Genome scans on experimentally evolved populations reveal candidate regions for adaptation to plant resistance in the potato cyst nematode Globodera pallida.

Delphine Eoche-Bosy, Matthieu Gautier, Magali Esquibet, Fabrice Legeai, Anthony Bretaudeau, Olivier Bouchez, Sylvain Fournet, Eric Grenier, Josselin Montarry

1 **Genome scans on experimentally evolved populations reveal candidate**

2 **regions for adaptation to plant resistance in the potato cyst nematode**

3 *Globodera pallida*

4

5 D. EOCHE-BOSY[1], M. GAUTIER[2,3], M. ESQUIBET[1], F. LEGEAI[4,5], A. BRETAUDEAU[4,6],

6 O. BOUCHEZ[7,8], S. FOURNET[1], E. GRENIER[1] and J. MONTARRY[1]

7

8 [1]IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, 35650, Le Rheu, France

9 [2]CBGP, INRA, IRD, CIRAD, Montpellier SupAgro, 34988, Montferrier-sur-Lez, France

10 [3]IBC, 34095, Montpellier, France

11 [4]IGEPP, BIPAA, INRA, Agrocampus Ouest, Université de Rennes 1, 35042, Rennes, France

12 [5]IRISA, GenScale, INRIA, 35042, Rennes, France

13 [6]IRISA, GenOuest COre Facility, INRIA, 35042, Rennes, France

14 [7]GeT-PlaGe, Genotoul, INRA, 31326, Castanet-Tolosan, France

15 [8]GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326, Castanet-Tolosan, France

16

19

20 Correspondence: Josselin Montarry, IGEPP, INRA, Agrocampus Ouest, Université de Rennes

21 1, 35650, Le Rheu, France; Fax: +33 2 23 48 51 50; E-mail: josselin.montarry@inra.fr

22

23 Running title: Molecular signatures of adaptation to resistance

Abstract

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

Introduction

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

Plant-parasitic nematodes are major agricultural pathogens causing severe damages in crops worldwide (Nicol *et al.* 2011), but sources of resistance against them remain scarce. Several studies have been conducted to optimize the management of these resistances (Djian-Caporalino *et al.* 2014; Barbary *et al.* 2015). Cyst nematodes are among the most

3

69 economically damaging plant-parasitic nematodes. Over the last few years, increased

70 knowledge about characteristics of cyst nematode populations allowed to better predict the

71 global evolutionary potential of these populations. In particular it was showed that cyst

72 nematode populations represent a real threat to the durability of plant resistances due to an

73 important passive dispersion (Picard *et al.* 2004; Plantard & Porte 2004; Alenda *et al.* 2014)

74 and favored expression of recessives virulence genes due to inbreeding (Montarry *et al.*

75 2015). The adaptation speed of nematode populations would also be modulated by the number

76 and the nature of genetic mutations required to overcome the plant resistance, and by the

77 presence and, if any, the frequency, of virulent individuals in fields.

78   The cyst nematode *Globodera pallida* (Stone) is a major pest of potato crop (Oerke *et*

79 *al.* 1994; van Riel & Mulder 1998) and *Solanum vernei*, a wild potato species, is an

80 interesting source of resistance against *G. pallida* as it is highly effective. The resistance,

81 which leads to the development of most nematodes into adult males, is explained at 61% by

82 the major QTL *GpaV$_{vrn}$* mapped on the potato chromosome V (Rouppe van der Voort *et al.*

83 2000). This source of resistance is to date the only one exploited in commercial potato

84 cultivars at the European level. The resistant cultivar Iledher, carrying *GpaV$_{vrn}$*, has been

85 registered in 2009 in the French catalogue as the first cultivar showing a high level of

86 resistance to *G. pallida*. However, Fournet *et al.* (2013) highlighted that *G. pallida*

87 populations were able to completely overcome the resistance of Iledher in only few

88 generations in experimental evolution. Identifying the determinants of virulence thus appears

89 to be crucial to better anticipate the speed of adaptation of *G. pallida* populations.

90   In organisms with relatively small genomes, such as RNA viruses with 10-15 kb

91 genomes, it is possible since several years to sequence the entire genome and to directly

92 compare sequences of virulent and avirulent strains in order to find the mutation responsible

93   for resistance adaptation (Meshi *et al.* 1988; Díaz *et al.* 2004; Ayme *et al.* 2006; Mardis 2008;

94   Janzac *et al.* 2010). However, this approach cannot be considered in eukaryotic organisms

95   whose genomes range from one to hundreds of thousands megabases. Genome scans

96   represent an efficient alternative to target genomic regions involved in adaptation in such

97   organisms, as they analyze genome-wide variations at the light of theoretical predictions

98   about the effects of selection, in order to detect locus specific signatures of positive

99   directional selection (Luikart *et al.* 2003; Storz 2005). Despite their strong potential to

100   elucidate the genetic bases of adaptation, genome scans have never been used to identify the

101   determinants of virulence in nematodes. However, Bekal *et al.* (2015) recently opened the

102   way for population genomic approach in cyst nematodes. They performed a whole genome

103   allelic imbalance analysis of SNP in *Heterodera glycines* inbred lines grown in the laboratory

104   for over 30 generations on resistant and susceptible soybean plants. This study revealed two

105   new candidate virulence genes: *HgBioB*, a gene encoded biotin synthase and *HgSLP-1*, a gene

106   that appears to have entered *H. glycines* genome via horizontal gene transfer and which

107   encode a protein containing a putative SNARE domain. Virulence is based on sequence

108   polymorphisms for the first gene and on a reduced copy number for the second gene. In

109   potato cyst nematodes, (a)virulence genes known to date were identified only through

110   candidate gene approaches. In *Globodera rostochiensis*, the venom allergen Gr-VAP1 was

111   identified as an avirulence gene product triggering a cell death response in tomato

112   (*Lycopersicon esculentum*) plants containing the *Cf-2* and *Rcr3pim* resistant genes (Lozano-

113   Torres *et al.* 2012, 2014). In *G. pallida*, only one avirulence gene, *Gp-RBP-1,* coding for a

114   SPRYSEC which interacts with the GPA2 resistant protein in potato, has been described

115   (Sacco *et al.* 2009). Virulence is due to a single amino-acid polymorphism, however this

116   mutation is widely distributed in European populations (Carpentier *et al.* 2012) and therefore

5

117   potato carrying *Gpa2* resistance gene is of very limited interest for control of *G. pallida*. On

118   the contrary, the resistant cultivar Iledher shows a high level of resistance to a wide range of

119   European *G. pallida* populations.

120   The aim of the present study was to investigate the *G. pallida* genomic regions

121   involved in the breakdown of the resistance of potato cultivar Iledher. By analyzing

122   microsatellite data, Eoche-Bosy *et al.* (2016) already showed that a genome scan on the *G.*

123   *pallida* lineages coming from the experimental evolution performed by Fournet *et al.* (2013)

124   was feasible. Here, we took advantage of the same biological material to perform a high

125   density genome scan using data from a whole genome resequencing of pools of individuals

126   (Pool-Seq, Futschik & Schlötterer 2010; Zhu *et al.* 2012; Ferretti *et al.* 2013; Gautier *et al.*

127   2013; Schlötterer *et al.* 2014) and a recent model testing for adaptive differentiation and

128   association to population-specific covariables (BayPass, Gautier 2015).

129

130   **Materials and methods**

131   *Study system:* Globodera pallida

132   *Globodera pallida* is a gonochoristic diploid organism with obligate sexual reproduction,

133   which achieves one generation per year in European climatic conditions (Jones 1950). This

134   obligate parasite enters the plant roots as second-stage juveniles (J2) and establishes a

135   specialized feeding structure, the syncytium (Jones & Northcote 1972), which is a severe

136   nutrient sink for the plant. In this species, sex is environmentally determined and depends on

137   the size and efficiency of the syncytium (Sobczak & Golinowski 2011). Adult males leave the

138   root to mate females, which can be fertilized by several males (Green *et al.* 1970;

139   Triantaphyllou & Esbenshade 1990). After mating, the females continue to feed from the

140 syncytium and when eggs development is completed, they die and form a cyst, enclosing

141 hundreds of eggs, which can stay viable for several years in soils.

142 The genome of *G. pallida* has been recently sequenced (Gpal.v1.0, Cotton *et al.* 2014).

143 It is available as an assembly of 124.7 Mb in 6,873 scaffolds, with a N50 of 122 kb and a GC

144 content of 36.7% (Cotton *et al.* 2014). Combining transcriptomic data with manual curation, a

145 total of 16,419 genes were predicted. The genome of *G. rostochiensis* has also been recently

146 sequenced (nGr.v1.0, Eves-van den Akker *et al.* 2016) and could be a more accurate

147 representation of a *Globodera* genome, as suggested by its higher completeness and low level

148 of gene duplication. This genome is available as an assembly of 95.9 Mb in 4,377 scaffolds,

149 with a N50 of 88 kb and a GC content of 38.1%. Annotation of *G. rostochiensis* genome

150 could also be of better quality as a manual annotation phase followed the initial phase of

151 automated annotation, resulting in the prediction of 14,378 genes.

152

153 *Selection of virulent and avirulent* G. pallida *lineages*

154 This study relies on the biological material coming from the experimental evolution

155 performed by Fournet *et al.* (2013). Briefly, nematode lineages were established from cysts of

156 two French natural *G. pallida* populations, SM (near Saint-Malo, Brittany, north-western

157 France) and N (from the island of Noirmoutier, western France) coming from infested fields

158 (Fig. 1). The lineages used here were obtained by rearing both populations during eight

159 successive cycles (*i.e.* eight generations) on the susceptible potato cultivar Désirée (D) and on

160 the resistant cultivar Iledher (I). The present study was conducted with the eighth generation

161 of the four lineages: two lineages adapted to the resistance of Iledher (named hereafter SMI

162 and NI), and two lineages which remained unadapted (named hereafter SMD and ND).

163

Pool sequencing

165   As the amount of DNA in a single *G. pallida* individual is very low, pooling individuals from

166   a same lineage was an efficient way to increase the amount of DNA while allowing the

167   accurate estimation of population allele frequencies (e.g., Gautier *et al.* 2013). To accurately

168   represent the genetic variability of each lineage, we chose to sample individuals from

169   different cysts, rather than different individuals from a same cyst, to constitute the pools. We

170   therefore sampled 300 cysts in each lineage, which was the maximum available, and crushed

171   them individually in sterile water. Two J2 (siblings) were sampled from each cyst and

172   attributed to two different pools, resulting in two pools of 300 individuals for each lineage,

173   *i.e.* two biological replicates. Water in samples was then vacuum-evaporated in a heated

174   Speed Vac Concentrator (MiVac, Genevac Ltd., Ipswich, UK). As the cuticle of the nematode

175   could prevent the lysis of tissues by the proteinase K, samples were stored at -80°C during

176   two hours and then rapidly heated at ambient temperature, just before DNA extraction. DNA

177   was extracted directly from each pool using Qiagen DNeasy Blood and Tissue Kit (Qiagen,

178   Hilden, Germany) following the manufacturer's instructions. A pooled DNA extraction has

179   been favored over an individual DNA extraction following by a pooling of the DNA in

180   equimolar proportions because this last strategy can lead to heterogeneity in the amount of

181   each individual DNA due to measurement (poor estimations of individuals DNA

182   concentration) or pipetting errors (Gautier *et al.* 2013). Extracted genomic DNA was

183   quantified using a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and quality was

184   estimated using a Nanodrop® ND-2000 Spectrophotometer (Thermo Fisher Scientific,

185   Wilmington, DE, USA). The DNA samples conformed to the required purity criteria

186   (A260/A230 and A260/A280 > 1.8) for gDNA library preparation for sequencing, but not to

187   the required concentration criterion, as the amount of DNA obtained from a 300 individuals

188 pool was still very low (*i.e.* 17 ng of DNA per pool on average). Paired-end libraries were

189 therefore constructed from the totality of DNA available in each pool (*i.e.* each biological

190 replicate in each lineage) contained in 100 µL AE buffer (Qiagen), using the TruSeq Nano

191 DNA Sample Preparation Kits (Illumina, San Diego, CA, USA), according to the

192 manufacturer's instructions. Briefly, DNA was fragmented using a Covaris M220

193 ultrasonicator (Covaris, Woburn, MA, USA) and, after a purification step, end repaired. A

194 bead-based size selection was performed, then DNA was A-tailed and ligated to indexed

195 sequencing adapters (60 bp on each side), each biological replicate being identified by a

196 different index. After a double-purification, libraries were enriched by eight PCR cycles,

197 followed by a final purification. Library profiles were controlled using a DNA High

198 Sensitivity chip on a BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA),

199 showing that the average size of the eight libraries was 513 bp (*i.e.* average insert size of 493

200 bp). Libraries were then quantified by qPCR on ABI7900HT (Applied Biosystems) in order to

201 pool them in equimolar proportions. The final pool was again quantified by qPCR on

202 ABI7900HT in order to load on each lane a volume corresponding to a DNA concentration of

203 8 pM. In order to sequence the eight libraries at an expected 150X coverage depth, four lanes

204 were necessary. Sequencing was performed on an Illumina HiSeq™ 2500 platform (Illumina).

205 Three lanes were sequenced with HiSeq v4 reagent kits (HiSeq PE Cluster Kit v4 and HiSeq

206 SBS Kit v4, Illumina) providing 2x125 bp paired-end reads and the fourth lane was

207 sequenced with TruSeq v3 reagents kit (TruSeq PE Cluster Kit v3 and TruSeq SBS Kit v3,

208 Illumina) providing 2x100 bp paired-end reads. Libraries preparation and sequencing were

209 performed at the GeT-PlaGe France Genomics sequencing platform (Toulouse, France). All

210 the obtained data were submitted to the BBRIC Archive network (https://bbric-

211 archive.toulouse.inra.fr/web/index.html) in Project Gpool.

Mapping and SNP calling

Reads from the 64 fastq files (4 lineages x 2 biological replicates x 4 lanes x 2 (paired-end)) were aligned to the *G. pallida* reference genome using Bowtie2 version 2.1.0 (Langmead *et al.* 2009; Langmead & Salzberg 2012) with default parameters. Duplicate reads were removed using Picard version 1.122.0 (http://broadinstitute.github.io/picard/) and remaining reads were realigned around indels using Realigner Target Creator and Indel Realigner from the Genome Analysis Toolkit, GATK version 2.8.0 (McKenna *et al.* 2010; DePristo *et al.* 2011; Van der Auwera *et al.* 2013). The 32 BAM files obtained were merged in a single mpileup file using SAMtools MPileup version 0.1.19 (Li *et al.* 2009) with minimum mapping quality and minimum base quality set to 20. The whole previous workflow was performed on the Galaxy platform (https://galaxyproject.org/, Blankenberg *et al.* 2010). The mpileup file was further processed to perform SNP calling and derive read counts for each alternative base, after discarding bases with a Base Alignment Quality score below 25. Genomic positions kept in the dataset had to show only two different bases across all the samples, however, the tri– allelic positions for which the third allele was represented by only one read were included in the analysis as bi-allelic SNPs (after filtering the third allele as a sequencing error). To evaluate the reproducibility of the sequencing results across the technical replicates (sequencing lanes), we carried out a Principal Component Analysis (PCA) based on the allele frequency counts (Fig. S1, Supporting information). A last filtering step based on the coverage and the minimum allele count (MAC) was performed: a SNP was kept in the dataset if i) it had a coverage of more than 30 reads and less than 300 reads (corresponding approximately to the 95% percentile of the empirical coverage distribution) in each sample and ii) the minor allele was represented by at least two reads in two different pool samples.

236

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

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

The allele count dataset was analyzed under all the models implemented in BayPass with all the parameter set to default, considering the virulence/avirulence status (coded as a binary variable with value -1 and 1 respectively) of each sample for the association analysis. First, analysis under the core model allowed to estimate the *XtX* for each SNP that were calibrated after analyzing a pseudo-observed dataset (POD) simulated under the inference model with parameters set equal to those estimated in the real data. Second, we relied on the Bayes Factor (BF) calculated under the auxiliary covariate model to evaluate association of SNPs to the virulent/avirulent status of the lineages. The auxiliary covariate model involves

the introduction of a binary auxiliary variable to classify each locus as associated or not. This allows to easily compute Posterior Inclusion Probability (and BF) for each locus while explicitly accounting for multiple testing issues. For each SNP, the Bayes Factor (denoted BF_{mc} as in Gautier 2015) was expressed in deciban units (dB) via the transformation $10 \log_{10}$ (BF). As a decision rule, we then followed the popular Jeffreys' rule (Jeffreys 1961) that quantifies the strength of evidence (here in favor of association of the SNP with the covariable) as 'strong' when 10 dB < BF < 15 dB, 'very strong' when 15 dB < BF < 20 dB and 'decisive' when BF > 20 dB.

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

As we were interested in identifying genomic regions involved in the adaptation to Iledher, we precisely had to consider only SNPs showing footprints of selection in the Iledher lineages (SMI and NI) and not in the Désirée lineages. It is in fact possible to pinpoint genomic regions involved in host adaptation (rather than resistance adaptation), which can occur both in nematode lineages having evolved on Iledher and in the ones having evolved on Désirée, as highlighted in Eoche-Bosy *et al.* (2016). For this purpose, we used the lnRH test (Kauer *et al.* 2003; Schlötterer & Dieringer 2005). LnRH is traditionally used for microsatellite datasets, but as it is not based on a particular mutation model (Kauer *et al.* 2003), it can also be applied to SNP datasets (Mc Evoy *et al.* 2006; Vasemägi *et al.* 2012).

284   This test, based on the assumption that markers linked to loci under selection will show

285   reduced levels of diversity within populations (Schlötterer & Dieringer 2005), uses the

286   expected heterozygosity H to compute the lnRH statistics. We merged allelic counts of

287   biological replicates and derived from it allelic frequencies at each SNP in SMI, SMD, NI and

288   ND. We computed the lnRH statistic for each locus in lineage pairs SMI/SMD and NI/ND by

289   calculating the natural logarithm (ln) of the gene diversity ratio $[(1/(1-H_{lineage1}))^2-1] / [(1/(1-$

290    $H_{lineage2}))^2-1]$. For lineages with monomorphic loci, one different allele was added to one

291   individual, as null values of heterozygosity prevent the estimation of lnRH due to division by

292   zero (Kauer *et al.* 2003). LnRH estimates were standardized to obtain a mean of 0 and a

293   standard deviation of 1. As the lnRH is approximately normally distributed under the null

294   hypothesis of neutrality (Schlötterer & Dieringer 2005), loci with lnRH values lower than -

295   1.96 were considered outliers at the 0.05 threshold, indicating reduced variability in SMI and

296   NI lineages as compared to SMD and ND, respectively. Only outliers in both lnRH pairwise

297   comparisons were considered, and common outliers to BayPass and lnRH were retained to

298   further investigation of their genic environment.

299

300   *Genic environment of outlier loci*

301   We searched for predicted genes located in a window of 120-kb around outlier loci (60-kb

302   either side) in the annotated *G. pallida* genome. The choice of this value was based on the

303   AFLP-based genetic linkage map of *G. rostochiensis* giving a physical:genetic distance ratio

304   of 120 kb.cM^{-1} (Rouppe van der Voort *et al.* 1999), as no data on the extent of recombination

305   and linkage disequilibrium was available in *G. pallida*. As the majority of virulence factors

306   identified in plant parasitic nematodes are effectors (Haegeman *et al.* 2012; Mitchum *et al.*

307   2013), we searched as a priority for genes coding for secreted proteins, *i.e.* harboring N-

308 terminal signal peptides in the predicted proteins, using SIGNALP v4.1 (Petersen *et al.* 2011).

309 We also performed a BLAST search of the predicted genes against the annotated *G.*

310 *rostochiensis* genome in order to potentially obtain more precise information on their

311 functional annotation. Eves-van den Akker *et al.* (2016) identified in *G. rostochiensis* and

312 subsequently in *G. pallida* genomes a dorsal gland promoter element motif, termed DOG

313 Box, which may be a strong predictor of secretion, and thus likely effector function. We thus

314 searched for the presence of the predicted genes identified around outlier loci in the *G. pallida*

315 putative DOG effectors list established by Eves-van den Akker *et al.* (2016).

316

317 **Results**

318 *Sequencing, mapping and SNP calling*

319 Sequencing was carried out on duplicates of the lineages SMI, SMD, NI and ND. Despite the

320 low amount of DNA used for each lineage, the generated sequencing data corresponded to the

321 expectations from a quantitative and qualitative point of view. Sequencing produced two

322 billion reads, *i.e*. 62.5 million reads/sample (biological replicate)/lane, and 76% mapped on

323 the *G. pallida* reference genome, but 24% of these mapping reads mapped more than once on

324 the genome. Duplicate filtering resulted in removing 13.5% of mapped reads. Remaining

325 reads covered 95% of the ungapped assembly at an average depth of 34X/sample/lane.

326 Processed mpileup contained 2,383,040 SNPs. PCA based on the allele frequency counts

327 showed that the main axes of variation of genetic variability across the different samples were

328 clearly associated with (i) the population origin and (ii) the virulence status. A very low

329 variation was observed among the sequencing replicates in each pool sample. This prompted

330 us to merge allele count data from the four sequencing replicates (lanes) for each of the eight

331 pool samples (Fig. S1, Supporting information). The final dataset filtered on coverage and

332 MAC consisted of read count data for 1,631,158 SNPs in the eight different selected lineages,

333 corresponding to a density of polymorphic sites of 16/kb on average.

334

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

336 Analysis of the dataset under the BayPass core model allowed us to estimate the scaled

337 covariance matrix of population allele frequencies Ω that quantifies the genetic relationship

338 among each pairs of populations. The resulting estimates of Ω accurately reflected the known

339 structure between samples, *i.e.* a clustering at the higher level by population geographical

340 origin, then by the virulence status and finally by biological replicate within each lineage (Fig.

341 2). *XtX* for each SNP were also estimated and calibrated by analyses of a POD containing

342 1,600,000 SNPs. At the 1% POD significance threshold, about 33,000 SNPs were identified

343 as overly differentiated (Fig. 3). However, analysis under the auxiliary covariate model

344 drastically reduced this outlier list, as less than 400 had also a $BF_{mc} < 20$ (355, 361 and 357

345 SNPs in each analysis, respectively) (Fig. 3). Overall, 275 outliers were shared by the three

346 analyses (outlier loci found in only one or two analyses actually exhibited *XtX* and/or BF_{mc}

347 values close to the thresholds, which explains that they were not found in all analyses).

348 Among these 275 SNPs showing an increase of genetic differentiation between Iledher and

349 Désirée lineages, 31 of them also showed a decrease of genetic diversity in both SMI and NI

350 lineages (Fig. 4) as identified by lnRH.

351

352 *Genic environment of outlier loci*

353 The 31 selected outlier loci were distributed on 23 different scaffolds. Among these outlier

354 loci for which the genic environment has been investigated, three outliers were located on

355 scaffolds which do not harbor any predicted gene, 16 were located in intergenic area and 12

356 were directly located into genes (Table S2, Supporting information). Among the latter, five

357 were located in exons, and four of them corresponded to synonymous mutations. The outlier

358 SNP located on the scaffold 1777 corresponded to a non-synonymous mutation that changes

359 an arginine into a histidine in GPLIN_001438500, which is annotated as a transcribed

360 hypothetical protein. Despite the fragmentation of the *G. pallida* genome in numerous

361 scaffolds that precluded a clear analysis of the physical distances among most of the identified

362 outliers, it appeared that several outlier loci were sometimes found in the same scaffold. The

363 most significant was the scaffold 988 which contains four outlier SNPs close to each other (<

364 200 bp). Four other scaffolds (44, 85, 182 and 283) harbored outlier SNPs close to each other

365 (< 200 bp) and a fifth scaffold (66) harbored also two outlier SNPs but that are 75 kb apart.

366 The number of outlier SNPs found in a scaffold was not dependent on the size of this

367 scaffold: for instance, scaffold 988, which is one of the smallest scaffold of interest identified,

368 contains four outlier loci. Overall, 258 predicted genes were identified in a 120-kb window

369 around the 31 outlier loci (Table S2, Supporting information). About 47% (121) of them have

370 unknown function in *G. pallida*, but 15 harbored a signal peptide, indicating that they

371 potentially encode for secreted proteins, and one, GPLIN_000314000, is similar to an effector

372 in *G. rostochiensis* (Cotton *et al*. 2014). Fourteen predicted genes with known functions also

373 harbored a peptide signal. Fifteen genes coding for proteins harboring a SPRY domain were

374 found at the proximity of eight outlier loci, and three of them were paralog of SPRYSECs

375 (Secreted protein with a SPRY domain), which are known in nematodes of the genus

376 *Globodera* to be involved in pathogenicity but also in (a)virulence (Rehman *et al.* 2009;

377 Sacco *et al.* 2009). Two of them, located on scaffolds 182 and 782, were paralog of RBP-1,

378 the only virulence gene identified in *G. pallida* (Sacco *et al.* 2009; Carpentier *et al.* 2012). It

379 is likely that some of the other genes coding for proteins harboring a SPRY domain could

380 actually be SPRYSECs, the absence of signal peptide being not ascertained as they are 5'

381 truncated (N's region) in the reference genome. Other genes were also interesting as they

382 encode proteins which have functions linked to *G. pallida* pathogenicity, e.g., cell wall

383 modifying proteins, or they encode proteins which are known to be secreted by the dorsal

384 pharyngeal gland of *G. pallida*. The pharyngeal glands are cells in which effectors are

385 produced, and thus can be regarded as a toolbox for infection (Eves-van den Akker & Birch

386 2016). Cyst nematodes have two sets, subventral and dorsal: the former are primarily active

387 while the nematode migrates through host tissue, while the latter are primarily active during

388 the sedentary parasitic stages (Endo 1987; von Mende 1997; Davis *et al.* 2000). Some genes

389 with unknown function in *G. pallida* had homologues with predicted function in *G.*

390 *rostochiensis* genome, part of them being putatively paralog of SPRYSECs or involved in

391 pathogenicity. Three predicted genes (GPLIN_000832500, GPLIN_001056700 and

392 GPLIN_001258100) harbored DOG Boxes in their promoter region, two of them having

393 unknown function and the third coding for a SPRYSEC protein.

394

395 **Discussion**

396 In this study, we investigated *Globodera pallida* genomic regions involved in the adaptation

397 to the QTL $GpaV_{vrn}$, a resistance factor against this potato parasite, by performing a genome

398 scan on Pool-Seq data derived from experimentally evolved lineages, adapted or not to the

399 resistant potato cultivar Iledher.

400 Elucidating the molecular bases of virulence is a central challenge in nematology. So

401 far, a few plant-nematode interactions have been studied to characterize the molecular

402 determinants of (a)virulence. In *G. pallida*, only one virulence gene has been described

403 (Sacco *et al.* 2009; Carpentier *et al.* 2012). *A priori* methods based on candidate genes to

404   identify new or unknown virulence genes are therefore not the most adapted ones at the

405   moment. Our study is the first to exploit genome scan to target genetic bases of adaptation to

406   plant resistance in nematode populations. We demonstrated in a previous study the feasibility

407   of a genome scan approach on our specific biological materiel coming from short

408   experimental evolution (Eoche-Bosy *et al.* 2016), which was achieved here by the

409   identification of 275 outlier loci and 31 candidates of a strong interest among them,

410   distributed on 23 scaffolds. Several genomic regions are therefore putatively linked to

411   adaptation to the resistance of Iledher, however this number should be put in perspective with

412   respects to the moderate quality of the *G. pallida* reference genome. In fact, the genome

413   assembly is still relatively fragmented (6,873 scaffolds), which could suggest that different

414   genomic regions identified as outliers in our study could in fact form only one or a few. Eves-

415   van den Akker *et al.* (2016) highlighted that some effector islands identified in *G.*

416   *rostochiensis* genome were split across different scaffolds in *G. pallida* genome. This

417   prevents to make realistic assumptions on the number of genes that could be involved in the

418   overcoming of Iledher resistance. However, we assume that only a few virulence genes

419   should be involved as few resistance genetic factors are expected in Iledher because of the

420   major effect of the QTL *GpaV$_{vrn}$* which is present in the potato cultivar Iledher and the fast

421   overcoming of this resistance observed in experimental evolution that support rather the view

422   of a mono- or oligogenic resistance.

423   The moderate quality of the genome involved other limitations in mapping of

424   sequencing reads, leading to the loss of a substantial part of the data. First, 24% of the

425   sequencing reads did not map on the reference genome, while in the same time, the assembly

426   contains 17% of N's and has a CEGMA completeness score of 74% for complete genes and

427   81% for partial genes (Parra *et al.* 2007; Eves-van den Akker *et al.* 2016). This suggests that a

428 large part of unmapped reads could actually belongs to the *G. pallida* genome, but

429 corresponding sequences would be absent from the assemblage. These missing sequences

430 could putatively comprise candidate genomic regions to adaptation. Second, heterozygosity of

431 the reference genome is suspected, as about 24% of the sequencing reads mapped more than

432 once on the genome, this hypothesis being supported by the difference between the length of

433 the assembly (124.7 Mb) and the estimated genome size (100 Mb). Sequencing data mapping

434 on these genomic regions were further discarded because of poor mapping score due to the

435 multiple positions mapping, and once again, genomic information about these regions and

436 their possible link to resistance adaptation were lost.

437 Pool-Seq is now commonly adopted in many studies, included population genomics

438 studies (Rubin *et al.* 2010, 2012; Clément *et al.* 2013; Ferretti *et al.* 2013; Fisher *et al.* 2013),

439 because it provides more accurate estimation of allele frequencies, at reduced sequencing and

440 library preparation costs (Futschik & Schlötterer 2010; Zhu *et al.* 2012; Gautier *et al.* 2013;

441 Rellstab *et al.* 2013; Schlötterer *et al.* 2014). Here, we emphasize the interest of using such an

442 approach for population genomic studies of non-clonal microorganisms, as it allowed to

443 increase the amount of DNA while keeping information about allele frequencies. The good

444 quality of our results also relies on library preparation protocol, which fit the

445 recommendations (Rhodes *et al.* 2014; Kofler *et al.* 2016). The TruSeq Nano DNA kit has

446 been shown to be highly accurate and even some duplicates were obtained due to the PCR

447 cycles, Kofler *et al.* (2016) showed that duplicates were also obtained with free-PCR library

448 preparation kits. Nevertheless, Kofler *et al.* (2016) highlighted the superior impact of the

449 genome quality over the library preparation protocol on the mapping quality. Numbers of

450 individuals in pools and depth of coverage used have also been determined to improve

451 sequencing results and allelic frequency estimations (Gautier *et al.* 2013; Kofler *et al.* 2016).

452   Consideration of a categorical population specific covariable in the genome scan

453   allowed us to refine the list of outlier loci. Analysis of association further conducted by

454   running the auxiliary covariate model allowed to retain only SNPs associated with the

455   virulence status of the lineages, and not those corresponding to other selective pressures (e.g.,

456   local adaptation), which helped to reduce the risk of a misleading biological interpretation, a

457   main limitation of genome scans (Pavlidis *et al.* 2012). Moreover, evaluating the reduction of

458   diversity allowed us to refine the outlier loci list to those linked only to the adaptation to the

459   resistant cultivar Iledher. A direct link between those candidate loci and the adaptation to the

460   resistance cannot yet be done at this stage. Indeed, footprints of selection detected in lineages

461   having evolved on Iledher might well be linked to adaptation to the plant resistance or to the

462   plant itself (including all its genetic background). Also, at this stage, we are unable to

463   definitively link the identified SNPs to the resistance factor $GpaV_{vrn}$ or to another resistance

464   factor present in the genetic background of the cultivar Iledher. Although these different

465   situations will only be deciphering by functional validation, there is a strong possibility that at

466   least some of the candidate genomic regions identified in this study are involved in adaptation

467   to Iledher resistance and in particular to its major resistance QTL $GpaV_{vrn}$.

468   Sixteen over the 31 outlier loci investigated for their genic environment have at least

469   one gene coding for a secreted protein in their neighboring region (even if all secreted

470   proteins are not effectors, all effectors are secreted protein) and several outliers are at the

471   vicinity of genes coding for proteins specifically secreted in the nematode dorsal gland.

472   Particularly, some of these secreted proteins are SPRYSECs effectors which act as a versatile

473   protein-binding platform for the nematodes to target a wide range of host proteins during

474   parasitism including plant resistance proteins (Rehman *et al.* 2009; Sacco *et al.* 2009; Diaz-

475   Granados *et al.* 2016). The fact that a considerable part of the outlier loci are found near genes

476 coding for a SPRY domain containing protein or even in or at the vicinity of SPRYSECs

477 cannot be only explained by the huge size of this family gene or the prevalence of the SPRY

478 domain in the *G. pallida* genome. In fact, 299 SPRY domain containing proteins are present

479 in the *G. pallida* genome (which can be moreover cluster into islands), *i.e.* 1 for 334 kb, and

480 30 of them are SPRYSECs, *i.e.* 1 SPRYSEC for 3.33 Mb (Mei *et al.* 2015; Eves-van den

481 Akker *et al.* 2016). When looking at the genic environment of outlier loci, we explored a total

482 of 2 Mb of the genome and we identified 15 SPRY domain containing proteins, *i.e.* 1 for 133

483 kb, and 3 SPRYSECs, *i.e.* 1 for 660 kb. This is more than expected by chance in both cases

484 and therefore support the view that some SPRYSEC proteins should be well indeed involved

485 in the adaptation of *G. pallida* to Iledher.

486 It can be noticed that the lnRH test also identified, among the 275 outlier loci, six

487 SNPs showing a decrease of genetic diversity in both Désirée lineages (data not shown).

488 These loci were distributed on four scaffolds, different from those on which the 31 outliers

489 were located. Although we did not consider them in this study, those results could further be

490 used to study adaptation to the potato cultivar Désirée. Only one of the 13 microsatellite

491 markers identified as outliers by Eoche-Bosy *et al.* (2016) was located on the same scaffold as

492 one of the 31 outlier SNPs identified here, and no one was located on the same scaffold as the

493 six outlier SNPs involved in Désirée adaptation. The microsatellite marker Gp235 was indeed

494 located at about 150 bp from the outlier SNP identified here on the scaffold 216. The fact that

495 the other outlier microsatellites were not found on the same scaffold that outlier SNPs is not

496 surprising, due to the small number of microsatellites used and to the fragmentation of the

497 genome.

498 By identifying genomic regions putatively involved in the adaptation to the resistance

499 from *S. vernei*, the present study has taken a further step towards understanding and

500　identifying the determinants of virulence. In the aim of pinpointing, among the candidate loci,

501　those actually involved in the adaptation to the potato cultivar Iledher, a first step of

502　validation could consist of studying allele frequencies of candidate SNPs in natural

503　populations showing different virulence level to Iledher, which could allow to validate some

504　of them whose variations in allelic frequencies would correlate with variations in virulence

505　level. Even if those loci are not the mutation responsible for the adaptation, they can be

506　enough linked to it to be used as molecular tools to determine virulence allele frequencies in

507　field populations before the deployment of resistant cultivars, and to study more accurately

508　the potential fitness costs or benefits linked to the mutation from avirulence to virulence.

509　Indeed, a recent study suggested that adaptation to Iledher involves an increase of fitness on a

510　susceptible potato cultivar (Fournet *et al.* 2016). However, this study did not test for

511　competition between avirulent and virulent individuals, because of the lack of molecular

512　markers allowing to follow (a)virulence alleles. And it was shown in the *Potato Virus Y* that a

513　cost of competitiveness could occur, even in absence of fitness cost in simple inoculation

514　(Janzac *et al.* 2010). The identified outlier loci will also allow to target more precisely the best

515　candidate genes involved in the adaptation, which will have to be functionally validated.

516

References

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.

Ayme V, Souche S, Caranta C *et al.* (2006) Different mutations in the genome-linked protein VPg of *Potato virus Y* confer virulence on the *pvr2³* resistance in pepper. *Molecular Plant-Microbe Interactions*, **19**, 557-563.

Barbary A, Djian-Caporalino C, Palloix A, Castagnone-Sereno P (2015) Host genetic resistance to root-knot nematodes, *Meloidogyne* spp., in Solanaceae: from genes to the field. *Pest Management Science*, **71**, 1591–1598.

Bekal S, Domier LL, Gonfa B, Lakhssassi N, Meksem K, Lambert KN (2015). A SNARE-like protein and biotin are implicated in soybean cyst nematode virulence. *PLoS ONE*, **10**, e0145601.

Blankenberg D, Von Kuster G, Coraor N *et al.* (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology*, **89**, 19.10.1-19.10.21.

Bonhomme M, Chevalet C, Servin B *et al.* (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, **186**, 241-262.

Carpentier J, Esquibet M, Fouville D, Manzanares-Dauleux MJ, Kerlan MC, Grenier E (2012) The evolution of the *Gp-Rbp-1* gene in *Globodera pallida* includes multiple selective replacements. *Molecular Plant Pathology*, **13**, 546-555.

Castagnone-Sereno P (2002) Genetic variability of nematodes: a threat to the durability of plant resistance genes? *Euphytica*, **124**, 193-199.

Clément JA, Toulza E, Gautier M *et al.* (2013) Private selective sweeps identified from next-generation pool-sequencing reveal convergent pathways under selection in two inbred *Schistosoma mansoni* strains. *PLoS Neglected Tropical Diseases*, **7**, e2591.

Coop G, Witonsky D, Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411-1423.

Cotton JA, Lilley CJ, Jones LM *et al.* (2014) The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology*, **15**, R43.

Davis EL, Hussey RS, Baum TJ *et al.* (2000) Nematode parasitism genes. *Annual Review of Phytopathology*, **38**, 365-396.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491-498.

Díaz JA, Nieto C, Moriones E, Truniger V, Aranda MA (2004) Molecular characterization of a *Melon necrotic spot virus* strain that overcomes the resistance in melon and nonhost plants. *Molecular Plant-Microbe Interactions*, **17**, 668-675.

Diaz-Granados A, Petrescu AJ, Goverse A, Smant G (2016) SPRYSEC effectors: a versatile protein-binding platform to disrupt plant innate immunity. *Frontiers in Plant Science*, **7**, 1575.

Djian-Caporalino C, Palloix A, Fazari A *et al.* (2014) Pyramiding, alternating or mixing: comparative performances of deployment strategies of nematode resistance genes to promote plant resistance efficiency and durability. *BMC Plant Biology*, **14**, 53.

Endo BY (1987) Ultrastructure of esophageal gland secretory granules in juveniles of *Heterodera glycines*. *Journal of Nematology*, **19**, 469-483.

Eoche-Bosy D, Gauthier J, Juhel AS *et al.* (2016). Experimentally evolved populations of the potato cyst nematode *Globodera pallida* allow the targeting of genomic footprints of selection due to host adaptation. *Plant Pathology*, doi: 10.1111/ppa.12646.

Eves-van den Akker S, Birch PR (2016) Opening the effector protein toolbox for plant-parasitic cyst-nematode interactions. *Molecular Plant*, **9**, 1451-1453.

Eves-van den Akker S, Laetsch DR, Thorpe P *et al.* (2016) The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights into the basis of parasitism and virulence. *Genome Biology*, **17**, 124.

Ferretti L, Ramos-Onsins SE, Pérez-Enciso M (2013) Population genomics from pool sequencing. *Molecular Ecology*, **22**, 5561-5576.

Fischer MC, Rellstab C, Tedder A *et al.* (2013) Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology*, **22**, 5594-5607.

Förstner W, Moonen B (2003) A metric for covariance matrices. In: *Geodesy-The Challenge of the 3rd Millennium* (eds Grafarend EW, Krumm FW, Schwarze VS), pp. 299-309. Springer-Verlag, Berlin/Heidelberg, Germany.

Fournet S, Eoche-Bosy D, Renault L, Hamelin FM, Montarry J (2016) Adaptation to resistant hosts increases fitness on susceptible hosts in the plant parasitic nematode *Globodera pallida*. *Ecology and Evolution*, **6**, 2559-2568.

Fournet S, Kerlan MC, Renault L, Dantec JP, Rouaux C, Montarry J (2013) Selection of nematodes by resistant plants has implications for local adaptation and cross-virulence. *Plant Pathology*, **62**, 184-193.

Futschik A, Schlötterer C (2010) The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, **186**, 207-218.

Gautier M (2015) Genome-wide scan for adaptive differentiation and association analysis with population-specific covariables. *Genetics*, **201**, 1555-1579.

Gautier M, Foucaud J, Gharbi K *et al.* (2013) Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology*, **22**, 3766-3779.

Green CD, Greet DN, Jones FGW (1970) The influence of multiple mating on the reproduction and genetics of *Heterodera rostochiensis* and *H. schachtii. Nematologica*, **16**, 309-326.

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205-220.

Haegeman A, Mantelin S, Jones JT, Gheysen G (2012) Functional roles of effectors of plant-parasitic nematodes. *Gene*, **492**, 19-31.

Janzac B, Fabre F, Palloix A, Moury B (2009) Constraints on evolution of virus avirulence factors predict the durability of corresponding plant resistances. *Molecular Plant Pathology*, **10**, 599-610.

Janzac B, Montarry J, Palloix A, Navaud O, Moury B (2010) A point mutation in the polymeras *Potato Virus Y* confers virulence toward the *Pvr4* resistance of pepper and a high competitiveness cost in susceptible cultivar. *Molecular Plant-Microbe Interactions*, **23**, 823-830.

Jeffreys H (1961) *Theory of Probability*, 3rd edn. Oxford University Press, Oxford, UK.

Jones FGW (1950) Observations on the beet eelworm and other cyst-forming species of *Heterodera. Annals of Applied Biology*, **37**, 407-440.

Jones MGK, Northcote DH (1972) Nematode-induced syncytium − a multinucleate transfer cell. *Journal of Cell Science*, **10**, 789-809.

Kauer MO, Dieringer D, Schlötterer C (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of *Drosophila melanogaster. Genetics*, **165**, 1137-1148.

Kofler R, Nolte V, Schlötterer C (2016) The impact of library preparation protocols on the consistency of allele frequency estimates in Pool-Seq data. *Molecular Ecology Resources*, **16**, 118-122.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357-359.

Langmead B, Trapnell Cole, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Lozano-Torres JL, Wilbers RH, Gawronski P *et al.* (2012) Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 10119-10124.

Lozano-Torres JL, Wilbers RH, Warmerdam S *et al.* (2014) Apoplastic venom allergen-like proteins of cyst nematodes modulate the activation of basal plant innate immunity by cell surface receptors. *PLoS Pathogens*, **10**, e1004569.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981-994.

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, **24**, 133-141.

McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, **40**, 349-379.

McEvoy B, Beleza S, Shriver MD (2006) The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Human Molecular Genetics*, **15**, R176-R181.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297-1303.

Mei Y, Thorpe P, Guzha A *et al.* (2015) Only a small subset of the SPRY domain gene family in *Globodera pallida* is likely to encode effectors, two of which suppress host defences induced by the potato resistance gene *Gpa2*. *Nematology*, **17**, 409-424.

Meshi T, Motoyoshi F, Adachi A, Watanabe Y, Takamatsu N, Okada Y (1988) Two concomitant base substitutions in the putative replicase genes of tobacco mosaic virus confer the ability to overcome the effects of a tomato resistance gene, *Tm-1*. *EMBO Journal*, **7**, 1575-1581.

Mitchum MG, Hussey RS, Baum TJ *et al.* (2013) Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist*, **199**, 879-894.

Montarry J, Jan PL, Gracianne C *et al.* (2015) Heterozygote deficits in cyst plant-parasitic nematodes: possible causes and consequences. *Molecular Ecology*, **24**, 1654-1677.

Nicol JM, Turner SJ, Coyne DL, den Nijs L, Hockland S, Maafi ZT (2011) Current nematode threats to world agriculture. In: *Genomics and Molecular Genetics of Plant-Nematode Interactions* (eds Jones J, Gheysen G, Fenoll C), pp. 21-43. Springer, Dordrecht, The Netherlands.

Oerke EC, Dehne HW, Schönbeck F, Weber A (1994) *Crop production and crop protection: estimated losses in major food and cash crop*. Elsevier Science BV, Amsterdam, The Netherlands.

Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**, 1061-1067.

Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution*, **29**, 3237-3248.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SIGNALP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, **8**,785-786.

Picard D, Plantard O, Scurrah M, Mugniéry D (2004) Inbreeding and population structure of the potato cyst nematode (*Globodera pallida*) in its native area (Peru). *Molecular Ecology*, **13**, 2899-2908.

Plantard O, Porte C (2004) Population genetic structure of the sugar beet cyst nematode *Heterodera schachtii*: a gonochoristic and amphimictic species with highly inbred but weakly differentiated populations. *Molecular Ecology*, **13**, 33-41.

Rehman S, Postma W, Tytgat T *et al.* (2009) A secreted SPRY domain-containing protein (SPRYSEC) from the plant-parasitic nematode *Globodera rostochiensis* interacts with a CC-NB-LRR protein from a susceptible tomato. *Molecular Plant-Microbe Interactions*, **22**, 330-340.

Rellstab C, Zoller S, Tedder A, Gugerli F, Fischer MC (2013) Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PLoS ONE*, **8**, e80422.

Rhodes J, Beale MA, Fisher MC (2014) Illuminating choices for library prep: a comparison of library preparation methods for whole genome sequencing of *Cryptococcus neoformans* using Illumina HiSeq. *PLoS One*, **9**, e113501.

Rouppe van der Voort JNAM, van der Vossen E, Bakker E *et al.* (2000) Two additive QTLs conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters. *Theoretical and Applied Genetics*, **96**, 654-661.

Rouppe van der Voort JNAM, van Eck HJ, van Zandvoort PM, Overmars H, Helder J, Bakker J (1999) Linkage analysis by genotyping of sibling populations: a genetic map for the potato cyst nematode constructed using a 'pseudo-F2' mapping strategy. *Molecular and General Genetics*, **261**, 1021–31.

Rubin CJ, Megens HJ, Barrio AM *et al.* (2012) Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 19529-19536.

Rubin CJ, Zody MC, Eriksson J *et al.* (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, **464**, 587-591.

Sacco MA, Koropacka K, Grenier E *et al.* (2009) The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death. *PLoS Pathogens*, **5**, e1000564.

Schlötterer C, Dieringer D (2005) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: *Selective Sweep* (ed. Nurminsky D), pp. 55-64. Landes Bioscience, Georgetown, USA.

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nature Review Genetics*, **15**, 749-763.

Sobczak M, Golinowski W (2011) Cyst nematodes and syncytia. In: *Genomics and molecular genetics of plant-nematode interactions* (eds Jones JT, Gheysen G, Fenoil C), pp. 61-82. Springer, Dordrecht, The Netherlands.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671-688.

714 Triantaphyllou AC, Esbenshade PR (1990) Demonstration of multiple mating in *Heterodera*
715   *glycines* with biochemical markers. *Journal of Nematology*, **22**, 452-456.
716 Van der Auwera GA, Carneiro MO, Hartl C *et al.* (2013) From FastQ data to high confidence
717   variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in*
718   *Bioinformatics*, **43**, 11.10.1-11.10.33.
719 van Riel HR, Mulder A (1998) Potato cyst nematodes (*Globodera* species) in western Europe.
720   In: *Potato Cyst Nematodes: Biology, Distribution and Control* (eds Marks RJ, Brodie BB),
721   pp. 271-298. CAB International, Wallingford, UK.
722 Vasemägi A, Nilsson J, McGinnity P *et al.* (2012) Screen for footprints of selection during
723   domestication/captive breeding of Atlantic Salmon. *Comparative and Functional*
724   *Genomics*, **2012**, 628204.
725 von Mende N (1997) Invasion and migration behaviour of sedentary nematodes. In: *Cellular*
726   *and Molecular Aspects of Plant–Nematodes Interactions* (eds Fenoll C, Grundler FMW,
727   Ohl SA), pp. 51-64. Kluwer Academic Publishers, Dordrecht, The Netherlands.
728 Zhu Y, Bergland AO, González J, Petrov DA (2012) Empirical validation of pooled whole
729   genome population re-sequencing in *Drosophila melanogaster*. *PLoS One*, **7**, e41901.
730

731 **Data accessibility**
732 Sequencing data were submitted to the BBRIC Archive network (https://bbric-
733 archive.toulouse.inra.fr/web/index.html) in Project Gpool.
734

735 **Author Contributions**
736 SF performed the experimental evolution. DEB, ME, SF and JM performed the experiments
737 according to a protocol elaborated jointly by DEB, EG and JM. OB performed the Illumina
738 sequencing at the GeT-PlaGe platform (Toulouse, France). DEB, MG, FL and AB analyzed
739 the data. DEB and JM wrote the text and prepared the figures. All authors edited the article
740 and have approved the current version.
741

742 **Supporting information**
743 Additional supporting information may be found on the online version of this article.
744 **Figure S1** Principal Component Analysis (PCA) of the sequencing data from eight *G. pallida*
745 pools sequenced on four lanes (technical replicates), based on read counts at 2,383,040 SNPs.
746 **Table S2** Putative functions of the predicted genes located in a 120-kb window centered on
747 the 31 outlier loci on the *Globodera pallida* genome assembly version Gpal.v1.0 (Cotton *et*
748 *al.* 2014).

Figure legends

751 **Fig. 1** Selection of the four experimental *Globodera pallida* lineages. Nematode lineages
752 were established from two French natural *G. pallida* populations, SM (near Saint-Malo,
753 Brittany, north-western France) and N (from the island of Noirmoutier, western France),
754 reared during eight successive cycles (*i.e.* eight generations) on the susceptible potato cultivar
755 Désirée (D) and on the resistant cultivar Iledher (I). Each sample name indicates its
756 geographical origin (SM for Saint-Malo and N for Noirmoutier) and the potato cultivar on
757 which it evolved (I for Iledher and D for Désirée).

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

765 **Fig. 3** SNP *XtX* as a function of the BF_{mc} for association with the virulence status covariable,
766 estimated in one of the three independent analyses. The vertical dotted line represents the 1%
767 POD significance threshold (*XtX* = 14.0) and the horizontal dotted line represents the 20 dB
768 threshold for BF_{mc}. Black dots represent the 275 outlier loci of interest (*i.e.* outliers for both
769 *XtX* and BF_{mc} values) and red dots represent the 31 outlier loci of strong interest (*i.e.* outliers
770 also identified with lnRH).

772 **Fig. 4** Example of increase of genetic differentiation (represented as *XtX* estimates computed
773 in BayPass across the whole dataset) and decrease of genetic diversity (represented as
774 expected heterozygosity H in each lineages having evolved on Iledher) in a candidate
775 genomic region. The scaffold represented is the scaffold 988 and the grey line indicates the
776 location of the four outlier SNPs on the scaffold.

777   **Figure 1 – Eoche-Bosy et al.**

778

Figure 2 – Eoche-Bosy et al.
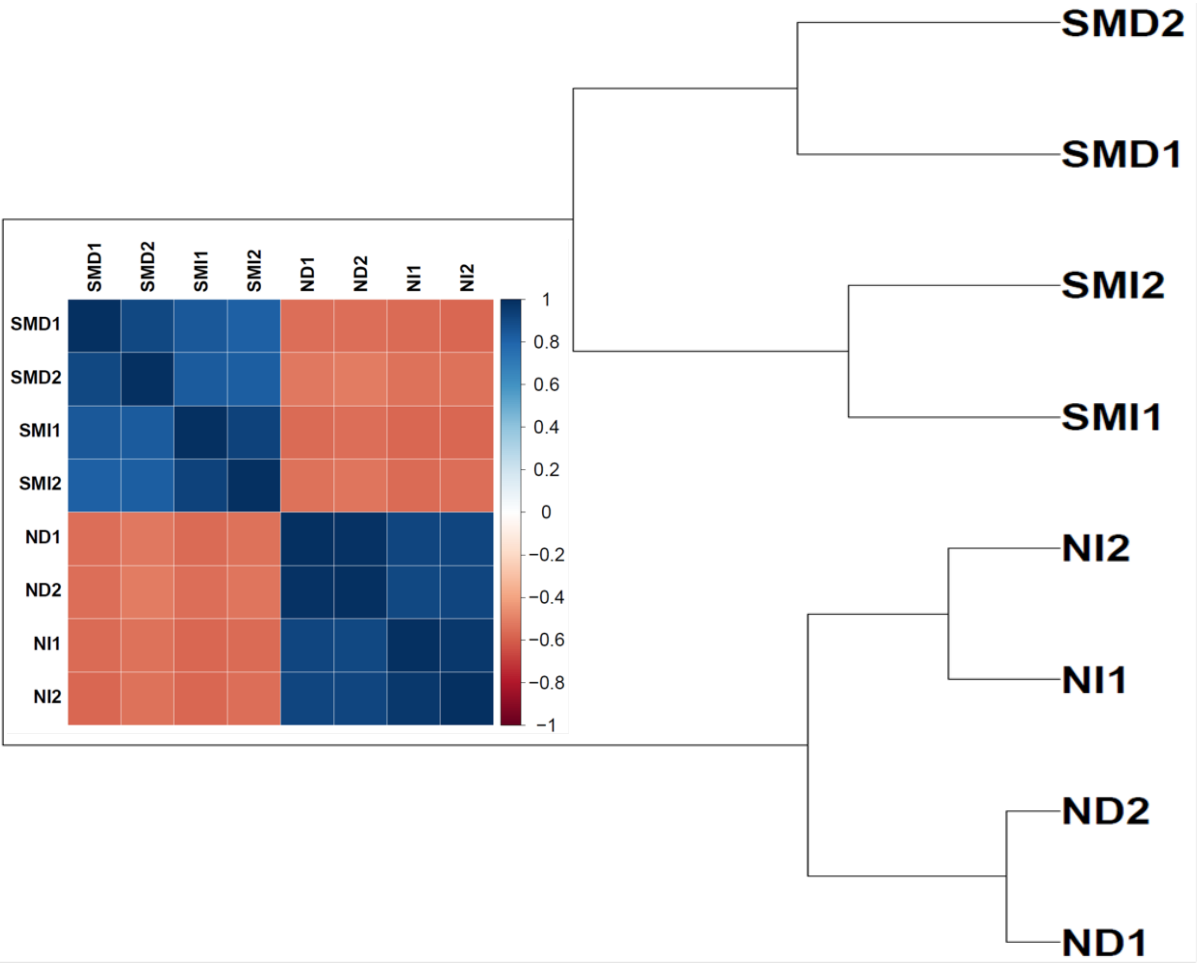
Figure 3 – Eoche-Bosy et al.
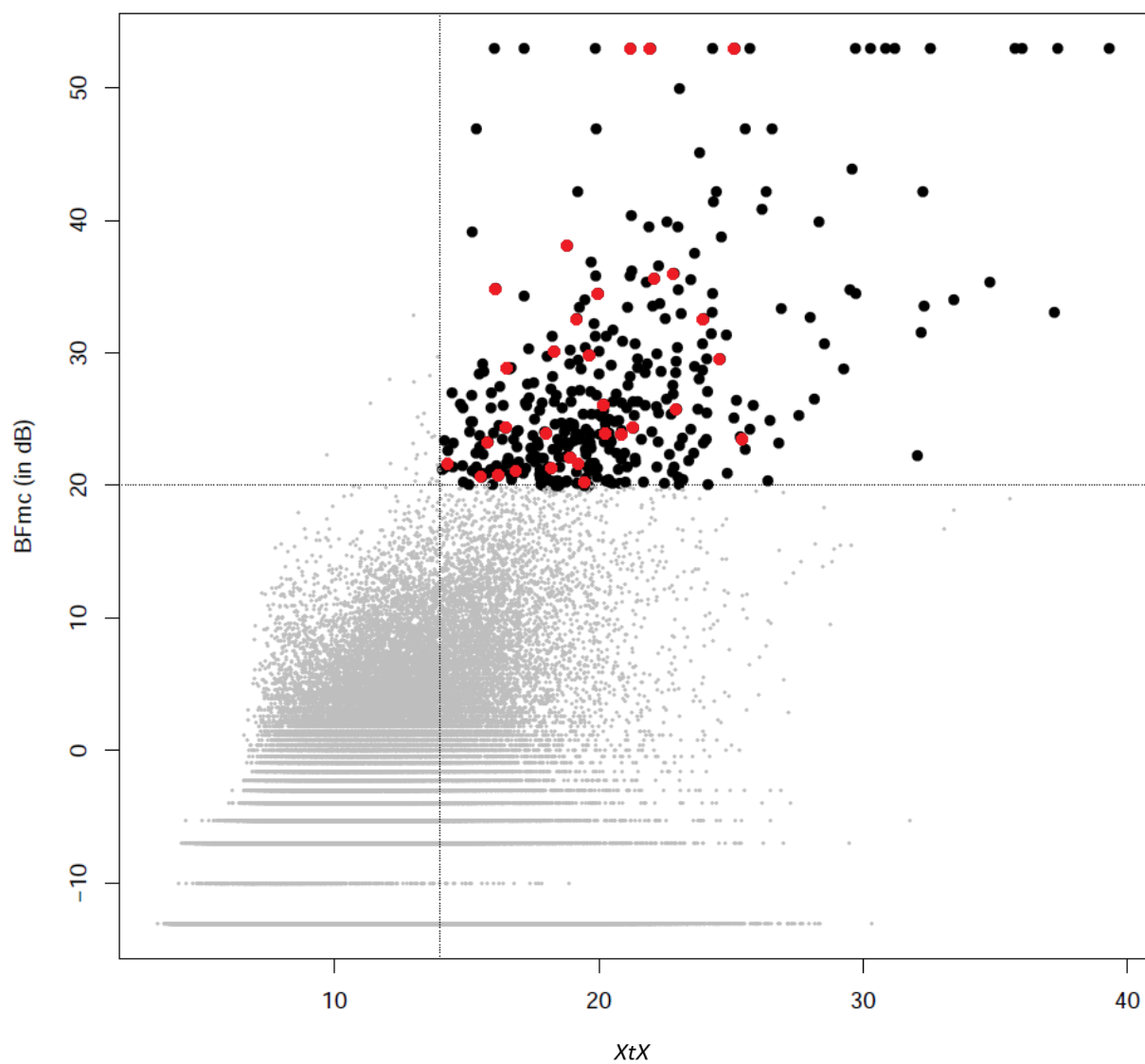
Figure 4 – Eoche-Bosy et al.

785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
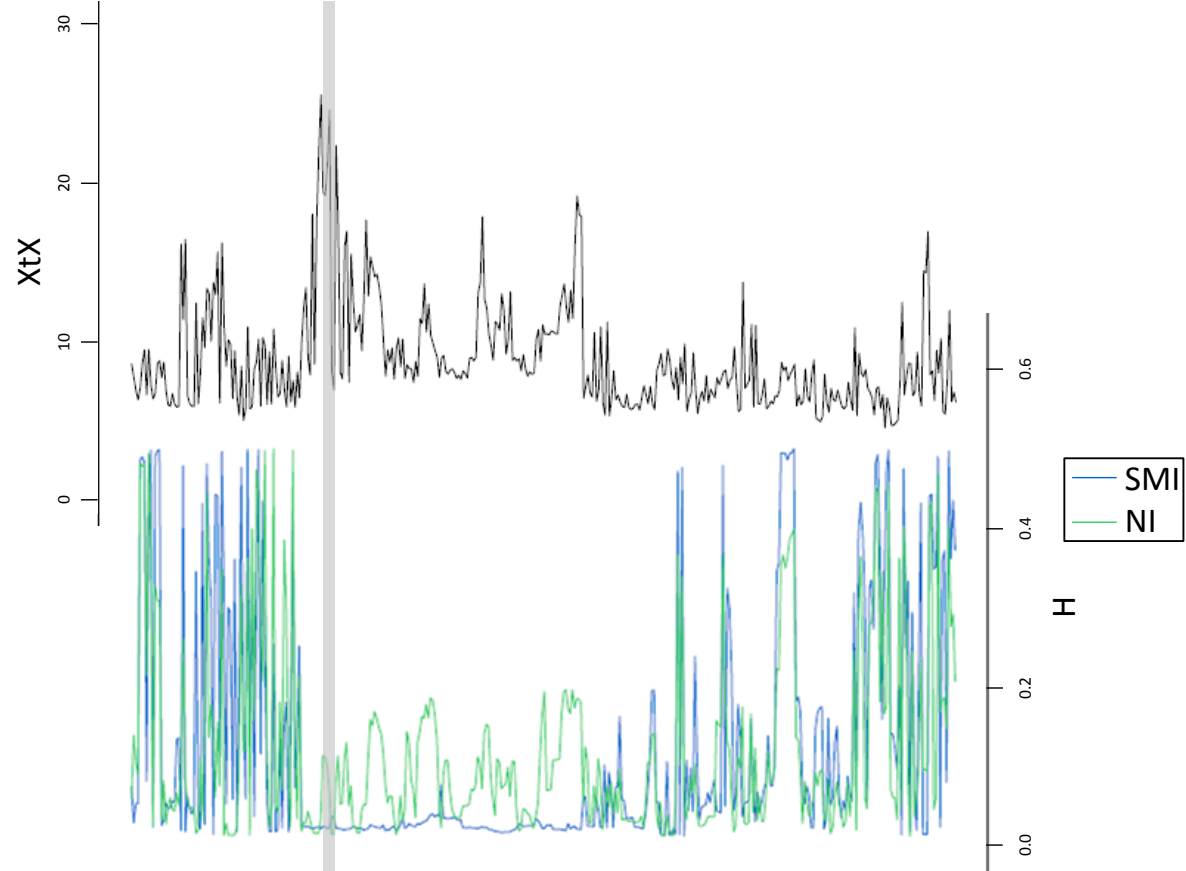
Supporting Information
Supplementary Figure S1

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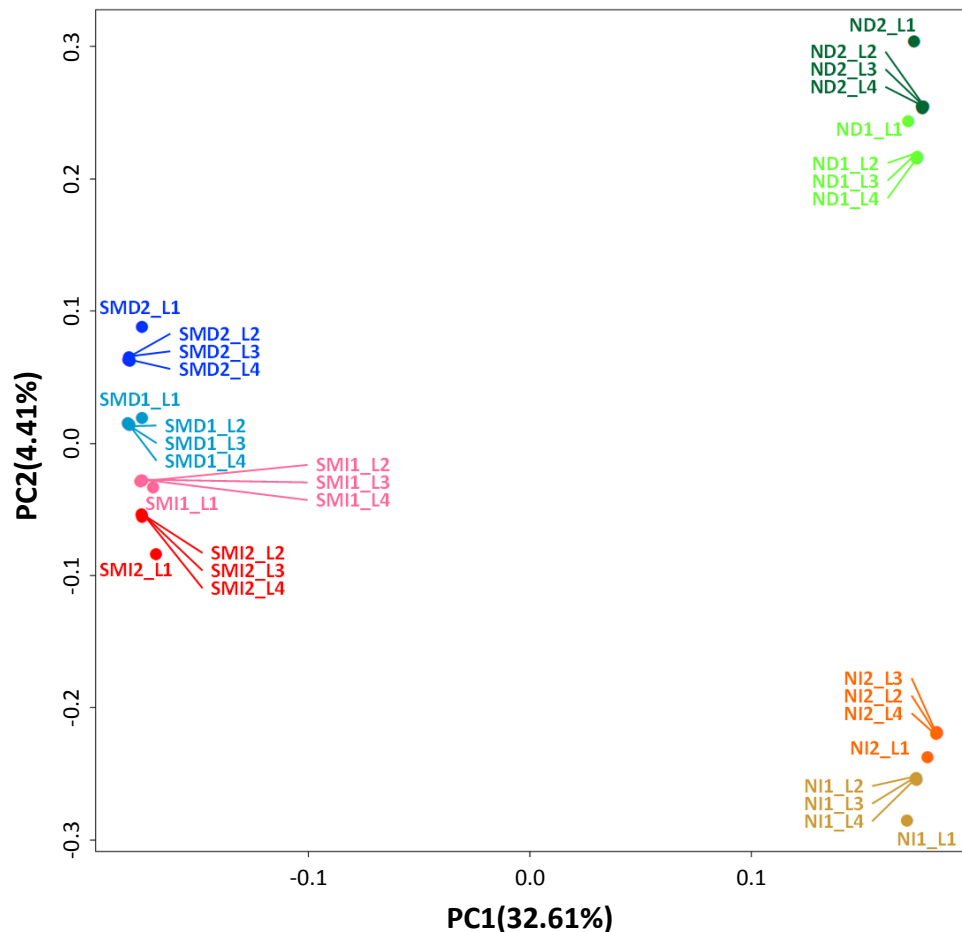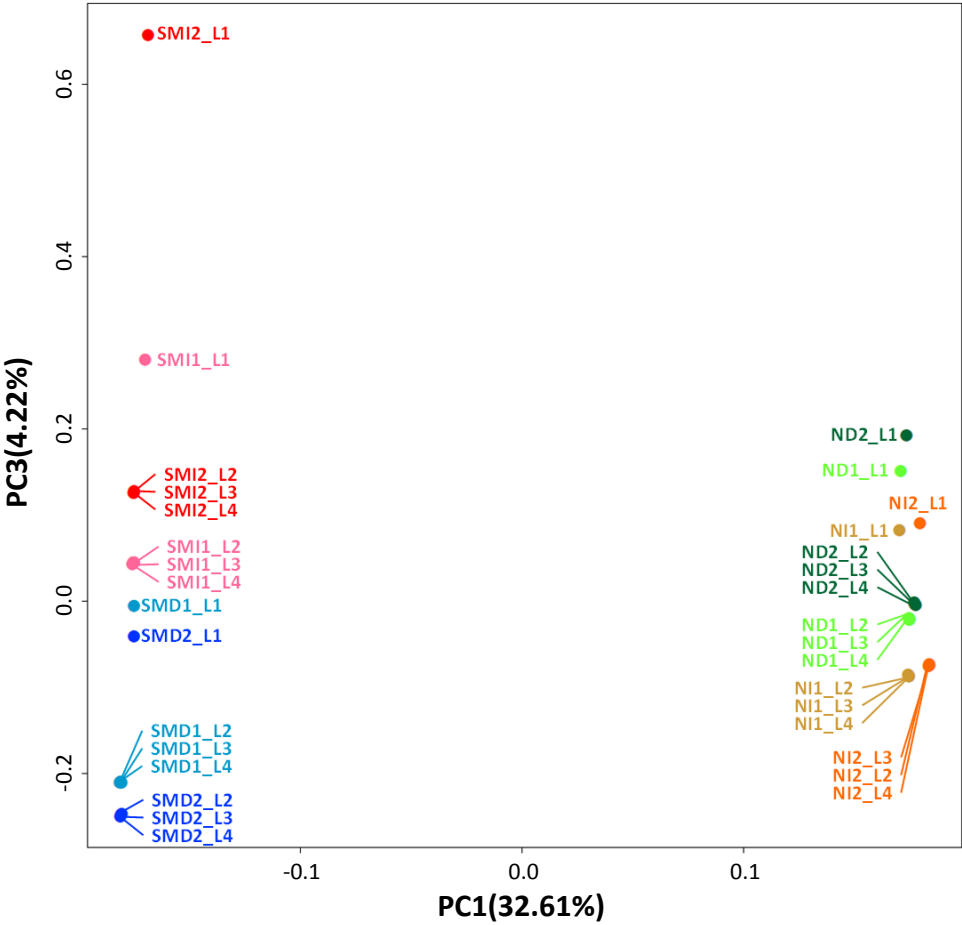
Supplementary Table S2

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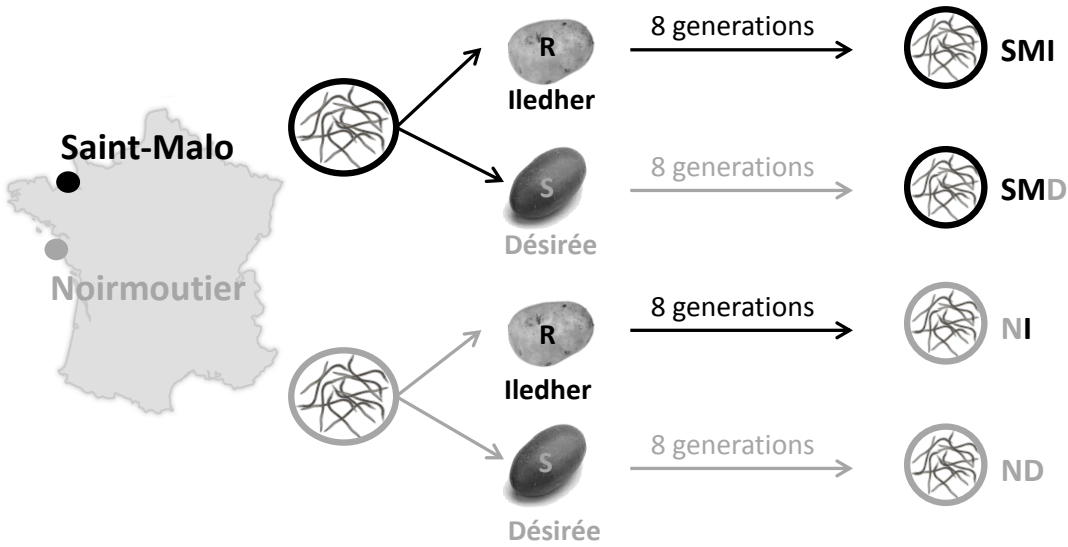lts of the BLAST search of the predicted genes against the annotated *G. rostochiensis* genome version nGr.v1.0 (Eves-van den Akker *et al.* 2016) are also shown. *: outlier SNP located in the gene, **: outlier SNP located in the exon.

Scaffold	Outlier locus	*Globodera pallida* genome				BLAST against *Globodera rostochiensis* genome	
		Gene ID	Location on the scaffold	Putative gene function	Peptide signal	Gene ID	Putative gene function
4	SNP_4_429075	GPLIN_0000 30900	367074-371195	actin protein 6	no	GROS_g1220 8	spliceosome
		GPLIN_0000 31000	371242-372608	nuclear Pore complex protein family member	no		
		GPLIN_0000 31100	398790-400048	transcription initiation factor TFIID subunit	no		
		GPLIN_0000 31200	403834-404255	beta 1 4 endoglucanase	no		
		GPLIN_0000 31300*	428992-438734	transcribed hypothetical protein	no	GROS_g0534 1	magnesium ion transport; sodium ion transmembrane transport; integral to plasma membrane; cation channel activity
		GPLIN_0000 31400	439019-440137	transcribed hypothetical protein	no		
		GPLIN_0000 31500	441856-451123	cysteine synthase	no		
		GPLIN_0000 31600	457740-458980	long chain fatty acid transport protein 4	no	GROS_g1336 3	fatty acid transport proteins
		GPLIN_0000 31700	460388-460633	transcribed hypothetical protein	no		
		GPLIN_0000 31800	462978-464011	beta 1,3 galactosyltransferase 2	no	GROS_g0045 0	galactosyltransferase
		GPLIN_0000 31900	470256-470486	transcribed hypothetical protein	no		
		GPLIN_0000 32000	471353-473130	transcribed hypothetical protein	no		
		GPLIN_0000 32100	474046-474331	transcribed hypothetical protein	no	GROS_g1071 4	protein binding; stress-activated map kinase interacting protein 1 (SIN1)

25	SNP_25_304232	GPLIN_000149900	243326-245017	solute carrier family 35 member B1	yes	GROS_g02151	transport
		GPLIN_000150000	245308-249627	mitogen activated protein kinase kinase kinase	no	GROS_g02150	MAPK signaling pathway
		GPLIN_000150100	251276-252209	superoxide dismutase (Cu Zn)	yes	GROS_g02149	peroxisome
		GPLIN_000150200	252398-253488	transcribed hypothetical protein	no	GROS_g02148	transmembrane
		GPLIN_000150300	254217-258092	transcribed hypothetical protein	no	GROS_g02146	RNA transport; negative regulation of DNA damage checkpoint
		GPLIN_000150400	258929-261360	T complex protein 1 subunit epsilon	no	GROS_g02145	cytoskeleton organization; posttranslational protein folding
		GPLIN_000150500	261650-263494	transcribed hypothetical protein	no	GROS_g14282	metabolism of xenobiotics by cytochrome P450
		GPLIN_000150600	263679-265256	protein vertebrate galectins	yes	GROS_g02143	galactose binding; signal transducer activity
		GPLIN_000150700	265891-266885	N alpha acetyltransferase 20	no	GROS_g02141	endocytosis
		GPLIN_000150800	266955-269813	transcribed hypothetical protein	no	GROS_g02140	protein homodimerization activity
		GPLIN_000150900	269952-271262	transcribed hypothetical protein	no	GROS_g02139	transmembrane
		GPLIN_000151000	271403-272406	transcribed hypothetical protein	no		
		GPLIN_000151100	273509-275230	fasciculation and elongation protein zeta 2	no	GROS_g02137	axon guidance
		GPLIN_000151200	275973-277291	protein kinase domain containing protein, ck worm protein kinase	no		
		GPLIN_000151300	279494-280946	calmodulin	no	GROS_g02136	calmodulin
		GPLIN_000151400	284171-284443	transcribed hypothetical protein	no		
		GPLIN_000151500	299966-302277	cyclin-dependent kinase inhibitor	no		
		GPLIN_000151600*	303454-310418	exportin 7	no	GROS_g02131	nuclear export signal receptor activity; protein export from nucleus; nuclear pore
		GPLIN_000151700	310648-311820	glutaredoxin 3	no	GROS_g02130	glutaredoxin-3
		GPLIN_000151800	313085-314369	transcribed hypothetical protein	no		
		GPLIN_000151900	315934-316953	transcribed hypothetical protein	no	GROS_g02129	transmembrane
		GPLIN_000152000	317223-324014	DEAD Box protein	no	GROS_g02128	ATP catabolic process

		GPLIN_000152100	328566-329000	transcribed hypothetical protein	no		
		GPLIN_000152200	329118-330907	transcribed hypothetical protein	no	GROS_g02126	chromatin binding
		GPLIN_000152300	335501-336396	immunoglobulin i set domain containing protein	yes	GROS_g02125	transmembrane
		GPLIN_000152400	336805-340764	transcribed hypothetical protein	no	GROS_g02124	Golgi apparatus
		GPLIN_000152500	341193-341716	transcribed hypothetical protein	no		
- -							
		GPLIN_000232300	163392-163732	transcribed hypothetical protein	no		
		GPLIN_000232400	168518-169562	transcribed hypothetical protein	yes	GROS_g04085	signal peptide
		GPLIN_000232500	169991-170378	transcribed hypothetical protein	no		
		GPLIN_000232600	171849-174837	transcribed hypothetical protein	yes	GROS_g04087	signal peptide
		GPLIN_000232700	180474-181761	phosphatidylcholine:ceramide	no	GROS_g04089	sphingolipid metabolism
		GPLIN_000232800	183265-184828	60S ribosomal protein L4	no	GROS_g04090	cell wall modification; ribosomal protein L1e signature
		GPLIN_000232900	184937-188002	palmitoyltransferase ZDHHC2	no	GROS_g04091	proteinS-acyltransferase
		GPLIN_000233000	196131-198830	transcribed hypothetical protein	no	GROS_g11539	Transmembrane
		GPLIN_000233100	198931-200169	transcribed hypothetical protein	no		
		GPLIN_000233200	200212-201447	transcribed hypothetical protein	no		
	SNP_44_221212	GPLIN_000233300	201499-202356	transcribed hypothetical protein	no		
44		GPLIN_000233400	202494-203706	transcribed hypothetical protein	no		
	SNP_44_221214	GPLIN_000233500	203859-205030	worm specific Argonaute NRDE 3	no		
		GPLIN_000233600	205120-205935	transcription factor TFIID, C-terminal DNA glycosylase, N-terminal	no		
		GPLIN_000233700	206066-211386	transcribed hypothetical protein	no		
		GPLIN_000233800	211708-212533	transcribed hypothetical protein	no		
		GPLIN_000233900	212735-214448	transcribed hypothetical protein	no		

		GPLIN_0002 34000	221431-221718	transcribed hypothetical protein	no		
		GPLIN_0002 34100	221849-222821	polynucleotide kinase 3' phosphatase	no		
		GPLIN_0002 34200	255627-261288	macrophage erythroblast attacher	yes	GROS_g0323 9	CTLH/CRA C-terminal to LisH motif domain; actin cytoskeleton; negative regulation of myeloid cell apoptotic process
		GPLIN_0002 34300	264876-267441	mitochondrial processing peptidase beta subunit	no	GROS_g0323 5	insulinase (peptidase family M16)
		GPLIN_0002 34400	268640-269749	U1 small nuclear ribonucleoprotein A	no	GROS_g0323 4	regulation of alternative nuclear mRNA splicing, via spliceosome
		GPLIN_0002 34500	270323-271985	guanine nucleotide binding protein subunit	no	GROS_g0323 3	small ribosomal subunit; ion channel inhibitor activity
		GPLIN_0002 34600	272436-274181	transcribed hypothetical protein	no	GROS_g0323 2 / GROS_g0323 8	mitochondrion; negative regulation of synaptic transmission, glutamatergic / ATP catabolic process; adenosinetriphosphatase
		GPLIN_0002 34700	274837-276164	receptor expression enhancing protein 5	no	GROS_g0323 1	protein binding
		GPLIN_0002 34800	276256-277074	G protein coupled receptor associated sorting	no		
		GPLIN_0002 34900	277860-282940	4 aminobutyrate aminotransferase, mitochondrial	no	GROS_g0323 0 / GROS_g0323 0	beta-Alanine metabolism / beta-Alanine metabolism
66	SNP_ 66_ 55507	GPLIN_0003 11400	1-3999	transcribed hypothetical protein	no	GROS_g1278 4	signal peptide
		GPLIN_0003 11500	19418-20911	transcribed hypothetical protein	no	GROS_g1426 0	RBP-4 protein [*G. pallida*]
		GPLIN_0003 11600	30359-32404	transcribed hypothetical protein	no	GROS_g1419 6	RBP-4 protein [*G. pallida*]; secreted SPRY domain-containing protein 18 [*G. rostochiensis*]
		GPLIN_0003 11700	36477-37252	transcribed hypothetical protein	no		
		GPLIN_0003 11800	45302-45915	transcribed hypothetical protein	no	GROS_g0402 3	riboflavin metabolism
	SNP_ 66_ 131126	GPLIN_0003 11900	46907-48176	BTB:POZ domain containing protein 3	yes	GROS_g1412 8	RBP-1 protein, partial [*G. pallida*]
		GPLIN_0003 12000	49696-50448	transcribed hypothetical protein	no	GROS_g1413 6	RBP-4 protein [*G. pallida*]; transmembrane
		GPLIN_0003 12100	53144-55475	protein containing SPRY domain	no	GROS_g1427 8	RBP-1 protein [*G. pallida*]
		GPLIN_0003 12200	57585-58590	transcribed hypothetical protein	no	GROS_g1118 9	peptidase family M41
		GPLIN_0003 12300	66233-69356	protein containing SPRY domain	no	GROS_g1413 0	truncated secreted SPRY domain-containing protein 15, partial [*G. rostochiensis*]; signal peptide

		GPLIN ID	Coordinates	Description		GROS ID	GROS Description
		GPLIN_0003 12400	71724-72660	transcribed hypothetical protein	no		
		GPLIN_0003 12500	92978-94740	protein containing SPRY domain	no	GROS_g1427 8	RBP-1 protein [*G. pallida*]
		GPLIN_0003 12600	95754-96913	protein containing SPRY domain	no	GROS_g1428 7	truncated secreted SPRY domain-containing protein 15, partial [*G. rostochiensis*]
		GPLIN_0003 12700	107749-107949	EF-hand 2 domain containing protein	no	GROS_g1428 1	EF-hand calcium-binding domain.
		GPLIN_0003 12800	111831-112806	transcribed hypothetical protein	no	GROS_g0572 9	protein SAX-7 ; transmembrane
		GPLIN_0003 12900	112920-115358	group 1 glycosyl transferase	no	GROS_g0172 7	glycosyl transferases group 1
		GPLIN_0003 13000	119091-120888	transcribed hypothetical protein	no		
		GPLIN_0003 13100	124022-124830	transcribed hypothetical protein	no		
		GPLIN_0003 13200* (SNP_66_13 1126)	129063-137156	Sensory AXon guidance family member (sax 7)	no	GROS_g0572 9	protein SAX-7; sensory AXon guidance ; transmembrane
		GPLIN_0003 13300	139135-140624	transcribed hypothetical protein	no		
		GPLIN_0003 13400	142713-144269	transcribed hypothetical protein	no		
		GPLIN_0003 13500	145772-146732	transcribed hypothetical protein	no	GROS_g1327 6	signal peptide
		GPLIN_0003 13600	154523-157608	beta 1,4 endoglucanase (cell wall modifying protein); putative GH5 cellulase (cellulose degradation)	yes	GROS_g0744 6	beta-1,4-endoglucanase, partial; cellulase
		GPLIN_0003 13700	158653-160495	transcribed hypothetical protein	no	GROS_g0210 5	poly(A) polymerase central domain; nuclear mRNA splicing, via spliceosome
		GPLIN_0003 13800	167056-168124	transcribed hypothetical protein	no	GROS_g0744 7	EB module
		GPLIN_0003 13900	168962-169322	transcribed hypothetical protein	no	GROS_g0112 6	transmembrane
		GPLIN_0003 14000	183685-184563	transcribed hypothetical protein, similar to *G. rostochiensis* effector 1106	yes	GROS_g1430 9	signal peptide; 1106 effector family [*G. rostochiensis*]
85	SNP_ 85_ 126066	GPLIN_0003 82000	67804-68769	glutaminyl peptide cyclotransferase	no	GROS_g0737 4	signal peptide; peptidase M28 domain containing protein [*Haemonchus contortus*]
		GPLIN_0003 82100	84608-86814	transcribed hypothetical protein	no	GROS_g1412 4	RBP-4 protein [*G. pallida*]
	SNP_ 85_ 126206	GPLIN_0003 82200	94832-100930	zinc finger protein	no	GROS_g1133 6	zinc finger protein [*Loa loa*]
		GPLIN_0003 82300	110433-110747	transcribed hypothetical protein	no		

		GPLIN ID	Position	Description	Secreted	GROS ID	GROS Description
		GPLIN_0003 82400	120748-122590	BTB:POZ domain containing protein 3	no	GROS_g1419 3	truncated secreted SPRY domain-containing protein 15, partial [*G. rostochiensis*]
		GPLIN_0003 82500	126792-129853	SPla RYanodine receptor SPRY	no	GROS_g1416 3	RBP-4 protein [*G. pallida*]
		GPLIN_0003 82600	142203-143446	transcribed hypothetical protein	no		
		GPLIN_0003 82700	143942-144891	transcribed hypothetical protein	no		
		GPLIN_0003 82800	151467-154962	E9 protein	no	GROS_g1412 3	signal peptide
		GPLIN_0003 82900	164278-168506	glutathione synthetase	no	GROS_g0539 0 / GROS_g1379 7	glutathione synthetase-like [*Maylandia zebra*] / glutathionesynthase; signal peptide
		GPLIN_0003 83000	169444-172268	Guanine nucleotide binding protein G(o) subunit	no	GROS_g0538 9	guanine nucleotide-binding protein alpha-3 subunit [*Brugia malayi*]
		GPLIN_0003 83100	174565-177704	ODR 3	no	GROS_g0538 9	guanine nucleotide-binding protein alpha-3 subunit [*B. malayi*]
		GPLIN_0003 83200	179602-190327	guanine nucleotide exchange factor for Ras	no	GROS_g0538 8	regulation of Ral GTPase activity
111	SNP_111_ 140151	GPLIN_0004 50000	79896-81897	UDP glucuronosyltransferase	yes	GROS_g0933 7	integral to plasma membrane
		GPLIN_0004 50100	94652-95140	hypothetical protein	no		
		GPLIN_0004 50200	98715-99013	transcribed hypothetical protein	no		
		GPLIN_0004 50300	120569-121060	hypothetical protein	no		
		GPLIN_0004 50400	126241-129974	protein containing SPRY domain	no	GROS_g1419 5	dendrite morphogenesis, truncated secreted SPRY domain-containing protein 15, partial [*G. rostochiensis*]
		GPLIN_0004 50500	132676-135723	inversin protein alternative, ankyrin repeat protein	no	GROS_g1419 2	protein binding
		GPLIN_0004 50600	133821-134846	BTB:POZ domain containing protein 3	no	GROS_g1419 5	dendrite morphogenesis, truncated secreted SPRY domain-containing protein 15, partial [*G. rostochiensis*]
		GPLIN_0004 50700	138326-138779	transcribed hypothetical protein	no		
		GPLIN_0004 50800	138941-139562	Nematode AStacin protease family member	no	GROS_g0742 2	molting cycle, collagen and cuticulin-based cuticle; Astacin
		GPLIN_0004 50900	144927-150255	NAD kinase domain containing protein 1	no	GROS_g1280 7 / GROS_g1280 8	NAD metabolic process / mRNA surveillance pathway
		GPLIN_0004 51000	153798-155515	transcribed hypothetical protein	no		

		GPLIN_0004 51100	156162-159524	Amino Acid Transporter family member (aat 5)	no	GROS_g12805 / GROS_g00275	transmembrane transport / protein digestion and absorption
		GPLIN_0004 51200	160262-161491	cyclin B	no	GROS_g04262	microtubule cytoskeleton
		GPLIN_0004 51300	162184-162860	transcribed hypothetical protein	no	GROS_g04261	bifunctional 3 -phosphoadenosine 5 -phosphosulfate synthase [*Ascaris suum*]
		GPLIN_0004 51400	163767-166872	serine:threonine protein phosphatase 5	no	GROS_g04260	transcription, DNA-dependent
		GPLIN_0004 51500	168900-173567	transcribed hypothetical protein	no	GROS_g04258	protein binding
		GPLIN_0004 51600	174095-175144	40S ribosomal protein S26	no	GROS_g04257	RNA metabolic process
		GPLIN_0004 51700	175416-177669	transcribed hypothetical protein	no	GROS_g04256	transmembrane
		GPLIN_0004 51800	177949-181975	DNA replication licensing factor MCM6	no	GROS_g04255	cell cycle - yeast
		GPLIN_0004 51900	182477-187262	SWI:SNF complex subunit SMARCC2	no	GROS_g04254	ATP catabolic process
		GPLIN_0004 52000	187774-189157	DNA replication complex GINS protein SLD5	no	GROS_g04253	nucleoplasm
		GPLIN_0004 52100	192011-196045	transcribed hypothetical protein	no	GROS_g04252	transmembrane
		GPLIN_0004 52200	197876-201179	transcribed hypothetical protein	no	GROS_g10891	galactosyl beta-1,3 N-acetylgalactosamine beta-1,3-glucuronosyltransferase activity
123	SNP_123_97233	GPLIN_0004 82300	35198-37842	KH domain containing protein	no	GROS_g06084	nematode larval development; masculinization of hermaphroditic germ-line
		GPLIN_0004 82400	59414-60834	transcribed hypothetical protein	no		
		GPLIN_0004 82500	61285-62959	nematode cuticle collagen N terminal domain	no	GROS_g11113	larval development ; transmembrane
		GPLIN_0004 82600	72586-74133	transcribed hypothetical protein	yes		
		GPLIN_0004 82700	79061-79808	transcribed hypothetical protein	no	GROS_g13551	signal peptide
		GPLIN_0004 82800	85115-87297	mucosa associated lymphoid tissue lymphoma	no		
		GPLIN_0004 82900	87613-89268	mucosa associated lymphoid tissue lymphoma	no		
		GPLIN_0004 83000	89592-91491	mucosa associated lymphoid tissue lymphoma	no	GROS_g11117	ubiquitin-protein ligase activity; peptidase activity
		GPLIN_0004 83100	98509-100826	sucrose hydrolase	no	GROS_g11374	beta-fructofuranosidase; signal peptide

41

		GPLIN	Coordinates	Description	Secreted	GROS	Description
		GPLIN_0004 83200	101689-102752	tail length tape measure protein	no		
		GPLIN_0004 83300	103018-107209	beta fructofuranosidase; secreted GH32 fructosidase	yes	GROS_g1139 7	beta-fructofuranosidase; plant-type cell wall modification
		GPLIN_0004 83400	108777-111065	transcribed hypothetical protein	yes		
		GPLIN_0004 83500	112509-112790	transcribed hypothetical protein	no		
		GPLIN_0004 83600	117038-117476	transcribed hypothetical protein	yes		
		GPLIN_0004 83700	119298-120112	mps one binder kinase activator 3 like	no	GROS_g0971 1	neuronal cell body; perinuclear region of cytoplasm; microtubule cytoskeleton organization
		GPLIN_0004 83800	120304-121636	ubiquitin interacting motif	no	GROS_g0971 2	perinuclear region of cytoplasm
		GPLIN_0004 83900	121782-123696	ATP synthase subunit beta, mitochondrial	no	GROS_g0971 3	oxidative phosphorylation
		GPLIN_0004 84000	123896-127752	protein arginine N methyltransferase 5	no	GROS_g0971 4	protein ubiquitination; negative regulation of mitotic cell cycle
		GPLIN_0004 84100	128956-130197	tryptophanyl tRNA synthetase	no		
		GPLIN_0004 84200	137371-139503	transcribed hypothetical protein	no	GROS_g0970 7	transmembrane
		GPLIN_0004 84300	141658-142945	receptor type tyrosine protein phosphatase eta	no	GROS_g0503 2	protein tyrosine phosphatase activity; salicylic acid mediated signaling pathway
		GPLIN_0004 84400	148275-154307	transcribed hypothetical protein	no	GROS_g0502 8	glycerophospholipid metabolism; phosphatidylcholine metabolic process; signal peptide
		GPLIN_0004 84500	154735-155098	transcribed hypothetical protein	no		
		GPLIN_0004 84600	155919-159554	mitotic checkpoint serine:threonine protein	no	GROS_g0428 7 / GROS_g0502 7	regulation of chromosome segregation; positive regulation of intrinsic apoptotic signaling pathway / mitotic cell cycle spindle assembly checkpoint
182	SNP_ 182_ 414	GPLIN_0006 26500	1034-2234	protein containing SPRY domain	no	GROS_g1424 1	regulation of Ran GTPase activity; RBP-1 protein [G.*pallida*]
		GPLIN_0006 26600	4596-5610	transcribed hypothetical protein	no		
		GPLIN_0006 26700	9510-11116	protein containing SPRY domain	no	GROS_g1418 4	regulation of Ran GTPase activity; RBP-4 protein [*G. pallida*]
	SNP_ 182_ 557	GPLIN_0006 26800	13854-21499	protein containing SPRY domain	no	GROS_g1419 6	cytoplasmic microtubule; RBP-4 protein [*G. pallida*]
		GPLIN_0006 26900	23491-25804	protein containing SPRY domain	no	GROS_g1419 5	truncated secreted SPRY domain-containing protein 15, partial [*G. rostochiensis*]
		GPLIN_0006 27000	27744-30562	transcribed hypothetical protein	no		

		GPLIN_0006 27100	41665-44164	paralog of RBP-1 protein	yes	GROS_g14256	ran GTPase binding; RBP-1 protein, partial [*G. pallida*]
		GPLIN_0006 27200	45726-46823	60S acidic ribosomal protein P1	no	GROS_g01370	ribosome
		GPLIN_0006 27300	49760-52726	transcribed hypothetical protein	no		
		GPLIN_0006 27400	55950-61422	serpin protein	yes	GROS_g09868	extracellular space
		GPLIN_0006 91700	593-1233	transcribed hypothetical protein	no		
		GPLIN_0006 91800	3867-5051	dorsal gland cell specific expression protein	no	GROS_g14298	dorsal gland cell-specific expression protein [*Heterodera avenae*]
		GPLIN_0006 91900	9497-12258	mimitin, mitochondrial	no	GROS_g05274	NADH ubiquinone oxidoreductase subunit NDUFA12
		GPLIN_0006 92000	13158-18286	succinyl CoA:3 ketoacid coenzyme A transferase	no	GROS_g05277 / GROS_g02487	butanoate metabolism / butanoate metabolism
		GPLIN_0006 92100	18536-19094	histone cluster 2, H3c2	no	GROS_g05278	protein complex
		GPLIN_0006 92200	19401-19965	H2A histone family, member X	no	GROS_g10756	positive regulation of DNA repair
		GPLIN_0006 92300	21646-27499	transcribed hypothetical protein	no	GROS_g02034	small GTPase mediated signal transduction
216	SNP_216_ 53947	GPLIN_0006 92400	27752-30375	transcribed hypothetical protein	no		
		GPLIN_0006 92500*	31556-57322	copine domain containing protein 2	no	GROS_g02032	plasma membrane
		GPLIN_0006 92600	57566-59215	transcribed hypothetical protein	no	GROS_g02032	plasma membrane
		GPLIN_0006 92700	66169-72222	solute carrier family 23	no	GROS_g02031 / GROS_g03645	nucleobase transport / plasma membrane
		GPLIN_0006 92800	78865-80575	serine:arginine rich splicing factor 4	no	GROS_g02030	mitosis
		GPLIN_0006 92900	83168-85506	T complex protein 1 subunit gamma	no	GROS_g02028	extracellular vesicular exosome
		GPLIN_0006 93000	85945-87738	T complex protein 1 subunit gamma	no	GROS_g02028	extracellular vesicular exosome
		GPLIN_0006 93100	88182-90397	heterogeneous nuclear ribonucleoprotein A1	no	GROS_g02026	spliceosome
		GPLIN_0006 93200	91404-93983	transcribed hypothetical protein	no	GROS_g02025	phosphatidylinositol-4,5-bisphosphate4-phosphatase.

43

		GPLIN_0006 93300	95369-98967	transcribed hypothetical protein	no	GROS_g0202 4	plasmin
		GPLIN_0006 93400	100056-103180	nicotinic acetylcholine receptor non alpha	yes	GROS_g0202 2	response to inorganic substance
		GPLIN_0006 93500	104161-105279	transcribed hypothetical protein	no	GROS_g0202 1	hypothetical protein Y032_0018g3584 [*Ancylostoma ceylanicum*]
		GPLIN_0006 93600	105754-107423	bifunctional aminoacyl tRNA synthetase	no	GROS_g0202 0	porphyrin and chlorophyll metabolism
		GPLIN_0006 93700	107723-110938	u3 small nucleolar rna associated protein 14	no	GROS_g0201 9	ribosome biogenesis in eukaryotes
		GPLIN_0006 93800	111764-113032	succinate dehydrogenase cytochrome b560 subunit	no	GROS_g0201 8	defense response to Gram-negative bacterium
		GPLIN_0006 93900	113418-114209	'cold shock' DNA binding domain containing protein	no	GROS_g0201 7	sequence-specific DNA binding
283	SNP_ 283_ 56739	GPLIN_0008 14000*	56408-57804	BTB:POZ domain containing protein At1g55760	no		
		GPLIN_0008 14100	58370-60428	BTB:POZ domain containing protein At1g55760	no		
		GPLIN_0008 14200	61441-67930	transcribed hypothetical protein	yes	GROS_g1355 9	signal peptide
		GPLIN_0008 14300	68894-70678	BTB:POZ domain containing protein At1g55760	no		
		GPLIN_0008 14400	72618-73953	transcribed hypothetical protein	no		
		GPLIN_0008 14500	75558-76793	BTB:POZ domain containing protein At1g55760	no		
	SNP_ 283_ 56740	GPLIN_0008 14600	78642-79828	transcribed hypothetical protein	yes	GROS_g0578 8	signal peptide
		GPLIN_0008 14700	84623-91635	zinc metallopeptidase	no	GROS_g0578 9	metalloendopeptidase activity; peptidase family M13
		GPLIN_0008 14800	95043-97768	myelin transcription factor 1 protein	no	GROS_g1429 5	zinc finger, C2HC type
		GPLIN_0008 14900	107592-114130	transcribed hypothetical protein	no	GROS_g1421 3	BTB/POZ domain
		GPLIN_0008 15000	115751-116921	G patch domain containing protein	no		
296	SNP_296_ 8846	GPLIN_0008 32100**	6442-11746	Na:H eXchanger family member (nhx 9)	no	GROS_g0464 1 / GROS_g1341 8	protein F32B5.6, isoform j [*Caenorhabditis elegans*] / sodium:hydrogen antiporter activity
		GPLIN_0008 32200	12647-13096	transcribed hypothetical protein	no		
		GPLIN_0008 32300	17388-19554	transcribed hypothetical protein	no		

		GPLIN_0008 32400	33723-34555	transcribed hypothetical protein	no		
		GPLIN_0008 32500	35455-36989	transcribed hypothetical protein; Gpa_Dog_0189	yes		
		GPLIN_0008 32600	44182-46613	transcribed hypothetical protein	yes	GROS_g0830 7	lysosomal membrane
		GPLIN_0008 32700	47366-49673	lysosome associated membrane glycoprotein	yes	GROS_g0830 7	lysosomal membrane
		GPLIN_0008 32800	50657-51448	transcribed hypothetical protein	no	GROS_g0830 8	aminoacyl-tRNAhydrolase.
		GPLIN_0008 32900	52058-53243	14-3-3 protein	no	GROS_g0596 3 / GROS_g0830 9	neurotrophin signaling pathway / neurotrophin signaling pathway
		GPLIN_0008 33000	54914-56106	14-3-3 protein	no	GROS_g0596 3 / GROS_g0830 9	neurotrophin signaling pathway / neurotrophin signaling pathway
		GPLIN_0008 33100	63249-68201	calcium binding EGF domain containing protein	no	GROS_g0831 0	hypothetical protein Y032_0091g2492 [*Ancylostoma ceylanicum*]
338	SNP_338_ 98803	GPLIN_0009 03300	71992-78704	transcribed hypothetical protein	no		
		GPLIN_0009 03400	86045-87892	transcribed hypothetical protein	no	GROS_g0349 0	response to chitin
		GPLIN_0009 03500	89309-92938	serpentine receptor, class T family member	no	GROS_g0664 5	transmembrane
		GPLIN_0009 03600**	97414-99406	dorsal gland cell specific expression protein	no	GROS_g1414 6 / GROS_g1422 4	dorsal gland cell-specific expression protein [*Heterodera avenae*] / transmembrane; dorsal gland cell-specific expression protein [*H. avenae*]
		GPLIN_0009 03700	102236-103533	protein MICAL 3	no	GROS_g0815 7	LIM domain
		GPLIN_0009 03800	104125-104725	LIM domain containing protein	no		
		GPLIN_0009 03900	105988-106776	transcribed hypothetical protein	no		
		GPLIN_0009 04000	107274-107909	protein MICAL 3	no		
		GPLIN_0009 04100	108895-109595	transcribed hypothetical protein	no		
340	SNP_ 340_ 56189	GPLIN_0009 04700	24813-25894	transcribed hypothetical protein	no	GROS_g0802 5	SSF52540
		GPLIN_0009 04800	35241-37595	nematode astacin protease family member	no	GROS_g0703 5	extracellular region; metalloendopeptidase activity; Astacin-like metalloendopeptidase [*Strongyloides ratti*]

		GPLIN_0009 04900	37694-40604	peptidase M12A, astacin	no	GROS_g12919	astacin (peptidase family M12A)
		GPLIN_0009 05000	42995-45732	cell death protein 3	no	GROS_g12555	execution phase of apoptosis
		GPLIN_0009 05100**	53759-57940	transcribed hypothetical protein	no		
		GPLIN_0009 05200	75502-76417	transcribed hypothetical protein	no		
349	SNP_349_ 8349	GPLIN_0009 14300	13395-18410	zinc finger protein 280D	no	GROS_g11336 / GROS_g02450 / GROS_g11223	zinc finger protein [*L. loa*] / signal peptide / zinc finger C2H2 type domain signature
		GPLIN_0009 14400	31215-32425	transcribed hypothetical protein	no		
		GPLIN_0009 14500	63737-66258	transcribed hypothetical protein	no	GROS_g10058	signal peptide
		GPLIN_0009 14600	69418-70547	ubiquitin carboxyl terminal hydrolase 48	no	GROS_g03356	ubiquitinylhydrolase1
474	SNP_ 474_ 28528	GPLIN_0010 56200	3568-10018	transcribed hypothetical protein	no	GROS_g11538	zinc finger C2H2 type domain signature.
		GPLIN_0010 56300	10515-10989	transcribed hypothetical protein	no	GROS_g11539	transmembrane
		GPLIN_0010 56400	11106-12182	transcribed hypothetical protein	yes	GROS_g11539	transmembrane
		GPLIN_0010 56500	13710-14450	transcribed hypothetical protein	no		
		GPLIN_0010 56600	14570-20232	transcribed hypothetical protein	no	GROS_g11540	signal peptide
		GPLIN_0010 56700	52073-52540	transcribed hypothetical protein; Gpa_Dog_0074	yes		
		GPLIN_0010 56800	56482-58809	transcribed hypothetical protein	no		
		GPLIN_0010 56900	58931-60252	transcribed hypothetical protein	yes		
		GPLIN_0010 57000	62667-67109	transcribed hypothetical protein	no		
504	SNP_ 504_ 47039	GPLIN_0010 82800	211-620	protein containing SPRY domain	no	GROS_g14163	RBP-4 protein [*G. pallida*]
		GPLIN_0010 82900	11855-12808	paralog of RBP-5 protein (33H17)	no	GROS_g14179	RBP-1 [*G. pallida*]; signal peptide
		GPLIN_0010 83000	14667-16919	transcribed hypothetical protein	yes	GROS_g13008	hypothetical protein TcasGA2_TC001495 [*Tribolium castaneum*]

		GPLIN_0010 83100	18255-23950	transcribed hypothetical protein	yes	GROS_g1045 6	thap domain-containing protein 4 [*A. suum*]; transmembrane
		GPLIN_0010 83200	25548-26905	transcribed hypothetical protein	no		
		GPLIN_0010 83300	32640-39130	serine:threonine protein kinase SIK3	no	GROS_g1045 4	serine threonine-protein kinase kin-29 [*A. suum*]
		GPLIN_0010 83400*	43996-50149	3 hydroxy 3 methylglutaryl coenzyme A reductase	no	GROS_g1045 2	hmg CoA reductase A [*Polysphondylium pallidum* PN500]; protein homodimerization activity
		GPLIN_0010 83500	54107-59069	transcribed hypothetical protein	no	GROS_g1045 1	peptidase M10A M12B domain containing protein [*H. contortus*]; transmembrane; matrixin
782	SNP_ 782_ 20136	GPLIN_0012 57900	8259-9831	transcribed hypothetical protein	no	GROS_g1414 6	dorsal gland cell-specific expression protein [*H. avenae*]; putative esophageal gland cell protein Hgg-20 [*H. glycines*]
		GPLIN_0012 58000	16237-17191	transcribed hypothetical protein	no	GROS_g1421 3	BTB/POZ domain
		GPLIN_0012 58100**	19137-20558	paralog of RBP-1 protein; Gpa_Dog_0077	yes	GROS_g1423 4	RBP-1 protein [*G. pallida*]; secreted SPRY domain-containing protein 16, partial [*G. rostochiensis*]
		GPLIN_0012 58200	25107-27756	ubiquitin 40S ribosomal protein S27a 1	no	GROS_g0756 8	protein ubiquitination
		GPLIN_0012 58300	31187-31435	transcribed hypothetical protein	no		
		GPLIN_0012 58400	33199-34624	protein containing SPRY domain	no	GROS_g1418 4	RBP-4 protein [*G. pallida*]; secreted SPRY domain-containing protein 9 [*G. rostochiensis*]
988	SNP_ 988_ 7923	GPLIN_0013 23400	2535-3808	transcribed hypothetical protein	no	GROS_g0156 8	protein of unknown function, DUF273
		GPLIN_0013 23500	12021-16917	serine:threonine protein kinase Nek6	no	GROS_g0133 4	serine/threonine-protein kinase Nek6, partial [*Bos mutus*]; cytokinesis; signal transducer activity
	SNP_ 988_ 8037	GPLIN_0013 23600	19109-26588	Motor AXon guidance family member (max 2)	no	GROS_g1007 7	protein BM-MAX-2, isoform i [*B. malayi*]; K04409 Axon guidance
	SNP_ 988_ 8045						
	SNP_ 988_ 8100						
1243	SNP_1243 _9865			no predicted gene			
1777	SNP_1777 _961	GPLIN_0014 38500**	864-1181	transcribed hypothetical protein	no		
		GPLIN_0014 38600	2326-4922	maternal protein pumilio	no	GROS_g1428 6	cytosol
		GPLIN_0014	6616-7313	transcriptional regulator ATRX	no		

		38700					
3221	SNP_3221_1475	GPLIN_001518200	1511-2451	dorsal gland cell specific expression protein	no	GROS_g14143 / GROS_g14139	transmembrane; dorsal gland cell-specific expression protein [*H. avenae*] / transmembrane; dorsal gland cell-specific expression protein [*H. avenae*]
3816	SNP_3816_941			no predicted gene			
5159	SNP_5159_970			no predicted gene			

811