite par Mascher *et al.* (2013), à la plateforme d'analyses génomiques de l'Institut de Biologie Intégrative et des Systèmes (IBIS, Université Laval, Québec, Canada), excepté le fait qu'un Blue Pippin (Sage Sciences, Beverly, MA, USA) a été utilisé pour réaliser la sélection de taille des fragments avant l'amplification PCR. Le couple d'enzymes *Mspl/Pst*I a été utilisé pour obtenir une représentation réduite du génome. Les librairies ont été barcodées puis multiplexées en proportions équimolaires, tout en traitant indépendamment les échantillons du JE1 et du JE2. Le séquençage a été réalisé sur une plateforme lon ProtonTM, à l'IBIS.

4.8.3 Traitement des données et SNP calling

Après démultiplexage avec le logiciel Sabre (https://github.com/najoshi/sabre), les données brutes issues du séquençage ont été traitées avec le logiciel Trimmomatic v0.32 (Bolger et al. 2014) afin de retirer les séquences des adaptateurs et d'éliminer les séquences inférieures ou égales à 32 pb (celles-ci ne représentant qu'en moyenne 1% du jeu de données). Le pipeline UNEAK, développé par Lu et al. (2013) et conçu pour identifier les SNPs sans génome de référence, a ensuite été utilisé en fixant un ETR (error tolerance rate) à 0.03, dans le but d'éliminer les erreurs de séquençage. UNEAK est un pipeline stringent dans lequel un seul SNP entre paires de séquences de 64 pb (préalablement tronquées à cette taille) est autorisé. Les SNPs ainsi identifiés ont ensuite été filtrés en fixant le taux minimum d'identification d'un SNP (i.e. le nombre de populations dans lequel le SNP doit avoir été détecté) à 0,80 (soit 20% de données manquantes autorisées) et la MAF (Minor Allele Frequency) à 0,01. L'implémentation actuelle d'UNEAK cesse d'enregistrer les occurrences d'un SNP une fois la couverture de 127 atteinte. Pour contourner cette limite, les comptages exacts de chaque allèle pour chaque SNP identifié par UNEAK ont été extraits à l'aide d'un script. Un filtre final similaire à celui réalisé dans l'étude précédente, portant sur la couverture et le MAC (Minimum Allele Count), a enfin été réalisé, en adaptant cependant les seuils maximum de couverture aux données GBS.

4.8.4 Détection des loci outlier

Les comptages alléliques obtenus ont été utilisés pour réaliser le scan génomique avec le logiciel BayPass, en suivant la même procédure que l'étude précédente.

4.8.5 Approche alternative pour la détection de SNPs

Le pipeline UNEAK étant relativement stringent, nous avons également souhaité savoir si un mapping des données GBS sur génome de référence permettrait de détecter un nombre de SNPs plus élevé. Les données démultiplexées ont ainsi été traitées suivant le même pipeline d'alignement et de détection de SNPs que l'étude précédente (voir partie **4.2.4**), excepté le fait que les données GBS ne sont pas pairées, ce qui a empêché la suppression des duplicas PCR (Sidhu *et al.* 2015). Le filtre final sur la couverture et le MAC ainsi que la détection de loci *outlier* ont été réalisé suivant le même protocole que celui utilisé pour les données obtenues avec le pipeline UNEAK.

4.9 Premiers résultats et discussion

L'analyse des résultats de l'approche GBS est toujours en cours et nous présentons ici seulement les premières tendances observées. Pour les deux JE, la quantité de données obtenues, et donc la couverture, était très hétérogène en fonction des échantillons. De manière générale, les marqueurs étaient faiblement couverts, et une part importante d'entre eux (entre 14 et 78% suivant les échantillons) a été supprimée de l'analyse suite aux filtres sur la couverture. Ainsi, un nombre final de 1865 SNPs a été obtenu pour le JE1 et 130 seulement

pour le JE2. Concernant ce dernier jeu de données, d'autres analyses UNEAK ne prenant en compte que des sous-parties du jeu de données (par exemple, uniquement les lignées ayant évolué sur lledher et Désirée) ont permis d'aboutir à la détection d'un nombre plus élevé de SNPs, cependant celui-ci ne dépassait jamais 1000. Pour le JE1, le fait de disposer de réplicas a permis de supprimer de l'analyse un des réplicas biologiques de la lignée SMD qui présentait le moins de données, permettant de doubler le nombre de SNPs détectés, en passant à 3611 SNPs. Globalement, la quantité de données obtenues et *a fortiori* de SNPs détectés est peu satisfaisante, en particulier pour le JE2. Elle pourrait être améliorée avec l'utilisation d'une autre enzyme, *ApeK*I, pour la préparation des RRLs. Une étude GBS réalisée sur d'autres données, hors cadre de la thèse, a en effet montré que pour un même échantillon initial, *ApeK*I permettait d'obtenir une quantité substantiellement plus importante de données que le couple d'enzymes *MspI/Pst*I. Cette proportion varie en fonction du filtre appliqué sur la couverture, de 14.8 fois plus de données à une profondeur de couverture de 50X.

Les premiers résultats du scan génomique suggèrent l'existence de phénomènes d'allele dropout dans les échantillons. En effet, en comparant les résultats obtenus avec le JE1 à ceux obtenus dans l'étude précédente, on constate qu'un pourcentage plus élevé de SNPs outlier a été détecté avec les données GBS (0.21%, soit 4 SNPs outlier, Fig. 4.4) qu'avec les données obtenues par reséquençage complet (0.017%). La position génomique de seulement la moitié des outliers ainsi détectés a pu être retrouvée (par mapping avec Bowtie2), mais aucun des deux n'a été retrouvé sur les mêmes scaffolds que les outliers détectés dans l'étude précédente. La problématique de l'allele dropout peut de manière générale être atténuée grâce à l'utilisation de logiciels détectant et corrigeant ces erreurs (e.g. GBStools, Cooke et al. 2016), ce qui sera par la suite appliqué à nos données. Pour détecter les régions génomiques impactées par l'allele dropout, ceux-ci se basent sur l'analyse de la couverture, puisqu'une diminution de couverture, due aux *reads* manquants associés à l'allèle nul, est attendue aux sites concernés (Gautier et al. 2013).

Le mapping des données GBS sur le génome de référence n'a pas permis d'améliorer sensiblement la détection de SNPs par rapport à l'utilisation d'UNEAK (2097 vs 1865 SNPs détectés). Outre la réalisation d'un nouveau séquençage avec une autre enzyme, une solution pour augmenter le nombre de SNPs détectés dans les jeux de données actuels serait éventuellement d'utiliser un autre pipeline qu'UNEAK. Dans notre cas, des pipelines utilisant un génome de référence et dont la meilleure performance dans la détection de SNPs a été démontrée (Torkamaneh *et al.* 2016) seraient applicables.

Pour connaitre la portion du génome analysée par cette approche GBS, les motifs des deux enzymes ont été recherchés dans le génome, révélant qu'environ 20% du génome était représenté dans les données de séquençage. Cette analyse n'a pas encore été réalisée pour l'enzyme ApeKI.

De manière générale, même si quelques critiques ont été émises, il ne faut pas omettre que les résultats obtenus sont très dépendants de l'enzyme utilisée, plutôt que de l'approche GBS globale. Tester le GBS avec différentes enzymes est donc une étape nécessaire avant de pouvoir conclure sur l'intérêt ou non de données GBS pour la réalisation de scans génomiques.



Figure 4.4 BF_{mc} (association avec la covariable statut virulent/avirulent de la lignée) en fonction de XtX. La ligne pointillée verticale représente le seuil de significativité à 1% (XtX=15.9) calculé à partir du POD et la ligne pointillée horizontale représente le seuil à 20 dB pour BF_{mc}. Les ronds noirs représentent les 4 loci *outlier*.

CHAPITRE 5

DISCUSSION GÉNÉRALE

CHAPITRE 5 DISCUSSION GÉNÉRALE

5.1 Synthèse

L'amélioration de la durabilité des résistances repose notamment sur la caractérisation d'éléments prédicteurs de la vitesse d'adaptation du pathogène, tels que le potentiel évolutif des populations ou encore l'existence de coûts de virulence. Ces éléments restent néanmoins des prédicteurs indirects, largement dépendants d'autres paramètres plus difficiles à quantifier (flux de gènes, taille des populations, dérive génétique, interactions compétitives) et donc entachés d'une marge d'erreur conséquente. L'identification des déterminants génétiques de l'adaptation permettrait de franchir un nouveau cap vers une gestion durable des résistances, en rendant possible, de manière beaucoup plus directe et exacte, la caractérisation du potentiel de contournement de chaque population de pathogène vis-à-vis d'une source de résistance.

Dans le cadre de cette thèse, nous nous sommes intéressés au nématode à kyste de la pomme de terre *Globodera pallida*, une espèce de quarantaine causant des pertes de rendement considérables sur une culture d'intérêt économique majeur à l'échelle planétaire. Peu de variétés de pomme de terre de consommation résistantes à ce nématode sont aujourd'hui disponibles, et leur utilisation est menacée par la capacité des populations de *G. pallida* à contourner ces résistances.

L'objectif de cette thèse était de mieux caractériser les bases génétiques d'une telle adaptation et ses conséquences sur la *fitness* du nématode, afin d'anticiper au mieux les contournements futurs et de préserver les résistances existantes. La stratégie d'étude mise en place a combiné plusieurs volets disciplinaires, tels que la nématologie, la biologie évolutive, la génétique et la génomique des populations, mobilisé plusieurs échelles d'étude, du phénotype au génotype, de la population au gène, et nécessité l'appropriation d'outils et de techniques moléculaires ainsi que de méthodes novatrices en génomique des populations.

Cette thèse aura ainsi permis de :

- 1. mettre en évidence les traits d'histoire de vie de *G. pallida* impactés ou non lors de l'adaptation à la résistance issue de *S. vernei* et d'identifier un bénéfice de *fitness* associé à la virulence ;
- 2. valider la faisabilité d'une approche de scan génomique sur un matériel biologique issu d'un processus court de divergence en évolution expérimentale ;
- 3. identifier chez le nématode des régions génomiques candidates à l'adaptation à la résistance de la plante hôte ;
- 4. tester une méthode alternative au scan génomique haute densité, dans l'objectif de généraliser et faciliter l'accès aux bases génétiques de l'adaptation aux résistances.

Dans ce chapitre final, nous reviendrons sur les originalités de notre étude, nous présenterons et discuterons les principaux résultats acquis tout au long de ce projet de recherche, ainsi que les limites auxquelles nous avons été confrontés, avant d'évoquer les perspectives de recherche et les implications agronomiques qui en découlent. Ce projet de thèse a permis de générer une quantité conséquente de données et a ouvert de nouvelles questions de recherches. Nous évoquerons ainsi également au travers de ce chapitre les résultats en cours d'acquisition et les études annexes envisagées.

5.2 Intérêt de la stratégie d'étude

5.2.1 Intérêt de l'étude des déterminants et des conséquences de l'adaptation pour une gestion durable des résistances

Prédire la vitesse d'adaptation des populations de pathogènes aux résistances de leur plante hôte est une étape indispensable dans l'optique d'une gestion durable des résistances variétales. Un premier prédicteur de la durabilité des résistances a été proposé par McDonald & Linde (2002) et déduit le potentiel évolutif des pathogènes de leurs caractéristiques biologiques et génétiques (Tableau 5.1). Selon ces auteurs, les agents pathogènes présentant de grandes capacités de dispersion, d'importantes tailles de populations et un mode de reproduction mixte (mêlant reproduction sexuée et asexuée) sont ceux possédant les plus grandes capacités adaptatives à une résistance génétique, c'est-à-dire le plus grand potentiel évolutif. D'après ces paramètres, les nématodes à kyste se présentent comme des pathogènes à faible potentiel évolutif, en grande partie du fait de leur très faible capacité de dispersion active. La taille efficace de leurs populations est de plus estimée comme étant relativement faible, de l'ordre de 100 (Jan et al. 2016), cependant celle-ci n'a été estimée qu'en milieu sauvage. Le milieu cultivé étant propice à la présence de fortes densités d'individus, elle pourrait y être beaucoup plus conséquente. Des études récentes montrent en outre que la migration des nématodes a été sous-estimée. En effet, il a été mis en évidence que les activités humaines pouvaient favoriser une dispersion passive des individus sur de grandes distances, de l'ordre de 150 km (Plantard & Porte 2004), et que les populations de nématodes se définissaient à l'échelle d'un bassin de production ou d'une coopérative agricole (Picard & Plantard 2006; Alenda et al. 2014) (Fig. 5.1). Ceci suggère finalement que les nématodes à kyste pourraient présenter un potentiel évolutif beaucoup plus important que ce qui avait été envisagé précédemment, prédisant ainsi une durabilité des résistances plus réduite. En effet, les grandes tailles de populations minimisant l'action de la dérive génétique, un allèle (de virulence) rare aura de grandes chances d'être conservé dans le temps, et c'est semble-t-il le cas des allèles de virulence dans les populations européennes de G. pallida. D'autre part, la dispersion passive des nématodes sur de grandes distances favorise l'introduction d'allèles (de virulence) rares dans des populations avirulentes. Le prédicteur proposé par McDonald & Linde (2002) ne permet qu'une estimation indirecte et globale (*i.e.* ne permettant pas de prédire la durabilité de résistances prises individuellement), et reste largement perfectible. Avec l'intensification de l'utilisation de la résistance variétale en agriculture, disposer de prédicteurs plus fiables et donc plus directs constituerait une avancée indéniable en matière d'utilisation durable des ressources génétiques.



Tableau 5.1 Tableau adapté de McDonald & Linde (2002) permettant d'estimer le potentiel évolutif d'un pathogène et donc le risque qu'il représente pour la durabilité des résistances, en fonction de son système de reproduction, de ses capacités de dispersion et de la taille efficace (Ne) de ses populations. 1 = petite Ne, 2 = Ne moyenne, 3 = grande Ne. Les chiffres à l'intérieur des cadres correspondent à l'échelle de risque. La croix verte indique l'emplacement théorique des nématodes à kyste, tandis que la croix orange indique leur emplacement potentiel dans les agrosystèmes, en prenant en compte la dispersion passive due aux activités humaines et en considérant une taille de population plus importante, favorisée dans les agrosystèmes.



Figure 5.1 Influence des pratiques agricoles sur la structure génétique des populations de *Globodera tabacum* (Alenda *et al.* 2014). Les populations se répartissent en deux clusters génétiques différents, dont la localisation correspond aux limites géographiques de deux coopératives agricoles.

Les études portant sur les facteurs d'avirulence chez d'autres agents pathogènes, et en particulier les virus, ont permis de mettre en évidence l'existence de relations entre durabilité des résistances et caractères impliqués dans le phénomène d'adaptation chez le pathogène.

D'une part, l'existence d'une relation entre durabilité des résistances et coûts de fitness associés à la mutation de l'avirulence à la virulence (Leach et al. 2001; Jenner et al. 2002) a été démontrée et s'est avérée importante pour la prédiction de la durabilité des résistances chez plusieurs pathosystèmes plante-virus (Lecoq et al. 2004; Boyd 2006; Fabre et al. 2009). En effet, nous avons vu que les gènes d'avirulence étaient d'abord impliqués dans diverses fonctions chez l'agent pathogène, pouvant être perdues lors de la mutation conduisant à la virulence (Vera Cruz et al. 2000). Chez les nématodes à kyste, nous avons évoqué le fait que les gènes d'avirulence étaient très probablement des gènes ayant une fonction dans le pouvoir pathogène, *i.e.* des effecteurs. Or, ces effecteurs sont nécessaires au parasitisme des hôtes résistants comme des hôtes sensibles, sur lesquels les individus virulents se retrouvent en compétition avec les individus avirulents. Si la virulence s'accompagne d'une perte ou d'une diminution de l'efficacité de la fonction assurée par le gène dans le parasitisme, les virulents peuvent alors être contre-sélectionnés par rapport aux avirulents sur un hôte sensible. Si l'effecteur a une importance majeure pour la fitness de l'organisme, le coût de la virulence peut être trop fort, même sur un hôte résistant et en l'absence de compétition, aboutissant là encore à la contre-sélection des virulents. On voit donc que la caractérisation des coûts de virulence est une étape importante dans la prédiction de la durabilité des résistances, qui peut aussi être exploitée pour la mise en place de stratégies de gestion des résistances. Dans le cas d'une diminution de la fitness des virulents sur un hôte sensible, l'alternance du déploiement de variétés résistantes et sensibles peut permettre de maintenir les populations de nématodes à des niveaux relativement bas, en contre-sélectionnant séquentiellement les individus avirulents (par utilisation de variétés résistantes) et les individus virulents (par utilisation de variétés sensibles). Dans le cas où le gène d'avirulence est impliqué dans une fonction primordiale chez le nématode et que la mutation aboutit à une perte trop importante de fitness, même sur hôte résistant, une utilisation continue de la résistance correspondante peut être assurée, en ne risquant probablement jamais un contournement de celle-ci. Au contraire, la mutation vers la virulence peut parfois n'être associée à aucun coût de virulence, et, plus rarement, être associée à un gain de fitness. Cette configuration s'avère plus problématique en termes de durabilité des résistances et dans ce cas, d'autres leviers sont à mobiliser. Il pourrait notamment être judicieux de favoriser des actions pouvant agir sur la balance sélection-dérive, en faveur de la dérive génétique, autrement dit des actions favorisant de petites tailles de populations et limitant les flux de gènes. En pratique, cela se traduit par un allongement des rotations culturales, l'utilisation d'autres résistances, une limitation des flux par la mise en place de mesures prophylactiques telles que l'utilisation de plants certifiés ou le nettoyage du matériel agricole.

D'autre part, l'étude des facteurs d'avirulence chez les virus a également souligné la relation existant entre la durabilité des résistances et le nombre et la nature (transition/transversion) des mutations impliquées dans l'adaptation (Fabre *et al.* 2009). En effet, plus le nombre de mutations nécessaires au contournement de la résistance est grand, plus les chances pour que ces mutations émergent simultanément dans une population de pathogènes sont minces. Ayme *et al.* (2006) ont montré que la virulence du PVY face au gène de résistance *pvr2*³ pouvait être déterminée par plusieurs mutations au sein de la VPg

(genome-linked viral protein), mais qu'une seule suffisait pour aboutir à la virulence, expliquant ainsi pourquoi cette résistance est facilement contournée. En revanche, la virulence du PVY face au gène *Pvr4* n'est déterminée que par une seule et unique mutation possible au niveau de la région codant pour la protéine NIb, et sa virulence face au gène de résistance *pvr2*² nécessite deux mutations au sein de la VPg, ce qui explique pourquoi ces résistances restent durables malgré leur déploiement depuis plus de 25 ans (Janzac *et al.* 2010; Fabre *et al.* 2012). Ainsi, caractériser le nombre de gènes impliqués dans le contournement de la résistance, le nombre de mutations au sein de ces gènes et la nature de ces mutations peut permettre d'évaluer la durabilité d'une résistance. Chez les nématodes à kyste, une fois les bases génétiques de la virulence déterminées, il serait possible d'aller encore plus loin dans l'anticipation des risques de contournements en caractérisant systématiquement au champ, à l'aide de marqueurs moléculaires, les fréquences d'individus virulents à différentes résistances, permettant ainsi d'affiner la stratégie de déploiement des résistances.

5.2.2 Intérêt de l'évolution expérimentale

Dans les études visant à déterminer les causes et conséquences de mécanismes, tels qu'ici l'adaptation, l'utilisation de populations naturelles peut entraîner un biais lors de l'interprétation des résultats. Des raccourcis entre paramètres mesurés et causes probables peuvent en effet être facilement faits, qui n'ont pourtant aucun fondement biologique.

Dans la partie de cette étude consacrée à la caractérisation des conséquences de l'adaptation sur la fitness de l'organisme, nous avons comparé plusieurs valeurs de traits d'histoire de vie chez des lignées avirulentes et virulentes à la résistance issue de S. vernei, provenant d'une même population initiale. Comme les deux lignées ont subi le même processus d'évolution expérimentale, les différences observées peuvent très probablement être attribuées à l'adaptation à la résistance de la plante hôte. En revanche, si nous avions considéré deux populations naturelles connues pour présenter des niveaux de virulence contrastés face à une plante hôte résistante, les différences phénotypiques observées auraient pu être attribuées à la différence de statut de virulence de ces populations, alors qu'elles pourraient être le résultat d'autres processus adaptatifs (i.e. adaptation locale par exemple). Pour pallier cette faiblesse des études utilisant des populations naturelles, il est de rigueur d'augmenter l'effort d'échantillonnage en prenant en compte plusieurs populations dans différents environnements, ce qui n'est pas toujours réalisable. L'évolution expérimentale représente donc une bonne alternative qui permet de contrôler efficacement les effets confondants. Néanmoins, il aurait été préférable d'avoir pu étudier au moins une seconde population indépendante pour conforter nos résultats. Dans le cas présent, les différences observées pourraient en effet être causées par le simple effet aléatoire de la dérive génétique.

Dans le cadre de l'approche de scan génomique, l'utilisation de ce matériel issu d'évolution expérimentale a présenté des avantages certains. L'étude simultanée de deux populations différentes, c'est-à-dire de deux réplicas indépendants, a en effet permis de donner du poids aux résultats répétés. De plus, les conditions environnementales homogènes dans lesquelles ont évolué ces lignées durant les huit années de l'expérience, conditions qui étaient également contrôlées durant les cinq dernières années, a permis de conforter l'hypothèse que la détection d'un signal dans une région génomique (en comparant une lignée virulente avec une lignée avirulente) était effectivement liée à l'adaptation à l'hôte

uniquement et non à un autre facteur environnemental confondant, comme cela peut être le cas lors d'études similaires sur des populations naturelles. De même, il était important que le témoin de l'étude subisse les mêmes contraintes que les échantillons auxquels il a été comparé, afin de différencier les effets dus à la sélection répétée imposée par les génotypes résistants de ceux liés au processus de multiplication en laboratoire. Enfin, *G. pallida* semblait être un modèle adapté à la réalisation de scans génomiques. En effet, pour que le scan soit précis et qu'il aboutisse à l'identification de régions génomiques restreintes, les organismes doivent présenter un fort brassage génétique. *C*'est le cas avec *G. pallida* qui se reproduit obligatoirement de manière sexuée et autorise la polyandrie.

Certains éléments liés à ce protocole expérimental restent néanmoins à considérer dans l'optique d'une approche de scan génomique. L'évolution expérimentale a provoqué une diminution forcée du taux d'infestation du sol dans chaque population, entre la cinquième et la sixième génération, particulièrement dans les lignées issues de la population N. En effet, les niveaux d'infestation sont restés constants au champ et en bac, mais la sélection de 10 kystes par pot à partir de la génération 6 a fait diminuer le nombre de J2 par gramme de sol. Comme nous l'avons vu, un tel goulot d'étranglement peut potentiellement engendrer une perte de diversité génétique et la fixation aléatoire d'allèles, mimant alors un évènement de sélection. En revanche, celui-ci a aussi pu être contrebalancé par un autre aspect du protocole : pour la production de chacune des trois dernières générations, l'ensemble des kystes de la génération précédente a été mélangé, avant d'y échantillonner les lots de 10 kystes. Alors que nous avons vu que dans la nature, les capacités de dispersion active du nématode sont très limitées et que celui-ci ne se déplace que de quelques centimètres autour de la plante, nous avons mélangé ici les kystes provenant d'un bac de 20L puis de quatre plantes différentes, ce qui a pu permettre de limiter les effets de la dérive génétique. De plus, nous avons également vu que la diversité génétique d'une population de nématode était retrouvée au niveau d'une plante. Ceci suggèrerait plutôt l'absence d'une perte de diversité génétique associée à la diminution de densité d'individus lors du passage en pots. La forte pression de sélection imposée par la résistance de la plante hôte au début de l'expérience a également provoqué dans toutes les lignées une diminution importante des tailles de populations. Ces réductions de taille de populations pouvant provoquer la détection de faux positifs dans les scans génomiques, elles restent des éléments importants à prendre en compte, de même que le déficit en hétérozygotes connu chez G. pallida et ayant les mêmes conséquences. Il est donc, dans ce contexte, d'autant plus important de disposer de réplicas biologiques, puisque la probabilité de fixation aléatoire d'un même allèle par dérive dans deux populations indépendantes est extrêmement faible.

Plusieurs espèces de nématodes à kyste et à galle ont déjà fait l'objet d'évolutions expérimentales visant à mieux caractériser la virulence, comme *G. pallida* (Turner 1990; Phillips *et al.* 1998), *Heterodera glycines* (Bekal *et al.* 2003), *H. schachtii* (Müller 1992) ou encore *Meloidogyne incognita* (Castagnone-Sereno *et al.* 2007). Notre étude pourrait ouvrir la voie à la réalisation d'études similaires visant à identifier les bases génétiques de ces contournements, et tirant profit du matériel issu d'évolution expérimentale déjà disponible. Le mode de reproduction parthénogénétique d'une grande partie des nématodes à galles pourrait se présenter comme une limite à la réalisation de scans génomiques sur ces espèces, cependant celui-ci ne s'oppose pas à une importante variabilité génétique chez ces espèces, déjà mise en évidence auparavant (Castagnone-Sereno & Danchin 2014).

5.3 Principaux résultats et limites soulevées

5.3.1 Hypothèse d'un allèle de virulence préexistant dans les populations

L'évolution expérimentale a mis en évidence la capacité de deux populations françaises de G. pallida à contourner la résistance issue de S. vernei. Sur la base de plusieurs éléments énoncés ci-après, nous émettons l'hypothèse que la (ou les) mutation responsable de la virulence préexistait dans les deux populations et qu'elle a été héritée d'une population ancestrale. C'est tout d'abord l'hypothèse la plus parcimonieuse, principe élémentaire en science, pour expliquer la présence d'une (potentielle) même mutation dans deux populations indépendantes. En outre, certaines populations de G. pallida présentes dans le berceau d'origine de l'espèce sont virulentes à la résistance issue de S. vernei (Hockland et al. 2012), renforçant l'hypothèse d'une mutation ancestrale. Les niveaux de virulence sont les plus forts au nord du Pérou et diminuent progressivement du nord au sud, en suivant la répartition des clades de l'espèce. Ainsi, les populations du clade I, d'où sont originaires les populations européennes, sont celles présentant les niveaux de virulence les plus faibles sur lledher, comparables à ceux observés dans les populations européennes. Cela suggère que l'allèle de virulence serait peu présent dans ce clade, et explique pourquoi certaines lignées de l'évolution expérimentale, dans lesquelles l'allèle devait être présent en fréquences très faibles voire nulles, ont été perdues. Cela démontre de plus, qu'au moins dans ces lignées, l'allèle n'a pas pu apparaître spontanément. En outre, l'apparition spontanée d'un allèle de virulence dans autant de lignées indépendantes (ayant contourné la résistance) et sur un nombre aussi faible de générations est difficile à envisager, d'autant que les nématodes présentent un taux de mutation relativement faible, estimé à 3,1 (± 1,9) x 10⁻⁹ mutations/site/génération chez Caenorhabditis elegans (Denver et al. 2009). On observe également chez G. rostochiensis des cas de populations incapables de s'adapter aux résistances, à l'image des populations anglaises restées avirulentes au gène majeur de résistance H1 malgré son utilisation depuis plus de 30 ans, au contraire des populations hollandaises qui l'ont contourné, suggérant que les populations anglaises ont été introduites sans l'allèle de virulence.

5.3.2 Conséquences de l'adaptation sur la fitness du nématode

Le concept de compromis évolutif laissait présager que l'adaptation à la résistance issue de *S. vernei* serait associée à un coût de *fitness* des individus virulents sur hôte sensible. Ceci a été classiquement observé dans différentes interactions hôtes-pathogènes (Laine & Barrès 2013), et serait cohérent avec la faible proportion d'individus virulents présents initialement dans les populations, ayant conduit à une réduction drastique du nombre de femelles dans les lignées confrontées aux génotypes de pomme de terre résistants en début d'évolution expérimentale. Étonnamment, c'est l'inverse que nous avons constaté, en identifiant un gain de *fitness* apparent chez les individus virulents. Ce résultat a soulevé plusieurs hypothèses cherchant à expliquer l'absence de coût et au contraire le gain associé à la virulence, mais a surtout questionné sur le fait que la mutation ne se soit pas plus répandue dans la population d'origine. Premièrement, la disponibilité de marqueurs moléculaires permettrait de vérifier l'absence de coût de compétitivité, paramètre que le protocole expérimental utilisé a volontairement éliminé en ne travaillant qu'en inoculation simple.

Plusieurs études sur champignons et virus ont en effet démontré l'existence de coûts de compétitivité dans des expériences de compétition confrontant des individus virulents et avirulents sur hôte sensible, ayant mené à une contre-sélection des individus virulents, même lorsqu'aucun coût de virulence n'était constaté en inoculation simple (Abang *et al.* 2006; Bahri *et al.* 2009; Janzac *et al.* 2010).

Il sera possible d'affiner les résultats de cette étude lorsque les bases génétiques de la virulence auront été identifiées. Ainsi, nous pourrons vérifier si les gènes candidats à l'adaptation permettent d'expliquer le gain de *fitness* constaté, en étudiant la cohérence entre la fonction de ces gènes et les modifications phénotypiques observées chez les individus virulents.

Une des explications proposées justifiant l'absence d'un coût de *fitness* est que celuici aurait déjà été compensé par une ou plusieurs autres mutations. Si le gain de *fitness* s'explique en effet par une mutation compensatrice ayant émergé et s'étant répandue dans les lignées virulentes suite à l'adaptation, celle-ci devrait aussi être détectée par scan génomique. Il restera cependant difficile de différencier dans le jeu d'*outliers* cette mutation des autres mutations potentiellement impliquées dans la virulence. Il serait en outre intéressant de vérifier si un gain de *fitness* est également observé dans les lignées virulentes aux autres génotypes de pomme de terre résistants portant *GpaV_{vrn}*, et si les traits sont impactés de la même manière.

5.3.3 Génomique de l'adaptation

Le cœur de ce projet de thèse reposait sur l'identification des régions génomiques de G. pallida impliquées dans l'adaptation aux résistances de sa plante hôte. La confirmation de de la faisabilité d'un scan génomique sur des lignées ayant subi un processus court de divergence en évolution expérimentale était un prérequis nécessaire. Dans cet objectif, un premier scan génomique faible densité avec marqueurs microsatellites a été réalisé, et a permis de confirmer qu'une faible proportion de loci outlier était détectée. Le brassage génétique réalisé au cours de huit générations a ainsi permis de restreindre suffisamment le signal de sélection au sein du génome. Ceci tend à confirmer l'hypothèse d'une mutation préexistante dans les populations, qui a déjà pu recombiner avec les régions avoisinantes et donc mener à un balayage sélectif localisé dans une région restreinte du génome (Przeworski et al. 2005). En outre, l'identification de loci outlier parmi les 53 marqueurs microsatellites a permis de les soustraire du jeu de données afin de pouvoir estimer de manière fiable (i.e. en n'utilisant que des marqueurs neutres) les indices de diversité et de différenciation génétique sur les lignées, dans le but de mieux caractériser notre matériel biologique. Nous avons ainsi pu montrer que les deux populations SM et N étaient suffisamment différenciées pour constituer deux répétitions indépendantes dans nos analyses, et nous avons également mis en évidence un déficit en hétérozygotes dans les populations, ce qui n'était pas surprenant, car classiquement observé dans les populations naturelles de nématodes à kyste (Montarry et al. 2015). Ce premier scan génomique faible densité a également permis de montrer que notre dispositif expérimental, à savoir l'étude de deux populations indépendantes, était efficace pour contourner le risque (bien réel, voir partie 5.2.2) de détection de faux positifs, en identifiant deux outliers communs aux deux populations indépendantes. Un deuxième scan génomique haute densité a ensuite permis la détection de 31 SNPs candidats répartis sur 23 des 6873 scaffolds que compte le génome de G. pallida (Cotton et al. 2014).

L'évolution des outils de la génomique des populations (tests et marqueurs) se reflète dans notre étude. En effet, nous avons dans un premier temps utilisé des marqueurs microsatellites peu densément répartis dans le génome, pour ensuite préciser notre approche avec des SNPs obtenus par reséquençage complet de génome. De même, nous avons dans un premier temps exploité des tests basés sur les F_{ST} , pour ensuite privilégier le test le plus récent faisant intervenir des covariables permettant d'affiner les analyses.

La démarche de détection des outliers entreprise, associée à notre dispositif expérimental, nous a assurés de cibler précisément les SNPs impliqués dans l'adaptation à Iledher. Lors du premier scan génomique faible densité, nous avions en effet montré que les outliers détectés pouvaient aussi bien être liés à l'adaptation à lledher qu'à l'adaptation à Désirée, les populations n'ayant en effet jamais été confrontées à ces deux génotypes de pomme de terre. Néanmoins, le recours à l'évolution expérimentale nous avait déjà permis d'éliminer d'autres facteurs potentiels d'adaptation. Dans les populations naturelles, de nombreux effets environnementaux confondants peuvent rendre plus difficile l'interprétation des loci outlier. Afin de nous intéresser uniquement aux loci sous sélection dans les lignées virulentes, nous avons ciblé les loci présentant une diminution de diversité allélique dans ces lignées, grâce au test InRH. Ce test s'est avéré efficace pour éliminer les loci impliqués dans l'adaptation à Désirée, cependant nous avons récemment constaté que cette approche pouvait aussi exclure une partie des loci impliqués dans l'adaptation à lledher. En effet, le test InRH est basé sur le calcul d'un rapport d'hétérozygoties attendues estimées au sein de chaque lignée. Cette approche est efficace pour détecter des outliers lorsque le locus concerné présente une hétérozygotie moyenne dans une des populations (non impactée par la sélection), et une hétérozygotie plus faible dans la population sous sélection. Cependant, un SNP peut présenter des fréquences alléliques totalement opposées entre deux populations (par exemple, p=0.9 et q=0.1 dans la population 1, et p=0.1 et q=0.9 dans la population 2), signe que ces populations sont sous sélection divergente, et pourtant l'hétérozygotie calculée pour chacune des populations sera égale et mènera InRH à considérer ce SNP comme non impacté par la sélection. Une réanalyse plus approfondie des listes d'outliers est donc planifiée, afin de vérifier si de bons candidats n'auraient pas été éliminés par la première analyse.

Même si nous avons pu cibler quelques régions candidates à l'adaptation à lledher, l'étude de scan génomique présente ici ses limites. En premier lieu, nous n'excluons pas le fait que des faux positifs figurent encore parmi les candidats retenus. En effet, même si le dispositif expérimental apparaît être relativement optimal pour une étude de scan génomique, il n'en présente pas moins les quelques défauts que nous avons évoqués plus haut. S'ajoutent à cela les contraintes imposées par l'organisme d'étude, et notamment le déficit en hétérozygote, qui peut également influencer le taux de faux positifs. Parmi les vrais positifs, nous n'avons pas définitivement ciblé les loci impliqués dans l'adaptation à la résistance issue de S. vernei, mais plutôt les loci impliqués dans l'adaptation à la plante hôte, dans sa globalité. Ainsi, certains des candidats pourraient en fait être impliqués dans l'adaptation aux caractéristiques générales de la variété de pomme de terre (e.g. pH, taux d'humidité, épaisseur des racines), d'autres pourraient aussi bien être impliqués dans l'adaptation à d'autres résistances que la résistance ciblée par notre étude. Il pourrait alors soit s'agir de résistances déjà présentes dans le fond génétique, soit de résistances provenant de S. vernei et introgressées lors des croisements successifs. En particulier, la présence du QTL mineur de résistance dans les différents génotypes résistants n'a jamais été vérifiée. Sa présence au sein d'Iledher et a fortiori l'adaptation possible des individus à ce QTL n'est donc
pas à exclure. Les prochaines étapes de tri et de validation, abordées plus loin, permettront certainement de dissocier parmi ces différents loci sous sélection ceux impliqués dans le contournement de la résistance de ceux conférant une plus grande valeur adaptative sur les génotypes de pomme de terre contournés.

La confrontation des *outliers* mis en évidence par les scans génomiques faible et haute densité a montré qu'un seul *outlier* microsatellite était situé à proximité d'un des SNPs *outlier* identifiés par BayPass. Cependant, il n'est pas étonnant de ne pas avoir obtenu plus de corrélation entre les résultats des deux analyses, étant donné le faible nombre de marqueurs microsatellites utilisés et le niveau de fragmentation relativement important du génome de référence.

La qualité moyenne de ce génome a pu poser certaines limites quant à une exploitation optimale de nos données et à l'interprétation des résultats du scan génomique. D'une part, un quart des données de séquençage n'a pas pu être aligné sur le génome de référence. Nous avons émis l'hypothèse que ceci était certainement dû aux 17% de régions manquantes dans le génome de G. pallida. Nous avons néanmoins cherché à caractériser plus en détails ces reads non alignés, et avons pour cela utilisé l'outil Kraken (Wood & Salzberg 2014), qui permet d'assigner un taxon à des séquences d'ADN génomique, en alignant ces données contre une banque de données génomiques. La Figure 5.2 montre les résultats obtenus suite à ces analyses. Il s'avère que la quasi-totalité des reads ne s'aligne pas non plus dans la banque de données, confirmant notre hypothèse de départ. Cette absence d'une partie du génome reste néanmoins problématique dans le cadre de notre étude, puisque l'on ne peut pas négliger le fait que cette partie manquante du génome puisse abriter la ou l'une des régions d'intérêt. Une solution serait alors de pouvoir identifier les SNPs sans génome de référence, BayPass n'ayant ensuite pas besoin d'avoir accès à la position génomique des SNPs. L'outil discoSnp (Uricaru et al. 2015) permet cela, et une étude est actuellement en cours pour identifier, à l'aide de cet outil, les SNPs dans le jeu de données issu du reséquençage complet de pools d'individus. Nous nous attendons ainsi à identifier une part plus importante de SNPs et surtout à pouvoir couvrir la quasi-totalité du génome lors d'une nouvelle recherche de régions sous sélection. D'autre part, le génome est également très fragmenté, ce qui peut avoir pour conséquence dans notre cas d'aboutir à l'identification de plusieurs régions génomiques impactées par la sélection, qui n'en formeraient qu'une ou quelques une en réalité. Le génome de référence de G. rostochiensis étant de meilleur qualité que celui de G. pallida, nous avons tenté d'aligner les différents scaffolds contenant des SNPs candidats contre celui-ci, afin d'éventuellement pouvoir rapprocher quelques régions. Ceci ne s'est pas avéré concluant, mais il sera peut être nécessaire de retenter l'expérience en alignant cette fois des séquences plus petites localisées autour des SNPs candidats. Ces limites démontrent la nécessité d'avoir accès à un génome de référence de meilleure qualité. Une équipe de chercheurs hollandais travaille actuellement à une nouvelle version du génome de référence de G. pallida, ce qui pourrait ouvrir des perspectives à nos travaux en permettant de réanalyser plus finement nos résultats. En parallèle, nous avons également de notre côté cherché à construire un génome de référence à partir de nos données de séquençage. Nous avons obtenu un génome plus fragmenté, par contre nous n'avons pas encore pu vérifier si nous étions parvenus à rapprocher certains scaffolds entre eux, l'idéal étant qu'il s'agisse de scaffolds sur lesquels sont positionnés les SNPs candidats.



Figure 5.2 Assignation des données de séquençage non alignées sur le génome de référence de *G. pallida* à différents taxons, avec l'outil Kraken (Wood & Salzberg 2014). La quasi-totalité des séquences n'a pas de référence dans les bases de données.

Le logiciel BayPass, que nous avons utilisé pour notre deuxième scan génomique, permet d'identifier les marqueurs génétiques soumis à sélection, mais également ceux associés à des covariables populations-spécifiques. Dans notre étude, nous avons seulement renseigné le statut virulent ou avirulent des lignées. Cependant, pour affiner nos résultats et dans le but d'établir le lien entre phénotype et génotype, nous aurions aussi pu utiliser des caractéristiques phénotypiques des individus, et précisément celles qui ont été utilisées pour évaluer le coût de virulence. Cette approche a été envisagée mais n'a pas pu aboutir car la population N ne présentait plus assez de matériel disponible (tout comme ce fut le cas pour son utilisation dans l'étude du coût de la virulence). Si elle avait pu être réalisée, elle aurait peut-être permis de révéler l'association entre les SNPs et les traits ayant été identifiés comme liés à l'adaptation.

L'état d'avancement de notre étude ne nous a pour l'instant pas permis de nous intéresser au rôle du fond génétique de la plante hôte dans la virulence du pathogène. Il serait en effet intéressant de savoir si les mêmes régions génomiques sont impliquées dans le contournement de plantes résistantes possédant le même QTL de résistance mais des fonds génétiques différents, ou si au contraire différents chemins adaptatifs peuvent exister. Ces résultats auraient un impact direct sur les stratégies de gestion des résistances. Si les chemins adaptatifs sont différents, alors il serait possible de déployer un même facteur de résistance (présent dans différents fonds génétiques) sur de larges échelles temporelles et spatiales, sans craindre que l'adaptation d'une population à l'une de ces résistances ne condamne l'ensemble des résistances provenant de la même source de résistance.

Chez la pomme de terre, le gène *GpaV* est en position colinéaire dans diverses espèces, notamment utilisées en tant que sources de résistance. Il serait urgent de déterminer, pour les mêmes raisons, si les mêmes régions génomiques sont ciblées par l'adaptation.

Au cours de ce projet, nous avons également cherché à tester une méthode alternative au reséquençage complet de génome, le GBS, dans l'objectif de tenter de rendre plus accessibles les approches de génomique des populations. Outre l'aspect financier, une limite à l'utilisation de ces approches peut résider dans l'absence d'un génome de référence de l'espèce étudiée. Le GBS peut constituer une première solution car il ne nécessite pas de génome de référence pour l'identification des SNPs. Si la volonté est de couvrir plus largement le génome, la solution peut reposer sur l'utilisation du logiciel discoSnp, ou encore peut-être sur l'utilisation d'un génome d'espèce proche. *Globodera pallida et G. rostochiensis* sont deux espèces sœurs avec des génomes de référence publiés, et nos données de séquençage sont actuellement en cours de traitement avec le logiciel discoSnp. Nous envisageons donc de comparer le nombre de SNPs découverts avec chacune des trois méthodes, c'est-à-dire en utilisant le génome de référence, le génome d'une espèce proche et en n'utilisant pas de génome de référence, pour statuer sur l'utilité de ces méthodes alternatives, lorsqu'il n'y a pas de génome de référence disponible.

5.3.4 Gènes candidats à l'adaptation

Il serait illusoire à cette étape de l'étude de chercher à cibler des gènes candidats à l'adaptation. La liste de SNPs candidats est encore conséquente et la fenêtre de recherche de gènes pourrait être importante. Cependant, nous avons émis au cours de nos travaux plusieurs hypothèses quant à la nature et au nombre de gènes impliqués, qu'il convient de rappeler ici.

Solanum vernei présente une résistance décrite comme polygénique, néanmoins la nature de la résistance d'Iledher et des autres génotypes résistants dans lesquels GpaV_{vrn} a été introgressé est plus difficile à appréhender. Seule la présence de ce QTL dans les génotypes résistants a été vérifiée. En revanche, le contenu de ce QTL reste inconnu et il pourrait tout aussi bien contenir un ou plusieurs gènes de résistance, il arrive en effet que les gènes de résistance soient présents en cluster dans le génome des plantes. De plus, les résistances présentes dans les fonds génétiques peuvent aussi représenter une part non négligeable de la résistance globale de la plante. Si le nombre et la nature des mutations déterminant la virulence sont des paramètres importants à intégrer dans une stratégie de gestion durable des résistances, le nombre et la nature des gènes de résistance le sont tout autant, puisque ces deux types de gènes sont liés au sein d'une dynamique coévolutive. Des connaissances supplémentaires sur les caractéristiques du QTL de résistance pourraient servir cette étude. Nous émettons cependant l'hypothèse d'une relation impliquant seulement un ou quelques gènes et d'une interaction plutôt de type gène-pour-gène. En effet, les interactions hôtenématode sont de manière générale considérées comme suivant des relations gène-pourgène (Castagnone-Sereno 2002), à l'image de G. rostochiensis et du gène de résistance H1

(Janssen *et al.* 1991).

De même, le type de gènes impliqués dans la virulence reste très hypothétique, néanmoins la majorité des gènes de virulence identifiés chez les nématodes phytoparasites sont des effecteurs (Haegeman *et al.* 2012; Mitchum *et al.* 2013), et c'est pourquoi nous avons orienté nos premières recherches « informatives » de gènes en ciblant des protéines sécrétées. Il s'est avéré que certaines des régions candidates identifiées étaient enrichies en SPRYSECs, or le seul gène d'avirulence connu à ce jour chez *G. pallida, Gp-RBP-1*, code pour un SPRYSEC qui interagit avec la protéine de résistance GPA-2 de la pomme de terre. Dans cette interaction, la virulence est due au polymorphisme d'un seul acide aminé.

Dans certains cas, il a également été montré que la virulence du nématode pouvait être due à la variation du nombre de copies de gène (CNV, *Copy Number Variation*) (Bekal *et al.* 2015; Castagnone-Sereno *et al.* 2016). Nos données de séquençage pourraient ainsi être exploitées dans une recherche orientée vers ce type de variations génomiques (CNV, Indels).

Même si les scans génomiques sont des approches sans a priori, l'étape d'investigation des gènes liés aux SNPs candidats peut avoir tendance à s'orienter vers les gènes ou fonctions déjà connues, comme c'est le cas ici. Or, d'autres types de gènes moins connus pourraient également être de bons candidats à la virulence. Nous nous sommes par exemple intéressés aux gènes codant pour des récepteurs nucléaires, car l'un d'eux a été trouvé dans la région investiguée autour du marqueur microsatellite Gr69. Il n'avait pas été évoqué dans l'article car cela restait trop spéculatif vis-à-vis de l'objectif de la publication qui était de valider la faisabilité de l'approche de scan génomique sur notre matériel expérimental, mais nous avons souhaité y revenir ici. Les récepteurs nucléaires sont une famille de facteurs de transcription impliqués dans la régulation des transitions dans les cycles biologiques des vertébrés, des insectes et des nématodes. Chez de nombreux nématodes, ils sont particulièrement impliqués dans la transition vers le stade *dauer*, un stade alternatif induit par des conditions environnementales hostiles et permettant au nématode d'y survivre. Chez le nématode libre Caenorhabditis elegans, il est connu que les récepteurs nucléaires sont impliqués dans la modulation du développement en fonction des conditions nutritives (Antebi 2015). Fait intéressant, chez G. pallida, comme nous l'avons vu précédemment, le déterminisme du sexe dépend aussi des conditions nutritives. De plus, la résistance issue de S. vernei agit précisément en masculinisant les populations de G. pallida. Nous pouvons donc émettre l'hypothèse que les récepteurs nucléaires seraient impliqués dans le dialogue moléculaire entre la plante résistante et le nématode, et qu'une mutation au niveau du récepteur nucléaire pourrait permettre de contourner la résistance de la plante. Cependant, des études complémentaires restent évidemment nécessaires pour mieux comprendre le développement et le déterminisme du sexe chez G. pallida ainsi que le rôle potentiel des récepteurs nucléaires dans ces processus. Néanmoins, aucun récepteur nucléaire n'a ensuite été retrouvé à proximité des 31 SNPs outlier.

5.4 Perspectives

5.4.1 Validation du lien génotype-phénotype dans des populations naturelles : une étude en cours

L'approche de scan génomique a permis d'identifier précédemment des SNPs candidats à l'adaptation à la résistance issue de *S. vernei*. Cependant, comme nous l'avons vu, cette liste porte certainement de l'information redondante mais comporte aussi probablement des SNPs impliqués dans d'autres types d'adaptation (adaptation à la plante hôte, adaptation à la résistance initialement présente dans le fond génétique). Afin de filtrer cette liste de candidats, une approche de validation du lien entre génotype et expression du phénotype est actuellement en cours. Elle a pour but de génotyper, au niveau des 31 SNPs candidats, des populations naturelles de *G. pallida* présentant des niveaux de virulence à lledher contrastés. Nous disposons en effet de données de phénotypage (*i.e.* nombre de kystes produits) de 20 populations naturelles représentatives de la diversité mondiale de *G. pallida* (six populations européennes, 12 péruviennes appartenant aux différents clades

génétiques présents dans le berceau d'origine de cette espèce, et deux chiliennes), sur la variété de pomme de terre lledher (Fig. 5.3). Ces données sont fortement contrastées : les populations sud-américaines appartenant aux clades IV et V sont naturellement très virulentes face à lledher, tandis que les populations des clades II et III sont moyennement virulentes. Enfin les populations du clade I, qui regroupe des populations sud-américaines et européennes, ne sont que très peu virulentes à lledher. L'approche retenue pour le génotypage est une approche d'Allèle-Spécifique PCR (AS-PCR, Gaudet et al. 2009). Le protocole expérimental utilisé est présenté en Annexe 4. Le but de l'AS-PCR est de concevoir des amorces de PCR spécifiques de chaque allèle du SNP candidat, ainsi qu'une troisième amorce commune, puis de réaliser une PCR quantitative afin de quantifier les fréquences alléliques de chaque SNP candidat dans les différentes populations. Si une corrélation entre les fréquences alléliques de certains SNPs candidats et les niveaux de virulence observés dans les populations est mise en évidence, alors elle permettra de valider l'implication de ces SNPs dans l'adaptation à la résistance issue de S. vernei. Rappelons que, contrairement aux lignées issues d'évolution expérimentale, les populations naturelles génotypées ici n'ont jamais été auparavant confrontées à lledher. Cependant, S. vernei est une espèce sauvage présente dans le berceau d'origine de G. pallida. Les SNPs ainsi validés seront donc très probablement bien liés à l'adaptation à GpaV_{vrn} et non à toute autre résistance présente dans le fond génétique d'lledher.

L'approche GBS présentée précédemment a aussi été menée sur les populations naturelles, hors cadre de la thèse. Cependant, l'exploitation des données de génotypage de ces populations peut aussi représenter un moyen de valider les SNPs candidats. En effet, si les populations naturelles ont été génotypées au niveau des SNPs candidats ou à des SNPs liés, et que ceux-ci sont effectivement impliqués dans l'adaptation à la résistance, alors les fréquences alléliques observées à ces SNPs sont susceptibles de refléter les différents niveaux de virulence. Comme le pipeline UNEAK ne nécessite pas de génome de référence pour l'identification des SNPs, ceux-ci n'étaient pas positionnés par rapport au génome de référence. Nous avons réalisé un mapping avec Bowtie2 des séquences de 64 pb issues du GBS, contenant chacune un SNP, sur le génome de référence de G. pallida. Aucune séquence ne comportait un des 31 SNPs d'intérêt. Cependant, nous avons conservé toutes les séguences situées à moins de 60 kb d'un des 31 SNPs d'intérêt. Une première analyse rapide du jeu de données a effectivement mis en évidence des corrélations entre les fréquences alléliques au niveau de certains SNPs et les niveaux de virulence des populations. Cette analyse doit être poursuivie, mais elle permet déjà de présager de l'obtention de résultats concluants avec l'approche AS-PCR.

5.4.2 À terme : des perspectives fondamentales et appliquées

Validation fonctionnelle des gènes candidats

Les régions génomiques candidates retenues précédemment seront étudiées plus en détails pour identifier le ou les gène(s) (et les variations alléliques sous-jacentes) impliqué(s) dans l'adaptation à la résistance. Ceci nécessitera tout d'abord une analyse bioinformatique autour des SNPs retenus, notamment par la recherche de peptides signaux. Il serait ensuite intéressant de réaliser une hybridation *in situ* des transcrits des gènes candidats, afin de



Figure 5.3 Nombre moyen de kystes produits par plante (n=4) dans chacune des 20 populations de *G. pallida* représentatives de la diversité mondiale, sur le cultivar lledher.

vérifier leur expression et de localiser la région du nématode dans laquelle ils sont exprimés. Nous nous attendons en effet à ce qu'en tant que potentiels effecteurs, ceux-ci soient exprimés au niveau des glandes salivaires. La recherche de motifs répétés et l'étude de pressions de sélection le long de la séquence de ces gènes permettront de cibler la nature de l'adaptation.

L'étape finale consistera à valider l'implication de ces gènes candidats par des approches fonctionnelles. En particulier, une approche d'expression transitoire par agroinfiltration de feuilles de pomme de terre (Bhaskar *et al.* 2009) issues des génotypes résistant et sensible, avait été envisagée dès le début du projet.

L'agro-infiltration repose sur l'utilisation d'un vecteur biologique, la bactérie Agrobacterium tumefaciens, introduit mécaniquement sous forme de culture liquide dans une plante hôte (avec une seringue ou en utilisant une pompe à vide), à travers les interstices naturels (stomates) ou artificiels (blessures) de la plante, généralement au niveau des feuilles. Agrobacterium tumefaciens possède un chromosome bactérien ainsi qu'un plasmide « Ti », composé en grande partie de gènes de virulence lui permettant de franchir les barrières des cellules végétales, ainsi qu'une région nommée « T-DNA », appelée à être intégrée dans le génome de la cellule végétale au contact de la bactérie. Le but est ici d'exploiter les propriétés de la bactérie, et en particulier du plasmide, pour intégrer les gènes candidats identifiés au génome des cellules végétales, afin qu'ils y soient exprimés. Pour cela, la séquence codante du gène candidat est liguée dans un vecteur d'expression. Les régions promotrices du gène, le peptide signal de même que les régions 5' et 3' régulatrices sont supprimées de la séquence. Le vecteur d'expression correspond au plasmide désarmé de la bactérie. Il possède généralement une cassette de transcription comprenant le promoteur et la séquence de terminaison CaMV 35S (Cauliflower mosaïc virus), un marqueur de résistance ainsi que les bornes du T-DNA. Après ligation, le plasmide est réintroduit par électroporation dans une souche d'A. tumefaciens désarmées portant un plasmide helper qui comprend deux gènes de virulence d'A. tumefaciens, virE et virG. Un inoculum d'A. tumefaciens ainsi composé de plasmides recombinés est ensuite agro-infiltré dans une plante résistante. Si l'interaction entre le produit du gène candidat (d'avirulence) et le produit de gène de résistance de la plante est compatible, une réaction d'hypersensibilité (HR) au niveau de la région infiltrée est observée. La réaction de nécrose doit ensuite être validée en vérifiant l'absence d'induction de HR sur variété sensible ou l'absence de nécrose en présence de la souche d'A. tumefaciens transformée avec un plasmide sans insert.

L'agro-infiltration est une technique particulièrement bien développée sur tabac. Lorsque le gène de résistance de la plante (autre que tabac) est connu, il est possible de faire s'exprimer les deux gènes d'avirulence et de résistance par co-infiltration dans une plante de tabac. Cependant, dans notre cas, le ou les gènes de résistance au sein du QTL *GpaV_{vrn}* n'étant pas encore caractérisés, cette approche sur tabac n'est pas possible à mettre en œuvre pour le moment. L'agro-infiltration est plus difficile à réaliser sur pomme de terre, de par la nature de ses feuilles. Des tests préliminaires ont en effet été réalisés sur la variété lledher au cours de ce projet de thèse, mais ils ne se sont pas avérés concluants, le protocole reste donc encore à améliorer. Une alternative à l'agro-infiltration pourrait être l'agro-infection (Du *et al.* 2015) sur pomme de terre. Celle-ci consiste à intégrer le gène candidat dans un vecteur d'expression comprenant PVX puis à introduire cette fois la bactérie dans la plante par blessure (lledher étant sensible au PVX). Dans ce cas, l'utilisation de constructions avec PVX entraîne souvent une expression systémique du gène candidat, mais elle peut aussi être locale (Ali *et al.* 2015). L'agro-infection pourrait aussi être réalisée sur tabac puis, une fois le virus multiplié, celui-ci serait prélevé et inoculé directement sur pomme de terre (par inoculation mécanique).

Si, à l'avenir, le ou les gènes de résistance chez la pomme de terre est (sont) identifié(s), une approche double-hybride validant l'interaction des deux produits de gène pourrait être envisagée. Si le gène d'avirulence est entre temps bien caractérisé, ce type d'approche pourrait alors être utilisé pour identifier, au sein d'un panel de gènes de résistance candidats, le ou les gène(s) effectivement impliqué(s) dans l'interaction.

De manière générale, la validation fonctionnelle représente une étape importante dans l'identification des bases génétiques de l'adaptation, permettant de démontrer qu'un gène sous-tend bien une variation phénotypique observée (Shimizu & Purugganan 2005; Hoekstra & Coyne 2007; Pavey *et al.* 2012). Si les outils de validation fonctionnelle tendent à se développer, ils restent néanmoins principalement appliqués aux organismes modèles. Les outils particulièrement utilisés aujourd'hui regroupent l'interférence ARN (RNAi, *RNA interference*) (Fire *et al.* 1998), une technique de *gene silencing* par dégradation de l'ARN messager, et les outils d'édition génomique, basés sur l'utilisation de nucléases modifiées et la reconnaissance de domaines sur l'ADN, tels que les *zinc-finger nucleases* (ZFNs), les *transcription activator-like effector nucleases* (TALENs) et le très médiatisé *clustered regulatory interspaced short palindromic repeats* (CRISPR)/*Cas9 endonuclease system* (Gaj *et al.* 2013; Wei *et al.* 2013). Dans le futur, des analyses de transgénèse devraient permettre de remplacer les allèles aux loci d'intérêts afin de démontrer l'importance fonctionnelle de ces substitutions.

L'étude des gènes candidats identifiés permettra d'améliorer les connaissances sur les bases génétiques de la virulence, et de déduire le nombre et la nature des mutations nécessaires au contournement de la résistance issue de *S. vernei*. Ces connaissances fondamentales seront en lien avec des perspectives appliquées, puisque nous avons vu que le nombre et la nature des mutations responsables de la virulence pouvaient être directement liés à la durabilité d'une résistance. L'étude des gènes candidats devrait également permettre de comprendre les conséquences phénotypiques de l'adaptation constatées dans le **Chapitre 2**.

Suivi de la fréquence des allèles de virulence

Dans un cadre appliqué, le ou les SNP(s) candidat(s) lié(s) à la virulence qui auront confirmé le lien entre génotype et phénotype constitueront un outil moléculaire permettant de suivre la fréquence des allèles de virulence dans les populations de *G. pallida*. Il ne sera en effet pas nécessaire d'avoir recours à la mutation précise responsable du contournement, puisque les SNPs validés devraient être physiquement assez proches du site impliqué pour en mimer le comportement.

Plusieurs applications de cet outil sont envisageables. D'une part, il pourrait servir à réaliser le diagnostic d'une parcelle avant tout déploiement de résistance, dans l'objectif d'adapter par la suite les stratégies de déploiement des résistances en fonction des fréquences de virulents mesurées. Ces stratégies auront auparavant dû être conçues et étudiées, et là encore l'outil développé pourrait jouer un rôle important. Il pourra par exemple permettre de suivre la fréquence des allèles de virulence dans des populations de *G. pallida* naturelles soumises à différentes pratiques culturales ou modalités de travail du sol, qui peuvent moduler le déficit en hétérozygotes des populations (Montarry *et al.* 2015), ainsi que dans des populations artificielles (évolution expérimentale) confrontées à différentes stratégies d'utilisation des génotypes résistants de pomme de terre.

Finalement, cet outil permettra de relier dynamique et génétique du contournement dans le but d'anticiper les futurs contournements et de concevoir des stratégies de déploiement des variétés maximisant la durabilité des résistances aux nématodes.

Les SNPs validés permettront aussi d'affiner les résultats de l'étude du coût de l'adaptation, en rendant accessible le suivi de l'évolution des fréquences des allèles de virulence et d'avirulence dans des situations de compétition.

5.5 Conclusion

Cette thèse a permis d'apporter des connaissances sur les déterminants génétiques de l'adaptation de *G. pallida* aux résistances de sa plante hôte et sur les conséquences de cette adaptation sur sa *fitness*. Nous savons à présent que l'adaptation de *G. pallida* aux résistances issues de *S. vernei* est susceptible d'entraîner un gain de *fitness* chez les individus virulents, connaissance à intégrer dans les schémas de gestion des résistances. Nous disposons également d'un nombre restreint de régions candidates à l'adaptation chez le nématode, ce qui constitue une progression vers la détermination des bases génétiques de la virulence et le développement d'outils de diagnostic et de conception de stratégies de déploiement des résistances. Les perspectives de ces travaux, tant fondamentales qu'appliquées, permettront *in fine* d'améliorer la gestion des résistances variétales dans le but d'augmenter leur durabilité.

Cette thèse avait pour cadre conceptuel le contournement des résistances et peut ainsi être replacée dans le schéma qui en décrit les différentes étapes (Gómez *et al.* 2009) (Fig. 5.4). En posant l'hypothèse d'une mutation déjà présente dans les populations, nous nous sommes affranchis de la première étape (facultative) d'apparition de la virulence. Cette thèse s'est ainsi positionnée à la seconde étape du contournement d'une résistance, qui correspond à la multiplication des individus virulents. Nous nous sommes en particulier intéressés à la sélection qui est la force évolutive déterminante lors de cette étape, et avons cherché à caractériser les bases génétiques permettant cette multiplication ainsi que les conséquences sur les traits d'histoire de vie du nématode. Les résultats de cette thèse permettront par la suite de développer des outils essentiels aux études portant sur la troisième étape du contournement, la dispersion, en permettant de suivre l'évolution des fréquences des individus virulents à l'échelle de la plante, de la parcelle ou du paysage.



Figure 5.4 Schéma bilan du positionnement de la thèse dans le cadre conceptuel du contournement des résistances (d'après Gómez *et al.* 2009). Les individus avirulents sont représentés en gris et les individus virulents en orange. Les flèches vertes représentent la migration, active ou passive, des individus pathogènes. Les forces évolutives intervenant à chacune des trois étapes sont renseignées en bleu.

RÉFÉRENCES BIBLIOGRAPHIQUES

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abang MM, Baum M, Ceccarelli S, Grando S, Linde CC, Yahyaoui A, Zhan J, McDonald BA (2006) Differential selection on *Rhynchosporium secalis* during parasitic and saprophytic phases in the barley scald disease cycle. *Phytopathology*, **96**, 1214-1222.
- Adoutte A, Balavoine G, Lartillot N, de Rosa R (1999) Animal evolution the end of the intermediate taxa? *Trends in Genetics*, **15**, 104-108.
- Agrios GN (2005) *Plant Pathology*, 5th edn. Elsevier Academic Press, Amsterdam, The Netherlands.
- Aguileta G, Refrégier G, Yockteng R, Fournier E, Giraud T (2009) Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution*, **9**, 656-670.
- Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*, **387**, 489-493.
- Aktar MW, Sengupta D, Chowdhury A (2009) Impact of pesticides use in agriculture: their benefits and hazards. *Interdisciplinary Toxicology*, **2**, 1-12.
- Alenda C, Montarry J, Grenier E (2014) Human influence on the dispersal and genetic structure of French *Globodera tabacum* populations. *Infection, Genetics and Evolution*, **27**, 309-317.
- Ali S, Magne M, Chen S, Obradovic N, Jamshaid L, Wang X, Bélair G, Moffett P (2015) Analysis of *Globodera rostochiensis* effectors reveals conserved functions of SPRYSEC proteins in suppressing and eliciting plant immune responses. *Frontiers in Plant Science*, **6**, 623.
- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST} -outlier method. *BMC Bioinformatics*, **9**, 323.
- Antebi A (2015) Nuclear receptor signal transduction in *C. elegans*. In: *WormBook* (The *C. elegans* Research Community ed.), http://www.wormbook.org.
- Antonivics J, Alexander HM (1989) The concept of fitness in plant-fungal pathogen systems. In: *Plant Disease Epidemiology* (Leonard KJ, Fry WE eds.), pp. 185-214. McGraw-Hill, New-York, USA.
- Arnold B, Corbet-Detig RB, Hartl D, Bomblies K (2013) RAD-seq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, 22, 3179-3190.
- Ayala FJ, Rzhetsky A, Ayala FJ (1998) Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 606-611.
- Ayme V, Souche S, Caranta C, Jacquemond M, Chadoeuf J, Palloix A, Moury B (2006) Different mutations in the genome-linked protein VPg of *Potato virus Y* confer virulence on the *pvr2*³ resistance in pepper. *Molecular Plant-Microbe Interactions*, **19**, 557-563.
- Bahri B, Kaltz O, Lecompte M, de Vallavieille-Pope C, Enjalbert J (2009) Tracking costs of virulence in natural populations of the wheat pathogen, *Puccinia striiformis* f.sp. *tritici*. *BMC Evolutionary Biology*, **9**, 26.
- Bailey S, Bataillon T (2016) Can the experimental evolution programme help us elucidate the genetic basis of adaptation in nature? *Molecular Ecology*, **25**, 203-218.
- Bakker J (2002) Durability of resistance against potato cyst nematodes. *Euphytica*, **124**, 157-162.

- Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen J (2014) Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends in Genetics*, **30**, 540-546.
- Barbary A, Djian-Caporalino C, Palloix A, Castagnone-Sereno P (2015) Host genetic resistance to root-knot nematodes, *Meloidogyne* spp., in Solanaceae: from genes to the field. *Pest Management Science*, **71**, 1591-1598.
- Barbary A, Palloix A, Fazari A, Marteu N, Castagnone-Sereno P, Djian-Caporalino C (2014) The plant genetic background affects the efficiency of the pepper major nematode resistance genes *Me1* and *Me3*. *Theoretical and Applied Genetics*, **127**, 499-507.
- Barker KR (1998) Introduction and synopsis of advancements in nematology. In: *Plant and Nematode Interactions* (Barker KR, Pederson GL, Windham GL eds.), pp. 1-20. American Society of Agronomy, Madison, USA.
- Barone A, Ritter E, Schachtschabel U, Debener T, Salamini F, Gebhardt C (1990) Localization by restriction fragment length polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Molecular and General Genetics*, **224**, 177-182.
- Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, **12**, 767-780.
- Beaumont MA (2005) Adaptation and speciation: what can F_{ST} tell us? Trends in Ecology & Evolution, **20**, 435-440.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969-980.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological sciences*, **263**, 1619-1626.
- Behrens E (1975) *Globodera* Skarbilovich, 1959 an independent genus in the subfamily Heteroderinae Skarbilovich, 1949 (Nematoda: Heteroderidae). *Vortragstagung zu Aktuellen Problemen der Phytonematologie*, **1**, 12-26.
- Bekal S, Domier LL, Gonfa B, Lakhssassi N, Meksem K, Lambert KN (2015) A SNARE-like protein and biotin are implicated in soybean cyst nematode virulence. *PLoS ONE*, **10**, e0145601.
- Bekal S, Niblack TL, Lambert KN (2003) A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. *Molecular Plant-Microbe Interactions*, **16**, 439-446.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05, logiciel sous Windows[™] pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
- Bendahmane A, Kanyuka K, Baulcombe DC (1997) High-resolution genetical and physical mapping of the *Rx* gene for extreme resistance to *Potato virus X* in tetraploid potato. *Theoretical and Applied Genetics*, **95**, 153-162.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **57**, 289-300.
- Bernard EC (1992) Soil nematode biodiversity. *Biology and Fertility of Soils*, **14**, 99-103.
- Bhaskar PB, Venkateshwaran M, Wu L, Ané JM, Jiang J (2009) *Agrobacterium*-mediated transient gene expression and silencing: a rapid tool for functional gene assay in potato. *PLoS ONE*, **4**, e5812.
- Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it...? Why are *F*_{ST} outliers sometimes so frequent? *Molecular Ecology*, **22**, 2061-2064.

- Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044-2072.
- Birch PR, Armstrong M, Bos J, Boevink P, Gilroy EM, Taylor RM, Wawra S, Pritchard L, Conti L, Ewan R, Whisson SC, van West P, Sadanandom A, Kamoun S (2009) Towards understanding the virulence functions of RXLR effectors of the oomycete plant pathogen *Phytophthora infestans*. *Journal of Experimental Botany*, **60**, 1133-1140.
- Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, **46**, 441-469.
- Blanchard A, Esquibet M, Fouville D, Grenier E (2005) *Ranbpm* homologue genes characterised in the cyst nematodes *Globodera pallida* and *Globodera mexicana*. *Physiological and Molecular Plant Pathology*, **67**, 15-22.

Blankenberg D, Von Kuster G, Coraor N, Guruprasad A, Lazarus R, Mangan M, Nekrutenko A, Taylor J (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology*, **89**, 19.10.1-19.10.21.

Blaxter ML, Ley PD, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, Vida JT, Thomas WK (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature*, **392**, 71-75.

Blumenthal T, Davis RE (2004) Exploring nematode diversity. Nature Genetics, 36, 1246-1247.

- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
- Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, SanCristobal M (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, **186**, 241-262.
- Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on common set of clones reveal modes on chromosomal evolution in potato and tomato. *Genetics*, **120**, 1095-1103.
- Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular Ecology*, **17**, 3583-3584.
- Bonin A, Nicole F, Pompanon F, Miaud C, Taberlet P (2007) Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation Biology*, **21**, 697-708.
- Bonin A, Paris M, Tetreau G, David JP, Després L (2009) Candidate genes revealed by a genome scan for mosquito resistance to a bacterial insecticide: sequence and gene expression variation. *BMC Genomics*, **10**, 551.
- Boucher AC, Mimee B, Montarry J, Bardou-Valette S, Bélair G, Moffett P, Grenier E (2013) Genetic diversity of the golden potato cyst nematode *Globodera rostochiensis* and determination of the origin of populations in Quebec, Canada. *Molecular Phylogenetics and Evolution*, **69**, 75-82.
- Boyd LA (2006) Can the durability of resistance be predicted? *Journal of the Science of Food and Agriculture*, **86**, 2523-2526.
- Bradbury D, Smithson A, Krauss SL (2013) Signatures of diversifying selection at EST-SSR loci and association with climate in naturel *Eucalyptus* population. *Molecular Ecology*, **22**, 5112-5129.
- Bradshaw JE, Meyer RC, Milbourne D, McNicol JW, Phillips MS, Waugh R (1998) Identification of AFLP and SSR markers associated with quantitative resistance to *Globodera pallida* (Stone) in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*) with a view to markerassisted selection. *Theoretical and Applied Genetics*, **97**, 202-210.

- Bronson CR, Ellingboe AH (1986) The influence of 4 unnecessary genes for virulence on the fitness of *Erysiphe graminis* f. sp. *tritici*. *Phytopathology*, **76**, 154-158.
- Brown JKM (2015) Durable resistance of crops to disease: a Darwinian perspective. *Annual Review of Phytopathology*, **53**, 513-539.
- Brown JKM, Tellier A (2011) Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual Review of Phytopathology*, **49**, 345-367.
- Browning JA, Frey KJ (1969) Multiline cultivars as a means of disease control. *Annual Review* of *Phytopathology*, **7**, 355-382.
- Brun H, Chèvre AM, Fitt BDL, Powers S, Besnard AL, Ermel M, Huteau V, Marquer B, Eber F, Renard M, Andrivon D (2010) Quantitative resistance increases the durability of qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *New Phytologist*, **185**, 285-299.
- Burlingame B, Mouillé B, Charrondière R (2009) Nutrients, bioactive non-nutrients and antinutrients in potatoes. *Journal of Food Composition and Analysis*, **22**, 494-502.
- Butlin R (2010) Population genomics and speciation. *Genetica*, **138**, 409-418.
- Caffier V, Lasserre-Zuber P, Giraud M, Lascostes M, Stievenard R, Lemarquand A, van de Weg E, Expert P, Denance C, Didelot F, Le Cam B, Durel CE (2014) Erosion of quantitative host resistance in the apple x *Venturia inaequalis* pathosystem. *Infection, Genetics and Evolution*, **27**, 481-489.
- Caromel B (2004) Cartographie génétique et étude de QTL conférant la résistance au nématode à kyste *Globodera pallida* (Stone) chez la pomme de terre (*Solanum tuberosum* ssp. *tuberosum* L.). Thèse de Doctorat, 133 pp., Université Paris XI Orsay, France.
- Caromel B, Aarrouf J, Chauvin L, Kerlan MC, Lefebvre V, Speck A (2011) *GpaV* gene resistant to nematodes in the Nightshade family. WIPO patent application wo / 2011 / 080329.
- Caromel B, Mugniéry D, Kerlan MC, Andrzejewski S, Palloix A, Ellissèche D, Rousselle-Bourgeois F, Lefebvre V (2005) Resistance quantitative trait loci originating from *Solanum sparsipilum* act independently on sex ratio of *Globodera pallida* and together for developing a necrotic reaction. *Molecular Plant-Microbe Interactions*, **18**, 1186-1194.
- Caromel B, Mugniéry D, Lefebvre V, Andrzejewski S, Ellissèche D, Kerlan MC, Rousselle P, Rousselle-Bourgeois F (2003) Mapping QTLs for resistance against *Globodera pallida* (Stone) Pa2/3 in a diploid potato progeny originating from *Solanum spegazzinii*. *Theoretical and Applied Genetics*, **106**, 1517-1523.
- Carpentier J, Esquibet M, Fouville D, Manzanares-Dauleux MJ, Kerlan MC, Grenier E (2012) The evolution of the *Gp-Rbp-1* gene in *Globodera pallida* includes multiple selective replacements. *Molecular Plant Pathology*, **13**, 546-555.
- Castagnone-Sereno P (2002) Genetic variability of nematodes: a threat to the durability of plant resistance genes? *Euphytica*, **124**, 193-199.
- Castagnone-Sereno P, Bongiovanni M, Wajnberg E (2007) Selection and parasite evolution: a reproductive fitness cost associated with virulence in the parthenogenetic nematode *Meloidogyne incognita*. *Evolutionary Ecology*, **21**, 259-270.
- Castagnone-Sereno P, Danchin EGJ (2014) Parasitic success without sex the nematode experience. *Journal of Evolutionary Biology*, **27**, 1323-1333.
- Castagnone-Sereno P, Mulet K, Da Rocha M, Danchin EGJ, Perfus-Barbeoch L, Abad P (2016) Genome-wide detection of copy number variations in avirulent versus virulent nearisogenic lines of the asexual root-knot nematode *Meloidogyne incognita*. 32nd Symposium of the European Society of Nematologists, Braga, Portugal, 28 August - 1 September 2016.

- Castagnone-Sereno P, Mulet K, Lachia C (2015) Tracking changes in life-history traits related to unnecessary virulence in a plant-parasitic nematode. *Ecology and Evolution*, **5**, 3677-3686.
- Cavalli-Sforza LL (1966) Population structure and human evolution. *Proceedings of the Royal Society of London B: Biological sciences*, **164**, 362-379.
- Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155-174.
- Chauvin L, Caromel B, Kerlan MC, Rulliat E, Fournet S, Chauvin JE, Grenier E, Ellissèche D, Mugniéry D (2008) Control of potato cyst nematodes *Globodera rostochiensis* and *Globodera pallida*. *Cahiers Agricultures*, **17**, 368-374.
- Cheatham MR, Rouse MN, Esker PD, Ignacio S, Pradel W, Raymundo R, Sparks AH, Forbes GA, Gordon TR, Garrett KA (2009) Beyond yield: plant disease in the context of ecosystem services. *Phytopathology*, **99**, 1228-1236.
- Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-microbe interactions: shaping the evolution of the plant immune response. *Cell*, **124**, 803-814.
- Clayton R, Storey M, Parker B, Ballingall M, Davies K (2008) *Impact of Reduced Pesticide Availability on Control of Potato Cyst Nematodes and Weeds in Potato Crops*. Potato Council Ltd, Kenilworth, UK.
- Clément JA, Toulza E, Gautier M, Parrinello H, Roquis D, Boissier J, Rognon A, Moné H, Mouahid G, Buard J, Mitta G, Grunau C (2013) Private selective sweeps identified from next-generation pool-sequencing reveal convergent pathways under selection in two inbred *Schistosoma mansoni* strains. *PLoS Neglected Tropical Diseases*, **7**, e2591.
- Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928-1933.
- Colosio C, Alegakis AK, Tsatsakis AM (Eds) (2013) Emerging health issues from chronic pesticide exposure: Innovative methodologies and effects on molecular cell and tissue level [Special Issue]. *Toxicology*, **307**, 1-146.
- Cooke TF, Yee MC, Muzzio M, Sockell A, Bell R, Cornejo OE, Kelley JL, Bailliet G, Bravi CM, Bustamante CD, Kenny EE (2016) GBStools: a statistical method for estimating allelic dropout in reduced representation sequencing data. *PLoS Genetics*, **12**, e1005631.
- Coop G, Witonsky D, Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411-1423.
- Cotton JA, Lilley CJ, Jones LM, Kikuchi T, Reid AJ, Thorpe P, Tsai IJ, Beasley H, Blok V, Cock PJA, Eves-van den Akker S, Holroyd N, Hunt M, Mantelin S, Naghra H, Pain A, Palomares-Rius JE, Zarowiecki M, Berriman M, Jones JT, Urwin PE (2014) The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology*, **15**, R43.
- Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133-3157.
- Danchin EGJ, Rosso MN, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, Abad P (2010) Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences*, **107**, 17651-17656.
- Davis EL, Hussey RS, Baum TJ, Bakker J, Schots A, Rosso MN, Abad P (2000) Nematode parasitism genes. *Annual Review of Phytopathology*, **38**, 365-396.

- Dawkins R, Krebs JR (1979) Arms races between and within species. *Proceedings of the Royal Society of London B: Biological sciences*, **205**, 489-511.
- de Boer JM, Borm TJA, Jesse T, Brugmans B, Tang X, Bryan GJ, Bakker J, van Eck HJ, Visser RGF (2011) A hybrid BAC physical map of potato: a framework for sequencing a heterozygous genome. *BMC Genomics*, **12**, 594.
- De Mita S, Thuillet AC, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, **22**, 1383-1399.
- Decraemer W, Hunt DJ (2006) Structure and classification. In: *Plant Nematology* (Perry RN, Moens M eds.) pp. 3-32. Cabi Publishing, Oxon, UK.
- Delourme R, Bousset L, Ermel M, Duffé P, Besnard AL, Marquer B, Fudal I, Linglin J, Chadoeuf J, Brun H (2014) Quantitative resistance affects the speed of frequency increase but not the diversity of the virulence alleles overcoming a major resistance gene to *Leptosphaeria maculans* in oilseed rape. *Infection, Genetics and Evolution*, **27**, 490-499.
- Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledo JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, Baer CF (2009) A genome-wide view of *Caenorhabditis elegans* basesubstitution mutation processes. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 16310-16314.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491-498.
- Díaz JA, Nieto C, Moriones E, Truniger V, Aranda MA (2004) Molecular characterization of a *Melon necrotic spot virus* strain that overcomes the resistance in melon and nonhost plants. *Molecular Plant-Microbe Interactions*, **17**, 668-675.
- Diaz-Granados A, Petrescu AJ, Goverse A, Smant G (2016) SPRYSEC effectors: a versatile protein-binding platform to disrupt plant innate immunity. *Frontiers in Plant Science*, **7**, 1575.
- Djian-Caporalino C, Molinari S, Palloix A, Ciancio A, Fazari A, Marteu N, Ris N, Castagnone-Sereno P (2011) The reproductive potential of the root-knot nematode *Meloidogyne incognita* is affected by selection for virulence against major resistance genes from tomato and pepper. *European Journal of Plant Pathology*, **131**, 431-440.
- Djian-Caporalino C, Palloix A, Fazari A, Marteu N, Barbary A, Abad P, Sage-Palloix AM, Mateille T, Risso S, Lanza R, Taussig C, Castagnone-Sereno P (2014) Pyramiding, alternating or mixing: comparative performances of deployment strategies of nematode resistance genes to promote plant resistance efficiency and durability. *BMC Plant Biology*, 14, 53.
- Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. *Genetics*, **161**, 1697-1711.
- Dong K, Opperman CH (1997) Genetic analysis of parasitism in the soybean cyst nematode *Heterodera glycines. Genetics*, **146**, 1311-1318.
- Dorris M, De Ley P, Blaxter ML (1999) Molecular analysis of nematode diversity and the evolution of parasitism. *Parasitology*, **15**, 188-193.
- Drake JW, Holland JJ (1999) Mutation rates among RNA viruses. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 13910-13913.
- Du J, Rietman H, Vleeshouwers VG (2014) Agroinfiltration and PVX agroinfection in potato and *Nicotiana benthamiana. Journal of Visualed Experiments*, **83**, e50971.

- Duforet-Frebourg N, Bazin E, Blum MGB (2014) Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution*, **31**, 2483-2495.
- Duncan LH, Robertson L, Robertson WM, Kusel JR (1997) Isolation and characterization of secretions from the plant-parasitic nematode *Globodera pallida*. *Parasitology*, **115**, 429-438.
- Dunnett JM (1961) Inheritance of resistance to the potato root eelworm in a breeding line stemming from *Solanum multidissectum* Hawkes. *Annual Report of the Scottish Plant Breeding Station*, **1961**, 39-46.
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, Neale DB (2010) Patterns of population structure and environmental associations to aridity across the range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics*, **185**, 969-982.
- Ehrlich PR, Raven PH (1964) Butterflies and plants a study in coevolution. *Evolution*, **18**, 586-608.
- Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169-175.
- Ellis PR (1968) Resistance to the potato cyst-nematode, *Heterodera rostochiensis*, in the plant genus *Lycopersicon*. *Annals of Applied Biology*, **61**, 151-160.
- Ellissèche D, Pellé R, Andrivon D (1999) Résistance variétale au mildiou : situation et perspectives. *La Pomme de Terre Française*, **510**, 16-22.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.
- Endo BY (1987) Ultrastructure of esophageal gland secretory granules in juveniles of *Heterodera glycines. Journal of Nematology*, **19**, 469-483.
- Eoche-Bosy D, Gauthier J, Juhel AS, Esquibet M, Fournet S, Grenier E, Montarry J (2017)
 Experimentally evolved populations of the potato cyst nematode *Globodera pallida* allow the targeting of genomic footprints of selection due to host adaptation. *Plant Pathology*, 66, 1022-1030.
- Ericson L, Burdon JJ (2009) Linking field epidemiological and individual plant resistance patterns in the *Betula pubescens-Melampsoridium betulinum* host-pathogen interaction. *Oikos*, **118**, 225-232.
- Ernst K, Kumar A, Kriseleit D, Kloos DU, Phillips MS, Ganal MW (2002) The broad-spectrum potato cyst nematode resistance gene (*Hero*) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. *Plant Journal*, **31**, 127-136.
- Eves-van den Akker S, Birch PR (2016) Opening the effector protein toolbox for plant-parasitic cyst-nematode interactions. *Molecular Plant*, **9**, 1451-1453.
- Eves-van den Akker S, Laetsch DR, Thorpe P, Lilley CJ, Danchin EGJ, Da Rocha M, Rancurel C, Holroyd NE, Cotton JA, Szitenberg A, Grenier E, Montarry J, Mimee B, Duceppe MO, Boyes I, Marvin JMC, Jones LM, Yusup HB, Lafond-Lapalme J, Esquibet M, Sabeh M, Rott M, Overmars H, Finkers-Tomczak A, Smant G, Koutsovoulos G, Blok V, Mantelin S, Cock PJA, Phillips W, Henrissat B, Urwin PE, Blaxter M, Jones JT (2016) The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology*, **17**, 124.
- Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theorical Population Biology*, **3**, 87-112.

- Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285-298.
- Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, **23**, 347-351.

Fabre F, Bruchou C, Palloix A, Moury B (2009) Key determinants of resistance durability to plant viruses: insights from a model linking within- and between-host dynamics. *Virus Research*, **141**, 140-149.

Fabre F, Montarry J, Coville J, Senoussi R, Simon V, Moury B (2012) Modelling the evolutionary dynamics of viruses within their hosts: a case study using high-throughput sequencing. *PLoS Pathogens*, **8**, e1002654.

Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*, **8**, 92-94.

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405-1413.

- Ferraz LCCB, Brown DJF (2002) An Introduction to Nematodes: Plant Nematology: A Student's Textbook. Pensoft Publishers, Sofia, Bulgaria.
- Ferretti L, Ramos-Onsins SE, Pérez-Enciso M (2013) Population genomics from pool sequencing. *Molecular Ecology*, **22**, 5561-5576.
- Fierer N, Strickland MS, Liptzin D, Bradford MA, Cleveland CC (2009) Global patterns in belowground communities. *Ecology Letters*, **12**, 1238-1249.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806-811.
- Fischer MC, Rellstab C, Tedder A, Zoller S, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2013) Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology*, **22**, 5594-5607.
- Fleming CC (1998) The evaluation and durability of potato cyst nematode resistance in the potato. In: *Potato Cyst Nematodes: Biology, Distribution and Control* (Marks RJ, Brodie BB eds.), pp. 197-208. CAB International, Wallingford, UK.
- Flor HH (1956) The complementary genic systems in flax and flax rust. *Advances in Genetics*, **8**, 29-54.
- Flor HH (1971) Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, **9**, 275-296.
- 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.
- Förstner W, Moonen B (2003) A metric for covariance matrices. In: *Geodesy-The Challenge of the 3rd Millennium* (Grafarend EW, Krumm FW, Schwarze VS eds.), pp. 299-309. Springer-Verlag, Berlin/Heidelberg, Germany.
- Fournet S, Eoche-Bosy D, Renault L, Hamelin FM, Montarry J (2016) Adaptation to resistant hosts increases fitness on susceptible hosts in the plant parasitic nematode *Globodera pallida*. *Ecology and Evolution*, **6**, 2559-2568.
- Fournet S, Kerlan MC, Renault L, Dantec JP, Rouaux C, Montarry J (2013) Selection of nematodes by resistant plants has implications for local adaptation and cross-virulence. *Plant Pathology*, **62**, 184-193.
- Fraile A, García-Arenal F (2010) The coevolution of plants and viruses: resistance and pathogenicity. *Advances in Virus Research*, **76**, 1-32.

- Fraile A, Pagán I, Anastasio G, Saez E, García-Arenal F (2011) Rapid genetic diversification and high fitness penalties associated with pathogenicity evolution in a plant virus. *Molecular Biology and Evolution*, **28**, 1425-1437.
- Frichot E, Schoville D, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687-1699.
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693-709.
- Futschik A, Schlötterer C (2010) The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, **186**, 207-218.
- Gabriel S, Ziaugra L, Tabbaa D (2009) SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Current Protocols in Human Genetics*, **2**, 1-18.
- Gaj T, Gersbach CA, Barbas CF 3rd (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, **31**, 397-405.
- Ganal MW, Simon R, Brommonschenkel S, Arndt M, Phillips MS, Tanksley SD, Kumar A (1995) Genetic mapping of a wide spectrum nematode resistance gene (Hero) against *Globodera rostochiensis* in tomato. *Molecular Plant-Microbe Interactions*, **8**, 886-891.
- Gandon S, Michalakis Y (2002) Local adaptation, evolutionary potential and host-parasite coevolution: interactions between migration, mutation, population size and generation time. *Journal of Evolutionary Biology*, **15**, 451-462.
- Gao BL, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS (2002) Molecular characterisation and expression of two venom allergen-like protein genes in *Heterodera glycines*. *International Journal for Parasitology*, **31**, 1617-1625.
- Gao BL, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS (2003) The parasitome of the phytonematode *Heterodera glycines*. *Molecular Plant-Microbe Interactions*, **16**, 720-726.
- Gaudet M, Fara AG, Beritognolo I, Sabatti M (2009) Allele-Specific PCR in SNP genotyping. *Methods in Molecular Biology*, **578**, 415-424.
- Gautier M (2015) Genome-wide scan for adaptive differentiation and association analysis with population-specific covariables. *Genetics*, **201**, 1555-1579.
- Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P, Kerdelhué C, Estoup A (2013) Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology*, **22**, 3766-3779.
- Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet JM, Estoup A (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, **22**, 3165-3178.
- Gebhardt C, Mugniéry D, Ritter E, Salamini F, Bonnel E (1993) Identification of RFLP markers closely linked to the *H1* gene conferring resistance to *Globodera rostochiensis* in potato. *Theoretical and Applied Genetics*, **85**, 541-544.
- Gebhardt C, Ritter E, Debener I, Schachtschabel U, Walkemeier B, Uhrig H, Salamini F (1989) RFLP analysis and linkage mapping in *Solanum tuberosum*. *Theoretical and Applied Genetics*, **78**, 65-75.
- Gheysen G, Fenoll C (2002) Gene expression in nematode feeding sites. *Annual Review of Phytopathology*, **40**, 191-219.
- Gilabert A, Wasmuth JD (2013) Unravelling parasitic nematode natural history using population genetics. *Trends in Parasitology*, **29**, 438-448.
- Goellner M, Wang XH, Davis EL (2001) Endo-β-1,4-glucanase expression in compatible plantnematode interactions. *Plant Cell*, **13**, 2241-2255.

- Gómez P, Rodríguez-Hernández AM, Moury B, Arand MA (2009) Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. *European Journal of Plant Pathology*, **125**, 1-22.
- Gompert Z, Buerkle CA (2011) A hierarchical Bayesian model for next-generation population genomics. *Genetics*, **187**, 903-917.
- Green CD, Greet DN, Jones FGW (1970) The influence of multiple mating on the reproduction and genetics of *Heterodera rostochiensis* and *H. schachtii*. *Nematologica*, **16**, 309-326.
- Grenier E, Blok VC, Jones JT, Fouville D, Mugniéry D (2002) Identification of gene expression differences between *Globodera pallida* and *G. mexicana* by suppression subtractive hybridization. *Molecular Plant Pathology*, **3**, 217-226.

Grundler F, Schnibbe L, Wyss U (1991) In vitro studies on the behaviour of second-stage juveniles of *Heterodera schachtii* (Nematoda: Heteroderidae) in response to host plant root exudates. *Parasitology*, **103**, 149-155.

- Grunewald W, Van Noorden G, Van Isterdael G, Beeckman T, Gheysen G, Mathesius U (2009) Manipulation of auxin transport in plant roots during rhizobium symbiosis and nematode parasitism. *Plant Cell*, **21**, 2553-2562.
- Guillot G, Vitalis R, Le Rouzic A, Gautier M (2014) Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. *Spatial Statistics*, **8**, 145-155.
- Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205-220.
- Guo F, Dey DK, Holsinger KE (2009) A Bayesian hierarchical model for analysis of singlenucleotide polymorphisms diversity in multilocus, multipopulation samples. *Journal of the American Statistical Association*, **104**, 142-154.
- Haasl RJ, Johnson RC, Payseur BA (2014) The effects of microsatellite selection on linked sequence diversity. *Genome Biology and Evolution*, **6**, 1843-1861.
- Haasl RJ, Payseur BA (2016) Fifteen years of genomewide scans for selection: trends, lessons and unadressed genetics sources of complication. *Molecular Ecology*, **25**, 5-23.
- Haegeman A, Jones JT, Danchin EGJ (2011) Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Molecular Plant-Microbe Interactions*, **24**, 879-887.
- Haegeman A, Kyndt T, Gheysen G (2010) The role of pseudo-endoglucanases in the evolution of nematode cell wall-modifying proteins. *Journal of Molecular Evolution*, **70**, 441-452.
- Haegeman A, Mantelin S, Jones JT, Gheysen G (2012) Functional roles of effectors of plantparasitic nematodes. *Gene*, **492**, 19-31.

Hales CN, Baker DJ (2001) The thrifty phenotype hypothesis. *British Medical Bulletin*, **60**, 5-20.

- Hawkes JG (1990) *The Potato, Evolution, Biodiversity and Genetic Resources*. Belhaven Press, London, UK.
- Hermisson J (2009) Who believes in whole-genome scans for selection? *Heredity*, **103**, 283-284.
- Hewezi T, Baum TJ (2012) Manipulation of plant cells by cyst and root-knot nematode effectors. *Molecular Plant-Microbe Interactions*, **26**, 9-16.
- Hijmans RJ, Spooner DM (2001) Geographic distribution of wild potato species. *American Journal of Botany*, **88**, 2101-2112.
- Hockland S, Inserra RN, Millar L, Lehman PS (2006) International plant health putting legislation into practice. In: *Plant Nematology* (Perry RN, Moens M eds.), pp. 327-345. CABI Publishing, Wallingford, UK.

- Hockland S, Niere B, Grenier E, Blok V, Phillips M, den Nijs L, Anthoine G, Pickup J, Viaene N (2012) An evaluation of the implications of virulence in non-European populations of *Globodera pallida* and *G. rostochiensis* for potato cultivation in Europe. *Nematology*, 14, 1-13.
- Hoekstra HE, Coyne JA (2007) The locus of evolution: evo-devo and the genetics of adaptation. *Evolution*, **61**, 995-1016.
- Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S (2009) Emerging concepts in effector biology of plant-associated organisms. *Molecular Plant-Microbe Interactions*, **22**, 115-122.
- Hohenlohe PA, Phillips PC, Cresko WA (2010) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences*, **171**, 1059-1071.
- Holterman M, Karssen G, van den Elsen S, van Megen H, Bakker J, Helder J (2009) Small subunit rDNA-based phylogeny of the Tylenchida sheds light on relationships among some high-impact plant-parasitic nematodes and the evolution of plant feeding. *Phytopathology*, **99**, 227-235.
- Huang YJ, Li ZQ, Evans N, Rouxel T, Fitt BDL, Balesdent MH (2006) Fitness costs associated with the loss of the *AvrLm4* avirulence function in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). *European Journal of Plant Pathology*, **114**, 77-89.
- Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153-159.
- Hugot JP, Baujard P, Morand S (2001) Biodiversity in helminths and nematodes as a field of study: an overview. *Nematology*, **3**, 199-208.
- Hugot JP, Gardner SL, Borba V, Araujo P, Leles D, Da-Rosa AAS, Dutra J, Ferreira LF, Araujo A (2014) Discovery of a 240 million year old nematode parasite egg in a cynodont coprolite sheds light on the early origin of pinworms in vertebrates. *Parasites & Vectors*, **7**, 486.
- Huijsman CA, Lamberts H (1972) Breeding for resistance to the potato cyst-nematode in The Netherlands. In: *Prospects for the Potato in the Developing World* (French ER ed.), pp. 161-171. International Potato Center, Lima, Peru.
- Hussey RS (1989) Disease-inducing secretions of plant-parasitic nematodes. *Annual Review of Phytopathology*, **27**, 123-141.
- Ingle RA, Carstens M, Denby KJ (2006) PAMP recognition and the plant-pathogen arms race. *Bioessays*, **28**, 880-889.
- Ishibashi K, Mawatari N, Miyashita S, Kishino H, Meshi T, Ishikawa M (2012) Coevolution and hierarchical interactions of *Tomato mosaic virus* and the resistance gene *Tm-1*. *PLoS Pathogens*, **8**, e10002975.
- Isidore E, van Os H, Andrzejewski S, Bakker J, Barrena I, Bryan GJ, Caromel B, van Eck H, Ghareeb B, de Jong W, van Koert P, Lefebvre V, Milbourne D, Ritter E, van der Voort JR, Rousselle-Bourgeois F, van Vliet J, Waugh R (2003) Toward a marker-dense meiotic map of the potato genome: lessons from linkage group I. *Genetics*, **165**, 2107-2116.
- Jacobs JME, van Eck HJ, Horsman K, Arens PFP, Verkerk-Bakker B, Jacobsen E, Pereira A, Stiekema WJ (1996) Mapping of resistance to the potato cyst nematode *Globodera rostochiensis* from the wild potato species *Solanum vernei*. *Molecular Breeding*, **2**, 51-60.
- Jan PL, Gracianne C, Fournet S, Olivier E, Arnaud JF, Porte C, Bardou-Valette S, Denis MC, Petit EJ (2016) Temporal sampling helps unravel the genetic structure of naturally occurring populations of a phytoparasitic nematode. 1. Insights from the estimation of effective population sizes. *Evolutionary Applications*, **9**, 489-501.

- Janssen R, Bakker J, Gommers FJ (1991) Mendelian proof for a gene-for-gene relationship between virulence of *Globodera rostochiensis* and the *H1* resistance gene in *Solanum tuberosum* ssp. *andigena* CPC 11673. *Revue de Nématologie*, **14**, 213-219.
- Janzac B, Fabre F, Palloix A, Moury B (2009) Constraints on evolution of virus avirulence factors predict the durability of corresponding plant resistances. *Molecular Plant Pathology*, **10**, 599-610.
- Janzac B, Montarry J, Palloix A, Navaud O, Moury B (2010) A point mutation in the polymerase of *Potato virus Y* confers virulence toward the *Pvr4* resistance of pepper and a high competitiveness cost in susceptible cultivar. *Molecular Plant-Microbe Interactions*, **23**, 823-830.
- Jasmer DP, Goverse A, Smart G (2003) Parasitic nematode interactions with mammals and plants. *Annual Review of Phytopathology*, **41**, 245-270.
- Jeffreys H (1961) *Theory of Probability*, 3rd edn. Oxford University Press, Oxford, UK.
- Jenner CE, Wang X, Ponz F, Walsh JA (2002) A fitness cost for *Turnip mosaic virus* to overcome host resistance. *Virus Research*, **86**, 1-6.
- Jensen NF (1952) Intra-varietal diversification in oat breeding. *Agronomy Journal*, **44**, 50-54.
- Jobson MA, Jordan JM, Sandrof MA, Hibshman JD, Lennox AL, Baugh LR (2015) Transgenerational effects of early life starvation on growth, reproduction, and stress resistance in *Caenorhabditis elegans*. *Genetics*, **201**, 201-212.
- Johnson R (1981) Durable resistance: definition of, genetic control and attainment in plant breeding. *Phytopathology*, **71**, 567-568.
- Johnson R (1984) A critical analysis of durable resistance. *Annual Review of Phytopathology*, **22**, 309-330.
- Jones FGW (1950) Observations on the beet eelworm and other cyst-forming species of *Heterodera*. *Annals of Applied Biology*, **37**, 407-440.
- Jones JDG (2001) Putting knowledge of plant disease resistance genes to work. *Current Opinion in Plant Biology*, **4**, 281-287.
- Jones JDG, Dangl JL (2006) The plant immune system. *Nature*, **444**, 323-329.
- Jones JT, Haegeman A, Danchin EGJ, Gaur HS, Helder J, Jones MGK, Kikuchi T, Manzanilla-Lopez R, Palomares-Rius JE, Wesemael WML, Perry RN (2013) Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*, **14**, 946-961.
- Jones JT, Kumar A, Pylypenko LA, Thirugnanasambandam A, Castelli L, Chapman S, Cock PJ, Grenier E, Lilley CJ, Phillips MS, Blok VC (2009) Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode *Globodera pallida*. *Molecular Plant Pathology*, **10**, 815-828.
- Jones MGK, Northcote DH (1972) Nematode-induced syncytium A multinucleate transfer cell. *Journal of Cell Science*, **10**, 789-809.
- Käll L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction the Phobius web server. *Nucleic Acids Research*, **35**, W429–W432.
- Kaloshian I, Williamson VM, Miyao G, Lawn DA, Westerdahl BB (1996) "Resistance breaking" nematodes identified in California tomatoes. *California Agriculture*, **50**, 18-19.
- Kang BC, Yeam I, Jahn MM (2005) Genetics of plant virus resistance. Annual Review of *Phytopathology*, **43**, 581-621.
- Kauer MO, Dieringer D, Schlötterer C (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of *Drosophila melanogaster*. *Genetics*, **165**, 1137-1148.

- Kerlan C (1996) Maladies à virus. In: *La Pomme de Terre : Production, Amélioration, Ennemis et Maladies, Utilisation* (Rousselle P, Robert Y, Crosnier JC eds.), pp. 232-260. INRA Editions, Paris, France.
- Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics*, **167**, 1513-1524.
- Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, **160**, 765-777.
- Kimura R, Fujimoto A, Tokunaga K, Ohashi J (2007) A practical genome scan for populationspecific strong selective sweeps that have reached fixation. *PLoS ONE*, **2**, e286.
- Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, **23**, 482-490.
- Kofler R, Nolte V, Schlötterer C (2016) The impact of library preparation protocols on the consistency of allele frequency estimates in Pool-Seq data. *Molecular Ecology Resources*, **16**, 118-122.
- Kort J, Ross H, Rumpenhorst HJ, Stone AR (1977) An international scheme for identifying and classifying pathotypes of potato cyst-nematodes *Globodera rostochiensis* and *G. pallida*. *Nematologica*, **23**, 333-339.
- Kreike CM, Dekoning JRA, Vinke JH, Vanooijen JW, Gebhardt C, Stiekema WJ (1993) Mapping of loci involved in quantitatively inherited resistance to the potato cyst-nematode *Globodera rostochiensis* pathotypes Ro1. *Theoretical and Applied Genetics*, **87**, 464-470.
- Kreike CM, Kok-Westeneng AA, Vinke JH, Stiekema WJ (1996) Mapping of QTLs involved in nematode resistance, tuber yield and root development in *Solanum* sp. *Theoretical and Applied Genetics*, **92**, 463-470.
- Kudla U, Qin L, Milac A, Kielak A, Maissen C, Overmars H, Popeijus H, Roze E, Petrescu A, Smant G (2005) Origin, distribution and 3D-modeling of Gr-EXPB1, an expansin from the potato cyst nematode *Globodera rostochiensis*. *Federation of European Biochemical Societies Letters*, **579**, 2451-2457.
- Laine AL, Barrès B (2013) Epidemiological and evolutionary consequences of life-history tradeoffs in pathogens. *Plant Pathology*, **62**, 96-105.
- Lange W, Müller J, De Bock TSM (1993) Virulence in the beet cyst nematode (*Heterodera schachtii*) versus some aliens genes for resistance in beet. *Fundamental and Applied Nematology*, **16**, 447-454.
- Langella O (2000) POPULATIONS 1.2: population genetic software, individuals or population distance, phylogenetic trees. [http://bioinformatics.org/~tryphon/populations/].
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357-359.
- Langmead B, Trapnell Cole, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.
- Lasserre F, Gigault F, Gauthier JP, Henry JP, Sandmeier M, Rivoal R (1996) Genetic variation in natural populations of the cereal cyst nematode (*Heterodera avenae* Woll.) submitted to resistant and susceptible cultivars of cererals. *Theoretical and Applied Genetics*, **93**, 1-8.
- Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548-1566.
- Leach JE, Vera Cruz CM, Bai J, Leung H (2001) Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annual Review of Phytopathology*, **39**, 187-224.
- Lecoq H, Moury B, Desbiez C, Palloix A, Pitrat M (2004) Durable virus resistance in plants through conventional approaches: a challenge. *Virus Research*, **100**, 31-39.

- Lee S, Lee EJ, Yang EJ, Lee JE, Park AR, Song WH, Park OK (2004) Proteomic identification of annexins, calcium-dependent membrane binding protein that mediate osmotic stress and abscisic acid signal transduction in *Arabidopsis*. *Plant Cell*, **16**, 1378-1391.
- Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *The American Naturalist*, **138**, 1315-1341.
- Leonard KJ, Czochor RJ (1980) Theory of genetic interactions among populations of plants and their pathogens. *Annual Review of Phytopathology*, **18**, 237-258.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175-195.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
- Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK (1999) Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics*, **152**, 1183-1202.
- Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, **23**, 2178-2192.
- Lozano-Torres JL, Wilbers RH, Gawronski P, Boshoven JC, Finkers-Tomczak A, Cordewener JH, America AHP, Overmars HA, Van 't Klooster JW, Baranowski L, Sobczak M, Ilyas M, van der Hoorn RAL, Schots A, de Wit PJGM, Bakker J, Goverse A, Smant G (2012) Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 10119-10124.
- Lozano-Torres JL, Wilbers RH, Warmerdam S, Finkers-Tomczak A, Diaz-Granados A, van Schaik CC, Helder J, Bakker J, Goverse A, Schots A, Smant G (2014) Apoplastic venom allergen-like proteins of cyst nematodes modulate the activation of basal plant innate immunity by cell surface receptors. *PLoS Pathogens*, **10**, e1004569.
- Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, **9**, e1003215.
- Lu SW, Chen S, Wang J, Yu H, Chronis D, Mitchum MG, Wang X (2009) Structural and functional diversity of *CLAVATA3/ESR* (*CLE*)-like genes from the potato cyst nematode *Globodera rostochiensis*. *Molecular Plant-Microbe Interactions*, **22**, 1128-1142.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981-994.
- Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, **24**, 133-141.
- Martins NE, Faria VG, Nolte V, Schlötterer C, Teixeira L, Sucena E, Magalhães S (2014) Host adaptation to viruses relies on few genes with different cross-resistance properties. *Proceedings of National Academy of Sciences of the United States of America*, **111**, 5938-5943.
- Mascher M, Wu S, St. Amand P, Stein N, Poland J (2013) Application of genotyping-bysequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in Barley. *PLoS ONE*, **8**, e76925.

- Maynard-Smith J, Haigh J (1974) Hitch-hiking effect of a favorable gene. *Genetics Research*, **23**, 23-35.
- McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, **40**, 349-379.
- McEvoy B, Beleza S, Shriver MD (2006) The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Human Molecular Genetics*, **15**, R176-R181.
- McHale L, Tan XP, Koehl P, Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. *Genome Biology*, **7**, 212.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytski A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297-1303.
- Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a userfriendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*, **26**, 403-404.
- Mei Y, Thorpe P, Guzha A, Haegeman A, Blok VC, MacKenzie K, Gheysen G, Jones JT, 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*. *Nematology*, **17**, 409-424.
- Meshi T, Motoyoshi F, Adachi A, Watanabe Y, Takamatsu N, Okada Y (1988) Two concomitant base substitutions in the putative replicase genes of tobacco mosaic virus confer the ability to overcome the effects of a tomato resistance gene, *Tm-1*. *EMBO Journal*, **7**, 1575-1581.
- Mimee B, Duceppe O, Véronneau PY, 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 frequency analysis. *Molecular Ecology Resources*, **15**, 1356-1365.
- Mitchum MG, Hussey RS, Baum TJ, Wang X, Elling AA, Wubben M, Davies EL (2013) Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist*, **199**, 879-894.
- Molinari S (2011) Natural genetic and induced plant resistance, as a control strategy to plantparasitic nematodes alternative to pesticides. *Plant Cell Reports*, **30**, 311-323.
- Montarry J, Cartier E, Jacquemond M, Palloix A, Moury B (2012) Virus adaptation to quantitative plant resistance: erosion or breakdown? *Journal of Evolutionary Biology*, **25**, 2242-2252.
- Montarry J, Hamelin FM, Glais I, Corbière R, Andrivon D (2010) Fitness costs associated with unnecessary virulence factors and life history traits: evolutionary insights from the potato late blight pathogen *Phytophthora infestans*. *BMC Evolutionary Biology*, **10**, 283.
- Montarry J, Jan PL, Gracianne C, Overall ADJ, Bardou-Valette S, Olivier E, Fournet S, Grenier E, Petit EJ (2015) Heterozygote deficits in cyst plant-parasitic nematodes: possible causes and consequences. *Molecular Ecology*, **24**, 1654-1677.
- Morand S (1996) Life-history traits in parasitic nematodes: a comparative approach for the search of invariants. *Functional Ecology*, **10**, 210-218.
- Mugniéry D, Balandras C (1986) Test de résistance de descendance de pomme de terre à *Globodera rostochiensis* Woll. *Potato Research*, **29**, 131-140.
- Mugniéry D, Fayet G (1981) Sex determination in *Globodera pallida*. *Revue de Nematologie*, **4**, 41-45.

- Mugniéry D, Fayet G (1984) Détermination du sexe de *Globodera rostochiensis* Woll. et influence des niveaux d'infestation sur la pénétration, le développement et le sexe de ce nématode. *Revue de Nematologie*, **7**, 233-238.
- Mugniéry D, Person F (1976) Méthode d'élevage de quelques nématodes à kyste du genre *Heterodera. Sciences Agronomiques Rennes*, 217-220.
- Mugniéry D, Plantard O, Fournet S, Grenier E, Caromel B, Kerlan MC, Picard D, Ellissèche D (2007) Evaluation of effectiveness and durability of resistance to *Globodera pallida* from *Solanum vernei*, *S. spegazzinii* and *S. sparsipilum*. *Nematologia mediterranea*, **35**, 143-153.
- Müller J (1992) Detection of pathotypes by assessing the virulence of *Heterodera schachtii* populations. *Nematologica*, **28**, 50-64.
- Mulvey RH, Stone AR (1976) Description of *Punctodera matadorensis* n.gen., n.sp. (Nematoda: Heteroderidae) from Saskatchewan with lists of species and generic diagnoses of *Globodera* (n. rank), *Heterodera*, and *Sarisodera*. *Canadian Journal of Zoology*, **54**, 772-785.
- Mundt CC (2014) Durable resistance: a key to sustainable management of pathogens and pests. *Infection, Genetics and Evolution*, **27**, 446-455.
- Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics*, **156**, 297-304.
- Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **367**, 409-421.
- Nadeau NJ, Jiggins CD (2010) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics*, **26**, 484-492.
- Nei M, Maruyama T (1975) Lewontin-Krakauer test for neutral genes. *Genetics*, **80**, 395.
- Nelson RR (1979) The evolution of parasitic fitness. In: *Plant Disease* (Horsfall JG, Cowling EB eds.), pp. 23-46. Academic Press, New York, USA.
- Nicholson G, Smith AV, Jonsson F, Gustafsson O, Stefansson K, Donnelly P (2002) Assessing population differentiation and isolation from single-nucleotide polymorphism data. *Journal of The Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 695-715.
- Nicol JM, Turner SJ, Coyne DL, den Nijs L, Hockland S, Maafi ZT (2011) Current nematode threats to world agriculture. In: *Genomics and Molecular Genetics of Plant-Nematode Interactions* (Jones J, Gheysen G, Fenoll C eds.), pp. 21-43. Springer, Dordrecht, The Netherlands.
- Nijboer H, Parlevliet JE (1990) Pathotype-specificity in potato cyst nematodes, a reconsideration. *Euphytica*, **49**, 39-47.
- Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, **103**, 439-444.
- Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walkingstick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316-336.
- Nürnberger T, Lipka V (2005) Non-host resistance in plants: new insights into an old phenomenom. *Molecular Plant Pathology*, **6**, 335-345.
- Oerke EC, Dehne HW, Schönbeck F, Weber A (1994) Crop Production and Crop Protection: Estimated Losses in Major Food and Cash Crop. Elsevier Science BV, Amsterdam, The Netherlands.
- Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119-127.

- Paal J, Henselewski H, Muth J, Meksem K, Menendez CM, Salamini F, Ballvora A, Gebhardt C (2004) Molecular cloning of the potato *Gro1-4* gene conferring resistance to pathotype Ro1 of the root cyst nematode *Globodera rostochiensis*, based on a candidate gene approach. *Plant Journal*, **38**, 285-297.
- Palloix A, Ayme V, Moury B (2009) Durability of plant major resistance genes to pathogens depends on the genetic background: experimental evidence and consequences for breeding strategies. *New Phytologist*, **183**, 190-199.
- Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, **6**, 445-464.
- Parlevliet JE (1979) Components of resistance that reduce the rate of epidemic development. *Annual Review of Phytopathology*, **17**, 203-222.
- Parlevliet JE (2002) Durability of resistance against fungal, bacterial and viral pathogens; present situation. *Euphytica*, **124**, 147-156.
- Parlevliet JE, Vanommeren A (1988) Accumulation of partial resistance in barley to barley leaf rust and powdery mildew through recurrent selection against susceptibility. *Euphytica*, **37**, 261-274.
- Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**, 1061-1067.
- Pavey SA, Bernatchez L, Aubin-Horth N, Landry CR (2012) What is needed for next-generation ecological and evolutionary genomics? *Trends in Ecology & Evolution*, **27**, 673-678.
- Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution*, **29**, 3237-3248.
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SIGNALP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, **8**,785-786.
- Phillips MS, Blok V, Ploeg A, Harrower BE (1998) Studies on an artificially fragmented population of potato cyst nematode *Globodera pallida*. *Nematologica*, **44**, 655-666.
- Picard D, Plantard O (2006) What constitutes a population for the plant parasitic nematode *Globodera pallida* in its native area (Peru)? *International Journal for Parasitology*, **36**, 115-122.
- Picard D, Plantard O, Scurrah M, Mugniéry D (2004) Inbreeding and population structure of the potato cyst nematode (*Globodera pallida*) in its native area (Peru). *Molecular Ecology*, 13, 2899-2908.
- Picard D, Sempere T, Plantard O (2007) A northward colonisation of the Andes by the potato cyst nematode during geological times suggests multiple host-shifts from wild to cultivated potatoes. *Molecular Phylogenetics and Evolution*, **42**, 308-316.
- Picard D, Sempere T, Plantard O (2008) Direction and timing of uplift propagation in the Peruvian Andes deduced from molecular phylogenetics of highland biotaxa. *Earth and Planetary Science Letters*, **271**, 326-336.
- Pineda O, Bonierbale MW, Plaisted RL, Brodie BB, Tanksley SD (1993) Identification of RFLP markers linked to the *H1* gene conferring resistance to the potato cyst nematode *Globodera rostochiensis. Genome*, **36**, 152-156.
- Pink DAC (2002) Strategies using genes for non-durable disease resistance. *Euphytica*, **124**, 227-236.
- 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. *Molecular Ecology*, **17**, 2208-2218.

- Plantard O, Porte C (2004) Population genetic structure of the sugar beet cyst nematode *Heterodera schachtii*: a gonochoristic and amphimictic species with highly inbred but weakly differentiated populations. *Molecular Ecology*, **13**, 33-41.
- Poinar G, Boucot AJ (2006) Evidence of intestinal parasites of dinosaurs. *Parasitology*, **133**, 245-249.
- Poinar GO, Poinar R (2003) Description and development of *Gastromermis anisotis* sp. n. (Nematoda: Mermithidae), a parasite in a quadritrophic system involving a cyanobacterium, midge and virus. *Nematology*, **5**, 325-338.
- Poulicard N, Pinel-Galzi A, Hebrard E, Fargette D (2010) Why *Rice yellow mottle virus*, a rapidly evolving RNA plant virus, is not efficient at breaking *rymv1-2* resistance. *Molecular Plant Pathology*, **11**, 145-154.
- Poulicard N, Pinel-Galzi A, Traore O, Vignols F, Ghesquiere A, Konate G, Hebrard E, Fargette D (2012) Historical contingencies modulate the adaptability of *Rice yellow mottle virus*. *PLoS Pathogens*, **8**, e1002482.
- Powers TO, Harris TS (1993) A polymerase chain reaction method for identification of five major *Meloidogyne* species. *Journal of Nematology*, **25**, 1-6.
- Prior A, Jones JT, Blok VC, Beauchamp J, McDermott L, Cooper A, Kennedy MW (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. *Biochemical Journal*, **356**, 387-394.
- Pritchard JK, Di Rienzo A (2010) Adaptation not by sweeps alone. *Nature Reviews Genetics*, **11**, 665-667.
- Pritchard L, Birch P (2011) A systems biology perspective on plant-microbe interactions: biochemical and structural targets of pathogen effectors. *Plant Science*, **180**, 584-603.
- Przeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic varation. *Evolution*, **59**, 2312-2323.
- Qin L, Kudla U, Roze E, Goverse A, Popeijus H, Nieuwland J, Overmars H, Jones JT, Schots A, Smant G, Bakker J, Helder J (2004) A nematode expansin acting on plants. *Nature*, **427**, 30.
- R Core Team (2014) R: a language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. https://www.R-project.org/.
- Rehman S, Butterbach P, Popeijus H, Overmars H, Davis EL, Jones JT, Goverse A, Bakker J, Smant G (2009) Identification and characterization of the most abundant cellulases in stylet secretions from *Globodera rostochiensis*. *Phytopathology*, **99**, 194-202.
- Rehman S, Postma W, Tytgat T, Prins P, Qin L, Overmars H, Vossen J, Spiridon LN, Petrescu AJ, Goverse A, Bakker J, 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. *Molecular Plant-Microbe Interactions*, **22**, 330-340.
- Reiss MJ (1989) *The Allometry of Growth and Reproduction*. Cambridge University Press, New York, USA.
- Rellstab C, Zoller S, Tedder A, Gugerli F, Fischer MC (2013) Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PLoS ONE*, **8**, e80422.
- Rhodes J, Beale MA, Fisher MC (2014) Illuminating choices for library prep: a comparison of library preparation methods for whole genome sequencing of *Cryptococcus neoformans* using Illumina HiSeq. *PLoS ONE*, **9**, e113501.

- Rice SL, Leadbeater BSC, Stone AR (1985) Changes in cell structure in roots of resistant potatoes parasitized by potato cyst nematodes. 1. Potatoes with resistance gene *H1* derived from *Solanum tuberosum* ssp. *andigena*. *Physiological Plant Pathology*, **27**, 219-234.
- Rice SL, Stone AR, Leadbeater BSC (1987) Changes in cell structure in roots of resistant potatoes parasitized by potato cyst nematodes. 2. Potatoes with resistance derived from *Solanum vernei*. *Physiological and Molecular Plant Pathology*, **31**, 1-14.
- Rickert AM, Kim JH, Meyer S, Nagel A, Ballvora A, Oefner PJ, Gebhardt C (2003) Firstgeneration SNP/InDel markers tagging loci for pathogen resistance in the potato genome. *Plant Biotechnology Journal*, **1**, 399-410.
- Ridley M (1996) Born to trade. New Scientist, 152, 34-37.
- Riebler A, Held L, Stephan W (2008) Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics*, **178**, 1817-1829.
- Ritter E, Debener T, Barone A, Salamini F, Gebhardt C (1991) RFLP mapping on potato chromosomes of 2 genes-controlling extreme resistance to *Potato virus X* (PVX). *Molecular & General Genetics*, **227**, 81-85.
- Robertson A (1975) Remarks on the Lewontin-Krakauer test. *Genetics*, **80**, 396.
- Robertson L, Robertson WM, Jones JT (1999) Direct analysis of the secretions of the potato cyst nematode *Globodera rostochiensis*. *Parasitology*, **119**, 167-176.
- Rodríguez F, Spooner DM (2009) Nitrate reductase phylogeny of potato (*Solanum* sect. *Petota*) genomes with emphasis on the origins of the polyploid species. *Systematic Botany*, **34**, 207-219.
- Roff DA (1992) *The Evolution of Life Histories: Theory and Analysis*. Chapman and Hall, New York, USA.
- Ross H (1986) *Potato breeding Problems and Perspective*. Parey VP, Berlin-Hambourg, Germany.
- Rotmistrovsky K, Jang W, Schuler GD (2004) A web server for performing electronic PCR. *Nucleic Acids Research*, **32**, W108-W112.
- Rouppe van der Voort JN, Kanyuka K, van der Vossen E, Bendahmane A, Mooijman P, Klein-Lankhorst RM, Stiekema WJ, Baulcombe DC, Bakker J (1999) Tight physical linkage of the nematode resistance gene *Gpa2* and the virus resistance gene *Rx* on a single segment introgressed from the wild species *Solanum tuberosum* subsp *andigena* CPC1673 into cultivated potato. *Molecular Plant-Microbe Interactions*, **12**, 197-206.
- Rouppe van der Voort JN, van Eck HJ, Draaistra J, van Zandvoort PM, Jacobsen E, Bakker J (1998) An online catalogue of AFLP markers covering the potato genome. *Molecular Breeding*, **4**, 73-77.
- Rouppe van der Voort JN, Wolters P, Folkertsma RT, Hutten RC, van Zandvoort PM, Vinke H, Kanyuka K, Bendahmane A, Jacobsen E, Janssen R, Bakker J (1997) Mapping of the cyst nematode resistance locus *Gpa2* in potato using a strategy based on comigrating AFLP markers. *Theoretical and Applied Genetics*, **95**, 874-880.
- Rouppe van der Voort JNAM, van der Vossen E, Bakker E, Overmars H, van Zandvoort P, Hutten R, Klein Lankhorst R, Bakker J (2000) Two additive QTLs conferring broad spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters. *Theoretical and Applied Genetics*, **101**, 1122-1130.
- Rouppe van der Voort JNAM, van Eck HJ, van Zandvoort PM, Overmars H, Helder J, Bakker J (1999) Linkage analysis by genotyping of sibling populations: a genetic map for the potato

cyst nematode constructed using a "pseudo-F2" mapping strategy. *Molecular and General Genetics*, **261**, 1021-1031.

- Rousselle-Bourgeois F, Mugniéry D (1995) Screening tuber-bearing *Solanum* spp. for resistance to *Globodera rostochiensis* Ro1 Woll and *G. pallida* Pa2/3 Stone. *Potato Research*, **38**, 241-249.
- Rousselle-Bourgeois F, Rousselle P (1996) Amélioration génétique. In: *La Pomme de Terre : Production, Amélioration, Ennemis et Maladies, Utilisation* (Rousselle P, Robert Y, Crosnier JC eds.), pp. 125-159. INRA Editions, Paris, France.
- Rousset F (2008) GENEPOP'007: a complete reimplementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.
- Rouxel T, Penaud A, Pinochet X, Brun H, Gout L, Delourme R, Schmit J, Balesdent MH (2003) A 10-year survey of populations of *Leptosphaeria maculans* in France indicates a rapid adaptation towards the *Rlm1* resistance gene of oilseed rape. *European Journal of Plant Pathology*, **109**, 871-881.
- Rubin CJ, Megens HJ, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg Ö, Jern P, Jørgensen CB, Archibald AL, Fredholm M, Groenen MAM, Andersson L (2012) Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 19529-19536.
- Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, Hallböök F, Besnier F, Carlborg Ö, Bed'hom B, Tixier-Boichard M, Jensen P, Siegel P, Lindblad-Toh K, Andersson L (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, **464**, 587-591.
- Russello MA, Kirk SL, Frazer KK, Askey PJ (2012) Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications*, **5**, 39-52.
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832-837.
- Sacco MA, Koropacka K, Grenier E, Jaubert MJ, Blanchard A, Goverse A, Smant G, Moffett P (2009) The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death. *PLoS Pathogens*, **5**, e1000564.
- Sacristán S, García-Arenal F (2008) The evolution of virulence and pathogenicity in plant pathogen populations. *Molecular Plant Pathology*, **9**, 369-384.
- Sanderson MJ (2003) Molecular data from 27 proteins do not support a Precambrian origin of land plants. *American Journal of Botany*, **90**, 954-956.
- Savary S, Teng PS (1994) La protection des cultures dans une agriculture durable. *La Recherche*, **25**, 1322-1329.
- Schafer JF (1971) Tolerance to plant disease. Annual Review of Phytopathology, 9, 235-252.
- Schenck S, Kendrick WB, Pramer D (1977) New nematode-trapping hyphomycete and a reevaluation of *Dactylaria* and *Arthrobotrys*. *Canadian Journal of Botany*, **55**, 977-985.
- Schiff CL, Wilson IW, Somerville SC (2001) Polygenic powdery mildew disease resistance in *Arabidopsis thaliana*: quantitative trait analysis of the accession Warschau-1. *Plant Pathology*, **50**, 690-701.
- Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics*, **160**, 753-763.

- Schlötterer C, Dieringer D (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: *Selective Sweep* (Nurminsky D ed.), pp. 55-64. Landes Bioscience, Georgetown, USA.
- Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals mining genome-wide polymorphism data without big funding. *Nature Review Genetics*, **15**, 749-763.
- Schoustra SE, Debets AJM, Slakhorst M, Hoekstra RF (2006) Reducing the cost of resistance; experimental evolution in the filamentous fungus *Aspergillus nidulans*. *Journal of Evolutionary Biology*, **19**, 1115-1127.
- Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615-629.
- Semblat JP, Rosso MN, Hussey RS, Abad P, Castagnone-Sereno P (2001) Molecular cloning of a cDNA encoding an amphid-secreted putative avirulence protein from the root-knot nematode *Meloidogyne incognita*. *Molecular Plant-Microbe Interactions*, **14**, 72-79.
- Shimizu K, Purugganan M (2005) Evolutionary and ecological genomics of *Arabidopsis*. *Plant Physiology*, **138**, 578-584.
- Sidhu G, Mohan A, Zheng P, Dhaliwal AK, Main D, Gill KS (2015) Sequencing-based high throughput mutation detection in bread wheat. *BMC Genomics*, **16**, 962.
- Simms EL, Triplett J (1994) Costs and benefits of plant-responses to disease resistance and tolerance. *Evolution*, **48**, 1973-1985.
- Skarbilovich TS (1959) On the structure of the systematics of nematode order Tylenchida Thorne, 1949. *Acta Parasitologica Polonica*, **7**, 117-132.
- Smant G, Jones J (2011) Suppression of plant defences by nematodes. In: *Genomics and Molecular Genetics of Plant-Nematode Interactions* (Jones JT, Gheysen G, Fenoll C eds.), pp. 273-286. Springer, Dordrecht, The Netherlands.
- Smant G, Stokkermans JP, Yan Y, de Boer JM, Baum TJ, Wang X, Hussey RS, Gommers FJ, Henrissat B, Davis EL, Helder J, Schots A, Bakker J (1998) Endogenous cellulases in animals: isolation of β-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 4906-4911.
- Smithson JB, Lenné JM (1996) Varietal mixtures: a viable strategy for sustainable productivity in subsistence agriculture. *Annals of Applied Biology*, **128**, 127-158.
- Sobczak M, Golinowski W (2011) Cyst nematodes and syncytia. In: *Genomics and Molecular Genetics of Plant-Nematode Interactions* (Jones JT, Gheysen G, Fenoll C eds.), pp. 61-82. Springer, Dordrecht, The Netherlands.
- Sosamma VK, Koshy PK (1997) Biological control of *Meloidogyne incognita* on black pepper by *Pasteuria penetrans* and *Paecilomyces lilacinus. Journal of Plantation Crops*, **25**, 72-76.
- Spire D, Rousselle P (1996) Origine socio-historique. In: *La Pomme de Terre : Production, Amélioration, Ennemis et Maladies, Utilisation* (Rousselle P, Robert Y, Crosnier JC eds.), pp. 25-48. INRA Editions, Paris, France.
- Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature*, **400**, 667-671.
- Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, 25, 705-712.
- Stearns SC (1989) Trade-offs in life-history evolution. *Functional Ecology*, **3**, 259-268.

- Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158-170.
- Stone AR (1973) *Heterodera pallida* n. sp. (Nematoda: Heteroderidae), a second species of potato cyst nematode. *Nematologica*, **18**, 591-606.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671-688.
- Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, **46**, 75-100.
- Subbotin SA, Mundo-Ocampo M, Baldwin JG (2010) *Nematology Monographs & Perspectives Vol. 8A, Systematics of Cyst Nematodes (Nematoda: Heteroderinae)*. Brill, Boston, USA.
- Subbotin SA, Sturhan D, Chizhov VN, Vovlas N, Baldwin JG (2006) Phylogenetic analysis of Tylenchida Thorne, 1949 as inferred from D2 and D3 expansion fragments of the 28S rRNA gene sequences. *Nematology*, **8**, 455-474.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.
- Takala SL, Smith DL, Stine OC, Coulibaly D, Thera MA, Doumbo OK, Plowe CV (2006) A highthroughput method for quantifying alleles and haplotypes of the malaria vaccine candidate *Plasmodium falciparum* merozoite surface protein-I 19 kDa. *Malaria Journal*, **5**, 31.
- Teixeira MA, Wei L, Kaloshian I (2016) Root-knot nematodes induce pattern-triggered immunity in *Arabidopsis thaliana* roots. *New Phytologist*, **211**, 276-287.
- Tellier A, Brown JKM (2007) Stability of genetic polymorphism in host-parasite interactions. *Proceedings of the Royal Society of London B: Biological sciences*, **274**, 809-817.
- Tellier A, Brown JKM (2009) The influence of perenniality and seed banks on polymorphism in plant-parasite interactions. *American Naturalist*, **174**, 769-779.
- Terauchi R, Yoshida K (2010) Towards population genomics of effector-effector target interactions. *New Phytologist*, **187**, 929-939.
- Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps? *Genome Research*, **16**, 702-712.
- The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, **475**, 189-195.
- The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635-641.
- Thiéry M, Mugniéry D (1996) Interspecific rDNA restriction fragment length polymorphisms in *Globodera* species, parasites of Solanaceous plants. *Fundamental and Applied Nematology*, **19**, 471-479.
- Thomas WK, Vida JT, Frisse LM, Mundo M, Baldwin JG (1997) DNA sequences from formalinfixed nematodes: integrating molecular and morphological approaches to taxonomy. *Journal of Nematology*, **29**, 250-254.
- Thrall PH, Burdon JJ (2003) Evolution of virulence in a plant host-pathogen metapopulation. *Science*, **299**, 1735-1737.
- Torkamaneh D, Laroche J, Belzile F (2016) Genome-wide SNP calling from genotyping by sequencing (GBS) data: a comparison of seven pipelines and two sequencing technologies. *PLoS ONE*, **11**, e0161333.
- Triantaphyllou AC, Esbenshade PR (1990) Demonstration of multiple mating in *Heterodera glycines* with biochemical markers. *Journal of Nematology*, **22**, 452-456.

- Trudgill D, Elliott M, Evans K, Phillips M (2003) The white potato cyst nematode (*Globodera pallida*) a critical analysis of the threat in Britain. *Annals of Applied Biology*, **143**, 73-80.
- Turner SJ (1989) New sources of resistance to potato cyst-nematodes in the Commonwealth Potato Collection. *Euphytica*, **42**, 145-153.
- Turner SJ (1990) The identification and fitness of virulent potato cyst-nematode populations (*Globodera pallida*) selected on resistant *Solanum vernei* hybrids for up to eleven generations. *Annals of Applied Biology*, **117**, 385-397.
- Turner SJ, Fleming CC (2002) Multiple selection of potato cyst nematode *Globodera pallida* virulence on a range of potato species. I. Serial selection on *Solanum*-hybrids. *European Journal of Plant Pathology*, **108**, 461-467.
- Turner SJ, Rowe JA (2006) Cyst nematodes. In: *Plant Nematology* (Perry RN, Moens M eds.), pp. 91-122. CAB International, Wallingford, UK.
- Turner TL, Hahn MW (2010) Genomic islands *of* speciation or genomic islands *and* speciation? *Molecular Ecology*, **19**, 848-850.
- Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM (2011) Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genetics*, **7**, e1001336.
- Tytgat T, Vanholme B, De Meutter J, Claeys M, Couvreur M, Vanhoutte I, Gheysen G, Van Criekinge W, Borgonie G, Coomans A (2004) A new class of ubiquitin extension proteins secreted by the dorsal pharyngeal gland in plant parasitic cyst nematodes. *Molecular Plant-Microbe Interactions*, **17**, 846-852.
- Uricaru R, Rizk G, Lacroix V, Quillery E, Plantard O, Chikhi R, Lemaitre C, Peterlongo P (2015) Reference-free detection of isolated SNPs. *Nucleic Acids Research*, **43**, e11.
- Van de Wouw AP, Howlett BJ (2012) Estimating frequencies of virulent isolates in field populations of a plant pathogenic fungus, *Leptosphaeria maculans*, using high-throughput pyrosequencing. *Journal of Applied Microbiology*, **113**, 1145-1153.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, **43**, 11.10.1-11.10.33.
- van der Hoorn RAL, Kamoun S (2008) From guard to decoy: a new model for perception of plant pathogen effectors. *Plant Cell*, **20**, 2009-2017.
- van der Vossen E, Rouppe van der Voort JN, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, Bakker J, Stiekema WJ, Klein-Lankhorst RM (2000) Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. *Plant Journal*, **23**, 567-576.
- van Riel HR, Mulder A (1998) Potato cyst nematodes (*Globodera* species) in western Europe. In: *Potato Cyst Nematodes: Biology, Distribution and Control* (Marks RJ, Brodie BB eds.), pp. 271-298. CAB International, Wallingford, UK.
- Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, **5**, 247-252.
- Vanderplank JE (1963) *Plant Diseases Epidemics and Control.* Academic Press, New York, USA.
- Vanderplank JE (1968) Disease Resistance in Plants. Academic Press, New-York, USA.

- Vasemägi A, Nilsson J, McGinnity P, Cross T, O'Reilly P, Glebe B, Peng B, Berg PR, Primmer CR (2012) Screen for footprints of selection during domestication/captive breeding of Atlantic Salmon. *Comparative and Functional Genomics*, **2012**, 628204.
- Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623-3642.
- Vera Cruz CM, Bai J, Ona I, Leung H, Nelson RJ, Mew TW, Leach JE (2000) Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 13500-13505.
- Vercauteren I, Engler JD, De Groodt R, Gheysen G (2002) An *Arabidopsis thaliana* pectin acetylesterase gene is upregulated in nematode feeding sites induced by root-knot and cyst nematodes. *Molecular Plant-Microbe Interactions*, **15**, 404-407.
- Vitalis R (2012) DETSEL: an R-package to detect marker loci responding to selection. In: Data Production and Analysis in Population Genomics: Methods and Protocols: Methods in Molecular Biology (Pompanon F, Bonin A eds.), pp. 277-293. Springer Science + Business Media, New York, USA.
- Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, **158**, 1811-1823.
- Vitalis R, Gautier M, Dawson KJ, Beaumont MA (2014) Detecting and measuring selection from gene frequency data. *Genetics*, **196**, 799-817.
- von Mende N (1997) Invasion and migration behaviour of sedentary nematodes. In: *Cellular and Molecular Aspects of Plant–Nematodes Interactions* (Fenoll C, Grundler FMW, Ohl SA eds.), pp. 51-64. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Waetzig GH, Sobczak M, Grundler FMW (1999) Localization of hydrogen peroxide during the defence response of *Arabidopsis thaliana* against the plant-parasitic nematode *Heterodera glycines*. *Nematology*, **1**, 681-686.
- Wang DYC, Kumar S, Hedges SB (1999) Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proceedings of the Royal Society of London B: Biological sciences*, **266**, 163-171.
- Watterson GA (1978) The homozygosity test of neutrality. *Genetics*, **88**, 405-417.
- Wei C, Liu J, Yu Z, Zhang B, Gao G, Jiao R (2013) TALEN or Cas9 Rapid, efficient and specific choices for genome modifications. *Journal of Genetics and Genomics*, **40**, 281-289.
- Wolfe MS (1985) The current status and prospects of multiline cultivars and variety mixtures for disease resistance. *Annual Review of Phytopathology*, **23**, 251-273.
- Wolfe MS (1993) Can the strategic use of disease resistant hosts protect their inherent durability? In: *Durability of Disease Resistance* (Jacobs T, Parlevliet JE eds.), pp. 83-96. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Wolters P, Vinke H, Bontjer I, Rouppe van der Voort J, Colon L, Hoogendoorn C (1999) Mapping of major genes for resistance to *Globodera pallida* in wild *Solanum* species. Plant & Animal Genome VII Conference, San Diego, USA, 17-21 January 1999.
- Woo JS, Imm JH, Min CK, Kim KJ, Cha SS, Oh BH (2006) Structural and functional insights into the B30.2/SPRY domain. *EMBO Journal*, **25**, 1353-1363.
- Wood DE, Salzberg S (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, **15**, R46.

- Wuyts N, Lognay G, Verscheure M, Marlier M, De Waele D, Swennen R (2007) Potential physical and chemical barriers to infection by the burrowing nematode *Radopholus similis* in roots of susceptible and resistant banana (*Musa* spp.). *Plant Pathology*, **56**, 878-890.
- Wuyts N, Swennen R, De Waele D (2006) Effects of plant phenylpropanoid pathway products and selected terpenoids and alkaloids on the behaviour of the plant-parasitic nematodes *Radopholus similis, Pratylenchus penetrans* and *Meloidogyne incognita*. *Nematology*, **8**, 89-101.
- Zhu Y, Bergland AO, González J, Petrov DA (2012) Empirical validation of pooled whole genome population re-sequencing in *Drosophila melanogaster*. *PLoS ONE*, **7**, e41901.
ANNEXES

ANNEXE 1 : SUPPORTING INFORMATION – ARTICLE 1

Data S1

Nematode females produce cysts which contain eggs, which may hatch and produce larvae.

We aim to model the dynamics of larvae production to better compare virulent and avirulent lineages obtained from experimental evolution on resistant and susceptible host genotypes, respectively.

Data We have

• the times at which newly hatched larvae are collected and counted:

$$\{t_i\} = \{t_0, t_1, t_2, \cdots, t_k\},\$$

• the number of larvae corresponding to each observation time:

$$\{n_i\} = \{n_0, n_1, n_2, \cdots, n_k\}.$$

• At the end of the experiment (time t_k), cysts are dessicated and unhatched eggs are counted: let their number be n_u , so $n_u + \sum_{i=0}^k n_i = N$ is the total number of eggs or potential larvae.

We are interested in the proportion of hatched larvae as a function of time, say H(t). Let $H_d(t_i) = \sum_{j=0}^i n_j / N$ be the observed dynamics.

Model We assume that each egg is viable with probability P (hatching probability). We also assume that the time-to-hatching is a random variable T with a cumulative distribution F(t) which depends on the parameter Θ , and which support is $[0, +\infty)$. Let $\theta = \{P, \Theta\}$ be the set of parameters to be estimated. Also, let $H_{\theta}(t) = PF(t)$ be the theoretical proportion of hatched larvae.

The probability to observe the data given a parameter set θ (its likelihood) is

$$L(\theta) = \underbrace{\begin{pmatrix} N \\ n_0 \end{pmatrix} H_{\theta}(t_0)^{n_0} (1 - H_{\theta}(t_0))^{N-n_0}}_{\text{Probability}(n_0 \text{ hatched before } t_0)} \times \prod_{i=1}^k \underbrace{\begin{pmatrix} N - \sum_{j=0}^{i-1} n_j \\ n_i \end{pmatrix} \left(\frac{H_{\theta}(t_i) - H_{\theta}(t_{i-1})}{1 - H_{\theta}(t_{i-1})}\right)^{n_i} \left(1 - \frac{H_{\theta}(t_i) - H_{\theta}(t_{i-1})}{1 - H_{\theta}(t_{i-1})}\right)^{N-\sum_{j=0}^{i} n_j}}_{\text{Probability}(n_i \text{ hatched between } t_{i-1} \text{ and } t_i | \text{ they did not hatch before})}}$$

Because the combinations do not depend on θ , we define a log-likelihood function as

$$LL(\theta) = n_0 \log(H_{\theta}(t_0)) + (N - n_0) \log(1 - H_{\theta}(t_0)) + \sum_{i=1}^k \left[n_i \log\left(\frac{H_{\theta}(t_i) - H_{\theta}(t_{i-1})}{1 - H_{\theta}(t_{i-1})}\right) + (N - \sum_{j=0}^i n_j) \log\left(1 - \frac{H_{\theta}(t_i) - H_{\theta}(t_{i-1})}{1 - H_{\theta}(t_{i-1})}\right) \right]$$

Assuming that the time-to-hatching is Weibull-distributed, we let $\Theta = \{\lambda, k\}$ and

$$F(t) = 1 - e^{-\left(\frac{t}{\lambda}\right)^k}.$$

The mean hatching time HT can then be expressed through a relationship involving the Gamma function:

$$HT = \lambda \Gamma \left(1 + \frac{1}{k} \right).$$

Data S2 Percentage of hatching (± standard error) measured during 70 days with a five-day interval [A] for the virulent and avirulent lineages after six generations and [B] for the virulent and avirulent lineages after ten generations. Lines represent the model adjusted to the hatching data for each lineage.



ANNEXE 2 : SUPPORTING INFORMATION – ARTICLE 2

Table S1 Polymorphic microsatellite loci used in this study. The eight identified outliers are shown in bold.

Locus ID	Repeat motif	PCR product size	Forward primer	Reverse primer
Gp106	ATTTT	195	TCTGTTCAGCGCACTTATGG	ATTTGATCGTTCCCTCGTTG
Gp109	ACGG	133	TCTCGCAGAAGGGAAAAGAA	TAAAAGACGGAAGAACGGGA
Gp111	TCGG	111	TCCATTTGTTTTGGGGACAT	CCGTGTCCGATAAATTCCTG
Gp118	TCCG	133	ACCGATGAAGAACATCGTCC	TCGTTCCGTCTTCGTAATCC
Gp135	GA	146	GCGAAATGAACGGTCGTAGT	ATTACATTGCCCAAATCGGA
Gp145	CGTC	253	TCACAAAACATTGGAGCTGAA	TTAAATGAGAAACGGGGGAAT
Gp108	AATC	240	TAACGGCTATCAGCCCAATC	TCGGCCAAAACGTAAAACTC
Gp112	AATG	247	GTTTTAAGCAGACAAGGCCG	ATCTCATAGCAATTTGCCCG
Gp116	CGTC	141	ATTCATTCGCAATGTTTCCC	TGGAAATGTGAGAAAGGGCT
Gp117	GCCC	118	GTCTATTGGCGGCACGTATT	TTCCAAATCCGCCATAATTG
Gp122	CATT	139	AGAGGGTGCCTTTGCTTCTT	ATTGAGTGCCAATAATCCGC
Gp126	GATT	192	GTTATTGTGGCGGATGGAAT	GTACTGTATGATGCCGGGCT
Gp100	ΤΤΤΑΑΑ	241	GAATGCTTTACGCGGCTAAC	GGCGACCGTTGAGGTAATAA
Gp101	CGGTGG	166	ACAATACAGCTTGATGGGGC	GCCATATTTTCAGTCCACCAA
Gp103	тсссс	247	TAATTACGCTCTGCCGGTTC	CAGTCGGTTGAGGCAATTCT
Gp110	TGAT	90	ATGAGATGAGCTTTGATGATGC	GCCATCTGATCACTGGCAC
Gp114	GATC	219	CTCCCCTTCCAACTCATTCA	GATTGGCTGTTCTTCGGAAA
Gp115	GAAT	176	CAGACTGGTTGTGTTGTGCG	GGTGGTGAGGTGGTTATGGA
Gp119	TTAT	143	TACGGCGAATTTGAGGTAGG	TATATTTTCGGGCATCTCGG
Gp121	ATTT	142	GGAAGGTGACGTGAGGACTG	CATTGCCTGTCCTTTCCTGT
Gp124	GGCC	147	CATTCACACATCGGCAAAAC	AATGCAACGAGACAATTCCC
Gp128	CCAC	146	TACCTTTTCCGATCCTTCCC	CCTTCCCATTCCGTCAGTAT
Gp132	CCGA	143	GGGAGCCGTCAGACAAATC	GTTGGAGTTGGGCCTACAAA
Gp133	GAAT	109	ATTCGGAATCTCGAACATGG	TGTTGCATTGCCTTTGCTTA
Gp149	СТ	90	AGCGCCAATTATCTCCACC	TACGGTTAAGCAACGAAGGG
Gp150	GT	305	TTAAACGTGCCCCAGAAAAC	AAAGGAATGGCAGGGAGAAT
Gp152	тс	130	TATATGCATTCAAGCCGCTG	AGCACAGAGGGTAAAGCGAA
Gp153	GA	91	TGGGGACCCTGTCTTAAATG	TTCTCTGGTGTCTTGCGTTG
Gp158	GA	114	ATCCAACGGAACAGCAAATC	CCGTATCACCCAATAATACGTTC
Gp177	тс	232	CCAAAACCGTCTCCAACACT	CATGGATTTTGCACTTCGTG
Gp178	тс	92	CAACTCTCCCCACTCAAAGC	CTTCCCCGCACAACACAAC
Gp179	GTC	311	TCTCGAAAGGCATCGAGAAT	CAGCTTACCGATTGGTCGTT
Gp194	ACA	289	AATGGCTTAAGCAAAGCGAA	CGACGCCGATTAATTTTGTT
Gp202	TTC	176	GGCACAATTGTTCAACACCA	CCCTCAAAAGAGCCTTTCCT
Gp234	СТ	197	AGCAAAAGGTCAAAAGCGAA	CTGATCGACTCTTCTCGGCT
Gp235	CG	248	AAAGCCGGGACAAATCTTTT	AAGGTCTACATCGCAGTGCC
Gp238	тс	142	ACATCGTCTGCGTCTTCTCC	TGCTTTCTTCTCTGTCCCGT

Annexe 2 : Supporting Info	rmation — Article 2
----------------------------	---------------------

Gp239	TG	90	CAATCTGCAGCTTCGTTCAG	TGGTTGGGGTTAGTTTTGTTTT
Gp245	CAA	195	AAAACAATGGCAAAAGGTGG	GGGCAGAGTCCAGTTGAAGT
Gp247	ATTA	106	GCCAGATTTGAGTGAATGAAGAG	CCAGTACCAGTACCAAAAGCAA
Gp248	TTCC	100	AGGAGGGACTGCTCACTTACC	TCAGATCGGGGCTTAAACAT
Gp249	CAA	106	GTCACTGAACTGCGGTGCT	CAGTTGCCCTTTGTCCTCAT
Gp250	CAG	120	TTCGCGTTAAATCTTTGTCAGA	TGTGATCACCAGGGAACTGA
Gp251	CTG	175	ATGTCAGCCAATGGCACTCT	CGTTGCATTTAATTTAATCATACAAA
Gp262	GGA	111	GCCCTGGTTTCCCACTTTAT	TCTCCAGATGTTGCAGATCG
Gp264	TTC	112	TGAATGCCTCCATATGAAATAA	TAATCACTGCGCCTTTGTCC
Gp265	GAG	118	AAAATTGGAGTGACGTTGGTG	GGAAGCTGTTCGTTGCTCTT
Gp274	CAA	102	GTCGCCTCGGATTTTCTACC	AATTGCCTCGGCTAATTCAC
Gp275	CAGAAG	112	GACGGAGGAGTCTTTTTCGAG	GGGAGGGGAGTGACTGTTCT
Gp280	ACA	154	GAGCCCGAATCATAAAATGC	CGGTCGGTGCTTAACACTTT
Gp286	ССТ	208	GTTCCATCTTGCACTCCGAT	CCATTTATCTTCCGGCTTCA
Gp306	TAA	108	TCCCCGTGACTATAAGAAATGAA	GAGGAGAGATGAGGACAATGAT
Gr69	AG	172	AAGAAGACGCCAAACTCGAA	ACAACAGTGCCAGCATTGAG

Table S2 Putative functions of the predicted genes located in a 240-kb window centred on the outlier microsatellite loci on the *Globodera pallida* genome assembly version Gpal v. 1.0 (Cotton *et al.* 2014). Genes likely to be involved in pathogenicity are shown in bold.

Outlier loci	Scaffold	window (kb) around the outlier loci	Gene ID	Location on the scaffold	Gene function	Signal peptide
		30-0	GPLIN_001015000	4336-7937	protein Kinase C family member (pkc	
		0-30	GPLIN_001015100	19264-21398	2) transcribed hypothetical protein	Yes, in an ortholog in <i>G. rostochiensis</i>
		0-30	GPLIN_001015200	21969-24632	transcribed hypothetical protein	
		0-30	GPLIN_001015300	27375-32597	vacuolar protein sorting associated protein 53	
		0-30	GPLIN_001015400	32675-35767	transcribed hypothetical protein	
		0-30	GPLIN_001015500	38175-43883	homocysteine methyltransferase	
		30-60	GPLIN_001015600	46773-49217	transcribed hypothetical protein	
Gp121	437	30-60	GPLIN_001015700	50138-58116	DOMON domain containing protein CG14681	Yes, in an ortholog in <i>G. rostochiensis</i>
-		30-60	GPLIN_001015800	58791-61387	CRAL TRIO domain containing protein	
		30-60	GPLIN_001015900	66465-70444	transcribed hypothetical protein	
		30-60	GPLIN_001016000	70498-73705	transcribed hypothetical protein	
		60-120	GPLIN_001016100	73859-76349	tetraspanin 9	
		60-120	GPLIN_001016200	76653-78072	Heat Shock protein family member (hsp 25)	
		60-120	GPLIN_001016300	79321-79672	transcribed hypothetical protein	Yes
		60-120	GPLIN_001016400	82420-82780	transcribed hypothetical protein	
		60-120	GPLIN_001016500	84706-85491	transcribed hypothetical protein	Yes
		120-60	GPLIN_000507300	3-1014	BTB:POZ domain containing protein 3	
		60-30	GPLIN_000507400	12768-16774	BTB:POZ domain containing protein 3	
		60-30	GPLIN_000507500	18294-19034	transcribed hypothetical protein	
		60-30	GPLIN_000507600	19710-20082	RBP 1 protein (SPRY domain)	
		60-30	GPLIN_000507700	20781-21854	transcribed hypothetical protein	
		60-30	GPLIN_000507800	22760-24230	RBP 1 protein (SPRY domain)	Yes
		60-30	GPLIN_000507900	36400-39044	transcribed hypothetical protein	
		30-0	GPLIN_000508000	40800-42953	transcribed hypothetical protein	Yes
		30-0	GPLIN_000508100	53263-53804	transcribed hypothetical protein	Yes
Gn101	130	30-0	GPLIN_000508200	56561-58483	acid phosphatase 1	Yes
Opioi	150	30-0	GPLIN_000508300	62417-63089	transcribed hypothetical protein	
		0	GPLIN_000508400	68187-71110	Chitin binding protein, peritrophin- A	Yes
		0-30	GPLIN_000508500	71712-74713	transcribed hypothetical protein	
		0-30	GPLIN_000508600	77806-78806	hypothetical protein	
		0-30	GPLIN_000508700	85277-88234	CLN (neuronal ceroid lipofuscinosis)	
		0-30	GPLIN_000508800	93678-96857	Transcription initiation factor TFIID subunit 8	
		0-30	GPLIN_000508900	99747-100054	hypothetical protein	
		60-120	GPLIN_000509000	175128-178645	transcribed hypothetical protein	
		60-120	GPLIN_000509100	187860-189159	transcribed hypothetical protein	
		60-120	GPLIN_000509200	189514-192621	transcribed hypothetical protein	

Annexe 2 : Supporting Information — Article 2

		60-120	GPLIN_000509300	196753-199437	Olfactomedin domain containing protein	
		60-30	GPLIN_000691700	593-1233	transcribed hypothetical protein	
		60-30	GPLIN_000691800	3867-5051	dorsal gland cell specific expression protein	
		60-30	GPLIN_000691900	9497-12258	mimitin, mitochondrial	
		60-30	GPLIN_000692000	13158-18286	Succinyl CoA 3 ketoacid coenzyme A transferase	
		60-30	GPLIN_000692100	18536-19094	histone cluster 2, H3c2	
		60-30	GPLIN_000692200	19401-19965	H2A histone family, member X	
		30-0	GPLIN_000692300	21646-27499	transcribed hypothetical protein	
		30-0	GPLIN_000692400	27752-30375	transcribed hypothetical protein	
		30-0	GPLIN_000692500	31556-57322	Copine domain containing protein 2	
		0-30	GPLIN_000692600	57566-59215	transcribed hypothetical protein	
		0-30	GPLIN_000692700	66169-72222	Solute carrier family 23	
		0-30	GPLIN_000692800	78865-80575	serine:arginine rich splicing factor 4	
		0-30	GPLIN_000692900	83168-85506	T complex protein subunit gamma	
		30-60	GPLIN_000693000	85945-87738	T complex protein 1 subunit gamma (stop codon in translation)	
		30-60	GPLIN_000693100	88182-90397	Heterogeneous nuclear ribonucleoprotein A1	
Gn235	216	30-60	GPLIN_000693200	91404-93983	transcribed hypothetical protein	
00200	210	30-60	GPLIN_000693300	95369-98967	transcribed hypothetical protein	
		30-60	GPLIN_000693400	100056-103180	nicotinic acetylcholine receptor non alpha	Yes
		30-60	GPLIN_000693500	104161-105279	transcribed hypothetical protein	
		30-60	GPLIN_000693600	105754-107423	bifunctional aminoacyl tRNA synthetase	
		30-60	GPLIN_000693700	107723-110938	u3 small nucleolar rna associated protein 14	
		30-60	GPLIN_000693800	111764-113032	Succinate dehydrogenase cytochrome b560 subunit	
		30-60	GPLIN_000693900	113418-114209	'Cold shock' DNA binding domain containing protein	
		60-120	GPLIN_000694000	116014-116682	transcribed hypothetical protein	
		60-120	GPLIN_000694100	117417-119253	leucine rich repeat family protein	
		60-120	GPLIN_000694200	132955-134135	transcribed hypothetical protein	
		60-120	GPLIN_000694300	136918-138507	transcribed hypothetical protein	
		60-120	GPLIN_000694400	139295-139936	small nuclear ribonucleoprotein sm d1	
		60-120	GPLIN_000694500	141034-142087	'Cold shock' DNA binding domain containing protein	
		60-120	GPLIN_000694600	142693-143362	transcribed hypothetical protein	
		60-120	GPLIN_000694700	144097-145945	leucine rich repeat family protein	
		120-60	GPLIN_000332800	15728-20702	transcribed hypothetical protein	
		120-60	GPLIN_000332900	23847-25135	transcribed hypothetical protein	
		120-60	GPLIN_000333000	33517-33765	transcribed hypothetical protein, potential novel effector (Cotton <i>et</i> <i>al.</i> 2014)	Yes
Gr69	72	120-60	GPLIN_000333100	37808-38524	transcribed hypothetical protein, potential novel effector (Cotton <i>et</i> <i>al.</i> 2014)	Yes
		120-60	GPLIN_000333200	41631-43207	transcribed hypothetical protein	Yes, in an ortholog in
		120-60	GPLIN_000333300	51748-55100	endothelin converting enzyme 1	Yes, in an ortholog in <i>G. rostochiensis</i>
		120-60	GPLIN_000333400	58437-59284	transcribed hypothetical protein	Yes, in an ortholog in
		120-60	GPLIN_000333500	60263-65962	transcribed hypothetical protein	G. 10310CIIICII3I3

Annexe 2 : Supporting Information — Article 2

120-60	GPLIN_000333600	66179-68416	inosine triphosphate	
120-60	GPLIN_000333700	70266-73721	transcribed hypothetical protein	
60-30	GPLIN_000333800	73913-75425	GPN loop GTPase 1	
60-30	GPLIN_000333900	75906-78617	CULlin family member (cul 2)	
60-30	GPLIN_000334000	78680-79208	CULlin family member (cul 2)	
60-30	GPLIN_000334100	80281-90568	yeast Swi2:Snf2 Like family member (ssl 1)	
60-30	GPLIN_000334200	90914-97637	transcribed hypothetical protein	Yes
60-30	GPLIN_000334300	99085-103120	rho GTPase activating protein 27	
60-30	GPLIN_000334400	103613-105097	transcribed hypothetical protein	
30-0	GPLIN_000334500	116429-117800	derlin2	
30-0	GPLIN_000334600	118095-122616	x ray repair cross complementing protein 5	
30-0	GPLIN_000334700	122855-124168	transcribed hypothetical protein	Yes, in an ortholog in <i>G. rostochiensis</i>
30-0	GPLIN_000334800	125116-126325	50S ribosomal protein L10	
30-0	GPLIN_000334900	126909-128163	Golgi SNAP receptor complex	
30-0	GPLIN_000335000	128299-130198	short transient receptor potential channel	
30-0	GPLIN_000335100	130255-133634	transient receptor potential ion channel A	
0-30	GPLIN_000335200	134545-142651	Calcium BiNding protein , family member	Yes, in an ortholog in G. rostochiensis
0-30	GPLIN_000335300	144114-145652	cyclin fold protein 1 variant b	
0-30	GPLIN_000335400	146043-147083	transcribed hypothetical protein	
0-30	GPLIN_000335500	156572-157764	transcribed hypothetical protein	
0-30	GPLIN_000335600	158929-159554	V type proton ATPase subunit D	
0-30	GPLIN_000335700	160800-163557	CCR4 NOT transcription complex subunit 1	
0-30	GPLIN_000335800	163874-165157	transcribed hypothetical protein	
30-60	GPLIN_000335900	166323-172199	CCR4 NOT complex component, Not1 family protein	
30-60	GPLIN_000336000	172822-174187	transcribed hypothetical protein	
30-60	GPLIN_000336100	174407-175828	coenzyme Q binding protein COQ10 B	
30-60	GPLIN_000336200	176940-179832	60 kDa heat shock protein, mitochondrial	
30-60	GPLIN_000336300	180199-180632	10 Kda heat shock protein, mitochondrial	
30-60	GPLIN_000336400	182328-183590	Citrate synthase, mitochondrial	
30-60	GPLIN_000336500	185215-187201	Probable citrate synthase, mitochondrial	
30-60	GPLIN_000336600	190031-195579	Phosphatase with EF hands family member (pef 1)	
60-120	GPLIN_000336700	196191-203281	Phosphatase with EF hands family member (pef 1)	
60-120	GPLIN_000336800	206124-206781	transcribed hypothetical protein	
60-120	GPLIN_000336900	209401-211944	tetratricopeptide repeat protein 28	
60-120	GPLIN_000337000	212063-214158	transcribed hypothetical protein	
60-120	GPLIN_000337100	216704-217929	transcribed hypothetical protein	
60-120	GPLIN_000337200	218819-221811	transcribed hypothetical protein	
60-120	GPLIN_000337300	222167-223914	transcribed hypothetical protein	
60-120	GPLIN_000337400	225000-226908	Serpin protein	
60-120	GPLIN_000337500	231760-236344	nuclear hormone receptor family member (nhr)	
60-120	GPLIN_000337600	238575-241019	transcribed hypothetical protein	

Annexe 2 : Supporting Information — Article 2

		60-120	GPLIN_000337700	249433-251731	transcribed hypothetical protein	
		60-30	GPLIN_000815100	66-2728	membrane associated tyrosine	
		60-30	GPLIN_000815200	3861-4509	Ground Like (grd related) family member	
		60-30	GPLIN_000815300	6711-7147	DNA directed RNA polymerase I,II and III	
		60-30	GPLIN_000815400	7430-8923	transcribed hypothetical protein	
		60-30	GPLIN_000815500	12114-13871	transcribed hypothetical protein	
		60-30	GPLIN_000815600	14262-15998	transcribed hypothetical protein	
		60-30	GPLIN_000815700	16165-17945	transcribed hypothetical protein	
		60-30	GPLIN_000815800	19024-19696	transcribed hypothetical protein	
		60-30	GPLIN_000815900	19024-19696	transcribed hypothetical protein	Yes
		60-30	GPLIN_000816000	21121-22593	transcribed hypothetical protein	
		30-0	GPLIN_000816100	22887-26298	transcribed hypothetical protein	
		30-0	GPLIN_000816200	28708-39520	ATP bindinding cassette, subfamily C(CFTR:MRP)	
		30-0	GPLIN_000816300	40724-41292	signal recognition particle 72 KDA protein	
		30-0	GPLIN_000816400	41415-45058	TBC1 domain family 0	
		30-0	GPLIN_000816500	45621-47571	transcribed hypothetical protein	
		30-0	GPLIN_000816600	48113-49031	mitochondrial ATP synthase F chain	
		30-0	GPLIN_000816700	49418-50368	Thymidylate kinase	
o	204	30-0	GPLIN_000816800	50962-54096	PR domain zinc finger protein 16	
Gp124	284	0-30	GPLIN_000816900	58224-60095	histone chaperone asf1	
		0-30	GPLIN_000817000	60910-64493	calcium binding egf domain containing protein	
		0-30	GPLIN_000817100	65808-66410	40S ribosomal protein S16	
		0-30	GPLIN_000817200	66585-67756	DNA directed RNA polymerase III subunit RPC8	
		0-30	GPLIN_000817300	69240-71471	v type proton atpase subunit h	
		0-30	GPLIN_000817400	71759-73577	hemKmethyltransferase family	
		0-30	GPLIN_000817500	73665-80076	BTB:POZ domain containing protein9	
		0-30	GPLIN_000817600	80246-81143	histone deacetylase complex subunit SAP18	
		0-30	GPLIN_000817700	81811-83699	dynein light chain	
		30-60	GPLIN_000817800	86603-88929	UDP glucuronosyltransferase ugt 50	
		30-60	GPLIN_000817900	89069-89757	transcribed hypothetical protein	
		30-60	GPLIN_000818000	91543-92142	transcribed hypothetical protein	Yes
		30-60	GPLIN_000818100	92467-96333	transcribed hypothetical protein	
		30-60	GPLIN_000818200	96933-99082	fatty acid elongation protein 3	
		30-60	GPLIN_000818300	101205-103646	ROLIER: helically twisted, animals roll when	
		30-60	GPLIN_000818400	105795-107397	transcribed hypothetical protein	
		30-60	GPLIN_000818500	107680-111411	transcribed hypothetical protein	
		60-120	GPLIN_000818600	119262-120071	transcribed hypothetical protein	Yes, in an ortholog in <i>G. rostochiensis</i>
		60-30	GPLIN_000075500	18181-34682	neurogenic locus notch protein	Yes
		60-30	GPLIN_000075600	38275-38565	transcribed hypothetical protein	
Gp153	12	60-30	GPLIN_000075700	40603-41614	transcribed hypothetical protein, potential novel effector (Cotton <i>et</i> <i>al.</i> 2014)	Yes
		60-30	GPLIN_000075800	41701-42358	60S ribosomal protein L22	
		30-0	GPLIN_000075900	43048-50296	histone lysine N methyltransferase EZH1	

		30-0	GPLIN_000076000	50854-52783	histone lysine N methyltransferase	
		30-0	GPLIN_000076100	53816-55370	39S ribosomal protein L28, mitochondrial	
		30-0	GPLIN_000076200	56376-57173	transcribed hypothetical protein	
		30-0	GPLIN_000076300	57723-58602	H ATPase V type subunit	
		30-0	GPLIN_000076400	60040-67075	transcribed hypothetical protein	
		30-0	GPLIN_000076500	71290-75008	transcribed hypothetical protein	
		0-30	GPLIN_000076600	98114-99546	transcribed hypothetical protein	
		30-60	GPLIN_000076700	106448-109756	vasa	
		30-60	GPLIN_000076800	110273-113549	src substrate cortactin	
		30-60	GPLIN_000076900	113983-114845	yrdC domain containing protein	
		30-60	GPLIN_000077000	120471-120716	transcribed hypothetical protein	
		30-60	GPLIN_000077100	121338-121729	transcribed hypothetical protein	
		30-60	GPLIN_000077200	122348-124160	karyopherin alpha subunit	
30-60		30-60	GPLIN_000077300	124521-126316	transcribed hypothetical protein	
		30-60	GPLIN_000077400	126451-127608	dorsal gland cell specific expression	
		30-60	GPLIN_000077500	129213-137135	RNA dependent RNA polymerase Family family	
		60-120	GPLIN_000077600	137480-139042	retinal rod rhodopsin sensitive cGMP	
		60-120	GPLIN_000077700	140268-141907	transcribed hypothetical protein	
		60-120	GPLIN_000077800	143516-144461	transcribed hypothetical protein	
		60-120	GPLIN_000077900	147638-160920	Trithorax group protein osa	
		60-120	GPLIN_000078000	161613-163877	transcribed hypothetical protein	
		60-120	GPLIN_000078100	165674-173174	DEgenerin Like family member (del 3)	
		60-120	GPLIN_000078200	175201-176045	60S ribosomal protein L17	
		60-120	GPLIN_000078300	176573-177123	serine:threonine protein phosphatase PP1	
		60-120	GPLIN_000078400	181159-182292	Phosducin protein	
		60-120	GPLIN_000078500	182497-187870	Chromosome transmission fidelity protein 18	
		60-120	GPLIN_000078600	189511-195365	transcribed hypothetical protein	
		60-120	GPLIN_000078700	195900-197539	Iron sulfur cluster co chaperone protein HscB	Yes, in an ortholog in <i>G. rostochiensis</i>
Gp135	4348		n	o predicted ge	ene on this scaffold	
Gp249			not found	l on the new g	enome version	

ANNEXE 3 : SUPPORTING INFORMATION – ARTICLE 3

Figure S1 Selection of the four experimental *Globodera pallida* **lineages.** Nematode lineages were established from two French natural *G. pallida* populations, SM (near Saint-Malo, Brittany, north-western France) and N (from the island of Noirmoutier, western France), reared during eight successive cycles (*i.e.* eight generations) on the susceptible potato cultivar Désirée (D) and on the resistant cultivar Iledher (I). Each sample name indicates its geographical origin (SM for Saint-Malo and N for Noirmoutier) and the potato cultivar on which it evolved (I for Iledher and D for Désirée).



Figure S2 Principal Component Analysis (PCA) of the sequencing data from eight *G. pallida* pools sequenced on four different lanes (technical replicates), based on read counts at **2,383,040 SNPs.** Each sample name indicates its geographical origin (SM for Saint-Malo and N for Noirmoutier), the potato cultivar on which it evolved (I for Iledher and D for Désirée), the number of the biological replicate (1 or 2) and the number of the lane (L) on which it has been sequenced. The first axis explains 32.61% of the variance and clusters the samples according to the population geographical origin. Even if the second and third axes explain a small part of the variance, *i.e.* 4.41% for PC2 and 4.22% for PCA3, they tend to cluster N and SM samples, respectively, according to their virulence status. Note however that for each of the eight pools, replicate sequenced on the first sequencing lane systematically depart from the three others. This patterns most probably results from a lower sequencing coverage for lane 1 (that was sequenced in 2x100 bp paired end) compared to the three others (that were sequenced in 2x125 bp paired-end).





Table S1 Putative functions of the predicted genes located in a 120-kb window centered on the 31 outlier loci on the *Globodera pallida* genome assembly version Gpal.v1.0 (Cotton *et al.* 2014). Relevant results of the BLAST search of the predicted genes against the annotated *G. rostochiensis* genome version nGr.v1.0 (Eves-van den Akker *et al.* 2016) are also shown. *: outlier SNP located in the gene, **: outlier SNP located in the exon.

affold	Outlier	Globodera pallida genome						BLAST against Globodera rostochiensis genome	
SC	locus	Gene ID	Location on the scaffold	Putative gene function	Peptide signal	Phobius prediction	Gene ID	Putative gene function	
		GPLIN_00 0030900	367074- 371195	actin protein 6	no		GROS_g1 2208	spliceosome	
		GPLIN_00 0031000	371242- 372608	nuclear Pore complex protein family member	no				
		GPLIN_00 0031100 GPLIN_00	398790- 400048 403834-	transcription initiation factor TFIID subunit heta 1 4	no				
		0031200	404255	endoglucanase	no			magnesium ion	
		GPLIN_00 0031300 *	428992- 438734	transcribed hypothetical protein	no		GROS_g0 5341	transport; sodium ion transmembrane transport; integral to plasma membrane; cation channel activity	
4	SNP_ 4	GPLIN_00 0031400	439019- 440137	transcribed hypothetical protein	no				
·	429075	GPLIN_00 0031500	441856- 451123	cysteine synthase	no				
		GPLIN_00 0031600	457740- 458980	long chain fatty acid transport protein 4	no		GROS_g1 3363	fatty acid transport proteins	
		GPLIN_00 0031700	460388- 460633	transcribed hypothetical protein	no				
		GPLIN_00 0031800	462978- 464011	galactosyltransferase 2	no		GROS_g0 0450	galactosyltransferase	
		GPLIN_00 0031900	470256- 470486	transcribed hypothetical protein	no				
		GPLIN_00 0032000	471353- 473130	transcribed hypothetical protein	no			a sector late dite a	
		GPLIN_00 0032100	474046- 474331	transcribed hypothetical protein	no		GROS_g1 0714	protein binding; stress-activated map kinase interacting protein 1 (SIN1)	
		GPLIN_00 0149900	243326- 245017	solute carrier family 35 member B1	yes	trans- membrane	GROS_g0 2151	transport	
		GPLIN_00 0150000	245308- 249627	mitogen activated protein kinase kinase kinase	no		GROS_g0 2150	MAPK signaling pathway	
		GPLIN_00 0150100	251276- 252209	superoxide dismutase (Cu Zn) transcribed	yes	secreted	GROS_g0 2149	peroxisome	
	SNP_	0150200	253488	hypothetical protein	no		2148	transmembrane	
25	25_ 304232	GPLIN_00 0150300	254217- 258092	transcribed hypothetical protein	no		GROS_g0 2146	negative regulation of DNA damage checkpoint cytoskeleton	
		GPLIN_00 0150400	258929- 261360	T complex protein 1 subunit epsilon	no		GROS_g0 2145	organization; posttranslational protein folding	
		GPLIN_00 0150500	261650- 263494	transcribed hypothetical protein	no		GROS_g1 4282	metabolism of xenobiotics by cytochrome P450	

		GPLIN_00 0150600	263679- 265256	protein vertebrate galectins	yes	secreted	GROS_g0 2143	galactose binding; signal transducer activity
		GPLIN_00 0150700	265891- 266885	N alpha acetyltransferase 20	no		GROS_g0 2141	endocytosis
		GPLIN_00 0150800	266955- 269813	transcribed hypothetical protein	no		GROS_g0 2140	protein homodimerization activity
		GPLIN_00 0150900	269952- 271262	transcribed hypothetical protein	no		GROS_g0 2139	transmembrane
		GPLIN 00	271403-	transcribed			2100	
		0151000	272406	hypothetical protein	no			
		GPLIN 00	273509-	fasciculation and			GROS 20	
		0151100	275230	elongation protein zeta 2	no		2137	axon guidance
		GPLIN 00	275973-	protein kinase domain				
		0151200	277291	containing protein, ck worm protein kinase	no			
		GPLIN_00	279494-	calmodulin	no		GROS_g0	calmodulin
		0151300	280946		110		2136	cannoudann
		GPLIN_00	284171-	transcribed	no			
			284443	nypotnetical protein				
		0151500	302277	kinase inhibitor	no			
		GPLIN 00						nuclear export signal
		0151600	303454-	exportin 7	no		GROS_g0	receptor activity;
		*	310418	·			2131	protein export from
		GPLIN 00	310648-				GROS 20	nucleus, nuclear pore
		0151700	311820	glutaredoxin 3	no		2130	glutaredoxin-3
		GPLIN_00	313085-	transcribed				
		0151800	314369	hypothetical protein	no			
		GPLIN_00	315934-	transcribed	no		GROS_g0	transmembrane
		0151900	316953	hypothetical protein	110		2129	transmembrane
		GPLIN_00	317223-	DEAD Box protein	no		GROS_g0	ATP catabolic process
		0152000	324014	· · · ·			2128	
		GPLIN_00	328566-	transcribed	no			
		GPUN 00	329000	transcribed			GROS a0	
		0152200	330907	hypothetical protein	no		2126	chromatin binding
		010100		immunoglobulin i set				
		GPLIN_00	335501-	domain containing	yes	secreted	GROS_g0	transmembrane
		0152300	330390	protein			2125	
		GPLIN_00	336805-	transcribed	no		GROS_g0	Golgi apparatus
		0152400	340764	hypothetical protein			2124	
		GPLIN_00	341193-	transcribed	no			
		0152500	341/16	nypotnetical protein				
		GPLIN_00	163392-	transcribed	no			
		GPLIN 00	168518-	transcribed			GROS 00	
		0232400	169562	hypothetical protein	yes	secreted	4085	signal peptide
		GPLIN 00	169991-	transcribed				
		0232500	170378	hypothetical protein	no			
		GPLIN_00	171849-	transcribed		concord	GROS_g0	cignal pontido
	SND	0232600	174837	hypothetical protein	yes	secreted	4087	signal peptide
	20 20	GPLIN_00	180474-	phosphatidylcholine:c	no		GROS_g0	sphingolipid
	221212	0232700	181761	eramide	110		4089	metabolism
44		GPLIN_00	183265-	60S ribosomal protein	20		GROS_g0	cell wall modification;
44	CNID	0232800	184828	L4	10		4090	signature
	5NP_ 11	GPLIN_00	184937-	palmitoyltransferase	no		GROS_g0	proteinS-
	 221214	0232900	188002	ZDHHC2	10		4091	acyltransferase
		GPLIN_00	196131-	transcribed	no		GROS_g1	Transmembrane
		0233000	198830	hypothetical protein	-		1539	
		GPLIN_00	198931-	transcribed	no			
		0232100	200109	transcribed				
		0233200	200212-	hypothetical protein	no			
		GPLIN 00	201499-	transcribed				
		0233300	202356	hypothetical protein	no			

		GPLIN_00	202494-	transcribed	20			
		0233400	203706	hypothetical protein	no			
		GPLIN 00	203859-	worm specific				
		0233500	205030	Argonaute NRDE 3	no			
				transcription factor				
		GPLIN 00	205120-	TEIID, C-terminal DNA				
		0233600	205120		no			
		0233000	203933	grycosylase, N-				
			206066	transarihad				
		GPLIN_00	206066-	transcribed	no			
		0233700	211386	hypothetical protein				
		GPLIN_00	211708-	transcribed	no			
		0233800	212533	hypothetical protein				
		GPLIN_00	212735-	transcribed	no			
		0233900	214448	hypothetical protein	no			
		GPLIN_00	221431-	transcribed	20			
		0234000	221718	hypothetical protein	110			
		GPLIN_00	221849-	polynucleotide kinase				
		0234100	222821	3' phosphatase	no			
								CTLH/CRA C-terminal
								to LisH motif domain:
		GPLIN 00	255627-	macronhage			GROS Ø0	actin cytoskeleton:
		0234200	261288	ervthrohlast attacher	yes	secreted	3239	negative regulation of
		0234200	201200	ci ytin oblast attacher			5255	myoloid coll apontotic
				unite als au duial				process
		GPLIN 00	264876-	mitochondriai			GROS g0	insulinase (peptidase
		0234300	267441	processing peptidase	no		3235	family M16)
				beta subunit				
								regulation of
		GPLIN_00	268640-	U1 small nuclear	no		GROS_g0	alternative nuclear
		0234400	269749	ribonucleoprotein A	110		3234	mRNA splicing, via
								spliceosome
			220222	guanine nucleotide			CROS a0	small ribosomal
		GPLIN_00	270323-	binding protein	no			subunit; ion channel
		0234500	271985	subunit			3233	inhibitor activity
								mitochondrion:
								negative regulation of
							GROS_g0	synantic transmission
		GPLIN_00	272436-	transcribed			3232 /	synaptic transmission,
		0234600	274181	hypothetical protein	no		GROS_g0	giulamalergic / ATP
							3238	catabolic process;
								adenosinetriphosphat
								ase
		GPLIN_00	274837-	receptor expression	no		GROS_g0	protein hinding
		0234700	276164	enhancing protein 5	no		3231	protein binding
			276256	G protein coupled				
		GPLIN_00	270250-	receptor associated	no			
		0234800	277074	sorting				
							GROS g0	
		GPLIN 00	277860-	4 aminobutyrate			3230 /	beta-Alanine
		0234900	282940	aminotransferase,	no		GROS g0	metabolism / beta-
		0201000	202010	mitochondrial			3230	Alanine metabolism
				transcribed			GROS g1	
		0211/00	1-3999	hypothetical protein	no		2794	signal peptide
				transcribed			2764 GROS a1	PPD 1 protain [C
		GPLIN_00	19418-20911		no		GRUS_gI	RBP-4 protein [G.
		0311500		hypothetical protein			4260	pallida]
								RBP-4 protein [G.
		GPLIN 00		transcribed			GROS g1	pallida]; secreted
	SNP_	0311600	30359-32404	hypothetical protein	no		/196	SPRY domain-
	66_	0311000		hypothetical protein			4150	containing protein 18
	55507							[G. rostochiensis]
<i>cc</i>		GPLIN_00	20477 27252	transcribed				
00		0311700	304/7-37252	hypothetical protein	no			
	SNP	GPLIN 00		transcribed			GROS g0	
	66	0311800	45302-45915	hypothetical protein	no		4023	riboflavin metabolism
	131126	GPLIN 00		BTB:POZ domain			GROS Ø1	RBP-1 protein, partial
		0311900	46907-48176	containing protein 3	yes	secreted	4128	[G. pallida]
				source protein o			1120	RBP-4 protein [G
		GPLIN_00	49696-50449	transcribed	no		GROS_g1	nallida].
		0312000	+5050-50448	hypothetical protein	10		4136	transmombrana
		CDUN 00		protoin containing			CPO5 ~1	
		GPLIN_00	53144-55475	protein containing	no		GRUS_g1	KBP-1 protein [G.
		0212100	5511155175	CDDV domain			1770	nallidal

		GPLIN_00	57585-58590	transcribed	no		GROS_g1	peptidase family M41
		0312200		hypothetical protein			1105	truncated secreted
		GPLIN_00 0312300	66233-69356	protein containing SPRY domain	no		GROS_g1 4130	SPRY domain- containing protein 15, partial [<i>G.</i> <i>rostochiensis</i>]; signal peptide
		GPLIN_00	71724-72660	transcribed	no			
		GPUN 00		nypotnetical protein			GROS a1	RBP-1 protein [G
		0312500	92978-94740	SPRY domain	no		4278	pallida] truncated secreted SPBY domain-
		GPLIN_00 0312600	95754-96913	protein containing SPRY domain	no		GROS_g1 4287	containing protein 15, partial [G. rostochiensis]
		GPLIN_00	107749-	EF-hand 2 domain	no		GROS_g1	EF-hand calcium-
		0312700	107949	containing protein			4281	binding domain.
		GPLIN_00	111831-	transcribed	no		GRUS_gU	protein SAX-7;
			112800	group 1 glycosyl			5729 GROS d0	ducosul transforaços
		0212000	112920-	group I grycosyr	no			group 1
		GPLIN 00	119091-	transcribed			1/2/	Broup T
		0313000	120888	hypothetical protein	no			
		GPLIN 00	120000	transcribed				
		0313100	124830	hypothetical protein	no			
		GPLIN 00	12.000	in political protein				
		0313200	120062	Sensory AXon			6006 -0	protein SAX-7; sensory
		*	129063-	guidance family	no		GROS_g0	AXon guidance;
		(SNP_66_ 131126)	137156	member (sax 7)			5729	transmembrane
		GPLIN_00	139135-	transcribed	no			
		0313300	140624	hypothetical protein	110			
		GPLIN_00	142713-	transcribed	no			
		0313400	144269	hypothetical protein				
		GPLIN_00	145772-	transcribed	no		GROS_g1	signal peptide
		0313500	146732	hypothetical protein beta 1,4 endoglucanase (cell			3276	
		GPLIN 00	154523-	wall modifying			GROS g0	beta-1,4-
		0313600	157608	protein); putative GH5 cellulase (cellulose degradation)	yes	secreted	7446	endoglucanase, partial; cellulase
								poly(A) polymerase
		GPLIN_00 0313700	158653- 160495	transcribed hypothetical protein	no		GROS_g0 2105	central domain; nuclear mRNA splicing, via spliceosome
		GPLIN 00	167056-	transcribed			GROS g0	
		0313800	168124	hypothetical protein	no		7447	EB module
		GPLIN_00	168962-	transcribed	20		GROS_g0	transmomhrano
		0313900	169322	hypothetical protein transcribed	no		1126	transmentprane
		GPLIN 00	183685-	hypothetical protein,			GROS g1	signal peptide; 1106
		0314000	184563	similar to <i>G.</i> <i>rostochiensis</i> effector 1106	yes	secreted	4309	effector family [G. rostochiensis]
								signal peptide;
	SNP_ 85_ 126066	GPLIN_00 0382000	67804-68769	glutaminyl peptide cyclotransferase	no		GROS_g0 7374	peptidase M28 domain containing protein [<i>Haemonchus</i>
0-		60					05.53	contortus]
85		GPLIN_00	84608-86814	transcribed	no		GROS_g1	квР-4 protein [G.
	SNP_		04027	hypothetical protein			4124 GPOS ~1	pullaaj zinc finger protein
	85_	0382200	<u>74032-</u> 100930	zinc finger protein	no		1336	
	126206	GPLIN 00	110433-	transcribed			1330	[]
		0382300	110747	hypothetical protein	no			

		GPLIN_00 0382400 GPLIN_00 0382500 GPLIN_00 0382600 GPLIN_00	120748- 122590 126792- 129853 142203- 143446 143942-	BTB:POZ domain containing protein 3 SPla RYanodine receptor SPRY transcribed hypothetical protein transcribed	no no no no		GROS_g1 4193 GROS_g1 4163	truncated secreted SPRY domain- containing protein 15, partial [<i>G.</i> <i>rostochiensis</i>] RBP-4 protein [<i>G.</i> <i>pallida</i>]
		GPLIN_00 0382800	144891 151467- 154962	E9 protein	no		GROS_g1 4123	signal peptide
		GPLIN_00 0382900	164278- 168506	glutathione synthetase	no		GROS_g0 5390 / GROS_g1 3797	glutathione synthetase-like [<i>Maylandia zebra</i>] / glutathionesynthase; signal peptide guanine nucleotide-
		GPLIN_00 0383000	169444- 172268	binding protein G(o) subunit	no		GROS_g0 5389	binding protein alpha- 3 subunit [<i>Brugia malayi</i>]
		GPLIN_00 0383100	174565- 177704	ODR 3	no		GROS_g0 5389	guanine nucleotide- binding protein alpha- 3 subunit [<i>B. malayi</i>]
		GPLIN_00 0383200	179602- 190327	guanine nucleotide exchange factor for Ras	no		GROS_g0 5388	regulation of Ral GTPase activity
		GPLIN_00 0450000	79896-81897	UDP glucuronosyltransferas e	yes	trans- membrane	GROS_g0 9337	integral to plasma membrane
		GPLIN_00 0450100	94652-95140	hypothetical protein	no			
		GPLIN_00 0450200	98715-99013	transcribed hypothetical protein	no			
		GPLIN_00 0450300	120569- 121060	hypothetical protein	no			
		GPLIN_00 0450400	126241- 129974	protein containing SPRY domain	no		GROS_g1 4195	dendrite morphogenesis, truncated secreted SPRY domain- containing protein 15, partial [G. rostochiensis]
		GPLIN_00 0450500	132676- 135723	inversin protein alternative, ankyrin repeat protein	no		GROS_g1 4192	protein binding
111	SNP_111_ 140151	GPLIN_00 0450600	133821- 134846	BTB:POZ domain containing protein 3	no		GROS_g1 4195	dendrite morphogenesis, truncated secreted SPRY domain- containing protein 15, partial [G. rostochiensis]
		GPLIN_00 0450700	138326- 138779	transcribed hypothetical protein	no			
		GPLIN_00 0450800	138941- 139562	Nematode AStacin protease family member	no		GROS_g0 7422	molting cycle, collagen and cuticulin- based cuticle; Astacin
		GPLIN_00 0450900	144927- 150255	NAD kinase domain containing protein 1	no		GROS_g1 2807 / GROS_g1 2808	NAD metabolic process / mRNA surveillance pathway
		GPLIN_00 0451000	153798- 155515	transcribed hypothetical protein	no			
		GPLIN_00 0451100	156162- 159524	Amino Acid Transporter family member (aat 5)	no		GROS_g1 2805 / GROS_g0 0275	transmembrane transport / protein digestion and absorption
		GPLIN_00 0451200	160262- 161491	cyclin B	no		GROS_g0 4262	microtubule cytoskeleton

		GPLIN_00 0451300	162184- 162860	transcribed hypothetical protein	no		GROS_g0 4261	bifunctional 3 - phosphoadenosine 5 - phosphosulfate synthase [<i>Ascaris</i>
		GPLIN_00	163767-	serine:threonine			GROS_g0	<i>suum</i>] transcription, DNA-
		0451400	166872	protein phosphatase 5	no		4260	dependent
		GPLIN 00	168900-	transcribed			GROS g0	
		0451500	173567	hypothetical protein	no		4258	protein binding
		GPLIN 00	174095-	40S ribosomal protein			GROS ØD	RNA metabolic
		0451600	175144	S26	no		4257	process
			175/16	transcribed				process
		0451700	177660	hypothetical protein	no		4256	transmembrane
			177040	DNA roplication			4250 GROS 40	
		04F1800	177949-	licensing factor MCM6	no			cell cycle - yeast
		0451800 CRUN 00	101975				4255	
		04F1000	102477-	subunit SMARCC2	no		4254	ATP catabolic process
		0451900	18/202	SUDUIIIL SIMARCC2			4254	
		GPLIN_00	187774-	DINA replication			GROS_g0	and a start start
		0452000	189157	complex GINS protein	no		4253	nucleoplasm
			102011	SLD5			6006 - 0	
		GPLIN_00	192011-	transcribed	no		GROS_g0	transmembrane
		0452100	196045	hypothetical protein			4252	
								galactosyl beta-1,3 N-
		GPLIN 00	197876-	transcribed			GROS g1	acetylgalactosamine
		0452200	201179	hypothetical protein	no		0891	beta-1,3-
								glucuronosyltransfera
								se activity
								nematode larval
		GPLIN 00		KH domain containing			GROS g0	development;
		0482300	35198-37842	protein	no		6084	masculinization of
								hermaphroditic germ-
				the second second				line
		GPLIN_00	59414-60834	transcribed	no			
		0482400		nypotnetical protein				
		GPLIN_00	C1205 C2050	nematode cuticie			GROS_g1	larval development ;
		0482500	01282-02959	domain	no		1113	transmembrane
				uomam				
		GPLIN_00	72586-74133	transcribed	yes	trans-		
				transerihed		memprane	CDO5 ~1	
		GPLIN_00	79061-79808	hypothetical protein	no			signal peptide
		0482700		mypothetical protein			3221	
		GPLIN_00	0E11E 07207	humphoid tissue	20			
		0482800	85115-87297	lymphoma	110			
				mucosa associated				
		GPLIN_00	97612-90269	lymphoid tissue	20			
		0482900	87013-85208	lymphoma	110			
				murosa associated				ubiquitin-protein
	SNP_	GPLIN_00	89592-91491	lymphoid tissue	no		GROS_g1	ligase activity:
123	123_	0483000	05552 51151	lymphoma	no		1117	nentidase activity
	97233			lymphoma				heta-
		GPLIN_00	98509-	sucrose hydrolase	no		GROS_g1	fructofuranosidase:
		0483100	100826	such obe injui clube			1374	signal peptide
		GPLIN 00	101689-	tail length tane				8
		0483200	102752	measure protein	no			
				beta				beta-
		GPLIN 00	103018-	fructofuranosidase:			GROS g1	fructofuranosidase:
		0483300	107209	secreted GH32	yes	secreted	1397	plant-type cell wall
				fructosidase				modification
		GPLIN 00	108777-	transcribed				
		0483400	111065	hypothetical protein	yes	secreted		
		GPLIN 00	112509-	transcribed				
		0483500	112790	hypothetical protein	no			
		GPLIN 00	117038-	transcribed				
		0483600	117476	hypothetical protein	yes	secreted		
				'				neuronal cell body;
								perinuclear region of
		GPLIN_00	119298-	mps one binder kinase			GROS_g0	cytoplasm;
					110			
		0483700	120112	activator 3 like			9711	microtubule
		0483700	120112	activator 3 like			9711	microtubule cytoskeleton

Annexe 3 : Supporting Information - Article 3

		GPLIN_00 0483800 GPLIN_00 0483900	120304- 121636 121782- 123696	ubiquitin interacting motif ATP synthase subunit beta, mitochondrial	no no		GROS_g0 9712 GROS_g0 9713	perinuclear region of cytoplasm oxidative phosphorylation
		GPLIN_00 0484000	123896- 127752	protein arginine N methyltransferase 5	no		GROS_g0 9714	protein ubiquitination; negative regulation of mitotic cell cycle
		GPLIN_00 0484100	128956- 130197	tryptophanyl tRNA synthetase	no			
		GPLIN_00 0484200	137371- 139503	transcribed hypothetical protein	no		GROS_g0 9707	transmembrane
		GPLIN_00 0484300	141658- 142945	receptor type tyrosine protein phosphatase eta	no		GROS_g0 5032	protein tyrosine phosphatase activity; salicylic acid mediated signaling pathway glycerophospholipid
		GPLIN_00 0484400	148275- 154307	transcribed hypothetical protein	no		GROS_g0 5028	metabolism; phosphatidylcholine metabolic process; signal peptide
		GPLIN_00 0484500	154735- 155098	transcribed hypothetical protein	no			0 1 1
		GPLIN_00 0484600	155919- 159554	mitotic checkpoint serine:threonine protein	no		GROS_g0 4287 / GROS_g0 5027	regulation of chromosome segregation; positive regulation of intrinsic apoptotic signaling pathway / mitotic cell cycle spindle assembly checkpoint
		GPLIN_00 0626500	1034-2234	protein containing SPRY domain	no		GROS_g1 4241	regulation of Ran GTPase activity; RBP-1 protein [G. <i>pallida</i>]
		GPLIN_00 0626600	4596-5610	transcribed hypothetical protein	no			, ,
		GPLIN_00 0626700	9510-11116	protein containing SPRY domain	no		GROS_g1 4184	regulation of Ran GTPase activity; RBP-4 protein [<i>G. pallida</i>]
	SNP_ 182	GPLIN_00 0626800	13854-21499	protein containing SPRY domain	no		GROS_g1 4196	cytoplasmic microtubule; RBP-4 protein [<i>G. pallida</i>] truncated secreted
182	414	GPLIN_00 0626900	23491-25804	protein containing SPRY domain	no		GROS_g1 4195	SPRY domain- containing protein 15, partial [<i>G</i> .
	182_ 557	GPLIN_00 0627000	27744-30562	transcribed hypothetical protein	no			Tostochiensisj
		GPLIN_00 0627100	41665-44164	paralog of RBP-1 protein	yes	secreted	GROS_g1 4256	ran GTPase binding; RBP-1 protein, partial [<i>G. pallida</i>]
		GPLIN_00 0627200	45726-46823	60S acidic ribosomal protein P1	no		GROS_g0 1370	ribosome
		GPLIN_00 0627300	49760-52726	transcribed hypothetical protein	no			
		GPLIN_00 0627400	55950-61422	serpin protein	yes	secreted	GROS_g0 9868	extracellular space
		GPLIN_00 0691700	593-1233	transcribed hypothetical protein	no			dorsal gland cell
216	SNP_216_ 53947	GPLIN_00 0691800	3867-5051	dorsal gland cell specific expression protein	no		GROS_g1 4298	specific expression protein [Heterodera avenae]
		GPLIN_00 0691900	9497-12258	mimitin, mitochondrial	no		GROS_g0 5274	NADH ubiquinone oxidoreductase subunit NDUFA12

		GPLIN_00 0692000	13158-18286	succinyl CoA:3 ketoacid coenzyme A transferase	no		GROS_g0 5277 / GROS_g0 2487	butanoate metabolism / butanoate metabolism
		GPLIN_00 0692100	18536-19094	histone cluster 2, H3c2	no		GROS_g0 5278	protein complex
		GPLIN_00 0692200	19401-19965	H2A histone family, member X	no		GROS_g1 0756	positive regulation of DNA repair
		GPLIN_00 0692300	21646-27499	transcribed hypothetical protein	no		GROS_g0 2034	small GTPase mediated signal transduction
		GPLIN_00 0692400	27752-30375	transcribed hypothetical protein	no			
		GPLIN_00 0692500 *	31556-57322	copine domain containing protein 2	no		GROS_g0 2032	plasma membrane
		GPLIN_00 0692600	57566-59215	transcribed hypothetical protein	no		GROS_g0 2032	plasma membrane
		GPLIN_00 0692700	66169-72222	solute carrier family 23	no		GROS_g0 2031 / GROS_g0	nucleobase transport / plasma membrane
		GPLIN_00 0692800	78865-80575	serine:arginine rich splicing factor 4	no		GROS_g0 2030	mitosis
		GPLIN_00 0692900	83168-85506	T complex protein 1 subunit gamma	no		GROS_g0 2028	extracellular vesicular exosome
		GPLIN_00 0693000	85945-87738	T complex protein 1 subunit gamma beterogeneous	no		GROS_g0 2028	extracellular vesicular exosome
		GPLIN_00 0693100	88182-90397	nuclear ribonucleoprotein A1	no		GROS_g0 2026	spliceosome
		GPLIN_00 0693200	91404-93983	transcribed hypothetical protein	no		GROS_g0 2025	phosphatidylinositol- 4,5-bisphosphate4- phosphatase
		GPLIN_00 0693300	95369-98967	transcribed hypothetical protein	no		GROS_g0 2024	plasmin
		GPLIN_00 0693400	100056- 103180	nicotinic acetylcholine receptor non alpha	yes	trans- membrane	GROS_g0 2022	response to inorganic substance
		GPLIN_00 0693500	104161- 105279	transcribed hypothetical protein	no		GROS_g0 2021	hypothetical protein Y032_0018g3584 [Ancylostoma ceylanicum]
		GPLIN_00 0693600	105754- 107423	bifunctional aminoacyl tRNA synthetase	no		GROS_g0 2020	porpnyrin and chlorophyll metabolism
		GPLIN_00 0693700	107723- 110938	u3 small nucleolar rna associated protein 14	no		GROS_g0 2019	ribosome biogenesis in eukaryotes
		GPLIN_00 0693800	111764- 113032	dehydrogenase cytochrome b560 subunit	no		GROS_g0 2018	defense response to Gram-negative bacterium
		GPLIN_00 0693900	113418- 114209	'cold shock' DNA binding domain containing protein	no		GROS_g0 2017	sequence-specific DNA binding
		GPLIN_00 0814000 *	56408-57804	BTB:POZ domain containing protein At1g55760	no			
	SNP_	GPLIN_00 0814100	58370-60428	BTB:POZ domain containing protein At1g55760	no			
283	56739	GPLIN_00 0814200	61441-67930	transcribed hypothetical protein BTB:POZ domain	yes	secreted	GROS_g1 3559	signal peptide
203	SNP_ 283	GPLIN_00 0814300	68894-70678	containing protein At1g55760	no			
	56740	GPLIN_00 0814400	72618-73953	transcribed hypothetical protein BTB:POZ domain	no			
		GPLIN_00 0814500	75558-76793	containing protein At1g55760	no			

Annexe 3 : Supporting Information — Article 3

		GPLIN_00 0814600	78642-79828	transcribed hypothetical protein	yes	trans- membrane	GROS_g0 5788	signal peptide
		GPLIN_00 0814700	84623-91635	zinc metallopeptidase	no		GROS_g0 5789	metalloendopeptidase activity; peptidase family M13
		GPLIN_00 0814800	95043-97768	myelin transcription factor 1 protein	no		GROS_g1 4295	zinc finger, C2HC type
		GPLIN_00 0814900 GPLIN_00	107592- 114130 115751-	transcribed hypothetical protein G patch domain	no		GROS_g1 4213	BTB/POZ domain
		0815000	116921	containing protein	no			
		GPLIN_00 0832100 **	6442-11746	Na:H eXchanger family member (nhx 9)	no		GROS_g0 4641 / GROS_g1 3418	protein F32B5.6, isoform j [<i>Caenorhabditis</i> <i>elegans</i>] / sodium:hydrogen antiporter activity
		GPLIN_00 0832200	12647-13096	transcribed hypothetical protein	no			
		GPLIN_00 0832300	17388-19554	transcribed hypothetical protein	no			
		GPLIN_00 0832400	33723-34555	transcribed hypothetical protein	no			
		GPLIN_00 0832500	35455-36989	transcribed hypothetical protein; Gpa_Dog_0189	yes	secreted		
296	SNP_296_ 8846	GPLIN_00 0832600	44182-46613	transcribed hypothetical protein	yes	secreted	GROS_g0 8307	lysosomal membrane
	0010	GPLIN_00 0832700	47366-49673	membrane glycoprotein	yes	trans- membrane	GROS_g0 8307	lysosomal membrane
		GPLIN_00 0832800	50657-51448	transcribed hypothetical protein	no		GROS_g0 8308	aminoacyl- tRNAhydrolase.
		GPLIN_00 0832900	52058-53243	14-3-3 protein	no		GROS_g0 5963 / GROS_g0 8309	neurotrophin signaling pathway / neurotrophin signaling pathway
		GPLIN_00 0833000	54914-56106	14-3-3 protein	no		GROS_g0 5963 / GROS_g0 8309	signaling pathway / neurotrophin signaling pathway hypothetical protein
		GPLIN_00 0833100	63249-68201	calcium binding EGF domain containing protein	no		GROS_g0 8310	Y032_0091g2492 [Ancylostoma ceylanicum]
		GPLIN_00 0903300	71992-78704	transcribed hypothetical protein	no			
		GPLIN_00 0903400	86045-87892	transcribed hypothetical protein	no		GROS_g0 3490	response to chitin
		GPLIN_00	89309-92938	serpentine receptor,	no		GROS_g0	transmembrane
		0903500		class T family member			6645	dorsal gland cell-
338	SNP_338_ 98803	GPLIN_00 0903600 **	97414-99406	dorsal gland cell specific expression protein	no		GROS_g1 4146 / GROS_g1 4224	specific expression protein [<i>Heterodera</i> <i>avenae</i>] / transmembrane; dorsal gland cell- specific expression protein [<i>H. avenae</i>]
		GPLIN_00	102236-	protein MICAL 3	no		GROS_g0	LIM domain
		GPLIN_00	104125-	LIM domain	no		8157	
		0903800	104725	containing protein	110			
		0903900	106776	hypothetical protein	no			
		GPLIN_00 0904000	107274- 107909	protein MICAL 3	no			
		GPLIN_00 0904100	108895- 109595	transcribed hypothetical protein	no			

		GPLIN_00	24813-25894	transcribed hypothetical protein	no		GROS_g0 8025	SSF52540
		GPLIN_00 0904800	35241-37595	nematode astacin protease family member	no		GROS_g0 7035	extracellular region; metalloendopeptidase activity; Astacin-like metalloendopeptidase
340	SNP_ 340_ 56189	GPLIN_00 0904900	37694-40604	peptidase M12A, astacin	no		GROS_g1 2919	[<i>Strongyloides ratti</i>] astacin (peptidase family M12A)
	50105	GPLIN_00 0905000	42995-45732	cell death protein 3	no		GROS_g1 2555	execution phase of apoptosis
		GPLIN_00 0905100 **	53759-57940	transcribed hypothetical protein	no			
		GPLIN_00 0905200	75502-76417	transcribed hypothetical protein	no			
CNID	SNP_349_	GPLIN_00 0914300	13395-18410	zinc finger protein 280D	no		GROS_g1 1336 / GROS_g0 2450 / GROS_g1 1223	zinc finger protein [<i>L.</i> <i>loa</i>] / signal peptide / zinc finger C2H2 type domain signature
549	8349	GPLIN_00 0914400	31215-32425	transcribed hypothetical protein	no			
		GPLIN_00 0914500	63737-66258	transcribed hypothetical protein	no		GROS_g1 0058	signal peptide
		GPLIN_00 0914600	69418-70547	ubiquitin carboxyl terminal hydrolase 48	no		GROS_g0 3356	ubiquitinylhydrolase1
		GPLIN_00 1056200	3568-10018	transcribed hypothetical protein	no		GROS_g1 1538	zinc finger C2H2 type domain signature.
		GPLIN_00 1056300	10515-10989	transcribed hypothetical protein	no		GROS_g1 1539	transmembrane
	SNP_ 474_ 28528	GPLIN_00 1056400	11106-12182	transcribed hypothetic al protein	yes	trans- membrane	GROS_g1 1539	transmembrane
		GPLIN_00 1056500	13710-14450	transcribed hypothetical protein	no			
474		GPLIN_00 1056600	14570-20232	transcribed hypothetical protein	no		GROS_g1 1540	signal peptide
		GPLIN_00 1056700	52073-52540	transcribed hypothetical protein; Gpa_Dog_0074	yes	secreted		
		GPLIN_00 1056800	56482-58809	transcribed hypothetical protein	no			
		GPLIN_00 1056900	58931-60252	transcribed hypothetical protein	yes	secreted		
		GPLIN_00 1057000	62667-67109	transcribed hypothetical protein	no			
		GPLIN_00 1082800	211-620	protein containing SPRY domain	no		GROS_g1 4163	RBP-4 protein [<i>G.</i> <i>pallida</i>]
		GPLIN_00 1082900	11855-12808	paralog of RBP-5 protein (33H17)	no		GROS_g1 4179	RBP-1 [<i>G. pallida</i>]; signal peptide
		GPLIN_00 1083000	14667-16919	transcribed hypothetical protein	yes	trans- membrane	GROS_g1 3008	TcasGA2_TC001495 [<i>Tribolium castaneum</i>] thap domain-
504	SNP_	GPLIN_00 1083100	18255-23950	transcribed hypothetical protein	yes	secreted	GROS_g1 0456	containing protein 4 [<i>A. suum</i>]; transmembrane
504	504_ 47039	GPLIN_00 1083200	25548-26905	transcribed hypothetical protein	no			
		GPLIN_00 1083300	32640-39130	serine:threonine protein kinase SIK3	no		GROS_g1 0454	serine threonine- protein kinase kin-29 [A. suum] hmg CoA reductase A
		GPLIN_00 1083400 *	43996-50149	3 hydroxy 3 methylglutaryl coenzyme A reductase	no		GROS_g1 0452	[Polysphondylium pallidum PN500]; protein homodimerization activity

		GPLIN_00 1083500	54107-59069	transcribed hypothetical protein	no		GROS_g1 0451	peptidase M10A M12B domain containing protein [<i>H.</i> <i>contortus</i>]; transmembrane; matrixin
		GPLIN_00 1257900	8259-9831	transcribed hypothetical protein	no		GROS_g1 4146	dorsal gland cell- specific expression protein [<i>H. avenae</i>]; putative esophageal gland cell protein Hgg- 20 [<i>H. glycines</i>]
		GPLIN_00 1258000	16237-17191	transcribed hypothetical protein	no		GROS_g1 4213	BTB/POZ domain
782	SNP_ 782_ 20136	GPLIN_00 1258100 **	19137-20558	paralog of RBP-1 protein; Gpa_Dog_0077	yes	secreted	GROS_g1 4234	RBP-1 protein [G. pallida]; secreted SPRY domain- containing protein 16, partial [G. rostochiensis]
		GPLIN_00 1258200	25107-27756	ubiquitin 40S ribosomal protein S27a 1	no		GROS_g0 7568	protein ubiquitination
		GPLIN_00 1258300	31187-31435	transcribed hypothetical protein	no			
		GPLIN_00 1258400	33199-34624	protein containing SPRY domain	no		GROS_g1 4184	RBP-4 protein [G. pallida]; secreted SPRY domain- containing protein 9 [G. rostochiensis]
	SNP_ 988_ 7923	GPLIN_00 1323400	2535-3808	transcribed hypothetical protein	no		GROS_g0 1568	protein of unknown function, DUF273 serine/threonine-
	SNP_ 988_ 8027	GPLIN_00 1323500	12021-16917	serine:threonine protein kinase Nek6	no		GROS_g0 1334	protein kinase Nek6, partial [<i>Bos mutus</i>]; cytokinesis; signal
988	SNP_ 988_ 8045	GPLIN_00 1323600	19109-26588	Motor AXon guidance family member (max 2)	e no		GROS_g1 0077	protein BM-MAX-2, isoform i [<i>B. malayi</i>]; K04409 Axon guidance
	SNP_ 988_ 8100							
1243	SNP_1243 _9865				no predicted g	ene		
	SND 1777	GPLIN_00 1438500 **	864-1181	transcribed hypothetical protein	no			
1777	_961	GPLIN_00 1438600	2326-4922	maternal protein pumilio	no		GROS_g1 4286	cytosol
		GPLIN_00 1438700	6616-7313	transcriptional regulator ATRX	no			
3221	SNP_3221 _1475	GPLIN_00 1518200	1511-2451	dorsal gland cell specific expression protein	no		GROS_g1 4143 / GROS_g1 4139	transmembrane; dorsal gland cell- specific expression protein [<i>H. avenae</i>] / transmembrane; dorsal gland cell- specific expression protein [<i>H. avenae</i>]
3816					no predicted ge	ene		
5159	אור 5159 _970				no predicted ge	ene		

	Scaffold	Position on the scaffold
SNP 1	1	441870
SNP 2	241	23119
SNP 3	382	96312
SNP 4	435	39853
SNP 5	435	39854
SNP 6	435	39892

Table S2 Location of the six outlier loci indicative of genomic regions putatively involved inDésirée adaptation.

ANNEXE 4 : Protocole expérimental de l'AS-PCR

Le principe de l'allèle-spécifique PCR (AS-PCR) qui a été retenu est basé sur l'utilisation en PCR temps réel de deux amorces sens ciblant chacune un allèle (*i.e.* présence du site polymorphe - SNP - à l'extrémité 3' de l'amorce) et d'une amorce antisens commune aux deux allèles. Chaque couple sens/antisens sera utilisé dans une réaction distincte.

A) Définition des amorces avec le logiciel Primer 3 (version 0.4.0)

Allè	le 1	Site SNP		
	Amorce sens AS-PCR	1		
5'G	AAGGTGTGCGTTCATT	GTGT 3'matrice ADN	TCGATTCGACTTGTGGAAGGTC	
С	TTCCACACGCAAGTAA	CACAmatrice ADN 3'	AGCTAAGCTGAACACCTTCCAG	5 '
			Amorce commune antisens	
Allè	le 2			
	Amorce sens AS-PC	R 2		
5'G	AAGGTGTGCGTTCATT	→→ GTG <mark>C</mark> 3'matrice ADN	TCGATTCGACTTGTGGAAGGTC	
С	TTCCACACGCAAGTAA	CACAmatrice ADN 3'	AGCTAAGCTGAACACCTTCCAG	5 ′
			Amorce commune antisens	

Consignes données à Primer 3 :

- Fragments amplifiables de 80 à 130 pb
- Taille des amorces comprise entre 18 et 30 pb (idéalement 20-22 pb)
- Tm des amorces compris entre 60 et 68°C ; Pour un SNP donné, les 3 amorces ne doivent pas avoir des Tm différents de plus de 4°C
- % de GC compris entre 40 et 60.

B) Matrice ADN

L'extraction des ADNs se fera avec le kit Qiagen DNeasy Blood and Tissue Kit sur 50 larves issues d'un pool de 10 larves de 5 kystes, selon le protocole du fournisseur. Vingt populations de la collection de *Globodera pallida* issue du témoin sensible Désirée, représentatives de la diversité mondiale de *G. pallida* et utilisées pour les expérimentations de phénotypage sur la variété résistante lledher, seront utilisées pour servir de matrice ADN à l'AS-PCR.

C) L'AS-PCR

Les réactions de PCR quantitative contenant 2 μ l de matrice ADN, 0,4 μ M de chaque amorce et 1 X SybrGreen I Master (Roche) seront réalisées sur le système LightCycler 480 PCR en temps réel (Roche). Le programme de PCR quantitative consistera en une dénaturation initiale à 95°C pendant 5 min, suivie de 45 cycles de 95°C pendant 15 sec, puis 40 sec à la température d'hybridation/élongation établie pour chaque couple d'amorces testé. Une analyse de phase de dissociation de 60°C à 95°C sera réalisée à la fin de chaque essai pour déterminer si un seul amplicon a été généré pour chaque jeu d'amorce utilisé. Les courbes d'amplifications seront aussi comparées entre chaque allèle des SNPs analysés.

RÉSUMÉ ABSTRACT

pallida est un important pathogène de la pomme de terre, vis-à-vis duquel un QTL majeur de résistance, GpaVvrn, a été identifié chez Solanum vernei. Cependant, la capacité des populations de G. pallida à s'adapter à cette résistance en quelques générations seulement a été mise en évidence par évolution expérimentale. Dans ce contexte, ce travail de thèse avait pour objectifs (1) d'étudier les traits d'histoire de vie du nématode impactés par l'adaptation, afin de tester l'existence éventuelle d'un coût de virulence. et (2) d'identifier les régions génomiques impliquées dans l'adaptation, par une approche originale combinant évolution expérimentale et scans nomiques sur des lignées virulentes et avirulentes. Contre toute e 3. verner entrainait une augmentation de la *nich*ess des individus irulents sur hôte sensible. Nous avons également pu identifier des égions génomiques candidates à l'adaptation à la résistance de a plante hôte, contenant des gènes codant pour des effecteurs t notamment des SPRYSECs, connus chez les nématodes à kyste our être impliqués dans la suppression des défenses des sever

Mots-clés : coût de virulence, durabilité des résistances, évo-

highlighted through experimental evolution. In this context, the traits impacted by adaptation, in order to test for potential existence of a virulence cost, and (2) to identify genomic regions involved in adaptation, through an original approach combining experimental evolution and genome scans on virulent and avirulent lineages. Unexpectedly, we highlighted that adaptation to resistance from S. vernei leads to an increase of virulent individual's fitness on susceptible host. We were also able to pinpoint candidate genomic regions to adaptation to host plant resistance, containing genes oding effectors, and especially SPRYSECs, known in cy odes to be involved in suppression of host defense but als natode virulence. These results will ultimately be useful i

Keywords: cyst nematode, experimental evolution, genome scan, Pool-Seq, resistance durability, virulence cost.



AGROCAMPUS OUEST • Institut national supérieur des sciences agronomiques, agroalimentaires, horticoles et du paysage 65 rue de Saint-Brieuc - CS84215 - F-35042 Rennes Cedex Tél.: 02 23 48 50 00 www.agrocampus-ouest.fr

