Université de Montréal

Étude des déterminants génétiques de la pathogénicité chez les nématodes du genre *Globodera*

par Michael Sabeh

Département des Sciences Biologiques

Faculté des Arts et des Sciences

Thèse présentée
en vue de l'obtention du grade de doctorat

en Sciences Biologiques

Juillet 2019

Cette thèse intitulée

Étude des déterminants génétiques de la pathogénicité chez les nématodes du genre *Globodera*

Présentée par

Michael Sabeh

A été évaluée par un jury composé des personnes suivantes

Dr. Jacques Brodeur
Président-rapporteur

Dr. Marc St-Arnaud
Directeur de recherche

Dr. Benjamin Mimee
Codirecteur

Dr. Frédéric Pitre
Membre du jury

Dr. Mario Tenuta
Examinateur externe

Résumé

Les nématodes à kyste de la pomme de terre (*Globodera rostochiensis* et *G. pallida*) sont des phytoparasites causant des pertes de rendement considérables, évaluées à plus de 9 % de la production mondiale de pomme de terre. Comprendre les éléments génétiques essentiels permettant aux nématodes d'infecter la pomme de terre devrait pouvoir mener à l'élaboration d'une méthode de lutte durable et efficace contre ces nématodes parasites. Pour identifier les éléments génétiques essentiels à l'infection de la pomme de terre, nous avons séquencé leurs transcriptomes, ainsi que ceux d'espèces apparentées, mais incapables de parasiter la pomme de terre, *G. tabacum* et *G. mexicana*. En comparant ces deux groupes, nous avons identifié les gènes impliqués dans la spécificité parasitaire pour l'hôte. Près de 25 % des gènes codants pour des effecteurs, des protéines dont le nématode se sert dans différentes étapes du parasitisme, ainsi qu'une multitude de gènes régulateurs ont été trouvés significativement surexprimés chez les deux espèces capables de parasiter la pomme de terre. L'expression de ces gènes était différente lorsque les larves furent exposées à l'exsudat racinaire de tomate, suggérant que le nématode est en mesure d'adapter l'expression de ces gènes effecteurs selon l'hôte par l'entremise des gènes de régulation en réponse au signal chimique de l'exsudat racinaire. Aussi, plusieurs variants non-synonymes communs à chacun des groupes ont été identifiés, ainsi que des gènes non caractérisés uniques aux nématodes à kyste de la pomme de terre, ce qui pourrait contribuer à la capacité de ces nématodes à pouvoir parasiter la pomme de terre. Une comparaison semblable a aussi été réalisée entre différents pathotypes de *G. rostochiensis*, afin de comprendre la différence de virulence entre ces pathotypes sur les cultivars de pomme de terre possédant le gène de résistance H_1. Le gène de résistance H_1 est utilisé dans les cultivars, qui sont disponibles commercialement et confère une résistance contre les pathotypes Ro1 et Ro4. Une courte liste de gènes candidats possédant des variants uniques aux pathotypes avirulents, ainsi que des gènes uniques à ces pathotypes ont été identifiés et permettront la validation future du gène d'avirulence, reconnu pas le gène de résistance H_1. Étonnamment, les analyses ont permis de montrer que plusieurs gènes d'effecteurs ont un niveau d'expression différent selon le pathotype et corrèle avec le degré

de virulence de ces pathotypes. Enfin, l'identification de gènes de référence a dû être réalisée, afin de normaliser les données d'expression obtenues par RT-qPCR. Onze gènes candidats ont été sélectionnés à partir de donnée de séquençage d'ARN et de gènes validés dans d'autres études. Trois gènes de référence ont été sélectionnés basés sur la stabilité de leur expression. Les résultats présentés ici permettent un avancement certain dans la compréhension de la manière dont les nématodes parasites utilisent leurs effecteurs et comment cela leur permet de définir leur gamme d'hôtes. Le but ultime de l'étude des nématodes phytoparasites est le développement d'un moyen de lutte efficace et la compréhension du mécanisme d'infection est une étape préalable pour contrer ce parasite complexe.

Mots-clés : Nématode phytoparasite, Phytopathogène, *Globodera*, Transcriptomique comparative, Effecteur, Pathogénicité

Abstract

Potato cyst nematodes (*Globodera rostochiensis* and *G. pallida*) are plant parasite causing considerable yield losses, estimated at more than 9% of global potato production. Understanding the genetic characteristics allowing the nematode to infect potato should lead to the development of sustainable and effective control methods against these parasitic nematodes. To identify the genetic elements essential for potato parasitism, I sequenced their transcriptomes, as well as those of related species unable to parasitize potato, *G. tabacum* and *G. mexicana*. By finding differences between these two groups, it may be able to identify the genes involved in the specificity of host. Nearly 25% of the genes coding for effectors, proteins used in different stages of parasitism, as well as a multitude of regulatory genes were found to be significantly overexpressed in the two species able to parasitize potato. The expression of these genes was different when the larvae were exposed to tomato root exudate suggesting that the nematode can adapt the expression of these effector genes according to the host through regulatory genes, in response to the chemical signal of the root exudate. Also, several non-synonymous variants common to each group have been identified, as well as uncharacterized genes unique to potato cyst nematodes, which could also contribute to the ability of these nematodes to be able to parasitize potato. A similar comparison was also made between different pathotypes of *G. rostochiensis* to understand the difference in virulence on potato cultivars carrying the H_1 resistance gene. The H_1 resistance gene is used in commercially available cultivars, and confers resistance against the Ro1 and Ro4 pathotypes, but remains susceptible to the Ro2, Ro3 and Ro5 pathotypes. A short list of candidate genes with variants unique to avirulent pathotypes, as well as genes unique to these pathotypes, were identified and will allow future validation of the avirulence gene. Surprisingly, gene expression analyzes have shown that the expression of several effector genes has a different expression for each pathotypes and correlates with the degree of virulence of these pathotypes. Finally, the identification of reference genes had to be performed, to normalize the expression data obtained by RT-qPCR. Eleven candidate genes were selected from RNA sequencing data and previously validated reference genes. Three reference genes were selected based on the stability of their expression. The results presented here allow important progress in understanding how parasitic nematode use their effectors and how this allows them to define their host range. The final goal of plant-parasitic nematodes

studies is the development of effective control means and the understanding of the mechanism of infection is a preliminary step for this complex parasite.

Keywords: Phytoparasitic nematode, Phytopathogen, *Globodera*, Comparative transcriptomics, Effector, Pathogenicity

Table des matières

Liste des tableaux

Liste des figures

Liste des sigles

ADN (DNA) : Acide désoxyribonucléique

ADNc (cDNA) : Acide désoxyribonucléique complémentaire

ARN (RNA) : Acide ribonucléique

ARNdb : Acide ribonucléique double brin

ARNm (mRNA) : Acide ribonucléique messager

AP : Pathotypes avirulent

BLAST : Basic Local Alignment Search Tool

DEG : Gène différentiellement exprimé

HIGS : Inactivation de gène induit par l'hôte

J2 : Nématode juvénile de 2e stade larvaire

NKPT : Nématode à kyste de la pomme de terre

PCN : Potato cyst nematode

PCR : Réaction en chaîne par polymérase

pJ2 : Larve parasitique de 2^{e} stade

PRD : Exsudat racinaire de pomme de terre

QTL : Locus de caractères quantitatifs (Quantitative Trait Loci)

RT-qPCR : Réaction en chaîne par polymérase quantitative en temps réel

RNA-seq : Séquençage haut débit de l'ARN

TRD : Exsudat racinaire de tomate

VP : Pathotypes virulent

Liste des abréviations

% : Pourcent

°C : Degré celsius

bp : Paires de bases

Cq : Cycle de quantification

Cv : Cultivar

µL : Microlitre

µg : Microgramme

µm : Micromètre

e.g. : exempli gratia (par exemple)

et al. : et alii (et les autres)

p. ex. : par exemple

À Anna et Kelly

Remerciements

Je tiens tout d'abord à remercier énormément mes deux directeurs de thèse, Benjamin Mimee et Marc St-Arnaud, pour m'avoir guidé et conseillé tout au long de mes cinq années d'études graduées, pour m'avoir permis de devenir un chercheur et d'accumuler des connaissances et compétences qui me suivront tout au long de ma carrière. Merci à vous deux de m'avoir fait confiance et de m'avoir encouragé à continuer avec vous mon doctorat.

Je remercie tous les membres de mon équipe de recherche en nématologie du Centre de Recherche et Développement de St-Jean-sur-Richelieu d'AAC pour toute l'aide, le soutien et les conseils qu'ils m'ont offerts, ainsi que pour leur amitié. Un merci spécial à Nathalie Dauphinais, Pierre-Yves Véronneau et très spécialement Marc-Olivier Duceppe pour l'aide cruciale et pour m'avoir transmis des compétences essentielles pour mener à terme mes projets.

Merci aussi à AAC, au USDA et à l'Université de Montréal pour le support financier tout au long de mes études graduées et aussi au différents professeur/chercheurs de l'Université de Montréal et d'Agriculture et Agroalimentaire Canada, qui ont donné de leur temps pour participer aux différents comités qui m'ont évalué.

Je tiens aussi à remercier mes parents et ma famille pour l'énorme support, l'intérêt et les encouragements durant toutes ces années. Merci de m'avoir guidé et conseillé dans mon choix de carrière. Finalement, un merci tout particulier à ma femme pour m'avoir écouté, accompagné, encouragé, soutenu et pour avoir tout partagé tout au long de mes études.

Listes des Publications et Communications

Publications:

1. **Sabeh M**, St-Arnaud M, Mimee B. Comparative transcriptomic analysis of *Globodera rostochiensis* pathotypes. *En préparation*
2. **Sabeh M**, Lord E, Grenier E, St-Arnaud M, Mimee B. What determines host specificity in hyperspecialized plant parasitic nematodes? *BMC Genomics 2019*, 20:457.
3. **Sabeh M**, Duceppe M-O, St-Arnaud M, Mimee B. Transcriptome-wide selection of a reliable set of reference genes for gene expression studies in potato cyst nematodes (*Globodera* spp.). *Plos One* 2018, 13:e0193840.
4. Duceppe M-O, Lafond-Lapalme J, Palomares-Rius JE, **Sabeh M**, Blok V, Moffett P, Mimee B. Analysis of survival and hatching transcriptomes from potato cyst nematodes, *Globodera rostochiensis*, and *G. pallida*. *Scientific Reports* 2017, 7:3882. (En annexe)
5. Eves-van den Akker S, Laetsch DR, Thorpe P, Lilley CJ, Danchin EG, Da Rocha M, Rancurel C, Holroyd NE, Cotton JA, Szitenberg A, Grenier E, Montarry J, Mimee B, Duceppe M O, Boyes I, Marvin J M, Jones L M, Yusup H B, Lafond-Lapalme J, Esquibet M, **Sabeh M**, et al. The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology* 2016, 17:124. (En annexe)
6. **Sabeh M**. La révolution du séquençage de prochaine génération, *Dire*, 2015, 24: 2 (Article de vulgarisation)

Communications:

1. **Sabeh M**, Lord E, St-Arnaud M, Mimee B. 12/12/2018. Transcriptomic particularities among species of *Globodera*, QCBS Annual Meeting, Montreal.
2. **Sabeh M**, Mimee B. 06/03/2018. Identifying genes associated with host specificity among *Globodera* species. *Globodera* Alliance Annual Meeting, Cornell University, Ithaca, NY.
3. **Sabeh M,** Lord E, St-Arnaud M, Mimee B. 21/06/2017**.** Effector-triggered immunity vs non-host resistance against potato cyst nematodes, a transcriptomic analysis. Annual Meeting of the Canadian Phytopathological Society and Canadian Society of Agronomy, Winnipeg.
4. **Sabeh M**, St-Arnaud M, Mimee B. 08/06/2017. What determines the host range of potato cyst nematodes? Annual Meeting of Quebec Plant Protection Society, Montreal.
5. **Sabeh M**, Grenier E, St-Arnaud M, Mimee B. 21/07/2016. Transcriptomic particularities among species of *Globodera*. Joint Meeting of the Society of Nematologists and the Organization of Nematologists of Tropical America, Montreal.
6. **Sabeh M.** 18/06/2015. Transcriptomic study of *Globodera* spp. Ma thèse en 180 secondes 2015 de la Société de Protection des Plantes du Québec. Québec • Second Prix •
7. **Sabeh M**, Duceppe MO, St-Arnaud M, Mimee B. 16/06/2015. Selection of a reliable set of reference genes for *Globodera rostochiensis*. Annual Meeting of Quebec Plant Protection Society, Quebec.

CHAPITRE 1 : Introduction

Solanum tuberosum

Actuellement la cinquième culture en importance dans le monde et la troisième pour consommation humaine, la pomme de terre est le principal aliment non céréalier au monde. Elle est cultivée dans plus de 160 pays et représentait une production mondiale de plus de 376 millions de tonnes en 2016 (FAOSTAT, 2018), dont près de 5 millions de tonnes, au Canada. Au Québec, la pomme de terre représente la troisième plus importante culture (Les Producteurs de pommes de terre du Québec, 2015). La pomme de terre est riche en glucides, fibres, vitamines B et C, potassium et en antioxydants et est un des aliments recommandés par les Nations Unies pour atteindre la sécurité alimentaire mondiale (Camire et al., 2009). Le pathogène le plus important de la pomme de terre est le champignon *Phytophthora infestans* qui fut responsable des famines en Europe dans les années 1840. Plusieurs autres pathogènes tels des virus, bactéries, champignons, insectes et nématodes (p. ex., *Alternaria solani*, *Thanatephorus cucumeris*, virus de l'enroulement de la pomme de terre, virus Y de la pomme de terre, nématode à kyste de la pomme de terre, doryphore de la pomme de terre, *Phthorimaea operculella, Streptomyces scabies*) occasionnent des dommages estimés à 40% de la production mondiale (Pandey et al., 2005). Ces dommages restent très élevés malgré les efforts déployés, principalement à cause de la difficulté à contrer certains pathogènes, notamment les nématodes et les virus (Oerke, 2006).

Nématodes phytoparasites

Les nématodes sont des organismes très diversifiés, constituant une très grande partie du règne animal et se retrouvent dans une multitude d'environnements. Vivant le plus souvent librement et jouant un rôle essentiel dans les réseaux trophiques de leurs habitats, certaines espèces ont divergé vers un mode de vie parasitaire, se nourrissant aux dépens d'animaux ou de végétaux. Les plus connus sont *Enterobius vermicularis* et *Ascaris lumbricoides*, des parasites humains extrêmement répandus, infectant plus d'un milliard d'individus à travers le monde,

ainsi que *Caenorhabditis elegans*, un organisme modèle en biologie moléculaire (Van Megen et al., 2009). La divergence vers un mode de vie parasitaire s'est produite au moins 15 fois indépendamment et a nécessité une adaptation importante, tels le développement d'organes spécialisés et l'acquisition de facteurs de virulence (Blaxter & Koutsovoulos, 2015).

Les nématodes parasites de plantes comptent plus de 4100 espèces, dont au moins un nématode parasite connu pour les plus grandes cultures. Ils sont responsables de pertes économiques estimées à 157 milliards $US chaque année (Abad et al., 2008). Les plus dommageables sont les Heteroderidae, incluant les nématodes du genre *Globodera* spp., *Heterodera* spp. et *Meloidogyne* spp. (Decraemer & Hunt, 2006). L'adaptation à un mode de vie phytoparasitaire a nécessité tout d'abord le développement d'une structure anatomique spécialisée appelée le stylet (Quist et al., 2015). Cette structure, dure et pointue, est utilisée pour percer les membranes et capter les nutriments. Ce mode de vie a aussi nécessité le développement ou l'acquisition de protéines effectrices ou effecteurs, des protéines excrétées par les nématodes phytoparasites et qui sont essentielles à chaque étape du mécanisme d'infection de la plante hôte. Ces molécules sont généralement produites par les glandes œsophagiennes (sous-ventrale et dorsale) et sont excrétées par le stylet et par les amphides. Les effecteurs permettent la dégradation de la paroi cellulaire végétale (p. ex., β-1,4 endoglucanases, pectate lyases et expansines), le détournement du système immunitaire de la plante (p. ex., superoxide dismutase, chorismate mutase, peroxiredoxine, glutation peroxidase) et la formation d'un site de nutrition dans la racine (p. ex., annexine, calreticuline) (Davis et al., 2009). Lors de leur découverte, certains de ces effecteurs n'avaient jamais été observés dans le règne animal, ils ont plutôt été acquis par transfert horizontal principalement de bactéries phytopathogènes (Haegeman et al., 2011). Comme les effecteurs sont impliqués dans la majeure partie des interactions du parasite avec son hôte, la plupart des études portant sur les nématodes phytoparasites sont concentrées sur ce sujet ces dernières années; toutefois, leur rôle exact, ainsi que leurs interactions avec les hormones de la plante hôte restent encore très mal connus (Haegeman et al., 2012; Smant et al., 2018). En effet, à cause de leur petite taille, de leur long temps de reproduction et du fait qu'ils soient biotrophes obligatoires, il s'est longtemps avéré difficile d'étudier les nématodes phytoparasites (Jones et al., 2013).

Globodera spp.

Les nématodes à kyste de la pomme de terre (NKPT), *Globodera rostochiensis* (Wollenweber) et *G. pallida* (Stone), sont des phytoparasites originaires d'Amérique du Sud et aujourd'hui présents dans la plupart des pays où l'on retrouve leurs hôtes. Ils furent récemment introduits au Québec (2006), mais avaient déjà été introduits à Terre-Neuve (1962) et en Colombie-Britannique (1965), où ils ont nécessité la mise en place de zones de quarantaine instaurées par l'Agence Canadienne d'Inspection des Aliments, afin d'en limiter la dispersion (Mimee et al., 2014; Olsen & Mulvey, 1962; Orchard, 1965). En se nourrissant des racines des plantes hôtes, ces nématodes sont responsables de pertes de rendement pouvant atteindre 90% (Nicol et al., 2011). Les espèces *G. mexicana* (Campos-Vela) et *G. tabacum* (Lownsbery & Lownsbery) sont génétiquement similaires aux NKPT, mais possèdent un spectre d'hôte différent; elles parasitent également des espèces de la famille des Solanacées, mais ne peuvent pas se développer sur la pomme de terre. Leur hôte principal respectif est la morelle noire (*S. nigrum* L.) et le tabac (*Nicotiana tabacum* L.). Le polymorphisme de certains effecteurs (p. ex., pectate lyase, chorismate mutase) a été associé à la capacité de certaines espèces de *Globodera* et d'*Heterodera* à se développer sur des cultivars différents (Bekal et al., 2003; Blok et al., 2006; Geric Stare et al., 2012). Des variations dans le niveau d'expression de certains effecteurs pourraient aussi mener à une pathogénicité différente.

L'éclosion des œufs des nématodes du genre *Globodera* est induite par l'exsudat racinaire d'une plante potentiellement hôte. Le signal induit par l'exsudat racinaire provoque une perte de pression osmotique et la forte disponibilité en eau permet la réhydratation, l'activation, puis l'éclosion des larves contenues dans les œufs (Perry & Moens, 2011; Duceppe et al., 2017b). En absence de ce signal, 30% des œufs éclosent spontanément dans l'eau (Turner, 1996). La température optimale pour l'éclosion se situe entre 15°C et 27°C (Mimee et al., 2015a). Le nématode juvénile de 2^{e} stade (J2) éclot de l'œuf contenu dans un kyste et se dirige vers la racine, guidé par les concentrations d'exsudats racinaires. Il entre dans cette dernière grâce à des protéines spécialisées (protéines effectrices) et à l'action mécanique de son stylet, où il deviendra un nématode juvénile de 2^{e} stade parasitique (pJ2). Ensuite, il transformera certaines cellules de la plante en sécrétant des protéines effectrices mimant des hormones de croissance et de développement cellulaire végétal, induisant la fusion de jusqu'à 300 cellules végétales afin

de créer un site de nutrition, appelé syncytium (Haegeman et al., 2012). Les protéines effectrices induisent une augmentation significative de l'activité métabolique à l'intérieur du syncytium afin de fournir un apport continu de nutriments au nématode, durant tout son développement (J3, J4 et adulte) (Bohlmann, 2015). Le développement de femelles est favorisé par des conditions nutritionnelles et environnementales favorables puisqu'elles ont besoin d'atteindre une plus grande taille, ainsi que de produire des centaines d'œufs, ce qui induit également la formation d'un syncytium de près de 10 fois plus gros que celui des mâles (Chauvin et al., 2008). Ensuite, la femelle adulte reste attachée à la racine, alors que le mâle adulte retourne dans le sol, où il fertilisera une ou plusieurs femelles encore attachées dans la racine, guidé par des phéromones sexuelles. La femelle fécondée produit jusqu'à 300 œufs avant de mourir. Son corps formera un kyste de couleur doré à brun foncé, pour *G. rostochiensis* et brun pâle à crème pour *G. pallida*. À l'intérieur de celui-ci, les œufs peuvent survivre jusqu'à 20 ans en attente d'un hôte potentiel (Jones et al., 2013; Sabaratnam, 2012). Dans un climat tempéré, une seule génération est produite par saison de croissance (Mimee et al., 2015a). Dans une interaction incompatible, comme c'est le cas pour *G. mexicana* sur *S. tuberosum*, le nématode éclot lorsqu'il reçoit le signal chimique contenu dans l'exsudat racinaire de pomme de terre, il se dirige alors vers les racines et la pénètre. À ce stade, il tente de former son site de nutrition, mais n'y parvient pas. Il ne pourra donc se nourrir et compléter son cycle de vie (Thiéry et al., 1997).

Les populations de *G. rostochiensis* sont classées en cinq pathotypes (Ro1 à Ro5), et celles de *G. pallida* en trois pathotypes (Pa1 à Pa3) (Kort et al., 1977). La classification repose sur leurs capacités de reproduction sur sept lignées de pomme de terre. Nijboer *et al.* (1990) ont proposé une réorganisation de cette classification comme suit : le pathotype Ro1 (pour les pathotypes Ro1 et Ro4), Ro3 (pour les pathotypes Ro2 et Ro3) et Ro5 pour *G. rostochiensis,* ainsi que le pathotype Pa, incluant les pathotypes Pa1 à Pa3 de *G. pallida* (Tableau 1.1). Bien que cette proposition ne soit pas utilisée, elle met de l'avant la grande difficulté encore aujourd'hui, de distinguer certains pathotypes qui sont indiscernables morphologiquement et qui ne se différencient pas suffisamment en termes de spécificité d'hôte non plus(Geric Stare et al., 2012; Nijboer & Parlevliet, 1990). Pour l'espèce *G. rostochiensis*, seul le pathotype Ro1 est actuellement présent au Canada et son développement est réprimé efficacement par le gène de résistance H_1, retrouvé chez certains cultivars de pommes de terre. Ce gène de résistance

provoque une réaction d'hypersensibilité lorsque le nématode tente de créer un site de nutrition à l'intérieur de la racine. Le nématode ne peut donc pas compléter son cycle de vie (Dalamu et al., 2012). Ce gène de résistance est efficace contre le pathotype Ro1 et Ro4; cependant, les pathotypes Ro2, 3 et 5 ne sont pas affectés. Récemment, des cultivars résistants ont été développés contre le pathotype Ro2, présent dans l'état de New York (Dandurand, L-M., en préparation) (USDA, 2008).

Tableau 1.1 : Classification des pathotypes de *Globodera rostochiensis* et *G. pallida* utilisés actuellement (Kort et al.,1977) (A) et classification proposée par Nijboer *et al.* (1990) (B); + signifie susceptibilité (Pf/Pi > 1 .0) et - signifie résistance (PI/Pi <1 .0).

	A								B			
Lignées différentielles	Ro1	Ro2	Ro3	Ro4	Ro5	Pa1	Pa2	Pa3	Ro1/4	Ro2/3	Ro5	Pa
S. tuberosum ssp. tuberosum	+	+	+	+	+	+	+	+	+	+	+	+
S. tuberosum ssp. andigena (H_1)	-	+	+	-	+	+	+	+	-	+	+	+
S. kurtzianum, 60 .21 .19	-	-	+	+	+	+	+	+	±	±	+	+
S. vernei, 58 .1642.4	-	-	-	+	+	+	+	+	±	-	+	+
S. vernei, 65 .346 .19	-	-	-	-	-	+	+	+	-	-	-	+
S. vernei, 62.33 .3	-	-	-	-	±	-	-	+	-	-	±	±
S. multidissectum, P 55/7	+	+	+	+	+	-	+	+	+	+	+	±

Méthodes de lutte actuelles

À cause de leur persistance dans le sol, dû à la protection apportée par la structure du kyste, les NKPT sont très difficiles à contrôler en milieux agricoles. Les méthodes actuelles de contrôle incluent : les règlements stricts des agences de contrôle (p. ex., le nettoyage de l'équipement agricole, la gestion des pommes de terre récoltées), la désinfection des sols grâce aux nématicides et aux fumigants (néfastes pour l'environnement, toxiques et dont plusieurs sont interdits au Canada), la solarisation, ainsi que les pratiques agronomiques comme l'utilisation de cultures pièges et les rotations de cultures (efficaces à long terme seulement, à cause de la persistance de kyste et de la présence d'hôtes secondaires dans les agroécosystèmes) et l'utilisation de cultivars résistants (Chauvin et al., 2008).

Actuellement, près d'une vingtaine de locus de résistance au NKPT ont été identifiés dans le génome de la pomme de terre, dont certains confèrent une résistance monogénique presque totale contre certains pathotypes. Les gènes de résistance H_1 (efficace contre Ro1 et 4), H_2 (contre Pa1), Gpa2 (contre Pa2 et 3), *Gro1-4* (contre Ro1) et *GroVI* (contre Ro1) provoquent une réaction d'hypersensibilité accompagnée d'une accumulation de lignine autour du syncytium, ce qui empêche le développement du nématode (Finkers-Tomczak et al., 2009). Le plus utilisé commercialement, le gène de résistance H_1, offre une résistance totale contre les pathotypes Ro1 et 4, une résistance partielle/faible contre les pathotypes Ro2 et 3, mais aucune résistance contre le pathotype Ro5. D'autres locus de résistance présentent une résistance polygénique, tels les locus H_3 (contre Pa2 et 3) ou QTL (Quantitative Trait Loci ou Locus de caractères quantitatifs (Chauvin et al., 2008)). Toutefois, une utilisation répétée de ces cultivars résistants peut mener à la sélection de sous-populations adaptées, insensibles aux gènes de résistance, comme observée dans l'état de New York (USDA, 2008). Cette pression de sélection pourrait également faciliter l'émergence d'espèces résistantes, si elles sont déjà présentes de façon cryptique, comme ce fût le cas avec *G. pallida* au Royaume-Uni (Dalamu et al., 2012). De plus, des populations avirulentes peuvent parfois s'adapter aux gènes de résistance et peuvent ainsi devenir virulentes, comme c'est le cas des populations de *G. pallida* face au QTL de résistance $GpaV_{vrn}$ (Eoche-Bosy et al., 2017).

Les progrès récents dans l'étude des nématodes phytoparasites rendent possible l'exploitation d'aspects spécifiques de leur relation avec leur hôte afin de mettre au point des stratégies de contrôle. Celles-ci pourraient permettre à la plante de prévenir l'invasion du nématode, de réduire l'efficacité du déplacement du nématode dans les tissus, ainsi que de réduire la capacité de se nourrir, de se développer ou de se reproduire du nématode. La découverte de gènes essentiels au parasitisme du nématode pourrait aussi être utilisée pour développer de nouveaux moyens de contrôle (Fosu-Nyarko & Jones, 2015). L'utilisation d'interférence par ARN pourrait, par exemple, réduire la pathogénicité du nématode. Une étude, employant cette technique afin d'évaluer l'importance de gènes d'effecteurs candidats, a montré une réduction importante du taux de reproduction du nématode parasite et pourrait s'avérer extrêmement efficace en ciblant un gène essentiel au parasitisme (Ali et al., 2017).

Transcriptomique et RT-qPCR

Les interactions moléculaires entre différents organismes peuvent être subtiles et ne mènent pas toujours à un phénotype facilement observable, par exemple, une réaction d'incompatibilité ou d'hypersensibilité. Même dans ces cas, il est difficile de mettre en évidence les éléments génétiques responsables. Les avancées dans les techniques de séquençage d'ARN à haut débit et d'analyses bio-informatiques, permettent la comparaison de l'expression et des séquences de gènes d'intérêts afin d'identifier des particularités propres à certaines populations. Ces profils d'expression et variants pourraient permettre de détailler les associations entre certains gènes d'effecteurs et/ou des régulateurs et la pathogénicité chez *Globodera* spp.

L'analyse de données de séquençage d'ARN à haut débit nécessite un traitement bio-informatique complexe afin d'identifier des gènes d'intérêt, ici impliqués dans la pathogénicité des différentes populations de *Globodera* spp. Des analyses d'expression des gènes peuvent être réalisées à partir de données de séquençage d'ARN, à l'aide de logiciels bio-informatiques spécialisés (p. ex., DESEQ2, EDGER, CORSET) (Davidson & Oshlack, 2014b; Hardcastle & Kelly, 2010; Love et al., 2014b). Il est ainsi possible de comparer l'expression de gènes entre différentes populations ou différents groupes afin d'identifier les gènes différentiellement exprimés communs à ces groupes. L'identification de variants, ou mutations, génétiques peut aussi être réalisée avec des données de séquençage, grâce à des logiciels spécialisés (p. ex., FREEBAYES, SAMTOOLS MPILEUP, SNPEFF), qui identifient chaque variant possible (polymorphisme nucléotidique, insertion / délétions et variant complexe), ainsi que leur effet sur une protéine (Garrison & Marth, 2012a; Li, 2011). Néanmoins, à cause des nombreuses étapes complexes et nécessaires, et des erreurs de séquençage, il s'avère encore nécessaire de valider l'expression ou le variant de gènes d'intérêts par RT-qPCR ou séquençage Sanger afin d'obtenir des résultats concluants.

La RT-qPCR, la technique la plus communément utilisée, est considérée comme étant la plus fiable pour analyser le niveau d'expression de gènes (Nolan et al., 2006). La technique consiste à mesurer l'augmentation de la quantité d'acides nucléiques après chaque cycle d'amplification et détermine avec précision la concentration initiale de transcrits d'un gène dans l'échantillon (Nolan et al., 2006). Cependant, la préparation des échantillons, de l'extraction

d'ARN jusqu'à la réaction de PCR, mais aussi la quantité et la qualité initiale du matériel génétique extrait de chaque échantillon, peuvent introduire des variations dans la quantification de l'expression des gènes. C'est pourquoi il est courant d'avoir recours à des gènes de référence pour normaliser les données obtenues (Chervoneva et al., 2010). Un bon gène de référence doit avoir une expression constante durant toute l'expérience sans être influencé par les traitements expérimentaux (Hoogewijs et al., 2008; Zhang et al., 2012). Il est courant d'utiliser plusieurs gènes de référence afin de minimiser la variation de l'expression de ces gènes; cependant, comme la disponibilité de matériel génétique est généralement restreinte en conditions expérimentales, on doit limiter le nombre de gènes de référence utilisés (Vandesompele et al., 2002). C'est pourquoi la sélection d'un petit groupe de gènes de référence fiables est essentielle. Afin d'aider la sélection de gènes de référence fiables, plusieurs algorithmes ont été développés pour calculer et évaluer la variation d'expression des gènes de manière objective. BestKeeper (Pfaffl et al., 2004), NormFinder (Andersen et al., 2004), geNorm (Vandesompele et al., 2002) et la méthode delta Ct (Silver et al., 2006) sont toute des méthodes de calculs qui permettent d'évaluer la stabilité de gènes candidats. Par exemple, geNorm possède un algorithme qui calcule l'écart-type du ratio d'expression après une transformation logarithmique, tout en formant une matrice de calcul en comparaison par paires (Chervoneva et al., 2010). RefFinder (Xie et al., 2012b) est un outil qui rassemble tous ces algorithmes afin de fournir un classement final des gènes candidats. Un ensemble de gènes de référence peut ensuite être validé en normalisant l'expression de gène dont l'expression est connue. Le niveau d'expression des gènes d'intérêt étant calculé en fonction de l'expression des gènes de références, le choix de ces gènes représente une étape extrêmement importante. Le niveau d'expression est calculé comme étant lié au changement relatif avec les gènes de référence, à l'aide de la méthode $2^{-\Delta\Delta CT}$, afin d'en connaître l'expression relative (Livak & Schmittgen, 2001). Étant la méthode la plus fiable, le RT-qPCR est utilisé pour valider l'expression obtenue par d'autres techniques (p. ex., séquençage d'ARN), qui ne sont pas encore aussi éprouvées.

Hypothèse et Objectifs de la thèse

Mon travail de doctorat a comme cadre général l'étude des différences transcriptomiques entre les espèces et pathotypes du genre *Globodera*, afin d'identifier les caractéristiques pouvant

expliquer les différences de virulence et de pathogénicité de ces phytoparasites. Les principaux enjeux, les chapitres, l'hypothèse et les objectifs spécifiques sont présentés ci-dessous.

Les nématodes à kyste du genre *Globodera* parasitent des plantes hôtes d'une grande importance économique, principalement la pomme de terre, mais aussi la tomate, l'aubergine et le tabac. Le nématode à kyste de la pomme de terre est l'un de ses principaux ravageurs avec des pertes de production évaluées à plus de 9% de la production mondiale (Nicol et al., 2011). Il est essentiel de comprendre les mécanismes moléculaires liés à la spécificité parasitaire, ainsi que ceux liés au pouvoir pathogène, afin de développer des moyens de contrôle durables et efficaces. Dans le cadre de cette thèse, l'objectif général était de comprendre les caractéristiques génétiques qui permettent à certaines espèces du genre *Globodera* de parasiter la pomme de terre alors que d'autres en sont incapables et d'expliquer les variations de virulence entre les pathotypes de *Globodera rostochiensis*. Cette thèse a pour but de tester l'hypothèse suivante :

> *Des mutations dans leurs séquences et/ou un taux d'expression spécifique de gènes codant pour des effecteurs expliquent le pouvoir pathogène ou les différences de virulence entre certaines espèces ou pathotypes du nématode à kyste de la pomme de terre.*

L'objectif spécifique du Chapitre 2 était d'identifier des gènes de référence fiables pour la normalisation des données d'expression obtenues par RT-qPCR. En effet, afin d'obtenir des données d'expression fiables et d'éliminer les variations liées au processus d'obtention du matériel biologique, il est nécessaire de recourir à un contrôle interne lors des prises de données. Ces contrôles internes doivent être minutieusement sélectionnés et leur stabilité doit avoir été validée au préalable. En plus d'être utilisée dans les chapitres suivants, la publication de ces gènes de références permettra à d'autres équipes de recherche d'utiliser directement ces gènes, dans un contexte similaire, ou encore, de réutiliser la méthode de sélection de gènes de références afin de l'appliquer à un autre système. En effet, cette méthode permet de sélectionner des gènes à partir d'une plus grande banque de gènes candidats, puisque des données d'expression issues de séquençages d'ARN (RNA-seq) sont utilisées pour effectuer une présélection, afin de ne pas se limiter aux gènes de référence traditionnels qui ne sont souvent pas optimaux. Ce chapitre a été l'objet d'une publication dans le journal Plos ONE.

L'objectif spécifique du Chapitre 3 était d'analyser les différences transcriptomiques entre quatre espèces du genre *Globodera*, afin d'identifier les éléments liés à la capacité de parasiter la pomme de terre. Les nématodes *G. rostochiensis* et *G. pallida* sont des parasites de la pomme de terre et bien que faisant partie du même genre et étant très proches génétiquement, *G. tabacum* et *G. mexicana* sont incapables d'infecter cette plante. Une analyse du séquençage de l'ARN de larves J2 suite au contact avec l'exsudat racinaire de pommes de terre a été réalisée afin d'identifier des différences, chez des gènes essentiels au processus de parasitisme, entre les espèces. L'expression de gènes d'intérêt fut également validée par RT-qPCR et les gènes de référence sélectionnés au Chapitre 2 ont été utilisés afin de normaliser les données d'expression. L'expression de ces gènes a aussi été mesurée après l'exposition à l'exsudat racinaire de tomate, afin de voir si l'expression en serait modifiée par la nature de l'hôte potentiel présent. En plus d'établir qu'il existe une différence dans la régulation de la transcription des gènes d'effecteurs lorsqu'ils sont exposés à l'exsudat racinaire de pomme de terre, ce chapitre démontre que le groupe d'effecteurs utilisés par le nématode phytoparasite pourrait être spécifique à l'hôte potentiel. Ce chapitre a été l'objet d'une publication dans le journal BMC Genomics.

L'objectif du Chapitre 4 était d'effectuer une comparaison transcriptomique des cinq pathotypes de *G. rostochiensis*, afin d'expliquer les degrés de virulence différents observés sur des hôtes possédant le gène de résistance H_1. L'analyse transcriptomique a été réalisée afin d'identifier des caractéristiques associées au degré de virulence de chaque pathotype. Une courte liste de candidats, gène unique ou possédant un variant unique aux populations avirulentes, pour l'identification du gène d'avirulence fût établie. L'article découlant de cet objectif est en préparation. Une vérification supplémentaire, par PCR, de la présence de gènes uniques aux pathotypes avirulents chez d'autres populations permettra de valider l'importance de ces gènes, autant pour la caractérisation du mécanisme moléculaire d'infection que pour le développement d'un test diagnostic permettant d'identifier directement la présence d'un nématode virulent. Cette vérification sera effectuée sur huit populations supplémentaires et sera réalisée avant la publication de l'article.

La thèse comprend également deux articles présentés en annexe, qui rapportent des travaux réalisés au cours de ce doctorat et qui furent essentiels à la réalisation du projet. L'Annexe 1 est un article rapportant le premier assemblage et la description du génome et du transcriptome de

Globodera rostochiensis. Cette analyse a permis d'identifier des gènes issus de transfert horizontal, ainsi qu'un motif présent dans le promoteur d'une grande proportion de gènes d'effecteurs. Cette publication fut le produit d'un projet de coopération internationale d'une dizaine d'équipes de recherche, réparties dans six pays. L'article offre une référence avec un génome complet annoté, comportant l'identification de plusieurs gènes d'effecteurs, permettant une meilleure analyse des données de mon objectif principal de thèse (Chapitres 3 et 4). En tant que coauteur, ma contribution fût de préparer une partie du matériel biologique (pathotypes Ro2,3,4,5 ainsi que plusieurs étapes du cycle de vie du pathotype Ro1), d'effectuer l'extraction de l'ARN en vue du séquençage, ainsi que de participer à l'annotation du génome (Eves-Van Den Akker et al., 2016). L'Annexe 2 est un article présentant une analyse des changements d'expression des gènes durant le cycle de vie de *Globodera rostochiensis*, qui a permis d'identifier des gènes dont une grande variation d'expression coïncide avec des moments clés du cycle de vie, notamment l'éclosion et le stade parasitaire J2. Le gène Neprilysin qui avait été identifié comme étant impliqué dans le processus d'éclosion, fût également identifié comme étant significativement surexprimé chez les espèces pathogènes au Chapitre 3, ce qui nous a poussés à investiguer davantage les niveaux d'expression de ce gène dans différentes conditions. Cet article fut réalisé parallèlement aux travaux rapportés au Chapitre 1, puisque les mêmes données de séquençage furent utilisées. Les gènes de référence utilisés lors de la validation des résultats de cet article sont en outre issus du Chapitre 1. En tant que coauteur, ma contribution fût de préparer du matériel biologique, d'effectuer l'extraction de l'ARN en vue du séquençage, de valider l'expression de gènes d'intérêt au RT-qPCR, ainsi que de participer à l'écriture de l'article (Duceppe et al., 2017).

CHAPITRE 2 : Sélection d'un ensemble fiable de gènes de référence pour l'étude d'expression génique chez le nématode à kyste de la pomme de terre.

Ce chapitre a été publié dans la revue *PLoS ONE*

Sabeh M, Duceppe MO, St-Arnaud M, Mimee B (2018) Transcriptome-wide selection of a reliable set of reference genes for gene expression studies in potato cyst nematodes (*Globodera* spp.). PLOS ONE 13(3): e0193840.

Résumé

Les analyses d'expression génique relatives par qRT-PCR nécessitent un contrôle interne pour normaliser les données d'expression des gènes d'intérêt et éliminer la variation introduite par la préparation de l'échantillon. Un gène de référence parfait devrait avoir un niveau d'expression constant sous toutes les conditions expérimentales. L'avènement du RNA-Seq et la disponibilité de bases de données publiques pour de nombreux organismes ouvrent la voie à la découverte de meilleurs gènes de référence pour les études d'expression. *Globodera rostochiensis* est un nématode phytoparasite limitant le rendement des cultures de pomme de terre. Le but de notre étude fut d'identifier un ensemble fiable de gènes de référence pour étudier l'expression des gènes de *G. rostochiensis.* Les niveaux d'expression génique déterminés à partir d'une base de données RNA-Seq, ainsi que des gènes de référence précédemment publiés pour d'autres nématodes ont été utilisés pour identifier des gènes de référence potentiels. Onze gènes candidats ont été validés avec une analyse qRT-PCR. Refinder, un logiciel pour évaluer des gènes de référence candidats a été utilisé sur les données d'expression du séquençage d'ARN et de l'analyse RT-qPCR. Trois gènes, GR, PMP-3, et aaRS, se sont révélés très stables dans les conditions expérimentales de cette étude et ont été utilisé pour normaliser l'expression de trois gènes dont l'expression était déjà connue et sont ainsi proposés comme gènes de référence pour de futurs travaux.

Transcriptome-wide selection of a reliable set of reference genes for gene expression studies in potato cyst nematodes (*Globodera* spp.)

Michael Sabeh[1,2], Marc-Olivier Duceppe[3], Marc St-Arnaud[2], Benjamin Mimee[1]

[1] Saint-Jean-sur-Richelieu Research and Development Centre, Agriculture and Agri-Food Canada, St-Jean-sur-Richelieu, Québec, Canada, [2] Biodiversity Centre, Institut de recherche en biologie végétale, Université de Montréal al and Jardin botanique de Montréal, Montreal, Québec, Canada, [3] Ottawa Laboratory Fallowfield, Canadian Food Inspection Agency, Ottawa, Ontario, Canada

Abstract

Relative gene expression analyses by qRT-PCR (quantitative reverse transcription PCR) require an internal control to normalize the expression data of genes of interest and eliminate the unwanted variation introduced by sample preparation. A perfect reference gene should have a constant expression level under all the experimental conditions. The advent of RNA-Seq and the availability of public datasets for numerous organisms are opening the way to finding better reference genes for expression studies. *Globodera rostochiensis* is a plant-parasitic nematode that is particularly yield-limiting for potato. The aim of our study was to identify a reliable set of reference genes to study *G. rostochiensis* gene expression. Gene expression levels from an RNA-Seq database were used to identify putative reference genes and in addition to reference genes publish for other nematodes, 11 candidate reference genes were validated with qRT-PCR analysis. Refinder, a software for evaluating candidate reference gene was used with data from RNA-seq and qRT-PCR analysis. Three genes, GR, PMP-3, and aaRS, were found to be very stable within the experimental conditions of this study and were used to successfully normalize three genes with known patterns of expression being, therefore, proposed as reference genes for future work.

Background

Quantitative reverse transcription PCR (qRT-PCR) is the most commonly used technique to measure the expression level of a particular gene and is considered to be the most accurate and reliable method so far (Kozera & Rapacz, 2013; Radonić et al., 2004). However, sample preparation, from RNA extraction to complementary DNA (cDNA) synthesis, can introduce biases in the quantification. To overcome these sources of variation, reference genes are often used as internal controls to normalize the expression data (Hoogewijs et al., 2008). However, it is now recognized that conventional housekeeping genes used as references (e.g., glyceraldehyde-3-phosphate dehydrogenase and β-actin) are not systematically appropriate, owing to their variability in some conditions (Bustin & Nolan, 2004; Dheda et al., 2005; Huggett et al., 2005). It is therefore important to select and validate good reference genes based on their expression stability within all the experimental conditions, in order to ensure valid results (Kozera & Rapacz, 2013).

Until now, microarrays were frequently used to test large numbers of genes simultaneously for the selection of good reference genes. Although this is not a bad approach, the technique requires previous knowledge of the nucleotide sequences of each candidate gene. With the advent of next-generation sequencing and, especially, RNA sequencing (RNA-Seq), this problem is now overcome, and quantifying a large number of gene transcripts without previous knowledge of their gene sequences can be done routinely. The RNA-Seq method yields millions of reads that can be assembled to generate a transcript database, which contains the sequences and expression levels of all expressed genes at a given time. Because RNA-Seq is quantitative, it can be used to study RNA expression (Nagalakshmi et al., 2008; Wang et al., 2009). Therefore, analyzing RNA-Seq data from different treatments should allow the identification of reliable reference genes for further qRT-PCR analyses, a strategy that has rarely been used to date, and to our knowledge, never applied to *Globodera* spp.

The potato cyst nematodes (PCNs), *Globodera rostochiensis* Wollenweber and *G. pallida* Stone, are plant-parasitic nematodes that affect exclusively Solanaceae plants, including potato (*Solanum tuberosum* L.), tomato (*S. lycopersicum* L.), and eggplant (*S. melongena* L.) (Bélair, 2005). Present in over 75 countries around the world (Yu et al., 2010) and recognized as

quarantine organisms, PCNs limit plant growth and are among the most economically damaging nematodes (Jones et al., 2013). They are responsible for estimated yield losses of 9% of the world's production of potatoes, currently the fifth most important crop in the world, with an annual production of more than 368 million metric tons (FAOSTAT, 2015) (Food and Agriculture Organization, 2016).

The nematodes enter the roots of the host plant, where they transform one of the plant cells to create a complex feeding site to pump plant nutrients. This will limit plant growth and eventually causes heavy yield loss. After maturation and fecundation, PCN females will dry to form a cyst containing up to 300 eggs. Inside their protective shell, these eggs can survive more than 20 years in the soil without a suitable host (Den Nijs & Karssen, 2004; Turner, 1996). Because of the strength and durability of the cysts, PCN populations are difficult to control with currently available control strategies (Chauvin et al., 2008). Hatching, host penetration, and establishment of the feeding cell are key stages of the life cycle that we need to focus on in order to find new PCN control methods. Studying gene expression during these stages could highlight essential genes against which control strategies could be developed. Therefore, we need a good technique to estimate gene expression levels during these key stages, as well as a reliable set of reference genes with constant expression across the experiment.

Many studies have been published reporting reference genes to analyze the expression of the plant gene transcripts, such as in potato roots infected with *G. pallida* (Castro-Quezada et al., 2013) and in giant cells and syncytia induced by *Meloidogyne incognita* and *Heterodera schachtii* (Hofmann & Grundler, 2007), but to our knowledge, no studies have yet focused on finding reference genes for PCN species. The aim of our study was to identify a set of potential reference genes for *G. rostochiensis* based on expression data obtained by RNA-Seq and to test reference genes previously reported in *Caenorhabditis elegans* for their expression in *G. rostochiensis*. In addition, the gene expression stability of the selected candidates was also evaluated in *G. pallida*.

Materials and Methods

RNA-Seq dataset

This study took advantage of a recently published transcriptome dataset from a *G. rostochiensis* hatching experiment conducted by our team (Duceppe et al., 2017a). These data (NCBI bioproject accession number PRJNA274143) include sampling at different life stages, including dry cyst, hydrated cyst, hydrated cyst soaked in potato root diffusate (PRD) for 1 h, 8 h, 24 h, 48 h, and 7 d, and fully hatched infective larvae (J2). Library preparation and sequencing were performed at McGill University and Génome Québec Innovation Centre (Montréal, Québec, Canada) using the TruSeq RNA sample prep kit v2 (Illumina) and a HiSeq 2000 sequencer (Illumina, San Diego, California, United States). All eight samples were multiplexed and sequenced in one lane for 100 bp paired-end reads. Two replicates were processed and assembled into a *de novo* transcriptome using the Trinity assembler (for details, see Duceppe et al., 2017).

Selection of candidate reference genes

A set of 15 genes commonly used as references in the model nematode *C. elegans* (Hoogewijs et al., 2008; Taki & Zhang, 2013; Zhang et al., 2012) were selected for evaluation in *G. rostochiensis*. These genes were *pmp-3*, *Y45F10D.4*, *tba-1*, *cdc-42*, *csq-1*, *eif-3*, *ama-1*, *mdh-1*, *gpd-2*, *act-1*, *act-2*, *F35G12.2*, *rbd-1*, *rgs-6*, and *unc-16*. The nucleotide sequences of these genes were retrieved from GenBank (Benson et al., 2009) and searched in the *G. rostochiensis* RNA-Seq transcriptome using BLAST to identify orthologues. Only the genes expressed in all the experimental conditions were kept for further analyses. The RNA-Seq dataset was also screened to identify transcripts with constant expression values in all the experimental conditions. This measurement of stability was based on standard deviations and expression variation through treatments, using log-transformed quantiles from normalized data. Only the four most stable genes were kept for further analyses.

Expression stability analysis of candidate reference genes

RNA extraction and cDNA synthesis

The nematode populations used in this study were from a greenhouse rearing of *G. rostochiensis* pathotype Ro1, initially isolated from St-Amable, Québec, Canada (obtained from Guy Bélair, AAFC), and *G. pallida* pathotype Pa2/3 from Noirmoutier, Vendée, France (obtained from Éric Grenier, INRA). The experimental conditions used were the same as in Duceppe et al. (Duceppe et al., 2017a): dry cyst, hydrated cyst, hydrated cyst soaked in PRD for 1 h, 8 h, 24 h, 48 h, and 7 d, and fully hatched infective larvae (J2). Each *G. rostochiensis* sample contained approximately 1000 cysts, while the *G. pallida* samples contained only approximately 100 cysts because of low availability. Each sample was homogenized in 650 µL of RLT Plus buffer (Qiagen, Hilden, Germany) with one 6-mm zirconium grinding bead and 200 µL of 1-mm zirconium beads using the PowerLyzer 24 Homogenizer (Mo Bio, Carlsbad, California, United States) before RNA extraction. Total RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. All samples were treated with DNase (DNase I, New England Biolabs, Ipswich, Massachusetts, United States). A 2100 Bioanalyzer system (Agilent Technologies, Santa Clara, California, United States) was used to assess RNA concentration and purity. First-strand cDNA was synthesized with the SuperScript II reverse transcriptase (Invitrogen, Carlsbad, California, United States) according to the manufacturer's instructions, from 0.5 µg of total RNA and using oligo $(dT)_{18}$. Three replicates were made for each treatment.

Primer design and qRT-PCR

Primers were designed using PrimerQuest tool (Integrated DNA Technologies, Inc., Coralville, Iowa, United States) based on the sequences retrieved from the *G. rostochiensis* RNA-Seq dataset. Target fragments lengths were designed between 84 and 130 bp. Primer information for the candidate reference genes is listed in Table 2.1 Reactions were prepared using QuantiTect SYBR Green PCR kit (Qiagen) and amplified on a Mx3000P qPCR System (Agilent Technologies) in a final volume of 20 µL according to the manufacturer's instructions. Melting curve analyses were done following the amplification cycles in order to examine the

specificity of the reactions. Amplification efficiencies were calculated with dry cysts using the Real-time PCR Miner algorithm (ver. 4.0) (Zhao & Fernald, 2005).

Tableau 2.1: Description of genes and primers used in this study.

Gene symbol	Selection	Gene description	Primer sequence (5′3′) forward/reverse	Amplicon length (bp)	Amplification efficiency (%)
PMP-3	*C. elegans*[1]	Putative membrane transporter	CTGGTTGCTGAGCAGGATAA/ GATGAAGCCCGATTGGTAGAA	102	83
Y45F10D.4	*C. elegans*	Putative iron-sulfur cluster assembly enzyme	CCAAGCAGCACTGAGTGATTA/ CATGATCCGCCGGGTTTATT	116	84
CSQ-1	*C. elegans*	Calcium-binding protein	GGTTGTGTTCTTCAACGATGTG/ ACCCTCAGCCTTTGTTCTTT	102	85
EIF-3	*C. elegans*	Eukaryotic initiation factor 3	CCGCCAGTCCTATGTCATTTA/ CTTCTTCGGGTCGGTGTATTAT	130	87
AMA-1	*C. elegans*	AMAnitin resistant family member	CTCCAAGCTCTCCACGTTATT/ GGCGAAGTTGGACTGTATGT	118	84
MDH-1	*C. elegans*	Malate dehydrogenase	GCTGGACAAATTGGCTATTCAC/ GAATGTCGAGGAGAACGAGAAC	94	86
Act-1	*C. elegans*	Actin-1	TGTAACCCACACTGTACCAATC/ TTCATGAGGTAGTCGGTCAAATC	99	89
aaRS	RNA-Seq[2]	Aminoacyl tRNA synthetase	CGGATTTACGGACCTTGTCTAC/ GGGAATCCGTCACGCTTAAT	84	86
mce1	RNA-Seq	mRNA capping enzyme	CCCGCATAAACTCCCATCTT/ CTTCACACCGATTTGCCTTTC	118	85
GR	RNA-Seq	Glutathione reductase	TTGAGAGACCATGCCGATTAC/ GAGTTGAGACGCCGAATGT	101	80
ArgRS	RNA-Seq	Arginyl-tRNA synthetase	GCCAACGCAAGAACCTTTAC/ GCGACGTCGGGATGATATTT	109	83
NEP-1	Validation[3]	Neprilysin NEP-1	GCTGAAATGGTGGAGAAAGTG/ TTTGACGCCCGAGTAGAAG	457	82
cht-2	Validation	Chitinase	ACAACTATTATGGCGGAAGGAG/ GGTGTTGAGTGAATCAGAAGGA	94	83
eng	Validation	β-endoglucanase	CTCATACCCACACGTTCTCTAC/ TAGCCTGATTTGGACTTGGG	106	85

[1] Reference genes previously reported in *C. elegans* and having orthologues in *G. rostochiensis*.

[2] Contigs from the analysis of the *G. rostochiensis* RNA-Seq database with stable expression levels across all experimental conditions.

[3] Genes known to have different levels of expression in *G. rostochiensis* and used to validate the candidate reference genes.

https://doi.org/10.1371/journal.pone.0193840.t001

Data analyses

RefFinder (Xie et al., 2012a), a wrapper tool that integrates the statistical algorithm of BestKeeper (Pfaffl et al., 2004), NormFinder (Andersen et al., 2004), and geNorm (Vandesompele et al., 2002), and the ΔCt method (Silver et al., 2006), was used to compare and rank the tested candidate reference genes. Based on the rankings from each program, it assigns an appropriate weight to an individual gene and calculates the geometric mean of their weights for the overall comprehensive ranking. Gene expressions from both the RNA-Seq database (read numbers) and qRT-PCR data (Cq values) across all treatments were compared. The variability of each gene across all treatments and replicates was also directly observed by plotting the distribution of the raw Cq values from the qRT-PCR experiment.

Validation of reference genes

Relative expression analyses of three genes with published expression data (*NEP-1*, *cht-2*, and *eng*) were performed using the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001) in order to validate the selected reference genes. The genes *GR*, *PMP-3*, and *aaRS* were used as a reference set to normalize the expression of the targeted genes. Dry cysts were used as the calibrator to calculate the fold changes for the other treatments. Gene expression was also normalized using the *Act-1* gene for comparison.

Results

Selection of candidate reference genes

Among the 15 putative reference genes previously reported in *C. elegans* and selected for this study, 11 were found to have orthologues in *G. rostochiensis* (*PMP-3*, *Y45F10D.4*, *tba-1*, *cdc-42*, *CSQ-1*, *EIF-3*, *AMA-1*, *MDH-1*, *gpd-2*, *Act-1*, and *Act-2*). However, four of them (*tba-1*, *cdc-42*, *gpd-2*, and *act-2*) were eliminated because they had no expression values for at least one experimental condition in the RNA-Seq experiment (data not shown). Four contigs from the analysis of the RNA-Seq database (*aaRS*, *GR*, *mce1*, and *ArgRS*) were also selected as candidate reference genes because their expression levels were stable across all experimental conditions. Specific primers were designed for these genes (Table 2.1).

Expression stability analysis of candidate reference genes in *G. rostochiensis*

According to the distributions of raw Cq values, the genes with the lowest ranges were *GR*, *AMA-1*, *MDH-1*, and *aaRS*. The genes *EIF-3* and *Act-1* were the worst candidates, with Cq values spanning several units (Fig 2.1). However, a comparison of the distribution of the raw Cq values is not sufficient to evaluate the expression stability of candidate reference genes. Overall gene expression stability for both methods was assessed with RefFinder (Fig 2.2). *GR*, *PMP-3*, and *aaRS* expression were found to be the most stable, while *EIF-3*, *Act-1*, and *CSQ-1* showed the highest variability across treatments. Details of this ranking based on individual algorithms are given in Table 2.2. All four methods gave a roughly similar ranking except

BestKeeper with the qRT-PCR data. The genes *GR*, *PMP-3*, and *aaRS* were rated among the four most stable genes across both methods.

Figure 2.1: Variability of the expression level (span of Cq values) in *G. rostochiensis* of the 11 candidate reference genes across nine combinations of development stages and time of exposure to potato root diffusate, as measured by qRT-PCR. The median of three replicates of the nine conditions is represented by the line inside the box, 50% of the values are inside the box, the upper and lower edges represent the upper and lower quartile, respectively, and the circles represent outliers (< Q1 – (1.5 x IQR); > Q3 + (1.5 x IQR)).

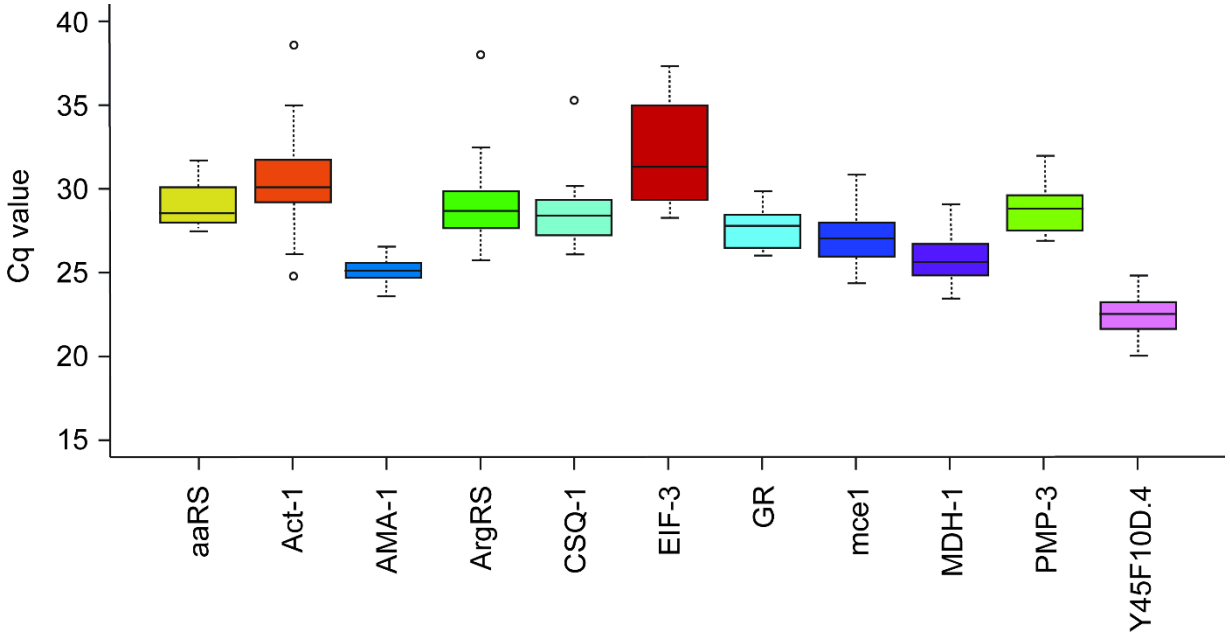


Figure 2.2: Summation of the comprehensive ranking values (variation mean) of gene stability as calculated by RefFinder for RNA-Seq and RT-qPCR variation analyses for 11 candidate reference genes. Expression variations were calculated on two RNA-Seq replicates and three qRT-PCR replicates for the eight developmental stages of *G. rostochiensis*.

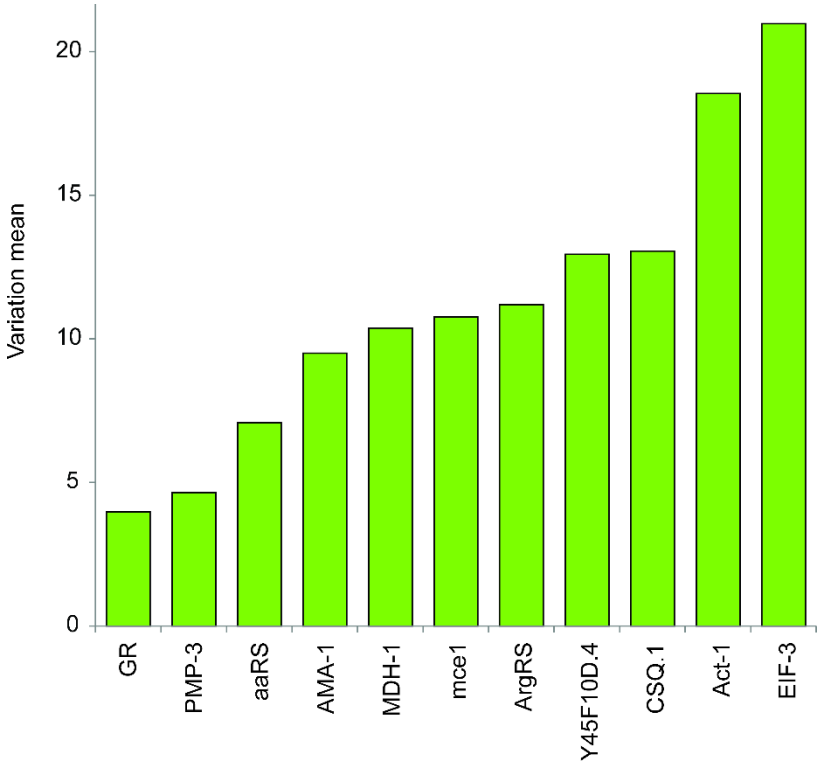


Tableau 2.2: Comparison of rankings of 11 candidate reference genes from RNA-Seq and qRT-PCR data, given by RefFinder in *G. rostochiensis*.

Method	ΔCt	BestKeeper	NormFinder	geNorm	Comprehensive ranking	Ranking
RNA-seq	*GR*	*ArgRS*	*GR*	*ArgRS* | *PMP-3*	*GR*	1
	aaRS	*PMP-3*	*mce1*		*ArgRS*	2
	PMP-3	*aaRS*	*AMA-1*	*aaRS*	*PMP-3*	3
	ArgRS	*GR*	*aaRS*	*GR*	*aaRS*	4
	mce1	*MDH-1*	*PMP-3*	*MDH-1*	*mce1*	5
	MDH-1	*mce1*	*ArgRS*	*mce1*	*AMA-1*	6
	AMA-1	*AMA-1*	*MDH-1*	*AMA-1*	*MDH-1*	7
	CSQ-1	*Y45F10D.4*	*CSQ-1*	*Y45F10D.4*	*CSQ-1*	8
	Y45F10D.4	*CSQ-1*	*Y45F10D.4*	*CSQ-1*	*Y45F10D.4*	9
	Act-1	*Act-1*	*Act-1*	*Act-1*	*Act-1*	10
	EIF-3	*EIF-3*	*EIF-3*	*EIF-3*	*EIF-3*	11
qRT-PCR	*GR*	*Y45F10D.4*	*GR*	*aaRS* | *PMP-3*	*GR*	1
	PMP-3	*MDH-1*	*PMP-3*		*PMP-3*	2
	AMA-1	*GR*	*AMA-1*	*CSQ-1*	*AMA-1*	3
	CSQ-1	*Act-1*	*CSQ-1*	*AMA-1*	*aaRS*	4
	aaRS	*mce1*	*MDH-1*	*GR*	*Y45F10D.4*	5
	MDH-1	*AMA-1*	*aaRS*	*mce1*	*CSQ-1*	6
	mce1	*PMP-3*	*Y45F10D.4*	*Y45F10D.4*	*MDH-1*	7
	Y45F10D.4	*ArgRS*	*mce1*	*MDH-1*	*mce1*	8
	ArgRS	*CSQ-1*	*EIF-3*	*ArgRS*	*Act-1*	9
	EIF-3	*aaRS*	*ArgRS*	*EIF-3*	*ArgRS*	10
	Act-1	*EIF-3*	*Act-1*	*Act-1*	*EIF-3*	11

https://doi.org/10.1371/journal.pone.0193840.t002

Evaluation of reference genes in *G. pallida*

Candidate genes were also investigated in *G. pallida*, a closely related species to *G. rostochiensis*. Both PCN species share the same host plants and have very similar morphological characteristics. The qRT-PCR analysis was performed using the same experimental design and primers that were designed for *G. rostochiensis*. The results for *G. pallida* were similar to those obtained for *G. rostochiensis*. Details of this ranking based on individual algorithms are given in Table 2.3. Based on the qRT-PCR analysis, *AMA-1*, *GR*, and *PMP-3* were the most stable potential reference genes for expression analysis in *G. pallida*. The results are similar to those obtained for *G. rostochiensis* as we find nearly the same genes in the first, second and last tier of the ranking.

Tableau 2.3: Detailed rankings of 11 candidate reference genes from qRT-PCR data, given by RefFinder in *G. pallida*.

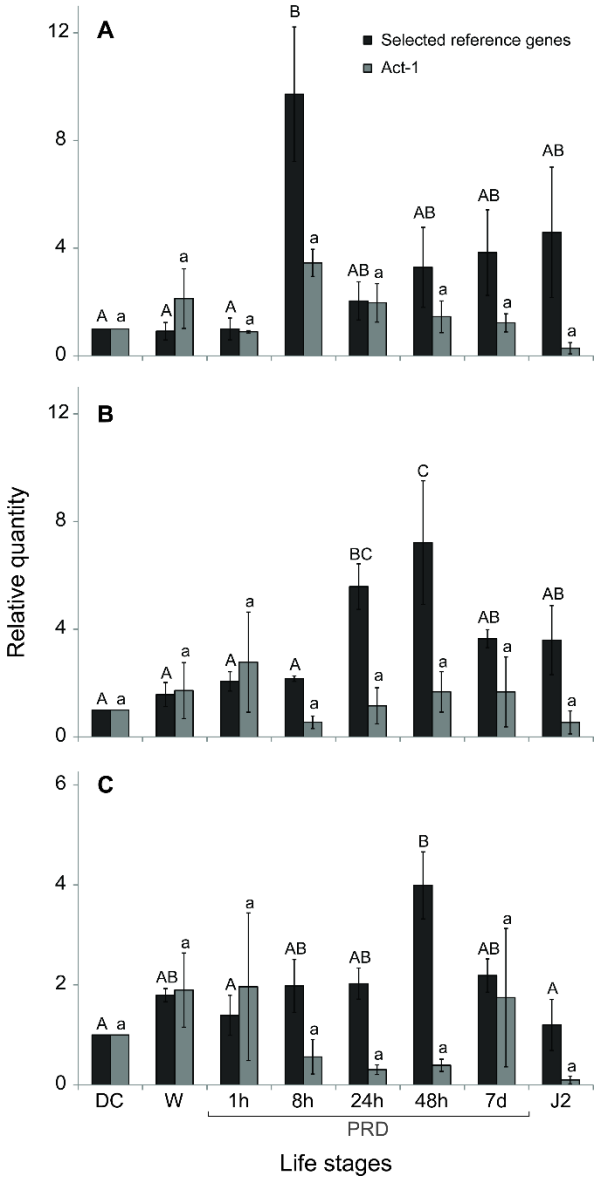
ΔCt	BestKeeper	NormFinder	geNorm	Comprehensive ranking	Ranking
AMA-1	*PMP-3*	*GR*	*AMA-1*	*AMA-1*	**1**
GR	*aaRS*	*AMA-1*	*mce1*	*GR*	**2**
mce1	*GR*	*aaRS*	*GR*	*PMP-3*	**3**
aaRS	*Y45F10D.4*	*PMP-3*	*Y45F10D.4*	*mce1*	**4**
Y45F10D.4	*EIF-3*	*mce1*	*MDH-1*	*aaRS*	**5**
PMP-3	*AMA-1*	*Y45F10D.4*	*aaRS*	*Y45F10D.4*	**6**
MDH-1	*mce1*	*MDH-1*	*PMP-3*	*MDH-1*	**7**
EIF-3	*CSQ-1*	*EIF-3*	*EIF-3*	*EIF-3*	**8**
CSQ-1	*MDH-1*	*CSQ-1*	*CSQ-1*	*CSQ-1*	**9**
Act-1	*Act-1*	*Act-1*	*Act-1*	*Act-1*	**10**

https://doi.org/10.1371/journal.pone.0193840.t003

Validation of reference genes

Validation of these recommended reference genes was performed using three genes (*NEP-1*, *cht-2*, and *eng*) with known patterns of expression. A qRT-PCR analysis was performed across the life stages of *G. rostochiensis* for each of these genes using the selected reference genes (*aaRS*, *PMP-3*, and *GR*) to normalize the data. The gene *NEP-1* was found to be overexpressed nearly 10 times after 8 h of exposure to PRD, *cht-2* was overexpressed 6 to 7 times after 24 to 48 h of exposure to PRD, and *eng* was overexpressed nearly 4 times after 48 h of exposure to PRD (Fig 2.3). In comparison, when normalization was performed using the *Act-1* gene, no significant difference was found in gene expression.

Figure 2.3: Expression of (A) *NEP-1*, (B) *cht-2*, and (C) *eng* assessed by qRT-PCR in dry cysts (DC), water hydrated eggs (W), hydrated eggs exposed to potato root diffusate (PRD) for 1 h, 8 h, 24 h, 48 h, and 7 d, and in J2 larvae. All expression levels were normalized using the geometric mean of *aaRS*, *PMP-3*, and *GR, our selected reference genes (in black) and Act-1 (in grey)*. Dry cyst was used as the calibrator for relative expression calculation (=1). Error bars represent the standard error of the mean, and significant differences among treatments are indicated by different letters (Tukey's test).



Discussion

Even though the PCNs, *G. rostochiensis* and *G. pallida*, are major threat to agriculture worldwide, many aspects of the infection process remain largely unknown, and several plant-nematode genomics analyses are currently underway. However, a set of reliable reference genes for the normalization of gene expression in qRT-PCR studies is still lacking. A perfect reference gene should have a constant expression and be unaffected by the experimental treatments. Because this is rarely met in practice, the use of at least three reference genes is generally recommended in order to obtain reliable expression data (Derveaux et al., 2010; Huggett et al., 2005). Previous studies often used common housekeeping genes as reference genes. However, it is now recognized that some of them are not suitable owing to their variability (Bustin & Nolan, 2004; Dheda et al., 2005; Huggett et al., 2005). In this study, we evaluated seven reference genes frequently used for *C. elegans* in addition to four other candidates selected from RNA-Seq expression data across a range of *G. rostochiensis* development stages and times of exposure to root exudates. This study demonstrates the feasibility of using RNA-seq expression data for the selection of RT-qPCR reference gene, as well as, to our knowledge, the first validation of reference gene published for PCNs. This work was carried out in parallel with an in-depth characterization of gene expression during hatching in *G. rostochiensis* which led to the identification of the NEP-1 gene in Duceppe et al. (2017). The lack of good reference genes adapted to cyst nematodes was hampering this kind of expression studies and although these genes were already used in Duceppe et al. (2017b), they need to be formally validated with recognized methods and on more genes.

Using RefFinder, the genes *GR* (glutathione reductase), *PMP-3* (putative ABC transporter), and *aaRS* (aminoacyl tRNA synthetase) were found to be the most stable among all the candidates. Therefore, they are the best available to be used as reference genes for normalization in qRT-PCR gene expression experiments for *G. rostochiensis* within the same experimental condition. Two of the three recommended reference genes (*GR* and *aaRS*) were identified based on RNA-Seq expression data analysis. This shows that RNA-Seq is an efficient approach to identify more stable and reliable reference genes and that public repositories like the NCBI Short Read Archive should be more exploited. The third gene, *PMP-3*, had already been selected as a reliable reference gene for *C. elegans* in previous studies (Hoogewijs et al., 2008; Zhang et al.,

2012). Our results showed that this gene was also appropriate in *G. rostochiensis*, and may be a good candidate overall since its expression was found to be very stable in experimental conditions that differed noticeably from those generally found in most *C. elegans* studies. RNA-Seq data mining was found to be a very useful strategy to identify new reference genes from a much larger pool of candidates than qRT-PCR.

The *PMP-3* gene encodes a peroxisomal membrane protein putative ABC transporter (Li et al., 2004), the *GR* gene encodes a glutathione reductase that is involved in oxidative stresses regulation (Mannervik, 1987), and the *aaRS* gene encodes an aminoacyl tRNA synthetase that is part of RNA translation (Consortium, 1998). Since none of these recommended reference genes belong to the same metabolic pathway, their combination decreases the probability that an experimental factor would affect the expression of all three reference genes, and therefore support their common use as a reference gene set.

The rankings given by the four calculation methods (ΔCt, BestKeeper, NormFinder, and geNorm) are slightly different because they rely on different algorithms and calculation approaches (Table 2.2). RefFinder proposes a comprehensive ranking that considers calculations from the four methods and was used here to select the reference genes. Some candidate reference genes that had been ranked at the top of the RNA-Seq analysis were not as stable when tested using qRT-PCR. This difference could be due to the sample preparation, in addition to the multiple bioinformatics steps required before the expression analysis could take place. Some genes that are routinely used as reference genes for other nematode species in the literature (e.g., *EIF-3* and *Act-1 (Espinola et al., 2014; Kozera & Rapacz, 2013; Radonić et al., 2004)*) were less stable than those that were found as the best potential reference genes in the present study. This suggests that different morphological and physiological stages of *G. rostochiensis* may influence the expression of these usual reference genes. This influence would not be surprising, considering the huge reshuffling of gene expression across the different developmental stages, spanning from dormant cyst to infective juvenile larvae (Cotton et al., 2014; Duceppe et al., 2017a; Eves-Van Den Akker et al., 2016).

Very similar results were found with the sister species, *G. pallida*, for the ranking of reference gene candidates according to their stability. This similarity was expected, given the close phylogenetic relationship between *G. rostochiensis* and *G. pallida*. Using the same

primers (for two out of three reference genes) for these two species would also increase the practicality of this set of reference genes.

To validate the set of reference genes, we used them to normalize the expression of the NEP-1, cht-2, and eng genes, whose expression is previously known, in multiple life stages of *G. rostochiensis.* The results were consistent with those previously reported. Duceppe et al. (Duceppe et al., 2017a) showed that NEP-1 was significantly overexpressed 6.7 times after 8h of exposure to PRD; cht-2 transcripts were 4.9 and 5.3 times more abundant 24 and 48h after exposure to PRD respectively and eng showed a 4.4 fold change after 48h in PRD. These data are very similar to our RT-qPCR results after normalization with the proposed set of reference genes (Fig. 2.3). In comparison, data normalization using the Act-1 gene, revealed no significant difference in gene expression according to a Tukey's test (Fig. 2.3). This result is probably explained by the high variation in the Act-1 Cq values across the treatments (Fig. 2.1) confirming that this gene is not stable enough to be used as reference. This was also confirmed in other organisms, for example expression levels of actin were found to be highly affected by biotic or abiotic stresses in potato (Nicot et al., 2005). These observations further support the efficiency and purpose of our selected set of reference genes. The NEP-1 gene, which is involved in the degradation of peptides and in post-transcriptional modification, was also found to be overexpressed prior to hatching in *C. elegans* (Spanier et al., 2005). The cht-2 gene, codes for a hydrolytic enzyme that degrades chitin. In plant-parasitic nematodes, chitin has been found only in the eggshell and was found to be overexpressed during hatching (Schwekendiek et al., 1999). The eng gene codes for beta-endoglucanase, a polysaccharide-degrading enzyme. That gene was also found to be expressed prior to hatching in *Globodera tabacum*, a closely related species (Goellner et al., 2000). In the present study, the normalized expression of these three genes was consistent with expectations from the previously reported studies. This confirms the efficiency of the selected reference genes.

Conclusion

In this work, we showed that the expression of the genes *GR*, *PMP-3*, and *aaRS* was stable in all tested stages of the PCNs nematodes life cycle and duration of exposition of hydrated cysts

in potato root exudates. We, therefore, recommend their use as reference genes in qRT-PCR analysis of PCNs nematodes. It is important however to remind that some life stages – J3 and J4 – were not tested in this work. Researchers willing to use these reference genes for sexual differentiation study, for example, should validate their stability in their experimental conditions. This study also demonstrated the benefit of using RNA-Seq expression data for the identification of novel candidate reference genes for qRT-PCR analyses as well as the concordance of the RNA-seq and RT-qPCR expression data.

Acknowledgements

The authors thank Guy Bélair and Éric Grenier for providing isolates of potato cyst nematodes.

Authors' contributions

Conceived and designed the experiments: MS, MOD, MSA, BM. Performed the experiments: MS, MOD. Analyzed the data: MS. Contributed reagents/materials/analysis tools: BM. Wrote the paper: MS, MOD, MSA, BM.

References

Andersen, C. L., Jensen, J. L., & Orntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research, 64*(15), 5245-5250. doi:10.1158/0008-5472.CAN-04-0496

Bélair, G. (2005). Les nématodes, ces anguillules qui font suer les plantes... par la racine. *Phytoprotection, 86*(1), 65-69. doi:10.7202/011717ar

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2009). GenBank. *Nucleic Acids Research, 37*(Database issue), D26-31. doi:10.1093/nar/gkn723

Bustin, S. A., & Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of Biomolecular Techniques, 15*(3), 155-166.

Castro-Quezada, P., Aarrouf, J., Claverie, M., Favery, B., Mugniéry, D., Lefebvre, V., & Caromel, B. (2013). Identification of Reference Genes for Normalizing RNA Expression in Potato Roots Infected with Cyst Nematodes. *Plant Molecular Biology Reporter, 31*(4), 936-945. doi:10.1007/s11105-013-0566-3

Chauvin, L., Caromel, B., Kerlan, M. C., Rulliat, E., Fournet, S., Chauvin, J. É., Grenier, E., Ellissèche, D., & Mugniéry, D. (2008). Control of potato cyst nematodes *Globodera rostochiensis* and *Globodera pallida*. *Cahiers Agricultures, 17*(4), 368-374. doi:10.1684/agr.2008.0217

Consortium, T. C. e. S. (1998). Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science, 282*(5396), 2012-2018. doi:10.1126/science.282.5396.2012

Cotton, J. A., Lilley, C. J., Jones, L. M., Kikuchi, T., Reid, A. J., Thorpe, P., Tsai, I. J., Beasley, H., Blok, V., Cock, P. J., Eves-van den Akker, S., Holroyd, N., Hunt, M., Mantelin, S., Naghra, H., Pain, A., Palomares-Rius, J. E., Zarowiecki, M., Berriman, M., Jones, J. T., & Urwin, P. E. (2014). The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology, 15*(3), R43. doi:10.1186/gb-2014-15-3-r43

Den Nijs, L., & Karssen, G. (2004). *Globodera rostochiensis* and *Globodera pallida*. *EPPO Bulletin, 34*(2), 309-314. doi:10.1111/j.1365-2338.2004.00734.x

Derveaux, S., Vandesompele, J., & Hellemans, J. (2010). How to do successful gene expression analysis using real-time PCR. *Methods, 50*(4), 227-230. doi:10.1016/j.ymeth.2009.11.001

Dheda, K., Huggett, J. F., Chang, J. S., Kim, L. U., Bustin, S. A., Johnson, M. A., Rook, G. A., & Zumla, A. (2005). The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Analytical Biochemistry, 344*(1), 141-143. doi:10.1016/j.ab.2005.05.022

Duceppe, M.-O., Lafond-Lapalme, J., Palomares-Rius, J. E., Sabeh, M., Blok, V., Moffett, P., & Mimee, B. (2017a). Analysis of survival and hatching transcriptomes from potato cyst nematodes, *Globodera rostochiensis* and *G. pallida*. *Scientific reports, 7*(1), 3882. doi:10.1038/s41598-017-03871-x

Espinola, S. M., Ferreira, H. B., & Zaha, A. (2014). Validation of suitable reference genes for expression normalization in *Echinococcus* spp. larval stages. *PLOS ONE, 9*(7), e102228. doi:10.1371/journal.pone.0102228

Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G., Da Rocha, M., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szitenberg, A., Grenier, E., Montarry,

J., Mimee, B., Duceppe, M. O., Boyes, I., Marvin, J. M., Jones, L. M., Yusup, H. B., Lafond-Lapalme, J., Esquibet, M., Sabeh, M., Rott, M., Overmars, H., Finkers-Tomczak, A., Smant, G., Koutsovoulos, G., Blok, V., Mantelin, S., Cock, P. J., Phillips, W., Henrissat, B., Urwin, P. E., Blaxter, M., & Jones, J. T. (2016). The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology, 17*(1), 124. doi:10.1186/s13059-016-0985-1

Food and Agriculture Organization. (2016). *FAO Statistical Pocketbook 2015*: The Food and Agriculture Organization of the United Nations.

Goellner, M., Smant, G., De Boer, J. M., Baum, T. J., & Davis, E. L. (2000). Isolation of Beta-1,4-Endoglucanase Genes from *Globodera tabacum* and their Expression During Parasitism. *Journal of Nematology, 32*(2), 154-165.

Hofmann, J., & Grundler, F. M. W. (2007). Identification of reference genes for qRT-PCR studies of gene expression in giant cells and syncytia induced in *Arabidopsis thaliana* by *Meloidogyne incognita* and *Heterodera schachtii. Nematology, 9*(3), 317-323. doi:10.1163/156854107781352034

Hoogewijs, D., Houthoofd, K., Matthijssens, F., Vandesompele, J., & Vanfleteren, J. R. (2008). Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in *C. elegans*. *BMC Molecular Biology, 9*, 9. doi:10.1186/1471-2199-9-9

Huggett, J., Dheda, K., Bustin, S., & Zumla, A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes and immunity, 6*(4), 279-284. doi:10.1038/sj.gene.6364190

Jones, J. T., Haegeman, A., Danchin, E. G., Gaur, H. S., Helder, J., Jones, M. G., Kikuchi, T., Manzanilla-Lopez, R., Palomares-Rius, J. E., Wesemael, W. M., & Perry, R. N. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology, 14*(9), 946-961. doi:10.1111/mpp.12057

Kozera, B., & Rapacz, M. (2013). Reference genes in real-time PCR. *Journal of applied genetics, 54*(4), 391-406. doi:10.1007/s13353-013-0173-x

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J. F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., & Vidal, M. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science, 303*(5657), 540-543. doi:10.1126/science.1091403

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods, 25*(4), 402-408. doi:10.1006/meth.2001.1262

Mannervik, B. (1987). The enzymes of glutathione metabolism: an overview. *Biochemical Society Transactions, 15*(4), 717-718. doi:10.1042/bst0150717

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science, 320*(5881), 1344-1349. doi:10.1126/science.1158441

Nicot, N., Hausman, J. F., Hoffmann, L., & Evers, D. (2005). Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany, 56*(421), 2907-2914. doi:10.1093/jxb/eri285

Pfaffl, M. W., Tichopad, A., Prgomet, C., & Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnology Letters, 26*(6), 509-515. doi:10.1023/b:bile.0000019559.84305.47

Radonić, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., & Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications, 313*(4), 856-862. doi:10.1016/j.bbrc.2003.11.177

Sabeh, M., Duceppe, M. O., St-Arnaud, M., & Mimee, B. (2018). Transcriptome-wide selection of a reliable set of reference genes for gene expression studies in potato cyst nematodes (*Globodera* spp.). *PLOS ONE, 13*(3), e0193840. doi:10.1371/journal.pone.0193840

Schwekendiek, A., Maier, T., Womack, C., Byrne, D., Boer, J. d., Davis, E., & Baum, T. (1999). Initial characterisation of endochitinase genes of the plant-parasitic nematode *Heterodera glycines*. *Journal of Nematology, 31*, 568.

Silver, N., Best, S., Jiang, J., & Thein, S. L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology, 7*, 33. doi:10.1186/1471-2199-7-33

Spanier, B., Stürzenbaum, S. R., Holden-Dye, L. M., & Baumeister, R. (2005). *Caenorhabditis elegans* Neprilysin NEP-1: an Effector of Locomotion and Pharyngeal Pumping. *Journal of Molecular Biology, 352*(2), 429-437. doi:10.1016/j.jmb.2005.06.063

Taki, F. A., & Zhang, B. (2013). Determination of reliable reference genes for multi-generational gene expression analysis on *C. elegans* exposed to abused drug nicotine. *Psychopharmacology, 230*(1), 77-88. doi:10.1007/s00213-013-3139-0

Turner, S. J. (1996). Population decline of potato cyst nematodes (*Globodera rostochiensis*, *G. pallida*) in field soils in Northern Ireland. *Annals of Applied Biology, 129*(2), 315-322. doi:10.1111/j.1744-7348.1996.tb05754.x

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology, 3*(7), Research0034. doi:10.1186/gb-2002-3-7-research0034

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics, 10*(1), 57-63. doi:10.1038/nrg2484

Xie, F., Xiao, P., Chen, D., Xu, L., & Zhang, B. (2012a). miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant molecular biology, 80*(1), 75-84. doi:10.1007/s11103-012-9885-2

Yu, Q., Ye, W., Sun, F., & Miller, S. (2010). Characterization of *Globodera rostochiensis* (*Tylenchida: Heteroderidae*) associated with potato in Quebec, Canada. *Canadian Journal of Plant Pathology, 32*(2), 264-271. doi:10.1080/07060661003740322

Zhang, Y., Chen, D., Smith, M. A., Zhang, B., & Pan, X. (2012). Selection of reliable reference genes in *Caenorhabditis elegans* for analysis of nanotoxicity. *PLOS ONE, 7*(3), e31849. doi:10.1371/journal.pone.0031849

Zhao, S., & Fernald, R. D. (2005). Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of computational biology, 12*(8), 1047-1064. doi:10.1089/cmb.2005.12.1047

CHAPITRE 3 : Qu'est-ce qui détermine la spécificité parasitaire chez un nématode phytopathogène hyperspécialisé?

Ce chapitre a été publié dans la revue *BMC Genomics*.

Sabeh M, Lord E, Grenier E, St-Arnaud M, Mimee B (2019) What determines host specificity in hyperspecialized plant parasitic nematodes? BMC Genomics 20:457.

Résumé

Chez les nématodes phytopathogènes hyperspécialisés, la capacité de parasiter un hôte dépend de facteurs de virulence spécifiques, appelés effecteurs. Ces protéines excrétées sont impliquées dans les mécanismes moléculaires du parasitisme et distinguent les nématodes virulents des espèces proches non virulentes. Les nématodes à kyste de la pomme de terre (*Globodera rostochiensis* et *G. pallida*) sont des nématodes phytoparasites parasitant de nombreuses espèces de solanacées, notamment la pomme de terre. Des espèces proches, *G. tabacum* et *G. mexicana*, sont stimulées par l'exsudat racinaire de pommes de terre, mais sont incapables de parasiter cette plante. Le but de l'étude est l'identification des gènes associés à la spécificité de l'hôte, en utilisant le séquençage d'ARN pour caractériser les différences transcriptomiques entre ces quatre espèces. Nous avons identifié sept transcrits uniques aux nématodes à kyste de la pomme de terre, y compris une protéine impliquée dans l'ubiquitination. Nous avons également trouvé 545 gènes différentiellement exprimés, y compris 78 gènes codant pour des protéines effectrices, ce qui représente un enrichissement plus de 6 fois supérieur à ce qui est retrouvé dans le transcriptome. L'analyse du polymorphisme des gènes a permis d'identifier 359 variants non synonymes homozygotes ce qui représente une preuve solide de sélection chez les nématodes à kyste de la pomme de terre. Nous avons pu démontrer que le déterminant de la spécificité de l'hôte réside dans la régulation de l'expression de gènes effecteurs essentiels, qui pourrait être sous le contrôle d'un seul ou de très peu de gènes régulateurs. De tels gènes pourront être des cibles prometteuses pour le développement de nouvelles sources de résistance durables contre les nématodes à kyste de la pomme de terre.

What determines host specificity in hyperspecialized plant parasitic nematode?

Michael Sabeh[1,2], Etienne Lord[1], Éric Grenier[3], Marc St-Arnaud[2], Benjamin Mimee[1]

[1]St-Jean-sur-Richelieu Research and Development Center, Agriculture and Agri-Food Canada, St-Jean-sur-Richelieu, Québec, Canada, [2]Biodiversity Center, Institut de recherche en biologie végétale, Université de Montréal and Jardin botanique de Montréal, Montreal, Québec, Canada, [3]INRA, UMR1349 IGEPP (Institute of Genetics, Environment and Plant Protection), F-35653 Le Rheu, France

Abstract

In hyperspecialized nematode parasites, the ability to exploit a particular host relies on specific virulence factors called effectors. These proteins excreted by the parasite are involved in the molecular mechanisms of parasitism and distinguish virulent pathogens from non-virulent related species. The potato cyst nematodes (PCN) *Globodera rostochiensis* and *G. pallida* are major plant-parasitic nematodes of numerous solanaceous plant species including potato. Their close relatives, *G. tabacum* and *G. mexicana* hatch when exposed to potato root diffusates but are unable to establish a feeding site on this plant host. This study aims at identifying genes involved in host specificity, using RNA sequencing to characterize transcriptomic differences among these four *Globodera* species. We identified seven transcripts unique to PCN species, including a protein involved in ubiquitination. We also found 545 genes that were differentially expressed between PCN and non-PCN species, including 78 genes coding for effector proteins, which represent more than a 6-fold enrichment compared to the whole transcriptome. Gene polymorphism analyses identified 359 homozygous non-synonymous variants showing strong evidence for selection in PCN species. Overall, we demonstrated that the determinant of host specificity resides in the regulation of essential effector genes expression that could be under the control of a single or of very few regulatory genes. Such genes are therefore promising targets for the development of novel and more sustainable resistances against potato cyst nematodes.

Background

Nematodes are a very diverse phylum of animals living in a wide range of environments. Most of them feed on microorganisms or organic matter detritus in a free-living mode of existence. However, some species have diverged towards a parasitic lifestyle on higher organisms such as plants and animals, often in complex obligate associations. This transition to parasitism has followed morphological adaptations, but also the acquisition of genes coding for excreted proteins giving them the ability to feed and survive on their host (Blaxter & Koutsovoulos, 2015). Different human- and animal-parasitic nematodes have been studied extensively but despite the importance of plant-parasite nematodes (PPN), many aspects of the infection process remain largely unknown (Bird et al., 2003). PPNs are plant parasites of great importance and represent a significant constraint on global food production causing yield losses estimated at $157 billion every year (Abad et al., 2008). Over 4,100 species of PPNs have been described to date exploiting all major cultivated crops in the world, the most damaging being the *Heteroderidae* including *Meloidogyne* spp., *Globodera* spp., and *Heterodera* spp. (Bird et al., 2003; Decraemer & Hunt, 2006). It is now generally accepted that after the development of an anatomical structure used for plant cell wall puncturing and nutrient uptake called the stylet, the ability of nematodes to parasitize plants was facilitated by the acquisition of bacterial genes through horizontal genes transfer (e.g. cellulases, pectate lyases, xylanases, galactosidases, and expansin-like proteins) (Blaxter & Koutsovoulos, 2015; Danchin et al., 2010b; Haegeman et al., 2011; Paganini et al., 2012; Quist et al., 2015; Smant et al., 2018). Host penetration, the establishment of a feeding site and suppression of host defenses are key steps of the infection process and are highly dependent on their set of specific secreted proteins called effectors, used by PPNs to manipulate the host to their benefit (Eves-Van Den Akker & Birch, 2016). Effectors are responsible for most of the interactions with the host and given their importance in the infection mechanisms, substantial research efforts have been directed to these molecules in recent years (Varden et al., 2017). Effectors are also typically involved in evolutionary arms races between plants and parasites (Anderson et al., 2010b). Although different plant-parasitic nematodes have a common arsenal of effectors, it is not yet known exactly how this arsenal differs among species and especially among closely related species showing different host ranges (Danchin et al., 2010b). It was previously shown that the sequence polymorphism of the

pectate lyase 2 effector among two potato cyst nematode species and one tobacco cyst nematode species can be associated to some extent to host specificity (Geric Stare et al., 2011b). Also, at the intraspecific level, the ability to grow on potato cultivars harboring the H_1 resistance gene was concordant with polymorphism in two effector genes (putative cellulose binding protein and 3H07 ubiquitin extension) found in virulent *Globodera rostochiensis* populations (Eves-Van Den Akker et al., 2016). Studies conducted on other plant parasites like fungi have also associated variants in effector genes with their host specificity, as shown by the specific non-synonymous variants in two genes that appeared to be crucial for *Zymoseptoria tritici* virulence on wheat (Poppe et al., 2015). It has also been shown that effector genes contain a greater proportion of non-synonymous mutations compared to other genes (Eves-Van Den Akker et al., 2016). To better understand this relationship, a formal comparison of genetic variation between closely related species having different host ranges would help to identify elements that are associated with host specificity.

The genus *Globodera* includes more than a dozen species parasitic to either *Solanaceae* or *Compositae* plants, which can be differentiated through their host range. All *Globodera* of Solanaceae species are parasitic on tomato and some solanaceaeous weeds but only *G. rostochiensis* (Wollenweber), *G. pallida* (Stone), and *G. ellingtonae* (Handoo, Carter, Skantar & Chitwood) are parasitic on potato. *G. tabacum* (Lownsbery & Lownsbery) is found in a dozen tobacco-producing countries (Alenda et al., 2014) and is parasitic on tobacco but not on potato, while *G. mexicana* (Campos-Vela) is mainly found in Mexico (Blanchard et al., 2005) and is parasitic on *Solanum nigrum* but not parasitic of either tobacco or potato.

Potato cyst nematodes (PCN), *G. rostochiensis* and *G. pallida*, are hyperspecialized plant-parasites, as they have a very limited number of potential hosts and can be considered as major agricultural threats as they are responsible for the loss of 9% of the world's potato production each year (Alenda et al., 2014; Blanchard et al., 2005). Originated in South America, PCN are now present in over 75 countries, where they are often considered as regulated quarantine organisms (Sobczak et al., 2011; Thorpe et al., 2014). The PCN infection process starts when the dormant eggs receive an appropriate chemical signal from the root diffusate of a potential host. The nematodes then hatch and move toward the roots. Using different secreted enzymes, they next enter the roots and transform a plant cell in the root inner cortex layers to establish a

complex feeding site, named syncytium, a highly metabolically active structure with enriched cytoplasm (Sobczak & Golinowski, 2011; Thorpe et al., 2014). The PCN induce a cascade of changes in host-gene expression, and cell fusion to form a syncytium. How the effectors cause the hypertrophy and endopolyploidization of feeding cells and their interplay with plant hormones is not yet fully understood (Haegeman et al., 2012; Smant et al., 2018). The important diversion of plant nutriments towards the nematodes limit plant growth and can eventually cause heavy yield loss. After maturation and fecundation, PCN females dry to form a cyst, a protective shell, containing up to 300 eggs able to survive more than 20 years in soil (Den Nijs & Karssen, 2004; Turner, 1996).

As *G. tabacum* shares a high level of genetic similarity with *G. rostochiensis* (Geric Stare et al., 2011b; Madani et al., 2008), and *G. mexicana* with *G. pallida* (Bossis & Mugniéry, 1993; Grenier et al., 2002), these species are therefore suitable candidates for a transcriptomic comparison analysis according to their parasitic status on potato plants. A transcriptomic study allows reduction of genomic complexity by sequencing only coding regions, in which a large proportion of significant functional variants are expected, also allowing identification of genes whose expression or allelic frequency can be correlated to a specific trait (Konczal et al., 2014). Other studies have focused on the discovery of effectors in plant-parasitic nematodes, including PCNs (Ali et al., 2015a; Ali et al., 2015b; Eves-Van Den Akker et al., 2014; Jones et al., 2009; Thorpe, 2012), as well as on the transcriptomic study of different pathotypes (Eves-Van Den Akker et al., 2016) or life stages (Duceppe et al., 2017b; Palomares-Rius et al., 2016) of *G. rostochiensis* and *G. pallida*, but to our knowledge, a direct transcriptomic comparison of pathogenic and non-pathogenic *Globodera* species on potato has never been done. The aims of this study were to characterize the transcriptomic differences between four *Globodera* species exposed to potato root diffusate and to identify genes putatively involved in host specificity using RNA sequencing to look at changes in gene expression and genetic variation between populations.

Materials and Methods

Globodera populations

In this study, eight populations representing four species were compared. Two PCN species, *G. rostochiensis* (populations St-Amable and Netherland) and *G. pallida* (populations Chavornay and Noirmoutier) were compared to two non-PCN species *G. tabacum* (populations 75181 and GV1) and *G. mexicana* (populations Tlaxcala and GM5) (Table 3.1). Potato plants cv. Snowden and tomato cv. MoneyMaker were grown in perlite, in 2 L containers, until they reached about 15 cm height. Root diffusate was harvested once a week, for six consecutive weeks, by the method of Fenwick (Fenwick, 1949). Perlite was drenched with tap water until saturation and the flowing liquid was collected. The procedure was repeated two more times and the total collected liquid was homogenized and filtered using milk filters (D-547, KenAG). Root diffusates were kept at 4 °C in dark plastic containers until use (<2 months). Three hundred cysts of each population were immerged in filtered distilled water (0.2 µm Nalgene 25 mm syringe filters, Thermo Scientific) for one week and then in filtered root diffusate (0.45 µm Nalgene 25 mm syringe filters, Thermo Scientific) for three additional weeks to induce hatching of second stage larvae (J2), used for the RNA extraction.

Tableau 3.1: *Globodera* populations and species used in this study.

Species	Potato host status	Population ID	Name and Pathotype	Origin
G. rostochiensis	Host	GrQC	St-Amable Ro1	Canada
		GrU1	Netherland Ro1	Netherlands
G. pallida	Host	GpA5	Pa3 Chavornay	Switzerland
		GpB1	Pa2/3 Noirmoutier	France
G. tabacum	Non-host	GtA1	*GV1*	United States
		GtA2	*75181*	Mexico
G. mexicana	Non-host	GmA1	Tlaxcala	Mexico
		GmA2	GM5	Mexico

RNA extraction and sequencing

Each sample was homogenized in 650 µl lysis buffer RLT Plus (Qiagen) with a 6 mm zirconium grinding bead and 200 µL of 1 mm zirconium beads in 2 ml tubes using the PowerLyzer 24 Homogenizer (Qiagen) and stored at −80 °C until RNA purification. Total RNA was extracted using RNeasy Mini Kit Plus (Qiagen) according to the manufacturer's instructions and stored at −80 °C. RNA was quantified, and its integrity assessed using a Bioanalyzer 2100 (Agilent Technologies) with the RNA 6000 Nano kit. All RNA samples had a RIN value ≥ 7. Libraries were generated using TruSeq Stranded mRNA Library Prep Kit (Illumina). Paired-end sequencing was done using the TruSeq SBS V3 2x125 bp chip on a HiSeq2500 sequencer (Illumina) at the McGill University and Genome Quebec Innovation Center in Montreal, Canada. All eight samples were multiplexed and sequenced on a single lane.

Sequences processing

Raw reads from all populations were trimmed using TRIMMOMATIC 0.36 (Bolger et al., 2014) with default parameters and were mapped to the *G. rostochiensis* transcriptome (assembly version nGr.v1.1) (Eves-Van Den Akker et al., 2016) using BWA-MEM 0.7.12 with default parameters (Li & Durbin, 2009). The *G. rostochiensis* transcriptome contains 14,309 putative genes. It was chosen for mapping and downstream analysis in order to work with a near complete transcriptome, to avoid contaminating sequences and because the genes providing the ability to grow on potato are theoretically included in it. To obtain up to date annotations for the reference transcriptome, we performed a conserved domain search using CD-SEARCH 3.16 with default parameters (Marchler-Bauer & Bryant, 2004). Predicted amino acid sequences were used as an input and were obtained using AUGUSTUS 3.3 (Stanke et al., 2008) with *Caenorhabditis elegans* as species parameter. In addition, sequence similarity search using NCBI (Geer et al., 2010a), KEGG (Kanehisa & Goto, 2000b), and UNIPROT (The Uniprot Consortium, 2017) databases were performed for unknown sequences of interest.

A phylogenetic analysis was performed using PHYLOGENY.FR (approximate likelihood ratio approach; bootstrap value = 100), a web-based wrapper-tool analysing phylogenetic relationships between molecular sequences (Dereeper et al., 2008), integrating MUSCLE 3.8.31 (Edgar, 2004), GBLOCKS 0.91b (Talavera & Castresana, 2007), PHYML 3.1/3.0 (substitution

model: HKY85) (Guindon et al., 2010), and TREEDYN 198.3 (Chevenet et al., 2006). The analysis was performed using the small subunit ribosomal RNA gene, a gene commonly used in nematode phylogenetic studies (Blaxter et al., 1998), and included a sequence from *C. elegans* for comparison (Accession number: NM_067514).

Quantitative analysis

Read counts for the statistical analysis was performed using CORSET 1.04 software with default parameters (Davidson & Oshlack, 2014b). Statistical analysis, including normalization and differently expressed genes (DEG) identification, was made using the DESEQ2 1.14.1 Bioconductor package in R (Love et al., 2014b). The eight populations were separated into two groups according to their host/non-host status on potato for DEG identification (GrQC, GrU1, GpA5, GpB1 vs GtA1, GtA2, GmA1, GmA2) using a parametric Wald test (DE; $P < 0.01$), a normalized minimum read count of 50 for all populations and a log2 fold change (log2FC) ≥ 1.

Variant analysis

Variant calling was done on all eight populations using FREEBAYES 1.0.2 software, a bayesian genetic variant detector designed to detect possible SNPs (single-nucleotide polymorphisms), indels (insertions and deletions), and complex events (Garrison & Marth, 2012a). Analysis was done using mapping files and the reference transcriptome as input with a minimum phred score of 30 and a minimum coverage of 10. BAYESCAN 2.1 (Foll & Gaggiotti, 2008) was then used to identify loci under natural selection, using allele frequencies as input and considering the pathogenic status on potato as the principal factor for selection. The method is based on locus-specific genetic differentiation (F_{ST}) outliers to detect candidate markers under selection. We relied on the "plot_bayescan" function in R provided with BAYESCAN to calculate a posterior odds threshold (FDR=0.05) and on a probability greater than 0.91, as this threshold indicates a strong evidence for selection (Jeffreys, 1998), to select outliers associated with the pathogenicity status of the population. Three analyzes were performed, giving different random initial seed values and only outliers present in all three analyses were kept.

The impact of these genetic variations on protein structure and cellular localization was evaluated to target the variants susceptible to lead to a difference in phenotype (Haegeman et

al., 2012; Mitchum et al., 2013b). SNPEFF 4.3 (Cingolani et al., 2012) was used to determine the impact (silent, missense or nonsenses) of the mutation while SIGNALP 4.1 (Petersen et al., 2011) and PHOBIUS (Kall et al., 2004b) were used to predict the presence of signal peptide cleavage sites and to determine the cellular localization of the proteins.

Validation by qPCR

Expression levels of genes of interest identified during the RNA-seq analysis were validated using qRT-PCR. Six candidate genes were chosen, based on their biological function: RBP-1 (Sequence ID: GROS_g14179.t1), putative effector SPRY domain-containing protein 19 (GROS_g14260.t1 and GROS_g14126.t1), pectate lyase 1 (GROS_g07968.t1), glutathione peroxidase (GROS_g02490.t1) and Peptidase M13 (GROS_g12349.t1). In addition, GR (glutathione reductase), PMP-3 (putative membrane transporter), and aaRS (aminoacyl tRNA synthetase) were used as a set of reference genes to normalized expression data (Sabeh et al., 2018). The transcription of these genes to mRNA was quantified in J2 larvae hatched after exposure to potato and tomato root diffusate, tomato plant is a compatible host for all the species under investigation and is used to determine if different root diffusate can affect the expression of effectors genes.

RNA extraction was performed as given above. First-strand cDNA was synthesized using SuperScript II reverse transcriptase (Invitrogen, Carlsbad, California, United States) according to the manufacturer's instructions, from 0.5 µg of total RNA and using oligo (dT)18. Three replicates were made for each treatment. Each sample was homogenized in supplied lysis buffer with a 6 mm zirconium grinding bead and 200 µL of 1 mm zirconium beads in 2 ml tubes using the PowerLyzer 24 Homogenizer (Qiagen) prior to extraction. Primers were designed using PrimerQuest tool (Integrated DNA Technologies, Inc., Coralville, Iowa, United States) based on the sequences retrieved from the *G. rostochiensis* transcriptome. Target fragments lengths were designed close to 100 bp. Primers information are listed in Table 3.2. Reactions were prepared using QuantiTect SYBR Green PCR kit (Qiagen) and amplified on a Mx3000P qPCR System (Agilent Technologies) for 45 cycles in a final volume of 25 µL according to the manufacturer's instructions. Melting curve analyses were done following the amplification cycles in order to examine the specificity of the reactions. Relative expression analysis was

performed using the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001). *G. rostochiensis* J2 larvae exposed to potato root diffusate was used as calibrator to calculate expression fold changes for all RNA samples.

Tableau 3.2: Primers information

Gene	SeqID	Sequence (5'3') forward/reverse
RBP-1	GROS_g14179.t1	GACGCCGTTTGCTTGTTCG / CTTTATTCTTGAGTTTGGTGT
SPRY domain-containing protein 19	GROS_g14260.t1	GCATTGATGGAAAGACGACAAC / GTTGCTGGTGGTTCTGATACT
SPRY domain-containing protein 19	GROS_g14126.t1	CGCGCCAAACAACAGTTAAT / GCATTTGTTCGGTCGCAAG
Pectase lyase 1	GROS_g07968.t1	GCTACTGGGTTCGGATACAA / GGCCAGATTGCGTGAAATAC
Glutathione peroxidase	GROS_g02490.t1	TCTACGACTTTGAGGTGGAAAC / GAAACGGGTTGAAGTCCAGATA
Neprilisin	GROS_g12349.t1	AATCACGCCGCCAAAGAA / CCAATGATGAGAGTGGTCGTAAA
Unknown	GROS_g09749.t1	CGCCCATCCCATTAGTGTT / CAACGACAAATCATGTTCTCCTC
Unknown	GROS_g10809.t1	AAATTCCGGTCGGCTCCT / TATTCCACGAACGGCTCCA
Polyubiquitin-B-like	GROS_g11284.t1	GCGACTGATCTTTAATGGGAAAC / CATCCTCCACGAAGACAAAGA
Unknown	GROS_g12023.t1	CGAATTGCCGGATGTTCTTG / CGTGTCAATTCGGTCGTAGAA
Unknown	GROS_g13375.t1	CGAGATGGTGTGATCAAGAAGA / TGACTGCGAGTTCGATTGG
Unknown	GROS_g13474.t1	CAGACAACACAGCACAACTTC / CTGAATCCCGGTCCTTGAAT
Unknown	GROS_g13669.t1	TTACGACTCCGCAAGTGTTC / TTGACTGCGGCGATTTCA

DNA from each species was also used to confirm the presence, in their respective genome, of seven genes for which transcripts were only observed in PCN species in the RNA-seq data. DNA extraction was performed on dry cyst using DNeasy Blood & Tissue Kits (Qiagen) according to the manufacturer's instructions and stored at −20 °C. Each sample was homogenized in supplied lysis buffer with a 6 mm zirconium grinding bead and 200 µL of 1 mm zirconium beads in 2 ml tubes using the PowerLyzer 24 Homogenizer (Qiagen) prior to extraction. Primer design and qPCR amplification were performed as given above.

Results

Sequencing and mapping

Exposure to potato root diffusate successfully induced hatching of a similar proportion of J2 larvae for all samples (data not shown). RNA sequencing of eight *Globodera* populations yielded a total of 233 M paired-end reads (2 x 125 bp). A mean of 29 M reads per sample, spanning from 24.1 M to 36.2 M reads was obtained (Table 3.3). The percentage of reads that successfully mapped to the *G. rostochiensis* reference transcriptome was on average 79.4% for *G. rostochiensis*, 74.2% for *G. pallida*, 69.3% for G. tabacum and 56.1% for *G. mexicana*. Horizontal coverage (breadth of coverage) was similar for all populations with reads mapping to 95.9% to 98.9% of the reference transcripts (Table 3.3). The phylogenetic analysis, performed using the small subunit ribosomal RNA gene, resulted in a greater genetic similarity between *G. rostochiensis* and *G. tabacum*, and between *G. pallida* and *G. mexicana* (Additional file: Fig 3.1).

Tableau 3.3: Sequencing yield and mapping statistics of all eight *Globodera* populations. Sequencing was done using HiSeq2500 sequencer, mapping and variant analyses were done using BWA-MEM and FREEBAYES softwares.

Sample ID	Sequenced Reads (M)	Mapped (%) *	Reference coverage (%) *	Total variants
GrQC	36.2	81.74	98.90	67 996
GrU1	28.1	77.02	98.42	68 778
GpA5	30.8	71.65	97.78	639 271
GpB1	24.1	76.73	97.26	608 293
GtA1	29.1	81.54	96.42	383 996
GtA2	27.5	57.01	95.92	365 750
GmA1	28.5	71.90	97.39	537 306
GmA2	28.2	40.25	96.75	500 035

* Reads were mapped on *Globodera rostochiensis* reference transcriptome (nGr.v1.1)

Quantitative analysis and differentially expressed genes identification

Seven transcripts were unique to *G. rostochiensis* and *G. pallida* and observe to be missing from the *G. mexicana* and *G. tabacum* transcriptomes (Table 3.4). Among these, six were coding for unknown proteins, and the remaining one (GROS_g11284.t1) was annotated as Polyubiquitin-B protein. To further investigate their functions, transcripts of unknown proteins were realigned to *G. rostochiensis* transcriptome to find similar sequences. Transcript GROS_g12023.t1, was found to be similar to GROS_13581.t1, which have a CHROMO domain (PFAM 00385), suggesting those genes could be implicated in the modification of the chromatin organization. In order to investigate if these transcripts correspond to missing genes in the *G. tabacum* and *G. mexicana* species, qPCR validation were performed on genomic DNA of the four *Globodera* species. Amplification products were obtained in all cases, except for amplification product corresponding to GROS_g11284.t1 and GROS_g09749.t1 were not detected for *G. mexicana* and amplification product corresponding to GROS_g12023.t1 was not detected for either *G. tabacum* and *G. mexicana*, suggesting in this last case a complete absence of the corresponding gene in the *G. mexicana* and *G. tabacum* genomes.

Tableau 3.4: Sequences expressed in PCN species (*G. rostochiensis* and *G. pallida*) and observe to be missing from non-PCN species (*G. tabacum* and *G. mexicana*) transcriptomes.

SeqID	Gene Description	Entry number (Organism) *
GROS_g09749.t1	Unknown	A0A183CCS7 (*Globodera pallida*)
GROS_g10809.t1	Unknown	A0A1I7SMH2 (*Bursaphelenchus xylophilus*)
GROS_g11284.t1	Polyubiquitin-B-like	A0A183CCZ8 (*Globodera pallida*)
GROS_g12023.t1	Unknown	A0A183C0B5 (*Globodera pallida*)
GROS_g13375.t1	Unknown	A0A183C870 (*Globodera pallida*)
GROS_g13474.t1	Unknown	A0A183BU61 (*Globodera pallida*)
GROS_g13669.t1	Unknown	A0A0K6FY64 (*Rhizoctonia solani*)

* Relates to UniProt database.

A total of 545 genes were found to be differentially expressed, 392 being up- and 153 down-regulated in PCN species (Fig 3.2; Additional file: Table 3.6). The most differentially expressed genes were coding for a SMC protein and a peptidase M13 with fold changes of 30.6 and 28.5.

Among DEGs, 216 were unknown genes, including 74 having a signal peptide for secretion. In the remaining 329 DEGs with known functions, 78 were known effector genes, 68 up- and 10 down-regulated in PCN populations (Fig 3.3). The fold change of these effector genes varied from 26.8 (putative effector SPRY domain-containing protein 19) to 2.2 (putative dorsal gland cell-specific expression protein), with a mean fold change of 6.6 for the 68 effectors genes up-regulated in PCN populations and 3.3 for the ten genes up-regulated in non-PCN populations. Fold change, as calculated by DESEQ2, is the average normalized reads counts of all populations of a group compared to the other (GrQC, GrU1, GpA5, GpB1 vs GtA1, GtA2, GmA1, GmA2).

Figure 3.2: Clustering of *Globodera* species based on expression value of 545 differentially expressed genes. Differential analysis was performed by group comparison according to their pathogenicity on potato; *G. rostochiensis* and *G. pallida* populations vs. *G. mexicana* and *G. tabacum* populations (GrQC, GrU1, GpA5, GpB1 vs GtA1, GtA2, GmA1, GmA2). Expression values are scores given by the pheatmap function (pheatmap 1.0.10 package in R) calculated using normalized read counts, as calculated by DESEQ2.

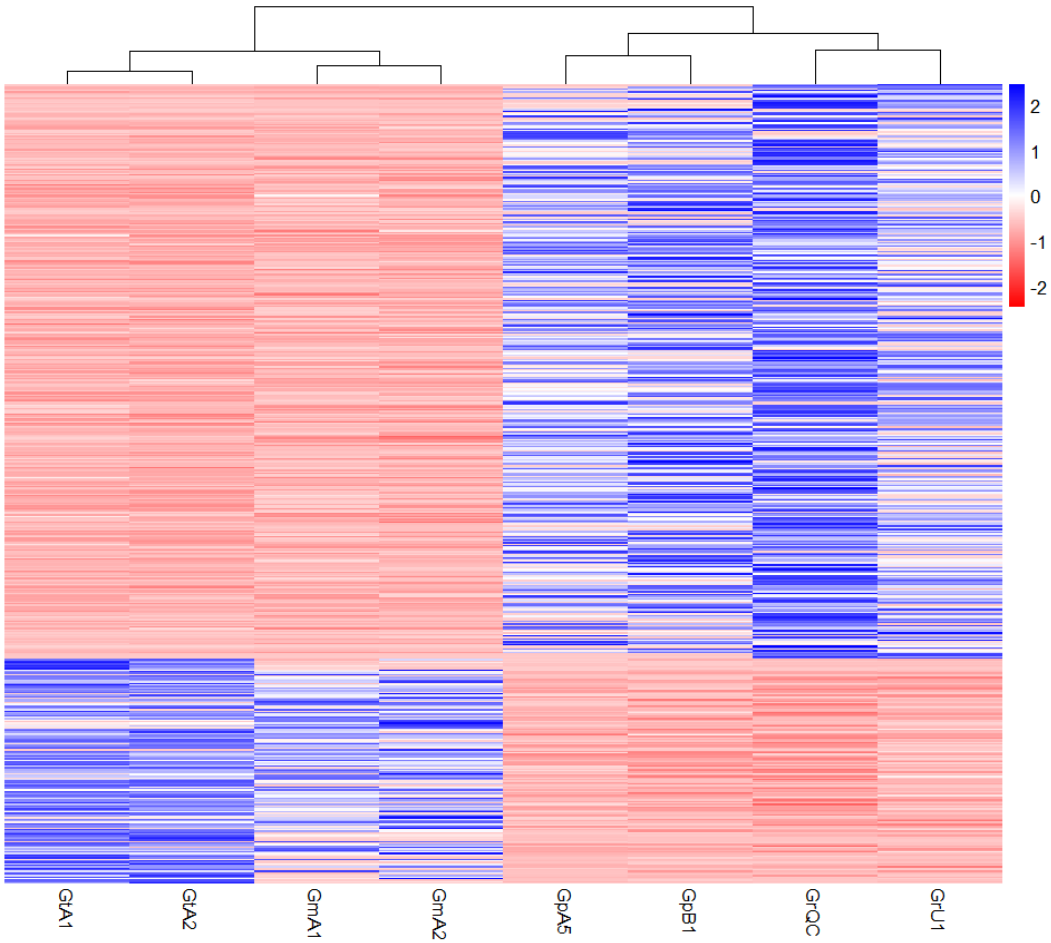
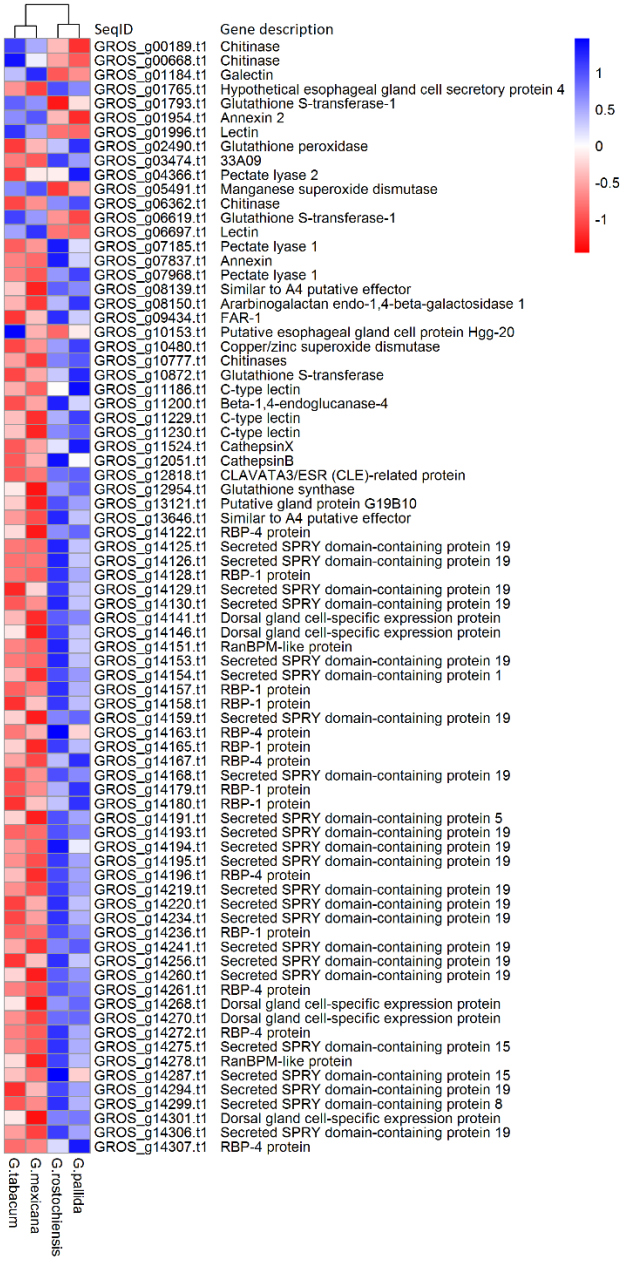


Figure 3.3: Differentially expressed effector genes between four *Globodera* species. Differential analysis was performed by group comparison according to their pathogenicity on potato; *G. rostochiensis* and *G. pallida* vs. *G. mexicana* and *G. tabacum*. Expression values are scores given by PHEATMAP function (PHEATMAP 1.0.10 package in R) calculated using mean normalized read counts of all populations for each species, as calculated by DESEQ2. SeqID and Gene descriptions correspond to sequences ID and annotation of the *G. rostochiensis* reference transcriptome (nGr.v1.1).



Variant analysis

The analysis of gene polymorphism between the samples identified 1,062,443 single nucleotide polymorphisms (SNPs), 63,455 insertions or deletions (Indels) and 21,161 complex events over 1,107,386 loci. As expected, *G. rostochiensis* had the least number of variants when compared to the reference transcriptome (6%), followed by *G. tabacum* (34%), *G. mexicana* (47%) and *G. pallida* (56%). Results were very similar between the different populations of each species. Using a Bayesian inference genome scan approach, we highlighted 1,181 genetic variants that were under selection in PCN species. Among them, 359 were homozygous non-synonymous variants in non-PCN populations (Additional materials: Table 3.7). Ten of these genes were coding for known effectors (Table 3.5) and 21 genes with unknown function contained a signal peptide for secretion. The effects of these gene variations were missense variants (325), frameshift (15), conservative inframe deletion (7), disruptive inframe deletion (4), disruptive inframe insertion (3), stop gained (2), conservative inframe insertion (2) and stop codon lost (1). However, despite the presence of a variant at the same position in all populations from the same group, these variants did not always have the same impact and therefore, no frameshift, stop gained or stop codon lost were shared by all non-PCN population. For example, contig GROS_g02285.t1 has a guanine duplication (134dupG) at position 134 causing a frameshift for all populations of *G. tabacum*, whereas populations of *G. mexicana* have a SNP at the same position (134G>A) only resulting in an amino acid modification.

Tableau 3.5: Homozygous non-synonymous variants in effector genes found only in non-PCN species (*G. mexicana* and *G. tabacum*). Variant analyses were done using FREEBAYES, BAYESCAN and SNPEFF softwares, with *G. rostochiensis* transcriptome as reference.

Gene description	SeqID	Pos[1]	Ref[2]	*G. mexicana*[3]	*G. tabacum*[3]
Chorismate mutase	GROS_g02441.t1	50 66 118	His Glu GluGlu	Pro Lys LysLys	Pro Lys LysLys
Skp1	GROS_g04817.t1	84	Ala	Val	Val
Ubiquitin carboxyl-terminal hydrolase	GROS_g05177.t1	810	GlyLeu	AlaLeu	GluMet
Pectate lyase 1	GROS_g07968.t1	212	Gly	Lys	Lys
PLP synthase	GROS_g08956.t1	280	Ala	Ser	Ser
Glutathione S-transferase	GROS_g10872.t1	64 1773	Pro MetTrpLysPro	Gln MetTrpLysSer	Leu Ser
β-1,4-endoglucanase	GROS_g11200.t1	386	Ile	Lys	Lys
Putative gland protein G19B10	GROS_g13121.t1	100	ArgLeu	HisLeu	SerSer
RBP-1	GROS_g14157.t1	90	Gly	Cys	Arg
RBP-1	GROS_g14180.t1	222	GluPhe	LysSer	LysPhe

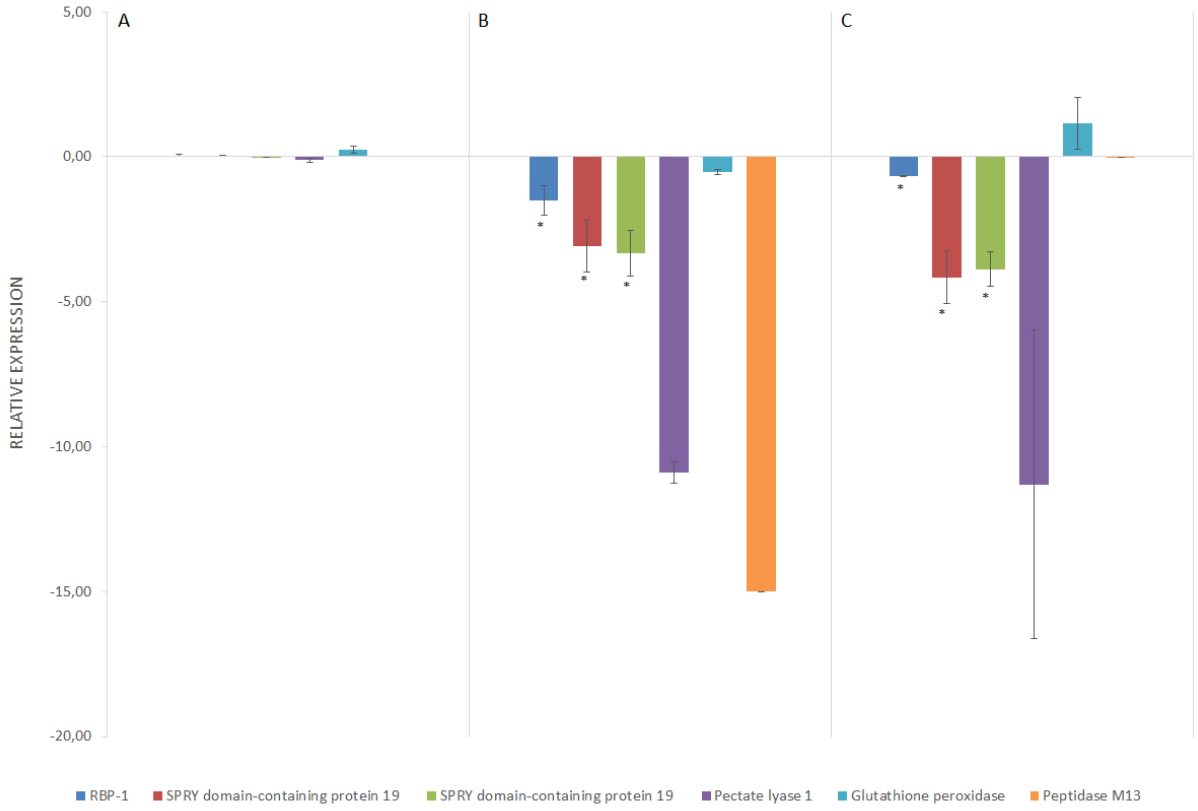
[1] Position refers to localisation in amino acid chain.
[2] Amino acid present in the reference.
[3] Amino acid predicted different from the *G. rostochiensis* reference transcriptome

qRT-PCR analysis

A subset of the differentially expressed effector genes was selected and tested by RT-qPCR to confirm their expression following exposure to root diffusate. These genes were selected to based on their biological function, to represent different steps of the infection process. Tested genes included RBP-1, two putative effector SPRY domain-containing protein 19, pectate lyase 1, glutathione peroxidase and Peptidase M13. When exposed to potato root diffusate, no significant difference was observed in the expression levels between PCN species for all the genes tested (Fig. 3.4A). In non-PCN species, they were all down-regulated with a mean fold change of -3.86, spanning from -0.53 to -10.88 (Fig. 3.4B). The qRT-PCR assay was not able to detect the presence of Peptidase M13 gene in non-PCN species, although it amplifies well on gDNA from these species (data not shown), confirming that the gene is not or very poorly expressed in non-PCN. The expression of these genes was also monitored after exposure to

tomato root diffusate in order to see the effect of a different host on effector genes expression for PCN species. The expression of glutathione peroxidase and peptidase M13 remained similar to when exposed to potato root diffusate, whereas RBP-1, the two SPRY domain-containing proteins 19 and pectate lyase 1 were down-regulated (respectively -0,67, -4.16, -3.88 and -11.19) (Fig 3.4C).

Figure 3.4: Expression of genes putatively associated with host preference in potato cyst nematodes. Change in the expression of selected effector genes after exposure to potato root diffusate in (A) potato cyst nematodes (PCN) species, *G. rostochiensis and G. pallida*, (B) non-PCN species, *G. tabacum* and *G. mexicana* or (C) PCN species exposed to tomato root diffusate. Expression was assessed by qRT-PCR and normalized using a set of reference genes (aaRS, PMP-3, and GR). The expression level of *G. rostochiensis* exposed to potato root diffusate was used as the calibrator for relative expression calculation. A default value of -15 was assigned to samples without detectable expression. Error bars represent the standard error of the mean of each group, and significant differences are indicated by an asterisk (*) (Tukey's test).



Discussion

In this study, we posit that genes essential for compatible host-parasite associations could be identified by comparing the transcriptomes of the infective stage of genetically similar *Globodera* species having a different host specificity. Specifically, we investigated the ability of four *Globodera* species to infect potato. In previous studies, plant-parasitic nematodes have often been compared to free-living nematodes to identify genes involved in plant parasitism (Cui et al., 2018; Mitchum et al., 2013b). Surprisingly, this resulted in the discovery of many genes that were previously never found in Metazoa (e.g. cellulase) (Haegeman et al., 2011). These genes, involved in plant cell wall degradation, defense suppression, feeding site establishment, and nutrient processes, were shown to be acquired from horizontal genes transfer mainly from bacteria (Haegeman et al., 2011) and are now considered as an effector set, promoting the ability of plant-parasitic nematodes to grow and feed on their host. Here, we compared closely related species and searched for genes directly related to host specificity.

Understanding in detail the molecular bases of pathogenicity is a major step in plant parasitic nematology, as it is a critical turning point that can lead to the development of truly effective control methods against these pathogens. Although the host shift between *G. tabacum* and *G. mexicana* may have occurred after their speciation (*G. tabacum* is genetically closer to *G. rostochiensis*, and *G. mexicana* to *G. pallida*), one can hypothesize that the same genetic variants might be involved, because the traits evolved from a similar genetic origin (Arendt & Reznick, 2008). We, therefore, assessed the changes in gene expression and gene polymorphism between potato-host and potato-non-host populations to find differences linked to host specificity, using RNA sequencing.

A total of seven predicted gene transcripts were unique to the PCN species, *G. rostochiensis* and *G. pallida*, as compared with the species non-parasitic to potato, *G. tabacum,* and *G. mexicana*. One encodes for an ubiquitin, a protein involved in protein degradation and in regulation of gene expression when associated with histones. Several ubiquitin proteins are known to be effector proteins (e.g. Ubiquitin extension protein, Ubiquitin carboxyl-terminal hydrolase), and studies have indicated that ubiquitins probably play critical roles in plant-

nematode interactions, promoting the survival of parasitic nematode (Chen et al., 2017; Chronis et al., 2013; Eves-Van Den Akker et al., 2016). Different ubiquitin-proteasome systems of human parasites were shown to serve as important virulence factors (Muoz et al., 2015). The other six predicted genes have unknown functions, but were all previously reported, including in other pathogens (Table 3.4), a result that support a putative role in parasitism. One of these genes (GROS_g12023.t1) was not amplified from the gDNA of *G. tabacum* and *G. mexicana* and presumed missing from their genome. Although no function was associated with this gene, a signature match was observed with a chromo domain. This protein structural domain is associated with chromatin organization and gene expression. This result may indicate that the loss of this gene may have impacted the expression of other genes involved in the nematode capacity to infect potato. The amplification of the other genes in the gDNA of *G. tabacum* and *G. mexicana* demonstrates that although the genes are present, they are not expressed under these conditions.

By comparing whole transcriptomic data, several genes and alleles whose expressions were correlated to the ability to parasitize the potato plant were identified. This includes 545 DEGs, among which 78 were known effector proteins (14.3%). Through the entire *G. rostochiensis* reference transcriptome (nGr.v1.1), 315 known or putative effector genes were identified based on previously published datasets (Chen et al., 2017; Eves-Van Den Akker et al., 2016) which represent 2.2% of the 14 309 predicted genes. Thus, almost 25% of these effector genes were up-regulated in PCN species in our experiment, representing a significant enrichment (6.5 times more represented) compared to the whole transcriptome. Among the 78 differentially expressed effector genes identified, 39 were linked to microtubule cytoskeleton organization, a key element in the feeding site establishment essential to plant-parasitic nematodes. Without successfully establishing a feeding-site, the nematode failed to feed itself and died without triggering a defense response. This corresponds to the situation observed for *G. mexicana* larvae that are able to invade the potato roots and migrate towards the vascular cylinder but failed to initiate a feeding site in potato roots. This enrichment of effectors expression in PCN species could be the result of a poor activation of the transcription of effector genes transcription in non-PCN species when exposed to potato root diffusate. The most significantly overexpressed DEG was a SMC_N family protein gene that was up-regulated 30.6 times in PCN populations. The

SMC (structural maintenance of chromosomes) family proteins are involved in chromatin and DNA dynamics (Strunnikov & Jessberger, 1999). Furthermore, it was recently shown that one of the first genes expressed in *G. rostochiensis* and *G. pallida* following exposure to potato root diffusate was coding for a neprilysin protein, a "*transmembrane metalloprotease able to activate/inactivate peptide hormones that could be involved in a cascade of events*" (Duceppe et al., 2017b; Sabeh et al., 2018). Interestingly, the second most highly overexpressed gene in our dataset was a peptidase M13 (GROS_g12349.t1), an unassigned homolog of the neprilysin gene, that was up-regulated 28.5 times in PCN populations. Among other up-regulated DEGs in PCN populations, 11 (3.4%) were also involved in the regulation of gene expression.

The chemical signals of potato root diffusate may not allow proper activation of the infective stage of J2 larvae for non-PCN populations. Up-regulation of several regulatory genes was not observed in J2 larvae following exposure to potato root diffusates, unlike PCN populations. The decreased expression of certain effector genes, when exposed to tomato root diffusate, in PCN species, as well as the increased expression of these genes in non-PCN species when exposed to tomato root diffusate, shows that the parasitic nematode may adjust its set of effectors for each potential host.

In addition, 359 non-synonymous variants showed evidence for selection between PCN and non-PCN species. Among these, ten were effector genes (Table 3.5) and 21 others with an unknown function contained a signal peptide for excretion. These non-synonymous variants may affect the function of these effectors and pathogenicity. Also, most amino acid replacements were non-conservative, with an amino acid replacement from another side chain group, which increase the impact on protein function (Dagan et al., 2002). Several cases of polymorphism of a single amino acid having a major impact on the function of a protein have been reported, including some in these highlighted effector genes. Secreted chorismate mutase is thought to alter plant cell development, cell growth, and plant defenses and is an important virulence factor found in many plant pathogens (e.g. *Meloidogyne javanica*, *Ustilago maydis*) (Djamei et al., 2011; Doyle & Lambert, 2003). It was shown that three single amino acid polymorphisms in *Heterodera glycines* chorismate mutase were associated with the ability to break host resistance on two particular soybean cultivars (Bekal et al., 2003).

In this study, we highlighted significant differences in gene expression and gene variation between PCN (*G. rostochiensis and G. pallida*) and non-PCN species (*G. tabacum and G. mexicana*), although they are extremely close genetically. These distinctions were particularly evident in effector genes, which were highly enriched among DEGs and whose expression was reliant on the host, despite the fact that all species share a similar effector gene set in their genome, and that only a few non-synonymous variants were found in effector genes of non-PCN species. Therefore, it seems that the determinant of host specificity may reside in the regulation of essential effector genes expression. Because neprilysin was recently suggested to be involved in the early response to root diffusate and was highly overexpressed in PCN species, it might be involved in parasitism. Ubiquitin and other genes unique to PCN, particularly those absent from non-PCN genome, also appeared as good candidates. We are strongly confident that genes responsible for the inability of non-PCN species to successfully infect potato plants are highlighted within our results and that a limited number of potential candidates have been identified. In future work it would be interesting to overexpress the neprilysin gene in *G. mexicana* and *G. tabacum* to verify if this induces an up-regulation of regulator or effector genes. These genes could become useful molecular targets to design new efficient strategies against PCN. Inhibition of the key regulators involved in host recognition may prevent the activation of infectious J2 larvae and avoid substantial yield losses.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Sequencing data were submitted to the NCBI Sequence Read Archive under the bioproject accession number SRP146253.

Competing interests

The authors declare no competing interests

Funding

This work was funded by Agriculture and Agri-Food Canada (AAFC) and the United States Department of Agriculture (USDA) through a Coordinated Agricultural Project (CAP) titled "Risk Assessment and Eradication of *Globodera* spp. in U.S. Production of Potato," supported by award #2015-69004-23634 from the National Institute for Food and Agriculture (https://www.globodera.org/).

Authors' contribution

MS, MSA, and BM conceived and designed all experiments; MS performed experiments; MS and EL performed bio-informatics analysis; BM contributed to materials/analysis tools; All authors discussed the results; MS wrote the manuscript with contributions from all the other authors. All authors read and approved the manuscript.

Acknowledgments

The authors thank Guy Bélair (Agriculture and Agri-Food Canada), Geert Smant (Wageningen University and Research Centre) and Gerrit Karssen (Wageningen University and Research Centre) for providing us with isolates of cyst nematodes.

References

Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M. C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T. R., Markov, G. V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Segurens, B., Steinbach, D., Tytgat, T., Ugarte, E., van Ghelder, C., Veronico, P., Baum, T. J., Blaxter, M., Bleve-Zacheo, T., Davis, E. L., Ewbank, J. J., Favery, B., Grenier, E., Henrissat, B., Jones, J. T., Laudet, V., Maule, A. G., Quesneville, H., Rosso, M. N., Schiex, T., Smant, G., Weissenbach, J., & Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature biotechnology, 26*(8), 909-915. doi:10.1038/nbt.1482

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. doi:10.1016/j.meegid.2014.07.027

Ali, S., Magne, M., Chen, S., Cote, O., Stare, B. G., Obradovic, N., Jamshaid, L., Wang, X., Belair, G., & Moffett, P. (2015a). Analysis of putative apoplastic effectors from the nematode, *Globodera rostochiensis*, and identification of an expansin-like protein that can induce and suppress host defenses. *PLOS ONE, 10*(1), e0115042. doi:10.1371/journal.pone.0115042

Ali, S., Magne, M., Chen, S., Obradovic, N., Jamshaid, L., Wang, X., Belair, G., & Moffett, P. (2015b). Analysis of *Globodera rostochiensis* effectors reveals conserved functions of SPRYSEC proteins in suppressing and eliciting plant immune responses. *Frontiers in Plant Science, 6*, 623. doi:10.3389/fpls.2015.00623

Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., & Singh, K. B. (2010b). Plants versus pathogens: an evolutionary arms race. *Functional Plant Biology, 37*(6), 499-512. doi:10.1071/FP09304

Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution, 23*(1), 26-32. doi:10.1016/j.tree.2007.09.011

Bekal, S., Niblack, T. L., & Lambert, K. N. (2003). A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. *Molecular Plant-Microbe Interactions, 16*(5), 439-446. doi:10.1094/MPMI.2003.16.5.439

Bird, D. M., Opperman, C. H., & Davies, K. G. (2003). Interactions between bacteria and plant-parasitic nematodes: now and then. *International Journal for Parasitology, 33*(11), 1269-1276. doi:10.1016/s0020-7519(03)00160-7

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*(1), 15-22. doi:10.1016/j.pmpp.2005.09.001

Blaxter, M., & Koutsovoulos, G. (2015). The evolution of parasitism in Nematoda. *Parasitology, 142 Suppl 1*, S26-39. doi:10.1017/S0031182014000791

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T., & Thomas, W. K. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature, 392*(6671), 71-75. doi:10.1038/32160

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Bossis, M., & Mugniéry, D. (1993). Specific status of six *Globodera* parasites of solanaceous plants studied by means of two-dimensional gel electrophoresis with a comparison of gel patterns by a computed system. *Fundamental and Applied Nematology, 16*(1), 47-56.

Chen, C., Cui, L., Chen, Y., Zhang, H., Liu, P., Wu, P., Qiu, D., Zou, J., Yang, D., Yang, L., Liu, H., Zhou, Y., & Li, H. (2017). Transcriptional responses of wheat and the cereal cyst nematode *Heterodera avenae* during their early contact stage. *Scientific Reports, 7*(1), 14471. doi:10.1038/s41598-017-14047-y

Chevenet, F., Brun, C., Banuls, A. L., Jacq, B., & Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics, 7*, 439. doi:10.1186/1471-2105-7-439

Chronis, D., Chen, S., Lu, S., Hewezi, T., Carpenter, S. C., Loria, R., Baum, T. J., & Wang, X. (2013). A ubiquitin carboxyl extension protein secreted from a plant-parasitic nematode *Globodera rostochiensis* is cleaved in planta to promote plant parasitism. *The Plant Journal, 74*(2), 185-196. doi:10.1111/tpj.12125

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin), 6*(2), 80-92. doi:10.4161/fly.19695

Cui, J. K., Peng, H., Qiao, F., Wang, G. F., Huang, W. K., Wu, D. Q., & Peng, D. (2018). Characterization of Putative Effectors from the Cereal Cyst Nematode *Heterodera avenae*. *Phytopathology, 108*(2), 264-274. doi:10.1094/PHYTO-07-17-0226-R

Dagan, T., Talmor, Y., & Graur, D. (2002). Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be

indicative of positive Darwinian selection. *Molecular Biology and Evolution, 19*(7), 1022-1025. doi:10.1093/oxfordjournals.molbev.a004161

Danchin, E. G. J., Rosso, M.-N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., & Abad, P. (2010b). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences, 107*(41), 17651-17656. doi:10.1073/pnas.1008486107

Davidson, N. M., & Oshlack, A. (2014b). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology, 15*(7), 410. doi:10.1186/s13059-014-0410-6

Decraemer, W., & Hunt, D. J. (2006). Structure and classification. In R. N. Perry & M. Moens (Eds.), *Plant nematology* (pp. 3-32). Wallingford: CAB International.

Den Nijs, L., & Karssen, G. (2004). *Globodera rostochiensis* and *Globodera pallida*. *EPPO Bulletin, 34*(2), 309-314. doi:10.1111/j.1365-2338.2004.00734.x

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., & Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research, 36*, W465-469. doi:10.1093/nar/gkn180

Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., Osorio, S., Tohge, T., Fernie, A. R., Feussner, I., Feussner, K., Meinicke, P., Stierhof, Y. D., Schwarz, H., Macek, B., Mann, M., & Kahmann, R. (2011). Metabolic priming by a secreted fungal effector. *Nature, 478*(7369), 395-398. doi:10.1038/nature10454

Doyle, E. A., & Lambert, K. N. (2003). *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. *Molecular Plant-Microbe Interactions, 16*(2), 123-131. doi:10.1094/MPMI.2003.16.2.123

Duceppe, M. O., Lafond-Lapalme, J., Palomares-Rius, J. E., Sabeh, M., Blok, V., Moffett, P., & Mimee, B. (2017b). Analysis of survival and hatching transcriptomes from potato cyst nematodes, *Globodera rostochiensis* and *G. pallida*. *Scientific Reports, 7*(1), 3882. doi:10.1038/s41598-017-03871-x

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Eves-van den Akker, S., & Birch, P. R. (2016). Opening the Effector Protein Toolbox for Plant-Parasitic Cyst Nematode Interactions. *Molecular Plant, 9*(11), 1451-1453. doi:10.1016/j.molp.2016.09.008

Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G., Da Rocha, M., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szitenberg, A., Grenier, E., Montarry, J., Mimee, B., Duceppe, M. O., Boyes, I., Marvin, J. M., Jones, L. M., Yusup, H. B.,

Lafond-Lapalme, J., Esquibet, M., Sabeh, M., Rott, M., Overmars, H., Finkers-Tomczak, A., Smant, G., Koutsovoulos, G., Blok, V., Mantelin, S., Cock, P. J., Phillips, W., Henrissat, B., Urwin, P. E., Blaxter, M., & Jones, J. T. (2016). The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology, 17*(1), 124. doi:10.1186/s13059-016-0985-1

Eves-van den Akker, S., Lilley, C. J., Jones, J. T., & Urwin, P. E. (2014). Identification and characterisation of a hyper-variable apoplastic effector gene family of the potato cyst nematodes. *PLOS Pathogens, 10*(9), e1004391. doi:10.1371/journal.ppat.1004391

Fenwick, D. W. (1949). Investigations on the emergence of larvae from cysts of the potato-root eelworm *Heterodera rostochiensis*; technique and variability. *Journal of Helminthology, 23*(3-4), 157-170. doi:10.1017/s0022149x0003248x

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*(2), 977-993. doi:10.1534/genetics.108.092221

Garrison, E., & Marth, G. (2012a). Haplotype-based variant detection from short-read sequencing. *arXiv preprint, Available from: arXiv:1207.3907*.

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., & Bryant, S. H. (2010a). The NCBI BioSystems database. *Nucleic Acids Research, 38*, D492-496. doi:10.1093/nar/gkp858

Geric Stare, B., Fouville, D., Sirca, S., Gallot, A., Urek, G., & Grenier, E. (2011b). Molecular variability and evolution of the pectate lyase (pel-2) parasitism gene in cyst nematodes parasitizing different solanaceous plants. *Journal of Molecular Evolution, 72*(2), 169-181. doi:10.1007/s00239-010-9413-4

Grenier, E., Blok, V. C., Jones, J. T., Fouville, D., & Mugniery, D. (2002). Identification of gene expression differences between *Globodera pallida* and *G.'mexicana'* by suppression subtractive hybridization. *Molecular Plant Pathology, 3*(4), 217-226. doi:10.1046/j.1364-3703.2002.00111.x

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol, 59*(3), 307-321. doi:10.1093/sysbio/syq010

Haegeman, A., Jones, J. T., & Danchin, E. G. (2011). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Molecular Plant-Microbe Interactions, 24*(8), 879-887. doi:10.1094/MPMI-03-11-0055

Haegeman, A., Mantelin, S., Jones, J. T., & Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene, 492*(1), 19-31. doi:10.1016/j.gene.2011.10.040

Jeffreys, H. (1998). *The Theory of Probability*: OUP Oxford.

Jones, J. T., Kumar, A., Pylypenko, L. A., Thirugnanasambandam, A., Castelli, L., Chapman, S., Cock, P. J., Grenier, E., Lilley, C. J., Phillips, M. S., & Blok, V. C. (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-Microbe Interactions, 10*(6), 815-828. doi:10.1111/j.1364-3703.2009.00585.x

Kall, L., Krogh, A., & Sonnhammer, E. L. (2004b). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology, 338*(5), 1027-1036. doi:10.1016/j.jmb.2004.03.016

Kanehisa, M., & Goto, S. (2000b). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research, 28*(1), 27-30. doi:10.1093/nar/28.1.27

Konczal, M., Koteja, P., Stuglik, M. T., Radwan, J., & Babik, W. (2014). Accuracy of allele frequency estimation using pooled RNA-Seq. *Molecular Ecology Resources, 14*(2), 381-392. doi:10.1111/1755-0998.12186

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods, 25*(4), 402-408. doi:10.1006/meth.2001.1262

Love, M. I., Huber, W., & Anders, S. (2014b). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Madani, M., Ward, L. J., & De Boer, S. H. (2008). Multiplex real-time polymerase chain reaction for identifying potato cyst nematodes, *Globodera pallida* and *Globodera rostochiensis*, and the tobacco cyst nematode, *Globodera tabacum*. *Canadian Journal of Plant Pathology, 30*(4), 554-564. doi:10.1080/07060660809507555

Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Research, 32*, W327-331. doi:10.1093/nar/gkh454

Mitchum, M. G., Hussey, R. S., Baum, T. J., Wang, X., Elling, A. A., Wubben, M., & Davis, E. L. (2013b). Nematode effector proteins: an emerging paradigm of parasitism. *New Phytol, 199*(4), 879-894. doi:10.1111/nph.12323

Muoz, C., San Francisco, J., Gutirrez, B., & Gonzlez, J. (2015). Role of the Ubiquitin-Proteasome Systems in the Biology and Virulence of Protozoan Parasites. *BioMed Research International, 2015*, 13. doi:10.1155/2015/141526

Paganini, J., Campan-Fournier, A., Da Rocha, M., Gouret, P., Pontarotti, P., Wajnberg, E., Abad, P., & Danchin, E. G. (2012). Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes. *PLOS ONE, 7*(11), e50875. doi:10.1371/journal.pone.0050875

Palomares-Rius, J. E., Hedley, P., Cock, P. J., Morris, J. A., Jones, J. T., & Blok, V. C. (2016). Gene expression changes in diapause or quiescent potato cyst nematode, *Globodera pallida*, eggs after hydration or exposure to tomato root diffusate. *PeerJ, 4*, e1654. doi:10.7717/peerj.1654

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods, 8*(10), 785-786. doi:10.1038/nmeth.1701

Poppe, S., Dorsheimer, L., Happel, P., & Stukenbrock, E. H. (2015). Rapidly Evolving Genes Are Key Players in Host Specialization and Virulence of the Fungal Wheat Pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). *PLOS Pathogens, 11*(7), e1005055. doi:10.1371/journal.ppat.1005055

Quist, C. W., Smant, G., & Helder, J. (2015). Evolution of plant parasitism in the phylum Nematoda. *Annual Review of Phytopathology, 53*, 289-310. doi:10.1146/annurev-phyto-080614-120057

Sabeh, M., Duceppe, M. O., St-Arnaud, M., & Mimee, B. (2018). Transcriptome-wide selection of a reliable set of reference genes for gene expression studies in potato cyst nematodes (*Globodera* spp.). *PLOS ONE, 13*(3), e0193840. doi:10.1371/journal.pone.0193840

Smant, G., Helder, J., & Goverse, A. (2018). Parallel adaptations and common host cell responses enabling feeding of obligate and facultative plant parasitic nematodes. *The Plant Journal, 93*(4), 686-702. doi:10.1111/tpj.13811

Sobczak, M., & Golinowski, W. (2011). Cyst Nematodes and Syncytia. In J. Jones, G. Gheysen, & C. Fenoll (Eds.), *Genomics and Molecular Genetics of Plant-Nematode Interactions* (pp. 61-82). Dordrecht: Springer Netherlands.

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics, 24*(5), 637-644. doi:10.1093/bioinformatics/btn013

Strunnikov, A. V., & Jessberger, R. (1999). Structural maintenance of chromosomes (SMC) proteins: conserved molecular properties for multiple biological functions. *uropean Journal of Biochemistry, 263*(1), 6-13. doi:10.1046/j.1432-1327.1999.00509.x

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol, 56*(4), 564-577. doi:10.1080/10635150701472164

The UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic acids research, 45*(D1), D158-D169. doi:10.1093/nar/gkw1099

Thorpe, P. (2012). *Bioinformatic and functional characterisation of Globodera pallida effector genes.* (Doctor of Philosophy Ph.D. Thesis), The University of Leeds,

Thorpe, P., Mantelin, S., Cock, P. J., Blok, V. C., Coke, M. C., Eves-van den Akker, S., Guzeeva, E., Lilley, C. J., Smant, G., Reid, A. J., Wright, K. M., Urwin, P. E., & Jones, J. T. (2014). Genomic characterisation of the effector complement of the potato cyst nematode *Globodera pallida*. *BMC Genomics, 15*, 923. doi:10.1186/1471-2164-15-923

Turner, S. J. (1996). Population decline of potato cyst nematodes (*Globodera rostochiensis, G. pallida*) in field soils in Northern Ireland. *Annals of Applied Biology, 129*(2), 315-322. doi:10.1111/j.1744-7348.1996.tb05754.x

Varden, F. A., De la Concepcion, J. C., Maidment, J. H., & Banfield, M. J. (2017). Taking the stage: effectors in the spotlight. *Current Opinion in Plant Biology, 38*, 25-33. doi:10.1016/j.pbi.2017.04.013

Additional materials

Figure 3.1: Genetic similarities of four *Globodera* species compared to *Caenorhabditis elegans.* Phylogenetic tree of the small subunit ribosomal RNA gene sequences from *Globodera rostochiensis*, *G. pallida*, *G. mexicana*, *G. tabacum* and *C. elegans*. Analysis was performed using Phylogeny.fr, bootstrap values (*) are given next to the nodes. The scale bars represent the number of substitutions per site.

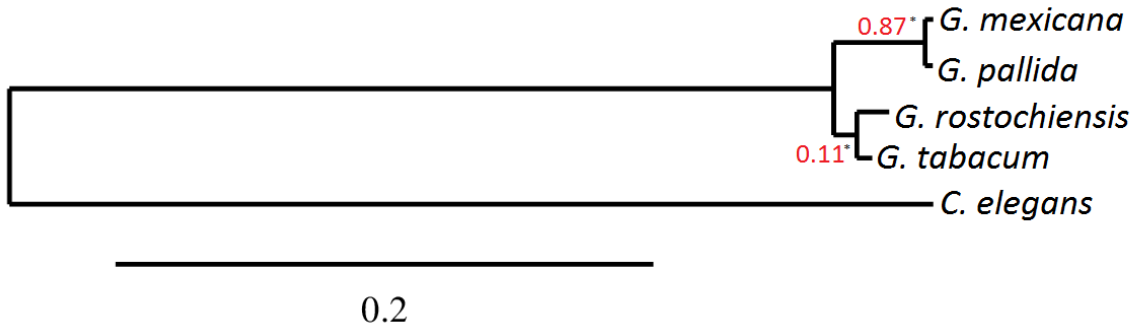


Tableau 3.6: Differentially expressed genes (DEG, P < 0.01) between PCN (*Globodera rostochiensis*, *G. pallida)* and non-PCN species (*G. tabacum*, *G. mexicana*); * indicatesknown effector genes. SeqID and Gene description corresponds to sequences ID and annotation of *G. rostochiensis* reference transcriptome (nGr.v1.1).

Gene description	SeqID	Fold Change
2-acylglycerol O-acyltransferase 2-A-like	GROS_g11514.t1	5.9
33A09 *	GROS_g03474.t1	7.9
ABC transporter family protein	GROS_g01146.t1	-2.8
acetylcholine receptor DES-2-like protein	GROS_g05548.t1	2.3
Acetylcholinesterase	GROS_g12528.t1, GROS_g09443.t1	2,1 to 2,5
Acyltransferase family protein	GROS_g00281.t1	7.0
ADAM-TS Spacer 1 family protein	GROS_g12554.t1, GROS_g13359.t1	11,3 to 13,6
Amidase	GROS_g01693.t1	4.0
Annexin 2 *	GROS_g01954.t1	-2.9
Annexin family and Annexin, type III family *	GROS_g07837.t1	15.8
AP-2	GROS_g05722.t1	2.4
Ararbinogalactan endo-1,4-beta-galactosidase *	GROS_g08150.t1	8.2
Astacin-like metalloendopeptidase	GROS_g13929.t1, GROS_g13937.t1, GROS_g09809.t1	3,2 to 9,1
Beta-1,4-endoglucanase-4 *	GROS_g11200.t1	4.8
Beta1,4-N-acetylglucosaminyltransferase	GROS_g07386.t1	-3.1
BTB poz domain-containing protein 2	GROS_g14218.t1	6.7
Bumetanide-sensitive sodium-(potassium)-chloride cotransporter	GROS_g01179.t1	3.1
Cathepsin B-like cysteine proteinase *	GROS_g12051.t1	4.4
Cathepsin X *	GROS_g11524.t1	3.6
Cbn-ccg-1 protein	GROS_g10576.t1	-3.2
CBR-ADT-1 protein	GROS_g09408.t1	3.5
CBR-DGK-5 protein	GROS_g01862.t1	-3.1
CBR-HUM-4 protein	GROS_g04142.t1	2.4
CBR-IGCM-2 protein	GROS_g07965.t1	2.7
CBR-IGCM-3 protein	GROS_g10563.t1	2.3
CBR-IRS-2 protein	GROS_g03974.t1	-2.1
CBR-NAS-11 protein	GROS_g10550.t1	6.0
CBR-NAS-14 protein	GROS_g09403.t1	4.7
CBR-NHX-9 protein	GROS_g11653.t1	-2.3
CBR-PTR-18 protein	GROS_g02953.t1	-5.5
CBR-TAG-263 protein	GROS_g00851.t1	4.4
CBR-TAG-271 protein	GROS_g02318.t1	-2.0
CBR-UGT-49 protein	GROS_g08532.t1	2.8
CBR-VAB-19 protein	GROS_g08473.t1	2.3
CD36 antigen domain containing protein	GROS_g11642.t1	-3.2
Cell death protein CED-3	GROS_g12555.t1	9.7
CG10600-PA	GROS_g03878.t1	-2.1
Chain A, Designed Armadillo-Repeat Protein	GROS_g10041.t1, GROS_g05714.t1	5,4 to 7,7
Chitinase *	GROS_g06362.t1, GROS_g10777.t1	3,8 to 10,7
Chitinase *	GROS_g00668.t1, GROS_g00189.t1	-2,6 to -3,3
Chloride channel protein ClC-Ka, Putative	GROS_g13867.t1	-3.8
CLAVATA3/ESR (CLE)-related protein 4C *	GROS_g12818.t1	4.2
COL-1	GROS_g09924.t1	4.3
Collagen alpha-6 chain	GROS_g09579.t1	-5.7
Conserved hypothetical protein	GROS_g13525.t1	10.6
Contactin	GROS_g02869.t1	-2.6
Copper/zinc superoxide dismutase *	GROS_g10480.t1	4.1
Coronin-like protein	GROS_g11801.t1	2.1

CRE-CLH-3 protein	GROS_g10994.t1	-3.7
CRE-NAS-4 protein	GROS_g05594.t1	-5.6
CRE-RIG-4 protein	GROS_g01772.t1	-2.3
CRE-TAG-341 protein	GROS_g07136.t1	3.8
C-type lectin *	GROS_g11230.t1, GROS_g11229.t1, GROS_g11186.t1	4,4 to 5,5
CUE domain	GROS_g13210.t1	18.3
Cyanate hydratase	GROS_g09531.t1	3.3
Cyclin domain containing protein	GROS_g10296.t1	3.8
Death-inducer obliterator 1	GROS_g08807.t1	3.5
DNA replication factor Dna2 domain containing protein	GROS_g03553.t1	-3.1
Dorsal gland cell-specific expression protein *	GROS_g14141.t1, GROS_g14146.t1, GROS_g14270.t1, GROS_g14268.t1, GROS_g14301.t1	2,2 to 3,5
Dynein Light Chain family member	GROS_g11992.t1	4.1
EB module family protein	GROS_g05093.t1	3.2
Excitatory amino acid transporter	GROS_g03064.t1, GROS_g03065.t1, GROS_g03753.t1	2,0 to 3,0
Exonuclease family protein	GROS_g01435.t1	-8.7
Expressed protein	GROS_g08299.t1	4.9
FAD-binding protein	GROS_g11633.t1	11.6
FAR-1 *	GROS_g09434.t1	3.6
F-box only protein 47	GROS_g06569.t1	-5.0
Flavin-containing monooxygenase fmo gs-ox3	GROS_g01621.t1	2.8
Foot protein 1 variant 1, partial	GROS_g10469.t1	2.2
Fringe glycosyltransferase	GROS_g03237.t1	-4.3
Galactokinase family protein	GROS_g12384.t1	6.2
Galactoside-binding lectin family protein *	GROS_g06697.t1, GROS_g01996.t1	-2,1 to -2,5
Galectin *	GROS_g01184.t1	-3.8
Gamma interferon inducible lysosomal thiol reductase GILT domain containing protein	GROS_g07806.t1	3.5
GI15478	GROS_g03980.t1	2.4
Glutathione S-transferase *	GROS_g10872.t1	3.1
Glutathione S-transferase *	GROS_g06619.t1, GROS_g01793.t1	-3,1 to -3,3
Glutathione synthetase *	GROS_g12954.t1	4.0
Glyoxalase domain containing 4	GROS_g06109.t1	-2.1
Ground region domain containing protein	GROS_g00605.t1	-2.8
Guanine nucleotide binding protein (G-protein)	GROS_g11261.t1	-4.9
High mobility group B protein 13-like	GROS_g14265.t1	2.0
Histidine acid phosphatase family protein	GROS_g09164.t1	-3.1
Histone H2B 8-like	GROS_g03924.t1	6.5
Histone-lysine N-methyltransferase EHMT1	GROS_g11502.t1	5.9
HSF-type DNA-binding domain containing protein	GROS_g13095.t1	-2.2
Hypothetical 24.7 kDa protein in POM152-REC114 intergenic region	GROS_g02832.t1	4.1
Hypothetical esophageal gland cell secretory protein 4 *	GROS_g01765.t1	11.0
Hypothetical protein	GROS_g13745.t1, GROS_g12347.t1, GROS_g11441.t1, GROS_g08271.t1, GROS_g09393.t1, GROS_g06567.t1, GROS_g13509.t1, GROS_g03396.t1, GROS_g14174.t1, GROS_g08959.t1, GROS_g07342.t1, GROS_g05782.t1, GROS_g13041.t1, GROS_g05246.t1, GROS_g09903.t1, GROS_g10584.t1, GROS_g02450.t1, GROS_g10861.t1, GROS_g08737.t1, GROS_g00118.t1, GROS_g03812.t1, GROS_g05710.t1, GROS_g03395.t1, GROS_g13200.t1, GROS_g08696.t1, GROS_g04487.t1, GROS_g04970.t1, GROS_g07200.t1, GROS_g08735.t1, GROS_g04866.t1, GROS_g00548.t1, GROS_g09550.t1, GROS_g14173.t1, GROS_g09206.t1, GROS_g06799.t1, GROS_g02649.t1, GROS_g11450.t1, GROS_g02245.t1, GROS_g08490.t1, GROS_g05636.t1, GROS_g04126.t1, GROS_g11099.t1,	2,0 to 8,2

	GROS_g03982.t1, GROS_g01612.t1, GROS_g11212.t1, GROS_g05569.t1, GROS_g05094.t1, GROS_g11353.t1, GROS_g12744.t1	
Hypothetical protein	GROS_g04510.t1, GROS_g09418.t1, GROS_g11202.t1, GROS_g10399.t1, GROS_g08159.t1, GROS_g00529.t1, GROS_g08369.t1, GROS_g13090.t1, GROS_g10449.t1, GROS_g10428.t1, GROS_g09486.t1, GROS_g05643.t1, GROS_g02426.t1, GROS_g07008.t1, GROS_g06549.t1, GROS_g13624.t1, GROS_g08934.t1, GROS_g07108.t1, GROS_g10754.t1	-2,0 to -8,3
Intraflagellar transport protein 52 homolog	GROS_g03100.t1	-3.7
Inversin protein alternative isoform	GROS_g14132.t1	4.1
Laminin-like protein C54D1.5 precursor	GROS_g08588.t1	-6.3
Lipase, class 3 family-containing protein	GROS_g03515.t1	-10.6
LOC100149074 protein	GROS_g02557.t1	4.5
Loechrig isoform VII	GROS_g13491.t1	-3.3
Low-density lipoprotein receptor-related protein 2	GROS_g00544.t1	3.6
Mak16 protein	GROS_g09892.t1	2.1
Manba-prov protein	GROS_g12832.t1	2.8
Manganese superoxide dismutase *	GROS_g05491.t1	-2.4
Multiple exostoses homolog 2	GROS_g09228.t1	-2.9
MutL homolog 1	GROS_g01051.t1	6.3
MutS domain III family protein	GROS_g03582.t1	5.2
MYND finger family protein	GROS_g12133.t1	2.5
Non-specificserine/threonineproteinkinase.n/a	GROS_g11095.t1	3.4
Nuclear hormone receptor, ligand-binding	GROS_g08841.t1	3.0
Other/TBCK protein kinase	GROS_g04855.t1	-2.0
P7E4 protein	GROS_g14133.t1	5.1
palmitoyl-CoA oxidase	GROS_g13778.t1	3.1
Pao retrotransposon peptidase family protein	GROS_g09279.t1, GROS_g13372.t1, GROS_g12921.t1, GROS_g12922.t1	-2,7 to -6,4
Patched family protein	GROS_g04808.t1	-4.4
Pax-6 transcription factor	GROS_g08493.t1	5.2
Pectate lyase 1 *	GROS_g07185.t1, GROS_g07968.t1	3,4 to 6,7
Pectate lyase 2 *	GROS_g04366.t1	5.4
Peptidase M13 unassigned homologues	GROS_g12349.t1	28.5
Phosphate transporter family protein	GROS_g07079.t1	-2.6
Phosphatidylinositol 4-phosphate 5-kinase type-1 gamma-like, partial	GROS_g13738.t1	7.1
Phosphoglycerate mutase family protein	GROS_g04915.t1	3.5
Plasmid Maintenance Protein containing protein	GROS_g01425.t1	-2.9
Potassium channel regulatory protein sup-10	GROS_g10105.t1	3.5
Predicted protein	GROS_g02428.t1	-3.2
Pregnancy-induced growth inhibitor	GROS_g03181.t1	-2.2
Protein AAT-6, isoform b	GROS_g00676.t1	-2.2
Protein ACS-22, isoform a	GROS_g09999.t1	-2.1
Protein Bm10998, isoform c	GROS_g10633.t1	-2.8
Protein Bm1342	GROS_g04245.t1	-2.2
Protein Bm2475	GROS_g09710.t1	-3.3
Protein Bm2855, isoform a	GROS_g05762.t1	2.2
Protein Bm2986, isoform b	GROS_g08766.t1	-2.2
Protein Bm4315, partial	GROS_g10486.t1	3.8
Protein BM-TTR-33, isoform b	GROS_g04409.t1	2.8
Protein BM-TTR-41, isoform d	GROS_g01450.t1	3.5
Protein C09B8.3	GROS_g12231.t1	-16.9
Protein CATP-2	GROS_g04410.t1	-2.0
Protein CBG10108	GROS_g11064.t1	5.0
Protein CBN-1, isoform b	GROS_g01081.t1	6.3
Protein CLK-1	GROS_g05914.t1	2.2

Protein COL-158	GROS_g06243.t1	3.8
Protein COL-164	GROS_g08818.t1	-3.2
Protein COL-3	GROS_g01183.t1	-7.3
Protein CYP-13B1, isoform c	GROS_g07387.t1	2.9
Protein D1053.3	GROS_g12043.t1	-2.4
Protein DAT-1	GROS_g01628.t1	2.3
Protein DHP-2	GROS_g07832.t1	2.3
Protein F41C3.5	GROS_g06427.t1	2.9
Protein GCY-20	GROS_g03198.t1	-2.0
Protein GLY-8	GROS_g05061.t1	3.8
Protein IMA-2	GROS_g02732.t1	2.6
Protein INSC-1	GROS_g07148.t1	-4.8
Protein K11D9.3	GROS_g12247.t1	2.4
Protein kinase domain containing protein	GROS_g13801.t1	3.4
Protein LIN-48	GROS_g03769.t1	-2.4
Protein LRON-2	GROS_g08563.t1	3.6
Protein MBOA-1	GROS_g05531.t1	5.9
Protein MLT-10	GROS_g02618.t1, GROS_g03611.t1	-4,5 to -5,6
Protein of unknown function DUF1647 domain containing protein	GROS_g13245.t1	2.3
Protein PMT-1, isoform b	GROS_g09108.t1	-3.2
Protein RHR-2	GROS_g09207.t1	2.3
Protein ROP-1	GROS_g02654.t1	-2.1
Protein SAX-1, isoform b	GROS_g12230.t1	-2.7
Protein THN-1	GROS_g05322.t1	9.1
Protein TKT-1	GROS_g05816.t1	7.2
Protein TWK-16	GROS_g11106.t1	2.6
Protein TWK-39, isoform a	GROS_g06797.t1	3.2
Protein WHT-2	GROS_g03574.t1	-2.2
Protein Y11D7A.9	GROS_g06487.t1	6.4
Protein Y45G12B.3	GROS_g00827.t1	-2.7
Protein Y53F4B.39, isoform a	GROS_g02198.t1	-2.9
Protein Y54F10AM.8	GROS_g07595.t1, GROS_g02392.t1	2,3 to 2,6
Protein ZTF-2	GROS_g09066.t1, GROS_g09071.t1	-2,3 to -3,0
Protein-tyrosine phosphatase containing protein	GROS_g13353.t1	3.1
Putative carbonic anhydrase 5 precursor	GROS_g03957.t1	2.1
Putative cytochrome b5	GROS_g01945.t1	2.1
Putative deoxyribonuclease TATDN3 isoform X1	GROS_g09441.t1	2.1
Putative esophageal gland cell protein Hgg-20 *	GROS_g10153.t1	-4.9
Putative gland protein G19B10 *	GROS_g13121.t1	3.7
Rac GTPase-activating protein 1	GROS_g06328.t1	6.0
RanBPM-like protein *	GROS_g14151.t1, GROS_g14278.t1	5,7 to 3,3
RBP-1 protein *	GROS_g14180.t1, GROS_g14165.t1, GROS_g14179.t1, GROS_g14157.t1, GROS_g14158.t1, GROS_g14236.t1, GROS_g14128.t1	3,1 to 9,4
RBP-4 protein *	GROS_g14261.t1, GROS_g14167.t1, GROS_g14163.t1, GROS_g14307.t1, GROS_g14122.t1, GROS_g14196.t1, GROS_g14272.t1	2,2 to 2,9
RGC/RGC protein kinase	GROS_g13452.t1	5.6
RIKEN cDNA 2410127L17	GROS_g04374.t1	-2.0
RP42	GROS_g10489.t1	6.5
SD03319p	GROS_g11738.t1	-2.9
Secreted glutathione peroxidase *	GROS_g02490.t1	4.3
Secreted SPRY domain-containing protein 1 *	GROS_g14154.t1	6.0
Secreted SPRY domain-containing protein 15 *	GROS_g14287.t1, GROS_g14275.t1	4,7 to 7,2
Secreted SPRY domain-containing protein 19 *	GROS_g14126.t1, GROS_g14234.t1, GROS_g14125.t1, GROS_g14168.t1, GROS_g14220.t1, GROS_g14153.t1, GROS_g14194.t1, GROS_g14306.t1, GROS_g14130.t1, GROS_g14219.t1, GROS_g14294.t1, GROS_g14193.t1,	2,2 to 26,8

	GROS_g14195.t1, GROS_g14129.t1, GROS_g14260.t1, GROS_g14256.t1, GROS_g14241.t1, GROS_g14159.t1	
Secreted SPRY domain-containing protein 5 *	GROS_g14191.t1	11.5
Secreted SPRY domain-containing protein 8 *	GROS_g14299.t1	8.4
Serine/threonine protein phosphatase PP1 isozyme 1	GROS_g06548.t1	2.8
Serine/threonine-protein kinase mph1	GROS_g00356.t1	3.0
Signal recognition particle receptor alpha subunit	GROS_g11781.t1	-2.0
Similar to A4 Putative effector *	GROS_g13646.t1, GROS_g08139.t1	3,7 to 18,4
SMC_N superfamily protein	GROS_g13269.t1	30.6
Solute carrier family 12 (sodium/potassium/chloride transporters), member 2	GROS_g01180.t1	3.0
transcriptional factor B3	GROS_g11661.t1	-5.4
Transthyretin-like family protein	GROS_g03623.t1	-3.3
Transthyretin-like protein 46	GROS_g09685.t1	-2.8
Trimethyllysine dioxygenase, mitochondrial	GROS_g12590.t1	-2.2
Tropomyosin	GROS_g03435.t1	2.9
Twik (KCNK-like) family of potassium channels, alpha subunit 7	GROS_g04094.t1	2.5
Ubiquitin-conjugating enzyme family protein	GROS_g11351.t1	2.4
Unknown protein	GROS_g07351.t1, GROS_g13249.t1, GROS_g05829.t1, GROS_g05830.t1, GROS_g12973.t1, GROS_g14226.t1, GROS_g12154.t1, GROS_g13774.t1, GROS_g13612.t1, GROS_g14134.t1, GROS_g11528.t1, GROS_g14227.t1, GROS_g13315.t1, GROS_g11984.t1, GROS_g05333.t1, GROS_g09553.t1, GROS_g05245.t1, GROS_g12599.t1, GROS_g11447.t1, GROS_g14228.t1, GROS_g09421.t1, GROS_g08885.t1, GROS_g13710.t1, GROS_g08272.t1, GROS_g14148.t1, GROS_g10493.t1, GROS_g03341.t1, GROS_g13497.t1, GROS_g13611.t1, GROS_g03754.t1, GROS_g01782.t1, GROS_g13319.t1, GROS_g09084.t1, GROS_g07611.t1, GROS_g08311.t1, GROS_g11553.t1, GROS_g07455.t1, GROS_g10479.t1, GROS_g10002.t1, GROS_g14137.t1, GROS_g14262.t1, GROS_g08958.t1, GROS_g13767.t1, GROS_g03511.t1, GROS_g08887.t1, GROS_g08332.t1, GROS_g14231.t1, GROS_g14138.t1, GROS_g01080.t1, GROS_g11190.t1, GROS_g08713.t1, GROS_g01796.t1, GROS_g01738.t1, GROS_g04899.t1, GROS_g13261.t1, GROS_g13795.t1, GROS_g09409.t1, GROS_g14251.t1, GROS_g08219.t1, GROS_g09923.t1, GROS_g13458.t1, GROS_g13270.t1, GROS_g14233.t1, GROS_g12719.t1, GROS_g08884.t1, GROS_g14036.t1, GROS_g14223.t1, GROS_g01648.t1, GROS_g02960.t1, GROS_g07335.t1, GROS_g06941.t1, GROS_g14254.t1, GROS_g11525.t1, GROS_g01716.t1, GROS_g00688.t1, GROS_g03337.t1, GROS_g01605.t1	2,1 to 19,6
Unknown protein	GROS_g01390.t1, GROS_g04932.t1, GROS_g08653.t1, GROS_g01178.t1, GROS_g08303.t1, GROS_g06185.t1, GROS_g01903.t1, GROS_g00211.t1, GROS_g00485.t1, GROS_g00210.t1, GROS_g05409.t1, GROS_g09067.t1, GROS_g10467.t1, GROS_g13188.t1, GROS_g12919.t1, GROS_g09525.t1, GROS_g08401.t1, GROS_g07009.t1, GROS_g06323.t1, GROS_g09562.t1, GROS_g09660.t1, GROS_g01786.t1, GROS_g07423.t1, GROS_g01485.t1	-2,3 to -11,4
Unknown protein, predicted to be in the cytoplasm.	GROS_g07274.t1, GROS_g03651.t1, GROS_g06594.t1, GROS_g06599.t1, GROS_g11605.t1, GROS_g10468.t1, GROS_g11725.t1, GROS_g11724.t1, GROS_g01182.t1, GROS_g11656.t1, GROS_g07373.t1, GROS_g10179.t1, GROS_g06660.t1	-2,1 to 18,9

Unknown protein, predicted to be in the cytoplasm.	GROS_g10647.t1, GROS_g03864.t1, GROS_g02747.t1, GROS_g04753.t1, GROS_g04695.t1, GROS_g04012.t1, GROS_g06136.t1, GROS_g08108.t1, GROS_g02085.t1, GROS_g07876.t1, GROS_g03164.t1, GROS_g05121.t1, GROS_g10501.t1, GROS_g09511.t1, GROS_g09245.t1, GROS_g11602.t1, GROS_g04913.t1, GROS_g05259.t1, GROS_g12083.t1	2,3 to 17,8
Unknown protein, predicted to be in the extracellular region.	GROS_g09162.t1, GROS_g12598.t1, GROS_g08092.t1, GROS_g12861.t1, GROS_g03497.t1, GROS_g06270.t1, GROS_g09419.t1, GROS_g12196.t1, GROS_g04648.t1, GROS_g09750.t1, GROS_g10389.t1, GROS_g09592.t1, GROS_g08318.t1, GROS_g05720.t1, GROS_g07580.t1, GROS_g02952.t1, GROS_g09065.t1, GROS_g05367.t1, GROS_g05366.t1, GROS_g10458.t1, GROS_g01780.t1, GROS_g04212.t1, GROS_g09554.t1, GROS_g08779.t1, GROS_g10006.t1, GROS_g11022.t1, GROS_g04661.t1, GROS_g06695.t1, GROS_g01718.t1, GROS_g09238.t1, GROS_g11715.t1, GROS_g02039.t1, GROS_g03714.t1, GROS_g01622.t1, GROS_g12714.t1, GROS_g09781.t1, GROS_g06694.t1, GROS_g06788.t1, GROS_g08772.t1, GROS_g09470.t1, GROS_g07620.t1, GROS_g12340.t1, GROS_g01949.t1, GROS_g08683.t1, GROS_g04874.t1, GROS_g06226.t1, GROS_g11109.t1, GROS_g04497.t1, GROS_g08057.t1, GROS_g12968.t1, GROS_g09435.t1, GROS_g08643.t1, GROS_g10477.t1, GROS_g07828.t1, GROS_g01956.t1, GROS_g01272.t1, GROS_g11582.t1, GROS_g00163.t1, GROS_g08189.t1	2,0 to 16,3
Unknown protein, predicted to be in the extracellular region.	GROS_g03331.t1, GROS_g11869.t1, GROS_g11112.t1, GROS_g09833.t1, GROS_g02427.t1, GROS_g01861.t1, GROS_g12924.t1, GROS_g04628.t1, GROS_g09366.t1, GROS_g07609.t1, GROS_g10079.t1, GROS_g05757.t1, GROS_g08998.t1, GROS_g08780.t1, GROS_g13006.t1	-2,3 to -11,6
Unnamed protein product	GROS_g07321.t1	14.0
Voltage-gated calcium channel, alpha subunit	GROS_g04969.t1	2.7
Xanthine/uracil permease family protein	GROS_g02031.t1	-2.3
Zinc finger C2H2 type domain signature.	GROS_g07450.t1	3.1
Zinc finger protein	GROS_g10614.t1, GROS_g05558.t1, GROS_g00603.t1, GROS_g02181.t1	2,9 to 4,5
Zinc metalloproteinase nas-10	GROS_g09810.t1	2.6
Zinc transporter ZIP11	GROS_g11274.t1	4.4
ZIP Zinc transporter family protein	GROS_g11275.t1	3.6
Zona pellucida domain-containing protein	GROS_g02236.t1	2.8

Tableau 3.7: Homozygous non-synonymous predicted variants in non-PCN populations that were located on a loci under selection; * indicates known effector genes. SeqID and Gene description corresponds to sequences ID and annotation of *G. rostochiensis* reference transcriptome (nGr.v1.1).

SeqID	Gene Description	Variant
GROS_g00043.t1	CRE-CGR-1 Protein	Val112Ile
GROS_g00096.t1	nuclear receptor NHR-34	Ile357Phe
GROS_g00119.t1	Unknown Protein	Ala289Thr
GROS_g00172.t1	Protein PDI-3	Val498Ile, Val498Ile
GROS_g00200.t1	Hypothetical Protein CAEBREN_31326	Thr1110Pro, SerThrThr1108PheThrPro, Ser1108_Thr1109del
GROS_g00225.t1	Elongation factor Tu GTP binding domain	His173Asn, His173Tyr
GROS_g00285.t1	Unknown Protein	His215Gln, His215Gln, Tyr218_His224del
GROS_g00374.t1	CBR-TYRA-2 Protein	Lys341Arg
GROS_g00441.t1	CCHC zinc finger Protein	Ala12Ser
GROS_g00460.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Met184Ile, Leu186Met, MetValLeu184IleValMet
GROS_g00473.t1	Hypothetical Protein CBG13916	Ser447Arg, Ser447Arg
GROS_g00495.t1	Alpha1,3-fucosyltransferase homologue	Arg234Lys
GROS_g00504.t1	Protein XRN-2	Asn349Asp, Asn349Val
GROS_g00529.t1	Hypothetical Protein LOAG_08316	Glu22Asp, Pro23Ser, Pro23Ser
GROS_g00555.t1	Hypothetical Protein CBG02531	Glu267Asp, Glu267Asp
GROS_g00694.t1	FKBP-type peptidyl-prolyl cis-trans isomerase-59, BmFKBP59	Gln179Glu
GROS_g00768.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Ser85fs, Thr87fs
GROS_g00784.t1	Protein AMT-1	Phe43Ser
GROS_g00802.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Asn34Lys, Asn34Lys
GROS_g00811.t1	Hypothetical Protein Y032_0093g2685	Gly46fs, Gly47fs, Leu48fs
GROS_g00821.t1	Hypothetical Protein Y032_0186g1094	Ser567Pro, Asn566_Ser567delinsLys, Asn158Asp, SerSerAsn156ArgSerAsp, SerSerAsn156ArgSerAsp
GROS_g00823.t1	Hypothetical Protein AaeL_AAEL013319	Cys50Arg
GROS_g00863.t1	unnamed Protein product	Ser78Pro, Ser78Leu, SerPro78ProLeu
GROS_g00929.t1	Histone deacetylase 11, partial	Asn340fs, Asn340Ser, Asn340Ser
GROS_g00954.t1	Cytoplasmic trna 2-thiolation Protein 2	Ser379Asn, Ser379Thr, SerSer379ThrAsn
GROS_g00969.t1	Protein F54D5.7	Thr51Ala
GROS_g01019.t1	Protein B0361.9	Asp119Glu, Asp119Glu, Arg117_Asp119del
GROS_g01127.t1	CBR-LIT-1 Protein	Thr52Ala, Thr52_Gln53delinsAlaAlaAla, Thr52Ala
GROS_g01177.t1	unnamed Protein product	Asn80Ser, Asn80Lys, Asn80Ile
GROS_g01224.t1	Hydrolase	Val241Ala, Val241Ser
GROS_g01234.t1	Protein RGS-2, isoform c	Gln104Arg, Pro103_Gln104insProGlnProGln, GlnGlnGln104ProGlnPro
GROS_g01271.t1	Larval opioid receptor	Ser619Pro
GROS_g01340.t1	Protein RSKN-2, isoform a	Pro37_Pro38delinsLeu, Pro38del

GROS_g01365.t1	Major facilitator superfamily MFS-1 domain containing Protein, partial	Gln327Lys
GROS_g01367.t1	CBR-ACR-17 Protein	GlyAsp581ValGly, GlyAspArg581ValGlyLys, Asp582Gly
GROS_g01384.t1	Protein CBR-RPY-1	His137Gln
GROS_g01497.t1	Ulp1peptidase	Asn264Ser
GROS_g01535.t1	Protein C25H3.11	Gly2410del, AsnLeuGln2495AspLeuHis
GROS_g01601.t1	DHHC zinc finger domain containing Protein	Ser252Gly
GROS_g01633.t1	Cuticule collagen	Gly163Asp, Gly163Val
GROS_g01647.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Asp184Asn
GROS_g01658.t1	C2 calcium-dependent membrane targeting and Copine domain containing Protein, partial	Arg9Gln, Arg9Pro
GROS_g01681.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Asp88Asn
GROS_g01763.t1	Hypothetical Protein BRAFLDRAFT_99981	Ser180Thr, Ser180Pro
GROS_g01778.t1	Hypothetical Protein CBG22514, partial	Val354Ile, ValVal354IleIle
GROS_g01790.t1	Protein RGS-7, isoform e	Ala321Thr, Ala321Thr, Ala321Ile
GROS_g01809.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Ala536Glu, Ala536Asn
GROS_g01849.t1	Hypothetical Protein ASU_10239	Gln289His
GROS_g01945.t1	putative cytochrome b5	Ala36Thr
GROS_g01955.t1	Hypothetical Protein ASU_09879	Val144Ile, Val144Leu
GROS_g01966.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Thr18Ser
GROS_g02008.t1	Unknown Protein	Lys90*, Lys90Glu, Leu91Phe
GROS_g02032.t1	Copine family Protein	Ala2831Glu, Ala2831Leu
GROS_g02084.t1	Hypothetical Protein Bm1_25695	Gln536His, Gln537del
GROS_g02095.t1	MutS domain III family Protein	Gly543Asp
GROS_g02103.t1	Hypothetical Protein LOAG_01892	Phe83Leu
GROS_g02162.t1	PREDICTED: UPF0364 Protein C6orf211 homolog	Phe308Tyr, Phe308Ser
GROS_g02163.t1	CRE-PST-1 Protein	Pro324Leu
GROS_g02230.t1	Hypothetical Protein CBG23528	Thr348Ala
GROS_g02257.t1	Hypothetical Protein CBG09661	Ser204Thr, Ser204Thr
GROS_g02285.t1	Unknown Protein	Leu46fs, Arg45Gln
GROS_g02286.t1	EGF-like domain containing Protein	Ala948Val
GROS_g02355.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Pro5Thr, Pro5Ser, Pro5Thr
GROS_g02364.t1	General transcription factor III polypeptide 5	Ser380Asn, Ser380Thr
GROS_g02368.t1	methyltransferase-like Protein 4	GluCysProPro150AsnPheProLeu, GluCysProPro150AsnCysProLeu, GluCysProPro150LysCysProLeu
GROS_g02381.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Gly95Ala, GlyGln95AlaLeu
GROS_g02420.t1	Protein pellino-like Protein 2, partial	His208Gln, His208Arg, His208Arg
GROS_g02438.t1	RecName: Full=Probable dihydropyrimidine dehydrogenase	Asp529Gly
GROS_g02441.t1	Chorismate mutase *	His50Pro, Glu66Lys, GluGlu118LysLys
GROS_g02461.t1	Hypothetical Protein Bm1_43530	Ala366Thr
GROS_g02503.t1	Protein CBG08513	Ala79Val
GROS_g02517.t1	Zinc finger, C2H2 type family Protein	Met239Ile, Met239Ile
GROS_g02519.t1	Protein D1037.1	Ala826Val, AlaPro826ThrLeu
GROS_g02521.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Asn34Ser
GROS_g02550.t1	Hypothetical Protein ASU_14044	Ala152Thr, AlaThrIle152ThrMetVal
GROS_g02554.t1	Hypothetical Protein NECAME_08926	Ala357Ser

GROS_g02565.t1	TPR Domain containing Protein	Phe1089Leu, LeuSerPhe1087PheProLeu, LeuSerPhe1087PheSerLeu
GROS_g02598.t1	Hypothetical Protein Y032_0011g1302	Ala678_Ala679delinsThr, Ala679_Ala680delinsVal, Ala680del
GROS_g02669.t1	Calcium binding EGF domain containing Protein	Lys711Asn
GROS_g02802.t1	Protein Y73F8A.5	Pro536Ala
GROS_g02804.t1	CBR-UNC-82 Protein	Ser1995Asn
GROS_g02862.t1	ribosomal Protein L16 containing Protein	Glu203Gln, Glu203Gly, Glu203Arg
GROS_g02947.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Ile274Val, IleIle273ValLeu
GROS_g02977.t1	Hypothetical Protein CRE_14851	Lys466Glu
GROS_g02989.t1	Protein C23H4.3	IleGlu796ThrAla, Ile796_Glu797delinsLys
GROS_g03016.t1	Hypothetical Protein	Ser908Cys
GROS_g03038.t1	Hypothetical Protein CRE_03460	Glu113Lys, Asp111_Asp112dup
GROS_g03062.t1	Hypothetical Protein Y032_0016g2918	ThrSerAspIlePhePheGln114IleSerAsp IlePhePheLys, ThrSerAspIlePhePheGln114IleSerAsp AsnPhePheLys, AspIlePhePheGln116GluIlePhePheLy s
GROS_g03091.t1	Unknown Protein	His787Tyr, His787del, Val106Ile, Val106Leu
GROS_g03117.t1	Pctaire class cell cycle kinase Protein 1, isoform b	Gly151Ser
GROS_g03179.t1	Hypothetical Protein Y032_0006g2968	Pro381Leu, Arg380_Pro381insSer, Pro381dup
GROS_g03191.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Ala190Pro, Ala190Pro, Ala190Val
GROS_g03279.t1	Arginine-tRNA-Protein transferase, C terminus containing Protein	Gly400Ser
GROS_g03282.t1	neurabin Protein 1	Pro191Ser, Pro191Ser
GROS_g03305.t1	Uncharacterized Protein F10E9.4	Val127Met, Val127Leu
GROS_g03404.t1	Lecithin:cholesterol acyltransferase family Protein	Ser177Leu, Ser177Leu
GROS_g03421.t1	Nuclear anchorage Protein 1	Ser807Ala, HisSer806ArgAla
GROS_g03466.t1	Unknown Protein	Asp591His, Asp591Tyr
GROS_g03518.t1	Hypothetical Protein Csp3_JD06.003	His29fs, Ile30Phe, His29_Ile30delinsLeu
GROS_g03528.t1	NLI interacting factor-like phosphatase family Protein	Ala198Thr, Ala198Pro
GROS_g03574.t1	Protein WHT-2	Ser52Asn
GROS_g03605.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Ala223Val, AlaAla223GlyVal
GROS_g03614.t1	Protein LGC-40	Thr428Ala
GROS_g03622.t1	Alcohol dehydrogenase class 3	Thr420Ile, Ser419Thr, SerThr419ProIle
GROS_g03632.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Arg69Gln, Lys68_Arg69insGln
GROS_g03646.t1	Protein GES-1	Ala335Gly
GROS_g03681.t1	Groucho/TLE N-terminal Q-rich domain containing Protein	Asn140_Met142del
GROS_g03798.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Arg104Lys, ArgArg103LysLys, ArgArgLeuLys103LysLysLeuArg
GROS_g03814.t1	Unknown Protein	Asp315fs, Thr314Ile
GROS_g03815.t1	Protein K01B6.3	Gln500Glu
GROS_g03861.t1	Protein CYP-33C2	Gly154Glu, Gly154Glu
GROS_g03913.t1	Carboxyl transferase domain-containing Protein	Ala801Pro, Ala801Ser
GROS_g03929.t1	Protein C10E2.6	Val185Ile, Val185Phe
GROS_g03937.t1	PDZ-domain Protein scribble	Glu463Asp, Gly465Arg, Glu463_Glu464delinsAspGluArg
GROS_g03957.t1	Putative carbonic anhydrase 5 precursor	Thr4Ile

GROS_g04002.t1	Zinc transporter SLC39A7	Glu28Gly, Glu28_Ser29delinsAsp
GROS_g04004.t1	phosphoglycerate mutase family Protein	Met145Val, Met145Leu
GROS_g04008.t1	Unknown Protein	Ser148Leu
GROS_g04047.t1	Zinc finger domain containing Protein	Val568Met
GROS_g04059.t1	DEAD/DEAH box helicase family Protein	Ser661Ala, Ser661_Gly666del
GROS_g04092.t1	Goliath Protein	Leu8Phe, Leu8_Ser9insThrIle, Leu8_Ser9insThrIle
GROS_g04100.t1	Hypothetical Protein LOAG_13260	Gln84Glu, Gln84Glu, Gln84Glu
GROS_g04105.t1	Hypothetical Protein LOAG_13467	Ala683Val
GROS_g04158.t1	RNase3 domain containing Protein	Gly49Glu
GROS_g04165.t1	Acyltransferase family Protein	Ala243Val, Leu245Ile, Ala243_Thr248del
GROS_g04197.t1	Protein NPR-31	Glu238Lys, Glu238Lys
GROS_g04293.t1	Hypothetical 60.2 kDa Protein T27F2.1 in chromosome V	Arg5Lys, Arg5Lys
GROS_g04303.t1	RibonucleaseD	Ser142Ile, SerGly142AsnCys
GROS_g04313.t1	Protein CBG18914	Val201Ile, Val201Ile
GROS_g04323.t1	Pip kinase Protein 2	Pro326Leu
GROS_g04330.t1	Kelch motif family Protein	Pro353Ser
GROS_g04337.t1	Protein ZK792.7	Leu45Ser
GROS_g04358.t1	Unknown Protein	Asn191Thr, Asn191Thr, Thr207Leu, Thr207Asn
GROS_g04365.t1	Domain of unknown function DUF148	Asp166Asn, Asp166Asn
GROS_g04410.t1	Protein CATP-2	Glu1016Ala, Glu1016Thr, Thr1021Asn, Thr1021Ile
GROS_g04416.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Ala333Val
GROS_g04439.t1	piezo-type mechanosensitive ion channel component 1	Gly1564Glu, Ser2202Thr, Ser2202Ala
GROS_g04450.t1	SH2 domain containing Protein	Leu3Ile, Leu3Pro
GROS_g04456.t1	Hypothetical Protein LOAG_02913	Ser93Phe
GROS_g04466.t1	BED zinc finger	Gly154Asp, Gly154Asp
GROS_g04474.t1	Unknown Protein	Gly6Arg, Ala8_Ala9delinsVal, GlyPro6ArgLeu, Gly54Glu, GlyIle54GluThr
GROS_g04484.t1	DNA excision repair Protein ERCC-6	Lys820Asn, Asp821Glu, Asp821_Glu823del, Ser856Asn, Ser856Asn, Ser908Asn, Ser908Asn
GROS_g04496.t1	Protein CYP-13A8	Val238Phe, Val238Phe
GROS_g04497.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Met37Thr
GROS_g04534.t1	Unknown Protein	Ser680Asn, Ser680Asn
GROS_g04564.t1	Unknown Protein	Ser512Thr
GROS_g04573.t1	Hypothetical Protein CRE_04077	Ser347Phe, Ser347Ile, Ser347Val
GROS_g04599.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Thr466Ile, Pro146Ser, Pro146Phe
GROS_g04658.t1	Androgen receptor, partial	Gly278Ser
GROS_g04783.t1	GTP-ase activating Protein for Arf containing Protein	Ser351dup, Ser351dup, Ser351_Leu352insGly
GROS_g04798.t1	CBR-VHA-19 Protein	Asn29fs, Lys28Arg, Lys28Arg
GROS_g04817.t1	Skp1 related Protein 18-like *	Ala84Val
GROS_g04885.t1	Cadherin domain containing Protein	Ala1214Gly
GROS_g04895.t1	Protein GCY-14	Ser369Thr, GluGlySer367AspGluThr, Ser369Thr
GROS_g04904.t1	Hypothetical Protein Bm1_10485	Thr650Pro, Ala233Val
GROS_g04912.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Ala54Thr
GROS_g04915.t1	phosphoglycerate mutase family Protein	Pro92Arg, Pro92Arg
GROS_g04952.t1	Protein RPN-2, isoform a	Ser678Gly
GROS_g04953.t1	family with sequence similarity 31, member B	Asp472Glu, Asp472Glu
GROS_g04971.t1	Protein K01A6.6, isoform b	Ala196Thr, Ala196Ile, Ala196Thr

GROS_g05130.t1	Unknown Protein	Val352Gly, Val352_Asn354delinsGly
GROS_g05168.t1	Protein Y39A3CL.4, isoform c	Glu119Asp
GROS_g05177.t1	Ubiquitin carboxyl-terminal hydrolase family Protein *	Gly810Ala, GlyLeu810GluMet, Trp145Ser, Trp145Leu
GROS_g05200.t1	Ubiquitin-conjugating enzyme family Protein	Gly638Ala, Gly638Ala
GROS_g05267.t1	rabenosyn-5	Leu448Arg, Leu448delinsAlaPro
GROS_g05268.t1	Unknown Protein	Val115Leu, Val115Phe
GROS_g05287.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Met409Ile, Met409Ile
GROS_g05351.t1	rapamycin-insensitive companion of mtor	Ile135Thr, IleVal135ThrLeu
GROS_g05364.t1	CRE-TAG-241 Protein	Ser290Leu, Ala289Val
GROS_g05388.t1	Hypothetical Protein Y032_0015g2813	Met184Val, Met184Leu
GROS_g05433.t1	PHD-finger family Protein	AlaValGluAla1313ValLeuGlyThr, Val1314del
GROS_g05434.t1	Hypothetical Protein CBG12638	Asn183Thr, Asn183Ser
GROS_g05467.t1	DNA damage-binding Protein 1	Met600Ile
GROS_g05554.t1	Piwi domain containing Protein	Val400Ile, Val400Cys
GROS_g05587.t1	Protein Bm13690	Met11Thr
GROS_g05635.t1	regulator of microtubule dynamics Protein 1	Phe163Ile, Phe163Leu
GROS_g05669.t1	Protein KQT-1, isoform a	Phe850Leu, Ser168Phe, Lys166_Ser168delinsAsn
GROS_g05680.t1	putative calcium binding EGF domain Protein	Phe1959Leu
GROS_g05729.t1	CRE-SAX-7 Protein	Ser680Pro, Ser680delinsAlaLeu, Ser680_Ala681del
GROS_g05747.t1	Hypothetical Protein	Met111Leu
GROS_g05843.t1	Cytochrome b5	Glu70Lys, ThrGlu69SerLys
GROS_g05935.t1	Protein E01A2.2, isoform a	Met396Val, MetVal396ValIle
GROS_g05985.t1	GJ22401	AsnGlnLysGlnSer250AspGlnLysArg Ala, AsnGlnLysGlnSerLys250AspGlnLys GlnThrTrp, Asn250_Lys255del
GROS_g06000.t1	ABC transporter substrate-binding Protein	Ser322Asn, Ser322Met, Ser322Asn
GROS_g06002.t1	Poly(A) polymerase and Nucleotidyltransferase domain containing Protein	Arg342His
GROS_g06045.t1	Protein GLNA-2	Ser822Asn, SerGly822ArgVal, SerGly822HisVal
GROS_g06101.t1	CBR-MML-1 Protein	Ile41Val, Ile41Val
GROS_g06103.t1	Hypothetical Protein ASU_00751	Ser27Gly, SerCys27GlyTyr
GROS_g06132.t1	spliced leader Protein	Asp88Glu
GROS_g06141.t1	Putative arsenical pump-driving ATPase	Ser199Ala, Ser199Thr
GROS_g06214.t1	ATPase	Val317Ile
GROS_g06247.t1	Protein C05D11.7, isoform a	Gln507Arg, Gln507Arg
GROS_g06252.t1	TolA Protein	Val229Ala
GROS_g06269.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Met250Leu, MetVal250SerPhe
GROS_g06278.t1	Protein polybromo-1	Gln23_Gln25delinsHisGlu, Gln26_Gln27insLys, Gln27del
GROS_g06301.t1	jmjC domain containing Protein	Met319Leu, Met319Val
GROS_g06325.t1	Unknown Protein	Asn59delinsAspAspValAspAsp, Asn59Asp, Asn59Tyr, Gly101Arg, AspGly100GluGlu
GROS_g06347.t1	Protein UNC-130	Val35Leu
GROS_g06379.t1	Hypothetical Protein SRAE_X000091800	Glu379Asp, Gln380del
GROS_g06404.t1	Hypothetical Protein CBG20766	Ala1256Val, Ala1256Val, Asp1901Ala
GROS_g06419.t1	Cadherin domain containing Protein	Val2625Ile
GROS_g06454.t1	Protein ACS-17	Ala698Gly, AlaVal698GlyIle
GROS_g06456.t1	Unknown Protein	Glu191Asp, GluArgPro191AspArgSer, Pro193Ser
GROS_g06478.t1	Hypothetical Protein CBG12481	Ile137Val

GROS_g06480.t1	piwi domain-containing Protein	Val284Ile
GROS_g06501.t1	phosphatidylethanolamine-binding Protein	Phe65Leu
GROS_g06506.t1	WW domain containing Protein	Val696Asp, Val696Ala
GROS_g06512.t1	Probable ATP-dependent RNA helicase A	Val602Ile, Val602Leu
GROS_g06534.t1	Unknown Protein	CysTyrVal337ArgTyrAla, CysTyrVal337ArgTyrAla, CysTyrVal337ArgTyrAla, Thr137Ala, Thr136_Ala138delinsAlaThr, Thr137_Ala138insThrAla
GROS_g06641.t1	CBN-UGT-62 Protein	Tyr538Phe, AlaTyr537SerTrp, Tyr538Trp
GROS_g06668.t1	Hypothetical Protein X975_02723, partial	Ile714Leu
GROS_g06727.t1	Hypothetical Protein CBG23685	Arg815_Ala827delinsSerVal, Thr821_Thr823delinsIleThrThrAla, AlaThr822ThrAla
GROS_g06736.t1	D-3-phosphoglycerate dehydrogenase	Ala287Thr
GROS_g06737.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Asp123Glu
GROS_g06759.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Leu42Pro
GROS_g06783.t1	Zinc transporter SLC39A7	Ala220Thr, AlaAsp220ThrAsn
GROS_g06786.t1	Bromodomain containing Protein	Gln1459Glu
GROS_g06844.t1	Unknown Protein	Gln38_Thr39insPro, Gln38dup, Gln37_Gln38del
GROS_g06918.t1	CBN-EAT-4 Protein	Pro613fs, Pro613Thr, Pro613Gln
GROS_g06931.t1	myosin-8	Ser630Leu
GROS_g06989.t1	Protein SAMS-3, isoform a	Lys122Arg
GROS_g07045.t1	CBR-EGL-8 Protein	Glu486Asp
GROS_g07062.t1	NADPH-dependent diflavin oxidoreductase	Ile19Thr
GROS_g07292.t1	DEAD/DEAH box helicase family Protein	Ala488_Leu496delinsGluSerArgArgAlaProLysMet, Asn489Ser, AsnGlyLysAlaLeuProLysLeu489SerGlyGluAlaLeuProAsnMet
GROS_g07316.t1	Receptor family ligand binding region containing Protein	Ala1070Gly, Ala1070Val
GROS_g07449.t1	UDP-glucose:GlycoProtein Glucosyltransferase containing Protein	Pro787Ser, Pro787Leu, Pro787Met
GROS_g07461.t1	Zinc knuckle family Protein	Pro281Ala
GROS_g07498.t1	FHA domain Protein, partial	Asp969Glu, Asp969Glu
GROS_g07510.t1	Hypothetical Protein CBG07084	Met20fs, Gln18His
GROS_g07530.t1	Hypothetical Protein Y032_0001g390	Leu74Ser
GROS_g07559.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Ala147Ser
GROS_g07578.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Asn105Asp, GlnAsn104ArgAsp
GROS_g07617.t1	CRE-UNC-22 Protein	Ser2185fs, Ser2185Val
GROS_g07676.t1	Hypothetical Protein	Thr2036Lys, Thr2037Ala
GROS_g07764.t1	Protein GGR-2, isoform a	Gly513Asp, Gly513_Ala514insAsp
GROS_g07810.t1	Unknown Protein	Thr47Asn, ThrHis47AsnAsn
GROS_g07846.t1	Protein GNRR-6	Gln395Lys, Gln395Lys
GROS_g07878.t1	Hypothetical Protein CBG03955	Glu45Gly, Glu45Gly, Glu45Arg
GROS_g07893.t1	Anaphase-promoting complex subunit 11 RING-H2 finger	Gly80Glu
GROS_g07916.t1	SWIM zinc finger family Protein	Gly753Cys, Gly753Thr
GROS_g07925.t1	mitochondrial RNA polymerase, partial	Thr460Ala
GROS_g07966.t1	Hypothetical Protein LOAG_04289	Lys90Asn, Lys90Asp
GROS_g07968.t1	Pectate lyase 1 *	Gly212Lys
GROS_g07971.t1	PREDICTED: uncharacterized Protein LOC100907797	Glu1978_Glu1979delinsAsp, Phe1976_Glu1977delinsIle, PheGluGluGlu1976IleGluGluGln

GROS_g07972.t1	PREDICTED: uncharacterized Protein LOC100877061	Asp1480Ala, Glu1481Gln, AspGlu1480AlaLys
GROS_g07996.t1	Hypothetical Protein CBG14695	Asn1225Thr, Thr1226Asn, Thr1226del, Val1490Leu, SerVal1489PheLeu
GROS_g08005.t1	Alcohol dehydrogenase transcription factor Myb/SANT-like	Ala432Thr, Ala432Ser
GROS_g08007.t1	Ubiquitinylhydrolase 1	Val7Ile, Val7Leu
GROS_g08019.t1	Myosin tail family Protein	Leu1180Ser
GROS_g08182.t1	zgc:100814 Protein	Ser166Asn, Ser166Thr
GROS_g08236.t1	CBR-RBC-1 Protein	Asp508Gly
GROS_g08307.t1	Lamp family Protein lmp-1	Leu327Ile, Leu327Pro, Leu327Asn
GROS_g08392.t1	Hypothetical Protein LOAG_00977	Val221Ile, Val221Phe
GROS_g08461.t1	APC-related Protein 1	Val721Ala
GROS_g08498.t1	Hypothetical RING finger Protein R06F6.2 in chromosome II, putative	Ser316Asn
GROS_g08513.t1	Hypothetical Protein Y032_0049g1745	Ser11Arg
GROS_g08547.t1	Hypothetical Protein CAEBREN_07082	Met743Lys, MetPheMet741IlePheAsn, MetPheMet741IlePheThr
GROS_g08592.t1	Hypothetical Protein LOAG_16930	Thr645fs, Thr645Asn
GROS_g08838.t1	LSM domain	Gln238Glu
GROS_g08843.t1	Unknown Protein	Ser210Thr, Pro211Ala, SerPro210ThrLeu
GROS_g08871.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Val87Phe
GROS_g08880.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Asn952Ser, Asn952Ser
GROS_g08912.t1	Hypothetical Protein CAEBREN_28137	Phe1042Leu, Ser1043Asn, Phe1042Leu
GROS_g08930.t1	Hypothetical Protein SRAE_X000091800	Ser357Asn, Ser357Thr
GROS_g08956.t1	PLP synthase *	Ala280Ser
GROS_g09011.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Gly482Glu, Gly482Val
GROS_g09066.t1	Protein ZTF-2	Asn293Ser
GROS_g09112.t1	PREDICTED: tartrate-resistant acid phosphatase type 5 isoform X3	Asn107Lys, Asn107_Lys110delinsLysGlyAsn, AsnAlaLeuLysAla107LysGlyLeuAsn Pro
GROS_g09199.t1	Regulator of nonsense transcripts 1 homolog	Ala267Val
GROS_g09241.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	AspGly359AsnGlu, AspGlyPro359AsnGlyAla, AspGlyPro359AsnGluLeu
GROS_g09371.t1	MATH domain containing Protein	Ser1371Pro, SerAla1371ProThr
GROS_g09408.t1	CBR-ADT-1 Protein	His1259Arg
GROS_g09411.t1	CG14616-PC	Cys772Phe
GROS_g09414.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Leu120Ser, Leu120Phe
GROS_g09436.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Val193Met
GROS_g09466.t1	Conserved Hypothetical Protein	Leu319fs, Glu322Asp, Leu319_Thr323delinsSer
GROS_g09540.t1	Unknown Protein	Ala533Thr, AlaGlu533ThrLys
GROS_g09592.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Lys187fs, Lys187Glu
GROS_g09622.t1	Hypothetical 35.6 kDa Protein ZK1073.1 in chromosome X, putative	Thr97Asn, Thr97Asn
GROS_g09643.t1	Unknown Protein	Ter83Ter
GROS_g09740.t1	CBR-ROM-5 Protein	Lys264Gln, Lys264Gln
GROS_g09847.t1	signal peptide peptidase family Protein	Ala297Ser, GlnAla296AlaSer

GROS_g09885.t1	Hypothetical Protein NECAME_09046	Ile312Asn, Ile312Ser
GROS_g09970.t1	Cytochrome P450	Ile275Val, Ile275Val
GROS_g10000.t1	Protein PYR-1	Arg1302Gly, Arg1302Gly
GROS_g10006.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Lys77Gln
GROS_g10017.t1	Protein Bm7351, isoform a	Ser1309Phe, ArgArgArg1797GlnArgGln, Arg1800dup, Arg1800del
GROS_g10176.t1	Rnp	Ser28Leu, Ser28Leu
GROS_g10204.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Gln29Glu, Gln29Glu
GROS_g10220.t1	CBR-PGP-4 Protein	Asp110Asn, Asp110Asn
GROS_g10297.t1	Unknown Protein	Ala1283Thr, Ala1283Ser
GROS_g10400.t1	Hypothetical Protein Bm1_25540	Glu977fs, Glu977Lys
GROS_g10410.t1	Exocyst complex component 7	Lys22Thr
GROS_g10512.t1	Phosphatase regulatory subunit family Protein	Pro160Ser, Pro160Ser, Ser27Pro, MetSer26IlePro
GROS_g10572.t1	Dolichol monophosphate mannose synthase	Lys138Asn, Lys138Asp
GROS_g10740.t1	Lipase family Protein	p.Leu22_Phe23insLeuIlep.Leu22_Phe 23del
GROS_g10811.t1	CBR-ECH-2 Protein	Thr7Ala, Thr7Ser, Ala232Asp, Ala232Gly
GROS_g10872.t1	Glutathione S-transferase *	Pro64Gln, Pro64Leu, Met1773_Pro1776delinsSer, Pro1776Ser
GROS_g10892.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Asn551Asp, Asn551del
GROS_g10893.t1	Unknown Protein	Ala117Thr
GROS_g10905.t1	CRE-HOE-1 Protein	Gln627His
GROS_g11003.t1	GDP-fucose Protein O-fucosyltransferase 1 precursor	Ser314Gly
GROS_g11029.t1	Unknown Protein, predicted to be outside the membrane, in the cytoplasm.	Val346Ile
GROS_g11034.t1	Unknown Protein	Lys674Arg
GROS_g11055.t1	Protein F01G10.10	Asn499Ser, Arg498_Gln500del
GROS_g11063.t1	Hypothetical Protein Bm1_47175	Val222Met, Val222Leu, ValVal222LeuIle
GROS_g11200.t1	Beta-1,4-endoglucanase-4 *	Ile386Lys
GROS_g11235.t1	Protein UNC-80, isoform a	Ser804Leu
GROS_g11440.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Gly117Glu
GROS_g11483.t1	Ankyrin repeat and BTB/POZ domain-containing Protein 2	Asn438Asp, AsnAsp438AspHis
GROS_g11518.t1	Protein NPR-20	Ala284delinsGluProSer, Ala284delinsGluProAsnSer
GROS_g11546.t1	Rab6-interacting Protein, putative	Asn1087_Asp1094delinsLysLysGluMetAsnAsnAsn, Lys1088_Asn1096del, Lys1088_Asp1094delinsAsn
GROS_g11565.t1	Protein C52D10.1	Met670Ile, Met670Ile
GROS_g11704.t1	Protein K07C11.4	Trp211Cys, TrpLys211LeuAsp
GROS_g11750.t1	Cyclin C	Val47Met, Val47Met, Val47Ile
GROS_g11752.t1	Protein E04A4.5	Thr321Ala
GROS_g11754.t1	Protein C35D10.6	Leu102Ser
GROS_g11762.t1	Protein PRO-3	Arg34Lys, Arg34Lys
GROS_g11994.t1	Protein F49E12.7	Arg208Gln
GROS_g12001.t1	Unknown Protein	Lys1803Arg, Lys1803Ser, Thr2051Ala
GROS_g12138.t1	Unknown Protein	Asp89Asn, Ser31Asn
GROS_g12241.t1	Protein ADOR-1, isoform a	Leu37Ile, Ile78fs, Ile78Leu, Ile78Phe
GROS_g12283.t1	Hypothetical Protein ASU_07802	Pro91Leu, Pro91Ser
GROS_g12326.t1	Protein NCX-1, isoform a	Val310Met

GROS_g12355.t1	Probable molybdopterin binding domain containing Protein	Val706Ile
GROS_g12361.t1	Unknown Protein	Pro172Gln, Pro172Ala
GROS_g12365.t1	Hypothetical Protein ASU_01837	Thr93Ala
GROS_g12374.t1	Protein UNC-10, isoform a	His590Arg
GROS_g12418.t1	Hypothetical Protein WUBG_12888, partial	Ser181Thr, Ser181Pro
GROS_g12448.t1	MMS19 nucleotide excision repair Protein homolog	Ala692Asp, Ala692Pro, Ala692Pro
GROS_g12533.t1	RE24065p	Ser205_Pro207delinsLeu, Pro207Ser
GROS_g12619.t1	PT repeat family Protein	Lys77Arg, LysGly77ArgArg
GROS_g12650.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Ala79Thr, Ala79Pro, AlaPro79ThrThr
GROS_g12678.t1	CBR-CHA-1 Protein	Ile357Val, Ile357Val, Ile357Val
GROS_g12698.t1	Hypothetical Protein, variant	Ala421Thr
GROS_g12800.t1	Hypothetical Protein Bm1_21125	Asn18Asp
GROS_g12824.t1	AT08590p	Asn3Ser
GROS_g12988.t1	Protein UBR-1	Thr76Ile
GROS_g13042.t1	Protein NPR-16, isoform b	Ser513Asn, Ser513Ile, Ser595Asn, Ser595Thr
GROS_g13121.t1	Putative gland Protein G19B10 *	Arg100His, ArgLeu100SerSer
GROS_g13161.t1	Ets-domain containing Protein	Ala193Val, Glu190_Gln195del
GROS_g13240.t1	Unknown Protein, predicted to be outside the membrane, in the extracellular region.	Gln128Pro, Gln128del
GROS_g13270.t1	Unknown Protein	Tyr252Ser
GROS_g13408.t1	Unknown Protein	Gly373Asp
GROS_g13411.t1	Tyrosine-Protein kinase abl-1	Gly577Ala
GROS_g13464.t1	Unknown Protein	Gly80Asp, Gly80Asn
GROS_g13468.t1	Unknown Protein	Arg6His, Arg6Leu
GROS_g13479.t1	ATP-binding cassette	Leu97Phe
GROS_g13520.t1	Calcium binding EGF domain containing Protein	Asp265Ala, AspPro265GlyGln
GROS_g13534.t1	Unknown Protein	LeuGluAla317ThrGluThr, LeuGluAlaThrSer317ThrGluThrAsnLeu, LeuGluAlaThrSer317ThrGluThrThrLeu
GROS_g13694.t1	Hypothetical tyrosinase-like Protein F21C3.2 in chromosome I, putative	Arg574Gln, Arg599Pro, Arg599Leu
GROS_g13722.t1	Protein CDH-8, isoform e	Gly342Asp
GROS_g13821.t1	Unknown Protein	Leu98Phe
GROS_g13853.t1	Hypothetical Protein LOAG_10199	Val58Ala
GROS_g13881.t1	Unknown Protein	Lys109Arg
GROS_g13977.t1	Unknown Protein	CysSerSer16*, Ser15_Ser17delinsArgCys, Ser17_Ser18insAsnThrSer
GROS_g14157.t1	RBP-1 *	Gly90Arg, Gly90Cys
GROS_g14180.t1	RBP-1 *	Glu222Lys, GluPhe222LysSer
GROS_g14183.t1	Unknown Protein	Trp469Cys, Trp469Arg, Asn497Lys, Asn497Lys
GROS_g14206.t1	putative Prion-like-(q/n-rich)-domain-bearing Protein Protein 51	Ser332Asn

CHAPITRE 4 : Analyse transcriptomique comparative des différents pathotypes de *Globodera rostochiensis*

Ce chapitre est en préparation pour publication

Résumé

Le nématode doré, *Globodera rostochiensis*, est un nématode phytoparasite causant d'importantes pertes économiques principalement sur la pomme de terre. L'espèce est divisée en cinq pathotypes (Ro1-5), différenciés en fonction de leur capacité à se développer sur des lignées de pommes de terre différentielles, y compris *Solanum tuberosum* ssp. *andigena*, un génotype possédant le gène de résistance H_1. Ce gène est largement utilisé pour des applications commerciales et confère une résistance aux pathotypes Ro1 et Ro4. Dans cet article, nous avons étudié les différences transcriptomiques entre les pathotypes qui sont virulents et avirulents sur des pommes de terre possédant le gène de résistance H_1 afin d'identifier les gènes candidats à l'avirulence. Nous avons identifié plusieurs gènes, incluant des gènes d'effecteurs connus (Glutathion peroxydase et deux Protéase à sérine) et des gènes non caractérisés possédant un peptide signal pour excrétion, qui étaient uniques aux pathotypes avirulents. Aussi, trois variantes homozygotes non synonymes dans les gènes effecteurs (deux protéines contenant des domaines SPRY et une protéine dont l'expression est spécifique à la glande dorsale) étaient associées aux pathotypes avirulents. Plusieurs gènes d'effecteurs impliqués dans l'évasion immunitaire étaient également significativement surexprimés dans les pathotypes virulents.

Comparative transcriptomic analysis of *Globodera rostochiensis* pathotypes

Michael Sabeh[1,2], Marc St-Arnaud[2], Benjamin Mimee[1]

[1]St-Jean-sur-Richelieu Research and Development Center, Agriculture and Agri-Food Canada, St-Jean-sur-Richelieu, Québec, Canada, [2]Biodiversity Center, Institut de recherche en biologie végétale, Université de Montréal and Jardin botanique de Montréal, Montreal, Québec, Canada

Abstract

The golden cyst nematode, *Globodera rostochiensis*, is a plant-parasitic nematode responsible for important economic losses in potato worldwide. The species is subdivided into five pathotypes (Ro1-5), differentiated by their ability to parasitize differential potato lines, including *Solanum tuberosum* ssp. *andigena*, a genotype harboring the H_1 resistance gene. This gene is widely used in commercial cultivars and confers resistance against pathotypes Ro1 and Ro4. We have studied transcriptomic differences between pathotypes that are virulent or avirulent on H_1-resistant potato to identify candidate avirulence genes. We found several genes, including known effectors (Glutathione peroxidase and two Serine proteases) and unknown genes harboring a signal peptide for excretion that were uniquely expressed in avirulent pathotypes. In addition, three non-synonymous homozygous variants in effector genes were associated with avirulence, including two SPRY domain-containing proteins and a dorsal gland cell-specific expression protein. Multiple effector genes acting on immune evasion were also significantly up-regulated in virulent pathotypes.

Background

Plant-parasitic nematodes (PPN) are pathogens of great importance affecting all major cultivated crops and causing damages estimated at 157 billion US$ each year (Abad et al., 2008; Nicol et al., 2011). The golden potato cyst nematode, *Globodera rostochiensis*, is a major threat to potato production and also parasitized other Solanaceous plants like tomato and eggplant. It causes yield losses estimated at 9% of the world potato production (Turner & Rowe, 2006) and is present in over 75 countries around the world, including Canada, United States and most of the European countries, where it is considered a quarantine organism (Yu et al., 2010). The most sustainable and effective way to control this parasite is the implementation of a crop rotation program combining non-host plants with resistant potato cultivars (Bélair et al., 2016).

Wild relatives of cultivated potato including *Solanum vernei* (Mai & Peterson, 1952) and *S. tuberosum* ssp. *andigena* (Ellenby, 1952), among others, are known to be resistant to the golden nematode since a long time. However, the selective pressure exerted by these different resistances has led to the evolution of various virulence phenotypes in *Globodera* populations. Outside of their center of origin in the Andes, populations of *G. rostochiensis* are classified in five pathotypes (Ro1-5), according to their capacity to grow on different potato lines. This classification was established in 1977 (Kort et al., 1977), but further studies revealed some discrepancies, probably explained by the fact that some nematode virulence genes are heterogeneous and some potato lines harbor polygenic resistance, thus yielding inconsistent results (Nijboer & Parlevliet, 1990). These authors suggested a reconsideration of the original classification and concluded that the only reliable recognizable pathotypes were Ro1/4, Ro2/3, and Ro5. Although the 1977 classification remained unchanged, this demonstrates the current lack of knowledge on the issue and the need for further studies on pathotype differentiation. One of the effective differential potato cultivar, *S. tuberosum* ssp. *andigena*, harbors the H_1 resistance gene (Gebhardt et al., 1993; Pineda et al., 1993). This monogenic resistance controls the pathotypes Ro1 and Ro4 while it is partially overcome by the pathotypes Ro2 and Ro3 and ineffective against the pathotype Ro5 (Nijboer & Parlevliet, 1990). Over the years, the H_1 resistance gene has been used in the production of most of the commercially available resistant potato cultivars.

The golden nematode has developed a specialized toolkit to evade plant natural defenses. Using its stylet, the nematode injects a set of secreted proteins called effectors that suppress host defenses and change the plant cellular organization to establish a nutrition site (Anderson et al., 2010a). These effectors are mainly produced in the dorsal and sub-ventral glands but also in the amphids cells (Davis et al., 2009) and most were acquired through horizontal transfer from other microorganisms (Danchin et al., 2010a). In incompatible reactions, when an avirulent nematode tries to establish on a cultivar carrying the H_1 resistance gene, the nematode successfully penetrates the potato root but is rapidly recognized by the plant, which initiates a hypersensitive response, causing the death of the cell and the nematode (Rice et al., 1985). It is assumed that the product of the H_1 gene recognizes a specific effector protein secreted by avirulent nematodes. In the pale cyst nematode, *G. pallida*, the avirulence protein RBP-1 was shown to induce cell death after recognition by the product of the *Gpa2* resistance gene (Sacco et al., 2009). It is not known exactly how the effector set differs between *G. rostochiensis* pathotypes, but previous research has shown that polymorphism of the pectate lyase 2 effector gene was associated with host range variation (Geric Stare et al., 2011a).

The identification of an avirulence gene and the development of diagnostic markers would greatly improve recommendations for the management of *G. rostochiensis*, by using cultivars resistant to pathotype present in a location. Also, highlighting the differences in the infection process of the virulent pathotypes could allow a better understanding of plant-nematode interactions and lead to the development of more resistant cultivars. The aim of this study was to characterize the transcriptomic differences between *G. rostochiensis* pathotypes to identify genetic elements conferring virulence/avirulence on a potato cultivar carrying the H_1 resistance gene.

Materials and Methods

Root exudate preparation

Potato plants cv. Snowden were grown in perlite, in 2 L containers, until they reached about 15 cm high. Potato root exudate was collected once a week, for six consecutive weeks, by the

method of Fenwick (Fenwick, 1949). The growth substrate was soaked with tap water until saturation and the leached liquid was collected. The collected liquid was used to repeat this procedure two more times. The final collected liquid was homogenized and filtered (KenAG, D-547). Potato root exudate was kept at 4 °C in dark plastic containers.

Nematode population and preparation

G. rostochiensis pathotypes Ro1, Ro2, Ro3, Ro4, and Ro5 were compared according to their virulence status on potato cultivars harboring the H_1 resistance gene (Table 4.1). Samples were received in the dormant cyst stage from collaborators. Three hundred cysts of each population were immersed in sterilized distilled water (0.2 µm Nalgene 25mm syringe filters, Thermo Scientific) for one week and then transferred to filtered potato root diffusate (0.45µm Nalgene 25mm syringe filters, Thermo Scientific) for three additional weeks to induce hatching of second stage larvae (J2), used for the RNA extraction.

Tableau 4.1: Pathotype, virulence status on potato cultivars carrying the H_1 resistance gene, and origin of the *G. rostochiensis* samples.

Pathotype	Virulence status[1]	Origin	Laboratory of origin
Ro1	−	Canada	Agriculture and Agri-food Canada
Ro2	+	France	Federal Research Centre for Cultivated Plants, Germany
Ro3	+	Netherlands	
Ro4	−	Netherlands	
Ro5	++	Germany	

[1] − denote avirulent, + low virulence and ++ high virulence.

RNA extraction and sequencing

The five samples of J2 larvae was homogenized 2 minutes in 650 µl lysis buffer RLT Plus (Qiagen) with a 6 mm zirconium grinding bead and 200 µL of 1 mm zirconium beads in 2 ml tubes using the PowerLyzer 24 Homogenizer (Qiagen) and stored at −80 °C until RNA purification. Total RNA was extracted using the RNeasy Mini Kit Plus (Qiagen) according to the manufacturer's instruction and stored at −80 °C. RNA was quantified, and its integrity assessed using a Bioanalyzer 2100 (Agilent Technologies) with the RNA 6000 Nano kit. All

RNA samples had a RIN value ≥ 7. Libraries were generated using the TruSeq Stranded mRNA Library Prep Kit (Illumina). Paired-end sequencing was done using the TruSeq SBS V3 2x100 bp chip on a HiSeq2000 sequencer (Illumina) at the University of Montreal Institute for Research in Immunology and Cancer in Montreal, Canada. All samples were multiplexed and sequenced on a single lane.

Sequence processing and assembly

Raw reads from all populations were trimmed using TRIMMOMATIC 0.36 (Bolger et al., 2014) to remove low quality sequences (Phred < 30) and Illumina adapters. Remaining reads were mapped to the *G. rostochiensis* transcriptome (assembly version nGr.v1.1) (Eves-Van Den Akker et al., 2016) using BWA-MEM 0.7.12 (Li & Durbin, 2009), with default parameters and unmapped reads were kept for a *de novo* assembly. A genome-guided assembly was performed using Trinity 2.2.0 (Grabherr et al., 2011), with default parameters. A *de novo* assembly was also performed using Trinity 2.2.0 (Grabherr et al., 2011), with default parameters using the unmapped reads from all five populations in order to include the sequences that were absent from the reference transcriptome. Five assembly iterations were performed with Cap3 12.21.07 (Huang & Madan, 1999) to merge these two assemblies. Reads were mapped to the final assembly using BWA-Mem 0.7.12 with default parameters (Li & Durbin, 2009). Contamination assessment was performed using Blobtools 1.0 (Laetsch & Blaxter, 2017) to remove undesirable contigs in the assembly based on GC-content of sequences, read coverage in sequencing libraries and taxonomy of sequence similarity matches. Assembly properties was assessed using Quast 2.3 (Gurevich et al., 2013).

Gene prediction was done using AUGUSTUS 3.3 (Stanke et al., 2008) with *Caenorhabditis elegans* as species parameter. Sequence similarity searches using NCBI (Geer et al., 2010b), KEGG (Kanehisa & Goto, 2000a), and UniProt (The Uniprot Consortium, 2017) databases were performed to identify unknown sequences of interest.

Quantitative analysis and differentially expressed genes identification

Read counts for the quantitative analysis was performed using the CORSET 1.04 software and its default parameters (Davidson & Oshlack, 2014a). Normalization and differently

expressed genes (DEG) identification were done using the DESEQ2 1.14.1 package in R using a parametric Wald test (DE; $P < 0.01$), a normalized minimum base count means of 10 for all populations and a log2 fold change (log2FC) $\geq (\pm)$ 1(Love et al., 2014a). For the analysis, pathotypes were separated into two groups according to their virulence on potato cultivars harboring the *H1* resistance gene for DEG identification; AP (avirulent pathotypes; Ro1, 4) vs VP (virulent pathotypes; Ro2, 3, 5).

Variant analysis

Trimmed reads were mapped to the transcriptome assembly using BWA-MEM 0.7.12 with default parameters (Li & Durbin, 2009). Mapping files were used for variant calling using FREEBAYES 1.0.2 software, a Bayesian genetic variant detector designed to detect SNPs (single-nucleotide polymorphisms), indels (insertions and deletions), and complex events (Garrison & Marth, 2012b), with a minimum Phred score of 30 and a minimum coverage of 10. BAYESCAN 2.1 (Foll & Gaggiotti, 2008) was then used to identify loci under natural selection, using allele frequencies as input and considering the pathotype virulence status on potato cultivars carrying the *H1* resistance gene as the principal factor for selection. The method is based on locus-specific genetic differentiation (F_{ST}) outliers to detect candidate markers under selection. We used the "plot_bayescan" function in R provided with BAYESCAN to calculate *a posteriori* odds threshold (FDR=0.05) and on a probability greater than 0.91, as this threshold indicates a strong evidence for selection (Jeffreys, 1998), to select outliers associated with the virulence status of each populations. Three analyzes were performed, giving different random initial seed values and only outliers present in all three analyses were kept.

The impact of these genetic variations on protein structure and cellular localization was evaluated to target the variants susceptible to lead to a difference in phenotype (Haegeman et al., 2012; Mitchum et al., 2013a). SNPEFF 4.3 (Cingolani et al., 2012) was used to determine the impact (silent, missense or nonsenses) of the mutations while SIGNALP 4.1 (Petersen et al., 2011) and PHOBIUS (Kall et al., 2004a) were used to predict the presence of signal peptide cleavage sites and to determine the cellular localization of the proteins.

Results

Sequencing and read processing

Exposure to potato root diffusate induced the hatching of J2 larvae from all pathotypes. A similar proportion of J2 larvae hatched in all samples (data not shown). RNA sequencing yielded 264M of 100 bp paired-ends reads for all five *G. rostochiensis* pathotypes, with a mean of 52.6M paired-end reads per sample, spanning from 46M to 57M (Table 4.2). Decontamination removed 12.6% of the contigs. The final assembly contained 135,911 contigs (Table 4.3), and the proportion of reads that successfully mapped to the assembly was similar for all populations spanning from 98.05% to 99.10% (Table 13).

Tableau 4.2: Sequencing yield and mapping statistics per populations.

Pathotypes	Sequenced Reads (M)	Mapped (%)	Variant counts
Ro1	46	98.80	416 043
Ro2	57	98.26	520 291
Ro3	52	98.56	505 796
Ro4	57	99.10	466 794
Ro5	51	98.05	529 637

Tableau 4.3 : Transcriptome assembly statistics

Contigs (n)	135,911
Contigs (n) (>= 1000 bp)	49,092
Largest contig (pb)	20,069
Total length (Mb)	146.3
GC (%)	47.3
N50 (pb)	2,307
Predicted genes (n)	104,476

Quantitative analysis

A total of 78,374 transcripts are common to all five pathotypes (57.95%). This core transcriptome represents 76% of the assembly length and from 49% (Ro5) to 75% (Ro1) of all sequencing reads. The AP (Ro1-4) shared 1,462 transcripts that were absent from VP (Ro2-3-5) while VP shared 5,865 transcripts not found in AP. Among these, 10 had a conserved protein domain matching to known effectors as well as a signal peptide for excretion, three were present only in AP (Glutathione peroxidase and two Serine proteases (Trypsin and Lipase)) and seven in VP (two Cathepsin, Cysteine protease, Galactosidase, Pectate lyase 1 and two Serine protease) (Additional materials: Table 4.5). Also, 24 unknown predicted genes only present in AP had a signal peptide for excretion and 111 only in VP (Additional materials: Table 4.6).

Differentially expressed gene analysis resulted in 4,377 differentially expressed genes (DEGs) (Fig. 4.1), 249 were up-regulated and 4,128 down-regulated in AP when compared to VP. Subsequently to BLAST, UniProt and Conserved Domain search, 50.3% of DEGs still remained unknown, including 213 transcripts harboring a signal peptide for excretion, all down-regulated in AP with fold changes spanning from -4 to -279. The most differentially expressed genes that could be identified were regulatory and maintenance component (e.g. GRPc12269, GRPc18978, GRPc27205, GRPc3467 and GRPc11733) with a mean fold change of 96 in VP (data not shown). Also, 60 DEGs coded for known or putative effector proteins (Fig. 4.2), 12 being up- and 48 down-regulated in AP. Fold change varied from -3 (RBP-4 protein) to -149 (Endoglucanase) for the down-regulated effector transcripts, and from 2 (putative dorsal gland cell-specific expression protein) to 11 (RBP-4 protein) for the up-regulated ones.

Figure 4.1 : Clustering of *Globodera* pathotypes based on the expression value of 4,377 differentially expressed genes. Differential analysis was performed by group comparison according to their virulence on cultivars carrying the H_1 resistance gene (Ro1, 4 vs Ro2, 3, 5).

Expression values are score given by PHEATMAP function (PHEATMAP 1.0.10 package in R) calculated using normalized read counts, as calculated by DESeq2.

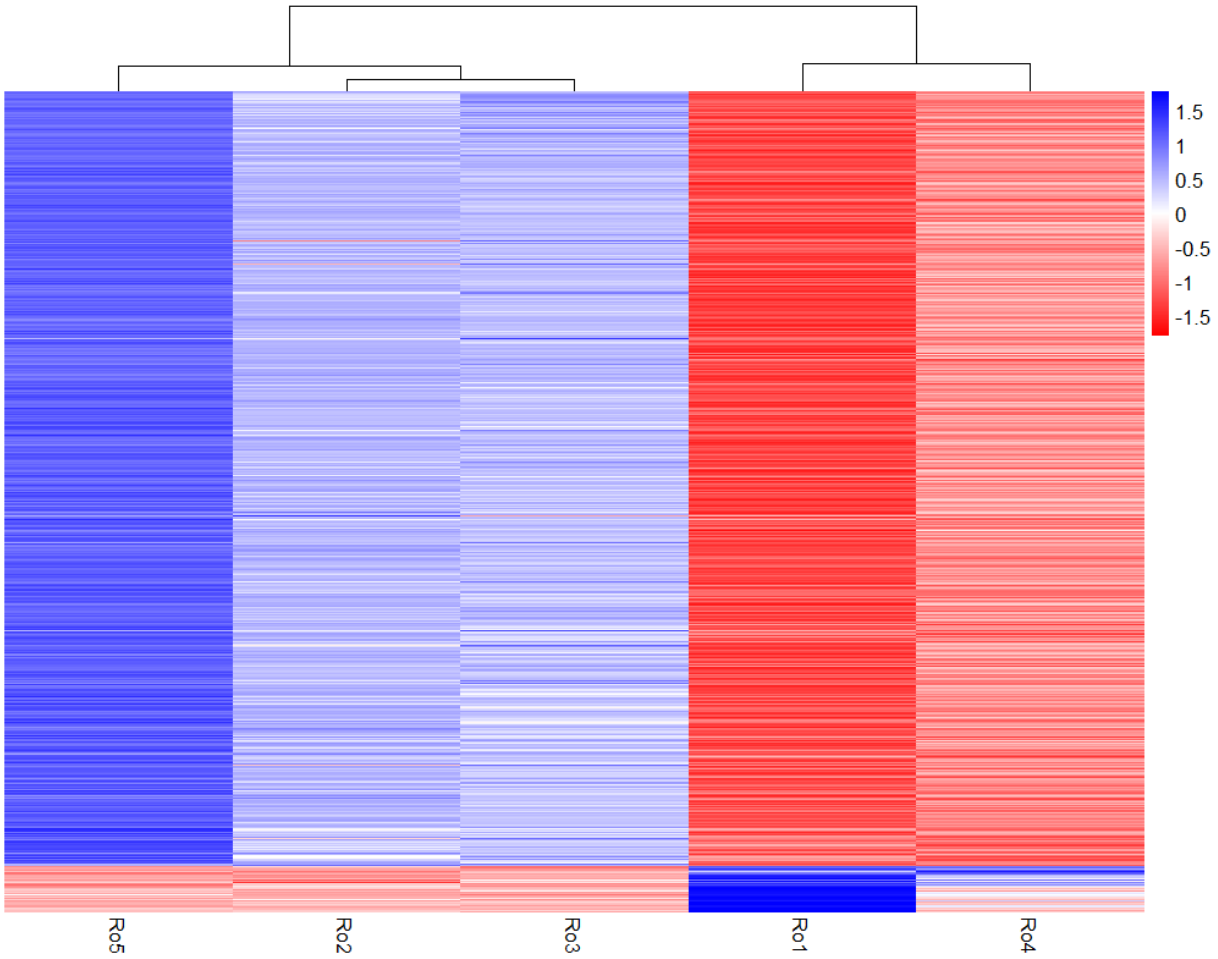
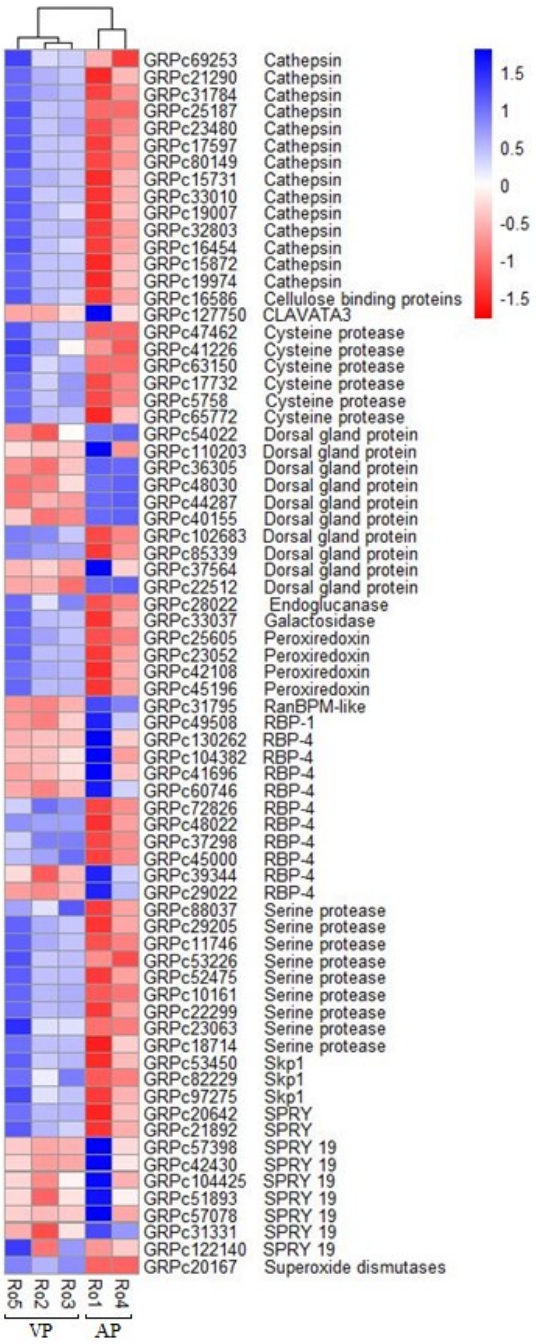


Figure 4.2 : Differentially expressed effector genes between *Globodera* pathotypes. Differential analysis was performed by group comparison according to their virulence on cultivars carrying the H_1 resistance gene (Ro1, 4 vs Ro2, 3, 5). Expression values are scores given by PHEATMAP function (PHEATMAP 1.0.10 package in R) calculated using normalized read counts with DESEQ2.

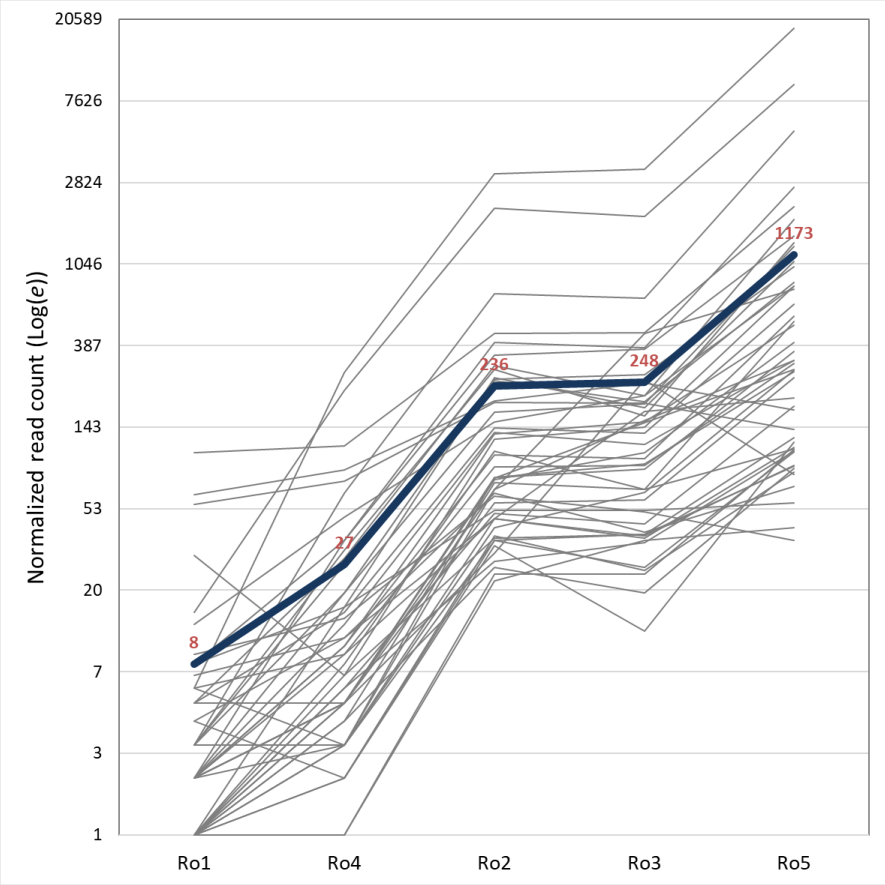


Differentially expressed genes in *G. rostochiensis* Ro5

Most of the up-regulated effector genes in VP were also significantly more expressed in Ro5 when compared to Ro2 and Ro3 (mean fold change of 3.8, Fig. 4.3). This was also true for other

DEGs, including regulatory genes (e.g. GRPc20260, GRPc21388) and 25 unknown genes harboring a signal peptide for excretion**.**

Figure 4.3: Expression level of up-regulated effector genes in *Globodera* pathotypes. Differential analysis was performed by group comparison according to their virulence on cultivars carrying the H_1 resistance gene (Ro1, 4 vs Ro2, 3, 5). Genes normalized read counts, as calculated by DESEQ2, are shown in grey, the average of which is shown in blue.



Variant analysis

Gene polymorphisms analysis identified 620,127 variants across all pathotypes. As expected, since the reference transcriptome was made using sequencing data from Ro1 pathotypes, the patothypes Ro1 had the least number of variants when compared with the assembly (67.1%), followed by Ro4 (75.3%), Ro3 (81.6%), Ro2 (83.9%) and Ro5 (85.4%). Among all variants, 70.9% were single nucleotide polymorphisms (SNPs), 15.9% insertions or deletions (Indels), and 13.2% complex events. A variant analysis using a Bayesian inference method highlighted 452 genetic variants under selection associated with the virulence status of the populations, including 358 non-synonymous variants, 8 of which were coding for known or putative effectors. The effects of these gene variations were missense (277), frameshift (32), stop gained (24), stop lost (22), disruptive inframe insertion (2), and conservative inframe insertion (1). Fifty non-synonymous variants under selection were homozygous for the reference allele in all AP and homozygous for the variant allele in all VP or the opposite. Only three homozygous non-synonymous variants were coding for known or putative effectors and one for an unknown gene harboring a signal peptide for excretion, including two predicted frameshifts occurring in VP (Table 4.4).

Tableau 4.4: Homozygous non-synonymous predicted variants in effector genes for which AP (avirulent pathotypes; Ro1, 4) allele differs from VP (virulent pathotypes; Ro2, 3, 5) allele.

SeqID	Gene description	Variant*
GRPc119280	Dorsal gland cell-specific expression protein	Leu*86*Pro
GRPc61928	Unknown	Ile*189*frameshift
GRPc78465	SPRY domain-containing protein 19	His*43*Arg
GRPc81549	SPRY domain-containing protein 19	Lys24 frameshift

* Amino acids preceding the figure are from the reference allele, the figure is the position and amino acids following the figure are those of the alternate allele.

Discussion

In this study, we posited that transcriptomic particularities linked to differences in virulence on cultivars carrying the H_1 resistance gene, could be identified using RNA sequencing data from virulent and avirulent *G. rostochiensis* populations. The H_1 resistance gene is widely used in commercial cultivars for resistance to potato cyst nematode but only confers resistance against the pathotypes Ro1 and Ro4.

The H_1 resistance gene induces a hypersensitive response (HR) in cells surrounding the nematode feeding site, thus limiting the nutrient flow and leading to the nematode death (Rice et al., 1985). It was previously shown that virulent *G. pallida* populations could evade this type of plant defense mechanism using a single amino acid polymorphism in the case of *Globodera pallida* RBP-1 protein, which is otherwise recognized by the *Gpa2* resistance gene (Sacco et al., 2009). This induction of HR following the recognition of a specific effector protein is called effector-triggered immunity and is the subject of a co-evolutionary arms race between hosts and pathogens (Diaz-Granados et al., 2016; Dodds & Rathjen, 2010). Screening for altered excreted effector proteins is therefore the best way to identify avirulence genes and was the aim of this research. The absence of a transcript or the presence of polymorphism in an effector protein of virulent pathotypes could explain why they do not trigger a HR response in plants harboring the H_1 resistance gene.

It was previously shown that effector genes were rapidly induced following hatching of juvenile (Duceppe et al., 2017b). Here, more than 75% of the genes contained in the reference transcriptome were expressed in the J2, including most of the known effectors. It is therefore probable that any avirulence gene would be expressed in this stage. The most striking result obtained while comparing the transcriptomic profiles between AP and VP was the high number of transcripts that were specific to each group. Among these, three transcripts coding for an excreted effector were only found in AP, a glutathione peroxidase (GpX) (GRPc65027), and two serine proteases (GRPc53895 and GRPc83341). Surface GpX protein usually metabolizes hydroperoxide substrates and should protect the parasite from damaging reactive oxygen species produced by the host defense mechanisms, but the role of secreted GpX is still uncertain (Jones

et al., 2004). Serine proteases are most likely implicated in the extracellular digestion of dietary proteins (Koritsas & Atkinson, 1994). In addition, 24 unknown genes with a signal peptide for excretion were only present in AP, which have the potential to be uncharacterized excreted effector genes. The assembly also contained seven genes coding for an excreted effector that were unique to VP, two cathepsins (GRPc51158 and GRPc95309), a cysteine protease (GRPc46107), a galactosidase (GRPc9097), a pectate lyase 1 (GRPc74716), and two serine proteases (GRPc63267 and GRPc65602). Two serine proteases were only found in AP and two only in VP, which suggests that VP may be using different serine proteases, in order to avoid recognition of the host's immune system. Those four serine proteases had no match on the *G. rostochiensis* genome (Eves-Van Den Akker et al., 2016) and only shared limited homology.

A non-synonymous variant in an avirulence gene could also prevent recognition and HR response in VP, like in the case of the *Gpa2* resistance gene mentioned above. Among genetic variants showing evidence of selection, 358 were non-synonymous and among these, 50 were homozygous across all pathotypes of each group, but only three were putative effector genes: two SPRY domain-containing protein 19 and a dorsal gland cell-specific expression protein (respectively GRPc78465, GRPc81549 and GRPc119280), in addition to one unknown gene with a signal peptide for excretion (GRPc61928). Two of these genes (GRPc61928 and GRPc81549) have a frameshift variant and most likely results in a non-functional protein. If the H_1 resistance gene interacts directly with one of them, the variant present only in VP could prevent the protein from being recognized. SPRYSEC effectors contains a SPRY domain, several dozen different version being found in *G. rostochiensis* in response to positive diversifying selection, and they are thought to be modulators of plant defense responses and could have other undefined roles (Diaz-Granados et al., 2016). SPRYSEC GpRbp-1 is the protein triggering the HR response in presence of the *Gpa2* resistance gene when a proline variant is present at position 187 (Sacco et al., 2009). Although the “dorsal gland cell-specific expression protein” is not a validated effector protein, the presence of a signal peptide in addition to the expression localization strongly suggest an effector function. A proline variant is, here, present at position 86 for AP as a replacement for a leucine. It was previously shown that genetic variants may be better conserved in geographically close populations than in distant populations

of the same pathotype (Boucher et al., 2013; Mimee et al., 2015b), showing the importance of validating these results in several populations from diverse locations.

Interestingly, in *G. rostochiensis* the degree of virulence is different between the virulent pathotypes Ro2/3 and Ro5 on H_1-resistant cultivars. Usually, when a single dominant gene is overcome, the resistance is entirely lost. This suggests that the resistance mechanism may be more complex, involving interaction with more than one resistance gene, and prompted the analysis of the differential expression between these pathotypes to see if the expression level of some genes could correlate the observed phenotypes. Several genes coding for effector proteins were significantly differentially expressed between AP and VP. Six dorsal gland cell-specific expression proteins were up-regulated in AP against two up-regulated in VP; four RBP-1/4 proteins were up-regulated in AP against four up-regulated in VP. The *G. rostochiensis* reference transcriptome contains multiple RBP proteins, as well as secreted SPRY domain-containing proteins and it is not yet fully understood what differences between them are and whether there is a link between them. Research about these effector genes should help us find out why some of these genes are overexpressed in AP while the others are overexpressed in VP. Besides, most of the DEGs coding for effectors were significantly up-regulated in VP. In total, 20 cysteine proteases (including 14 putative cathepsins), 9 serine proteases, 4 peroxiredoxins, and 3 Skp1 were up-regulated in VP. Although genes up-regulated or unique to VP cannot be recognized by the H_1 resistance gene because no hypersensitive reaction occurs, they can nevertheless contribute to immune evasion. For example, cathepsin, which two were unique to VP and 14 were significantly up-regulated in VP (mean fold-change of 32) is a cysteine protease believed to plays key roles in parasitic nematodes, acting in reproduction, development, invasion, pathogenesis and immune evasion (Li et al., 2015). More than 50 different cysteine proteases are expressed in the case of the ruminant parasitic nematode, *Haemonchus contortus*, being the most active protease of the excretory products and it is believed to be essential for the life cycle or pathogenicity of many parasitic nematodes (Bakker et al., 2004; Yatsuda et al., 2006). The presence of a peptide signal for excretion in nearly all of them (except for GRPc21290 and GRPc69253) also support the expected involvement in pathogenicity and their abundance, like in the case of *H. contortus*, suggests that they must be associated to virulence on cultivars carrying the H_1 gene. Several peroxiredoxin where also up-regulated in VP; this

antioxidant enzyme may be overexpressed here to compensate for the absence of a glutathione peroxidase, an enzyme with a similar function only found in AP. Normalized read counts for all DEGs coding for cathepsins and other cysteine proteases revealed that on top of being up-regulated in all VP compared to AP, they were significantly more expressed in Ro5 than in Ro2 and Ro3. This overexpression was about 90-times more for Ro5 compared to AP and 18 and 16-times more respectively for Ro2 and Ro3. This pattern was similar for most of the transcripts coding for serine proteases, peroxiredoxin and regulatory genes and is consistent with the higher virulence of pathotype Ro5 on cultivars carrying the *H1* gene. The reproduction factor - or female index – of this pathotype was shown to be more than 4-times higher than Ro2 and Ro3 (Nijboer & Parlevliet, 1990). This could indicate that the differential regulation of a specific group of effector genes confers greater virulence to pathotype Ro2, 3 and especially Ro5. The presence of a large quantity of up-regulated proteases in the VP also suggests that these enzymes may potentially be involved in the degradation of proteins in the signaling cascade induced by the activation of the HR response or of the product of the *H1* resistance gene. This would prevent the cell death associated with HR and also explain the greater virulence of the pathotype Ro5.

Although they are genetically very similar, populations of *G. rostochiensis* from different pathotypes are not able to parasitize *H1*-resistant potato cultivars with the same success. We showed that different transcriptomic profiles could potentially cause the different *G. rostochiensis* phenotypes. Of particular interest, three effector genes and several putative effectors were only expressed in avirulent pathotypes. In addition, non-synonymous homozygous variants were found in four genes that could potentially be the avirulence gene associated with the *H1* resistance gene. This work will require further investigation, including validation in other populations, and validation by agroinfiltration. Simple PCR validation in other populations could allow the characterization of the resistance mechanism and the development of a diagnostic tool. These results provide good candidates for being recognized by the *H1* resistance gene and as targets for the development of a diagnostic tool.

Acknowledgments

This work was funded by Agriculture and Agri-Food Canada (AAFC) and the United States Department of Agriculture (USDA) through a Coordinated Agricultural Project (CAP) titled "Risk Assessment and Eradication of *Globodera* spp. in U.S. Production of Potato," supported by award #2015-69004-23634 from the National Institute for Food and Agriculture (https://www.globodera.org/).

The authors also thank Björn Niere (Federal Research Centre for Cultivated Plants, Germany) for providing us with isolates of cyst nematodes, Étienne Lord for advices with bioinformatics analysis, Nathalie Dauphinais and Pierre-Yves Veronneau for technical support.

Authors' contribution

MS, MSA, and BM conceived and designed all experiments; MS performed experiments and bio-informatics analysis; BM contributed to materials/analysis tools; MS wrote the manuscript with contributions from all the other authors. All authors read and approved the manuscript.

References

Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M. C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T. R., Markov, G. V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Segurens, B., Steinbach, D., Tytgat, T., Ugarte, E., van Ghelder, C., Veronico, P., Baum, T. J., Blaxter, M., Bleve-Zacheo, T., Davis, E. L., Ewbank, J. J., Favery, B., Grenier, E., Henrissat, B., Jones, J. T., Laudet, V., Maule, A. G., Quesneville, H., Rosso, M. N., Schiex, T., Smant, G., Weissenbach, J., & Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature biotechnology, 26*(8), 909-915. doi:10.1038/nbt.1482

Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., & Singh, K. B. (2010a). Plants versus pathogens: an evolutionary arms race. *Functional plant biology : FPB, 37*(6), 499-512. doi:10.1071/FP09304

Bakker, N., Vervelde, L., Kanobana, K., Knox, D. P., Cornelissen, A. W. C. A., Vries, E. d., & Yatsuda, A. P. (2004). Vaccination against the nematode *Haemonchus contortus* with a thiol-binding fraction from the excretory/secretory products (ES). *Vaccine, 22*(5), 618-628. doi:10.1016/j.vaccine.2003.08.025

Bélair, G., Dauphinais, N., & Mimee, B. (2016). Evaluation of cultural methods for the management of the golden nematode (*Globodera rostochiensis*) in Quebec, Canada. *Canadian Journal of Plant Pathology, 38*(2), 209-217. doi:10.1080/07060661.2016.1189969

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Boucher, A. C., 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*(1), 75-82. doi:10.1016/j.ympev.2013.05.020

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin), 6*(2), 80-92. doi:10.4161/fly.19695

Danchin, E. G., Rosso, M. N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., & Abad, P. (2010a). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences, 107*(41), 17651-17656. doi:10.1073/pnas.1008486107

Davidson, N. M., & Oshlack, A. (2014a). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol, 15*(7), 410. doi:10.1186/s13059-014-0410-6

Davis, E. L., Hussey, R. S., & Baum, T. J. (2009). Parasitism Genes: What They Reveal about Parasitism. *Plant Cell Monographs, 15*, 15-44. doi:10.1007/978-3-540-85215-5_2

Diaz-Granados, A., Petrescu, A.-J., Goverse, A., & Smant, G. (2016). SPRYSEC Effectors: A Versatile Protein-Binding Platform to Disrupt Plant Innate Immunity. *Frontiers in Plant Science, 7*(1575). doi:10.3389/fpls.2016.01575

Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews Genetics, 11*(8), 539-548. doi:10.1038/nrg2812

Duceppe, M. O., Lafond-Lapalme, J., Palomares-Rius, J. E., Sabeh, M., Blok, V., Moffett, P., & Mimee, B. (2017b). Analysis of survival and hatching transcriptomes from potato cyst nematodes, *Globodera rostochiensis* and *G. pallida*. *Scientific Reports, 7*(1), 3882. doi:10.1038/s41598-017-03871-x

Ellenby, C. (1952). Resistance to the Potato Root Eelworm, *Heterodera rostochiensis* Wollenweber. *Nature, 170*, 1016. doi:10.1038/1701016a0

Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G., Da Rocha, M., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szitenberg, A., Grenier, E., Montarry, J., Mimee, B., Duceppe, M. O., Boyes, I., Marvin, J. M., Jones, L. M., Yusup, H. B., Lafond-Lapalme, J., Esquibet, M., Sabeh, M., Rott, M., Overmars, H., Finkers-Tomczak, A., Smant, G., Koutsovoulos, G., Blok, V., Mantelin, S., Cock, P. J., Phillips, W., Henrissat, B., Urwin, P. E., Blaxter, M., & Jones, J. T. (2016). The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology, 17*(1), 124. doi:10.1186/s13059-016-0985-1

Fenwick, D. W. (1949). Investigations on the emergence of larvae from cysts of the potato-root eelworm *Heterodera rostochiensis*; technique and variability. *Journal of Helminthology, 23*(3-4), 157-170. doi:10.1017/s0022149x0003248x

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*(2), 977-993. doi:10.1534/genetics.108.092221

Garrison, E., & Marth, G. (2012b). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907 [q-bio.GN]*.

Gebhardt, C., Mugniery, 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*(5), 541-544. doi:10.1007/bf00220911

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., & Bryant, S. H. (2010b). The NCBI BioSystems database. *Nucleic Acids Res, 38*(Database issue), D492-496. doi:10.1093/nar/gkp858

Geric Stare, B., Fouville, D., Sirca, S., Gallot, A., Urek, G., & Grenier, E. (2011a). Molecular variability and evolution of the pectate lyase (pel-2) parasitism gene in cyst nematodes parasitizing different solanaceous plants. *J Mol Evol, 72*(2), 169-181. doi:10.1007/s00239-010-9413-4

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data

without a reference genome. *Nature biotechnology, 29*(7), 644-652. doi:10.1038/nbt.1883

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics, 29*(8), 1072-1075. doi:10.1093/bioinformatics/btt086

Haegeman, A., Mantelin, S., Jones, J. T., & Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene, 492*(1), 19-31. doi:10.1016/j.gene.2011.10.040

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research, 9*(9), 868-877. doi:10.1101/gr.9.9.868

Jeffreys, H. (1998). *The Theory of Probability*: OUP Oxford.

Jones, J. T., Reavy, B., Smant, G., & Prior, A. E. (2004). Glutathione peroxidases of the potato cyst nematode *Globodera rostochiensis*. *Gene, 324*, 47-54. doi:10.1016/j.gene.2003.09.051

Kall, L., Krogh, A., & Sonnhammer, E. L. (2004a). A combined transmembrane topology and signal peptide prediction method. *J Mol Biol, 338*(5), 1027-1036. doi:10.1016/j.jmb.2004.03.016

Kanehisa, M., & Goto, S. (2000a). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res, 28*(1), 27-30.

Koritsas, V. M., & Atkinson, H. J. (1994). Proteinases of females of the phytoparasite *Globodera pallida* (potato cyst nematode). *Parasitology, 109*(3), 357-365. doi:10.1017/S0031182000078392

Kort, J., Stone, A. R., Rumpenhorst, H. J., & Ross, H. (1977). An International Scheme for Identifying and Classifying Pathotypes of Potato Cyst-Nematodes *Globodera Rostochiensis* and *G. Pallida*. *Nematologica, 23*(3), 333-339. doi:10.1163/187529277x00057

Laetsch, D., & Blaxter, M. (2017). BlobTools: Interrogation of genome assemblies [version 1; referees: 2 approved with reservations]. *F1000Research, 6*(1287). doi:10.12688/f1000research.12232.1

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, Y., Wang, K., Xie, H., Wang, D.-W., Xu, C.-L., Huang, X., Wu, W.-J., & Li, D.-L. (2015). Cathepsin B Cysteine Proteinase is Essential for the Development and Pathogenesis of the Plant Parasitic Nematode *Radopholus similis*. *International Journal of Biological Sciences, 11*(9), 1073-1087. doi:10.7150/ijbs.12065

Love, M. I., Huber, W., & Anders, S. (2014a). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Mai, W. F., & Peterson, L. C. (1952). Resistance of *Solanum Ballsii* and *Solanum sucrense* to the Golden Nematode, *Heterodera rostochiensis*, Wollenweber. *Science, 116*(3009), 224-225. doi:10.1126/science.116.3009.224

Mimee, B., Duceppe, M.-O., Véronneau, P.-Y., Lafond-Lapalme, J., Jean, M., Belzile, F., & Bélair, G. (2015b). A new method for studying population genetics of cyst nematodes based on Pool-Seq and genomewide allele frequency analysis. *Molecular Ecology Resources*, n/a-n/a. doi:10.1111/1755-0998.12412

Mitchum, M. G., Hussey, R. S., Baum, T. J., Wang, X., Elling, A. A., Wubben, M., & Davis, E. L. (2013a). Nematode effector proteins: an emerging paradigm of parasitism. *New Phytol, 199*(4), 879-894. doi:10.1111/nph.12323

Nicol, J. M., Turner, S. J., Coyne, D. L., Nijs, L. d., Hockland, S., & Maafi, Z. T. (2011). Current Nematode Threats to World Agriculture. In J. Jones, G. Gheysen, & C. Fenoll (Eds.), *Genomics and Molecular Genetics of Plant-Nematode Interactions* (pp. 21-43). Dordrecht: Springer Netherlands.

Nijboer, H., & Parlevliet, J. E. (1990). Pathotype-specificity in potato cyst nematodes, a reconsideration. *Euphytica, 49*(1), 39-47. doi:10.1007/bf00024129

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods, 8*(10), 785-786. doi:10.1038/nmeth.1701

Pineda, O., Bonierbale, M. W., Plaisted, R. L., Brodie, B. B., & Tanksley, S. D. (1993). Identification of RFLP markers linked to the H1 gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Genome, 36*(1), 152-156. doi:10.1139/g93-019

Rice, S. L., Leadbeater, B. S. C., & Stone, A. R. (1985). Changes in cell structure in roots of resistant potatoes parasitized by potato cyst-nematodes. I. Potatoes with resistance gene H1 derived from *Solanum tuberosum* ssp. andigena. *Physiological Plant Pathology, 27*(2), 219-234. doi:10.1016/0048-4059(85)90069-4

Sacco, M. A., Koropacka, K., Grenier, E., Jaubert, M. J., 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*(8), e1000564. doi:10.1371/journal.ppat.1000564

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics, 24*(5), 637-644. doi:10.1093/bioinformatics/btn013

The UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic acids research, 45*(D1), D158-D169. doi:10.1093/nar/gkw1099

Turner, S. J., & Rowe, J. A. (2006). *Cyst nematodes*. Wallingford, Oxfordshire: CAB International.

Yatsuda, A. P., Bakker, N., Krijgsveld, J., Knox, D. P., Heck, A. J. R., & de Vries, E. (2006). Identification of Secreted Cysteine Proteases from the Parasitic Nematode *Haemonchus contortus* Detected by Biotinylated Inhibitors. *Infection and Immunity, 74*(3), 1989-1993. doi:10.1128/IAI.74.3.1989-1993.2006

Yu, Q., Ye, W., Sun, F., & Miller, S. (2010). Characterization of *Globodera rostochiensis* (Tylenchida: Heteroderidae) associated with potato in Quebec, Canada. *Canadian Journal of Plant Pathology, 32*(2), 264-271. doi:10.1080/07060661003740322

Additional materials

Tableau 4.5 : Sequences of 10 effector genes only present in J2 larvae transcriptome of avirulent or virulent pathotypes having a signal peptide for excretion

>GRPc9097_Galactosidase (Virulent pathotypes only)

ATCAGTTGCATGATATCATTGGATCCCATTCACCTCGTTCATAGTGTGAGAAGATTATCGAAGAACATTCACACAACCAATCCTCATATCACCGCACACCTCCTTTCCATCTCAACAACTTGAGGTCCTCATCGATTAATTCCTCTGATTGGATTTTGAGCTAGGATGGTTCGATCTCGTTTGATACGATTACTCCTACTGGCATGTA
TCATCTGCGCCATACTCGTGGCCACATATGCCAAAGACGACCATCATCATGAGAAGCATCGGCCACGAACGAAACATACGAATGAGAAGAATTCACGCGTGTCAGAATCTCATCCATCACACACAACTCATCATGCGACTCCCACGACTGATGCATCAACTGCAACTGGCTCCATCTACACCTCTGAAATGGCAACTGGCTCACC
TGCACCATCATCATCCACTCCCGATCAACAACCATCAAGCACAACCCAAAAAACCACGACCACGGCCGATGCATCCTCTGCAACTGGCTCATCCGCATCACTATCATCATCCACTCCCGATCAACCGTCCTCACCACCGCCTCGCATCGGCATTCCTGTTCCACCTGTTCGTACGCCGCGTGTGATGCCTTCGCCTCCCAAAGGGCG
CCATGTGCCGCCGACGCAGCCACAACAACAGCCACCTCCTCTGCCCTCGTATGATCTTACCACTCATCGCACGACCAGTGCTCCAGTGAATCAGCCACCATTGGTGCCGAATCCGCCGCCGCAGGGGAATGTGCCACCCTTCAGCCAAGGCGCCGCCGCTTTCGCCACCTGTTGCTGGTGGTGAAACAACTAAAGCCACACCCAT
ACCGAGCACAACCACACCACCACAGGGGAATGTGCCGCCTTCAGCCAAGCCAACTTCACCACCTGTCGCGGGTGGCGAGACTACCAAGGCGACACCCGTCCCAAGCACAACCACACCACCACCACATGTCGACGATCATCAACGATTACATTTGGATGGCAATGCGACCGCTGCTGCCTTCGGTACCACGGACTTGAAAAGTT
CAGAGAGTGCCTCTGATGACGATATTATCGTCTCCTCAGATGATCCACATGAATTACCGATTCCCTCAAAGTCGAGTAATTCACAAGGAGGCATTCCTTATCCTGAGCCACATGATCCCATCATTCATGTCACACCACCACCGACGGGTGGATGGAGAGCAGGCGATCTCAGACAACGTGAACAGAAGATTGCAGAACAACAAC
AAGCGGCGAATCCAACTGCAACTGAGACAGCATCAGAGGATTCACCTAAATCGAACTCCTATATCTCACCGCATGGTCCTTCTGAGGAATTTCGTCGTAATGCCTCGAATTGGTTTGCACCACCTGATCTTCAAGGTCCGATGCCTTTAGTGGATGCTTTGACGTATCGTCCATCTGGCGCACCCGACTATTGGATGCCGGATGC
CGCCCATATTGAGATGTTGAACGAGCGTCATCGTGAGATTCATAATCGTCATGTTCATAAGCATGGTGCTGGTGCACGTACGACACGTTTAGCTGCGGTGGATTCGGCACCGATGGTGGAAATACGTGGTCAACAATTGTTTGTGGAAGGCAAACCCTTCTGGATTCAAGGCATCTGTTATAGTCCAGTACCGATTGGTGAGA
GTGTTTCCTTCGATCCGAAGGGTGATTATTTCACACCTGATTATGCTTATATCTGGCAACGTGATTTACCCTTAATTAAAGCCATGGGCGCGACCACTCTACGTATCTATGGCTGGAAGGCTGAGGCGGATCATACCATGTTCTTAGATGCTGTCCATGCTGCCGGTCTGAAGGTACTGGTCACATTCTATTTAGGTGATGCGAAA
GCGAATCCAGTCTTCACACAGGATCAACGGAATCAAATTCTGATCAATTTCGTGGCGCAAGTACATAAATACAGGGATCATCCTGCCATCTTGATGTGGAGTTTCGGTAATGAATTGAATGGTGCTTGGAACGGATTTATGAAACAATTTAGTGATGCCTTTACTTGTTGGTGGCAAGCTGGCTGTGCTGGTTATAGTGATACCA
ATTCGGATTGTCATTGGCAGAGTACATGTACTTACTATCAACTCTTCTCCTGGATCAATTCCGCTTGTCGTGCAGCGAAGATCGTCACACAACGACCCATCATCTCCGGGTTTGCTGATGTCGATTATATGGTCGGACCAACACCCTGGTTAGATAAAGTCGCGCGCTTTAATTATCTTCTACCTGATATGGATGCTTGGGCTGTCC
AGTTGTATCGTGGTTGGACGTTTGGTGGTTATTTTGGGATGTATCGTGGTGAATCATGGAAGCCGATGATTGTCACCGAATTTGGTGTTGATGCTTACAATGATGCTTGTGGGTGGCCCGAGAATAATCAAGGTGGTTGTTTCAACATGGCCGGTGAGGCGGCGGGTGGTGCCGATGATAACGGTCACTTCTGGGGCTGTGCT
GAGGGCGGTGAATGCGCCAAACCTGGCGTCGTCGCGCAGGCCGAATGGGATGTCCGCCTTGCGAACGAACTCATGGCACAGTACGTCGATCGTGGCGGAATCGTGTGGGGCGGCTTCCTCATGGCCTGGACGGATGAGTTCTGGAAGGGCGGAGGGACGCAAGATCTGTGTGCTTATCCATGTGCGACCTGGGATGAGGC
GTGGTGTCGTGGCGGCGGCATCGCAGCCTATCGACCGGGAGGGTCGGCTATGTGTAGCTGGAAGGCGCATTTCACATGTCCGAACTTTAATACCTGGTATCATGACCTATGTGGGTATTGGTTGGCGGCTGCACCCGATAATTATGTCAATGAAGAATGGTTCGGTATTACTACACTGTCTCGTGTGGTGTCCACAAAGTCA
ATCGGGATGGTGTCGATGGCGGTTATCGATTAGATTCACTCTATGTGCGTGCGGCTTATATTCAAGTGATGCAGCTCTGGTCGGGTCTGAGTGAGGTCGATACGAGTCATACATCATGCGATGCTTTACATCCATGTTGGCTCTGTAGCACCACCCACACCATCGATTATCTGAATGATGGCGCATGTGAGAAGGAGTGTATGA
TTCAACATCTACGTGTTGGACCTGGTAGTAGTGTGCCACAGACAACCAAATCGATTGGTAGTACCTTATTACACGGCTCCGATTCACCTGATTCATTTTGGAGTAAGTTTGGTTGCCGATCTTCATTGGTATGGGCATCATCCTACTTGCCTTATTAGTCTCCATCTATTTATGGCATAGACGACGAACAGCGAAATTAACTGCGC
TACAACAACCATTAGTACGACAAGGGTCTGGATCATAGCCACTCCTCTGGTCTTTTCCTCCTTTTTGAATGTCTTATTCGCTCGCCCCTTTTTGTCGCTTTATGTATTAATTGCATATTTTTCACACAG

>GRPc46107_Cysteine_protease (Virulent pathotypes only)

GAAGTTTGTAGCTGTTGTCTTGCTCGCAGTAGCAGTCGCTGTTATGGCTATGCCAAATGGTTACGACCACGTTAGACCAGCCGCCCTCGCTGAGGCTGAAGTCCAAAACATGTTCTCCGCCTTCATGGCTGAACACGAACGTGTATACGTCGGTGCTGAAAAGCAAAAGAGATTCGAAATCTTCAGAACTAACGTCGAAAAGGC
CCGTTACTTGAATGCCATCACCAGAAACAACGTTTATGGTGTCACCAAGTTCAGCGACTTGAGCGAACATGAATTCAACAGCATGTACCTCATGAAGCCAAAATCACCAGAGGAAGTCAGAAAGATGTTCCCAGCCAACACTCAAATCGTCAAGTCTTTCCCAACCAACGATTTACCAGAATCATTCGATTGGAGAGAAAAGGG
TGCCGTAACTGGTGTCAAGAACCAAGGTGCTTGCGGTTCATGCTGGGCTTTCTCCGCCACTGGTAACATGGAAGGTCAAGCCTTCATGAAGACAAAGAAGTTGGTCAGCTTGTCTGAACAACAATTGGTTGATTGCGATCACGAATGTGTCCAATACGAAGGTCAACAATCATGCGACGATGGTTGCAATGGTGGTTTGATGT
GGGCTGCTTTCCAACACGTCCAAAAGAACGGCGGTATCATGACCGAAGCTCATACCCATACGAAGGTATTGACGGTGACTGCAGATACAAGAACACCACCATCGGTGCCAAGGTATCTGGCTACAAGATGCTCCCAACCAAAGATGAAGATTTGGCTGCCTGGTTGGTTGCCAACGGTCCAGTCTCAGTTGCCATCAACGCT
GAGTGGCTCCAATTCTACTTTGGAGGTATCTCTGATCCATTATGGTGCAGCCCAACTGCCCTTGACCACGGTGTCTTGTTGGTAGGTTTCGGCAAGGGTAAGACCTTGTTGGGCTTCGAAGTTAACTACTGGATCGTCAAGAACAGCTGGTCAACTGGCTGGGGAGAGTCTGGTTACTTCAGAATTGCATGGGGCAAGTGCGG
TATCAACACTGTTCCATCCACTGCAGTTATTGCCTAAATATTTGAAATGATGGAAA

>GRPc51158_Cathepsin (Virulent pathotypes only)

AATGATCAAGGTCGCCCTTCTCTTGCTCGCCTTAGCTGCAGTCATCTTTGCCGACGCTGCATACTTTGCACAACCAGCCATTCACCACGATTTAATCGCTGAAATCAACGCTGACGAAACCTCAACCTGGACTGCTGGTGTTAACCCAGTTTTCAAGGACATGACTTTGGCTGACGTCAAGAAGCAACTTTTGGGTGCTCGTGCCA
TCAACTTGATCCCAGGAAACTACCAGCACACAGCTGGCAAGTACGAACTCCAAGACTTACCATCGCGTTTCGATGCCAGACAAAGTGGGGCTCTTGCATTCACCCAATCAGAAACCAACAACAATGCGGTTCATGCTGGGCATTCGCCGCCTCAGAAGTTCTTTCTGACCGTTCTGCATTGCCGGACGCAAGGAATGTCGGTGT
TTTGTCACACAATACCTCGTCTCATGCGACGACTCTGACTACGGCTGCGATGGTGGATATTTGGCCAACAGCTGGCAATTCCTCCAAAACACTGGTATCCCAACTGATGCCTGCTACTCATACACCTCTGGTGGAGGAGATTCCGGCCAATGCAAGACCACTTGCGACGACGGATCTGCCATCAAGACTTTCAAGGCTGACAAG
GTTGTAACTCCAACCACTCCAGCTGACATCCAAACTGAAATCTTCAACAACGGTCCAGTAGAAGTTGCCTTCTCTGTATACCGTGATTTCTTGTCCTACACCTCTGGTGTCTACTCACACAAGACTGGTGGTCTTTTGGGCGGTCACGCCGTCAAGGCCATTGGCTGGGGTGTCTCCAACGGAACTCCATACTGGATTATCGCCAA
CAGCTGGGGAGAATCCTGGGGACAGAACGGTTTCTTCTGGATCAAGAGAGGCTCCAACGAATGCGGTATCGAATCCAACGTCATTGCTGGTTCTCCAGCTTTGTAAATGGTTTAGTAAACTGTTTCTTGTG

>GRPc53895_Serine_protease (Avirulent pathotypes only)

GTTGACTGGAACTATTCATCCGTGGAGGTGGAGAAAGGTCGTGACAATTGTCTATCATCAAAAATTTGGTTAAGCCATGGAACGTAATATTTTATATCAGTATAGACACCAGGATAAGCAGCATGAGCGCATTTAACACCCCAGCTAACTAAGCCGGCAAGATTTCCTTGAGAGTCATAGAGAGGACCGCCTGAGTCGCCTTG
GCAAGAGTCTTTACCTGCGTCGCTGGCACAAAACATGCCGTCTTTAATATGAGCCTTCAGAGGATCACAACTTTGAAAAATATATGCGCTTTTCGCAGTGTGCATTGGGCAGAGCCTTGGAATCTCGTAGTACCCCAGCCAATTACCACAGATGGATCGCCGATTTGTTTGTTAAAGTGCTTAAGCGCGCCAATTTAATACTTT
GATCAGCGGGGAGAGGAATATCTGAACTGAAAAACAAAAGAGCAATGTCGTTTTCTAATGTCCTTGCATTGAAATTCTCGTGTGGGACTATCTCAATCGGTTTGACAGTGACGCCAGATTCAATGCAGCGATTGAGAACACCCAATTTTGCCGCAGAATTCTGAACAGAGACATCCCCATCAACACAATGAGCAGCAGTCAGA
ATGGCTCTCGGATGAATGAGGGAACCGCCGCACATATAACCTCCCAAACCGAGAAAAGCAATCGACGGGAATTCGTTGTGTGTTGCGTGTGCCCGCCCACAATTCTGGAAGATATGTTGACATGATCATCATGATCGGGAAGAACATAATGAATTCCTTTCAAATTGCTTCCTTTAACAAAGGACTCCTTTTTAACAAATGCAG
CAGTGTCATCATCAGCCACACCACCACATACCATCACACACAACACAACACCCAAAACCCTCAACCACATGGCGACA

>GRPc63267_Serine_protease (Virulent pathotypes only)

GACACTGCACAACGTCAACACCGCACGACGAATAACAAGACAGAGACCAGACCAACCAACCAAGCGGAGATCGAACATCACAACCAGACACTCGTTGTCTGAACCATCACCTCTAACGCATCACCCCATTCAAATTAATCCCCATACGTTCACCCATACCATACCCAATCCTACAACCTGTTTCAAATAATTTTTAAGTCCGAAGT
CCGCCAGCTTCCTCCATCAGCCCCGCCTGCCTACTCATTCACTTCATATGAAGATGGTTTCATCCGCACACGTGCTCCGCCTCCCTATTCTCATCTGCCTCCTCCCCCTCCTACTATCGCTATCACCCGTGGTCGACGCCACGCTATCGACGGACGTACCATTTCCGCCGAGAAACATCACCACTCCATACCGTTGGAACAAAGCTTC
ATTCTTCTTAATGTCTTCGGTCCTGCCAATCTGACCGGTTCGATTGCTATTGAGCCGGCTGTATTCGGTCAACATTTCCCTATTACAGGTACAACAGGTGGTAGTGTTCCTTATTTTACTGGCGAAGTTGTCCTCCCGAATGCTGCGGATGATATTCAAATGTGTTCCGACAATGTTCAACAACACCGAGTTTGGTGATACTTTC
ACAGGTAAAGTGTTACTTATTCGTCGTGGTGCTTGTGCGTTCGGTTCGAAGGCACAGAGAGCTCAAGCTGCCGGTGCAGCTGCAATGTTGGTATATGGTTGCCCTACGGTCG

>GRPc65027_Glutathione_peroxidase (Avirulent pathotypes only)

GTGCCCATCACAACATATATATGTCAGCGGGAGTAGTATTATCTACTACTCGTGATTGTCCCCTCGCAGGAAAAGGAGAAGAAAAATATTTGTATTGAATTGTCATGGAATTCTTCTGTTGATTATGAATGCGTGGTGCTGGTTATTCATTGAGGAAGGCCTCGACGTCTGATATGATTTCGAGGGGATTCACAGAGCTACGATA
ATGTTTACACAACACCAGAACGACTCACAAGGTACTTTGCAAAATTCCAATCAATTTGTTCACGATCTGAGGCTGTTTCAAGAATCGATACACAGGATGAGTCTCAGGGCCATTGACACTGATCTTAGACATCATGAGGAAATTGACTTTGTACTTTGAGCAGAGCACATGATGCTCTCAGAAGAGCCGGGCTCTTGTGAAAGA
AACTGATTGCAAGGAAAAGCAATGATCTCTAAACCACGCGGATGGCACTGTTCGAACAATTCGACCAATTGTTTGTAGTTTCGATCAGTTAATCCACAGTAGGATGCCACATTCACCACTAATGTGACTTTCCTTGAAATCGGGAGAAATCGATCTTTTCACCTTGAAGATCTTCTGCTGATAAGTCAAAGAAAGACTTGTATTC
ATGCTCAAGATTATCAGAACCCTGAACCAGATTGGCGAGATAAGATAACATTGCAGGCAATGATGGGACCAAGGCGACAACCAACACAAAAAAAA

>GRPc65602_Serine_protease (Virulent pathotypes only)

GTCGTTTTACACCCATAAATGATCATGGCTGCCGCACCCATCGCCTGTCGCTTACTCGCCTTCAAACCGAATGGGCATGTTCCACGACGAACGAGGAGGACCTTGCCAACTAATGACTGGCCGTATTCATCCGTAGAGCGAATGGTGGGATCACGACATAATTCAATATCATCAGATAAATTTGGAACAACAATCTCATTCAGGA
AATAAGGAATGTTAGTTCCAACTGTTGTGTTGGTAATTGGGAAATGTGTACCAAATACAGCCGGCTGAGTGGCGATCGAACCAATTAAAGAAGCAGGACCATCAATATTAAAGAAGGTGGAAGCACGTGTCGAACGATAGGGTGCGGTGGATGCACGAGGAGGAAACGTTAAAGTTGCCGTCACTTCGAGTGCAGTCACTA
AGAGGACCAGTGCGGAGGCGATGAACGGAGAGAGTCGCATGTCGACGAACTAGGGGGTGGTAAGTATTAAGAATAAGGATTATGGATGGGTGACTATTAAGGATGTGTGATATAATGTTCACAGTCTCTCTTGTCTTGGAGTTGTCAGCGCTGCTGCGCTGATTGTGTTGTTTGTGTTGTTAACAACACGTTGACTCTGTGACT
GTTAACTTCGTCGACTTCCGATAATTAAAAAGACAACACAATTAGTCTCAACTCTCACGAGTGAGACCAGCGCTGATGTGATGCACTGTTATTG

>GRPc74716_Pectate_lyase_1 (Virulent pathotypes only)

AGCTGATCGTTTTTCCAATGATTCACGCACACATAAATTGCGTATACATTGTCAGACCGATCGGAAGCCACTGATTTTCAAAAAAAAAGATTATTTGAAAATGTTAGCTAATCTCGTCACATTGGTCGCTTCTTTGGCCATCTTTGTGCTTTATAATTCCGCTATCGCCGGCCCTCTGAAGTGCCAAAGCGGAGTCTCCTTTCTGCT
GGAAGGCATATACGATAGCAAAATTTCGAACTGCACATATCCAACTGCCCAATATTGTTTCGAGGCCACCTGCGCCGCATTGGACCAGCCTGGAAACATTCGCAAATATTGGACATGCGTTGCTTCCACTGAAACAGGAGATTGTGCTCGTTATGAGAACACAATGATGAAATTGACCAAATCCGAGAAGGTTTCTTGCCAGTG
CAAGTATGGCGAAAAAGGGAAAAATTTGACAAATGTGCCGTCTTCTGCTGCCGAAAACAGACTTCGGTGCAAGGCTGGGGCAATGAATGCATCAGGAACAGGCAAAACATATTACAGCGATTGTATTGATGGATACAACTATTGCTATGCGGCTTACTGCGCGAAAGGGAAGGAGCATGTTCGAACATT

>GRPc83341_Serine_protease (Avirulent pathotypes only)

ATCTCCTGTACTGACTTCCATGTGTCATAAAAACCTTTATGAACTTTACAGCCAGAGCAATATGGATAATTCACTTGCGTCACATCAAAATCATTCAACCAATTCACAATGTCTGTGAATCCACGGAATGACACAACAATTTGAGCATAGTCAGGATTATAACCCACGTACACGAGCTCGCCGTCCTTAGTAAAAATGCGTGTCGT
GACAAAACCACTCACTGAAGCATTTTGCTGACACCATTTGCAATCCCACAACTGAATGGAGGAATCAGGGCAATATGCGGCATAAGAATACAAACAATAAGTATATCCTGTATCATAGCTCCACACGGTTGGATCAACAGGTTGTGCAGAACCTAAACCCACTGCCACCACACACAACACAACACACCCACAACATTGTCTC
ACAAATGCCTTATTAGGGTTCAATTTAGCAAAAAATAATGGAAAAAAGTGAAATAGAATGTAAAAAAAAAAATTAAAAAAAAATAAGGAAAAAAATGGAAAACGTTAATTATATTA

>GRPc95309_Cathepsin (Virulent pathotypes only)

GTTCATGACCTGGGTGGAGAGCGAAAGGTTCTTCTTGAGGAAGATGCAGTTGCGGTCGGACATGGCCTCGACGGATCCGAAGGCCCAGCATGAGCCGCAGTCGGACTGGTCGCGAATGCGCTGAATGTCGGAGCAATTGGGGAAGGCAGTGCGGCAGTCGTAGGAGGCGGGCACGGCAGCGAGGTCGACCTCGAACTCG
GTCTTCTCGGGGAGCTTGGGGCCGCCAACAAGGCGAGTGCCCATCATGGCCTTCACCTGGGCCTTAGTCTTGCGCGAGATGGCGCCGTTGTCGAGGCGGGCCTTCCAGCCAAAGTTGCCGTTGTCGATGAAGTCTACAAGCTCCTGCTTGACGGCAGGCTCATTGTCGCCGGCGAGGGGCTGAGGGATGACCTGAACGGCCA
GAGCGACGGCCGCGAGGCACACCAAGAGGATCAGCGCACGAG

Tableau 4.6: Sequences of 135 unknown genes only present in J2 larvae transcriptome of virulent or avirulent pathotypes having a signal peptide for excretion

>GRPc35063 (Avirulent pathotypes only)
GTTACATCGATGACGCCAAATTGTCTCAATTTCTCGAATCGGTGGCCGCACAGAGGTTGTCTTCCCCGGCTGCCACTTGTTTGAAGCTATGTAAAGCATTTTTGGAATCCAAACAATAATGAATTGAAAAGCTACTTCAATCATTTATTTTCCTTCATGGTCAGACATTAGCTTGGACTGAAATCTTTTCGGTTGTGTGAGTATGAT
CAAAAGGAACCCCATAATAAAGACATTCGAAAACAACACCCTCACTTTGACAGACAGTTCCTGTGATAGATTGGTCAGGAATCACAGTAGTGGTCTGCGGAGTTGAGTAGTCGACCAATTGAATCAAACCAGTTCGATGCCATAACTTGGTGTCTGAAGGGGTCGACAAGTCTGGGCTCTGAGCGCCCATTGCGTAGTTATAG
AGCATCGATGACGTGTACACGCCCGGTGAGGATTCGGTGACATTGTAAAGTCCGTTTGTCACATAAAAAGTGTAGGCATAAAGATTACGACCATCATTGGGTATCAAAAGTTCGGACAATTGGAACCCCAAATCCACGAAGAGTGACAAAGGCTTGATCATAGCGAACGCGACCAGTAGTGGTAATAGTTGCGCGATCGTGGA
TCACATCAGAATGGGTTCCTGATTTCTCCACAACTAATCGAGGCGAGGCAACGAAATTAGCATCAGATGTATTTGCAGCAATTGTGCTGAACGACACAAGTACATTGTAATCAACTGCCCAATTAAAGTCATCAGGAATATCTCCGGGTGGGGTTGTGCGGGCAACCTCAAAGACGAGCGTGCAAGAAAGCGTGACGCTATTC
AATTTAGAACAGGTCGAAATATCTATTTGGAGAGTCGAGGGCCAAATACTGCTGGACGTCGCATCACAAGTACCTTGAGGTTGCTTCTCAGGGTCGCAACTCATTTCCAAAGAAAAGCCATTCAAAATAGCGTCATACTGATCAAAACCAGCAGGGATTGGAAGGGATAACGGAATGGTATACGAAGAATCAGCATAAACTG
ATTGACTTTCATTATCATCACGAGTGGCAAGATCAGAAAATTGAAAACTGTAAGAATGACCACTCGTCAAAACAACCTTCGGAGAAAAAGCAGCCGAATAATAGATTATGGGATTTGCAATGTCTCCATCTGCTCCCGGTGTGAATGTCATGCTGGCGGTGACTTTCCCGGTCCATTCGCCACTATCAGAAATCGTATGCTGTTG
ACCTTCCAATTGAGCCTCCAAATTTCCCAATTTATGCCCATAGTGGAGGCCAGTCAAAGGGACAGTGCGCTGCCAGACATACTCGAACCCATTCCAAGTAAAGAATCCATTATTCGTTGGTTGTGTCTTTTGGAAGGAGGCCGCAACCCAAGCACCTAACAACAGAAGAACCAACCGCATAGCCAAAAAGATCGGAAG

>GRPc35849 (Avirulent pathotypes only)
GTTTTATCTGGTTATGGCGAAAGTTATTCTATCTTTGTTAGTTCTGGGCAAAGTTTGCTCAGGAGCCACAGAAACAGCTAATGATATTCTTAGTTTTTCAGATCAGAAGACCGTAGGTCTATCACCAACGGTTTCCCACTTCACTTCTATGCTCGCCACTGTCAAGGACTTCCTAGCTGTATTACGCAATAAAGAAGCCGAGGAAA
ATAATGTGCGTAATACCGCTGTATTTTTCGTTGAACGGCCCGTAGAGAATGCCCAACATTTTCTTCGTGGAAATAATCAGCAGAGGGTCTCTCCTCAAGACAATAAGCCATGCTCTAAGGCTAGCTCTATGAATGAAAACTCTCACAAATCTATGAATACTTTGGATGTTGAGAATGAGAGTGATTCTGCTGCATCTGAGTCGCA
GTATGGTTATTATTCATATGGTTCTGGAGGTGATGATGGTTCAAGTTCCACTGCTTATTCTTATGGTTATTATGATGATTTTTCATTTGATGATGACTTTTGGAACTATTATTCTTATGGATCTTATGGTTCATCATATTCTTCTGGCTCCTCCTCAGAAAGTTCTAGTTATGGTTCAACTGGGTCTTACAGTTATAGCAGCTCGTCATC
AACCACATCCACAGAAAGTTCTGGAACTTCTTCTGATAGTTCTGCTTCTACAAGTACCGGTTCTGAAAGCGACGACGCGGGTGGGTCATCTATGACATTCTCTGACTCAGGTTCAGCTGCAGATGGAGCCTCAACATCTCAACAAGATTCCTCCACTGCTGCAGAAGACGCAACAAGTGATGTGGATACAGACAGTGAAATTTC
ATCTTCCAGCAGCCTATCTGGCTTGCGTCATATCCATCATGCTCGGCCGCAACGCAGCGACAGACCGAACCAAAAACAAAGGCCACTTGATTATCAACATTTGCAAAAAACTATCAAAATGGAGAGACCACTCGATAGCCGCTGGGAGAAGACAAATAATGTACGTGACAGAACTAAAGACAATTCTATTATGACTACAAATTC
AAATACGGCTGCTGCTACTGGTGCTGCTGAGTCTACAGAAGAGACCGTGGAGTCGCGTCCAAGACGTTCTTATCTACGCAACGTCGAGAAACTGTAATAAACTTGCTCACTGAGTAAAAGTGTGTTATACGTTGTGAAGGTGGTATGTGAAAATGTTTATAGCAAATCTGCTGATAAGTAGGCCATGTTTTAATTTTGTAAATA
TTATGGGAGAAACTGACATTTAATGCTGAGTAAGGACATTATGTTGAGATATAGTTCAATTCTGCTGCTGTTGTGTTGTCTTTTCGAGGTACTTTGTATAATATTGCATCATGTAAAAAGGAAACTTAATCAATAAACAGACTACCGGAAAAAAAAA

>GRPc37651 (Avirulent pathotypes only)
CTCGCGATTTGCTCGATTGCCCCGACACTCCCCATCCCATCCCCACCGTCGCCATGCGTTCTCTACTTCTCTGCCTCCTCGGCCTCGCTTCGGCTGGCCTGCTCGTCGCTGGTGCGCCGATCAGCGGTCAGTCACCCCTCTGCCCATCGCTGCCGGTTCAACTCGAGTTGCGTACCTCGCCGACTCTGAAGAACAAGTATCAACTCG
TCTTCGCGTCGGTCGATCCCAACAATCATCAATCTGCTGATTTCGTCGATATCCATTATCGCATCAACAACATGGACTCTGTGAACATGCGCGTTGCGACTCAGGAGGCCATCACCGAGCAGCGTCAGGCGGCTAAGCAGGCGCAGGTCGATGGTATCTCGCTCCAGCCAGGTGACGCTCTCCGCGCCTATGCTACCTACCAGG
CGAAGGGCTATGCTTGCGACACTCCAGTTCACACCTTCTCTGCGCCAGAGTACGAGACCGACTCGGAGCTCGATTTCCACAACATCAACAACATGCGCAACTTCCGTGGCATCAACAACAACATGAACAACCGCAACCTGCGCAACGTGAACACTCGCCGCATCGAGAACGATTTCATGTCTCAGGCTGTTCTGGAGGAGACG
GCTCAACAGGGCGCCTTATGCCCGGCGATCGTCATCGATCAGGATATCGTGAAGATCCAGGGTCTCGACAACACTTACATGTTGACCTTCGAGAACACGAATCCGGCGAAGCAACTTCACTACGTCGACCTCCACTTGTCGACTGATGCTGCCAACGGATGGGACAACCTTCGTCTCGCTTCGGACATGCAGCTCCACATGGCT
CACAAGGTCATGCAGCCAGGCATCGTCATCAAGCCGACTGAGCAGCTGAAGTGGTATTGGTCATATCGCGCGACCGCTCCGAACGGCGAAGTCGTCGACTGCACCACGCCAGTCTCGGCAATCTCGCCACACCAGATCACCCACGAGATCACCCAGCAGCAGCTGAACGCGGCCGCTCGCTCGTAAGTGCTCCCGCTGATCTC
ATCAAGCAGGTTGGTCGTCGAAGATGACCCGACTTCCGACTCCGGTCTCAACACATTCTCGCTCGATGCTCGAAGTCCGATTTGATCGTCGAGCAAGAGTCTGCCTGATCCGTCTTACCTACACCCTCTCCCCATCCTCGCTGCGTATCGGCGCCCGCCTCTATTCTATTTCGGATTCGAACGACAGTTCACTCTTTCACTGATTCCT
CCGCACCCTTTCTTGTTCCGAGCATCCGACGAGCGCGCATGCCACCCCAAGTCCGTCGATCGATATTTGATCGATTCGATTCGTGTTCGCTTCGCAG

>GRPc38028 (Avirulent pathotypes only)
CCTCGAGAGTCACTTTCTCACACACATACACATGGCAGCGCGCGGTTTGTGGGGTGTGGCTTTGTTGTTGAGTTTGGTGGTGGTGTCGGTAATCGCCGATGACTGCTTGGATGAATCAGTGGCTGTGAAAGCAGTATCAGGAGCAAGTAAATTGGGAACATCGGTTTTGGTGTTGGGCCAACAAGCGACCACGAGCTCGGCG
GCCGCGACTGTGTTTCAGAATTTTGCTCCCGTGGCGGCGAGACGAACTTCCCCTTATTTTCAGAATTTGAATTTGGCATATGCAAATAACAACATTTACATTGTTGCTCGTTGCTATGCATTTTTTCCTGATTGTGGCAATCGTTTGATTTATCTTATGCGCAACGACGCTATCTATGACTCCAGTACAGATGAATATCAGCATTTGG
TTGTGTTTCTTTCCATCGCAGCTGTGGCCAATGGCACTGTCTGGCGGGTGACACCAACAACTTACCAAACCAACACTACATTTTGGTATGTCACTGGCAGCAGTGGTACTACCCAATTCGTTTCTTCACCAGTGAATGGGCTTAACAAAGTGTCCAAAGTTTATGTTAATGGGGCCAACGCTAATTCTTACTATGGAGCAGTGTGG
GCCAGCGAGCGTTCTAATGCAGAGCCATGTGACTTCTGTCTCCAACAAATCGCAGTTGTCAATGCTGTTCAATTTGCGGCGGCCAGCAATGATTTAGACGTTGCAGTGAATGCAACTTCCATCCCATCCACACAACGAGCAGCCTCAGCCAATTTGTTGCTGTATTTGTTGTCAACAATCAATGATAATGGTGCCGAGATCTATTC
GCTTTACTATGGCACAGACAATGCGACCTTTTACCAAGTCTCCGATTGCTTGTACCTCAGCTCTGTTGCTTGCAGCAAACGAAGCAGCAAATCGCGATACGTCTTTTCAGTGACTGACTCACTGAGCTCTAGCAATTTCATCAGTTACGAATGCGATGATTACAAGATTGGCTCCACTGCCATCTATACCACAGTGGCCTACAACC
CAACCACATCGACATGGTACCGTGTCAAAACCGGATGGACCGACTATTATGATGCCCCCTTGGCTGCCGGCGCCTACACTCGCACATTCACTAGCCAATTCAACGGAGGTGTTGTTGGGGGCGATGTGTCTCCTTCATACATTGAATCTCTTTGTGTTGCGACCACTGACTCATCCTCACACAAAATTTCTGCAACAATGGCATCG
CTGGCAGCTGTGCTCATGGCCGCTCTCTGGTTTTGAATGTCCATTTAGTCGTTTTAAATAAACAGATCGGAAGAGCGTC

>GRPc38956 (Avirulent pathotypes only)
TGTGCTCTTCCGATCTTGCGAGTGGGTGGATGTGGTTGGCTGTGTTGGGACTGTGTGTTGGTGTGGTGTTGGGCCAATGCTATAACCAAACCACAGCCATTCAATACTATAATAGTGGTGCTGAGTGTGATCAAGATTATTCACAGGAAGGGGAAGTGACAGTGTTTTGTAAAAACTTCCTTGCGTGCCGCAGTGGCGTGAAC
GGCCTTGCGCTGTGTGTGTCATCTAATGAAGCTATCTTCGCAGATTTCCCATACGGTCAGAATTGCACTTCAAACGCCGATTGTTTGTCACCTCAGTTCTGTGGGAGTGATCTCACCTGCGACTACCGCGACTTGATCAATGGCTACTCTTGTGAGAAAAATTCACAGTGTGCCTCCAACACTTGCATTAGCGGTGTATGTGTGAG
CAGTTTCGATACAATCAGCGATGGGTTAGGGTGCAGTCAGGCCAGCGATTGTTCAGTGAAGACATTCTGTGACGTCACCGGCACTTGCACTCCATCTCTGGGGGAGGGAGAGGATTGCAGTGAATACAGTAACTCCATTGACACTGCCAATTTCGTCTGTGGGCCTGGATTGGCGTGTTTGCCGTCAGATGACACCAATTACG
TCTGCAAACAGATCTTTTCCGATACTGTTGGGGAAATACTGTGGCCAGAGCATATATGCCTGTGACTATGGATTGTTTTGTGATTCCGCACAGAAATGCGCTTACTACCCTTCGTATGAAGAAGACATTCGATGCGACTCCTCTGATCCATCTGAGTGTCCAGCTGGATACCACTGCTCTTGTGAGGGGAGCTTGGATACGTCATC
ACAAGCCACGTGTTCAGCGTTTGTGAATCCGTTTTGTCCAAACTCGTTTGCTGCGTATTTGGATTGTCTCACTGCAAATAGCTGTCCATTCGATCCATACACGATTATCAACAATCCGGACAATGCCTTTCCCTCTGATTTCGCGGGTCCGCTCAATCGCTTTTCGTGTGCCGCACAACACTGCACCAAAGAATATAACAAATTTGT
GCGATGCCAAAATGATGGGTTTAAGTCTGATGATACGTTTTCCCCTGCGCTTATTCCGCAATGGTGTCTTGTTGGTGACGTCAAATTGGCTGGATGGCAGATAGCGCTGATTGTGCTTGGATCGATGGGGTTAGCAATGATAATCGCTGTAGTGATTTTAATTTTGAAAGACAATTCGGCGTACGAAGAATTGTAAAAATGGAA
AAATGTCCGATGGCTGTCCGTGCAGAGGGTGACCCCGCCAGGGATGGTCAAATGAA

>GRPc42507 (Avirulent pathotypes only)
AAAAACTTTTATTCAGTTTAAAAGAAAGCCGCCATCAACCCCGCCCCAACGACAATCGAAGATACAACCGCAGAGTTCGCGGGGGATACAAAATAACCGGTTGGGTAGACACCATCTCGGCCTAGATAACAATTGTTGCTGGTTGATAATGTGATCGCACTTGCCGCCACTCCACCGTCAAAAGGATAGGCAAATGTGCGAAT
GGTGGTGCCAATAGTCAATGCGGTGTAGGGATAAGTCCATCCACTTTTGAGAGTATACCATGGGCGTGTGCTCGTGGGGAAAGCAGCACCGTAAGAGGCTTGAGAGGTGAATGTTCCAGTAGAGTCAAGTGTGTATTTCACCAAACTTGTGTTGCCATACACAATGACGTTACGGACATAGAAAACGTACCCACCGCTGGTG
CTGGTACATTTTTCAAGAGGATTACGACAATCACCGATATCATAGAAATTTCCGTTATCGAATCCAACAAAAACTTCCAAGAAAATCTCTTGTGTGAATGAATTGACAGTGGCAATCATCTGTTTTGCATAAGTCTGAACATCTGCAGTGGTCGATATCGAATACCCACCCGACTCAGAAGAAGCCCAGCTAAGAGCCCAAGGTA
CAGGGAAGGAGCCTGTCAAACAAGGACTACATGGTTCAGATCCCACGCGGTCTGCTGCAACCTTCCCACCAGAGAATGAATAATAATAAGTCTCCCCAGTGGGTGTGTCTGTCCATTGGGTCCAGCCGGAACCGGTTTGCCACCATTGTTGATTAAGAGGATCGTAAGAAGAGCTGGATTTTCCGTAGCTGGAACCCAAGACT
CCGTTATTGTCAGTGTTGTAAGCATGAAGTAATGTATCGGAGAATTTTGCATTGTTTCGAGTGACAAAAATGTATTGGCTGGTCGCATTTGTGCAAAGAGTGTTTGTGTACCCAAGAGCCCAGCAATTCACAACTTGGTAGAAGTTGCCAGAGGTCTGAATACCATATGCCAAGGCCGGGGCATAGTCGCTCGAGGTTTTGGAA
AATACACCCCAAATCTGTGACAAAGCACGACTGAAATTTGTTCCTGTCTCCAAGCCGGAGATACTTGCAAGCAACTTCGCACTTGAAGGAGCCCAGCTGGCCGCCAAACATTCATCATCACACACCGCACACCCAACCACCACCAACACCACCAATG

>GRPc42674 (Avirulent pathotypes only)
GCCAGAATGAAGAGTGTGGTGTTGTTGGTTGTTTGTGCGCATTTGGCATTGGCCATCCCTGTTGGCTACACCACTTTCTTTCGAGGAGACTCGTGTCCGTCTGGATGGGCACCAACTTCAGCAGCAGAGGGACGGCTCATTGTCTCTGTATCCAACACAAATTTGAGTGGACTTGTGGTTGGAACACCACTGAATAACATGGAG
GACCGTACTCATGTGCACACCATTTCCTCCACTGTTTATTTGCCTGTCAAATCTGTCTCTGCGATTGGTTGTTGTAATCCAGACGGTGCGTGCTATGGCTCCTATCCAGTCACTGGCAAGACCACTTCAGAGCCCTCCGGTCTTCCTTTCGTTCAATTGTTGATGTGTACTCTCACTGCGGAAGTTGTTGGTGACATTCCCTACGGCA
CACTTGCATATTTCACCAACACAACTATTAAGCAATGCGAGGACATGAGCGGATGGGCTACCATCGCAAGCGCAAATGGTCGCTCAATCATCCCTGGCTACACCACTGGCATGTTCACTTCCGTCACTGCCCCATTGGCCAGTCAAGAGGATCGCCCTCACACTCACACCTTCGAGTCTAATTTCACTACGCAAGACCAATCTTTT
GCGGGTGTTGCAGGCTGCTGTGATAAGAAATTGGCTGAGGACTCAACTTACCATGTCAACGATGTGACCTCTTCATCATCGACCAATATTCCTTACATCCAACTCCTCACTTGTGTTTCCCAAACCCCAACATTTAATGTGGGACTCCCAAGCAATGCCTATATCTTCACTGAGGTGAGCTGCCCTCCTGGCTATACCATCTCAGAG
GTTTTGGCTGGACGCTTTTTGGTTTCTCTTCCAGCTGGAGGTGTTTCTGGTGCTGTCTTTGGCGGTCCATCTTTGCCTCAGGGAAGCACGATCGGTAACAATCACTCCCATCCCTTCAATTCTTACTTCACCACCGACTCCTGCGAAATCGGACTCGCCAGTGGCTGTTGTGGTGATGGATACGCCAAGACTGGTCAATACATTTCC
GAAGGGATCACTGCCACCACAGATGCTCACTTCCCATTCATTGCTGTGCCCATCTGCGAATTGGAATCCCAAGACAATTTCGAGTAAAGCACGTTTTGCCTTTGTTGTACGTCCGATCGAATAAAACAGATCGGAA

>GRPc45446 (Avirulent pathotypes only)
CGTCAGGCGGCATCATGAGTGTGTGGTTGTGGGTGTTCGTGTGTTGTGTGTTGGTGGGGAGTGCGAAATCGGCAATGACTTTTCAATTTCACAATCAACACAACAAGTTGGTGACTGTGTATTGGGAGAAGTATGGAAGCACCCCCACTGCACTTGGATTGCACAGGATGACTGACTTGAAGCCTGGTGAACGCCACACTATT
GAGTCGAACAAAGGTCACATCTTCGAAGTGCATGAGTCATCTACTATGACTTTCTTGGCCCTCGCCCAAGCTACAGATGTGGACCAAGAATTCATTATTGGTTCTGTTCGTCAGTGTGGTCTCTATGACACAGCAGTGCAAGTCAATGGCCTTCTCCCTTTACAAACAGACATCGCAGTGTGGAGGGATATGTTGCACTTCTCAC
GAGTGGCAATGCGTTTGTATCATTCTAAAATGTACAATCATCCATCAACTGAAGTGGCACATCAACGTGCTCAAAATGAAGCCCGCCGATTACGTTTAAACAAAGAACAGCCCGGGGTTCATAAAAATTCACCGAATTCGGATACGAAATCAAACGATTGCCTGAAAAAACATATCAGAAATTGCGTAAATTTTGGGATGAGA
ACCGCCACCTCATGGAACTCGAACAATGGGATGACAACAACCACCATGTCAACCACTGGGAAGCTGCTTCGTATATTCTTTGGCTCCCAAATGAAATTCAATCCGAATTGAGAAGAACCATGCGCCCCATTTTAGAGAAATGGGTGGACCAAGAGCTTGAGTACACTGCGTTGTATGGCATTCGTGTTTACACAAACAATAGCT
ATCTCCTCAACCACGTTGATCGATCAGCCACACATGCTGTCTCTGCCATCATTCAGGTTGACCAACGTGTGGATGAAGATTGGATGTTGGAAGTCATCGGTCATGACAACAAACCCAATTATTTGAAGTTAGTGCCTGGTGAGATTGCGTTGTATGAGAGTGCCACTGTCATTCACGGGCGAGAAACCAAACTCAACGGCGATT
ACTTTGCTAATGTTTTCATTCATTTCCGGCCTATGAGGAATTGGAATGCTGTTCATCTTTAATCAATAAGA

>GRPc45982 (Avirulent pathotypes only)
CTTGCTCATCGATGCAGCCCATGTTGTTAAACGTTGGCATCCTGATGATGATGATGATGAATGATGGATTTATTCATGCAAGCGAACTTAATCCACCCAATTGGCCTGCTGGAGTTTATGTGTTTGACCCAGCAACACCATCTTATACTCAAAGTATTGTTGACCAAGCATTCTCAGTGAATGGAGGTCATGACCCAGTCTTTAAC
GGCCAATGGAGCACCAATCGATATGCATTTTTATTCAAACCAGGAGTGCATAATGTGAAAGTGAATGTTGGATATTATACATCAGTGATTGGGTTGGGAACAAGTCCAAAAGATACAAGCATTCAACATATCATTTGTGAAAATGGGGATTTTGATTATTCAGGAGGAGCTTTAAGTAATTTTTGGAGATCAGTGGAGAATGT
GTATAACAAACCATCACTGACCTGGAACAATGAAGCAACACCAACCATGCTGTGGGCAGTGTCACAGGCATCACCATTGCGACGAATTTATGTGGAAGGGAATTTGGATTTGTGGGAATATAATTATGGATGTTGTGCTGGATATGCATCAGGCGGATTTCTGGCAAATTCAGTGGTGACAGGGCAAATCAGCAGCGGATCA
CAACAACAATGGATTACTCGAAACACGCAAATGGGGTCATGGTTGGGAGGAGTGTGGAATATGGTGTTTGTAGGAAACAGTGGGAGTGTGCCAGGAAATCAATGTCCACCAACAACAACAGTACAAACAACGCCAGTGATTGCAGAAAAACCATACATTGTGATTGATGAAAATGGCAAGTATTATTTGATGGTGCCGCCA
GTCGAGTACAACAAACAAGGTCCAAGCGATTTCTCCAAGAATCAAGACACACCAATTGATTTTGAAAATGTTTATGTGGCACAACCATCAGACAATGTGTCAGTGATCAATTCAAAAATCGCCAGTGGAGTTCACATTGTGTTTGGCCCTGGCATTTACAATTTGACAGGGACAATTGTGGTGAATGTGCCTAATTTGATTTTGT
TGGGGATTGGGTTTCCGACATTTGTTTCCAATAACGGAGAAGCGTGTTTTCGAATTGGG

>GRPc48268 (Avirulent pathotypes only)
CTTCCGATCTGTTCGTGTGTGGGCGGAGAGAAAATGAGGCTGTTGGTTGTTTTCGCGCTTACCCTTTGTGTTCAAGTTCTCCATGCTAAAGTCTATACGAATTCTTTGAGCTGGGACAGTGTTGCGAGTCAGGAGGTCTTGATTCCTGAAAACTACACGGGGACAGCCAATTTGACTTTATATCCCCTTGAAGATGAAGACGTCG
TCGTTTGCGTAAATTGGCTCGTTCCTGCTGAGGCTCCTGAAGATGTGAATCTGGCTGAAACCTGGAACGTGACTTGCCAATATGCTATAACAGTGCTCGCGGGGACTGATCCAGTCGTTCGTTCAGTTGCTCTCCGTGCCAATATCAATTACGTCTCTGCTGCTTTGATTAATGAAGATGCTGATAGTGATGGAAATAGCGATGA
AACCTTTGATTTCTCGTGGGAGCTCTCTTATGACTCTTGCCCTTCCGGAAGTTCTGGTCCTTCGTGCACCTCTGTTGTTGACGTTGACTACTCTTTGATCATCACTGCGAATGCATTTACTGTGTTCCGTTTGCCAACTCTTTTTGATAAAAGTGGGGGCCAATTCCTCCAATATATCTCATTCTCACTCCCGACTGCCGGAGCCGGAG
CCGAAATGGATGCTAGTAATGTTTTGATTTCTTATCGTCTGAATGGACCTGTGACTACATTGTCTGATGGTAATTGCACATCCGCCCAATTCGCCACTTGTATCCCGAATCCACCATTTTACTCTGCGTCCGAGACCTATTGGGAGATTAGTGTGTCAGACAAACTCTCCTTTCAAATTATTCCTATCTTTTGCAGTGATGGGAGGAT
TGGAGCTGATTGCAAAACGTCATTAATTGATTTTGAGAACGAGGTTGGAGTGCCCTTGACCCGAACAGGACCTGCCTCTTTTGTTTTCTATGGTTCTGATTTGTCCGTTGCTGTTGGTGCCTTGGATGGATACACCCAACAATCCACTGCTCCCTCTCTCGTGACTCAATTAGACGCACTGCCTGTTGTTTCT

>GRPc50352 (Avirulent pathotypes only)
GTCATGGGGTTGTGGCTTGTGGCCCTGTTGTGTGCGCAGCTTGCGTCCGCTGCACCCTCTTGGACCTTCGGGGACTACTCCCCTGGTCACCCGTCGAATGTGCTTGTCAACATTTATCCAGATGGTTATGTGGACGGTTCCGATTGTTTTGTCTCCAGCTGCCCGGAAGATGCAGTTGTGTTCCAATACATCGATCCCGTCTATGG
AAGTTGTGGGCAAAACGCTGATAAGCTCGCAATCACAACATACTCCAACGGCACTGGCACGCTTAACTTCGGCGGCCTTGTGAATTCTGTTTACTCAATTAATTGTCAAAGAAACCTCAGAGCAGCCTTTGAAAATCAAGGCTGGGTGTCCTTTACAAGTGAGTTAAGTTCCACTTTAACAATTAATTGGGTGGAGCACGGCCG
CTTGGACAGCCTTAAACAGAGAGTGTGTGATTGTGCTGACCCCTCCGCCCCTCCCGCTACAGTCCTCAATATTCTCACCTCTGATCTTTGCCAATCAGGTTGTAAAAATCCAAATAATGTCGTCCCACCCACCCCTACGCCCACTCCCTCTCCTACACCAACGCCCACTCGAACACCATCGCCGACCTTCACCCCCACCCCGGCCCCC
ACTCCATCCCCAGTGTTATGCCCTACTGGTGTGGATACACAGAATTGCTTAGAATCGACGGTATTCAATGCCACTCGCTGCAATCATTTGGACGACTGCGCGTGCCTTACAGTATTGATATTCAATAATGAATTCCCATGCAGCACAGCTAGCCAGTGCTTTTGCAATGGTGTCATTGACTTCTCCAAAGTGCCGTTCTGCCTGCA
AGGAAGCAGCGAATGTGCTGCGTGTGCAAATTTCAAATGGCCGGACGACTGCGCCCCTGCGTCCACATTGACTCTTTCAAGCCTTTTGGTTTTCTTTGTAACTTTAATCTATTTGTATGCTTTGTAAGAAGTAAAGTCGTTCTCCA

>GRPc56405 (Avirulent pathotypes only)
GGGGCATTTGCACACAACCATTCTTTTGTCTCAGAAATTCGGTTGTGCCTGAGTTTCTGTGAAGCGCTGTTGCATGCATGTGCCAATGACGTTGTTGTGGATAATGAATCCCCCAACGTTACTTCTGTCGATGAAAGATACAATGGAGATGCTGCCAAATTTTGCTATGATCTGTACTCCTCAGTCCAAAATGTTAAAGTTTTGATT
TCTCTTGAAACGTCGAAATCCACAATTAGATCATATTGCTTCTCCGATGAGCCAGAATGCACCGCAGATGACTTTGTCAGTTTCTACACTGACTGTGATCCTGTGAAGAAGACCAGAGCGGTGTACTTTGATTACAAAACAACCACCAACTGCAAAGGCGGAGTACCATTGCCTTCGCCGATTTTGGACCTTCCCTGTTATCAATC
GTGCGACCCTGGCTACTACCTTCCCATTGGGACCACTGAATGCGAGCAATGTCCAGCAGGAACACACAGTTTTGGTGGTTACTATAACGATGAGTGGAAGTCTTGGCCTGAGGAGTTGCATGTGACGACCTATTGCGAAAGCAACGACGGTACGGAGCCTGAAAGTGGCTCTCTGATAACGAATCAGACATCTGGTGGCTGC
GGTTGGCTTCTGAACAACACTTACATTGCCTCCGGAAATGTTGGGGACTCACAGCAGTCCGTTTTGGAAATGAAAGTTTTTCTTGTAAAAGAGGGTTTCGTCAACTTCCGTTTCAAGGTCTTCGCCGAACCACAATACGATGGACTTCTCTTTTACATTGACAACGTGAATAAGATGAAATTAGTCTCCCAAGCAACGGAGTGGG
CAGAGACGAGCTATCCCATTGGAG

>GRPc57564 (Avirulent pathotypes only)
GGAAGGCATCAGCCCAAAACTGCGCACATGTTGGCATTGCCCTTGCTTCTTTGTGTTGCCTTCGGCACTGTTGCCTTCGGCGATGATGTTCATTACTTTATCAAAGTGAACAGAGATGTATTTGTTCCGGCGAACACTGCTATCAGTTACTTGACGCGATACAACCATTGGTATCATTTGGCCTTCGACCAAAATGGCCACTGGGT
TTCCAATTGCTCCTCCTTGGAATTATCTGCGCCTTACAATCCTGTTGTGTGGGTCGCAGGAATTTACCAAAATTTGTTGACCGAAGAGGTTTGTAAGAGATCCCAAGAGGAACCTTTGCTGGTCGGAGATACACAAGCCTTCGATGACGCATTTTACTTGATTGAACTTCACCGACAAATTACTTTTGGGGATGATGGATTTACGA
TCCAAGCACAAACAAAATACGCTGTTGCCACTCCTGAGGGATGGTGTGCAGAGAATAAAGATGGAGATTGTATCATACTCTCTAGCCAATCTGGTCAAACCAACCGATTTGACCCAGATGGAAACCTTCGAGACGTGGATTATGTTGAAGACAAATACTTTGTTCGTGAGTACATTGATGAGTTTGGCGATGTGGGGTCGAGG
AAGAGGAGTGCGCCGAACAAAATAAAACTCCCCCCACGTATTCCTGAAGGCGCTGGTATTCACAGCTTACCGTTGGATGCGTCTGTGTTACAGCAATTGCTTGAAAAGACGTCTGTGTCCACCTCGGTGGTGAGTGAGGCCAAGACAGAGCTGTGAGATAACTGTACGCGGACTATATGAGTGTGCGATGTATCAAAGAGGA
ATAAAT

>GRPc60317 (Avirulent pathotypes only)
ACCGAGGATACATTGCCACCAGCCGGTTCGCTCGGCGGTCTCGAATAATGCGTGTGTACGGAGCCACACCAAGGTTAATCATTTGTGCTCGGGCAAAGGCAAGTGTCAGCAACAGAAAACTGAGCTGACCAAGAGGCGGTTCCATCCAATTCCATGTCCATGCACCGACCTCGAGGAAGGTCTCAAGGTCTTCTTGTTCTGGA
ACGTCTGTGGTGACAAAATATTTGTTGAACCAAAGCGCTGTGCGGAGATCAAAGACCATCGGCGCTGATACAATACCTGCGACCACACTCGTTGCAATCCCTATTTTGTATGGAAGTGTGCGCAAGCCATCGCCACTTTTGTTGTATTGCTCCATTTCACTGAATATTTTTTTCGCTGCAGAATAATCGATTTCGTCAACGGCCAT
GATCTCTCTCAAAAGACGCTCTTCTTTTGCTCTCTGATCACCCGCGCTGGCCATCGTGAGGATAAGATCATTAGACATTGATGAATAATTTTTTGGCATTGCTTTGGCAATTGCAAGAGTCGGCACAGGCTCAGAGGCGGAGTCCGAAAAATATCGTCGAGACCCCGTTGGCCGCACCAACTGCACAGACCACGTGGATATTTGT
CTGAGGTGAAGAGCAGAAACACTGCTAACCATCCCCTTCATCACCACTCCCCTTCCACCTCCCAACCCAAGCCAACTCGAGGAAAATCCTCGACACCAGAGGGTCCCGCGAGTGGCCAACGTGCACATGTGGATGATGAAGAAACGAACGACGACAAATGTCGG

>GRPc66581 (Avirulent pathotypes only)
CCTTCTCTTGCTTCTTCTCTTCTTCTTGTTCTTGCTGTCGTGCGAGGTCCAAGGACAGAGCAATGGCGAGTTCATCATCTTGTTCCTCTCTCAGCATTTGAGCCTCGGACTTGCGTGGTGGATGGAATGTGGGGATGGGATGATCGGATGATTGTAAGATGGTTTGGTGACGCTGTGTGAGGCCGCGTGAAATGGGGCGAAATT
GTGTGCGCTGTATGGGCGGGGGTGCACCCATACCAAAGATCCGTTCTCTAATTTCATTGATGGACATTGGCTCATTCACATTCGCACCCCAACCTCTTTCCCACCTCTGCCGTTCGATTCGGTCAAAGTCTCCCTCTCTCCCCTCAGACATATTCCCTGCGTTCTCTGCAGCCAAACTCTCAGCCAAAAGAGCATCCGATCTCCCCTC
CTCGAGAGGTAAGTCAGCATTGAGGGCATGGCTGTGTTCAGGTCCACTCTCTGTCACATTCCTGTGATTGGATGGAGTGTTTGGGGTTTGAGGAGGGGAGCCTCCAAGTTCGAGGCCACGCCGACGTTCAAAAAAAGGGCGGAATAAGTCATAAGTGAAAGAAAACAACACAACTATGAACCCTGTCCACAGCAGCAACCCA
CGCACCCACTCATCCATGTATCTGTCTTCAATAAAATTTTATCTTTCATAATCACATCATCACACATCATCACACA

>GRPc69615 (Avirulent pathotypes only)

CGGGTTTATTGTTAAAATATATTCTCGACAAAAAACTTAATCATATTGATCAAAGTTGTTATTGGCGGCTTGCTTCTTCTTCCAGTAGATGAATCCTCCAACACCAGCAACAATGATGACAACGACAAGGATAACTCCAGCGATGACCAGAGCCCAGACCCAAAGGTCACTATCGCCCGAAGTATTATTATCACCTCCGTCCGCCT
CTTCAGTTGCGGTTACCCAAGTCGGGGTAAGAGTGTAGGATACACTTTGTTCGGCAGCAGTACTGCTGTCACCAAACGCAACAAGGACCACTCCTGTGGCAGCTGTGAAGTTGCTTTCGAGAACTGAAACGGTAACAGTGGTTTCTCCATCTGCAACGCACGCACAGCCAAACTCATCGTTGCTAGTATCCAGAATTTGGAGAC
AGAAGTATTGAGTTGATGATGAAAGAGCAGCAGCGATAGTGGCAGAAACATTGAATGCCGCATAGGATTGATTGTTTGTGAACGAGTAACAGTCTGAAACACCTGCAGGATCAACGTAAGCTGGTTGCGTTAAAGTCACTTGCGAGGTTGGAATTCCATCGACGAGGGTGTTGGTGCAAGGAACAGACGTGCCGCAAATAT
CTGCGACGGCAACTAAGGCCAAGAGGGCAAGTACCAATACACG

>GRPc71504 (Avirulent pathotypes only)

CCTGCCTTTGTTGAACCCATCGTGACAACGATTGTTCCTTCTCTCGTGTGTCACTGGTTACCCTTCCTCACCACGATGAAGTCGTCTTCTCTGTTGAGCGTGAAGGTGGTCGCCATCATGGCCCTCCTTGCGTTGTTGGTTGCGCCGATCGTGTCGACTCTGGTCGCTTCGGTCGCACCCGAGCGTAAGGCACAGGATGTCGAGGA
TTGTGTCGCATGCCGATACATCTGGCTGCAAGTGGAGATGGACGTCGGCAACTCACAGATCGAGGAGAACATCTATGACTCGTTTACACAGAACTGCATTGAGGCCCAGAAGGCGCCGATCTTCTACCCGGCCTGCCAGGATATGTTTGATGGAATTGATGATATGATCGGAGACTACATGGATGGATACACCGTCAACCAG
GTCTGCGAGAACGCCCGCCTCTGCCGATAAACGTCATCAACAGATTGGGTGGATGGAATCGGGACAGGAGGAAGGACTGGCCATGCGAGAGGGCTTTGAGAACGGAATTCGCGCTCGATCGTTCGACTCGCTGACCAGCACACAACATCACAACATCACCTCCTCCTCCTGCTTCAACCTCTCACACCCATCATTTCTCTGCTG
CAGATCGGAAGAGCGTCGTGTA

>GRPc80671 (Avirulent pathotypes only)

GTGTGAGGATGTTCGTTGGGGTGGTGGTGTATGTGGTCAGTTGCCTTTTAGTGGCTTCCTCTGCGACCAAGTCCTATCCAGTCCCTGTTTTTATGTGGGCTGAGAATGGGAGGTATTTGCACCCTGGTAGGACATGGATTCCAGAAAGCAAAACAACAGACGATATTCGAAATTTTTTGTTAGCAATGAGTGATGGAAAAAGCC
AAATGACCCCTCCCTTTCCAACTGCACGGACCCCTATTTCCCCAGAAGTCATTATTCTCTTTCTTGAACCAGAGTTATTTTTTGGTCGAAGTGATTTGTACGGTGGTACTCGGCGTGCTCAGCAACAATCATCTTTGAAGAAATTAAAGCGAATGATTGAAGGATCTGAAGCACAACTCGTCGTTCCATACACATACGATCTTACTT
TAAGTGAGAGCGAGGATGACCGCGCGATCATCCTCACTGAAAAATGGGCAATCACTTTGGATGAATATTTGACAAACTGCGAACATAATTGTAAAGTTGTGGTGGAAGCCATGGACATGAGTGCAGTGAATGCTTG

>GRPc82159 (Avirulent pathotypes only)

TTCCGATCTATATCATAATTTTTATTGTTTGATTATTGAACAAAAAGCATCAATCCCACAGCTGCGAGAAGGGAAAATAAAGATACCGTTAATGCGCTTGAACTACTGTATTTCTTGAGAACACCGCCTGGTCGAATGAAGTTGACATCATGGCAATTTGAGAATACAAGATTGTACTGGTACGCTGGGGCAGCATCGCAGAGG
GCATTAAAGCAAGTGTCCGAACACCCGCCAGTCGGTCTCAAAACGAGGTAGCACGAGGAATTAGTGTAGGTGTACACACCAGACGCTGTCGCCGTGCAATTTGAAACATCAACAAGGGAGGCCGTCCAAGTGTAGTACAATCCGCTGTTGAAGCTCAATTGTTCACTAACATAGAGAGTATCGGCTCCAACTTTTTGGAGCGC
TGTCGAAGAGTAAGAAGATCGGAGTGTGCATTCCAAGTGGCAAGTGGTGTTCTGTGCAGAGTTTCCTGCAATGAAACACAACAGTACAAGGCCAAGAAGCCGAATGCCCTTGCAGCGCATGGCTTTGA

>GRPc83704 (Avirulent pathotypes only)

GGTGGAGGTGGTGGTGGTGGAGGTCGTGGTTGTTGTGGTGGTGGGATATGGACTGTAGCGATTACGCCAGGAGGTTGGAGAGATGACTTCAATTTCGGGTGTGATGATACTAAGGCCAGGAGAGGGGTGCTGGCCCTCTGGGGCAGCTCCAAATTGCAAAATAATATTTGGATGACGATAGGTGGGCTCAATGAATTTGA
GTTCAACAGTGATCTCAGAATCAGAGAGGGAAGGAAAGCGACCGCAATCCAAATATGATATGGCATGATCGAGTTGGGCTTTTCCGATAAAAAGCCGTGTCTCCCCCTGATTAAGGCGACAGAATAAATTGTAATATGCCGTCGGTGGTACTCCTTTTAAAGCAAAAGTAAGAAGGAAGACTTTCTCTGAAGTGGCCACAATC
TTTTGTTTGCCTTTCTCATTCTCATACTTCACCCACTTTTGTGTTGAATTCGAGCGACACCGAAATTCACCCATCAACAGACTCGCCCCTGCCGGCCCTGCAGATCGGAAGAGCACACGT

>GRPc84743 (Avirulent pathotypes only)

ATTTGAGAATTCAAGAAACTTTCATTTTTTTTTTCGATGCAGTTTACTGCTGATCACTGCAACAAATAAACGACACTTGCTCGCTGCAAGGACACAAACTTGGAAGAATAGGCCCTCCATAAATTTCTCTGAGGCAACGGCTGTATTGTTCATTCGACGATGTGGTATAACCAATGCAATCATTCAATGGGACATAGCTATTCGTCT
GACGATTGTAGACTTGAGCATAAGATGTATCGATATATCGGACATCTCCTTGGCCCAAGATCACGAAGAGAGTCTCGGTTTGGGCAACTACTGCCAGATCATCAGCAGTACAAACTCTTGATCCGGGGAACTCCAGTGCACATTTCGAGCCTGCTGTTTCATATCCATTCCCGCAAGTTGTCGAAGTGCAAGACGCAAGAGATC
CATTGTACTTCGCTTCGGTCACCCCTACAAGAGGAGAGACAATCGACGCGGCAAAAGAAAAATAAGAGACAAAGACAACAACGAGCAAGCACTGCATTGGGTCTA

>GRPc87482 (Avirulent pathotypes only)

CGTGTGCTCTTCCGATCTCTATTTTCCTGTTGTGGCCATGATGAAGTTCGTGTGCTTGGCGTTATGCATCCTCGGTATGGTGTGGGTAACTGAACAAGCTCAGTGCAGTCAATATTGCGTTTCTGGATCCGATTGCTGTTTCAAGGCCGAGGAGTGTTTGGCTTACATTTCAGCGTTTGGTCAATCTTCTTCCTTTGCCAAACAAAG
CTGCCAACTTTTCCGTGAATGCCCCCCTAACTATGATCTTGTTGGCCCTCTCTGCCAAAAGCAAGATAAGACTCTCAACTCCTTCCAAGATACTACCGAACAAGCCCCATTCACCAGCTTGGTTGACTGCTTCTCGTCCTCCGAAGGATATGGTAACTATGGAGATGACTTCTGTCTTTCTCCTTACCGAGAGGCCTTCCGTGAATG
CTTGCGCTCGCTTGAACGCAGAGGATAAATATAGTCGAATCTTTAAACTTTCGAATAAACTGTTTTGTTTCCATCAAAAAAA

>GRPc93120 (Avirulent pathotypes only)

TTTTATTGTTGGCTCTTTTTAATATTGTAGGTCGGCTTGGAGGTTCCCCAAAACGTTGAGCATTGATTATAATCTAAAGTCATATAGGACCCATCAGTGCAGGAGACCGAACTGCCAGAACAAGAGTTCGAACCAAAGCTGTAGGAATATTTTCCAGTCATTTCTTTTGGAGCCGTATAATCAGAGCACCAATTGACTTCAGGAGT
GTACCCAAAGCGATAATCGATCTCATCAACACCAACGAAGGGGCTCGTCAGGGTGAAGTTGGTGTAGGCGTACGGCTCAAGACGAATGGCCGGAAGGTAGATCACAGAACCTTGGCGAACATGACTGGTGGTGAGGTCCAAGTAATACTCACTGTTATTAATAATGAAGCCAACCACATCGTGTCCAGCCACAGTGAACAG
AAACAGAACCAAAGAAACAAACACCGCGCTCTTCATCGCTCTGCAAAACG

>GRPc101586 (Avirulent pathotypes only)

GTGCAGTGTGGTTGTGGTGCTCTGCTTCTTCGCGGCCACTGCGTTTGCACAGTCTCGGTCCCTTGCCGTCTATGCCACTGATCCCTGTTGCACGGAGGTGGCAGGGGTAGTGTCGTATTCTGTGGGCGTCAATTTGACCATGATTGCATCATCAAGCAACACATGTGCCATTGAACTCGCTTGCTTGACAAACCCGAAGTCAATT
CTTTGCCAAGCATTTGAGACAGGGATCACCACCGACAACAAGATTCTGAGTGTCACCAATGAAGAAACTACCATTCGTTATCAATCCATCACAGGGGAGCTGACCACAGAAACATACACCACTACGGATTGTATTAAGAGCTGGCTCTACCCGCATTGTTATGTGAAGTTTACAACAGCAGAAAACGTCAATAGCAACAA

>GRPc19002 (Virulent pathotypes only)

CTAACATATGGTCATGGGCTGAACGTGTTAACCTTGGAGTTGGTCCAAACGAATCGTCGATCGGGAGCGATCGATCCTCTGATCTGATCATGGCCTTTCAAACGATTCCGAGTGTGTTGATCACACTCCTCACCCATCTGTTGAGTGACTGTGATGCCATCTCGTTCATTCGTACATGTTCGGAGTTATTTCATCGTATGAAGGCG
ATCTATCAACTACGATCAACCTTCCAACTCCAATTCATCACTCCTCAGACAACCACTATTCAATCAACCAATCCACCCACCACCTTGACTCTACCATTCGGTGAGGTATGGAAACGACCAACTCATTTCATTGGATTAATTCATGACCTCATTATTAGTCTGCCAAATGAATCGGGTCGTGACCCGAGTGAATTGAATCGATTGAGA
GAATTGTTTGTTCGGCCCGTTTGGCCGAATGTGATACCAACAAGTGTCCGAAGAATCGAATTGGATGATACCTGGAAAAGAAACGGATCAACAGTAATCGAGTTGAATTTACCACCATATTTAGTTGAATTTAGTAGTTCGCAAGTCGACCATTCGATCGCGGAGTCACTCACCACCAATCAATTCGATCGACTCAAGAGTTTAA
CTTCCTTACATTTACCACTCTTTCAACTACCATTAGATCTATCACACTTGATTACATTACGTATATTAAAATTAGGTCGATGGAATCTACCGATCAACAACCTTCGTCTCCCAGATTCGTTGGAGAAGTTGAGTATGGGTAATTCTTTTTCTCAATCGATCGACAATTGGATTTTACCACCGAATTTAATCTACCTACAATTAAATGGT
GCTCGGCTAGAATCTTTGAAATGGAAGCAAGATCAGCTTCCATCATCGTTGACCGATTTGAATTATTACACGGATGTGCATATAGATCGAAGTCCGAGTGCCGATACTGAAATTCCCCAACAATTCGAATCCGTCTTAGGCCCGATCGATTCAACTCAATTACCGCATGATCTCGTTCGCCTCTCCGTACCTTGTTTATTTGAATCC
CCATTTAGTAAAGTTGTCTTACCACCCAGTTTACGCGAGATTAATATCTGCTGGCTGTGGTCCGATCGGTATCGTGATGACCTAGAGAAGCGCGGTCACTTTCAACTCCACTTCTCACAATTGCGATTACCAAATCAATTACAACGATTGAATCTTCGTTTCGACTCACTCTTCACATTACAAGATGTGCAGATGCCCGCATCATTA
ACAGACCTGGTCTATTCGGATAATGGTGAGACTTTCCCTCCGCTGATATTCAAAACGACCCCGAACTTACGAATGTTGGATTTGAGTCGTGGTGATGGTTGTCTCAATTTACAGGATATTGAATATTGGCCAAACTCGCTACGAATCTTCAAATTCTCTCCCTCCTTTGGTGATCCTCTAAAGGATTTAATCTTACCTGATCACTTA
GAGGAGTTAATTCTCACACCCTTCATGCCCTCGTCTACATTGCCTCGCATCCTACCGCGTCATCTCCAACTGTTACAATTTGATCAATTTAGTGAGAGTGGTTGTTTAGAGCAATTAAGTCATGAGCGACTACCGAACCTACGTACATTAATTATAGGCAATGAGTCGCCTGATGGCAAGTTACCGCGATTCGCCGATTCTCTGCG
GACGTTGGAATTGAATGACAATTCTCGAATGAATCCATCCATCCAGTCCTTAGTATCCGCCTTACCATCCCAGCTCACTTCACTTTCATTATGTGGCAGTTTTAATTGTCATTTATTTGGTGATATTTTACCTGGTTCGTTAACCCGCTTAACATTGGGTCGAGTGTTTGACTCGAACATTCAAGATTTAATTCTACCACCATTCATGC
GACAACTCATCTTTCAGAGGGATTCACGATTTAATCGTCTCATTCATCGGCTCATCTTACCACCCGGCCTCCTCTCCATCACCCTCCCAACCGGCTATAGTCACCCGATTCAATCACTCACACGACAACCGACTGCCATTCGTGATAAGGATGGTCAGGTCAGGCGCTTCCCGTACGGTCAACCACCGAAATGGTTATATCGTCCAC
CCATCGAATTGAACGACCAGCCAATTACAATATTCTTTTAAGGTGAGGTCGACTCCAACTGGACGAGGAGCACGCCAACACCATCCTTGCTTCGACTGGGATCAAGACTGTCGCACGATTCTCAGGATCAATTCCGGTCCCGCATCACTATTATTTGTTGTCTCGTATGACTCACTAGATTCGGACCCTCTAGTTAGG

>GRPc23330 (Virulent pathotypes only)

GCCAATTTCTTATATAAAATGCTATAACATGGATTCATCGTTTGCGTTCCTAATACCATAATCATCCGTGCTGGTCATTGCAAACCCTCTAGTCTAACAACCCGTAACCTAAAGACACAAATCATCCACCACAACAATTAAGCTCTCATCGCGAAGCAAATATTATTCCTCAGGTTGCCCAACTTTCGGAACCATCCAACTGGCGCAT
GCTCTAGGTATCTCCCTGAAAAAGGTAAAAGCGCCCATGAGGCCGGAAGAATCGAAATTTTGTCCGCAGCAATCATACAATCATGTCTGCGCTCGCACTTGGAGCGCGGGGGAATGATGATTTGGTTGAAGTACAGTCGATTATTAAGTTCTAGTGGGTTATCCTTGCACGCGATTGTTGCCTCCGCGAAAAGCACTTTATGCT
TCACGAAGGATGATGAAACACAATATGTTGAGTAAAAACAGAGAATATGGGGGGGAATAGTTCACAGCTCCGACACGAACCATCCCTTTTCGTCAACTTCTTAGAAGAGCTGGGCCAGAATGGCAACACCAGCAACACCGAGAAGAGCCATGTTGACGCTGGCGGAGCCAGCGGCACTATCCTTGCTCTTGTCGCCACTGCT
GCCCTTCTGCTGACCACCGTCAACACCCTCGTTGGCCTTGTCGTTGTTGCTCCTGCACTTGGTGCTTCCCTGAATGGCCTTGGTCTTCTGGAGGTTGAAGAAGAAGTCGCAGAAATTGGAGATATCGGTGGAAGACTGGACGTTAACGGATCCCTTGACGTCCTTAAGAGAAGGGAGCTCGACCTTGTTGAAGTCACCGAGGAG
ATCGATACCACCTGAAACGGTCTCGACCAAGGGGAAGCCGTCAATGGTCTCAAGAGCGGTGTTGTTCTTGATGGCCAAACCACCACCAATGCTCTCGAGGAGGGGCATGGAGATGTTGGCAAGGTTCTTGTTGTTGATGAATGAGACATCGTTGGTAACCTTGGTAAGGTTGTTGGCGGAGAAAGACTCAAGCTCAGTAGAG
TCAAACTTGATGGACTCGGAAACCTCAGCAAGAGCAGGGACGTTGAAAGTCTTGACATCAGAGATCTGGATCTCGGCAGCATACTCGAGCTTGGGCATGTCAATCTCCATGTTGTTTCCGTTGTTGTTGATGATGAGCTGGCCGGTGATGTTGACGAGATCGGAGTTGAACTTGGTCAGCTTGCTGTTGTCGTTGATTTGGAAG
TTCTCGATGGAGGCGACGCTGAGGCCACTCAAATCGCTGATGAAAGTATCGGTAATTCGGATGCTGTTGACCTTAGTAACGCCGTTGGAACCGAAGTTGATCTTGCTGAGCTGGGTCAACTTGATGAACTCGACCTCGTTCAATGTTCGGAGAGCGCCCATCTCAAGGGTGGCCAAGAATTCGAGGTTTTGAAGCTGGAACTT
GCCGCCAATGGCGTTAATGGTTGTGCTGGACAAAGAGATGAGCTTCGTAGCGTTGCTGATGATCAGATCCTGCTTGAGCTGCTTGGGTCCATTAATGACCAAGTTGCCCGCGACACTCTCGTCAACAGTAACGATGGCGTCGACGACATCGCAGTTGAAAGTAGGGGTAGGCTCGGTGATAGTAGTGGCCTTGGTGCAGGTC
GCAGCCGAGACGACACTGGCACCGAAAAGGGCGACCGCGGAGAGGACAGCGGCGGAACGCATTTTGATTCTTTTGAATTAGATCGGACTAAATCCCTGTTTTCTTGATGTTGTTTCTTGTGTTCCTCGGCAACGCGCTGGGGAAAACTGGAGGTTGCAGTAGCTCCTTTGGAGTATGGTGAGATACTGGATGAAGCGAGAAG
TTGAGTGCGAGAAGAATCGACTCTGTTTTTCCTTTAAACCTGCTATAACGAGTGGCGTGAATGGAAGAAAAATGTAATATTTGGACGATACGACGACAAGAACCGACGGACGAGAAGAAAGGAAGAGTGAGAAAAAGAGGAG

>GRPc23512 (Virulent pathotypes only)

CGGCTCTTTTTTTCTCTTGATCACAGCCTGTCACGCAGACTCTCGTCCACTCGGAAGTGCGTCCGATCCTAAGCTAACGCATTTTTCTTTATCATTTCACTGAAGGCTTTGTCATCATGACGCCATCTCTCCTTCGAATGTTCTTTTTCCTCGTCTTGCTCACTTCGAATTCCTTCGTCGTCGCAGTCGTCGCTGGTTCTCCACTGAATG
GTATGACACTCCCTTCTTTCTCCATCGTTGTTGGTAGTGGCACCAGTGTCGGTATTGGTGGTTCGTCACCACAATTTGCTGAATTCTACGGTCGGTTGTGGTGGCTTGGTGGTTATGCAACAAACGGTGTTGCCGACTGGGACAATTGTACCTCGAAGTGGACTAGTCATGAAGAGTTAGGATTCAAGAATAAGACGAATGGGG
CAGAATTACGTGACCGCAATGGATTCATTCAGATTTCGCCAAGCCAGGTATTGCTTACAGCAGGTGCGATCTTTTGTAGTAGCCCGTGTGTTGAGCGTAATCCAGCGGATGTCTGGCTGGTTCAAGCTTATCCGAATCATACGGTAACTTACACGATGGTAGATGCAGCGGCTTTATGGGGTCCACGAAGCGCGCATCAACATG
CCGTATGGCAAGGACAGCTTCTTCTCATCGGGGGGTGGAATTCTTCATATGTAACCAAGAATGATATTTGGTCAGCGTCGTTGGGGACACCTGATAGCAATCTCGCATTCGATGTATCCCATCTGTCGTGGCAATTGCAATCATTAGTTGCTCCGTTCGCTAATACCATAGAGTTCCATCTGTTAGTCTCGAATTCAACTCTCTATT
GTGTGAGCGATGACGGCAGTGGTACTATCCGTATCAACACCCTCGATTCATGGTCGCAGGGCTGGTCCGCACCGATTGATGTTTCTATCACACGAGTCGGCTATGCGATTCAGCTGTACAAGAATTTTATTGTACTCACTGGTGGTGCGATCTACAATCCGGGTGCAATTCCTCAGTCGACCGTGTATTCATTTTCGTTATTGGAT
CATCAGTGGCGGCAGGAGCTAAACGCGACAATGCCCGCCAATTGGCAGCATTCGATGATTGTGTTTGGTGGCCAGCTGTACATGAATATCTACGGAAGCGGGGGGAAAATGTGGCGTCTCACGATTGATGGTAATTCGAGTATTGTATGCCCACCTACACCAATAGACAGCAGTACCGGCAGTGTTGCTCAGACCAGCAGCA
GCTCAAGTGGACCGAATGTAATTGCAGGGACATCATCATCAACTGGTGGTACCGGCATGTCAACTGGTGCGACTCAGAGCAGCTCAACGGCAACCATATTGACGTCGACGGGAGCATCTAGCACAGGAACAGCCGCAACATCATCCGCGGGTCAATGCTTCGGTGCGGCAACTCATTCAACGATTATCATTCTGCTTCTCATC
GCAGCTGCAGTTGTCCAAGTGTTGCAGCAAATGTAAGCAGCAGACTGCCATATCATCGGGCCCTTCATGATTCACACACTGTACTTTGAATTTCCACTGAAAATAGTTCAGCAGATCGATAGCGTGATGCATACCCCTTAAACGTCCGATCCCTCAAATCAAGACCCTGGACAAAAGTCCCTTGTATCTGCAGCTGTAGAAAACA
ATCAAATGAGTACTGTTGCAATGCAACATAACAACATCGGACTCAAAGCAAAGAAACTACCAGCTAACGGTTCGAAGAACAGCACCAAGGCCTGGCTTCAGGGTCATCTTACCCGGCTCAGAGTTGAGGAGAGCATTCACACCCGAATGGCTGTAGATAATCATCGTATTATGCTTTTCTTCTTCCTCCTTCTTTCGTTTAATCGC
TGCAATTACTTGTGCATTGGTGGTCACCACATGACCACCTGATTTAGTCGTCATGGTGAGTTAGGCTCAAATCAAGAAGAGAACTCAGAATCAAATTAATTAGGAGAAGCCGTT

>GRPc26015 (Virulent pathotypes only)

GGCACATGCATAAATGTATGATTTCATTTTGAGCTGAACAAGATCAGTCTCGAGACATCGAAACATCCTTGGGAGACTGTTGCAAGGTTGGGTAAGAGTGCGAGTGGACGCATCAAACATCAAACGAGTAAAAAAGGAGGGGAGGACCACGACGGATGAAATCACCAATAATTGAAGTGTCTTTCAAGTCCAAACAAATTGT
GGCATGATGGGGTGGAAGGGAGGAGAAGAAGAAAAAAAAAAACCGGAGTGAGCGCCGGAATGGAGAAAAAGTAAAAACAAAAAGGTTGTTGAAGGCTTTGAACCACTACCAATCTTGCCAGATAGTAGACTTATTCACTATCAATTAAATTGCTCTCATCCTTTAATTGTGTAGCGAATAACGAAGAGAGATCAAATTGTGC
CCAGGCACGACTGAAGGCGGTACGATCCACTGAAGTCGAATTATCATTCGCAGCAAAGGAAAGATATAATTGATCATTATACAGATCGGCAACATAAGCATGTAAATTACCCGTTTGTACGATTGGTAAGATCTCGCCAATAATCTTCGCAGGTGTTAATTGGCCCCATAGCGACTCAATTTGTGCGGCCATGACAATGTTGTA
ATTCGGACAGAGCCAATCCATACCATGATAGACAGTACCATTAATCACAGGATGCCAAGTGCTGTTCACTGGATGCATATTAGTGTCATCCATAATGAGACATAAGTCAGCCGAATACGCGATACTATTAAATCGTCTATCCTTTCCATCACCAACGCCTAAAATCAAATCACATGTTCTGTGTGCTGATGCAACACGACTGACAC
CATCCATGCGTGTGCGATCGAATTGTAAGATATCACGTAAGAGATAAGTGAACGGAATGCCTTGTGTACTTTCATTACCGAAACTTGCATCAGGAAAGGAGACACCAATCTCATGAATACTCATCTGTTGATCATTCACACCAGTAATGGAACCAATCCAACCGGTCCAACCAACATTTAAAAAGGAGTTTTCTAATGGATTTGA
AGGATCGGCGTGATAAATCGTCATTTGATGATAATTCTGAAATGGTCCATCCATGATATAATCTAATGCACGGACTGTTAATAACCCTTCACCACTTGCTAGAGATTGATCCCATGCGCCAATCATACTGCAGTCGCCCTTGGTCATTTCACCTAACATCTGCAGGCGTCGAATAGTGTTGACATCCTCACCCGTCGAGTCAGCCA
TACCTTGGATCTCTTCGCCGTAATATGCACCTGTGAAAGGCTGAGTATCATAAGAGGTCAGATCTAAACCAGCATCAACCAGAATTTGTGCTTCAACTAAAGCTTTGGGCATGGGAACACCAGTCTTATTAATGGCATCGGCTGCTTGTGTGATTAAATAATTCATCATCTGATCCATGAATCCGGTCATACGCTCCTTCATCAAT
TGTCCATGTGCATAACCTTTCTCATATGGTGTACCATAGAGATGCACAATTGGTATGATTTGATCATCGACCTGCACTTCATAGAGTTGTCCATTCGGCACAGACGAGACGAAGGTGGGTGTAGGCACACCGATCGGATGAGTGTTCGGCTGGGCAGATGGCGATGGCCGGCCATGACAAAACTCACCCGAGACTTGAACGAA
GTGGGTGGTGCAGGAGAGAAGGGTTACGAGGACCAAGAGGGAGATGATGGTGGTAGTCATCGTGGTTGGTGCTGGATGAATGGAATAATTTCTTCTCCGACTCGAACAACTGTGAGTGTCACCCGGTTGATCAGTATGGGTGATCAACTTGACACCAATGTCACCTGAACTCACTCCTGCACTCT

>GRPc29298 (Virulent pathotypes only)

GGATTGTGGTCGTATTCTCGCTTGTGGTTGTAGTATTACTCACCCTTAAATGGAATCGTAAGAAGTGGCTGGAGACAGTCAGAGTCGAGGGGTCGAAGGGAGCGACAGTCAGCAGACCACAGATGAAGAAAGGTTTTGGTTAGCCTTGGGTGAATCAGATCCAATTACCCATCAATCAACAATTGGAATTCATCTCACTAAAG
ATGCCTCCACAGGTGGTTCAGTCGTACTGCCGATCACTCCTGATACCCCAGTCAACTGTTTCCAATTCGCACGTGCTCAACGTATTGTGGATCCTGCAACCATTTTCATTCGTGGTCATTCGGAATTGGGTAATCGTATGTATATCCGTGAGTCGGCTCATTGGCTATGGGAGGAGATTCTAATGCATAAGAAGGAGTATTTTGAC
GATCGACGGGCTTCAGGAATCATACCTATCAACCGTCTAATTCTGACAGGTAATTCAGGTGTAGGTAAAAGTATGAATGTAATTCAATATATCTGGCATCTTTTACATCCAGGACGGGGTCAAGATCGTCCAACGATTGTTTATCAAGATTTAGTGCACAAGGTGAATATTGGGTCTTGAAGGGTGCTCGATATGAAGAATGG
AGGGGCAGGAGTCATGTACCCGATTTGCAACAATCTGATATTAGTTTTTCTTACAGCATCCGAGGGTCAATTCAAGATTTATTGCAAGATCCCTCAACTTGGTATTTCATTGATGCACCTATGGGTGGTGAGATTGATACCTGCTTTGAATGCTTTGCACTGTGTGTTGCCCCTCTGACTACTCTCGAGTCATTGAAACGATTTCGC
AATTATCTGCCACCTTCCATCAGCTTTGTTAGTGCACCTATTTGGACGCTGGATGAATTAGAGATGGCGCGTTTCTGTTGTTATCCAACTCGAAGAGAACGAATTGGTTCCATACCTCTTCTTCCAGAGTCTACAGAATACGTCAAGGAAGATGAACGTGTCATTAATGGTATCCCTCACACTCGGATGTTGTCTCAATTTGGGATT
GTTGGTGGCCTCATTTGCCATGTGCTAACACCCAATGCTTTGACTGAGCGCCATTTCCAACAAGACTTAAATCTCTCACGTCAGATGATCGAGAAGATACTTTCTCATCCGCATCACGAACTCTTACATAATATCGATGCCAATTGGTTTCTTCATCGTGTTGTACGTGTGGTTAAGCCGAATCAACAAAGAAATCAGAATGAATTA
CCTTTCCGAGAAAAAACTTTACAATGGTCGAGTCCCTATCTGACGAAGTACTTTTATAAAACCGCTGTGGCCCTTCGATCACCTCGTATCAATTATCTCATGAATCTACGTATCTCAGTCAGAGAGAATCATCATATGCTTCGCTTCCCTGATGCCTTGTGTGAAGGTTTGCATCAACAATTTCTGAAGGATCGCTCGCGCTCCAAA
CAAATCAAACTCTGCACAGAGGAGGAGCTGGATGAACAATCACTGGCTACTTCTAGACCGGACTATCGTGACTAACTCCTCACAATCACATGTAGATCTAGATCGAATGAATCCTACATCATCCTCTGTCTTCCTATTTCACCACTTAGTTACCGTTGAACATCCAATTTTGTTTAAGTTTCTTCTCGAGCATCCATTTTGCAATGTG
CTATCAATT

>GRPc30370 (Virulent pathotypes only)

TTTTTTGAGAACTCACACACGACGTTTGAATGGTTAATGTCGAGATATAACTAGGTACAAAATTAAAAGTCCTCAAAAAAAATCGTTTCCTAAGAGAGTTCGAAATTGGTTGTTGATGAAATTCTTCTCAATTTGTACATCTCTCGATGAAGTTGACTAGGATAACGAGGAAGTGAGTTCGCAAATGTGAACGTACGGCAGTAAT
AAAAAAGATGTTTGTTATTTCAGTTAAACTAATATTAATGCTCTAAACCCGTTTGCAGTTGGTGTGACTGGTATCCCTTGTAATGCTGTCCATTTTGCCTTGTCCACACTATTACTCCAATTGAATGTCACGGTCATATCACCTTTGGCAGTTAACCATTGACCAATTAAAACATTGACAGCAGCCTCATTAGTAACATCACGTAACC
AAACTAAATTATGGAGTGGATTGCGATTCAATTCCCAAGCCGGAAAAGCTGAAATCGCATTTGCAAATGTGATCGTTCTTCCACCTGGTGTACCAGCTATATTTGGAGCTGCCATAAATGCCAACCAAAGTATACGTCGAATACGAAGACCTACGGCACTACCATTTGCAAAATCTGTCCAAAGTTGACTAATACAAGCAACGG
GGAGACGCATACCATAGGTACCACTTTTGGGAACATAAAAATTGAGTTCCCGGGCATCAGCAGACATCCATTCATTGATTTGAATATCCGCTGTCATAAGGCCATCACAACCTAGCAAGGTATTCATATCAGCTAAATGCATTTCTGTATTATACCAAATACCTTTACCATTCGTATCATCATGTTGAACGATAAAATTAATTGGAT
CCATGATCGGTGGTTCGTTCGCACTATTACATTGAAATTGATCACCCATTAAATGACCAACACGATCAGAAGCGGGAGGATCACAACCTTGATGATTTTGATGCTTGGCAGCACAGCGAAAATCCTTAAATATGTTTGTAATAATCGATCCAAACCCCGCTCGATCACCGGCACGCTGTGCTGTTGAACATGGTGTGAATGTGAC
ATGGGTAACACGACCGCGATTTAATCGACAAGGTGCAGGAGGTGTCGGTGGTATGGTAACAGGTACAACTGGACATCCAATAGCTGGTGTATCACTACCAGCTGCAACACCACCCTCTAAGGTAAATGTTGCACGATTACCTCCACCATGACGAGCAACCTGAATGACACCTGGATTGGTTTGTTCTTGATACCGTGGACTGT
GAGCCGGTTGACCATCATATTCAGCTGCTGCATCACGTGCTTGATCATAACGTGCCAACTGACGTAAGGAAGTCATACCCATACCAATAGACCAATGATTACGAATCAAACGTGCGTCATCATTTAATGTGCTTTGAGCATGAGTCATGGATTCACCCGCAATTTTCGAGTCGGATTGAGTTTCCATTTGTAGTACACCAGCAGC
AGCATTCGATTCATGCGAACGTAGACTCAACTCACGTAAATTGATACCATAAACTGCCTGGATCACACATACCCAAATGAGTAATATGATGATAGTCATGTACCTCATGTTGTCTACCATTCTCATCGCTTCAACTCTGAGACTGCAGTAATATGATACAGATGTGTGTGAGTG

>GRPc30457 (Virulent pathotypes only)

GCGATGCATACAACTACGAGTGAAGAAAGAAAGCCGCCGATACCGGTACTGATCGAGGCGGAGGAAAAGGGTGGAAGGTGGAGGGATGGACGGGACTAGTTGACGGACGCGAATTCATCCGAGTGTAATGGTGGTCGATAGTTTGATCGTTCCCCTGCTACTACTCTTCGTCTGATTAATTCATCGTCTTGTTGCGAGATTA
ATCGTCGTCCCCATTTCATATAGGCGAAGATGATGAAGCCGAAGAAGGCGACGACACCAATCAGAATACCAAACAACTCACCCCACGAAAGACCGACATCAGGTTTAGATGGATCATTCGATGGTTGGCCTGTGCCAGTACTCGAGGAGGCTGTATTGCAACTCGAGAAGGACACATCACCATTGGCTAAATTGTTGACTGTT
CGTGTGGATGAATCACCTAATTGAATGAATGCCGTTCGTGCATCACCCGTCGGCTGCTCCTCATACTTCACACCACCACCAACTATCAGCAGACGATCACTACTATCACAAACCACGCTAAAACTATCTGAATCAGGCACGACCGACCAGGGATAAGATAGACTGCATTTAGTCGAATCATCTGCACGTGTCCAACTGGCACCAT
TTTGACTCGAGTAAAGAAGATCAGGTACACCAGGAACAGGCACATTGATCTCTTGACCAAAGATCCAAATTGTCTGTTGCGAATCTTGACACAACTGTGGAGATGAGCGCCCAGTCCAGCTTGCAGCTGATGTGAGTTGCGTCCAACCGGCCGCGGGTGAATTTAAATCGAAACTCCAGACATCATTCATGAGATTATAACTTG
GCACTGCTTGACCACCGGCAACTAAGAGATGATTATCTTGTGTCACAATCGTACCAGGCCAAACTCGTGGTGGAAAGGTGGCCAATGAGATATTCGTCCAGGTTTGATAATTATTGCTTGATTTCGCTTGAAATAATGCGCCACTCGTTAAGACGTATAGGTTGTGTTGTCCACTACGATCGAAATGCATACGAAGATAGTTGTC
ACTCGTCCATAGTGGTGTTGGCCAACCGACTTGACTCGCGATTTGTATCCAAGTCTGACCTTGATCGTTTGAGCTCCATAAGTCGGTGAGAGCGGTCGAACCAGTCGCCGTTCCACCACTGATGTAGAGTGTATCCTGATCGACGGCGGTGGCCGCCCAGTAACGTGGCGGATATTTGATGTTACTGGCAGCTGCACAGTTCTG
ACATGACCAATGCGCACCTGAATCGCTACTCATCCAGACATCATTCGTGGCACCGGTTGAACTGAATCCATCAACTAGAAAGATTCGATCGTGACTATCGACAACAGCATTGAAACGAAGACGAGGTGCATAAATCGAACCAGTCGGCGGGAGTGAACTAGACCAATAGGAGATGTTGAGATTGAGTGGAGTGGCTGAAGTG
AAGGTATGAGAGATGAATGATAGGAGGAGAAGGAGGCGGAATTTTGAATGGCGACATGGAGTTGTCTGACCCATCTGCTTAGTCGGCGCGAATCGGTCAAGTTCACAGTGGAGAACTACTGACTGATGATTGAATGGCTGATTGATAAGTGCGATGATGACTCCCAAACCTTGTACGGTATG

>GRPc30832 (Virulent pathotypes only)

GTGCACTGCCATTGCCATTCATCTCAGAATCATCTGTCATCATGTCTAATCTCGCCATTGGTTTCATCGCCTTCTCATGTATATTCGCTCTGGGCACGGTGACAGCATCGAACATCACACAAGTGGAGCACGGAATTTATTTAGATAGCGCGTGTACGCAGCCACAACCATATTCGGTCTCGAACTCAACCGCAGGTGATTGCGTT
TCTAATTTCCAATCCTTTTACCCATCCGCCGACAATACTTTCGTCTATACTTATACTTACAACTACGGTGCTAACTGTGCTCTGGCCAATATTACCTCGGTCACCGCCTTTGCCGTTGGTGTGTGTGGTGTTTCAGCTCGTTCAAGTCCATCACAAGGTTTTAGCCTTTATAAGATCTTCACAGGCGCGGTCAACCCACCACCACCAAC
TCCATTTCCACCCACGGGCTTTACCACTACCACGCAATACATCAACACCACAGATGCCACACCAGATAATACTTGCTCGTCTGCTCGTACTCTGGTCTCTGCTTATCCCACTCCACCCGCTACCAGCCCATGTCACAAGTATCCTGATTCTGGTCAATCCGCTCAAGCATTCTTAGATGCCACTTGTACTACCTCCGTTGTTGCTGAAT
ATTCGGATCTGAGTTGTACCAATATGATTCGCGCGGACACCACTACGGTTGAATTCAAGTCTTGTTCATACGGTTATCGTCAAGTCTGCTCTGGTACTGGTGTCCCTGATACCGCTTCGTTATTTTATATTGGTTACAAGCAATATGATAACAACCAGTGTATCCTTAATTCCAATACTCGGACTACTGAGTCGAGTGATGGTGGAT
GTTTTCGTGGTTCCTCAATTAGTTGTGGCTTGGTCAATGGTACCGAGATGCTCCTGGCCAGACAATACCAAAATCTGAACTGCACCGGTAGTGTCACTGACATCGTCTACGTCGAAGCTGGTATGTGTGTCGTCGATAATGGTCGTGTGCAAAATGTGACCGCATTCTGTACAGCACCAAGCGGACCAGGCGCAAGTTCATCCT

[illegible]

>GRPc31319 (Virulent pathotypes only)

[illegible]

>GRPc32849 (Virulent pathotypes only)

[illegible]

>GRPc33093 (Virulent pathotypes only)

[illegible]

>GRPc33111 (Virulent pathotypes only)

[illegible]

>GRPc34660 (Virulent pathotypes only)

[illegible]

>GRPc35202 (Virulent pathotypes only)

[illegible]

>GRPc35275 (Virulent pathotypes only)

[illegible]

>GRPc39544 (Virulent pathotypes only)

[illegible]

>GRPc39663 (Virulent pathotypes only)

[illegible]

>GRPc39692 (Virulent pathotypes only)

[illegible]

>GRPc41370 (Virulent pathotypes only)

[illegible]

>GRPc42467 (Virulent pathotypes only)

[illegible]

>GRPc42794 (Virulent pathotypes only)

[illegible]

>GRPc44434 (Virulent pathotypes only)

[illegible]

>GRPc45265 (Virulent pathotypes only)

CGCAGAACCAATAGCTCAAACGTGAGCACAATCCGACACAAAATCGTGCTAGCCATCAATCGAGCGATAATAGATGGGGCATCCTTGACAGCAACAAAGAGGACACTCGAGAAGATCAAAGTGCCAAAGATGATGTGACACGCCGTGAATATGGTAGTGAGAAGGGACATGGCCGAGAAGACGAGTGACTCTGGATGGAC
TTTAAAGATGATCCTGGTGCGTCGCACGAGAATGTCAAACTGGCTCTGGGCCAGGCTGACCGGCCATGAAGCCCCTTGTTTCACATCGTCCCTGGGCTTGAGCCTCGTACTGACTCGGAGGAGGAGAGCAAAGCCGGTAAAGACATGAATGACGAAGCCGAGAGAGGCCCAAAGGAGAGCTGCGTATGGGATCACTGGCAT
GGCGATGAAGATGGATTGCGCGCCAAGTCGATATGTCAACTCGCCAATGTTGGCGATGGCAGCTGCGGCAAGAGCATATTCAACGACGATGATGATGGGGTCGAGACGTCGGATGACATGAGGACGAATCTTGGATCGATCTCTCAGACTTTCAATCGCAGCTGAGTACTCTTGTGCTCGATACGCATTGATGGCCGGCGAT
CCCACCGCCAAGCCCAACGCAAGAAGATGCCGACGCCCAACGACAGACAGGATCGCGATATCTTGTGGCCGAACCGCCAACTGGGCTACAATCGACGGCGTGAGACCCAGCATGATCTGAGCAGCACCCATCTTGTATTTTGCCAGCTCAGGCATGTTCTCCAACATACACTCAACAACCTCGTGTGTCATGGTGCCCATGCC
GAAGATGCGCAGCATCGGACTGACGCCGGAGCTGTGGTGCTTGTTGATCAGGTACGAGTCATACTGCTCGGAGCAGCTGTCGCGGATGATGGTCTGTAGGCCTTCGTTATAGTCGGGGTAGAAGTGCTCGAGTGGATTGCTCCAGGTTGCGGATACGGTTGTGATGAAGCATGATACGATGCTCAGTGCCAGACTGAGGGT
TCGGAGAAGTTGGTTGGGTGGCGCCATTGAGGTGAGTCCCAGTGCTTCGGATTACAGATATCCAACGTCAGGTCCTTGCTTGAGCAGAG

>GRPc47439 (Virulent pathotypes only)

GCTGGATCTAGACCGAAACTTATTCAGCTGCAGTAGAGTCGCGGCTGTTGTATGGGATGCATGAATAAAAGGTGACGCGATCGACAATTGAAGAGGGGAGAAGCGACGAAGAAAGATAGAGTGAGAGAGATGGATAGTGTAGTGAGGATGGTGACTGGACAGCTATTGGATCATTTTGATTGATTTCTGTTGGCAGTCCTT
GCTGATCCACGCATCGTTAAAAACTACATTCAAGAGTGGCATTATTGACACATTGGCCCAATTGACCAGTGTATTGAAAGACGGAAAATCCACTACAGCCTTTCGGTGATGCGACAGGATAGCCGGTGATTGAGAACGAATTGGCCTGTTGATCGGCAGCGACATAGATTCGTTCAGCTAAGGTACTCTCAGTACAAGTTCCTG
GTGTCATACCGAAGGAATTTTCTAAAAATGAGAAGCATTCGGCATCGGCATACATCTTGTAAATACATGGTGCAACGATAGGTGAGACGACTTGAAATGAATTAAATTGATCAATTGCTTGATTACATGCATTACCTTGACCGACGGCACTGGTACCGAACGAATCGCAGTTCGTATCATAAAAGAGTTCCAACTGATAAAAGT
TGGTTTGATTATTGACAAGCACGGTCAGTGATTGATAACCAAAGGTATTGGTGGGTAGACCAGTGGTGCAGACACCTGGTTGTACGTCCATTTCCACGATAGCTGCGTTGCTACAAGGTTGATTCGCTGGGTAGAAGTATACTGAGACATAGTTGGGGTCGATGATACGATGAGCACGGTGTGGATGAAGATGAGTTTCAGG
AGTTGGAGTTGGCTTCGAGTGTGGCTTGGTGGTGAGATGACGATGAACACGATGGGAGGCATGAGCTTGAACTGCGCAGATGCACGCGGCTGCCACGACGGACACGACGAGGAGGGAGCGAAGGAGAGACATGATGGGGGAAGGTTGGAGATGAAAATGTGAGTCAGGGAGTCAGAAAGTGTGTGTGAGTGATGAGT
GTGAGTGTGTGTATGAGTGGTGGGTCTCGTTG

>GRPc47613 (Virulent pathotypes only)

CCCTGACTCAGTCGAGTGGACGCCACCTTCGCAGTGGTTAGTTCTTAGTCGATCAGATACATAAAATCGGAAACCTGACTTCTGACTTCTGACTTGCACCTTCGCAGGACCTCGCTCGTTTTTCACAATGGCGCAGCTCCATCATCACATCATATTTACACTTGGTCTGGTAGCTACATTACTTTCTCAATCACTTCCATGCCATTCG
GTACGACTTGTAAGAGTTGATACGACTGCTTCTCCGACTGCCTGTGACGGTACCGAATATAGCGGAATTAGCTTCGTTTACCCCAATGGTACTTGTCTAACGAAGTCGACCACTTCGTCATCTCTATTAACCTGCTATCCTGATGGGTCTATCCAATGGTTTGATTTTACGGGTTCAGGTAATTGCACCAACAATGCCAACGTCATT
TATAACACGGGTAATTTAACCAGCACTGCAACTTCACAGAATCGTTGCATGCCCGTTTATCGTATGGATGGACAATATCAATACAGTTTGATTGTTGCATGTGATGGTACAGGTACACCAGCTACTGTTAGTCCAGGGTCACTGTATTACAAGGGTAGTAATAAACCATTAACCAGTGATACTAATATTACCAACGTGGAGTATT
ACACTTCATCTGATTGTTCTGGTCTACCACGATCCACTCTCAATAATATTCCCAATGGTGGTTGTTTACATGATCCCTTAGGTCCCGGTACTGCACAACGTTTATCATGCCGTTATGATGTTGATAAGGGAAATTTTCAATATGTAGAAGCCTGGGCTGGTCAAGCCTCATTAAGTGTACCAAATAATTGTCCATCGAGCACGATCA
ATTTTGACAATGGTAATATGGATGGATATCGTAACACAAATTATGGGGATTGTATTACTCTTTATGCAACAGATATTCCTTATGGCTCTTATAAATTTACATGTCCCTCCATGGCACCGATACCAGTGAGTAGTAGTAGTAGTAGTAGTAGCAGTAGTAGTACTGGTTCTACCCCTTCTTCATCTTCGTCCAGTTCTACATCATCATCTTC
G

>GRPc47941 (Virulent pathotypes only)

AGCCTTGCGGCCATTGTGACTCAGCTGCAGACCGCCGTCACGCAGGCCAACACCACGCTCCAGACCCTGACGGCCCAGCTGTCAACCATGTCCGCGTGCGCTGCCTCTGGCATGCTCATGGCGGCCAACGGCTCGTGCGTGACTGTCACGGCAGCCATTGCTGGGACGTGCACCAAGACCAACAACACGGTGTCGCCTGCGTG
CGCAGCTGGATTTGCGCCCGTGTCCTCGACCACCGACGTGTCGCAGTGCGCCGGTGTCAACCCGACGACCACCTTCGCTTGCAGCGCCGTCTCCTGCCCCGAGCTCAGCCAGGGCATCAACGTGCCCAGTGGCTGCAAGTGCCTGCCCGGCATCAACGGCACCGTCACCCCCATTACGAATGCGCCCTACTATGCCAACACCTG
CACGCACATTAAGTGCCCGGCCAACAGCTATGGTGACAATATCTTCTCTGGCTGCCAGTGCTCGCAGGGTTACAACGGTACCGTCACGCCGGTTGATTACCCTCCGTATTACGTCAGCACCTGCATGCTGACGCCGTGTCCCGCCTACAGCAACGGTGTCAACATTGTCAGCGGGTGCACGTGTAACTCTGGCTATGTCGGCGC
GGTCCAGCTGACCTCGTACACCCCTTACTATACTTCCAACTGCACGCCCGTGTGCGCGTTCGGCTCTGTGCTCTACCGCACCCCCGGCACGTACCAGTGGACCGCGCCCAGCAACTGCGCCAACTTCCCCATCACCGCTGTCCATGTCGTCGCTGTCGGTGGCGGCGGTGGTGGCGGTACCTACTCTGGCGTCTACGGCCAGAAC
GGTGGCACGTCGTACTTCATCTCGACGGGCTGGGTCTATGGCTATGGCGGCTATGCCGGCACCAGCGGCGGCTCGTACGGCGGTGATGGCGGTGGCTACGGTGGCAACGGTCCTGCTGGCGGTGGTGGCGCTGGAGGCTACAACGGTCAGGGCGGTTATGGTGGCAACTACAACACATACTCCAGCAACAACGACCCCCA
GGGTGGCGGCGG

>GRPc51259 (Virulent pathotypes only)

ATGCGGAAGTGTCGATGTATGCATGGATGAATTGTGACTGTTCATAAAAGTATGAGTTACAAAAGTTTAAAACCAGCCGCTAATAAGGATGGACCAGAATCTTTCGCATGACCCACTTTAGCATACTTAATACCTGATGCAATAATGGCATTCGAGACGTATGTAGCACAATCACCTAATTCAGCACCAGCGGGTTTGATATTCT
TGAGAGTGATCTCAGCGATATGTTCAGCACGTGTGAGACATGGGGTAGGAGCGGAGGATGAAGTACTACTGTTTGTTGAACTTGATGAGGATGGTGCAGCAAACGATTGCTGTACAACTGGTTTCATTGTTGCTGTCACTTGGTCATGAGTTCTGGAATCGTATTTGACGTGGGATTGCCCGTGGCAGCCGTAGCCACAGCTG
ATGTAATGGCTGTTGCGGTTGCCGCCGCAGCTGCCGGCGCCCCAACCAAAACTGCACCAGCCCCACCTGTCACAGCACCAACACCAGCCGCAGCTACAACAGTGACAACCTTACCCAAGAACGATGGCTTTTGACCTGGTTGCAGCTTGGGGCCAAAGAGACCATATTTACGACCGAATAAACCCATACTGGCGTGATGGGTG
CAGCAACACAATATCTGTGCGGATAACCAGAGACGGGTGCGAGCAGAGACGTGAACAATCCAGCCATCATCGAAATGACCTGAACAAACGACCTCGAGCTGCGTAACCTAGACTAGCCATCAGAGTCGACCGGTTACACTCGGCCCTTCTCCACAAGTGGCCGGCGGATTCATACAGATTCTTACGGTCCTAATCAATGGCGC
GTGACTCGAACTGACAAATAATATAATGGCAATCAGCTAGAATGAAGTAATGAATCGCCCTGGATATATAAGTATTGGATATATTTGAAAAGGGAATGTTTAGTACAGTCAGAATTCGATAGTTCTTCGTTCGTAT

>GRPc52395 (Virulent pathotypes only)

AGGGACTGCAATGTTTCGAATCAACTCGCTTAGAAGAAAGAGAGAATCGTGGTGACGAGCCAAAAAGGGGCTCCATTCAAGAAGGCTCCAATGAAGATGAGCCATGGCAAGTACAAAACTTGTCCATCAAAACGATGTGCCAAGGAGAAATGCGCATCCTTCCAAAACTCGTAGCCACAATTGACATTGAGGTATTGGATAT
CAGAAGGTTCAACATGGGCAGGTCCTTTCTTCGACTTTTTCGATTGAAAGGCAGGAAACCACTCATAAATTTCAGTGCTGGGAAGCTTGATAGCAAACGGAATTGCGACACGTGCCCAAACACTGAGCAGCGAGACGACGAACACAGATGATTCGAAGCTATAGGTGGGCCAACGAGGTCCGAGAACCACAAAGGAGAGTG
GCAAATACCACAGATGATGATGACAAGTGGCCTTCTTGACCCAAGAAGTGTCAGGCCAAGTCAAGTAGCGGGCTGCGCCAATGGGGAATTTGCCGGTGGTTAAATATCCACCAATGTCCACGTACCAAAGCAACTGGTCACACAGGACATTGATCATGGCGATACTCAAAAGACGCGGCCTCGAAACAAGACAACCAATCCC
ACTCATTAGCAGGCCGACGTTACAAGTCCAGGACATTTCGTACAGAATGACATTGCCAGTCTCAAACCACCGAAGAATGCACAACATGGCCAAATAGCCGAGAAAAAAGAGGCCGATTTTCGGAAAGTAGAAATCGTCTGACAGAACGACTCCCTCCCGTGCGTCGATGTAATCACTCAGCGCAATTGAGAGGAGGGGAAAC
GCGGTCCACGCGATGAGAAGAGTCATCAATGGTCTCCATACTGATATGGATGCGCTTGCGCTCGAGTGGCCCATCTGTGAGGTGTGTGCTGAGAGGCCAAATCTCCGGTGGACACCGG

>GRPc52542 (Virulent pathotypes only)

GTGGTGTAGTGCTGCTCATCGAGTTCCTCAACGCATTCCAATGACTCCCACCGTCTGTTTGCTTGGTCTGACCATCCTCCTCGGTGCATTACAATGCCATTCTGCCCGGCTTGTAAGAGTTGATACAATTGCTTCTCCGACTGCATGTGATGGTGACGAGTACAGCGCAAGTAGTACCATTTTCCCCAATGGTACTTGTCTATCGGG
TTCTACCACCTCATCTCTATTAACCTGCTATCCTGATGGGTCTATCCAATGGTTTGATTTTACGGGTTCAAATAATTGTACCAATAATGCCAATGTATTTTTGAAAGCTGGTAATTTAACTAGTGCTGCAAACCGTTGCATGCCCGTTTATCGTATGAATGATGGACAATATCAATACAGTTGGATTGTTTCATGTGATGGTGCAGGT
ACACCAGCTACAGTTAGTCCCCAATCACTCGCTTACAGGGGTAGTGCTCAACCATTAACAAGTGATACCAATATTACGAACGTCGAGTATTACACCACGGCTGATTGCTCTGGTAATCCACGTTCCACCCTCAGTACTATTCCCAATGGTGGTTGTTTAAATGATCCATTAGGTCCAGGTACTGCCCAACGTTTATCATGTCGACAT
GATATTGATAGTGGTGATTTTCAATATGTGGAAGTATGGGCTGGTCAAGCGTCATTAAGTGTGCCAAGTAATTGTCCAACACGGACCGTCAACTTTGAGAATGGTAATATGGACGGATATCGTAGAACAAATTATGGAGATTGTATTACCCTTTGGCTATCAGATACTCCTTATGGCTCTTATAAATTTATTTGTCCATCGATGAT
GCCACCACCAGTGAGTAGTAGTAGTAGTAGTAGTAGCAGTAGTAGTACTGGTTCTACCCCTTCTTCATCTTCGTCCAGTTCTACATCATCATCTTCG

>GRPc52554 (Virulent pathotypes only)

CAAAATACGGGTTCCTATTCTCGCCCAAATATCTGAGCAGTTATGAAGATGCCAAGGTATCCTAAATTTGTCCACGGAAAACAAAAAAGTTAAGGTACAGCCATATATGCAATGGCCCTCTTCTTCTAAATCATGCATTGCCACACATTACATCCTAATTTCCCGTGACTCCTTCAACATAAAACACATCATGTTTACCAAGTCCAA
ATTTACAGAGCAAAGACGAACGCCACGAGGCCGAGGAGGGCAGCGGAGCCGACGGCAACGGCGGGAGCAGCGTTGGGGGTGGTGGAGGTGACGGCAGCGGAGGCAGAGGCAGCGGCGCTCGCGAGGGCGGAGGCGGCAGAGGAAGCGGCGTGCGAGGCGGAGGCCAGGGCGCCGGTCTTGGAGGCGACGGCAGC
AGTGCCGGTCTGGGATCCGCCGGCGGCAGCGCTGGAGCCGGAGCCGCCGTTGTTGGGGGTGGAGCCCTCGGTCGAGTCGAAGTAGTGCTTCTTGATGCAGGCCTGGGTGCACTCCTCGTACTTCTTGGTCTGGGCCTCGGAGCCGTCGCCCTGGGGGCAGTCGGCGACGCAGTTGACGGTGTCGTTGGCCTGCTTCTCGTT
GGGGGCGGGGACGCTGATGCACTTGGACTGGCAGTCGACGTCACCAGCGTCACACTTGTCGAGGCAAGAGTAGATGGATGACTGGGCAGGGTCGAGGGCGTAGGTCTGAGTGGCAGCCTGGGCGGGAGTGGTAGTCTGGGCGCTGGCAATGGCGGCGAAAGCACCGGCGACGAGGATGGAGAAACGCATGTTGATGG
TTTCTTTGTTGGGTTATTACAGAATTAAGTCTTGTAAGGTTTGTTTTTTAATAACGACAGTAGACTCGAGATGAAAGAGCGTAGAAGGATATGAATCGGTCGGTCTAGCGAGGGAAGTGGAG

>GRPc52865 (Virulent pathotypes only)

ACAAATAAATTTATAAGATCCATAAGGTTTAGAAGCCAAGTACAACGTCATACAGCCCCCAAATTTGGGTGCAACAAAACCAGTCCAACCTTGAATATCACCATGTGGTGCACAAGTCGTACTAGTTGGTTGACCCGAAATGGGTTGACCTACCCATACCTCAGCAAACTGAATCTTCCCATAAAATGATGGATCGTTCATACAG
GACAGACGAATACCAGCACCAGGATCTTGAGGAGTGAGATTAGTAATGCAAGCACCGTTAGTGAAATTAGCATAACTTGTGAGCGGTGCACCTGAACAATTTGCATTCGACCAATATTGTACATTGGTTACGTACGTATCATTTGTTAAAGGGAGGGTGTCGGAATATACAGGTTGAGTATGAGGATCGGATGTGACGGTATC
CCCAGATCCATCACATGCGAGAATAACACTTTCAGCGAGAGCATTATTCATGTAGTACACAGGCATACACTTATTGATGGAGGTCGCTGTTAAATTACCACTAACGAATGGAGATGTACATGTAGAATCACTGTACCGAGTAAACGATTGTGACCCATCTGCACCACATGTGAAAATGTTGGATGAGGTGGAACTCGTAGCAAG
ACAAACACCCGATTGATAGTTATGACTATCAGCGCTCACGCTACCATTACATGATGTAGGGGTCGTATTATCCACCCCTAGAATAACAACATGGATAGATTGAACGTGGAGTATGGACCAATCACCAATCAGACTCAAAAGACAAATAAAAGCCAGCTGAGTGAAAGAATTCATGGTAGATAATGGATGAATGGATGATGATA
GGCTGGATGGAAGACAAAATTGGAGACGACAGTTACTCTATGGCACATGCATGAGTCCAGGATAGCTGTACGTACAGCGAGAAGGAGTCGAGGGTTGCAG

>GRPc53066 (Virulent pathotypes only)

GGGTGAACTAAAATAAGCCGAATTCAAATGAAGTCGAAGATAAATCACACGAACGGGTGAACCGCTCACGGAACTTGAAGCTAAACTTGAACGTGGTAAACAATCGTTCGAGCCGACCTGGTCACCATCATACCGATTCGAGCTGGTGGTTGGCAGTGATGTTGTTGACTCCGGACAATCCGGTGCTTGGCCAACAAAGAGG
TGAGATAAATCGACATTCCAGGCGGTGTTGACACCGAGTGTCAATCGATTTGAACCATACCACCCGTCACTAAAGGATTGGTCACTCGTCGTTGTGCCGCCACCTAAAGAATTCGACGTTGCCACTAAAGCATCGAAATCAGGATATCCGTACTATTCACAGGTTGAAAGGGTAGTAGTTTCGCCGCCATCGGTACTACGCCAC
ACGAAGGTGCTGCACCAAAGAATCCAGCACCACAGGCACAACCTTGATAAGCATAGAAGCTGGGATCTGTACGAAGCACAGTACCGAGATAGATGGTGAATGAGATGGTAGGACAGACGAAATCACCGAGATATTGTAAATTGGTCGAGTCTAAGGTGAACATATCCGGTAGGGGTGAACCGGCTTCAAGGTATGGATTAC
CTAATAGATTGACGTAACGAAGTTGATTCACTTCAACGAATTGACCAATGATACTAAAGCCAACTAATGAACCATCTAGGGTCAAATTATTGTTCGATAGATCAAGAACTTGTAGAGAATCCGATCTCACCCACGAATATCGTGCGATGTGATTATCGTTTAGTGTGAGGGTTGCTAACGAACCAGTCGGTGGTAGGCTCTGGA
GCCAATCACAACTAAAGTTATTGTGGTGGAGTGCGAGTGAGGTTAATGAGGCGGAGAGTAGTGAGGAGATGATGACACCGAGTTCATTCGTCGATTGGCC

>GRPc53804 (Virulent pathotypes only)

ATCAAGGGCAAGGTGGCTCTCCTTTGTTAGTTCAGACCTCGGCTTGACTTTAATTCGAGTGCCGGTGAGAGTTAATGGTTAATCGCCACCTAAATAGTTGATCACTATCTGACTCAAGTCGGGTATGAGATAAAGAGTTAATTGTTGATGGATGGTGGGCCAATGATGTTTGGTCCAATTCAATTTCATTTCTGCGAGTAGATCG
ATCAACTCACCCTGATATTCTTGTTCATCATTTGTCTTAAAATACAACGGATTTAAGTATTCAACAAATTCGCTCATTCTCATACCATTGATACGAATTGGTATCGGACGACTGGAAGGTCTGTGTTGAAAAACGAAAGTTGATGGGCCGATTGTTCGAATTGGTTGAAATAAGTCACTCGTTTGGAATCTAGGTGTGGTAAACAA
TGTTTTGATCCAGTGATACACTTCCGGATGTTTGGAACGATTGGTATAATACGTTACAATCGCACCTTCTAACAACGTACCTAAACATTCACCTTCGTGTTCTACCAAATAAGTGATATGAGTCAAATAGTTGAACCATACATCATCCATTACATCCAACTGACTTGATGAAGAGTTAAACTGAACGAATGATAACAACCATAAAC
GAATTGTCTTGACCATACTCAGATGACAATCTCTTAACATTGCATCTAATTCATCAGTTAATTCATCCAATGGAAAAGGATAGAATTGGTCATCATATACATCGCGTAATTGGTATGAGTAGATCGTTGTTAATTGTTGGCGCTTATCAGATGACTGATTCAGTTGAATAATTTGATTTAGTTGACTCATGAGAATATGCCATAGTT
TGTTCATGAATTGATCGAATTGTTCTTGGTCCTCATTCGATTTATGTTCTTCCATTTGTCGGTGGTGCGGT

>GRPc54577 (Virulent pathotypes only)

AAACCTATGGAATTGACTTGTTCAAAATTGATGGGTCCAGAAACACAAGAATTTTCGTCGGCGATCGCGTGAATAGCGATAAGTTCGATCAGTCAATCGTTGACACCACCGAGTATGAAACGAATTGGAAGAGGTATCAACCATGATGGAAGGATGTGGAGAGTGAGGGTATGCGGACGAAAAGTGTGCGCCACGTGAGCG
ATGGGATAGAGTGGCAGTATCACACTGATTCGGGATGTTTAATCAATAAATTCGTCTAATTCGGCATCATCAAAATCATCACGTTCATCATCCATCGTCTCATCCTTCTCCTCCTCTTGAGCGTGACGTGATTTGGGTGTGGTAAGGAGTAGGATTTGTTTCATCGCTTCTTGGACCTTCTTCTTACGTTCACTCGTATGAGTCAAC
ATCGATAAATTCTTGCGTTTCATATCGCCCGTATCAACTTTCGTGAGTGAGCGTTCGGCTTGTTCGAGTAGTGTTGGTCGTGGTAATTTTTTCTGTAATGCCTTGGCACGCTTGATCGCTGAGGTGGATAAGCCCGTCTTACGAAACCCAGGCGAACGATTCGTGCGAGCAACAGAAGTCGAAGGTGCTTTCAATTTAGAGCCAG
GGGTCTTCTTGGAGGAGGGGGCAGCTGTGTGATAGGTGCAAGAGGAAAGTCAGCGAGGAGATCCTCACGCTTGACTGAACGCGACGCGGTGGATGGAGCCACTGATTTCGCACTGCCTGATTTCGAAGGTGCGGCAGCTGAACTCGTGGACGATGAGCTCGTGCTTGTTGTGAGCTGAACGTGCTGCTGGAGATGCTTAC
GAGCGAGTGTCAACATGATGATGAGGAGGACAGGATGAAAGATTAATACTTGTGGGTCAGAGATAG

>GRPc55119 (Virulent pathotypes only)

AAACAACTCACCTCAGTCTTTACCGTGAAACAACACACTCATTATGAAGACCGCGACACTCATCTCCCTCGTCTCCCTCGCCCTCTCGGCCTCGGCCAAGATCTGCAAGCCCAGCGACATTGACTACTGCGACAAGAAGCTCCAGTGCATCCCCAAGGCTACGCTCGAGGCCATCGTACCCAAGTACGTCGCCGCCTTTGCGGGC
ATCGAGGACGGCGGGAAGCAGGTCCGCGAGGTGTTTGACGAGAATGTCGAGATTTACTCGCAGACGCAGTGGTGGGTGTCTGGCGCACCCGATGTCATCGCCGCCCACGCCGAGAAGGATGACTTTCCCCCATCCTCAAGGGCACAGCCCAAATGATCCAATTCAACTCGCACCCGTACAACCCCGACGCCTTCATCGTCG
GCCCCGTCGCCTACGGCTGCAACACCTTCAGCTTCTACTGGAAGGGCGACTTCAGCATCCCTGGCATCGGTGTCCCGCGCGGCCGCACCAACGGCATCGACATGGTCTTCCTCGACTACAAGACGGGGCTCGTCACGACGGCGTACTCTGAGTACAACAACCTGAACCAGATGTACAACTATGGTGCGCACTTGACGTGGAGT
GAGAACCCGACGTGTTGTGAGTGTCCTACGCCGTTTGATCCGTCGTGCAAGTGCCCGGCGTCGAATTCGACATGTCCGGTTACCAGATAGGGAGGAGATGAGGGGTGTTGATGGGTCTGGATGTATTTCATTTCGACTGCACGGGGTTTCTCGATAGGTCTCAGTTGATACTATTGCTGAGAAGTCAACTCCGTGGCGTTGAC
TTGCATTGTATAAATCACTTGACAAGAACTCACATCACGCACGCAACTTACAAACG

>GRPc55920 (Virulent pathotypes only)

TGCTTATTTGACTGATCGACGACGATTTAGTCGCGTAATTGCACGTATTCAGCATCCGAGTTGGTCACGTGTGTTCGCTGGGATTGGCGTCATTCCGATCTACCATCTCAGTGTTGAGGCACAACTCCTCACACAACTGCTCACAGATGACTCAAGCACAAATCGTTTCTCTACTACATCACTCACCATACCATTTATCTCGGCCAG
TTCCCTATTTGGTGCTACCACTTCACCCACCTCGATCGCAGCTGCCGCTGGTATAGGTCTGAGTGAGCTGATGTCCTCGACCATTCATTCCGAGGATGATTTCCTGTGTGCGATGCAAGATTTGGCGCCATCCACTCGCCCTGAGTTCCTTCGTCTACATCCCACACCACCTGCACACCTATCCGCTGAAGCCATGCAAAAGTCGTT
GACCCGTCCCCAATTCTTGAGTGTCGTTTCTGAGGCTCATCGCGATGGGTATATGACCGCGATTGCTGCCTTACATCACCATACAACGAATTCAACCAAGAATGAAGTAGCTGATGAAGCTCAACGTGCACAGCAGGATGTTGCATTCGGTGGTGGATTATCAGCCATGCGACATCAACAACGATTACAACGGGCAGCATCAAC
AGTCAGTCAGCAACCACCCAATTCGATCCCATCACCTACACGACCAACAGCTAATGATGATGATGAGAACAATGTACGAGCAATATTTCATGACTAAAATAGTGCCCGTAATCGATTCATTCGTATTTCATGGTGGAGTCCAAGACCGATCCATCACCATCATCATTCCATTCGAATGAAGCTCCTAGACCTTTCATCTGCAATTGTTTA
TCTCTCACACCCATTGTTACATACGCTGG

>GRPc56708 (Virulent pathotypes only)

TGTAGATGCCAATTGAAGACCATTCGGAAATTGTCGTGTATCACAAGACCGCCTCTCATTGCAACATGTGCCATAATAATTATATGTTTCACAACAATTGCTAAAACCTGTACATTGTGTCAGATAGTAATAAGAACTACTACTACAACCGGCACCGCAATATGACCCTAATGGATTGGCACATTGATTTGTATCGCGTAAATATT
CCCAACCATCCGGACAGCTGCAGATCGTACCATAGGGCGGATTAATTAATACTAACGTCTGATTCAGATTAGGACATGAAACTGTAAATTGGTTTGAGGTGAGTGCAGTTTGTGCATGGGTGAAATGCACTCCCAACGTTAACGTAAAGATACACCACAGCCAACCAGACATTCGGGGGAATGGAAAATATCGTGATCGTAAT
ACTACTGACGCCGGCTCTAAATGGCACCACAATGAATAAAGGTAAGGAACAGTCTAAAGAATCAGGCAGGTGAGTGTAAAGAGTTACTTCGTTGCGATGTAAGGTAGTCCAGTCGGCTCGCTGAGCGGCTGACTCGGGACTCGTTCCCACGTTCCCACCTCCAGAGGAATCAATGACGAACCAAAGAGGTGGTATGGCGAG
GCGATGAAGTGGGTTGAGTGTTAATTGGGTGCCGGTGATTGGAAGGATAGTTGGAGAAGTTTGGTCTGTTGTGGAGAGAGGGAAACAGTTTGGCTCACCGTACCAAGTACCACAGCGGGTCGGGAGGTAACAGGTGGTCAGGAGAGGTCGACGATGGAAATGTGTTTGACAACAAGTATTCGAATTAACACCATAATCAG
GGCAGGAACCAGGAACACGACCCCAAAGA

>GRPc57784 (Virulent pathotypes only)

ATACGATAATTGACTCGCTGTTAAGTTCCAGAGCTTCTCAGTATTGAAGCCAAAGAAAAAGAAGCCACGATTATAATTGACGGGAAAATAATTATACATGGCACCATCCGGCTGTGGTACGGTCACGAGTGAAATGGTTTGATAGTATGATAAGAAGAAGTCGTCGCTGGCTGGTTGGAATAAGATCGCGCCACCACTTACAT
CGATTTGAATCCACTCAATCCACGCGCCATAGTACTCAAAAGAGAACGGTATGGTATGACTTAAACAGAATTGATTCTGTGGTAGATAATCTCCCCTGCTTGGTGAGGGGGGTCCATCACACGCCAGTTCAACATACCAGGTCACGGGCTGCCAGTTGATACCACCCGGCTGATAATAGGTGGGAAATTCCGATGGTGAGATT
GTCGAACCGGTCGTAGTGAAGGGCGCACTTTCCACACCATATGCAGTACCAGTAACTTGACCATCAACGAGTTGCCAACATTGACAACAAGAGTAGAGGAGAGTGAACCAGCAGATGAGGGTGACAGAACGACAAAATAGAAGGTGCGTCACCAAGGGCTGCTGGGGATGGCTCATTCCCCGCGCACTCAACTAAGGATGA
GACTCACTCAACTCACGTGCGTGATGAGTGGGGCGCGTCTGGATGACCGTTGGATGAAGTGGGTGGGTCCTGTGCGTATCAGTTGTGTGGAGTTGAGGAGTCTGTGTGATCCGCTGGTGGATTGAGAATGAGCCATGCAAGGTGTTCAAGCTTGAGCAGATGAGGTTGATAAAGACAACACTGAATGAGGGCAGTCACAG
AGTTGGCACTGCG

>GRPc59178 (Virulent pathotypes only)

CGCAAATCTTATTCAAAGGCGGTATCAATGTTCATCCATGTATCTTCAGTCACCAAATGACCTCTAGCTGTCTAAATTCATGCACTCCTTTGCTACTCTTGCTCATGAATAAACACCCATATCGTTCTTCCTTGAGTGGAGTTTCGAAAAGGGTAGCGCATGATGAGTGGTGGCGTTCGGTTCCGTCTGCGATGTGAGTATGAGTT
GGAAGGAGATGACGGTTCAGGGATTGATTCGATTCCAGGTTGAGCTACGAGATGAACTTGGTGAATGAGTGAATGAGTTGAGGGGAGTATAGTGGTCAAGGTGCGATTTGAGTATTGAAATTGTTGGTGTTCAGTCACAGTTGTCTGGTTGTTTGATTGTTGTGTCCGATTGGGATTTGGCTCCCTTCAAAGACGAGCTCGTC
TTGGTCATTGAGTAGAGACGTTTCTTCATTTCCATGATGAACACGTTTGAAGGGCATCCCAACAAGAGAAAGAGAATGACGATGGGCATCATCCGGTAGAGTCACAGTGATTCCACACCATGATGATGATGATGGTGATGTACTTGTGTGTGATTTACTTGTTGTTGCGTGACCCGTGGTGCCCGCTTCATTCCTGCAAAAGCA
GTTAATGCTGAATCACTTCCACGTATCAAACTTGGTGCGTTGAGTAGATGGGGGGATGGGTGTCGTGATGGCTGAGTACTCATATGATGGTGTGAGAAGTGATAGCCTTTCGCCACCACTTAATGAAATTCAGCGGGCACTGCTTCATCCTTCTCTTCGTCCTTCCATGGATGGTTTCGCTTG

>GRPc59431 (Virulent pathotypes only)

GGGAGAAGAAGTCGAGCCTTGCTTCACCAGAAGAGATGAAGGCCGTCTTGTTTGTTTGCGCCTTGCTGGTCGCTGTCGCAGCTGCCCAAGAAGTGGCATTTGCTAATGGAGCTCGCTTCCAACAAGTGGGCCTCGGAGTGAAGGGCAACTGCTTCCCTGATTGCAAGTGGAATTGCGATAGCCCATCTTGTCCCGCCGTTTGT
GAGCCTCAGTGCGAGAAGCCCGCTTGCGAGCTCCGCTGCGAGAAGCTCGGTGAGACTGCTTGCGATATCCGTTGCGAGAAGCCCAAGTGCGAGGTTCGTTGCTCTCGAAAGAACTGCCAAGCGAGCGGCTGCCCTCTGTGCGAGAACGTCTGCCTGCCCGCTAAGTGCCAAACTGTCTGCACTCCCGCTGTCCCCAACTGCC
GACCCACTTGCTCTGCTCCCAAGTGCACCTGGTCCTGCAAGAAGCCCGCTTCTTGCCCCAAGCCCAAGTGCAACCTCGTTTGCAACTCCCGTCCTACCCCCACTTGCGGCGCCAAGCCCCAGTGCTGCAGCTGCCACAAGTTCAACAACATGGCCGCCGCCGTCGAGCTCGCTTCTTCCACCCATCTCCAGAACCATGGCGCTCTC
CCATCCGACGAGCACATGCCCGCCCTTGTCGAGGTTGCTCACGAGATCCGTCTCCACTCCGAGTCTGGCGCCGAGTCTCAGTGCTGCCCTTGCGCTTCTGAGGAGGAGGCTTCTTTCTAAACAAAACAAAACACACTTTAACTGCTACGCGTGATCCATCCATACATCTACCATTGATCAA

>GRPc59836 (Virulent pathotypes only)

GGAGGCGCAATTGATGAATCGGAGGGAGGGAGAAGAAGGGAGCCAGTATGAGCGAGAGAAATGGAGCCGACTTTAGACACCTGTGGTTGTGGAAGGGAATCCCATTGGTGGTGGTTGTGAATGGATGGCATGCTGTTGTTGAGGTGGGTAATTATATGGAGCTGCGCCGGGTGCAAATTGAGGTTGAGGCGGTTGTTGC
TGTTGGACGTAAGGATTCATTCTCATCGAATTCTGTAACTCTAATGCCTTATTTCTCTTGTGTGAACGATAACAAAGCCAACCAATAACACCAATGAGAATGAGACCACCGAAGATGCAACCAACAATGATACCACTCTTGGCAGCTGATGAGATGGCTGGTTGTGGTCCAGCTGCGGTACAATCGATGTAAGCACTATCAGAC
GGTGAGCCATCGGTGAATTTGAGGGGAATACATGTTACTCCATCACCTGCACCTGATCCAGTCGCAACGAGATCTTGACAGTTGCTACCCAGAAATAGTACGGCTGAGGTGACAAAGTTGTTGTCATCGCATTGAACCTTGGCGCTCACACGACTACTGATGGAAATACAATCGCCACTGGCCAGATTAGTGGCAGACACGAG
GTAGGTGCTTGGTCCGCAATTCGAACTGTCGAACAGTTCAGCACTCAACCCGTGGACTAATTCAGGGATGTAGCTGACGAGGATCAACGTGGCCACGAGATAGGTGAGGATACACATGGTTGTGATTGAGAGAGTGAGTGTGTGACCTGTCGGTTGCAGGCCAAACCAATTCAAACACAG

>GRPc60029 (Virulent pathotypes only)

TATTTGTATGCCTTCTAAACATGCACGTACAATATGCAAATATACAGAGATTTCTAATAAATTCGTGTCTGATTGCGTTGGGGTGTGATCTGGATAAACTCTCGAACAATGAATGGGCGGGGAAATCCCGCGTGAATAAATTAATATCGTAAACCAGCGACTCGCTTCTCAGCTGGATAGAAGTATCGCGTGCATCTATCTTTTCC
ACCAAAGAAACATTGTTTGACTGGATAGAAGCAGGCACGAATACATTTCATAATTAATTTTAATGGCAACAATACTAATTTAATCGTGTGATAGATGCACGTCTCCATACAAGCAACACAGATAGCCACTTGATCACGAAAACAGTAACCGATGACGCACACCACAATGAGAATGACGAACACAAAGATTAACCAAACCCATTG
ACTATGCAATAATTTAGTGATGGGATCCGACATTTCTGCATGGACGGCAATATTCAATCCAGGTGGTGCGGCACCTGTAGCTGCACCAGTACTCGCTCCAGCTGCATGTCGAATGGAGGATTTGAGTTGATTGAACGAGATGGAGGAAACGACTTCATGTTGGTTTTGGTCTATCCGCCGTTGGTGAGTGAGTGAAGTGGCTG
AAGTGAATGTCAGATAGCAGCAGCTGCTGGCTGCAAGTAGTAGGATGGTGCAGAGGAGGAGACATCGATTCGAGATCGAGGAGCACCGTATCATACGTGTGGAGCTCATCTGGAGAGAATCAACACTGCTGGACAATCAGCTATCGTACTCCGTCTTCGTCAACGCGAG

>GRPc61629 (Virulent pathotypes only)

CAAGGAGAAATGAGAGCCATTGTGCTTGCCACGGTTGTTGCCCTTGCGGCCGCCGTCCCGTCCTCCGTCCAAATCTGTCGCCGCTATGATGACGTTTACCAAGTTGGCTCAACTGTCTACACCAGCGGAGGCTCTGCACTTCCTTCCGGAACTGTCCTCTCGCCCTACGACAATGTCGCCTGGGTCACCAAGACCGTCGGCGCTG
ACTGGGAACTGGCCGAAAACGAACTGGCCGGACCTTGCAGCACCGCCAACCCCGTGCAGTACTACTGCCACGCCTACAACTGGTTGCACGAGGGTGACATCAACGACATTTCCCAGCTCAACTACAACTGTGGTGCCGGCTACATCCGCATCCTGGACAACTCGGCTCACTATTCAGTGCCCATCTACAACGTCACTGACGGTC
CCACTTTCCTCAACTGCTTCAAGTACGACATTGTCAACCTGTACGGCATGTGCTCCAACCCAACCCCCGTTTGCAACCCTGGCGACAATGTTGAGAACCCCGCCGAGCTGGTGCAAGTCCTGCCCCAAGTTTATACGGATTTCAAGGCTGCTATCCCTGGGGCCCTCCTCAACCCCCGCGATGGTCTGTGCGGTGAGACCCCCGC
CCCTACTCCAGCCCCGACTCCTTCTCCCACTCGCTCGCCTACTCCCCAGCCGACTCCTTCTCCTACCCCCAACCCCACTCCTCAGCCCACTCAACGACCTACTCCCCAGCCTACCCCTCAACCCACTCCTCAGCCCACTCCCCAG

>GRPc63379 (Virulent pathotypes only)

AATAGTGTTAATTCTGAATGTGTGAGATAGTTCACAATCAACTGAACTTCAATCTGAACCAATCGTTGCATGGTTGCGTGTGAGTTGAGTCTGATTCTAGATGAGGTCTGTTGGTCGAGCTGCACGGTTAGTCGTATTCGTTCGTGTTCTGTTCTGTTCGCCTAGCCTGTATTATCATGTAATCGTGCGTTGTGTATTTGGTTGTGT
AGTGTCAATCACAGAATCTGGAGCCTGGTGCTTCTCGGCTGAGCCAACCGGTTCGGAATTCTTATCGATGAGATCGAGAAATTCGGTATGGTACAATCCGCAATAAAATGTAAATCCACGAAAACCATATCGATCCCACAGTCTACAGAGCAGCAATTGGTGCATCCGTTCGTGTTGCTGATCGTTGTGCGTTACAGCTCGTTC
GAGGGCTTCGATGCTTGCTACATTTTGATGGAGAGCAGCCATACCAGCTAGATTGTTGAGAAAAGCAGCTAGGACAAACCATTGATGACTGCTCCAGTGCAAGAGACGAGGTTCATCTGGATTTCCAACGAAAGAAACTTGAGTGAGTTTAAATAAAGCGGCAGGTATGGAGAGAGGTATGAACATGAAGACAAGCATCAT
GAAGCAACACTCTCGAAAGAGAATAAGTAGCTCTCCACGACGCTCCCTAAGGATTTGCCATCGGAGAGGGAACAAACGCCAACCAGTTGTTAGCCAAGGTAATGCCAGAAAAGCGTTGAAG

>GRPc63433 (Virulent pathotypes only)

TGTTATTACATTGAAGATTGGTTCTTGTTGTGAAGCGCCACCCACGTATTTATTGATGGGTAACGCATTCCTAATATTCTTAATGCTTTTATTTGTGTTGATCTCATTTGGTGTCACCTGGTCTACAGCCACTTCGAATACAACCGATTATACCAATGAATTTTATATTGGTCACTACAAGGAATATACGAATACAAGTATTATCACG
GAGGATTATACTTCCGCCACCGATTGGCGTCAACAATGCAAGATTGATAATATCCGTGTCCCGGTCTGTCTCGGTCTGAGTATTCCTTTTCCTTTCCTGATCTTCTATTTTATATGGAAGCGCCGCCAGGATCACTTTGATCCGAATATGCGTCGTGGTGCCATCCTGTGCTTATTAATCTTTGGTGCTTGGATGTTTATGGCCTGTG
CATTCTGGTGGCAAGGGAGTTATAATTTTATCAAGGCTGCTACCATTGACCCACCAACAGCAACTGGCACACTTGCATTAAGAAATGTCACCTTCTCAGGTGGTTGGTGGTTATCCATGGTTAATTTATTCTTTACTGTCATTTTAATAATTGTTCAAAGTACGGCCTATGGTGAAGAGAAATGTTGTGTCTGCTCATGTTAACACA
CACACGAGCTCACACCTAACCATCGATATCATCGGTACTGATCTCCCTCGATCAACCCCCCACTTGTCCATCAATATGTTCATATCAGTCAATTAGAATTAATCTACAA

>GRPc63595 (Virulent pathotypes only)

CGGATGTCTTTCTTTGACTCTAATAATTCCTTATGATCTAAAGAGTTTGATCCAGAGACTCACAAATTGGTTTGATCGAATAATAATCCAGATGAGTTCATTTCATGATTTAATGGTGGTGGTGGTGGTGGTGTTGATGAACAGGCGAGTGGTGTTGGTAATGTCCAATTTTGCACACAACTATTCGGTGGAATTGTAACACTTAA
TTCCCCTTGGGTTTCAGCACAGGTAAATAAAACATTCGAAGTGTAAAAGGTCGGGGAGCCAGTTGCCCAGCATGATTGTGCACCTTGAATCTGATACATAATTCCGAGTTCAGGATTGGATGGATCAACCCAACCCCAAACAGGATCTGATGAAGCACTCCATTCACCCGCAATCCATGCTGGTTGCGAAGGATAATTAATGTC
AATCTGACAGTTCGACGCAGAGGGCTGTACGTCGGTACAGCGAGTTGCTTGCTCTAATGGTCCACATACATGCATCGTGTAAGCATAATTACTATCACTCGCGTTAAAATCATCACTCGCTTGTAATGAACTTAAATCAAAACCATTCCAACTGCACGTGGGAGCAGCAGTCACAGCAGTCGTAGCAGCAACACAGACGACGAC
ACCGACGAGGATGAAAGAAGACAAGATCTGAAGACGAGTCATGATGGATGAAAGTGTTCGGTTTGGGTCGAATTCTCCTGAATCAATGGTCAACGTGTGTGTTTGTGTGGTACAG

>GRPc64219 (Virulent pathotypes only)

GTCAGGAGTGTTGGAGAAGTGTAAACTATGGCGACTCAATAGACCACTATTCAAAGAATCAATATGACGCCCTTTATAGTGAGTGCATTTATATGCACTCTTATAGGCATACTGATCTTAATCAACCTATCCACCTCATCAGCCAATCACGCTTCCTTCTCCTCCTCTCCTCCTCAACCTGACGCTACCACCATCCACACCCATCCTT
CACCACCTTCACTCGATCATCTCCTCTCGACACCCACCTACACGTTCACCCCCATCTCCCATCGTTATCTCGTCGGCTTTCATGCCTACGCCCTTGCTGAAGAGCATGAAACGACCTTGCGCACTCTTTGGTCACAACTCCCAACCATACATCAACCCGAAGAGGAATGCATTGAAATATTCCCATTCACATGCCATTCACCACCACG
AGTCAACTCTACCACGCATCATCAAGAACAATGTGCCTCTATCGTCGGTACTTGCACCCTTCAACCACTCCCACTCGATTGGCAGATGATCCATCGCACCAATCCTGGTTCATCCCTACCAACCGATTTCTTACTCGTAGAACTCCAGGGACCGGCAACCATCTACCAAACTGAAGCTGAAGTGCGTGTTCGAATGAATGAGATAT
GGACGGAATTAAAGCAACAATATCCCACGATTATTAAATATGTGACCGAAGAGAAGCGTGTCACCAGGTCTCTCATGGAATTCATTGATCAGGATACGG

>GRPc64229 (Virulent pathotypes only)

AAATAATTGTAATTGCTTTCGGTGTGAAAAAATGAATAAGGAGGCGAGAGGAAATAATAATAAAGCACAGGTCAGGATAATTAATCAGCAGAAAGTACCAGCGTTATAATGATAACTAACTATTAACGTTATAATGGTGCCCCATGACATATGTGTGAGTCATAATCGTGTTCGAATGTCCATTAGTCTTAGTCGTTCATGATT
AATTGTGTCGCCACTGTACTAACGTAATGGTCCAAACAGCAATGGTGATGATGAGACTAAAATTGATTGGGTGTGTCAATACTGCACTACTGGATTGTATACCTGTCGATGATGATGAGGCCACCCGTCCGGCTGTACTACTTCCACCATTATCAGTTGTACCAGTTGAACTAATTGGTGTGGGTGCGATCGCATAAGGTGGTA
AGGTTGAACATGAATAGATTTGATTATTAGTGCATGTATTCACTGTGTGGGTCGAAGCTGATGACTGAGCATTACTTGAACCACACCCTTGATAATAGTATGTCTGAGTCGATACAAAGGGTGTACCCGATTGATATGGTACATCACAAAATGCATGAATCGATTGAGTACCATAGCCTGGGATGGGTTGGTGCGTATCCGCAC
ACGTTCCAGTACCCAACCAATACATCGAGAGTTGTTCATCCAGTGCTTGACTCGGATCACAAGTACCACTCTGATCAGAGGTAAATTGCGTGTAACCAATCGTCAAAGT

>GRPc64396 (Virulent pathotypes only)

GCTGGTACCACTTGTTTGTTTAGATTTGGCTGGTTCGACCTGTAAATGCCTCGGTACAATATAATTCATCCAAATTCGATCACGCACCGTCTGAAAAGTCCAAGTCAATCCGATTAATAATGCAACCACCACACCAGTAAAGAGGCCGAAATGTGTGGGTGTGGTGAAGTGTGCAAAGAATGGATAAGGCCACACTTGAAAGGC
AGAGATATAACGCTTATGATAACTGGCATAAAGGAAGAGAGAAGGAATGGAGATGCAGAGAAGACGCTCCAGCTGACGATTCACTTTGACATGGGGAAATGAGTGTAATTCAATCAGCATTAAAATGATTGGTAGGGTATGCTTCATCTGTTCCTCGAAAGGTGTGGTCGGGTCAGTCGGAGTTTTCACTCCTGATGGATGA
GCAAAGAATAAACCCCAGAAGATCAAAGTCATACCACTGCATAAAACAAAATTAATTGAAAATATATAACTCAACGCTTGACGTAAGCGATAAACATTCGAGACCGACCCCCCTGCGCTCCAAGAGAGTAATATATTCAATAGCCATAAACCGTACAATAAGAGATTTCCAATTAAATTCCATGCTGTCAAAAAATACGGGACT
GCGTCTAGAGTATAATTCCCTAAGAATAATGCATATAGAATGAGTACATAAGCAGAGGCATGAAAAAGGAAACTCCACGTGGATAAATGTGGATCCTTAGCAGAACGTG

>GRPc65971 (Virulent pathotypes only)

AATTAATAACCTTCAACTCCTCCTCGTCATGTCTTCCCTCTTTCATCCTGTTTCATCCATCCTGGAGAAGTTACTAAAATCGGATGATGTCGATGTGCGGTATATGTGTCTGCGTGATTTAGCTGGTGAATTGAATACACTCATCCAACAACAACTCGCCACTCAACCACTCCATCCATCCACCTCAGCTGGCATCGCATCCGCCACT
TCAACCAATTCAGCTGGTACTGTCAGTTCAATCAGTGCACCCGTCACCAATCTACCACCTATTTCACACCTATCGCCCTTAGATGCTACGACTCAACAACGATTATGTAAGGGACTATTAGCACAGTTTATGTGCACACAAGGAGAGATACATGAAGCGGCGATTCGATGTATACCTGGCCTTCTCTATTCAGTTGGTGATGGTAG
TCTTCTCCCACTGATTGATTCCCTACTCCTTCATACATTCGATCCTAAGACCGCTTTACCACCTGTCCATCTCGATTCTGAGGAGATGGCAGATGAGCGCGAGAAGGAGGCACGGCTAGAGAGCTTAAGAATGAAAGCCATGAATACAATTAAGGTAATTTGGAAAGAATTGATCAATGAGGAGGCGAGAAGTAAAGTCGCAA
AAAAAATTATTCCACCCGTGATTAAAATTATTCTCCACACGACCGAAAAAGGGAAAACCAAGAGTGCTGAACCTCCCGTAG

>GRPc66247 (Virulent pathotypes only)

TTTTTGGTGATATTAAAAATTGCATTGAACTTTTTAAAAAAAGTTTATCTTTGTTTGAGACGGGACAGTTAGTAGTTAGACAACGCGAAATTCAAGCATCATGAAACTTTTGGTTTTTGCGATTGTTGGAGCCATTCTTTTTGATATCGTCGCCATGACTTCCGAGATGAAATTTGTCAATGAACTGGCGTTGTCCAAGGAGGAGA
AGGAGCGTCGGCTGGCGAAGGCTAAATTGCGCATCGTTGAAATCAACAAAAAGGCGACGACATGGAAGGTAAAGAAGAGGGAATGCCAAATGGGGCCAAAGAAATCTGATCCAGTCCGAACTTTCCTGCGAGTTCAGGCCGAGCTGACTCCATTGTCTCTGCTTGATGAGGCTGATCTTCGCTGTTACAGGATGGACAGCC
GAAGAGGGGCGAAGGACGACGGAAGTCCCCCAGGCAAGTCAATGGGAGTGGGCGCTTTAATTGACAAAAATGCGGGAATTTAAGGCGATCCGGAGGCGTACAACAAGGACAGTACTGAAAATTCGGAGGGAATGCCGGGGACATGCCGCCCAATGTAAGAAATAGGGCAAATGGATATTGTTCCACTTTAGCATTGTTTT
CAATGTAAAATTTTCGGTACGAATCCCGATCCGGTCTCACACATAAACTTTTCATTTAAATGATTTTAATCTTGACATGGAATTGCAG

>GRPc66666 (Virulent pathotypes only)

GGTCGATTGAGCGGTCTTATCCAACTTAAATCAAACCGACTGTTTTGTTCATTCATACACCAATCATTCCATCTTGTGCGTCAACATGTCGTCTGGCACCTCCACCCTCATCACCCTCCTCGTACTCACAGTGACTCTGGGTCTCAGTCTGACTCCAGTGCAAGGTTTCTCGTATACTTCATACACAGATTCGAGTTGCTCGCATGTG
TTGGATGGTGGTAATGGGTATCAGAATAATACGTGTAAAGCTGACTCGTGGTGGATAGGTGGAACTGAGATCACTTGTGTCAATTTAACTTCCTCGACTGCTGCTTGGCATGTCAAATATTTTGACGATTCACGCTCATGTAGTTCAGGTGTCACACATGAGACGGATGGTGTCACTGGGCAATGTGTCACACATGAGACGGAT
GGTGTCACTGGGCAATGTGTCTATTTTAAGTTTGCTATTGGTTCTTCTTACGTAATGGTTGACTGCTTCGCTGATAGTTACGGAGGCGCAACTGCACCTGGTTTAATTGCTGGTCCAATTGTATCTGTTCTATTACTATTGGCATCTACAATCTACTGCTGCTACGCACATCTGTGGTAAGAGTACCTCGATTTCGAAATCCTTGGCT
TTCTAGCCTGGTTCCTTTGACATTTTCATTTTTAATTTTTACAGATCAATTGTTGTCCGAGTACCATTT

>GRPc67665 (Virulent pathotypes only)

GAGCGTATGTAAGATGAAGGATAATTCGACCTCGCGTGGTAGATGGGAAATGTCACATTCCAACCATTCCGTCGATGAAGCGTGTTTGTCGTTCGACTTGTACGAAAGTTGACAGACTTGTTTGCGTATCCGTCGCAGCTGCTGCTGCTGCCTTCGTTGAGATGATAAATCGCTCCTTAAATTCTTCGGCGCACATGGAATGGG
TGCCCATTCAGCAAATAAGGTGAAGGTGGAGTGAGTGGAGGTGGGTGCCGGTACAGACCATCGAATCCAGTGGGTGTGTTCATCAGCATGGATCGAAGCCGCTGTCGTTGATGATGGTCGTGTGTGGTAGATCAATTCAGGTATCAGCAAGGAGTCCGAATTAAATTCAGCCGCATCCTCCACGGACCCATGAATTCCGGCC
GACATACACCACTCTGTGGTAATCGTGAGCCATATGAAGAGGCGGAGAATGAATGCGCCACTTCCTCCTTGTTGGTCCGTTCGCGATCGTCGCATGGTGATAATATCACAGCCTACTCGAATCCCGGAACTCTCGCGTGGCTCAAGAGGATCGGTTAGTTGGCAACACAGAGGATACTGGATGAGTTGCAACGCATGTGCGAG
AGGTAGAAGGATGGTGCAGTGCGAATGACGAGGCAGTGGAGAGTCGTTCTGTCGTTGGCTTGCTGTCT

>GRPc67758 (Virulent pathotypes only)

TTCCCCCACATCGATCGGATCTGGTGTCGTTGGTCACCGTTGATCGGCTACCCATCATGTAGTCCTGTTATGTCCACCCTGGTGTTGGTGGACACACCGATCATTGTATATGAAATCGACTGCACCACTGTGAACAAGAGTCACCCAAGACTGCCATCCATCGATCGATCCATACCATGTTACGATCCAAAGCCTCACCTCGCCTT
CAAGGCCATCCCTCCCTTCCGACCACAGTTCCGCCTTCCTCCACTACCGCCGCCTCTTCCTCCGCTCTGCCACCTCTGTCACCTTTTCATGCACGAGAGAGTGTCGCCATGTTAGCTGCTCGACGCTTGTCTGAGGCGGTAGCGAAGGCTGAGGCCGCGAAACCAGTATGGCCGAAGAGGGTCACCATACCGGTTATGGGATTGA
CATTGCTGGCGACAGTGACACTATTAGTCATCTTCTCAGTCGGAGCGGCACACTGTACGACGGAGAAGCAGATCTTAGATTTCCAGTCGAATGATGCCACAACAACGATGCGTGAGATTACGAGTGCCATCTGTGCCATATGGATTGATATCAGATATTGGTTAATTGCTATGTTAAGTGTGAATACGGTTACGCTCATTATATT
ATTAGGTTGTCTCGCGCATGAATGTTGTCAATTAAATGTGACCGTCACTAAATGTTCCGG

>GRPc68472 (Virulent pathotypes only)

GGCGATTTTTGGTTTATTTTATGGCATTTATGCAATGGCGGATTGGATGTATCCCATAGCACCCTCCTTGGCGGAGTTCAAGACTTGCTTGCCGAGTCCCTTGAGGAAACCTCCAGCGGACTTCAACAACTTCTTGAAGAAACGAGCATCAGATTGGACACGCTTGATGGACTTCTTGACGTCAGTCTTCTTGGCGGCCTTCTTGC
TTCCCTTCTTGGATCCCTTCTTGGAGCCCTTGCGGGAACCGCGTCTGGAGCCACGTCTGGATCCCTTCTTGGAGCCCTTCTTGCCCTTGAGTTCACGAGTGCATCTTCTCTTGACGAGTTGTTGCAAGTTCTTGACGGCTCTCTTTTCTTCACGAGCAGCGAGGGTTCTGAGGACCTTCTTGAGGGTGGCAACTTGCTTGTTGTCTC
CCTTCTTGTCGGTGGAGATCTTCTTAATCAAATCCTTGAATTGTCTGACACGGTCGGCAACTCTCAAGATTCTTGGGTTTCTTTCGAGGAGACCCTTGAAGTACTTGGCAGCCTTCTCGTTGGAGATGGTTACGTTCTTAGCAGCAACAACTGGTTTGACGAGTTTGGCTGGTTTGACGGCCTTTGTTTGGGCGAAGACAACAGC
GATGATGGCAATGAATGCGACTAGGGCGATGGCTCTCATTTCGTTTTTTT

>GRPc68766 (Virulent pathotypes only)

ATCATAACGAATGAACGAACGATATATTAACGCAATAGCCAACACGACCAAACGAGGGACGACGAACGACGAACGACTCGCACCACCCATCATTCACTCAGCGAGTCAGTGAAAGTGTCCAGTACCAGATTCAATCAGATTCCTTCATGCGTAAACGCTTGGTCTTGTTCACAGCCGCCTGTCTGACGGGGATCACAGGTGGA
CAATATGTACGTGTCAAACGATTGGATGAGTCAGAAGGTGTCGGAAATGGTGTGAGTCCAATTCAGCGGTATCGTCCTTCGTATTGGAATTGGTTGAATGTGGGTGGTGATCAACTCTACTACTATCTCGGTTCATACGCACGAAAATTCATCGACCCAGAGACATCACATACAGCATCCATACGAATGAGTGGTGCAAATCGA
ACCATCAGAGGATTACTCGGTCTGATACCACCGATCTATGACTATCCTCGTTTACATTCCGAACCATTTCAAACATCCACCACTAACACCAACGACAATCATACCAGCAACAACTTGAATACGAATTCCCATTCCCATGACCAATTGAAAGTGTTACATTTTCGACATCCAATTGGTTTAGCTGCTGGTTGTGATAAGAATGGTGAG
GGTATAGATGGACTATTAGAGATGGGTTTTGGTTCGGTTGAGCTTGGTAG

>GRPc68839 (Virulent pathotypes only)

GACGAATCGACCTCGGAACATAGGTGAAGGAATGACCATTTGAAAGACCTGAGTTGCATCGTAGTTGTTTAAGAGACTCTGCGCGAAGGCGGAGGATGAACCCAGACCAAGTGCTCGACTCCATCGTCGTTGTTGTCTCGTCATCATACCACCAGCGAATCTCATCGGACGTCGGAAGGACTGCTGAGATTGACCCACTCGAC
TACGGGAGCTGGAACTAACACGAGCACTCACCCCATTGATTGGTCGCGAAGATAGATTTAGGAAGGAGATGATACATACACCACCACTACAGCTAAAATCTTGTTTATAATCAATGGCTGGACAGGTGGCAATGGATTCTTGTTCAACAGCCGACGCGGTGAGTGCAAGTGACGAGATGACCAAAAGGAAGCAGATGGTACA
TGGGGGTGACATGATCGTTGGAGAAGAGGGAGGATGAACAGTGGTGAATTGAGGTAAAATTGCGATGTTGTGACGAATGAATCCGTTAGGTCGTATGACGAGCAGCGGATGATCGATGAGCTAGGTGAATTTCACATGACTGACTCATGCTTCGATCCGATCCGATGGCACAGTGTCAACTTCCCCTGAATCAAGCTCCTGG
CTCACGGTCATCGTTTGTCACAGTTGATCAATTCCCTCCCTCCAATTCGTCTCCGT

>GRPc69973 (Virulent pathotypes only)

TGCTTGACTAGTTCGACGCTCGTGCTACTGAGCCAGTCCAGAACTGTTACTGAGAACGAAGAGACGAGACAATTCCACTCTCACCCCCACCACCACCCTCCAGTCCTCATTGATCACACAGCGACCATCCCAGACACGTTCAGGATGAGTACGATTTATCGTTTCATATTCATATGCCATTTGGTTGGTATTGGTTGGTTGATATC
GCTAGTCTTGGCGACTGATCCTGTTCAAATTCAACCACTACCACCATGGCCGACTATTAAACCAACTGGTGTGACCAGTACCACACCGATCACTCAGCATGTGACCGTCGCGAGTGGTAATTCGGCTGTGTTCAACTTTTCACGACCTGCTCCAGTGAGAAATAGATTTCAATATCAATTTAATTATAATATTACGGCTGGTTCCA
CCACTTACTTAACTGCACTCCATTCATGCTCGTCCTCTTGCTCGATTATTATACCAGGAAATCCATATGGAAGCGGACCCTACCTCTATGATGTTGAACAGAGCATTACTGGTCAGACGGTTTCAGGTGTGTTTAATCCAAGATTTCCGTTCGCTCAGAGTTTTAGTCAAAATGTTGGTTACGGTGATTGTTTATGTAATTCTACTAA
TCTTGCCAATGAAGTCTGGTATTTTCGTTCG

>GRPc70062 (Virulent pathotypes only)

AGCATGTGGCTGATGGGCTTCCTCTACTTCGTGGCCCGCATTGGCTTCTTCTACTGGGCTGCTCAGCAGAAGGGTGCTCAGATCATCTACGAGAAGGCTCTGCTGCCCGCCTTCAACAAGATCGAGGAGGCCGTCGAGGGCAAGAAGGCCTAAGCTGCCTCGCCGCCCCGGCACTGCCTCAGACACACATACACTCTGTTTCTC
GCACCGATGCACGCCCCCCCGTCCCCATGTTCAAAGCAACGGTGGGCCCCTCCCCTCTTTCGGGGAATGCCGGCCCTTCCCTCCCACGCGGACTGCAAATCCTCTTCATGTCGTCGTCGCATCAGTATTCATGGCGAACCCGTCCAACTGGAGTTCCTTTCTAGTGCCGTCTTCCTCAATTGACAAGTGTCTGTTTTGTTTTTTCCCC
ATTTGCAAGTGCCACTCCCTCTCAAATCCGTATCCCATCAAACCCTTTTTGCCTTTGCTCTTCTGTTCTACGCCCGTTGATTTAGCTGCCAGCGAGCATGGCTTGGGTGCTGTGTTTTGATCCTGTCAAACGCGTTGCACAGCGCTCTCGCTCATGGCTCTAGTGGCTAAGCCTTTGTGCGTCTTGTTTCGTGCCTGTCCTGCTGTCCT
CGGCGAGTCCACACCATCAAACTTTGCCA

>GRPc71146 (Virulent pathotypes only)

GGTATATTTATTCGTCATGGCATGAAGTATCGAGATGAGTATGGTTGATGAGAGTTGTGTATGCAATAGGTTGAGTAGATGATTTAATAGTTCGAAAGTATCATCACTAAATGAGTGTGGAATCCATGCTGTGAAGGATTCCAACATCGTGATTAAAGCAGCCGAAAATTCTTGATCAGCTAAGACCTCAACTTCAGACATCAT
ACCTAAGACAGCCACCAGCAACTGTAATAACTCTTCTTTAGGTGCACTATGTTGCATATGATTCACTGTGATCTGGATTAATGGCAACACAAATCGATGAAAGCTGGCGGCATTCGATAAAGTCGATGACGCACGCAGTTGAAAGGTTGACTCTGTTGAGAGTGGAAGAATAGGTAAGTTGAATGCTGGTGGTGGATGAGAG
ATATTGGTTAAAGTGGTGCCATCATCATTCGAGATAGGAGCGATGGTATCCGATGACTTGATTGACTGTTGAATTGGTACAGCTCGTGTCAGAGGTAGGAGTAATTGCAGAACAGTCCAATATTGGCGCATGAGTAGAATCGAGCGTAGAGTGGCTCGGCGACATTCTTCAGCTAACAGTCCACTAATACAATTCGTGAGTGT
TGCGAATTCATCCGCAATCTTTATATCCT

>GRPc71403 (Virulent pathotypes only)

TCTTAGGTGACTGTTGCTGGTATTTCATTCATTCTCATTTGATTTGTCACACTCACTCTCCAGTCGACTGATCCACCCATTCGATTGAACCACCCCACTCGACTCTCACCACTCCCTCCCTTGATCCACCACCACCACACAACTCGCCCCCATGGGTAATACCTTAACGGCACCAGCGTTGGTCTCAACCCAACAAGATCTGCCAACAC
TATTATTAGATATACCAGCATATCTACACCATTCAGTCTTAGGAGGTGGTAAAATTTTCAAGACCCTTCAATGTCTCCATGAGGATGGATTTGTAGTGATTAAGCTATATGTACGTCGTTCGAGTGACTTCGATTTAAAGCGATATGAAGCGAAATTAGATGAGTTGTGTTCAAAGCTCCCACTCGCCACCTGTCCATCCATACTTA
ATCTAGTCAAGATTGAAATCACCGATCGAGCTGTATATTTAGGACGACAATATTTCGGCTCCAATTTAGTTGACCGTTTCCACGTCCATCCCAGTTTAACCAGTATTGAAAAGCAATGGTTGACCTATCAAATGCTACATGCCGTACATCAGCTACACTCACACGGGATTCGTCATGGTGATATTAAGACAGAAAATTTCCTCTTG
ACAAGTTGGAATT

>GRPc71590 (Virulent pathotypes only)

GCCTGATCCAACCATACCCATTACCAAGCCTCGAAGCGAATTAAACAGTGGAGAGTCTAGTGCGCAATCTGCGACTAATTCTTCAGTCACTTTCGGAGATGAGGCGGTGGATCAGATCAACCTCTCACTCGGCTCTGACACCAATTTTCGTCCTCCTCCTCACATGCTGCCTATCAGCCATCCCCTCCATCGATGGTACTGCATTCC
CCTCTCCACGTGGTGATCCAGATGCGCATCATGGCCATCGACCAACCGGTCATAAACATGCATTATTAGAATTCGTTGAAACAGCGGAACCAATCACTCCACCTAGCCGTGATTGGATCAGTTCCGATTCAAAGTCATGGTTAGCACATGGTGATTTAGCTGCCGCGTTCGCTCTGACGCCATTTCGCACACTTGAGGTACCAAT
TTATGTCAATATACTCGCCGTTGGATTAGATGGCTTCGGTAATCTTGGAGTCGATATCTTTGAAGAGGAGTTGGAGGAGTGGATGCAACATCTCGATCATATCGTCTCCCATCAAGTCCTGAAATCGAATGAGGCTGATCAGAATGAGGAGGAGCGCGATGTCGATCAATTGGAACATACTTCACGATTGACCAAACATGGCA
AGAGTCGTGCATTTAAG

>GRPc71656 (Virulent pathotypes only)

GGATAAATGACGATTAAGGACCAAGACACTGAGTCACCCATTCGTTTCTCGTCTCTTCATCCTGACCTTCACACGTATAGTATCATCCAAATCGTTCGAAACAGATCCATTCATTCACGCGTGCGAATCCATCTCGTCGTCTGCTGACATGATGTCAGGTGCGACCGCGTCATCCCTCACCACCCCACTCAAACAATCAGCCAAAC
GAAGCATAACACTCATATGCCAGATATTGAGACGAACACTCAGATTGAACGAACAGGTGATACAACCAGAATATATTCGACGATGCAAACAGAATCAACTCACCTCAACCAATGACACGAATGAGGATGACCGTACAGGTGAACGCATTCAGGAAGGTCTATGGCATGCCATACATCATACCTTTCTATTGGAATTACATCAAC
CTGTACAATCATCTACCCATTCTACTCCCAGTGGCTCGTCCTTCTCGCTGACTCAATTGGAGGAATACTGGTTGCAACGCGGTAGTGAGGAGGAATTAGTTACATTCCTAGTCGAACAGTTACACCAATGGAATTATCATGGATTGTGTCAAAAAGACGCACATGCTTCTCAGCATACCGTACCACCGCCTGTGATCATCACCTGT
CGTCAGCAGCTCCTGG

>GRPc71704 (Virulent pathotypes only)

AGCGAGGCATTCGCGCTAACGATGATGCGATAGAAGATCCAGGTCGGATCGGTCTGGCTGATTGAGGGAAGGTCACACGCGAATAATAAGACATGTGAACCTGAGGGAATGCGGACTGTCGCGTTGGACGTCATGGTCTCTCTGGGGAAGAGGCTGGAGAACGAGACAAGGTCGACGCGATCATGTGGACAATTGACTCG
GGGCCAACTGTCGGTATTGGGTACACCAGGCACCACACCATCGTAGCCGGTCCACCCGGTATATGGTGCTGATTGTGAGAATACTTGTACCTTCTCGCTGCAACCCAATGCAACAAAAGTGCTGGGTTGACTGGTGGCTGTGGTGATGGCCGCTTTCAGTTGGTCACCATCCGCGAATGTACTGATATCTGGATTCGATGGCGT
GATGAGATCACGCGCGGCATTGGTACCACATGTGGGTATGGCATCGGAGGTGGTGATGGTGAATAGACATATGGCCACGACCAGACAGGCAACAACTCGACCCACGTGATTGGTTGACATGACTGACGGCTGCACGGTCAATGTGTGCAGTCTGATCGTCGGGGTCTTGTTAACGAGTCAACCGCGAGGCCAAGACTAATC
TGAATGAAAATAGAAGTTGACACTTG

>GRPc72085 (Virulent pathotypes only)

ATTTGAAAAGCGCCAAAGATGCGTTGTCTGGGCTGTTCTTGTCTCTCTTGGATATCGGCCCTTGTGTTGATCTTGGCCTTGCTGGCCACATGGCATGGTGTGTTTGACACGATAACATTCAGCCAGGGCAACTCACCAACTGTCAAGTTTGCCTTTCGAGATGCACAGGGAAAGTACAGCGAACTCGCTGGAAATTTGTTTTGGG
AGATCGGCGAAGCTGTTGCCTTTTGTGGTGATGACGAGGCCTTCATTGGTCTGATGGGTGTCTACTACGATCATCCGGATGAATGCTATCCAAACTCTTCCTGTCGCATGCGAATTGGAGCTGCCATCAACTCCCAAGCTTCTTCCGAATTGGTCAAGCAACTGACTGACAAGGGTCTCACTGTCCTTGACCTCAATCCCACTCGA
AGCTACTTTACCTCGATTCCTTATCGCACTTGTTTATCGATTTTTGTTGGCATCTTCAAGGTGTTCCCTGCGTTTGAGACTTACGCGAAGCTGCACTCAATTGACGGCGCCGTTGTTGCCGAGATTTATCACGAAGACACCATGGAGTACTTCTTCTTGGACAAAAACGTTCTTGATGTCAGCTATGCTCTTGCTTCCAGTTCCTCAT
AGTTCTTG

>GRPc72513 (Virulent pathotypes only)

AAACGGAACGGCTCTTTTGCACATCTGTGCAATCAATCTTGTAGCTGTCCTTCCGCGCGTTTCTAATCCAGACCACAATCAACGTCGGTCGGATCAGCCATCTATGGACACGTTTGGGACAGTCAAGTCTGAGTCAACCGTTGTCCTACTAATGCGACCTTCACATTTCTATGGGCTCATTCCGACTCCTCCTCATGGTTGGTTATT
ACTATTGGTAGCGATCTTAACTTTCACCCATGTGATACCAACCGATGAGTGGGTGACGAAGGTGAACGCGACCACAGATATGGGTATCCTTGATCCGATGAATATATCGTCAACGACTTTCGGTCATTTAGTTGATTACCCAGGCTATTTACAGTATGAACAAAACGCGAATGTCTGTCTCGTTCAAACTCAACTCACACCCATG
AAGGATGGAGTCGTCTATGCGCTGAAAATATCACCGCACACATTTTATAAAATCCCAACCGATATTGAAATCATTCAGGTACGAGGTTGGTCGAATAATGAGTCGGCATCAACGATGCCTGATTTAACTATTATTAATTCCGTACCGATTACAATTGCTTCGCCGGATAGTGTCAACAATTTTATCTATTATTTTAATTCGAGTCAA
TTA

>GRPc74775 (Virulent pathotypes only)

TTCCAACCATTCTGAAATACGAGACTATCTGAATGAGTATTCGGAGCATTCACGATATTACCACTATTCGAAATTGCATTGCCCTTCCACTTAAAATTAATCTGGACACCATTGAGGGTCACCTCCACAGTTGACATCGGATCCTCAGTAATTGGATACGCGACCGTACTAAATAGTGCATCGGTGAATTGACATGAGCATTGATA
ATGGAACCAATTTGAAGCGAATGTTGGTACGAGTGTAAAGCCATTACTAAAGGTTAGGCTCTTTAATAAGCACATCATATTAGTGTCAACAGACCCATCGGGTGGCGTATTGATCTGGACCACCTGAGTGGCTTGTGCAAAGACAGCTAGAGTAATGAATAGCACGCCTAATAAGGTGACCTTCATCTTCTCGTACGAATGGAC
GACGGTAAGTTCGGCTGTAGGTGACTGTGAGTGGGACTGTGAACTGGTGTTGGGTAGTTGGGCGAGAATTGGTAGCTTGAGTGAATTCGAATGACTTGTTCGTTAGGTATTGATAATGTTGATACTCGTTTGAGTTGTGATTCGAACAGCATTCCTCTGCCAGCGCCACACAGAGCAGGTCTGCACTCCG

>GRPc74837 (Virulent pathotypes only)

ACCATAACAATCAATCACTCCATTTTCAACTCCTTTCTCGTGAAGAATTCCCTTCATTCAACCTTTACACGACCAGTTGTCATCCCTCGATCATTTCCGTTCACGTCCTCCGATCCTATCTTTGCTTACTAATTCAGCGCTTCGAGAATGGCAGCCAACAAGAAGACACCACTTCTGATTGTTGCACTTCTGTTCATGTCATTTGCACTC
TATCTCATACTATCCACACCAGCCGAGGAAGAGGAAGTAATTACGACCACGACAGTCACTTCGACCGGTGAAACGATCACCACCACTTCGACTATCCCATACGTTCATCCTTGTCCTTATGCGGATATACTCGGTCTGGATGAACATTCTGAGAATCCTCATCTAGCTGGTGGACCACTCCCACCTATGCCACCGACACAGAGGA
AGCTACATGCGGAAGAAGAAGACCTTCGATCCGACGCAGGTCGTACATGCTGAGAAGGAGAAAGTTGATGAGGTCGATCATAACCCACATTATATTGGCCGATCAGATTTGACTCATCGCCACCGATTCCACCAGCACGATTGAAAGAAGAATTGTAATCTGGAAGTTCATGGATCAAAGTCGA

>GRPc74957 (Virulent pathotypes only)

GCCAGACCACGTGTCCACCGCTACTCCACTTCTGCAATTCACACCTAACCGATTGCTTCTGCGGACAACCAACGTATTCACTACCTTTCTCACACCAGTTGAGTGGGTGATCGCGATTCGTTCACAGTTCGTCTTCTCCTTCATCACCTCGACTCACCCATAAAATCACATCTCGCCGTCCATTCCAATTGAATCGACCATGCGACTC
ATCTATCTCATCACCATTCTGTTGATGATAGCGAGACAATTCACACATGCAACGAACACGTTTCAAAATCCGCTGGGTGGCACGAATGGTACTTCCTCACCTGTGTGTTATTCCCATTGGCAATCAATCGGACCATTCCATGTCGCACAGACTTATCCAGGTATTAACGATCCAACTGATCCTGAGGATATCGGTTCAGCGAGTGA
TATCTGGTTAGCCGGCTCTACAGCCTATTTAGCATCTGTCAATGGTGGGGTATGGCGAACGAATATTAGTGGTACGGAAGCATATGGTGGACCACACTGGGAGGCACTAACAGATCATATGCCATGTACCAGTATTAGTGCTATCAGTGGTTCGAGGGATGACACAGATATTATTGTAGCCGGG

>GRPc75160 (Virulent pathotypes only)

TTCTCATTCACTCATTCTGAGACGAGTGCATTGAGAGTACATCCCACGATCATAGACAACTATTGAACCAAGTCCTTCGATTGTTCAAGTTAGTAGCCATGCCGCCACCACCGCCTCCATCGACTGCTGCAACTGACGATGTGGCGATGTCCTCATCCACTCTACCAACTTGTTCGCTTGAAAATAACAGTGACTGTTGCTCGGCTA
CAAGCATGGTCATTGTTGCAACCAGCTGTGGTGAGAAGAAAATCACGACCTGCTGCAAGTCACCTCCACAACCATCCATCGCAGTGGCTTCCGCAGTCGCAGTCGCTGAGGCAAATCTGGACAACTCCATCAACGCTTGCAACAATCCTGCTGTGACCAAGCCAACTATCGCAACTTGTTGCATCGATGATGCAGCCAAAGGGC
AACCATGCTCCGGGATTGCACCTTGTTGTCAAGCGAATCAGGTGGTGAAGATTATCATACCAAGTTCCTTCAATGATGACAATCAGCCATCAACAGGGCAAACAACCTCAAATGAATGGAGTTCCATTTTCGGTATTGATCTCAAATTACCAGTCAACAAATCAAACACGACCAATCAAAAGCAG

>GRPc75211 (Virulent pathotypes only)

AGAATCCAAATTAATTGCTAATGAATGAACGCAAAATGGGAGTCCGAGACACGAGCTAGTGCACAAGGAGTGAAAGGAGTGGAGCGATCGAGATATGTACAGAGAACTTGATGGACCGTTGATTTCGGATCCGCTAAAGACGCGTATGTGTGAATTAATCGATCCAGTTGCTGTTTGTTGTGTTGTTCTTATGGCGTGGGTG
CCACGAGTAATGATGTACAAGCGACCATACATGTCCCCTGAGCTGCACTACAGGCAGCAGCAGCCCCTGTTAACCAAGCCCACCAACCGACTGGACCACTCGTACCAGCAACAATACCAGACCCACTTAAACAGGTGACATAGCCAGCATTACAGGCGGTATAACAAAGACCTTGAGCAATTGGACCACCGATACAACTAGTC
GGTACCAAGGAGAGGAGACTGAACCAGATGAGCAAGTATTGAACCAGAAGGAGACGACTCATGAGTGGGTGACTGGATACAATGGATAGAGAGCAGAATGAATAAACGAGAGAAGTCAACACTTGAATTGAACTTGAAGAAGGAACTTGTGCCAGTCAATGGATGGATTAGGTGAGGATGAAGATGTAAG

>GRPc75350 (Virulent pathotypes only)

TATGACTATGTCAATGTTTGGGGGGTGCGGGGTGAAGAACAATAGAGTCCGCAGAGGGGAGAGTTTGTAGGTGGCCACTGCCTCCCAGCAGCAAATCCAGCCAAGCGGCAAACGCTGTGGACGCGCTACTATCTGCTGCATGCTGCCAGACTTCGTATTTGCGGTGTTTCGCTTTGAACGCGATCACGGACACAGTCTTGGA
ATTGAGATTGAGAGAGAAATTAAACTGCTGAGCGAACAGCCGGACATCGACACTTGACGATGAGACACGAGCAAACCAAAGAGTCTTTTGCCTTTGAAATTTGAGCGAGAATGTCTCCAACATCTTCTCAACAGCGACTTTGTTGCCCGGGGTTTGGAGCATCAAAATGGTGAATCCATTGTTGTTGCCCAGTAAACGCGCCTC
CTTGCCGGGCACCAGATCCAGCACGAGCGAAGGGGACTCGGGCTCGCTGCTGCCACTGCCGCTGGTGGTGCTCTCGCTACTCCTATGAGACTCTTGGGACCTGGAAGAGGAAGCCGAGGATTTGCCGGATTTGGAGGAAGAACCTTGCTGTTGCTGCTTCTTGGTGGAAGAGGAAGAGGAAGAGGAAG

>GRPc75372 (Virulent pathotypes only)

ATCAGCAGGAGTCGGGTGTTTAATACCATTCATCTTATGGTATCATGCGGCATCCTCCTCATCCTCCGATGAAGTCAAGGAAACAGAGTTGCAACGAATTGGCCTGACACCTATCCCGGCCCACTATGTTCATCACGTCGAACAGTCCGTAGCAGCATGTCGACGCCAGTCACAACCAATCAGCAGGACGAGTACCACA GATGC
GACGATCACACAGGAAATGCGGGCCACGGCATTGAATTCGAATGCTCGATCCACTAAAGTCTATGCATCCATCGAAACATGTCCACCATTCATCGATAGTGATGATTCAGAATCGAATAGTTCATTCAACTCTTCGATCCATCTCTTGGCCATTATCCATCACCTGCCAATCTACGTGATTTATGGCGTCTCGATGGAGTGATAC
CGGATGAACAGCTCACGATTAGAGCTGAACAACCATTCGTTTGGATCGAACCGAAATTACTTCACAAGACGACATCGACAATGACCACACCAGGTAAGGTCGATCCGATAAATGCCCACCGTCAACCGAATATGATTGTCCGTGTGGGTGAACGATTCACCGTTCACCTCCAAACCAATTTCCA

>GRPc75458 (Virulent pathotypes only)

AATTGGAGCGCCTCCTCTCCTCTCAGGTCGCACTCATCCTCATGACTCACATACTACCAATGACGACGTCGTGTGGTCGCCTGCCCTCCTCCTCGCCCTATGCCTACTGACGATCACACCTCGGCAGCAGAGTCAGTCCACTCTCAATCATGTTCACCTCTACCCTCCACCTCCTTCTGTGCAAATGTGCTGCCATATGCGGCCTCAAA
TGTGACCGTACTCGGTCCAAATGGTTGTACTGATCCTCAAGAAGCTGCATTGAAACAACTACCAAATGTGATCGCAGTCCTGAATACGATCCCCGATTTACCCGCGACCTGTGCTGCTGCCGTCTCCGAATTCAATTGTGTGAAAGTCTTTCCTGATTGTACGGGTGCACCACCTTGTGCGAGTATGTGTGATACGGTCAACACA
TATTGTGGAGCTCAGATCGTTGGGCTACTCACCGAATTTGGTATGGACTTCAATTGTTCACATTATACCACAACGAGTTGTTACGTCCCCCTGCACTACCGACTGTGGTGCCGCCAGCACCTCAATGTGCCAACTATACCGCCTCAGGCGGCACCTTTTGTTCTAGTGTGATTGGTAGT

>GRPc75514 (Virulent pathotypes only)

TGCACATTGACCACCACCCTGTACCACTACCACCACTAAAACACGAGGACCAACGAGTGTCAAAACAAATCACCTGCCAAATAACTTGAACCGTCGAACCACTAATGAAACCATTCGTAGTAGCAGAAATTAAACCACTATCCCATTCAAGATAAGCACGAGGAATACAGGCTTCAGGTCCACCGACTAAACTCGAACTATTGGT
ACATGTGTAAACTTGTCCGCTAATTTCTTCATTTCTACTCGTACAGCCGAGAGTACATGAGCTTCCGGTTGGTAGTGTGGATGAACTGCAGGTACCCACAGTCGCTCCACCTGGTAAATTCATTTGAAGCGTCGTTAGGGAACATGGTAGTCCTTGACCATGACTAAGGTCAATGAGAATCAACACCAATATCAATAACCATTGG
ACGCATATCTTCATTCTGGAGTGGAATGGGTGAGGAGACATGGTGGGAGAAATTGGCACGTGCTGGGTACTCTCTTCCACCTGATGGACTAACAACACACAGGAAGAGACACGGGTGGATTAAAAGGACGGACATCGCGACAGTCAGCGAGTGAGATGGACGAGTCACACAGGCTGAAGTTGTT

>GRPc76422 (Virulent pathotypes only)

TTTGTATACTCGATTAATCTAAGGCGACAGGATGAAGCTGGCGGTGGTCTTCGTCTTGTGTGTCGCGTTCGCGGCCGCAGCGGCGCGCGCGGAGGACGGCGGCAGCGCGGAGAAGCCGGCGGCCAAAGCCGGCGGGTGAGAAGGCGCTGGCGGATTTGGAGAAAACGCCGACGCAACGC
GCGGTGGATGCGTTCAACAAATTGCAGGAGGAACGTAAGAAGTTGAACCCCATCAAAGGCCGAAACAGGAGAAGAAGGAGCTGCCGCATGCGTCGGAGAAACTCAAAGTGCCGATCCCGATCGCGCATCCACGACCGTACCTGCTGAAGCCCATCCACAGCGTGTTTAAGTGCGACACGCTCTTGGACAAGAGCGCAG
ATCGCGCGCTGCACCGAGTGCTTCCGCAAAGGGCTGCCGGCGCCGTGGGAGGTGATCGGGTATCAGGCCAACAACGCGCAGTGCCGCGTGGGTCACACGATCATCCCCGTGGCGCTGTTGGACGCCGGTGCGGCCGCCACCGCCGCGAGCGCGGAGCAGATTATGGCCGAAGACGCCGCCGCGA

>GRPc76681 (Virulent pathotypes only)

TTTTTATTGTCAGCGCCGCTCTTGGCAAGTCTGTTCTCATCCATCGATCGTGTGCGGGTGGAGGTTGTGTTGGGCCGTCTGTGTTCGATCGCCACGTGTTCACAATCACTCGATCAACAGTGCGCTCATCCTGACTCTGACTCCATTCATCCAGACAGTCAGAGTCAGACGGTCGTCGACTCTTGGCGTGCCTGTGTTCACCCAGC
CTGGACCCATCATCACTCGAGAATTGAACACCATGCTGACCAGGAATAGAGCAAACATGTCAGCATCAGTCGCGTCATTACCAGCAGCAAACATCAACTCGTGGCAAACGAAAGAGGCAGATACAGCTGTAGAGGCGGATTCAACGGTGACGAAAAATCAAAAAAGTCGCATATGAATTCTCCCAATTATGAAGAACAAG
AGGATGCTGAGATGAAGGACACGGAATCATCAACAACGAGTACAACAACTACTGCCTCAGCCACTCTAGCTGCTATACCACCACGGTGACGTATTCCTCTGTTGAATGTCAAGCAGCCATCAATCGGTTGTTACGTCAACCATTGCCAAGTGTACTCGCGGATATGGTGACCGA

>GRPc76891 (Virulent pathotypes only)

CGTCCGGCTAGCGGAGTCCACTGCTGCTGGGCAAAAGTATCCCGTACCAGGACCGGAAAGTGCGACTTTACTTGGATTGTTGCATTTCACTTTGGTCGCAACTGATTCTTGATTGATACCATCAG GACATAACATAATCTATCTTGATCCCCAAGGATTGCGCCATCCAGAATAACCAAATGCTCACATCTTTTTGCAGAAATCA
GCGACAAAAGCTGGGACACCGGGGGCAGTAGCAAGAGCAGCATACCAATTTCCAATAGGAACAGCGACATTATTGGGTGCACCGGTCAGATGAGCATGGCTGAGCGCTTCGGTACGGGACATTAGTAGAGCTTAAGTGGCTTAAGAGTTGGCATCGTTTCACACGAGTATCACGTTGAACAGGTGCAACTAATTCATCATG
GACAGCCATCTTTAATCCCTCAAGGGGATCAGGTCTGAGCGAACCAAGTAGGACGACTCTTACCAACACTCATCAAAGTCATCACGTATAGCATGGAAGAGTATTGTACCAGCTTCACATTCCAATCCCTCACACCGGGAGTGATAGCCTGACCAGTGACCTGATCAACAGTG

>GRPc77274 (Virulent pathotypes only)

CGTGCGATGATGTTACAATGCCTCGAACAATTTTTAGCCTTATGTCCTTTCGCTGCCACGGGTTCAGCCACAACAGGCTCATCCGACACGAAAACCTCCTCCACTTCATCATCCATGTCCTCACTTCTTCATTCATCTGCTCTCAATTCATTGATTTATCCTGATCCGACTCTCTTAGCTAGTATGGATACTTCGACCTCATCCTCCGAT
TCAGCTTCAAGTATTCTAGCCGGCACCAAATCGATCCCACTCAAATACCAGCGGAAGAATGCGCAGCAGATGAAATTCTATTCACCACCTGGTACACAACAACTAGCACAGCAAATTCCCAACTCAAATCTGCCTCCCCTCTCCAATCCTTTCTGAATCAACTCCTCCCCTTATCGATCGAAATATTGGCTGGAGAGTTTCGGTCTAGCC
TCTTCACAGACCTCGACCAATTCATCCGGTGGTGCGCCCTCCACCTCCTCGACTACACCATCGAAAGCCACCGCCTCATCGGCTGATCTACTCGCGAATTTAAGTTCTATCTTAAAGTTGATGCACTTAAGCTTAGATATCTTACGTGGACAACTATTGG

>GRPc77524 (Virulent pathotypes only)

TTTTTGGTGATATTAAAAATTGCATTGAACTTTTTAAAAAAGTTTATCTTTGTTTGGAGCGGGACAGTTAGTAGTTAGACAACGCGAAATTCAAGCATCATGAAACTTTTGGTTTTTGCGATTGTTGGAGCCATTCTTTTTGATATCGTCGCCATGACTTCCGAGATGAAATTTGTCAATGAACTGGCGTTGTCCAAGGAGGAGA
AGGAGCGTCGGCTGGCGAAGGCTAAATTGCGCATCGTTGAAATCAACAAAAAGGCGACGACATGAAGGCCGAGCTGACTCCATTGTCTCTGCTTGATGAGGCTGATCTTCGCTGTTACAGGATGGACAGCCGAAGAGGGGCGAAGGACGACGGAAGTCCCCCAGGCGATCCGGAGGCGTACAACAAGGACAGTACTGAA
AATTCGGAGGGAATGCCGGGGACATGCCGCCCAATGTAAGAAATAGGGCAAATGGATATTGTTCCACTTTAGCATTGTTTTCAATGTAAAATTTTCGGTACGAATCCCGATCCGGTCTCACACATAAACTTTTCATTTAAATGATTTTAATCTTGACATGGAATTGCAG

>GRPc77841 (Virulent pathotypes only)

GGTGAGAGGTGCACCGGTACAGTTCCCATTCGACCAATACTGTACGTTAGTGACAAATGCATCGGGTGATAATGGGAATGTATTGGAATAGATAAGTTGTGACTGTGGATCGTCGGTGACGGCTGGTCCACCACCATAACAATTGATAATAGCACTAGTCCAGTAAGCTCCATGGATGTCATACAATGGCATACACTTATTGA
TCGAACCGGCTGTTAAATTACCTAACACAGCTGGTGTTGTACAAGTGGCATCGTGGTCATAATCGATACTCGTTTGAGAACCGTCTGGATAGCAGGTAATGAGCTTGGAGGATGACTCAGATGTGGCGAAACACGTACCGTTTGGATAGTTGGTGGCATCAACAATCGGAGCAGCATCACAGGCGGTAGGCGTACTGCCCGA
AATATCGACCTCGTGAACAGCATGAACCTGGACTACGCCAGTGAGGACAACGAAACAAACCACGAACGAAACCAACTGCACCGTGGTTGTCGCAATGATGGATAAACGCATGGTTGAGGCAGCGATGTTGATGTCTAACGATCTCGGCTAAAGAACACTCGAAGGAG

>GRPc79432 (Virulent pathotypes only)

GTTTTCCTTGCCTGTGTACTCGTCTCGTCACTCGGCTCGCCCTCGCCCAACGTAGTGAAGGAGTGGCGCCATGTGGCTGTTGATCCCATCGCCGCCAGCAAAAACTACACGAGCGTGGCCTATTTGTGGTACCAGGTTGCGGGCAACACCACGCGCCACCTGGAAGCTTTCTTCGAGCAATATCACGGCGGGCAATGTGATAACA
GCCTTCACCGAGTCATTCATCAGCGGGCCTCAGAACACTGGTGTTATCATGGGCCACCATCACTGGCGTCTCTGTAGGCTTCGTGCTCAGCTACTCTTCCAATCCGGCGACGACATGACATGGAACTTCTCCTTTACGCAGGGCGGTGGTGGTGGCAGCTTCAGTCGACCCAGCTCCTTTGTCATGCAGGATGTCAGCCTCTTCTT
CACCATCCCACCCGTCGCAGCATGCAACTCCTACATGGAGATCATGAACTCGTACATTCAGACACCTTGAAGACATGCCTCTTGTTGTGCATCCATCATAAGGGAAAAGTAGGACGCTCCCCCCTGCCGCTCATTTTTTCCAAGTCAG

>GRPc82346 (Virulent pathotypes only)

GAGTGTCATCTGGTTGGTGTCTACCTGCTGCTGTCGTTGTTGCCGCACTGAGTTCCATTATTGCCACGCAAGCGCAAGTTGGTCAATGGTCTCGATGGACGATCGGTGGTTCAGTTGGGCTGTTGGTTGTTAATGGATGTGGTGTTGTACCAACGAGGGTATGATGATCAATATCAATTTAACACCAGCGGCATCAACGATTCA
GCAAATGACCTGTTCGATAGTTGAATGGATCAATGGAGGACACATGGATTAGATGTGTCAAATGAATGAACTCACCGCAATGCCATGAATGAGTGAGGAAGGAAGGGAAAAAAAATGCCACCGCATAAAGAATCTACCAGAGAGACGAATCGGCTTTTGATGACAGTGAAGGTCGGTAGGAGTTAGACATCTGACGTGC
AAAAAGGAATGCATTCAGGTGAGCGAACGGACGCATCGAGAGTGAACTCGACGACTGGTACACTCACACGAACTACACATTTGATGAATGAACGAAGGGAACGGAAGGAAGCGGATGAGGAACGGAGAGA

>GRPc83430 (Virulent pathotypes only)

GTGTCGAAGCGAAGGGTTTCATCCTTTCAACGATTTGATGGTATGGGCGCTAGCACTCGTCCACTAGCCACAAGTGACCAATTATCGCTACTCATGGATTGTATCAAAGATTTGAGGGGTTTTCTACCTCTTATTAAAAAACATCGAGTAAGTGGTGGTCAATTTCTATCTTCGACTACCAAGGAGATTGATTCATTAGTTGAAA
AGGATCCGACACTTAAACCAATTCGCGGTAGATTAACTTGTTTCAAAGGGTTCATCAATTTTGCTGTACGTGTGACTTCTCGATGTGAACTAGCAACCAAAGAGAGAGAGGAGCAAGTAGCCGCATCCAACTCATGTGTTTCGCCACCATCTCAAAAGACCAAGGAAATTGTCACCAACTCCTCACCTGGTGACACATCATCCAC
ACGCACTTCCGCACACCACCTGACACCGCCAACAGATACCTCCTCCACGGCTGCATCACTATCGTTCACACCTTCCACCGCCCAAATCGCCCTTGTCCCAACCATCTCTAACAA

>GRPc84112 (Virulent pathotypes only)

TAGAGTCACATACAACCCATGGTAACAGATTTAGATAGAGTGAAGAAGTGGGGTCAGATGATGAACCATCATATTGTCGTGTATCAGTACATGCGATAGCGGATAGTACGGTTTGAGCGCATGGGAATAGGAAGAAGTATAATAATGCTAGGCATACACCATACATACGTGGAA AGTTGAGTAATATACCCGATTCATGACT
GTTGTTATCAACGTGTGTGATTGATTCATTTGGTATGTTAACAGAGGTAGATGATGAGTCGAGGGCGGATGAATCAAGTCCTAATTGAATCGCGTCCCGTAGTGAATACGAATTTGATGGAACAGAGGATGGTGAGGCAACTGGTCCAATTGATACAACTGCGCTAGCATGACCACGACGATATTGTTGAAATAGATATATCA
GAATACAGACCAGACTGACGAGAGCAGGTGCTAATAGACTCATCCACACTTTATGCTGAATCGACCATGCATGATTCAGACATTCGATTGAAATTAGTGATGAGAGTGAAAAGCCT

>GRPc85642 (Virulent pathotypes only)

GCAAAACACCGAATAATTGAACTCGACCTGATCATGACACCCGACCGCGCCCGCAGGACACTTCATGCAGTGTGATTATCTCCCTACCTACATCCTAGTATTCGCGCCTCTTTTCTCGTATTCATCCTCTGAACTCGCATCTCTCCATCCCATCCCAGTCGTGAGTGAGTCGAGTCGTCCGCTAGTCACCATGCGACCCTGCTTTCCT
CTCTTACTCTTCCTCATCCTAGCACTATCATGTGTATTCGTCCATGGTGGTGAACGCGGTGAGTCATCTTCACCAGTTAGTAGTGTCGGTCAATACTCCACCCATCACTTCGATCAGCCGCAACAACAATCAGAGAAGATGACCCTCACGGAACAACAGTCATATCAACTACTTCAACAGCAAGTGACGGGTCATGATTTAACACC
AACGAATCAATATCAATTAATCTCAGGTGGCGCTATACCATCACAATTTACCACCTCTGAAAAGACTCAACCACCCTTCCTCCCTCGTCCACCTG

>GRPc85860 (Virulent pathotypes only)

CTGCCACACTTGATAATACCTCGCTTGCATCCTCGGTCGTAATGACGGTCAATGTTGTTAAGATTTGTTTCGTAGTGTCACTGGTGGTGGTTTGAGTCGTACCAAAACCGGTTTCAATCGTAGCGGCTTCAGGCACGAGTATCACGTGGCTCGTATCCGCTAATACGGTTTGATCGAGTGTGAGACCGGCATCACCGGTGACAAC
TAGACTGGCACATAGTGTCATCAACAGTATCAACAACATCACCAATTGATGACTGTTCAGACGGCGTCTCATGTTGGGACCAGATCAGTTCGGATGTGGCGGTCGGATCGAAGAAGGGAGGAGACAAAAAGAAAAGGTTGAAAAACAGTTGGGACGTGGTGATGCTTGAATGAATGAAGAAGGTGGAGAGCAGTCGATTCTCT
GCTTGAGTTGGGAGTTGGTTGTCTCCGGAAGCTTGACTGTTGTGATTTTGTGGCTTGGTCCCTTTGTGTGTGTTTGAGGTTCGGTTCGGTTCGGTTCGTG

>GRPc86326 (Virulent pathotypes only)

ATGCATGTAATCCAACTGCACCCAATCCCTCCTATAATGCAATGGCACTAGCATCGGTCACTACAGTCGGTACATGGACCAGCCTAGTCACCGCGACGGATTGTCCAGGACCAGGTGCAGGTGTCGTCATCTACACGGAAGGGACCAATCCTGCATATGTCGTAGCATCATGTGGTGGCACTATCACGAATGCTGGTTCATTAA
TTGCGATCGATCCGAAAATACCTGGTACAATCATCAAAATGGCTGCAAATGCATGTACCAACATAGTATCCATGGCAGAAGATACATCCACCGGTGCAATCTATTTCGCTTGTGGTGGTGGAACGAATTCAGCTGCTGTTATGGTACTATCGTCGATTCAATCAGGTGCGAGTGCTTCACAAGTATTATCAACATCCGATTGTCC
AACTCCGAAACATCTCGTCTATAACAGTGTCGGCAAGTACGTATGTGTCGTATGTGGGTCGAATAAAGTCGTACAGGTATCTACACAGACCTACG

>GRPc87170 (Virulent pathotypes only)

GATTGATCCACGTCTCATGCCTCGTACGATCGGCCGATTGATGGTGATGCTGTTGCTGACACCATTGCCGGACTCGTTTCAATTTACCCATGTCTTTCAATTTACCCATGAGGCGATTCCCTTCATCATCGTCTACGGTAAGCTATTATCGATTCCTCGTGAGGATATTATTCGAATGAATCTGATCACACTGAATCTACTCTTGTGC
GATCAGCACCACTCCCTGACTCCCGCCAACTTTTCTCAGCTACTTCAAGCTTGTGTAGAACACAACATCCCAACACGCCTTTTTCTCTTTTTGAAGTCAAATCATCCATTTCTTTCGACCGACGCACAATTAGATCTGGTTCTGCCTTGTTTAGCATATTTTCGTTATGGTCCACCCGCACTCCGCTCGGTCGTTTTTACTTCATCAATG
ATCCAGCGATTGATGTCAATCAAGAATAATTCCGACGAAGGTCGTCGTGCATTGGGTGAGTCTAATCCATTACTGTTAAT

>GRPc87740 (Virulent pathotypes only)

CAACTCAACCGTTCCTCCTTCCACGCCAATCATGTCCTCCTCCAACTCATCCACTACCTTGAAGCCTGTTGTTGCTCCTCCCGACCCTCCTCGATGTGGTGCACCGAGATCCGTCCTTCTCTAAACCTTAAGGGATTACTTGACCCCAATGTCAAATCATCGTGGGATGAGCATGGTACACGATGTAATGAAATCTGGTTCAACCATC
GCTTGGATGAGGGACCTGCACTTCGCGACTATTGATAAACTCCACCATTTACTTCCTCTAATTTCGTTCGATCTCCAATCGGAAACCAGACTTCTCACTTCACACTTCTAGATCATCTCATCACCTCGTCGACTTCGATTCAATTCTTCTATTAAGTTCGAATGGCAGAGCGGTCGAAGAAACGCGCGGCGTCTCCACTCCACCCAG
AGACAGAGGTACCAGAATCATCTGCAAAGAAACACCGCGCCACCATGAAACCAAGATGTCAATTCGGACCAAGTTGTTA

>GRPc88129 (Virulent pathotypes only)

AGGATGTCTGTGCGATTATTTAATCGTGCTAATTGAAATTTGAATACAGCCCATACGGATGGACCATAGGCTTGAGTATGATAGGAAGTAATTGGTAGTGGATTGGCTGGGTCATATGGTGGACTTGAGACATTTGCAACCGTGCTCGGATAAGTGAATTGAAACATACAACCTAGTCGAATCACTGCCGGTGGTTGTGTCAG
TTGTAGTAAAGAGCGACTACCACCACCCGAATGATCGAATAGTTGATGGTGAGAGTATGGAATGGAGGTAGTAGCGAGAGAGCCCATGACCATTCCATCCAAGAAGGAGGAGGATTCCCACCTGCACGGATGGATGGATGATCATCCTATTGGTTCTGATCACAGTCGACGCAGGTTCTGATCTGTCGATGTTAGTGCATGCA
TGGAATGGGGATGGTTCACAATCGTACTAATCGTGTGTGCATGAGGGTGGTCGTGTGATTGGTGTGGTGTGATTGGGGTGGTGTGG

>GRPc92478 (Virulent pathotypes only)

GGATGGTGCTCGACCATGAAGGCAACTTTTGGATTGCTATGCCGGGCGGGTCTCAAATCTGTCGCTTTACTCCTTCCGGTGAGTTGTTACAATCAATTCGACTGCCTGTCTCAAAAGTGACCTCACTCTGCTTCGGTGGCGATCAATTCCAGGATCTATTTATCACCTCAGCTAGTAAGGGTATCAATCGTGCTAAGGAGCCATTA
GCTGGGAGTGTATTTGTAATTCAGAATATTGGTCCACGTGGTAAGGAGATGAACCCATTAAAAGGCGCTTCTCAATCCATATCCAACCAACGCGATCTGATTCAGCATCAGTTCTTGACATCAATTCCACCTACGCCACAACCGCGATTGATCTCCAAATTGTAAGAGAGAGGCACGGAGACGGTGCGTCAAAGCCCCCCGAC
CGAATGATCTTATGCCTATAGGCCTAGCCTATTAAATGTTATCCTCTTTGTG

>GRPc92619 (Virulent pathotypes only)

GCTGTGTTCAATTTGGATGTCAACCACCTGCGAACCCCACGATTCAAACGGACTCTTGAGAACCACAACCAACAACAATCTCACTTCACGCATCTTTTGCACGACACCAAATGACATCAATTAAGAAGGTTGTTCGTGAGATGTCATTAACCACGTCCTCCTCATCATCGATGATACGATTGGGTTTGATCAGTCTGCTGCTGGTC
TTGTTCCTCTCCTCCTCCTCCTCCTCCTTCATGACCCCAACCACTGCGATCACGACGTGCACCACTGTGAATGCAACACGTGCAGATGTAGTCAATATTGCCATCTTCTTCTGGAGTGCAACGCCACATTATAATCAGACGGTCCATTATCACAATCCAATCCTCGACAACTATGTCGCACCGTCCATCATCGAAAAACCGGTTCTCC
AATCCTCGACAACTATGTCGCACCGTCCATCATCGAAAAACCGGTTC

>GRPc93123 (Virulent pathotypes only)

TCTTGTTCTGCGTGCATTGGATACTGTAGACTGTTGTACTGAAGCCTTCATTGCGTCTTGTGGCGCCCAATCCAACCTCACCTCTCACCACCTCATCTAACCCGAGTCTAGTCAGTCATCATGTCTGCTACGCCAACCGGTACAACAACGCAGCCTGCACCAGGGACATCAACGGGATCGTCAACCATGACGACGAGCACAACTAC
AGCCAATTCATCAGACAGTAGCAGCACAAGCGGTATCTTCATCTCACGCACAGCCATGGAATTCTTTACGACTGCGGCACCTAAATTACACGCAGCTGTTCGTGAGCTCCAGGCCGAGAATAAACTCTTGCGTGAGGAGAATAGTAAATTACGTGAGGAGGTTGAGATCTTAGATGGTATCTCACACTTCCTGAGAGAGATGTC
CAAGACCGATTTAAATTATGCACAATTCATGGCCAAGAAGACCGCGG

>GRPc95279 (Virulent pathotypes only)

CCGGCAACTAGCATCGTCGAAATTCCGCTGAGAGTAAGCGCAAAGACGGCCGATGTGGCAGGGTCACCGGCACATGAAACGATGAGGTGCCCTGCGTACGGATTCCACACTAGACTATTCGGTGCGCATAATGAGGCATCGAGTAGCGTCTGTACCTCGCCATTGGTGTTGATGGCGACCACACCCGGCCATGCACATGCAG
CAAAGATGGTACCATCAGGATGGAAAGCAAGGGCTGCAGGTGCAGTGCACTCAGCGCCCGTGAAAGCTAAGGACTCAGCAGAGCTCAAACCAATGGCAATGGACCACAAGCTCCACAGACCGAGATACCAACCAATCACTAAGCGGAAAGGCATCGTAGGATTTGTGTTTGTAAGGTCGCTGCGACTGTACGTTCGAGTGC
AGCGTATGAGTTAAGACATAACCCTATCAGGAATGAATCGG

>GRPc95850 (Virulent pathotypes only)

GCATCGGAGTAGTCATCATGATCTGATCATATCGTACACCGGATTGACGACGGCCATGATGGTAGGCGGCGTTTGGTGAGGAGGTCTCACTCTCAGGAATGGTATGTGGTAGATGGAGTGCGGAGGCATGCAGATACGAGGACGAAGATGGTGAGATATTCGACGTATGAACATGATGCTGCTGATGATACTGTGCATACG
AGGCTGGTGAGAGTGTATCACTAAATGAATCACTTGTAGTAGGTGATGTACTCCTTTGCTCTTCATTCATCTCCGCCTGTCCTGCGGTGGTAAGATGCTCAGCCTCGGCCGCTTCCTCCGCTTCCTCAATTAATTCGTCTAAGAAGCCCTCCTCCTCCTCGACTTGTTCCTTCGAGCCGACCCACACACCAATTACGGTCAGGATGA
CACCAACACCAAAGAGTGCACCTTGAACCGGG

>GRPc97904 (Virulent pathotypes only)

GTTTTAGTGCAGGTGGGAGTGCGGCTAATGGTGCCTCTTTGATAAGATCAATATTATCGCGGTGAAAGAAACTGATAGTCAGTTCCTTTAAACGGGTTAAGTGTGATAGTTGTTGATAATCCGTCCAGTCGACCGGGCACTTCGATAAATCATCAAATTCGAAACGAAGAGTGGTCAACGATTTCAAATGCACCAATGATGCGA
TCAGTTCAGGCCACCACATACCATTGGATAAGCGTAAGCAGAGAGTTGTTAATTTGGTTGGCCAGGTGGTTGGTTGGTTGATGTTGTGGGCTCGGTGTACATCATAGGAGCAACTGAAGGTGGTTAGGTTAGGTAAGAGTGGCAGGAATTGGTTGGCAAATTCATACTCCTGCGTGTTGAAATATCGTGCAGGTGTACGTGG
TTTAATCGTCGTGAAATGATG

>GRPc98458 (Virulent pathotypes only)

CACTCAACACATTCAACATGAACAAGATTTTGATCATCGCCGCCATCGTCGCAGTCTTCGCCACTGCTGCCTTTGCTCAACAATGCAATGCCTCAAGCGCCACTAGCTGCGCCAACACCTGGTCAACTTGCGTCACCAGCGGTGGTCTCGACGTTAACAAGATCTGCGCTTGCTACGGTCCTTACTTCTCTTGCCTCCGTAGCGCCA
ACTGCTTGACTGATTCCCAATTGACCAGCATCCAGCAATCCTGCTCTCAATCCGGCTGCACTACCGCTCAATGCAGCTCTGCCGTCTCCATCTCCGGCGTTTTCGGCTTGTTCGTCGCTGCCGTTGCCGCTTTCTTCTAAGCGAAGTTACTAGCAACGTATTCATGTTTGCGAGCGCAATCATGTATTTTACTGATTTTCAATAAACA
ATCGGTTAG

>GRPc98717 (Virulent pathotypes only)

ATGACACCATTAGTGAGCCATTTCATTTCCATACTATTAGCTGAGATATCACCGATTGATGCGGAGGCCTCGGTCTCTGATTTACCCTTGTGGGATACACTGCGAGAATTGGATCCTGCTCAACCCACGGATCAAATGAAATTGCAAGCTTTAACTGCCATTATGGATTTAGCAATTGAAGTCGCGGATGATGAAATTGAGGAGT
TTGAAGTCTTACTGGATAAGAATGAGGAGATCATGAAGGCATTAGAGTATATGAGCAACACGATCCCAAAATTGCTGAACTCGTTGCCTTCTTCAATGACATCTGCTTCCACAAGCTCCACAACAGCCGCATCTGAAGCACCTGATTCGGTTCAAGATCAATTACAACGCGAATTAACATTATTGAATGATTTTATTGCCAGTTGC
CAATCTACGGG

>GRPc99396 (Virulent pathotypes only)

TACACAATGCACGACTCGCCCATCTGCCTCTTCGCTTCTCACTTTTGCTCTTAACATGATCAACCAACCATTTCATCGGCTGTTAGTGTGGGCAGTCATCACGTTAGTGCAGCTGCATTGTCTGACCTTTGCTGCAGCGGGTGCTGCAGGACCTGGTGGTCGTGGATTAGATCCGCTGCCTCCTTCTGTCACCATTGTTGGTTTGGG
CGAGCCATGCACGCTGACATGGAATTCAACTGATGTCGATTTACGATGGTGCGGTGCCGGCCTCAGTTGCACCTATGTATCAGCTTATAATAGCACCGACGGCGATGATCCTATCTTTGCTGAGTATTATCGACAGGCTGAACGGTGCAGCTTCTGTGAAGGTGCTTCATACGACATGCAGCGAATGGCACAATGCAATACCAA
TGGTATG

>GRPc99719 (Virulent pathotypes only)

GTAAAAAGTCAGGGAAGGATACCATTCAAACGTGTGAACGTTGCTATTTCCTCCAGCCTGAACGTCAGATGTGGATGTCCCGTGCTCATTGCTGTCCGAATGGTGGTGGTGTGGCTGATCTCAAACATGTCAATGATATTGGGGATCGTAAATGTAGTAAACCGAATTGTACCTGTCCTATCAGTGCTGAGGATCAAAAGAAG
CGTGATGAGAGGGCGAAGGAAGTGGATCATAGTCGACCAATGAAGGATCGAGCGAATCAAGCGAAGTGGAAAGAACAATTTAAGAAGGCAGGATATATATGTGAACGATGTCCAACTTATGCAGAGTGTATGACATGTTATCTCAAATTCTCACAAATTCGTCTAACAGCCGAATCGAGTCAAGCATTTAGGCAAGTCATTG
CTGGTAAGAAG

>GRPc100595 (Virulent pathotypes only)

ACCAGTCCAATTACCAGCTTCACATTGATAGGCACCATTCGTGAATGAAGTCTGATATAAACCGGCGATACAACCAACCAAACAGTAACTACCATCTGCTAAGGAATTATTTGCATTAATCGTAGCACAGGAGACATAATCATTACCATAATTCACACGACCATCTAACGTTTGTCCTCTTGTCAATGGTGGTGGAAAATTACAGT
CTTGACCAGCAACCATACTCACCGTGAGGGTGAATAGGAGGGCGACGAGGCCCAACCAACGAATGGAGCAAGTCATCTAGCTTCTGAAAGACTGTGAGCTAGAATCTGAACGGGTCAGACGTCGGATGGGAGTTGGGTGTGAGGAGGGGTTGCGGGGAGGTGACGGTTGCAGTGGGAGTACGAATGCGGACGTGACGC
GTATTCG

>GRPc100968 (Virulent pathotypes only)

GGGTTTGAGGATGACGTTTCCGGTGCTGCATTTGACGCATTTAGCGGTGCATCTAGACTTTGTCGCGGTGAGATACGATTATGATGATGTGGGGCAGTCTGAAGTGTATTGGATGATTGTGGTGTCGAAGGTGCAGTCGGCTTCTTGTTGCCACTCGGACTACTCGTATTCGAGACAGTCGTCGACTGCTGTCGTTCCTTCAAT
TCTTCCACACGCTCATTAACAGTCTTAAATTCCTCCTCCACGCGTGCACGGTATGATTCACTTGATAATTGTGCATGTTGAAATTCGACAATCTGAGTCTGAACGGAGTGTTCTAAACGATGAAATAAGGATTGTAAATGATCGAACTCATGGCGCGTCACAAGGGCGTGTCCATCCTTACCTCCGCTACTGCGTGGACTCATG

>GRPc104513 (Virulent pathotypes only)

CGGTACCGTAAGTGTTGATGAACTCAATTTCGAGCTCCTCCAGCGCCTCCGCCACGATCTCGTCCTGCGTCTCCGCGTCGAGCCCGAACGTACCGGCTTCCGAGAGCGACGCGAAGTCCAGCGCCTCCGCCGAAGCGCCCACTTCCATGAGCGACAGCGACGACAGGTCCACGGCGGCGGCGCCAGCGGTGGCAGCGGCAGA
CACCTCCATAGCGACGACCGGCGCGAGCTCCACCGCCTCCGCCGCCTGCGCGTCGGCGCTGGTGCTGGCGACCGTCACCCAATCGGCCCCGTCCGCGCGCACACCTGACGCCGCAGCGCACACCGCCAGCGCCAGCAAAACGGCGATGCTTCCGCTCCGCATCGCGTCGTTTCTTAGTTTATCGAG

>GRPc105415 (Virulent pathotypes only)

TTGAATTCGTGCAGACAGATTGACGATTTCACCCATTTTGGACCATCAATTGCACACAATGAGGGACAAGAGTACCTTCACCACCATACTGCTACTGGTGTTGGCTATCAGCACGAGTATCTCCTCAGTCGTCGCCTTCGACTGTTCCTCAGTTCAAGATTGTGCCACCTGTCTCGCCACACCCGGCTGTGGATTCTGTGCTGGTA
CAACCACAGGATCATGTGTATCTGGAAACAGAACTGGACCATTCAATCCAACAGATTGTCCTGCACCAGCATGGATAGCCACCGCCAATCAATGTGCCCATCGTCCATGCTTCACACCGACCCTACCAGCGGATACCTCCTATGGTACTTGTCTGACATCGACCGGTTGGTCAACGG

>GRPc105801 (Virulent pathotypes only)

CCTCCGCTCCTGGCGGCTCCGCCCATTCCAACGGCGCGATATCTGTACTTCATCCTTGACTGACGGAGCTGGTACGTCATTCGAACAGATACCCGATTCATACGTGACACTGAATGGATATGGATCACCACCCACACACTGTGTGTCCTCTGCTTGATAGACTGATAGTACAGCTTTTGCTGGTGTCTGACCGAACACAGTTGAC
GACACCTGCAACACCATCAAACACCAGAGGAGGAGAGTAGGTTTCATTGTTCGAGAGGAGACGTGTTGATCTACTGAGGGGGAAACAACGTCGAGAATGAAGATCAATGGTCGACTAGTCTCGTCGATTAGTGGTATTGTTGGGTGATCGGTGGTGTTGTGCACGGGTGTGATTG

>GRPc108599 (Virulent pathotypes only)

CGAGCATCAGCATTAGCATCGGCATCAGCATGCGCGAGCACGAGTGCAGCAGCGAGGTGACCGGCAGCGAGCGAGCACGGCAGCGAGCAGCATTAGCATGCGCGAGTACGAGCACGGCATCGAGCACGGCAGGAGCACAGCAAGCATCAGCATCAGCATGCGCAAGCACGAGTACGGGAGCGAGATGCACGGCAG
CAAGCACAGCAGCGAGCATCAGCATCAGCATGCGCGCGCACGAGCACGGCATCGAGCACGGCAGGAGCACAGCAAGCATCAGCATCAGCATGCGCAAGCACGAGTACGGGAGCGAGATGCACGGC AGCAAGCACAGCAGCGAGCATCAGCATCAGCATGCGCGCGC

>GRPc110086 (Virulent pathotypes only)

ATGGGGATGGAATTTCAGTCGCTGGAGGTAGCACACCAATTGAACCATCATCAAGTCGCTTCAAGGCAGGCGTGGAGATGGATCTCGCGAAATGGCCCGATGCATCGCCTAATGAAGCAGGTTGTTGTGACGGTATCGCGCCATAATTCATCGATTGTCCAAAGTCCGGTCGTGGTATCCTCGCCCACGAGTTGGTCCGCCTG
AGGTTCCGTCTCGCGGCTTCTTCACGCGCACGTGCGATGGCCTCCTCATCATCCAGCTCCTCGGGGTTCTTCCAGAAGTAGACGACAGCGACCAAGGGCAGCATCATCAATGATTCTAACATGAAGCCGTAATGCCAGCCGAATTTATCGAAA

>GRPc110860 (Virulent pathotypes only)

TATTGGTCGTTGTGTTACACGCCATCCTTGCTCCTCTCCTCCAACTCGCCTCTCCTCTCCTCATCTCATCTGATCTGATCGGACTCCTTCCACAACAGCGCACACCTACAAGCAATCGTGCAGCAGCTGCACCAATTAACCAATTAAATCGAGTTGAGATGTTTGGTGGACTGAAGACAGCGATGGCAAATGCACATCAACCACAGC
TGCAACAACCACTACAACACTCATCACAAGCATTCCAACGACCGAGACAACACCACCCGCCGGTGCAGTTGCAACCATCACCACAACCACATACAAATTCCAATGAACAACCGTCAATGATGGATGATGATGATGCAGCTACGAG

>GRPc115391 (Virulent pathotypes only)

CGTACCGTCCTTCGCCATTCCATTCCTTTTCTATTCACCATGAACTACTCACACCTATTCCTGGTGGTGGTGAGTGTTCTCTATGGCGTCACAGCTACTGCCACTGCCGCCTCCTCACCCAACCATGCAGCCGCTGTCGTTGACTCAACTCCCCAAACTACCCTATCCATGTACGTGTATCACAATCTCGACTGTTCAGGCGAGCCAA
GTGAAATTTCCAACATGACCATCGGTGAATGTGTCAGTGATGGTGTATACTTCTATCCATCGAATGATGGTCAATTTGTATACAGCTTTAGTACGATTGATCTTACTTGTGAACAGGCG

>GRPc117804 (Virulent pathotypes only)

GTCGAGGTGGCAGAATCCCAGGCAATTGGCCCTCCTGTGGTACGTAGGTGGTCGTAGCAGCGATTGTGATGATCGCGGATCAATCAGTTTGTTTGGGTTGAGCCATCTTGGAGCGCATTCGGTGGTGTCGTCGTCCATGATACGCCTCATGTTTCACCAGTCCGACCTCGGCCTCATTTAATTTATTCTCGACCGGGTTGTAGG
TGAACTCGGATTCGGAATTGTGAAGTGATGAACTCATAGTCGAAAGGCTCACCTTCGTCTCGGCGTGAGTGAGGATAGTGCAGAAGCACAAAGAGATGAGAGCCACCAAA

CHAPITRE 5: Discussion générale

Cette thèse avait pour but l'identification d'éléments génétiques directement liés à la capacité de certaines espèces de nématodes (*G. rostochiensis* et *G. pallida*) de pouvoir parasiter la pomme de terre, alors que d'autres espèces très proches (*G. tabacum* et *G. mexicana*) en sont incapables, ainsi que d'identifier les éléments génétiques liés à la différence de virulence des différents pathotypes de *G. rostochiensis* (Ro1-5) sur les cultivars de pomme de terre possédant le gène de résistance H_1. Le projet a également permis de contribuer à l'assemblage et l'annotation du génome et du transcriptome de *G. rostochiensis*, à l'identification de nouveaux gènes effecteurs potentiels, ainsi qu'à l'étude de la dynamique de l'expression génique des NKPT à différents stades du cycle de développement (dormance, éclosion, etc.) afin d'identifier les voies métaboliques activées durant les stades clés (p. ex., éclosion des larves). Ce travail a globalement permis d'approfondir les connaissances sur les interactions moléculaires entre les NKPT et leur hôte principal, *Solanum tuberosum*.

L'analyse du génome de *G. rostochiensis* (Annexe 1) a permis l'identification de 91 protéines acquises par transfert horizontal de gène. Plusieurs de ces protéines avaient déjà été identifiées dans d'autres espèces et associées à la capacité des nématodes de parasiter les plantes (Blaxter & Koutsovoulos, 2015; Danchin et al., 2016; Haegeman et al., 2011; Quist et al., 2015; Scholl et al., 2003; Yang & Luo, 2013). Ces protéines ont, ici aussi, en grande partie une fonction liée au parasitisme, telles la dégradation de la paroi végétale et l'évasion du système immunitaire (p. ex., cellulase, pectate lyase, chosismate mutase) contribuant au pouvoir pathogène du nématode. Également, l'analyse des régions promotrices en aval de gènes d'effecteurs des glandes dorsales a permis de trouver un court motif retrouvé chez 77% de ces gènes d'effecteurs. Ce motif fut utilisé afin d'identifier de nouveaux gènes d'effecteurs potentiels, ainsi que deux gènes possédant ce motif dans leurs régions promotrices, en plus d'avoir un signal peptide pour excrétion et une expression coïncidant avec un rôle dans le parasitisme. Un tel motif avait déjà été retrouvé chez des organismes phytopathogènes et des gènes d'effecteurs potentiels avaient été identifiés de cette manière chez *Fusarium oxysporum* (Schmidt et al., 2013). L'identification des gènes d'effecteurs était cruciale puisque l'hypothèse la plus généralisée afin d'expliquer les

différences de virulence entre les espèces et pathotypes de NKPT est la présence de mutations ou de différences d'expression dans ces gènes. L'analyse des différents stades de développement des NKPT a d'ailleurs permis de démontrer que la plupart de ces gènes étaient induits lors de l'éclosion et que les larves J2 pouvaient donc être utilisées dans notre étude (Annexe 2). L'éclosion est une étape clé du cycle de vie puisque le nématode en dormance doit détecter la présence d'un hôte potentiel et enclencher son processus d'éclosion. Après le contact avec l'exsudat racinaire, le gène le plus surexprimé était la néprilysine, une métalloprotéase impliquée dans la dégradation des peptides et dans les modifications post-traductionnelle. Ce gène de régulation fut aussi identifié comme l'un des plus surexprimés chez les nématodes capables d'infecter la pomme de terre au Chapitre 3. Son expression a également été corrélée au potentiel d'éclosion de *G. pallida* suite à l'exposition à différents éliciteurs (Hoysted et al., 2018). Chez *C. elegans*, cette protéine est aussi surexprimée juste avant l'éclosion (Spanier et al., 2005). L'expression de ce gène pourrait donc être à la tête d'une cascade d'induction génique menant à la synthèse des effecteurs. La plupart des gènes impliqués dans le parasitisme étaient justement surexprimés suite au contact avec l'exsudat racinaire à la suite de la néprilysine (p. ex., chitinase, endoglucanase, expansine).

Les données de séquençage d'ARN utilisées pour l'étude de la dynamique des transcriptomes des NKPT à différentes étapes du cycle de vie furent également utilisées afin de sélectionner un groupe de gène de référence, pour la normalisation de données d'expression RT-qPCR chez *G. rostochiensis* (Chapitre 2). Cette méthode a permis la sélection des meilleurs candidats disponibles, puisqu'elle prend en compte tous les gènes exprimés durant les étapes du cycle de vie du nématode utilisé ici. Un groupe de trois gènes de référence (GR, PMP-3, and aaRS) fut sélectionné en analysant la stabilité de l'expression des gènes candidats de données de séquençage d'ARN et de RT-qPCR. L'utilisation de plus d'un gène de référence est essentielle afin d'assurer la stabilité des rapporteurs; malheureusement nombre d'études utilisent encore un seul gène de référence ou des gènes de références non validés ce qui remet en cause la validité de leurs résultats (Dheda et al., 2005; Huggett et al., 2005; Zhou et al., 2017). Ici, en utilisant deux techniques quantitatives pour le choix et la validation de ces gènes de référence, en plus d'avoir normalisé avec succès les données d'expression de gènes connus, les résultats devraient être hautement fiables.

L'étude des éléments génétiques liés à la pathogénicité sur la pomme de terre chez les espèces du genre *Globodera* (Chapitre 3) fût conçue afin d'identifier les éléments génétiques uniques aux NKPT en les comparant à des espèces très proches, mais incapables de se développer sur la pomme de terre (*G. tabacum* et *G. mexicana*). Les résultats ont montré que certains gènes n'étaient pas exprimés par *G. tabacum* et *G. mexicana* lorsqu'exposés à l'exsudat racinaire de pomme de terre contrairement à *G. rostochiensis* et *G. pallida*. Un de ces gènes, similaire à une polyubiquitine-B, serait potentiellement un effecteur. En effet les ubiquitines sont impliquées dans la survie du nématode lors de l'infection de l'hôte (Chronis et al., 2013; Haegeman et al., 2012). Les autres gènes uniques au NKPT n'ont pas de fonction connue. Une comparaison similaire avait déjà été réalisée par le passé, où les transcriptomes des larves J2 de *G. pallida* et de *G. mexicana* furent comparés par hybridation soustractive (*Suppression Subtractive Hybridization*), cependant, aucun gène unique à une espèce ne fut identifié (Grenier et al., 2002). L'analyse de gènes différentiellement exprimés a permis de mettre en évidence que près de 25% de tous les gènes d'effecteurs, en addition de plusieurs gènes régulateurs, étaient différentiellement exprimés et que la majorité d'entre eux était significativement surexprimée chez les espèces *G. rostochiensis* et *G. pallida*. En moyenne, l'expression de ces gènes était 6,6 fois plus importante par rapport aux espèces incapables de se développer sur la pomme de terre, *G. tabacum* et *G. mexicana*. Une grande quantité de gènes d'effecteurs sous-exprimés chez les espèces non pathogènes sont impliqués dans l'organisation du cytosquelette (Mei et al., 2018), un mécanisme important dans la création du site de nutrition et qui correspond à l'étape du processus d'infection que les espèces non pathogènes sont incapables d'accomplir (Thiéry et al., 1997). L'incapacité d'infecter la pomme de terre pourrait être le résultat d'une mauvaise activation de la transcription des gènes d'effecteurs suite au contact avec l'exsudat racinaire de pomme de terre chez *G. tabacum* et *G. mexicana*. Telle que mentionnée plus haut, l'analyse a également identifié un gène de régulation, une métalloprotéase (néprilysine), aussi identifiée comme étant impliqué dans le mécanisme menant à l'éclosion des larves J2, suite au contact avec l'exsudat racinaire d'une plante hôte (Annexe 2). Cette protéine pourrait aussi être impliquée dans l'activation de l'expression des gènes d'effecteurs. En effet, ce gène est exprimé 28 fois plus chez *G. rostochiensis* et *G. pallida* que chez *G. tabacum* et *G. mexicana*, et pourrait induire non seulement l'éclosion des larves J2, mais aussi l'expression de gènes d'effecteurs. Une autre métalloprotéase avait déjà été mise en cause dans le processus d'éclosion chez

G. rostochiensis et *Heterodera glycines*; le gène était surexprimé juste avant l'éclosion des larves (Kovaleva et al., 2004). Afin de vérifier si l'expression de cette néprilysine et d'un sous-groupe de gènes d'effecteurs surexprimés était la même sur différents hôtes potentiels, l'expression de ces gènes fut mesurée par RT-qPCR à la suite du contact avec de l'exsudat de tomate et de pomme de terre. Curieusement, la surexpression des gènes d'effecteurs, comme observée lorsqu'exposés à l'exsudat de pomme de terre, n'a pas été observée avec après l'exposition à l'exsudat de tomate, suggérant que le nématode phytoparasite serait capable d'ajuster l'expression de ses gènes d'effecteurs dépendamment de chaque hôte potentiel et que ce mécanisme serait induit par le signal chimique de l'exsudat racinaire de l'hôte. L'expression de la néprilysine fût similaire lorsqu'en contact avec l'exsudat racinaire de tomate, mais était toujours sous-exprimé chez les espèces non pathogènes, et ce, peu importe l'exsudat racinaire, démontrant l'importance probable de ce gène pour *G. rostochiensis* et *G. pallida*. Ici, les résultats devraient être très fiables, puisque quatre populations de chaque groupe furent utilisées dans les analyses, en plus d'avoir confirmé l'expression obtenue du séquençage d'ARN de plusieurs gènes d'intérêt par RT-qPCR. Ces résultats appuient l'hypothèse de départ selon laquelle des gènes codant pour des effecteurs possèdent des mutations dans leurs séquences et/ou un taux d'expression spécifique expliquant la pathogénicité de certaines espèces de *Globodera*. En effet, le taux d'expression d'un groupe de gène d'effecteurs est, ici, certainement en cause et explique la différence de pathogénicité entre ces espèces. Cette étude est la première à montrer une différence de régulation majeure dans l'expression d'un grand groupe de gènes effecteurs chez des nématodes phytoparasites et il serait intéressant de poursuivre la comparaison en y ajoutant différents hôtes (p. ex., tabac) lors de l'induction de l'éclosion des larves J2. Les résultats RT-qPCR comparant l'expression de gènes suite au contact avec l'exsudat racinaire de pomme de terre et de tomate sont très prometteurs, mais ne sont réalisés que sur peu de gènes. L'expression des gènes effecteurs était différente dépendamment de l'hôte et il serait intéressant d'effectuer une comparaison par séquençage d'ARN afin de vérifier qu'un grand groupe de gènes d'effecteurs se comportent ainsi afin d'identifier les gènes régulateurs impliqués.

L'analyse comparative des cinq différents pathotypes de *G. rostochiensis,* quant à leur différence de virulence face au gène de résistance H_1 (Chapitre 4), a permis l'identification de

plusieurs effecteurs significativement surexprimés. Vingt protéases à cystéines, jouant un rôle dans la reproduction, le parasitisme et l'évasion du système immunitaire étaient exprimées en moyenne 33 fois plus chez les pathotypes virulents, contribuant vraisemblablement à leur virulence. De plus, ces gènes surexprimés chez les pathotypes Ro2, 3 et 5 avaient une expression près de cinq fois supérieure pour le pathotype Ro5 comparé aux pathotypes Ro2 et 3, ce qui concorde avec le meilleur taux de reproduction observé pour le pathotype Ro5. Cependant, puisque le gène de résistance H_1 induit une réaction d'hypersensibilité, la reconnaissance d'un gène, ou mutation, spécifique est nécessaire et doit être unique aux pathotypes avirulents (Ro1 et 4). Ainsi, les résultats ont démontré que trois gènes codant pour des effecteurs étaient uniquement exprimés par les pathotypes Ro1 et 4 et que quatre gènes d'effecteurs possédaient des mutations non synonymes homozygotes seulement dans ces mêmes pathotypes. Étonnamment, un grand nombre (24) de gènes non caractérisés possédant un signal peptide pour excrétion étaient aussi uniques aux pathotypes Ro1 et 4, ce qui nécessiterait des investigations futures. Une comparaison similaire fût précédemment réalisée entre les génomes de deux populations de *G. pallida*, afin d'en identifier des variants liés à la différence de virulence sur les cultivars possédant le QTL de résistance $GpaV_{vrn}$ (Eoche-Bosy et al., 2017). Étonnamment, aucun variant situé dans une région codante ne fut rapporté. Ici, le séquençage d'ARN permet la recherche de variants directement dans les gènes exprimés, augmentant donc le pouvoir de détection de variants fonctionnellement important (Piskol et al., 2013). Ces gènes et mutations uniques aux pathotypes avirulents nous permettent d'établir une courte liste de gènes candidats qui devront être directement testés sur un cultivar de pomme de terre possédant le gène de résistance H_1 afin de confirmer s'il y a une réaction d'hypersensibilité induite chez l'hôte. Pour ce faire, il est possible d'induire l'expression des gènes d'intérêt en utilisant une technique d'agroinfiltration, qui provoquera une réponse d'hypersensibilité s'il y a reconnaissance du gène d'avirulence (Sacco et al., 2009). Cette technique a déjà été employée afin d'étudier l'effet de plusieurs effecteurs potentiels sur la plante hôte (Ali et al., 2015a; Ali et al., 2015b). L'hypothèse de départ, proposant que des mutations et/ou un taux d'expression spécifique de gènes d'effecteurs expliquent la différence de virulence entre les pathotypes, sera validée si l'une des mutations uniques associées au pathotype avirulent induit la réponse immunitaire sur un cultivar résistant. Si la réponse immunitaire est plutôt induite par un gène unique aux pathotypes

avirulents, l'hypothèse sera confirmée seulement si le gène est présent dans le génome et qu'il n'est simplement pas exprimé dans ces conditions.

La recherche d'éléments génétiques liés à une caractéristique observable en utilisant le séquençage d'ARN présente beaucoup de limites et d'incertitudes. Ces incertitudes sont d'autant plus présentes du fait d'étudier un organisme non modèle, dont les assemblages de génome et de transcriptome disponibles ne sont pas complets et dont une grande proportion de gènes reste encore non caractérisée. Plusieurs facteurs peuvent avoir une influence sur les adaptations génétiques sous sélection liées à certains groupes de populations, parmi lesquelles on peut compter l'hôte, l'environnement et des caractéristiques phénotypiques, et il peut donc s'avérer difficile d'associer un marqueur génétique à un phénotype particulier. Afin de s'assurer d'obtenir des résultats fiables, il est encore plus essentiel d'avoir un bon design expérimental. En ciblant principalement les gènes d'effecteurs et des gènes potentiellement impliqués dans la pathogénicité, nous souhaitions éviter de sélectionner des gènes qui seraient liés à une autre variable. De ce fait, nous manquons aussi des gènes importants, dont leur fonction n'a pas encore été associée à la pathogénicité du nématode. L'identification de gènes non caractérisés possédant un signal peptide pour excrétion, permet néanmoins de conserver ces gènes potentiellement importants dans le mécanisme moléculaire de la résistance de l'hôte. De plus, la disponibilité de populations, pathotypes et espèces du genre *Globodera* fut également un facteur limitant. On ne retrouve que *G. rostochiensis* Ro1 au Québec, et l'acquisition d'autre espèce ou pathotypes doit se faire avec la collaboration d'équipes de recherche étrangères. L'utilisation de plus de populations lors du séquençage d'ARN dans les Chapitres 3 et 4 aurait permis de circonscrire davantage l'identification de gènes d'intérêt, et d'ainsi obtenir des résultats plus fiables. La validation des gènes d'intérêt issue du Chapitre 4, par PCR, permettra tout de même de valider ces gènes sur des populations supplémentaires en nécessitant moins de matériel biologique et à moindre coût.

Les résultats obtenus aux chapitres 3 et 4 suggèrent la présence d'un ou de quelques gènes de régulation induisant une cascade menant à l'activation de la transcription d'un grand groupe de gènes d'effecteurs. Ces gènes de régulation devraient vraisemblablement être activés par le contact de composés présents dans l'exsudat racinaire de l'hôte et différents groupes de gènes

d'effecteurs pourraient ainsi être exprimés selon des patrons d'expression différents tout dépendants de l'hôte présent. La présence d'un gène régulateur, responsable d'initier l'expression à large échelle des gènes de virulence, chez des pathogènes humain et animal, fut souvent proposée sans pouvoir être démontrée (Coulter et al., 1998; Kohler et al., 2002). Le passage de ces pathogènes, comme pour les NKPT, par plusieurs environnements différents durant leur cycle de vie nécessiterait plutôt plusieurs gènes de régulation afin d'induire l'expression de gènes d'effecteurs au bon moment (Thomas & Wigneshweraraj, 2015). Dans ce cas, le signal chimique de l'exsudat racinaire semble déjà induire l'expression d'une grande quantité de gènes d'effecteurs impliqués dans la formation du site de nutrition chez les espèces de NKPT. De l'autre côté, les espèces *G. tabacum* et *G. mexicana* semblent avoir perdu la capacité d'activer cette cascade d'événements, perdant ainsi également la capacité d'infecter cet hôte. En effet, l'arbre phylogénique, présenté au chapitre 3, suggère une acquisition de la capacité d'infecter la pomme de terre antérieurement à leur spéciation. Chez les pathotypes de *G. rostochiensis*, Ro5 semble aussi avoir un meilleur taux de reproduction corrélé avec un plus haut taux d'expression d'un groupe de gène d'effecteurs, suggérant aussi une plus grande activation de leur transcription. Ces observations correspondent avec les résultats rapportés à l'Annexe 1, qui montrent un motif spécifique identifié dans la région promotrice de 31 gènes d'effecteurs connus suivant le même patron d'expression, et ont permis l'identification de plusieurs nouveaux gènes d'effecteur potentiels. Comme ici, aucun gène régulateur ne fut identifié, mais son existence fut également proposée (Eves-Van Den Akker & Birch, 2016). De plus, chez *G. rostochiensis*, plusieurs gènes d'effecteurs de mêmes familles furent localisés sur des îlots génomiques qui étaient pour la plupart composés de gènes exprimés uniquement dans la même glande, dorsale ou subventrale (Annexe 1). Les gènes d'effecteurs de la glande subventrale sont principalement exprimés durant les premières étapes de l'infection et ceux exprimés dans les glandes dorsales sont plutôt exprimés durant la phase sédentaire (Endo, 1987). Les îlots génomiques denses en gènes d'effecteurs sont donc potentiellement composés de gènes exprimés au même moment et dont l'expression est induite par peu de gènes régulateurs, d'où la différence d'expression observée pour des groupes de gènes impliqués dans les mêmes processus (Chapitre 3 et 4). L'implication de la néprilysine dans l'éclosion et la régulation de l'expression de gènes d'effecteurs est très probable, mais l'idée d'une seule cascade d'activation induite par l'exsudat racinaire et résultant en l'éclosion des larves et l'expression des gènes

d'effecteurs est probablement trop simple. La dynamique d'expression des gènes d'effecteurs doit être sophistiquée et coordonnée en réponse à plusieurs signaux de l'hôte.

Perspectives

L'expression d'un groupe de gènes d'effecteurs suivant tous un patron transcriptomique similaire, tel qu'observé dans l'étude comparative entre les différents espèces et pathotypes, ainsi qu'en accord avec les conclusions de l'étude présentée à l'Annexe 1, suggère l'existence d'un ou de quelques gènes régulant cette expression. Le gène néprilysine est ici suggéré et serait unique aux espèces *G. rostochiensis* et *G. pallida* selon les résultats obtenus. La validation de ce gène régulateur serait une avancée remarquable dans la recherche sur les nématodes à kyste de la pomme de terre, ainsi que pour l'ensemble des nématodes phytoparasites. Cela permettrait non seulement de comprendre davantage le mécanisme moléculaire de l'infection de l'hôte, mais permettrait aussi le développement de moyens de contrôle plus efficaces et durables. Une méthode de contrôle exploitant l'inactivation de ce gène régulateur permettrait d'empêcher l'expression d'un groupe de gènes d'effecteurs essentiels à l'infection de façon telle que les NKPT seraient incapables d'infecter la pomme de terre. Ce moyen de contrôle serait envisageable grâce à la méthode HIGS (Host Induced Gene Silencing, Inactivation de gène induit par l'hôte), basée sur l'ARN interférence pour empêcher l'expression du gène chez le parasite (Ghag, 2017; Yin & Hulbert, 2015). Un inconvénient de cette méthode serait qu'elle implique l'utilisation de plantes transgéniques qui soulèvent beaucoup d'inquiétudes dans la population. Cependant, cette approche offre de bonnes perspectives, ayant déjà permis une réduction marquée de la pathogénicité du nématode phytoparasite *Meloidogyne incognita* ou dans le contrôle de maladies virales (Yadav et al., 2006). De plus, cette méthode permet une évaluation rapide de son efficacité en laboratoire par la répression du gène d'intérêt sur une plante hôte par agroinfiltration.

Tel que nous avons prévu avant la publication de l'article découlant du Chapitre 4, la validation des variants et gènes uniques chez d'autres populations de *G. rostochiensis* permettra de réduire davantage la liste de gènes d'avirulence potentiels. La validation d'un gène d'avirulence induisant une réaction immunitaire chez les cultivars possédant le gène de

résistance H_1, parmi la courte liste de gène candidat, constituerait également une avancée majeure vers une meilleure compréhension de cette résistance. Le gène d'avirulence pourrait être utilisé comme outil dans l'identification du gène de résistance H_1 chez des espèces proches (Mantelin et al., 2017). La validation permettra aussi de développer un test moléculaire diagnostic capable de déterminer si un nématode de l'espèce *G. rostochiensis* est virulent ou avirulent sur un cultivar de pomme de terre possédant le gène de résistance H_1. Présentement, seul un test de croissance sur ce cultivar peut différencier ces nématodes. Un test moléculaire permettrait d'obtenir rapidement les détails de la virulence d'un nématode afin de développer directement une stratégie de contrôle.

Le développement d'une nouvelle méthode de contrôle des NKPT les empêchant d'infecter les cultures de pomme de terre est très certainement l'objectif ultime de tout programme d'étude sur ces nématodes phytoparasites et la compréhension des mécanismes d'infection est une étape préalable obligatoire. Ainsi, les résultats présentés ici contribuent à cette compréhension et permettent certainement un avancement majeur dans l'étude des nématodes phytoparasites du genre *Globodera*.

BIBLIOGRAPHIE

Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M. C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T. R., Markov, G. V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Segurens, B., Steinbach, D., Tytgat, T., Ugarte, E., van Ghelder, C., Veronico, P., Baum, T. J., Blaxter, M., Bleve-Zacheo, T., Davis, E. L., Ewbank, J. J., Favery, B., Grenier, E., Henrissat, B., Jones, J. T., Laudet, V., Maule, A. G., Quesneville, H., Rosso, M. N., Schiex, T., Smant, G., Weissenbach, J., & Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature biotechnology, 26*(8), 909-915. doi:10.1038/nbt.1482

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. doi:10.1016/j.meegid.2014.07.027

Ali, M. A., Azeem, F., Abbas, A., Joyia, F. A., Li, H., & Dababat, A. A. (2017). Transgenic Strategies for Enhancement of Nematode Resistance in Plants. *Frontiers in Plant Science, 8*(750). doi:10.3389/fpls.2017.00750

Ali, S., Magne, M., Chen, S., Cote, O., Stare, B. G., Obradovic, N., Jamshaid, L., Wang, X., Belair, G., & Moffett, P. (2015a). Analysis of putative apoplastic effectors from the nematode, *Globodera rostochiensis*, and identification of an expansin-like protein that can induce and suppress host defenses. *PLOS ONE, 10*(1), e0115042. doi:10.1371/journal.pone.0115042

Ali, S., Magne, M., Chen, S., Obradovic, N., Jamshaid, L., Wang, X., Belair, G., & Moffett, P. (2015b). Analysis of *Globodera rostochiensis* effectors reveals conserved functions of SPRYSEC proteins in suppressing and eliciting plant immune responses. *Frontiers in Plant Science, 6*, 623. doi:10.3389/fpls.2015.00623

Andersen, C. L., Jensen, J. L., & Orntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research, 64*(15), 5245-5250. doi:10.1158/0008-5472.CAN-04-0496

Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., & Singh, K. B. (2010a). Plants versus pathogens: an evolutionary arms race. *Functional plant biology : FPB, 37*(6), 499-512. doi:10.1071/FP09304

Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., & Singh, K. B. (2010b). Plants versus pathogens: an evolutionary arms race. *Functional Plant Biology, 37*(6), 499-512. doi:10.1071/FP09304

Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution, 23*(1), 26-32. doi:10.1016/j.tree.2007.09.011

Bakker, N., Vervelde, L., Kanobana, K., Knox, D. P., Cornelissen, A. W. C. A., Vries, E. d., & Yatsuda, A. P. (2004). Vaccination against the nematode *Haemonchus contortus* with a thiol-binding fraction from the excretory/secretory products (ES). *Vaccine, 22*(5), 618-628. doi:10.1016/j.vaccine.2003.08.025

Bekal, S., Niblack, T. L., & Lambert, K. N. (2003). A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. *Molecular Plant-Microbe Interactions, 16*(5), 439-446. doi:10.1094/MPMI.2003.16.5.439

Bélair, G. (2005). Les nématodes, ces anguillules qui font suer les plantes... par la racine. *Phytoprotection, 86*(1), 65-69. doi:10.7202/011717ar

Bélair, G., Dauphinais, N., & Mimee, B. (2016). Evaluation of cultural methods for the management of the golden nematode (*Globodera rostochiensis*) in Quebec, Canada. *Canadian Journal of Plant Pathology, 38*(2), 209-217. doi:10.1080/07060661.2016.1189969

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2009). GenBank. *Nucleic Acids Research, 37*(Database issue), D26-31. doi:10.1093/nar/gkn723

Bird, D. M., Opperman, C. H., & Davies, K. G. (2003). Interactions between bacteria and plant-parasitic nematodes: now and then. *International Journal for Parasitology, 33*(11), 1269-1276. doi:10.1016/s0020-7519(03)00160-7

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*(1), 15-22. doi:10.1016/j.pmpp.2005.09.001

Blaxter, M., & Koutsovoulos, G. (2015). The evolution of parasitism in Nematoda. *Parasitology, 142 Suppl 1*, S26-39. doi:10.1017/S0031182014000791

Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T., & Thomas, W. K. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature, 392*(6671), 71-75. doi:10.1038/32160

Blok, V. C., Pylypenko, L., & Phillips, M. S. (2006). Molecular variation in the potato cyst nematode, *Globodera pallida*, in relation to virulence. *Commun Agric Appl Biol Sci, 71*(3 Pt A), 637-638.

Bohlmann, H. (2015). Introductory Chapter on the Basic Biology of Cyst Nematodes. *Advances in Botanical Research, 73*, 33-59. doi:10.1016/bs.abr.2014.12.001

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Bossis, M., & Mugniéry, D. (1993). Specific status of six *Globodera* parasites of solanaceous plants studied by means of two-dimensional gel electrophoresis with a comparison of gel patterns by a computed system. *Fundamental and Applied Nematology, 16*(1), 47-56.

Boucher, A. C., 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*(1), 75-82. doi:10.1016/j.ympev.2013.05.020

Bustin, S. A., & Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of Biomolecular Techniques, 15*(3), 155-166.

Camire, M. E., Kubow, S., & Donnelly, D. J. (2009). Potatoes and Human Health. *Critical Reviews in Food Science and Nutrition, 49*(10), 823-840. doi:10.1080/10408390903041996

Castro-Quezada, P., Aarrouf, J., Claverie, M., Favery, B., Mugniéry, D., Lefebvre, V., & Caromel, B. (2013). Identification of Reference Genes for Normalizing RNA Expression in Potato Roots Infected with Cyst Nematodes. *Plant Molecular Biology Reporter, 31*(4), 936-945. doi:10.1007/s11105-013-0566-3

Chauvin, L., Caromel, B., Kerlan, M. C., Rulliat, E., Fournet, S., Chauvin, J. É., Grenier, E., Ellissèche, D., & Mugniéry, D. (2008). Control of potato cyst nematodes *Globodera rostochiensis* and *Globodera pallida*. *Cahiers Agricultures, 17*(4), 368-374. doi:10.1684/agr.2008.0217

Chen, C., Cui, L., Chen, Y., Zhang, H., Liu, P., Wu, P., Qiu, D., Zou, J., Yang, D., Yang, L., Liu, H., Zhou, Y., & Li, H. (2017). Transcriptional responses of wheat and the cereal cyst nematode *Heterodera avenae* during their early contact stage. *Scientific Reports, 7*(1), 14471. doi:10.1038/s41598-017-14047-y

Chervoneva, I., Li, Y., Schulz, S., Croker, S., Wilson, C., Waldman, S. A., & Hyslop, T. (2010). Selection of optimal reference genes for normalization in quantitative RT-PCR. *BMC Bioinformatics, 11*, 253. doi:10.1186/1471-2105-11-253

Chevenet, F., Brun, C., Banuls, A. L., Jacq, B., & Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics, 7*, 439. doi:10.1186/1471-2105-7-439

Chronis, D., Chen, S., Lu, S., Hewezi, T., Carpenter, S. C., Loria, R., Baum, T. J., & Wang, X. (2013). A ubiquitin carboxyl extension protein secreted from a plant-parasitic nematode *Globodera rostochiensis* is cleaved in planta to promote plant parasitism. *The Plant Journal, 74*(2), 185-196. doi:10.1111/tpj.12125

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin), 6*(2), 80-92. doi:10.4161/fly.19695

Consortium, T. C. e. S. (1998). Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science, 282*(5396), 2012-2018. doi:10.1126/science.282.5396.2012

Cotton, J. A., Lilley, C. J., Jones, L. M., Kikuchi, T., Reid, A. J., Thorpe, P., Tsai, I. J., Beasley, H., Blok, V., Cock, P. J., Eves-van den Akker, S., Holroyd, N., Hunt, M., Mantelin, S., Naghra, H., Pain, A., Palomares-Rius, J. E., Zarowiecki, M., Berriman, M., Jones, J. T., & Urwin, P. E. (2014). The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology, 15*(3), R43. doi:10.1186/gb-2014-15-3-r43

Coulter, S. N., Schwan, W. R., Ng, E. Y., Langhorne, M. H., Ritchie, H. D., Westbrock-Wadman, S., Hufnagle, W. O., Folger, K. R., Bayer, A. S., & Stover, C. K. (1998). *Staphylococcus aureus* genetic loci impacting growth and survival in multiple infection environments. *Mol Microbiol, 30*(2), 393-404. doi:10.1046/j.1365-2958.1998.01075.x

Cui, J. K., Peng, H., Qiao, F., Wang, G. F., Huang, W. K., Wu, D. Q., & Peng, D. (2018). Characterization of Putative Effectors from the Cereal Cyst Nematode *Heterodera avenae*. *Phytopathology, 108*(2), 264-274. doi:10.1094/PHYTO-07-17-0226-R

Dagan, T., Talmor, Y., & Graur, D. (2002). Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. *Molecular Biology and Evolution, 19*(7), 1022-1025. doi:10.1093/oxfordjournals.molbev.a004161

Dalamu, Bhardwaj, V., Umamaheshwari, R., Sharma, R., Kaushik, S. K., Joseph, T. A., Singh, B. P., & Gebhardt, C. (2012). Potato cyst nematode (PCN) resistance: Genes, genotypes and markers- An update. *Sabrao Journal of Breeding and Genetics, 44*(2), 202-228.

Danchin, E. G., Guzeeva, E. A., Mantelin, S., Berepiki, A., & Jones, J. T. (2016). Horizontal Gene Transfer from Bacteria Has Enabled the Plant-Parasitic Nematode *Globodera pallida* to Feed on Host-Derived Sucrose. *Molecular Biology and Evolution, 33*(6), 1571-1579. doi:10.1093/molbev/msw041

Danchin, E. G., Rosso, M. N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., & Abad, P. (2010a). Multiple lateral gene transfers and duplications have promoted

plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences, 107*(41), 17651-17656. doi:10.1073/pnas.1008486107

Danchin, E. G. J., Rosso, M.-N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., & Abad, P. (2010b). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences, 107*(41), 17651-17656. doi:10.1073/pnas.1008486107

Davidson, N. M., & Oshlack, A. (2014). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol, 15*(7), 410. doi:10.1186/s13059-014-0410-6

Davis, E. L., Hussey, R. S., & Baum, T. J. (2009). Parasitism Genes: What They Reveal about Parasitism. *Plant Cell Monographs, 15*, 15-44. doi:10.1007/978-3-540-85215-5_2

Decraemer, W., & Hunt, D. J. (2006). Structure and classification. In R. N. Perry & M. Moens (Eds.), *Plant nematology* (pp. 3-32). Wallingford: CAB International.

Den Nijs, L., & Karssen, G. (2004). *Globodera rostochiensis* and *Globodera pallida*. *EPPO Bulletin, 34*(2), 309-314. doi:10.1111/j.1365-2338.2004.00734.x

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., & Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research, 36*, W465-469. doi:10.1093/nar/gkn180

Derveaux, S., Vandesompele, J., & Hellemans, J. (2010). How to do successful gene expression analysis using real-time PCR. *Methods, 50*(4), 227-230. doi:10.1016/j.ymeth.2009.11.001

Dheda, K., Huggett, J. F., Chang, J. S., Kim, L. U., Bustin, S. A., Johnson, M. A., Rook, G. A., & Zumla, A. (2005). The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Analytical Biochemistry, 344*(1), 141-143. doi:10.1016/j.ab.2005.05.022

Diaz-Granados, A., Petrescu, A.-J., Goverse, A., & Smant, G. (2016). SPRYSEC Effectors: A Versatile Protein-Binding Platform to Disrupt Plant Innate Immunity. *Frontiers in Plant Science, 7*(1575). doi:10.3389/fpls.2016.01575

Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., Osorio, S., Tohge, T., Fernie, A. R., Feussner, I., Feussner, K., Meinicke, P., Stierhof, Y. D., Schwarz, H., Macek, B., Mann, M., & Kahmann, R. (2011). Metabolic priming by a secreted fungal effector. *Nature, 478*(7369), 395-398. doi:10.1038/nature10454

Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews Genetics, 11*(8), 539-548. doi:10.1038/nrg2812

Doyle, E. A., & Lambert, K. N. (2003). *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. *Molecular Plant-Microbe Interactions, 16*(2), 123-131. doi:10.1094/MPMI.2003.16.2.123

Duceppe, M.-O., Lafond-Lapalme, J., Palomares-Rius, J. E., Sabeh, M., Blok, V., Moffett, P., & Mimee, B. (2017). Analysis of survival and hatching transcriptomes from potato cyst nematodes, *Globodera rostochiensis* and *G. pallida*. *Scientific reports, 7*(1), 3882. doi:10.1038/s41598-017-03871-x

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Ellenby, C. (1952). Resistance to the Potato Root Eelworm, *Heterodera rostochiensis* Wollenweber. *Nature, 170*, 1016. doi:10.1038/1701016a0

Endo, B. Y. (1987). Ultrastructure of Esophageal Gland Secretory Granules in Juveniles of *Heterodera glycines*. *Journal of Nematology, 19*(4), 469-483.

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*. *Molecular Ecology, 26*(18), 4700-4711. doi:10.1111/mec.14240

Espinola, S. M., Ferreira, H. B., & Zaha, A. (2014). Validation of suitable reference genes for expression normalization in *Echinococcus* spp. larval stages. *PLOS ONE, 9*(7), e102228. doi:10.1371/journal.pone.0102228

Eves-van den Akker, S., & Birch, P. R. (2016). Opening the Effector Protein Toolbox for Plant-Parasitic Cyst Nematode Interactions. *Molecular Plant, 9*(11), 1451-1453. doi:10.1016/j.molp.2016.09.008

Eves-van den Akker, S., Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G., Da Rocha, M., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szitenberg, A., Grenier, E., Montarry, J., Mimee, B., Duceppe, M. O., Boyes, I., Marvin, J. M., Jones, L. M., Yusup, H. B., Lafond-Lapalme, J., Esquibet, M., Sabeh, M., Rott, M., Overmars, H., Finkers-Tomczak, A., Smant, G., Koutsovoulos, G., Blok, V., Mantelin, S., Cock, P. J., Phillips, W., Henrissat, B., Urwin, P. E., Blaxter, M., & Jones, J. T. (2016). The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology, 17*(1), 124. doi:10.1186/s13059-016-0985-1

Eves-van den Akker, S., Lilley, C. J., Jones, J. T., & Urwin, P. E. (2014). Identification and characterisation of a hyper-variable apoplastic effector gene family of the potato cyst nematodes. *PLOS Pathogens, 10*(9), e1004391. doi:10.1371/journal.ppat.1004391

Fenwick, D. W. (1949). Investigations on the emergence of larvae from cysts of the potato-root eelworm *Heterodera rostochiensis*; technique and variability. *Journal of Helminthology, 23*(3-4), 157-170. doi:10.1017/s0022149x0003248x

Finkers-Tomczak, A., Danan, S., van Dijk, T., Beyene, A., Bouwman, L., Overmars, H., van Eck, H., Goverse, A., Bakker, J., & Bakker, E. (2009). A high-resolution map of the Grp1 locus on chromosome V of potato harbouring broad-spectrum resistance to the cyst nematode species *Globodera pallida* and *Globodera rostochiensis*. *Theoretical and applied genetics [Theoretische und angewandte Genetik], 119*(1), 165-173. doi:10.1007/s00122-009-1026-1

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*(2), 977-993. doi:10.1534/genetics.108.092221

Food and Agriculture Organization. (2016). *FAO Statistical Pocketbook 2015*: The Food and Agriculture Organization of the United Nations.

Fosu-Nyarko, J., & Jones, M. G. K. (2015). Chapter Fourteen - Application of Biotechnology for Nematode Control in Crop Plants. In E. Carolina & F. Carmen (Eds.), *Advances in Botanical Research* (Vol. Volume 73, pp. 339-376): Academic Press.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint, Available from: arXiv:1207.3907*.

Gebhardt, C., Mugniery, 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*(5), 541-544. doi:10.1007/bf00220911

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., & Bryant, S. H. (2010a). The NCBI BioSystems database. *Nucleic Acids Research, 38*, D492-496. doi:10.1093/nar/gkp858

Geric Stare, B., Fouville, D., Sirca, S., Gallot, A., Urek, G., & Grenier, E. (2011). Molecular variability and evolution of the pectate lyase (pel-2) parasitism gene in cyst nematodes parasitizing different solanaceous plants. *Journal of Molecular Evolution, 72*(2), 169-181. doi:10.1007/s00239-010-9413-4

Geric Stare, B., Lamovšek, J., Širca, S., & Urek, G. (2012). Assessment of sequence variability in putative parasitism factor, expansin (expB2) from diverse populations of potato cyst nematode *Globodera rostochiensis*. *Physiological and Molecular Plant Pathology, 79*, 49-54. doi:10.1016/j.pmpp.2012.04.002

Ghag, S. B. (2017). Host induced gene silencing, an emerging science to engineer crop resistance against harmful plant pathogens. *Physiological and Molecular Plant Pathology, 100*, 242-254. doi:10.1016/j.pmpp.2017.10.003

Goellner, M., Smant, G., De Boer, J. M., Baum, T. J., & Davis, E. L. (2000). Isolation of Beta-1,4-Endoglucanase Genes from *Globodera tabacum* and their Expression During Parasitism. *Journal of Nematology, 32*(2), 154-165.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology, 29*(7), 644-652. doi:10.1038/nbt.1883

Grenier, E., Blok, V. C., Jones, J. T., Fouville, D., & Mugniery, D. (2002). Identification of gene expression differences between *Globodera pallida* and *G.'mexicana'* by suppression subtractive hybridization. *Molecular Plant Pathology, 3*(4), 217-226. doi:10.1046/j.1364-3703.2002.00111.x

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol, 59*(3), 307-321. doi:10.1093/sysbio/syq010

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics, 29*(8), 1072-1075. doi:10.1093/bioinformatics/btt086

Haegeman, A., Jones, J. T., & Danchin, E. G. (2011). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Molecular Plant-Microbe Interactions, 24*(8), 879-887. doi:10.1094/MPMI-03-11-0055

Haegeman, A., Mantelin, S., Jones, J. T., & Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene, 492*(1), 19-31. doi:10.1016/j.gene.2011.10.040

Hardcastle, T. J., & Kelly, K. A. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics, 11*(1), 422. doi:10.1186/1471-2105-11-422

Hofmann, J., & Grundler, F. M. W. (2007). Identification of reference genes for qRT-PCR studies of gene expression in giant cells and syncytia induced in *Arabidopsis thaliana* by *Meloidogyne incognita* and *Heterodera schachtii*. *Nematology, 9*(3), 317-323. doi:10.1163/156854107781352034

Hoogewijs, D., Houthoofd, K., Matthijssens, F., Vandesompele, J., & Vanfleteren, J. R. (2008). Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in *C. elegans*. *BMC Molecular Biology, 9*, 9. doi:10.1186/1471-2199-9-9

Hoysted, G. A., Bell, C. A., Lilley, C. J., & Urwin, P. E. (2018). Aphid Colonization Affects Potato Root Exudate Composition and the Hatching of a Soil Borne Pathogen. *Frontiers in Plant Science, 9*(1278). doi:10.3389/fpls.2018.01278

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research, 9*(9), 868-877. doi:10.1101/gr.9.9.868

Huggett, J., Dheda, K., Bustin, S., & Zumla, A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes and immunity, 6*(4), 279-284. doi:10.1038/sj.gene.6364190

Jeffreys, H. (1998). *The Theory of Probability*: OUP Oxford.

Jones, J. T., Haegeman, A., Danchin, E. G., Gaur, H. S., Helder, J., Jones, M. G., Kikuchi, T., Manzanilla-Lopez, R., Palomares-Rius, J. E., Wesemael, W. M., & Perry, R. N. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology, 14*(9), 946-961. doi:10.1111/mpp.12057

Jones, J. T., Kumar, A., Pylypenko, L. A., Thirugnanasambandam, A., Castelli, L., Chapman, S., Cock, P. J., Grenier, E., Lilley, C. J., Phillips, M. S., & Blok, V. C. (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-Microbe Interactions, 10*(6), 815-828. doi:10.1111/j.1364-3703.2009.00585.x

Jones, J. T., Reavy, B., Smant, G., & Prior, A. E. (2004). Glutathione peroxidases of the potato cyst nematode *Globodera rostochiensis*. *Gene, 324*, 47-54. doi:10.1016/j.gene.2003.09.051

Kall, L., Krogh, A., & Sonnhammer, E. L. (2004a). A combined transmembrane topology and signal peptide prediction method. *J Mol Biol, 338*(5), 1027-1036. doi:10.1016/j.jmb.2004.03.016

Kall, L., Krogh, A., & Sonnhammer, E. L. (2004b). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology, 338*(5), 1027-1036. doi:10.1016/j.jmb.2004.03.016

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research, 28*(1), 27-30. doi:10.1093/nar/28.1.27

Kohler, S., Foulongne, V., Ouahrani-Bettache, S., Bourg, G., Teyssier, J., Ramuz, M., & Liautard, J. P. (2002). The analysis of the intramacrophagic virulome of *Brucella suis* deciphers the environment encountered by the pathogen inside the macrophage host cell. *Proceedings of the National Academy of Sciences, 99*(24), 15711-15716. doi:10.1073/pnas.232454299

Konczal, M., Koteja, P., Stuglik, M. T., Radwan, J., & Babik, W. (2014). Accuracy of allele frequency estimation using pooled RNA-Seq. *Molecular Ecology Resources, 14*(2), 381-392. doi:10.1111/1755-0998.12186

Koritsas, V. M., & Atkinson, H. J. (1994). Proteinases of females of the phytoparasite *Globodera pallida* (potato cyst nematode). *Parasitology, 109*(3), 357-365. doi:10.1017/S0031182000078392

Kort, J., Stone, A. R., Rumpenhorst, H. J., & Ross, H. (1977). An International Scheme for Identifying and Classifying Pathotypes of Potato Cyst-Nematodes *Globodera rostochiensis* and *G. Pallida*. *Nematologica, 23*(3), 333-339. doi:10.1163/187529277x00057

Kovaleva, E. S., Masler, E. P., Skantar, A. M., & Chitwood, D. J. (2004). Novel matrix metalloproteinase from the cyst nematodes *Heterodera glycines* and *Globodera rostochiensis*. *Mol Biochem Parasitol, 136*(1), 109-112. doi:10.1016/j.molbiopara.2004.03.001

Kozera, B., & Rapacz, M. (2013). Reference genes in real-time PCR. *Journal of applied genetics, 54*(4), 391-406. doi:10.1007/s13353-013-0173-x

Laetsch, D., & Blaxter, M. (2017). BlobTools: Interrogation of genome assemblies [version 1; referees: 2 approved with reservations]. *F1000Research, 6*(1287). doi:10.12688/f1000research.12232.1

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27*(21), 2987-2993. doi:10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J. F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., & Vidal, M. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science, 303*(5657), 540-543. doi:10.1126/science.1091403

Li, Y., Wang, K., Xie, H., Wang, D.-W., Xu, C.-L., Huang, X., Wu, W.-J., & Li, D.-L. (2015). Cathepsin B Cysteine Proteinase is Essential for the Development and Pathogenesis of the Plant Parasitic Nematode *Radopholus similis*. *International Journal of Biological Sciences, 11*(9), 1073-1087. doi:10.7150/ijbs.12065

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods, 25*(4), 402-408. doi:10.1006/meth.2001.1262

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Madani, M., Ward, L. J., & De Boer, S. H. (2008). Multiplex real-time polymerase chain reaction for identifying potato cyst nematodes, *Globodera pallida* and *Globodera rostochiensis*, and the tobacco cyst nematode, *Globodera tabacum*. *Canadian Journal of Plant Pathology, 30*(4), 554-564. doi:10.1080/07060660809507555

Mai, W. F., & Peterson, L. C. (1952). Resistance of *Solanum Ballsii* and *Solanum sucrense* to the Golden Nematode, *Heterodera rostochiensis*, Wollenweber. *Science, 116*(3009), 224-225. doi:10.1126/science.116.3009.224

Mannervik, B. (1987). The enzymes of glutathione metabolism: an overview. *Biochemical Society Transactions, 15*(4), 717-718. doi:10.1042/bst0150717

Mantelin, S., Thorpe, P., & Jones, J. T. (2017). Translational biology of nematode effectors. Or, to put it another way, functional analysis of effectors – what's the point? *Nematology, 19*(3), 251-261. doi:10.1163/15685411-00003048

Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Research, 32*, W327-331. doi:10.1093/nar/gkh454

Mei, Y., Wright, K. M., Haegeman, A., Bauters, L., Diaz-Granados, A., Goverse, A., Gheysen, G., Jones, J. T., & Mantelin, S. (2018). The *Globodera pallida* SPRYSEC Effector GpSPRY-414-2 That Suppresses Plant Defenses Targets a Regulatory Component of the Dynamic Microtubule Network. *Frontiers in Plant Science, 9*, 1019-1019. doi:10.3389/fpls.2018.01019

Mimee, B., Andersen, R., Bélair, G., Vanasse, A., & Rott, M. (2014). Impact of quarantine procedures on weed biodiversity and abundance: Implications for the management of the golden potato cyst nematode, *Globodera rostochiensis*. *Crop Protection, 55*, 21-27. doi:10.1016/j.cropro.2013.10.003

Mimee, B., Dauphinais, N., & Belair, G. (2015a). Life Cycle of the Golden Cyst Nematode, *Globodera rostochiensis*, in Quebec, Canada. *Journal of Nematology, 47*(4), 290-295.

Mimee, B., Duceppe, M.-O., Véronneau, P.-Y., Lafond-Lapalme, J., Jean, M., Belzile, F., & Bélair, G. (2015b). A new method for studying population genetics of cyst nematodes based on Pool-Seq and genomewide allele frequency analysis. *Molecular Ecology Resources*, n/a-n/a. doi:10.1111/1755-0998.12412

Mitchum, M. G., Hussey, R. S., Baum, T. J., Wang, X., Elling, A. A., Wubben, M., & Davis, E. L. (2013). Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist, 199*(4), 879-894. doi:10.1111/nph.12323

Muoz, C., San Francisco, J., Gutirrez, B., & Gonzlez, J. (2015). Role of the Ubiquitin-Proteasome Systems in the Biology and Virulence of Protozoan Parasites. *BioMed Research International, 2015*, 13. doi:10.1155/2015/141526

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science, 320*(5881), 1344-1349. doi:10.1126/science.1158441

Nicol, J. M., Turner, S. J., Coyne, D. L., Nijs, L. d., Hockland, S., & Maafi, Z. T. (2011). Current Nematode Threats to World Agriculture. In J. Jones, G. Gheysen, & C. Fenoll (Eds.), *Genomics and Molecular Genetics of Plant-Nematode Interactions* (pp. 21-43). Dordrecht: Springer Netherlands.

Nicot, N., Hausman, J. F., Hoffmann, L., & Evers, D. (2005). Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany, 56*(421), 2907-2914. doi:10.1093/jxb/eri285

Nijboer, H., & Parlevliet, J. E. (1990). Pathotype-specificity in potato cyst nematodes, a reconsideration. *Euphytica, 49*(1), 39-47. doi:10.1007/bf00024129

Nolan, T., Hands, R. E., & Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature Protocols, 1*(3), 1559-1582. doi:10.1038/nprot.2006.236

Oerke, E. C. (2006). Crop losses to pests. *Journal of Agricultural Science, 144*(1), 31-43. doi:10.1017/S0021859605005708

Olsen, O. A., & Mulvey, R. H. (1962). The discovery of golden nematode in Newfoundland. *Canadian Plant Disease Survey, 42*, 253.

Orchard, W. R. (1965). Occurence of the golden nematode on Vancouver Island, British Columbia. *Canadian Plant Disease Survey, 45*, 3.

Paganini, J., Campan-Fournier, A., Da Rocha, M., Gouret, P., Pontarotti, P., Wajnberg, E., Abad, P., & Danchin, E. G. (2012). Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes. *PLOS ONE, 7*(11), e50875. doi:10.1371/journal.pone.0050875

Palomares-Rius, J. E., Hedley, P., Cock, P. J., Morris, J. A., Jones, J. T., & Blok, V. C. (2016). Gene expression changes in diapause or quiescent potato cyst nematode, *Globodera pallida*, eggs after hydration or exposure to tomato root diffusate. *PeerJ, 4*, e1654. doi:10.7717/peerj.1654

Pandey, S. K., Singh, S. V., & Sarkar, D. (2005). Potato (*Solanum tuberosum*) for sustaining food and nutrition security in developing world. *Indian Journal of Agricultural Sciences, 75*(1), 3-18.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods, 8*(10), 785-786. doi:10.1038/nmeth.1701

Pfaffl, M. W., Tichopad, A., Prgomet, C., & Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnology Letters, 26*(6), 509-515. doi:10.1023/b:bile.0000019559.84305.47

Pineda, O., Bonierbale, M. W., Plaisted, R. L., Brodie, B. B., & Tanksley, S. D. (1993). Identification of RFLP markers linked to the H1 gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Genome, 36*(1), 152-156. doi:10.1139/g93-019

Piskol, R., Ramaswami, G., & Li, J. B. (2013). Reliable identification of genomic variants from RNA-seq data. *The American Journal of Human Genetics, 93*(4), 641-651. doi:10.1016/j.ajhg.2013.08.008

Poppe, S., Dorsheimer, L., Happel, P., & Stukenbrock, E. H. (2015). Rapidly Evolving Genes Are Key Players in Host Specialization and Virulence of the Fungal Wheat Pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). *PLOS Pathogens, 11*(7), e1005055. doi:10.1371/journal.ppat.1005055

Quist, C. W., Smant, G., & Helder, J. (2015). Evolution of plant parasitism in the phylum Nematoda. *Annual Review of Phytopathology, 53*, 289-310. doi:10.1146/annurev-phyto-080614-120057

Radonić, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., & Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications, 313*(4), 856-862. doi:10.1016/j.bbrc.2003.11.177

Rice, S. L., Leadbeater, B. S. C., & Stone, A. R. (1985). Changes in cell structure in roots of resistant potatoes parasitized by potato cyst-nematodes. I. Potatoes with resistance gene H1 derived from *Solanum tuberosum* ssp. andigena. *Physiological Plant Pathology, 27*(2), 219-234. doi:10.1016/0048-4059(85)90069-4

Sabaratnam, S. (2012). Patato cyst nematode. Retrieved from http://www.agf.gov.bc.ca/cropprot/goldennema.htm

Sabeh, M., Duceppe, M. O., St-Arnaud, M., & Mimee, B. (2018). Transcriptome-wide selection of a reliable set of reference genes for gene expression studies in potato cyst

nematodes (*Globodera* spp.). *PLOS ONE, 13*(3), e0193840. doi:10.1371/journal.pone.0193840

Sacco, M. A., Koropacka, K., Grenier, E., Jaubert, M. J., 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*(8), e1000564. doi:10.1371/journal.ppat.1000564

Schmidt, S. M., Houterman, P. M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., Boeren, S., Takken, F. L. W., & Rep, M. (2013). MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics, 14*(1), 119. doi:10.1186/1471-2164-14-119

Scholl, E. H., Thorne, J. L., McCarter, J. P., & Bird, D. M. (2003). Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. *Genome Biology, 4*(6), R39. doi:10.1186/gb-2003-4-6-r39

Schwekendiek, A., Maier, T., Womack, C., Byrne, D., Boer, J. d., Davis, E., & Baum, T. (1999). Initial characterisation of endochitinase genes of the plant-parasitic nematode *Heterodera glycines*. *Journal of Nematology, 31*, 568.

Silver, N., Best, S., Jiang, J., & Thein, S. L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology, 7*, 33. doi:10.1186/1471-2199-7-33

Smant, G., Helder, J., & Goverse, A. (2018). Parallel adaptations and common host cell responses enabling feeding of obligate and facultative plant parasitic nematodes. *The Plant Journal, 93*(4), 686-702. doi:10.1111/tpj.13811

Sobczak, M., & Golinowski, W. (2011). Cyst Nematodes and Syncytia. In J. Jones, G. Gheysen, & C. Fenoll (Eds.), *Genomics and Molecular Genetics of Plant-Nematode Interactions* (pp. 61-82). Dordrecht: Springer Netherlands.

Spanier, B., Stürzenbaum, S. R., Holden-Dye, L. M., & Baumeister, R. (2005). *Caenorhabditis elegans* Neprilysin NEP-1: an Effector of Locomotion and Pharyngeal Pumping. *Journal of Molecular Biology, 352*(2), 429-437. doi:10.1016/j.jmb.2005.06.063

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics, 24*(5), 637-644. doi:10.1093/bioinformatics/btn013

Strunnikov, A. V., & Jessberger, R. (1999). Structural maintenance of chromosomes (SMC) proteins: conserved molecular properties for multiple biological functions. *uropean Journal of Biochemistry, 263*(1), 6-13. doi:10.1046/j.1432-1327.1999.00509.x

Taki, F. A., & Zhang, B. (2013). Determination of reliable reference genes for multi-generational gene expression analysis on *C. elegans* exposed to abused drug nicotine. *Psychopharmacology, 230*(1), 77-88. doi:10.1007/s00213-013-3139-0

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol, 56*(4), 564-577. doi:10.1080/10635150701472164

The UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic acids research, 45*(D1), D158-D169. doi:10.1093/nar/gkw1099

Thiéry, M., Mugniéry, D., Bossrs, M., & Sosa-Moss, C. (1997). Résultats de croisements entre *Globodera pallida Stone* et *G. "mexicana"* Canlpos-Vela : héritabilité du développement sur pomme de terre et notion d'espèce. *Fundamental and Applied Nematology, 20*(6), 551-556.

Thomas, M. S., & Wigneshweraraj, S. (2015). Regulation of virulence gene expression. *Virulence, 5*(8), 832-834. doi:10.1080/21505594.2014.995573

Thorpe, P. (2012). *Bioinformatic and functional characterisation of Globodera pallida effector genes.* (Doctor of Philosophy Ph.D. Thesis), The University of Leeds,

Thorpe, P., Mantelin, S., Cock, P. J., Blok, V. C., Coke, M. C., Eves-van den Akker, S., Guzeeva, E., Lilley, C. J., Smant, G., Reid, A. J., Wright, K. M., Urwin, P. E., & Jones, J. T. (2014). Genomic characterisation of the effector complement of the potato cyst nematode *Globodera pallida*. *BMC Genomics, 15*, 923. doi:10.1186/1471-2164-15-923

Turner, S. J. (1996). Population decline of potato cyst nematodes (*Globodera rostochiensis, G. pallida*) in field soils in Northern Ireland. *Annals of Applied Biology, 129*(2), 315-322. doi:10.1111/j.1744-7348.1996.tb05754.x

Turner, S. J., & Rowe, J. A. (2006). *Cyst nematodes*. Wallingford, Oxfordshire: CAB International.

USDA. (2008). Golden Nematode Ro2 Eradication in Livingston and Suffolk Counties, New York. *Environmental Assessment*.

van Megen, H., Holovachov, O., Bongers, T., Bakker, J., Helder, J., van den Elsen, S., Holterman, M., Karssen, G., & Mooyman, P. (2009). A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology, 11*(6), 927-950. doi:10.1163/156854109x456862

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology, 3*(7), Research0034. doi:10.1186/gb-2002-3-7-research0034

Varden, F. A., De la Concepcion, J. C., Maidment, J. H., & Banfield, M. J. (2017). Taking the stage: effectors in the spotlight. *Current Opinion in Plant Biology, 38*, 25-33. doi:10.1016/j.pbi.2017.04.013

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics, 10*(1), 57-63. doi:10.1038/nrg2484

Xie, F., Xiao, P., Chen, D., Xu, L., & Zhang, B. (2012). miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant molecular biology, 80*(1), 75-84. doi:10.1007/s11103-012-9885-2

Yadav, B. C., Veluthambi, K., & Subramaniam, K. (2006). Host-generated double stranded RNA induces RNAi in plant-parasitic nematodes and protects the host from infection. *Molecular and Biochemical Parasitology, 148*(2), 219-222. doi:10.1016/j.molbiopara.2006.03.013

Yang, Y., & Luo, D. (2013). The origin of parasitism gene in nematodes: evolutionary analysis through the construction of domain trees. *Evolutionary Bioinformatics, 9*, 453-466. doi:10.4137/EBO.S13032

Yatsuda, A. P., Bakker, N., Krijgsveld, J., Knox, D. P., Heck, A. J. R., & de Vries, E. (2006). Identification of Secreted Cysteine Proteases from the Parasitic Nematode *Haemonchus contortus* Detected by Biotinylated Inhibitors. *Infection and Immunity, 74*(3), 1989-1993. doi:10.1128/IAI.74.3.1989-1993.2006

Yin, C., & Hulbert, S. (2015). Host induced gene silencing (HIGS), a promising strategy for developing disease resistant crops. *Gene Technology, 4*, 130. doi:10.4172/2329-6682.1000130

Yu, Q., Ye, W., Sun, F., & Miller, S. (2010). Characterization of *Globodera rostochiensis* (Tylenchida: Heteroderidae) associated with potato in Quebec, Canada. *Canadian Journal of Plant Pathology, 32*(2), 264-271. doi:10.1080/07060661003740322

Zhang, Y., Chen, D., Smith, M. A., Zhang, B., & Pan, X. (2012). Selection of reliable reference genes in *Caenorhabditis elegans* for analysis of nanotoxicity. *PLOS ONE, 7*(3), e31849. doi:10.1371/journal.pone.0031849

Zhao, S., & Fernald, R. D. (2005). Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of computational biology, 12*(8), 1047-1064. doi:10.1089/cmb.2005.12.1047

Zhou, Z., Cong, P., Tian, Y., & Zhu, Y. (2017). Using RNA-seq data to select reference genes for normalizing gene expression in apple roots. *PLoS One, 12*(9), e0185288. doi:10.1371/journal.pone.0185288

Annexe 1

The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence

Sebastian Eves-van den Akker[1], Dominik R. Laetsch[2], Peter Thorpe[3], Catherine J. Lilley[4], Etienne G. J. Danchin[5], Martine Da Rocha[5], Corinne Rancurel[5], Nancy E. Holroyd[6], James A. Cotton[6], Amir Szitenberg[7], Eric Grenier[8], Josselin Montarry[8], Benjamin Mimee[9], Marc-Olivier Duceppe[9], Ian Boyes[10], Jessica M. C. Marvin[4], Laura M. Jones[4], Hazijah B. Yusup[4], Joël Lafond-Lapalme[9], Magali Esquibet[8], **Michael Sabeh**[9], […] Peter E. Urwin[4], Mark Blaxter[2] and John T. Jones[3,16]

[1] Division of Plant Sciences, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK. [2] Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, UK. [3] Cell and Molecular Sciences Group, Dundee Effector Consortium, James Hutton Institute, Dundee DD2 5DA, UK. [4] Centre for Plant Sciences, University of Leeds, Leeds LS2 9JT, UK. [5] INRA, University Nice Sophia Antipolis, CNRS, UMR 1355-7254 Institut Sophia Agrobiotech, 06900 Sophia Antipolis, France. [6] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK. [7] School of Biological, Biomedical and Environmental Sciences, University of Hull, Hull HU6 7RX, UK. [8] INRA, UMR1349 IGEPP (Institute for Genetics, Environment and Plant Protection), 35653 Le Rheu, France. [9] Agriculture and Agri-food Canada, Horticulture Research and Development Centre, 430 Bboul. Gouin, St-Jean-sur-Richelieu, Québec J3B 3E6, Canada. [10]Sidney Laboratory, Canadian Food Inspection Agency (CFIA), 8801 East Saanich Rd, Sidney, BC V8L 1H3, Canada. […] [16]School of Biology, University of St Andrews, North Haugh, St Andrews KY16 9TZ, UK.

Cet article a été publié en juin 2016 dans la revue *Genome Biology* (17:124)

Abstract

Background: The yellow potato cyst nematode, *Globodera rostochiensis*, is a devastating plant pathogen of global economic importance. This biotrophic parasite secretes effectors from

pharyngeal glands, some of which were acquired by horizontal gene transfer, to manipulate host processes and promote parasitism. *G. rostochiensis* is classified into pathotypes with different plant resistance-breaking phenotypes.

Results: We generate a high quality genome assembly for *G. rostochiensis* pathotype Ro1, identify putative effectors and horizontal gene transfer events, map gene expression through the life cycle focusing on key parasitic transitions and sequence the genomes of eight populations including four additional pathotypes to identify variation. Horizontal gene transfer contributes 3.5 % of the predicted genes, of which approximately 8.5 % are deployed as effectors. Over one-third of all effector genes are clustered in 21 putative 'effector islands' in the genome. We identify a dorsal gland promoter element motif (termed DOG Box) present upstream in representatives from 26 out of 28 dorsal gland effector families, and predict a putative effector superset associated with this motif. We validate gland cell expression in two novel genes by in situ hybridisation and catalogue dorsal gland promoter element-containing effectors from available cyst nematode genomes. Comparison of effector diversity between pathotypes highlights correlation with plant resistance-breaking.

Conclusions: These *G. rostochiensis* genome resources will facilitate major advances in understanding nematode plant-parasitism. Dorsal gland promoter element-containing effectors are at the front line of the evolutionary arms race between plant and parasite and the ability to predict gland cell expression a priori promises rapid advances in understanding their roles and mechanisms of action.

Disponible sous : https://doi.org/10.1186/s13059-016-0985-1

Annexe 2

Analysis of survival and hatching transcriptomes from potato cyst nematodes, *Globodera rostochiensis* and *G. pallida*

Marc-Olivier Duceppe[1,4], Joël Lafond-Lapalme[1,2], Juan Emilio Palomares-Rius[3,5], **Michaël Sabeh**[1], Vivian Blok[3], Peter Moffett[2] & Benjamin Mimee[1]

[1]Agriculture and Agri-Food Canada, 430, Boulevard Gouin Saint-Jean-sur-Richelieu (Québec), J3B 3E6, Québec, Canada. [2] Département de Biologie, Université de Sherbrooke, Sherbrooke, J1K 2R1, Canada. [3] Cell and Molecular Sciences, The James Hutton Institute, Invergowrie, Dundee, DD2 5DA, United Kingdom. [4] Present address: Canadian Food Inspection Agency, Ottawa Laboratory Fallowfield (OLF), 3851 Fallowfield Road, Ottawa, Ontario, K2H 8P9, Canada. [5]Present address: Institute for Sustainable Agriculture (IAS), Spanish National Research Council (CSIC), Avenida Menéndez Pidal s/n, 14004 Córdoba, Campus de Excelencia Internacional Agroalimentario, ceiA3, Spain.

Cet article a été publié en juin 2017 dans la revue *Scientific Reports* (7:3882)

Abstract

Potato cyst nematodes (PCNs), *Globodera rostochiensis* and *G. pallida*, cause important economic losses. They are hard to manage because of their ability to remain dormant in soil for many years. Although general knowledge about these plant parasitic nematodes has considerably increased over the past decades, very little is known about molecular events involved in cyst dormancy and hatching, two key steps of their development. Here, we have studied the progression of PCN transcriptomes from dry cysts to hatched juveniles using RNA-Seq. We found that several cell detoxification-related genes were highly active in the dry cysts. Many genes linked to an increase of calcium and water uptake were up-regulated during transition from dormancy to hydration. Exposure of hydrated cysts to host plant root exudates

resulted in different transcriptional response between species. After 48 h of exposure, *G. pallida* cysts showed no significant modulation of gene expression while *G. rostochiensis* had 278 differentially expressed genes. The first *G. rostochiensis* significantly up-regulated gene was observed after 8 h and was coding for a transmembrane metalloprotease. This enzyme is able to activate/inactivate peptide hormones and could be involved in a cascade of events leading to hatching. Several known effector genes were also up-regulated during hatching.

Disponible sous : https://www.nature.com/articles/s41598-017-03871-x