



**HAL**  
open science

# Experimentally evolved populations of the potato cyst nematode *Globodera pallida* allow the targeting of genomic footprints of selection due to host adaptation

Delphine Eoche-Bosy, Jean-Pierre Gauthier, A. S. Juhel, Magali Esquibet, Sylvain Fournet, Eric Grenier, Josselin Montarry

## ► To cite this version:

Delphine Eoche-Bosy, Jean-Pierre Gauthier, A. S. Juhel, Magali Esquibet, Sylvain Fournet, et al.. Experimentally evolved populations of the potato cyst nematode *Globodera pallida* allow the targeting of genomic footprints of selection due to host adaptation. *Plant Pathology*, 2017, 66 (6), pp.1022-1030. 10.1111/ppa.12646 . hal-02625021

**HAL Id: hal-02625021**

**<https://hal.inrae.fr/hal-02625021>**

Submitted on 9 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Experimentally evolved populations of the potato cyst nematode *Globodera***  
2 ***pallida* allow the targeting of genomic footprints of selection due to host**  
3 **adaptation**

4

5 D. EOCHE-BOSY, J. GAUTHIER<sup>1</sup>, A.S. JUHEL<sup>2</sup>, M. ESQUIBET, S. FOURNET, E.  
6 GRENIER and J. MONTARRY\*

7

8 *IGEPP, INRA, Agrocampus Ouest, Université de Rennes 1, 35650 Le Rheu, France.*

9

10 \*Correspondence: Josselin Montarry, josselin.montarry@inra.fr

11

12 Running head: Footprints of selection in *G. pallida*

13

14 **Keywords:** experimental evolution, genome scan, outlier, resistance, virulence

15

16

---

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

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

17 **ABSTRACT**

18 In the current agronomical context of pesticide use reduction, deciphering the genetic bases of  
19 pathogen adaptation to plant resistances is of major importance to improve their durability.  
20 Knowledge of the virulence gene frequencies in pathogens populations could indeed allow the  
21 prediction of their durability before deployment. *Globodera pallida* is a major pest of potato  
22 crop for which a promising resistance QTL, *GpaV<sub>vm</sub>*, has been identified in *Solanum vernei*.  
23 An experimental evolution study, in which *G. pallida* lineages evolved on resistant or  
24 susceptible potato genotypes for up to eight generations, previously showed that *G. pallida*  
25 was able to rapidly overcome *GpaV<sub>vm</sub>* resistance. However it was not known if enough  
26 genetic mixing occurred in these lineages to be able to detect islands of differentiation in a  
27 genome scan approach. Here, we investigated this question using 53 polymorphic  
28 microsatellite markers distributed along the genome and three different tests based on genetic  
29 differentiation and heterozygosity. We identified eight outlier loci, indicative of genomic  
30 regions putatively involved in host adaptation. Several loci were identified by multiple  
31 detection methods and/or in two independent adapted lineages. We also showed that some  
32 identified candidate genomic regions seem to be also involved in the overcoming of nematode  
33 resistance in a genotype harbouring the same resistance QTL in a different genetic  
34 background. These results validate the feasibility of a genome scan approach on a biological  
35 material coming from short experimental evolution, and encourage to perform a high  
36 coverage genome scan using whole genome resequencing.

## 37 INTRODUCTION

38 Host-parasite interactions constitute an interesting setting in which biological adaptation to a  
39 new environment, *i.e.* the host, can be studied. The identification of the molecular bases of  
40 host adaptation in parasites becomes a central challenge for human and environmental health  
41 (Aguileta *et al.*, 2009); this is also true for plant parasites, responsible for severe damages in  
42 agricultural crops worldwide.

43 A comprehensive approach to analyse the molecular bases of adaptation is first to  
44 identify genomic regions showing footprints of selection, before investigating more precisely  
45 these regions by searching for candidate genes potentially involved in the adaptation. To do  
46 so, population genomics developed methods based on genome wide variation, called genome  
47 scans, to separate locus-specific effects from genome-wide effects (Luikart *et al.*, 2003; Storz,  
48 2005). Demographic processes affect the genome in its entirety, whereas selection effects are  
49 expected to be locus-specific. Loci potentially involved in selection therefore show a non-  
50 neutral pattern of variation, by exhibiting significantly more differentiation among  
51 populations and lower diversity within populations (Luikart *et al.*, 2003; Storz, 2005). Due to  
52 the process of genetic hitchhiking (Maynard-Smith & Haigh, 1974), markers linked to the  
53 selected loci will be also influenced by selection, leaving a genomic signature which can be  
54 detected by genotyping markers distributed along the genome (Luikart *et al.*, 2003). The  
55 comparison between the distributions of summary statistics estimated from simulations under  
56 the null hypothesis of neutrality and from the dataset allows to detect outlier loci (*i.e.* loci  
57 departing significantly from the neutral distribution). In natural populations, confounding  
58 environmental effects could lead to false positive outliers. Coupling genome scans and  
59 experimental evolution is therefore a promising approach, as it allows to control the selection  
60 pressure imposed on populations and then to establish a direct link between this selection  
61 pressure and the phenotypic and genetic changes observed (Turner *et al.*, 2011). Repeating the

62 search for outliers in different populations allows to strengthen to the presumed implication in  
63 adaptation of outliers found several times (Nosil *et al.*, 2008). Compared to genome scan  
64 approach conducted on natural populations, fewer independent replicates should be then  
65 necessary when using experimental lineages to target the genomic regions impacted by the  
66 selection pressure. However, sufficient genetic mixing should also occur during the  
67 experimental evolution process (Storz, 2005).

68 Plant-parasitic nematodes are major agricultural pathogens causing severe damages in  
69 crops and therefore considerable economic losses (Nicol *et al.*, 2011). The cyst nematode  
70 *Globodera pallida* (Stone) is an important pathogen of potato and is classified as a quarantine  
71 pest on both EPPO and USDA quarantine lists. *G. pallida* is a diploid organism with obligate  
72 sexual reproduction, which achieves one generation per year in European climatic conditions  
73 (Jones, 1950). The use of nematicides has enabled for decades the control of these pathogens.  
74 However, their use is now strictly limited or even prohibited, as they constitute a threat for  
75 environmental and human health (Clayton *et al.*, 2008). The currently privileged alternative to  
76 control *G. pallida* populations is the use of resistant potato cultivars. Because the breeding of  
77 resistant potato genotypes is a long process, only few registered cultivars show a high level of  
78 resistance to *G. pallida*. A promising resistance factor, the QTL *GpaV<sub>vm</sub>* responsible for the  
79 development of most nematodes into adult males, which do not cause symptoms by  
80 themselves, has been introgressed from *Solanum vernei* (a wild species close to the cultivated  
81 potato *S. tuberosum tuberosum*) into several potato genotypes differing by their genetic  
82 background. The resistant parents all harbour QTL *GpaV<sub>vm</sub>*, mapped on the potato  
83 chromosome V and explaining 61% of the resistance in biparental segregating populations  
84 (Roupe van der Voort *et al.*, 2000), and possibly a minor QTL mapped on the chromosome  
85 IX and explaining 24% of the resistance (Roupe van der Voort *et al.*, 2000), but whose  
86 presence was not ascertained (Fournet *et al.*, 2013). One of the resistant genotypes, cultivar

87 Iledher, has been registered in 2009 in the French catalogue as the first cultivar showing a  
88 high level of resistance to *G. pallida*. In an experimental evolution study, Fournet *et al.* (2013)  
89 exposed two independent French *G. pallida* populations to several potato genotypes  
90 harbouring *GpaV<sub>vm</sub>* and differing by their genetic background over eight parasite generations.  
91 They demonstrated that the resistance from *S. vernei* could be overcome by both *G. pallida*  
92 populations, and thus generated lineages adapted to each resistant potato genotype. Moreover,  
93 they showed that the plant genetic background impacts the durability of resistance.

94 In order to design strategies for the sustainable management of plant resistance, it is  
95 necessary to study the evolutionary potential of pathogen populations, including the genetic  
96 bases of their adaptation to resistant cultivars. However, an important limit lies in the scarcity  
97 of molecular data available about *G. pallida* response to selective pressures imposed by host  
98 genotypes. Only one virulence/avirulence gene has been described in *G. pallida*, and is  
99 specific to a particular nematode population (Sacco *et al.*, 2009). Recently, a study based on a  
100 whole genome allelic imbalance analysis of SNP in *Heterodera glycines* populations grown  
101 on resistant and susceptible soybean plants revealed two new candidate virulence genes  
102 (Bekal *et al.*, 2015) and has opened the way for population genomic approach in cyst  
103 nematodes. In this last study investigations were carried out from controlled matings between  
104 *H. glycines* inbred lines which have been maintained in the laboratory for over 30 generations.  
105 With the aim to validate the feasibility of a genome scan approach on a biological material  
106 coming from short experimental evolution, we performed a genome scan on *G. pallida*  
107 lineages virulent and avirulent to Iledher coming from two different populations, using 53  
108 polymorphic microsatellites markers distributed along the genome. We applied three different  
109 methods to identify outlier loci. Moreover, we also compared the candidate genomic regions  
110 identified using lineages adapted to Iledher and to another potato genotype harbouring the  
111 same resistance QTL in a different genetic background.

112

## 113 **MATERIALS AND METHODS**

### 114 **Selection of virulent and avirulent *G. pallida* lineages**

115 Nematode lineages were established from cysts of two French natural *G. pallida* populations,  
116 SM (near Saint-Malo, Brittany, north-western France) and N (from the island of Noirmoutier,  
117 western France) coming from infested fields (3 and 82 J2 per gram of soil for SM and N,  
118 respectively). All details about the experimental evolution protocol were given by Fournet *et*  
119 *al.* (2013). The lineages used here were obtained by rearing both populations during eight  
120 successive cycles on the susceptible potato cultivar Désirée (D) and on two resistant cultivars:  
121 Iledher (I) and the potato genotype 360.96.21 (360), both possessing the resistance QTL  
122 *GpaV<sub>vm</sub>*. Because of the high level of resistance of the potato genotypes used during the  
123 experimental evolution, the three first generations were made directly into the fields, the next  
124 two generations were made in tanks under greenhouse conditions, and the three last  
125 generations in pots. Unfortunately, the intermediate generations as well as several lineages,  
126 including the lineage from the N population evolving on the potato genotype 360.96.21, were  
127 not available anymore. The present study was thus conducted with the eighth generation of  
128 the five remaining available lineages, named hereafter: SMD, SMI, ND, NI and SM360. As a  
129 consequence, the comparison for the adaptation to the same resistance QTL in different  
130 genetic backgrounds (Iledher and genotype 360.96.21) could only be performed with lineages  
131 from the SM population.

132

### 133 **DNA extraction and microsatellite genotyping**

134 For each lineage, 50 cysts from the eighth generation of the experimental evolution protocol  
135 were randomly sampled and individualized. As the amount of DNA in a single larva was not  
136 sufficient to perform the genotyping with the whole set of microsatellites markers, three J2

137 (siblings) per cyst were used and each one was genotyped with a subset of markers (*i.e.* the  
138 three siblings DNA were not pooled). Extraction of genomic DNA of each J2 was performed  
139 following Boucher *et al.* (2013). Presence of DNA and species identity of these samples were  
140 checked by PCR-RFLP of the ITS region as previously described (Boucher *et al.*, 2013).  
141 Samples were genotyped using 202 microsatellite markers. A large part of these markers were  
142 developed directly from one of the first versions of the *G. pallida* genome, composed of  
143 17,095 supercontigs (version gpal.201011.supercontigs.fasta, Sanger Institute:  
144 <http://sanger.ac.uk>). The QDD software (Megléczy *et al.*, 2010) was used to select sequences  
145 and design primer pairs for sequences containing perfect microsatellite motifs with at least  
146 four repeats and flanking regions without tandem repetition. Among those markers, we  
147 discarded small PCR products (<100 nucleotides), and only one marker per supercontig was  
148 retained. Eight supercontig sequences containing pathogenicity genes, *i.e.* genes involved in  
149 the production of molecules interacting with the host plant (Haegeman *et al.*, 2012) (PL1,  
150 PL2, ExpB1, ExpB2 and four paralogs of RBP1) and 3446 ESTs sequences (available on  
151 NEMBASE4: <http://www.nematodes.org/nembase4/>) were also screened for microsatellite  
152 content using the software MSATCOMMANDER v0.8.2 (Faircloth, 2008), which was also used to  
153 design primers. Eleven microsatellite markers developed for *G. rostochiensis*, which cross-  
154 hybridize with *G. pallida* (Boucher *et al.*, 2013), were also selected. To avoid unexpected  
155 hybridization, selected primers were tested on the whole *G. pallida* genome by e-PCR  
156 (Rotmistrovsky *et al.*, 2004), and only primer pairs which cannot hybridize on another part of  
157 the genome (allowing up to three mismatches) were retained. Overall, a total of 191  
158 microsatellite markers were selected, to which were added 11 markers previously developed  
159 by Montarry *et al.* (2015).

160 The 202 microsatellite markers were amplified in 5  $\mu$ L Polymerase Chain Reaction  
161 (PCR). The mix included 2  $\mu$ L of genomic DNA previously diluted at 1/3, 0.125 U of Taq,



162 PCR buffer at 1× final concentration, 2 mM of MgCl<sub>2</sub>, 0.2 mM of each dNTP (all products  
163 from Promega, Madison, WI, USA), 0.25 μM of forward and reverse primer for each loci, and  
164 0.25 μM of a universal primer (M13) labelled with either 6-FAM™, VIC™, NED™, or  
165 PET™ fluorescent dyes (Dye-GTGCTGCAACATTTTGCTG, Applied Biosystems, Foster  
166 City, CA, USA). Amplification steps were an initial denaturation at 94 °C for 5 min, followed  
167 by 20 cycles consisting of a denaturation step at 94 °C for 30 s, an annealing step at 57 °C for  
168 90 s and an elongation step at 72 °C for 30 s, then 20 cycles consisting of a denaturation step  
169 at 94 °C for 20 s, an annealing step at 53 °C for 30 s and an elongation step at 72 °C for 30 s,  
170 and a final elongation at 60°C for 30 min. Amplifications were conducted on C1000™  
171 Thermal Cyclers (Bio-Rad Laboratories, Hercules, CA, USA). After amplification, the PCR  
172 products from eight loci – four pairs of loci labelled each with a different dye and loci of the  
173 same pair labelled with the same dye but showing nonoverlapping amplicon sizes (2 μL of  
174 PCR products with 6-FAM™, VIC™ and NED™ fluorescent dyes and 4 μL of PCR products  
175 with PET™ fluorescent dyes) - were added to 5 μL H<sub>2</sub>O. Two microliters of this dilution were  
176 then added to 4.95 μL of HiDi™ Formamide (Applied Biosystems) and 0.05 μL of  
177 GeneScan™ 500 LIZ® Size Standard (Applied Biosystems). Electrophoresis of amplified  
178 fragments was carried out in a capillary sequencer ABI PRISM® 3130xl (Applied  
179 Biosystems). For each marker, peaks of the different runs were automatically binned using  
180 GENEMAPPER® v4.1 (Applied Biosystems) and the allele assignation was additionally  
181 proofread manually.

182         After genotyping, all monomorphic microsatellite markers and those with too many  
183 missing data were removed. The final dataset consisted of 53 polymorphic markers (13 di-, 14  
184 tri-, 21 tetra-, two penta- and three hexa-nucleotide motifs; Table S1, see Supporting  
185 Information) successfully amplified from 42 “individuals” in SM360, 47 “individuals” in NI  
186 and ND to 48 “individuals” in SMI and SMD (as mentioned above, data for each “individual”

187 actually come from three larvae of a single cyst) and contained in average less than 10%  
188 missing data. Despite the fact that the markers were identified on a previous version of the  
189 genome, the 53 polymorphic markers were distributed on 47 different scaffolds of the last  
190 genome version (Cotton *et al.*, 2014).

191

## 192 **Detection of outlier loci**

193 On one hand, two complementary  $F_{ST}$  outlier tests of selection were performed: DETSEL  
194 v1.0.2 (Vitalis, 2012) and the FDIST approach (Beaumont & Nichols, 1996) implemented in  
195 ARLEQUIN v3.5.2, which presents few differences with the original version (Excoffier &  
196 Lisher, 2010). Both methods rely on the estimation of the joint distribution of summary  
197 statistics, by means of stochastic simulations of neutral gene genealogies in a simple  
198 population model. FDIST and DETSEL differ by the underlying demographic model: while  
199 FDIST considers an island model of population structure, *i.e.* a set of populations with constant  
200 and equal deme size that are connected by gene flow, DETSEL considers a pure divergence  
201 model, in which an ancestral population splits into two daughters populations. Analyses were  
202 run separately for each lineage pairs SMI/SMD, NI/ND and SM360/SMD. In ARLEQUIN, for  
203 each analysis, 100,000 coalescent simulations in 100 demes were performed. In DETSEL,  
204 1,000,000 simulations were performed for each pairwise comparison, assuming no bottleneck  
205 before the split and using the following parameters: maximum allele frequency = 0.99;  
206 mutation rate  $\mu = 10^{-4}$  (corresponding to the average mutation rate for microsatellite markers –  
207 Selkoe & Toonen, 2006); time since population split  $t = 8$  generations; ancestral population  
208 size  $N_e = 10^4$ ,  $10^5$  and  $10^6$  (as recommended by Vitalis, 2012; because of the uncertainty in  
209 actual  $N_e$ , data have to be analysed using a wide range of values for this parameter). We used  
210 for all simulations the stepwise mutation model (SMM). However, additional simulations  
211 were performed using the infinite allele model (IAM), a mutation rate  $\mu = 10^{-5}$  and/or

212 assuming a bottleneck, and provided the same outlier list (data not shown). For both  $F_{ST}$ -  
213 outlier tests, a locus was classified as a significant outlier if it laid outside the 95% and 99%  
214 confidence envelopes and, as we focused on the loci putatively involved in divergence  
215 between adapted and non-adapted lineages, we only considered the loci falling above the  
216 upper confidence limits.

217 On the other hand, we used a third test based on the reduction in heterozygosity: the  
218 lnRH test (Kauer *et al.*, 2003; Schlötterer & Dieringer, 2005), specifically designed for  
219 microsatellite loci. This test is based on the assumption that microsatellites linked to loci  
220 under selection will show reduced levels of diversity between populations (Schlötterer &  
221 Dieringer, 2005). We computed the lnRH statistic for each locus in each lineage pair  
222 SMI/SMD, NI/ND and SM360/SMD by calculating the natural logarithm (ln) of the gene  
223 diversity ratio  $[(1/(1-H_{\text{lineage1}}))^2-1] / [(1/(1-H_{\text{lineage2}}))^2-1]$ , where H is the expected  
224 heterozygosity. For lineages with monomorphic loci, one different allele was added to one  
225 individual, as null values of heterozygosities prevent the estimation of lnRH due to division  
226 by zero (Kauer *et al.*, 2003). LnRH estimates were standardized to obtain a mean of 0 and a  
227 standard deviation of 1. As the lnRH is approximately normally distributed under the null  
228 hypothesis of neutrality (Schlötterer & Dieringer, 2005), loci with lnRH values lower than -  
229 1.96 (resp. upper than 1.96) and -2.58 (resp. 2.58) were considered outliers at the 0.05 and  
230 0.01 threshold, respectively, indicating reduced variability in the numerator lineage (resp. the  
231 denominator lineage) as compared to the denominator one (resp. the numerator one).

232

### 233 **Genic environment of outlier loci**

234 There is no available data on the extent of recombination and linkage disequilibrium in the *G.*  
235 *pallida* genome. However, a genetic linkage map constructed with AFLP exists for the species  
236 *G. rostochiensis* (Roupe van der Voort *et al.*, 1999) and indicates an overall physical/genetic

237 distance ratio of 120-kb/cM. Thus, we searched for predicted genes located in a window of  
238 240-kb around outlier loci (120-kb on both sides) in the annotated *G. pallida* genome (version  
239 nGp.v1.0; Cotton *et al.*, 2014). That genome version was available as an assembly of 125-Mb  
240 in 6,873 scaffolds with a N50 of 122-kb. Combining transcriptomic data with manual  
241 curation, a total of 16,419 genes were predicted. As the majority of virulence factors  
242 identified in plant parasitic nematodes are effectors (Haegeman *et al.*, 2012), we searched as a  
243 priority for genes coding for secreted proteins. We thus searched for the presence of N-  
244 terminal signal peptides in the predicted proteins using SIGNALP v4.1 (Petersen *et al.*, 2011).  
245 We also performed a BLAST search of the predicted genes against the annotated *G.*  
246 *rostochiensis* genome (version nGr.v1.0; Eves-van den Akker *et al.*, 2016), for which an  
247 expert annotation has been performed (contrary to *G. pallida* genome which has only an  
248 automated annotation), in order to potentially obtain more precise information on their  
249 functional annotation.

250

## 251 **Genetic diversity and differentiation**

252 The genetic characteristics of the main lineages of this study (SMD, SMI, NI and ND) were  
253 estimated on three datasets: one dataset containing all 53 microsatellite markers retained, one  
254 containing the markers putatively neutral (*i.e.* without the outlier loci detected) and one  
255 containing only the outlier loci (combining results of SMI/SMD and NI/ND comparisons).  
256 Genetic diversity was estimated through allelic richness ( $A_R$ ) and unbiased gene diversity ( $H$ ).  
257 Allelic richness  $A_R$  corresponds to the mean number of alleles per locus, and was estimated  
258 on a reduced sample of 20 individuals using the rarefaction method implemented in  
259 POPULATIONS 1.2.32 (Langella, 2000). Gene diversity  $H$  corresponds to the average  
260 probability across loci to draw at random different alleles in the same population, and was  
261 computed using GENETIX 4.05.2 (Belkhir *et al.*, 1996-2004). The fixation index  $F_{IS}$  was also

262 computed using GENETIX, and the statistical significance of  $F_{IS}$  values for each lineage was  
263 tested using the allelic permutation method (10,000 permutations). Genetic differentiation  
264 among all lineage pairs was estimated by calculating pairwise  $F_{ST}$  values using GENEPOP v4.0  
265 (Rousset, 2008).

266

## 267 **RESULTS**

### 268 **Detection of outlier loci**

269 We performed a genome scan on the eighth generation of five lineages coming from two  
270 independent *G. pallida* populations (SM and N). Two of them were virulent to Iledher (SMI  
271 and NI), one was virulent to another resistant genotype harbouring *GpaV<sub>vrn</sub>* in a different  
272 genetic background (SM360) and two were avirulent (SMD and ND). Each lineage was  
273 genotyped with 202 microsatellite markers. Fifty-three polymorphic markers were finally kept  
274 to perform the genome scan with three different methods: FDIST, DETSEL and lnRH. Analyses  
275 were run separately for each lineage pairs SMI/SMD, NI/ND and SM360/SMD. A locus was  
276 considered to be putatively under positive selection if at least two outlier tests were significant  
277 for that particular locus or if that locus was found outlier in at least two pairwise comparisons.

278 Eight outlier loci were found across the three pairwise comparisons and using the three  
279 outlier detection tests (Table 1). Most of them were found with the three methods and at the  
280 0.01 threshold. Gp249 in SMI/SMD and Gp235 in SM360/SMD were only detected with  
281 FDIST and DETSEL, and Gp121 in NI/ND and Gp235 in SMI/SMD, were found only with  
282 DETSEL (Table 1). In order to lower the bias on the mean neutral  $F_{ST}$  estimation induced by  
283 outlier loci present in the dataset, and subsequently the simulation of the expected distribution  
284 under the null hypothesis, it is advocated to run the analysis a second time after removing  
285 outlier loci (e.g. Antao *et al.*, 2008). This procedure is particularly justified for the NI/ND  
286 comparison, where a locus shows a very high  $F_{ST}$  value (Gp101 with  $F_{ST} = 0.75$ , Figure 1B).

287 We ran therefore each analysis a second time after removing the outlier loci at the 0.01  
288 threshold and found that Gp121 in NI/ND and Gp235 in SMI/SMD were outlier at the 0.05  
289 threshold in FDIST.

290 In the comparisons between lineages virulent and avirulent to Iledher (*i.e.* SMI/SMD  
291 and NI/ND), seven outlier loci were detected (Figures 1, 2, Table 1). Among them, one  
292 outlier, Gp121, was shared by both populations (Figure 3). Among the six other outliers,  
293 found in only one population, one of them, Gp101 (found with the three outlier tests in NI/ND  
294 comparison) was also detected as an outlier at the 0.05 threshold in both  $F_{ST}$ -based tests in the  
295 SMI/SMD comparison after the second run, and can therefore be considered as a second  
296 outlier shared by both populations (Figure 3).

297 In the comparison SM360/SMD, four outliers were found (Table 1). Three of them  
298 (Gp121, Gp235 and Gr69) were shared with the SMI/SMD pair (Figure 3). Gp121 was also  
299 shared, as seen above, with the NI/ND pair.

300 LnRH indicates in which lineage the reduction of variability - and therefore the  
301 potential positive selection - happened. In SMI/SMD comparison, a reduction of variability  
302 was observed in the SMD lineage at the Gp121 locus, while this diversity loss happened in the  
303 SMI lineage at the Gp124 and Gr69 loci (Figure 2A). In the NI/ND comparison, reduction of  
304 variability happened in ND for Gp101 and NI for Gp153 (Figure 2B). In SM360/SMD  
305 comparison, it happened in SMD for Gp121 and Gp135 and in SM360 for Gr69.

306

### 307 **Genic environment of outlier loci**

308 We investigated the predicted genes located in a window of 240-kb around outlier loci (120-  
309 kb on both sides). We searched as a priority for genes coding for secreted proteins (*i.e.*  
310 harbouring a signal peptide). Predicted genes were also blasted against the expert-annotated  
311 *G. rostochiensis* genome.

312           Microsatellites markers were developed on a previous version of the *G. pallida*  
313 genome. However, when we performed a BLAST search of outlier loci against the annotated  
314 genome version, Gp249 was not found and must therefore have been discarded from the new  
315 genome version. The scaffold where locus Gp135 was localized did not contain any predicted  
316 gene. Overall, 187 predicted genes were identified around the six remaining outlier loci  
317 (Table S2, see Supporting Information). About 44 % of these corresponded to genes with  
318 unknown functions in *G. pallida*, but ten of them harboured a signal peptide, indicating that  
319 they potentially encode secreted proteins. Nine others were orthologs of predicted genes in *G.*  
320 *rostochiensis* which harbour a signal peptide. Although their function is unknown, three  
321 genes, one localized near Gp153 and two localized near Gr69, were identified by Cotton *et al.*  
322 (2014) as potential novel effectors. Among the identified genes with known functions, five  
323 harboured a signal peptide. One of them is described as SPRYSEC, which is well-known in  
324 nematodes of the genus *Globodera* to be involved in pathogenicity (e.g. Sacco *et al.*, 2009).  
325 Another gene is also described as SPRYSEC but does not harbour a signal peptide.  
326 Interestingly, both genes are localized near Gp101, which has been identified as an outlier in  
327 both NI/ND and SMI/SMD comparisons. Two genes described as calcium binding proteins, a  
328 protein family also potentially involved in pathogenicity, were also found near Gp124 and  
329 Gr69.

330

### 331 **Genetic diversity and differentiation**

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

337 Genetic diversity was higher for the *G. pallida* lineages coming from the population of  
338 Saint-Malo (SMD and SMI) than for the lineages from the population of Noirmoutier (ND  
339 and NI), both for allelic richness and unbiased gene diversity (Table 2). The comparison  
340 between the three datasets showed clearly that genetic diversity ( $A_R$  and  $H$ ) was lower for the  
341 seven outliers than for the 46 neutral markers, and intermediate for the combined dataset  
342 (Table 2).

343 A highly significant heterozygote deficit relative to Hardy-Weinberg equilibrium was  
344 observed in all lineages: in the dataset with the 46 neutral markers,  $F_{IS}$  values ranged from  
345 0.085 to 0.165 (Table 2). For all lineages, the  $F_{IS}$  was higher for the outlier than for the  
346 neutral loci, and this was mainly due to the Gp101 locus as revealed by a jackknife  
347 resampling performed on the outlier dataset (data not shown).

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

356

## 357 **DISCUSSION**

358 In this study, we tested if a genome scan approach on a biological material coming from a  
359 short experimental evolution is efficient to target genomic footprints of selection due to host  
360 adaptation in the potato cyst nematode *Globodera pallida*. The genome scan performed on *G.*  
361 *pallida* lineages virulent and avirulent to Iledher, and coming from two independent



362 populations, highlighted seven different outlier loci, which constitute as many candidate  
363 genomic regions putatively linked to host adaptation.

364 As previously recorded in natural populations of several plant parasitic nematode  
365 species (Montarry *et al.*, 2015), all the lineages showed a heterozygote deficit. It is known that  
366 heterozygote deficit is likely to increase the rate of false positives (De Mita *et al.*, 2013), so  
367 we cannot ignore the fact that some of the outlier loci identified in our study could be false  
368 positives. However, several factors can be considered to rule on the status of good candidate  
369 for adaptation of these loci, and by extension of the linked genomic regions. For example, most  
370 of the outliers were found in two or three tests. To avoid false positives, it is in fact  
371 recommended to performed different outlier detection tests (Luikart *et al.*, 2003), and to keep  
372 only outliers detected by at least two methods.

373 Another criterion useful to regard an outlier locus as a good candidate for selection is  
374 its identification in both independent populations. In our specific case, the strong bottleneck  
375 imposed to lineages exposed to the resistant potato genotypes could have induced genetic drift  
376 and therefore a risk of detecting loci fixed only by drift. However, the use of two independent  
377 populations as repeated samples provides stronger support to outlier loci found in both  
378 populations, as the probability that the same loci were fixed by chance only in two  
379 independent populations is very small (here  $P = 1/53 * 1/53 = 0.00036$ ). This particularly  
380 applies here to the Gp121 locus, which was also found in the third comparison, *i.e.*  
381 SM360/SMD, and to some extent, to the Gp101 locus (Fig. 3). When studying natural  
382 populations, this strategy is also interesting to avoid the detection of outlier loci selected by  
383 other local environmental variables, e.g. climatic or pedological, in some populations. Here,  
384 we also avoided this pitfall by using lineages coming from an experimental evolution, which  
385 allows to control the environment, at least for the five last generations which were performed  
386 under glasshouse (see Fournet *et al.*, 2013), and thus breeds additional confidence that the

387 selection signal detected is actually linked to the external factor differentiating lineages within  
388 a pair, *i.e.* here the resistant or susceptible status of the host plants. Even if 16 genes were  
389 found in a 240-kb window around Gp121, none of them was known as an effector and none of  
390 them harboured the recently identified DOG-Box which may be a predictor of secretion  
391 (Eves-van den Akker *et al.*, 2016), but four could be candidate to play a role in pathogenicity.  
392 For Gp101, among the 21 genes found close to this locus, five of them harbour a signal  
393 peptide and two are known as SPRYSEC, which are effectors responsible for the suppression  
394 of plant defences but also for nematode virulence.

395         Results obtained with InRH indicate that, regarding the loci Gp121 and Gp101, the  
396 signal of selection detected could correspond to an adaption to the cultivar Désirée rather than  
397 to Iledher. It is not surprising as, before the experiment, none of the potato genotypes was  
398 cultivated in either of the population source fields. On the other hand, InRH also supported the  
399 results obtained for three outlier loci, Gp124, Gr69 and Gp153, for which the decrease of  
400 diversity happened in the lineages virulent to Iledher. The Gp153 locus could be an interesting  
401 candidate, as it was identified as an outlier in the NI/ND pair with all three tests and a gene  
402 close to this locus was described as a potential novel effector by Cotton *et al.* (2014). Two  
403 genes described as calcium binding proteins were found near Gp124 and Gr69, and could also  
404 represent good candidates for adaptation to plant resistance, based on their homology with  
405 some types of proteins, the annexins and the calreticulins, which are effectors allowing the  
406 suppression of plant defence reaction, and are also calcium binding proteins (Lee *et al.*, 2004).  
407 It should be emphasized that at this stage and with the genome coverage used in this study we  
408 cannot investigate as it should be the precise genomic regions involved in the adaptation to  
409 the QTL *GpaV<sub>vm</sub>*.

410

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

421

422 This work shows that genome scans on experimentally evolved lineages is efficient to detect  
423 genomic regions impacted by selection and putatively involved in host adaptation. It was  
424 important to confirm that this approach was feasible in non-model organisms with a  
425 generation time much longer than those of model organisms such as *Caenorhabditis elegans*.  
426 In fact, since *G. pallida* completes only one generation per year, the experimental evolution  
427 performed here involved a short process of divergence, *i.e.* only eight generations. If  
428 adaptation was clearly attested at the phenotypic level (Fournet *et al.*, 2013), the risk that the  
429 signal of selection was still too widespread along the genome could prevent the genetic bases  
430 of this adaptation to be targeted. Eight generations could indeed not be sufficient to allow  
431 enough recombination events in the genome and thus not restrict enough the selection signal  
432 to a small window, resulting in the detection of an excessive number of outlier loci. Here, it is  
433 not the case, as 6% microsatellite loci on average were detected as outliers, showing that the  
434 nematode lineages used in this study have undergone sufficient genetic mixing in only eight  
435 generations. This is consistent with what is classically found in genome scan studies, where 2-

436 10% of markers were typically reported as departing from the neutral expectation (Nosil *et*  
437 *al.*, 2008). Another argument based upon the assumption of the pre-existence of the  
438 mutation(s) in the populations, *i.e.* standing variation, could also have predicted that the  
439 selective sweep would be restricted enough to allow the detection of a small number of  
440 outliers. Several facts support the hypothesis that the individuals virulent to *GpaV<sub>vrn</sub>* were  
441 already present in the wild populations before the start of the experiment. Bakker (2002)  
442 observed that the introduction of many resistant cultivars in Europe has been unsuccessful  
443 because of *G. pallida* and *G. rostochiensis* populations showing high level of virulence,  
444 without ever being exposed to these resistance genes in Europe, suggesting that the virulence  
445 genes were already present in the initial populations introduced from the Andean region of  
446 South-America. This is in line with the fact that some native *G. pallida* populations from  
447 South-America are also virulent to *S. vernei* (Hockland *et al.*, 2012). This hypothesis was also  
448 supported in the context of our study, as the overcoming of several resistant genotypes was  
449 attested in less than eight generations for both independent populations (Fournet *et al.*, 2013).  
450 These results could be more related to a shared pre-existing mutation rather than the  
451 independent appearance of a new mutation in both populations in a small time interval. In a  
452 situation where a novel mutation is selected and rapidly fixed in the population, the diversity  
453 is strongly reduced on a large region because recombination cannot occur in a small period.  
454 On the contrary, in the case of an adaptation from a pre-existing polymorphism, which would  
455 occur here, the favourable variant is ancient and has been able to recombine, so the selective  
456 sweep is located in a restricted region (Przeworski *et al.*, 2005).

457         The results of this study are encouraging as they have given a first insight into the  
458 possibilities that genome scan approaches offer to infer the genetic bases of host adaptation in  
459 nematodes and more largely in sexually reproducing plant parasites. The coupled approach  
460 used in this study is close to the ‘evolve and resequencing’ (E&R) studies which combine

461 experimental evolution and genome resequencing to identify loci involved in adaptation and  
462 analyse the trajectories of these loci during adaptation (e.g. Turner *et al.*, 2011). However,  
463 E&R studies use time series data, which was not possible in our case as the parental material  
464 of the lineages was not available anymore. These studies seem to be mostly performed on  
465 model species like *Drosophila melanogaster*. More generally, the approaches of population  
466 genomics are still underused in the phytopathology area in order to elucidate the determinism  
467 of the adaptive potential of pathogen populations.

468

469 In the present study, only 53 microsatellites loci were genotyped in a 100-Mb genome. Even  
470 if some candidate genes were highlighted, they remain very hypothetical and it is still not  
471 possible to target the molecular bases of the adaptation. It is now necessary to use a more  
472 refined approach, using markers more densely distributed along the genome, *i.e.* Single  
473 Nucleotide Polymorphisms obtained by whole genome resequencing. With the recent  
474 publication of the *G. pallida* genome (Cotton *et al.*, 2014), a Pool-Seq approach has started on  
475 the lineages studied here. Using such an improved coverage we will be able to better target  
476 the genomic regions impacted by selection by identifying outliers that cluster physically close  
477 to each other and may be also close to one of the outlier loci reveals in the present study.  
478 Before pinpointing the mutation involved in the overcoming of the resistance, advanced  
479 results could already be used to develop efficient strategies of deployment of plant resistance.  
480 Indeed, the efficiency and durability of a plant resistance could be predicted before its  
481 deployment by a proper characterization of the nematode populations, including their genetic  
482 variability and particularly virulence genes frequencies. SNP linked to the virulence genes  
483 could therefore be used to follow the evolution of virulence allele frequencies in field  
484 populations or coming from experimental evolution, confronted to different resistance  
485 management strategies.

486

## 487 **ACKNOWLEDGMENTS**

488 The authors gratefully acknowledge Didier Mugniéry and Marie-Claire Kerlan who took part  
489 in the experimental evolution. We kindly acknowledge Didier Andrivon, Fabien Halkett and  
490 two anonymous referees for comments on previous versions of this manuscript and Sylvie  
491 Bardou-Valette for her help in the microsatellite genotyping. DEB is supported by the INRA  
492 department SPE and the Région Bretagne through a three year PhD grant.

493

## 494 **REFERENCES**

- 495 Aguilera G, Refrégier G, Yockteng R, Fournier E, Giraud T, 2009. Rapidly evolving genes in  
496 pathogens: Methods for detecting positive selection and examples among fungi, bacteria,  
497 viruses and protists. *Infection, Genetics and Evolution* **9**, 656-70.
- 498 Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G, 2008. LOSITAN: a workbench to  
499 detect molecular adaptation based on a  $F_{ST}$ -outlier method. *BMC Bioinformatics* **9**, 323.
- 500 Bakker J, 2002. Durability of resistance against potato cyst nematodes. *Euphytica* **124**, 157-  
501 62.
- 502 Beaumont MA, Nichols RA, 1996. Evaluating loci for use in the genetic analysis of  
503 population structure. *Proceedings of the Royal Society of London B: Biological sciences*  
504 **263**, 1619-26.
- 505 Bekal S, Domier LL, Gonfa B, Lakhssassi N, Meksem K, Lambert KN, 2015. A SNARE-like  
506 protein and biotin are implicated in soybean cyst nematode virulence. *PLoS ONE* **10**,  
507 e0145601.
- 508 Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F, 1996-2004. GENETIX 4.05, logiciel  
509 sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations,  
510 Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

511 Boucher AC, Mimee B, Montarry J, Bardou-Valette S, Bélair G, Moffett P, Grenier E, 2013.  
512 Genetic diversity of the golden potato cyst nematode *Globodera rostochiensis* and  
513 determination of the origin of populations in Quebec, Canada. *Molecular Phylogenetics*  
514 *and Evolution* **69**, 75–82.

515 Brun H, Chèvre AM, Fitt BDL, Powers S, Besnard AL, Ermel M, Huteau V, Marquer B, Eber  
516 F, Renard M, Andrivon D, 2010. Quantitative resistance increases the durability of  
517 qualitative resistance to *Leptosphaeria maculans* in *Brassica napus*. *New Phytologist* **185**,  
518 285-99.

519 Clayton R, Storey M, Parker B, Ballingall M, Davies K, 2008. Impact of Reduced Pesticide  
520 Availability on Control of Potato Cyst Nematodes and Weeds in Potato Crops. Kenilworth,  
521 UK: Potato Council Ltd.

522 Cotton JA, Lilley CJ, Jones LM, Kikuchi T, Reid AJ, Thorpe P, Tsai IJ, Beasley H, Blok V,  
523 Cock PJA, Eves-van den Akker S, Holroyd N, Hunt M, Mantelin S, Naghra H, Pain A,  
524 Palomares-Rius JE, Zarowiecki M, Berriman M, Jones JT, Urwin PE, 2014. The genome  
525 and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant  
526 parasitism by a cyst nematode. *Genome Biology* **15**, R43.

527 De Mita S, Thuillet AC, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y, 2013.  
528 Detecting selection along environmental gradients: analysis of eight methods and their  
529 effectiveness for outbreeding and selfing populations. *Molecular Ecology* **22**, 1383-99.

530 Eves-van den Akker S, Laetsch DR, Thorpe P, Lilley CJ, Danchin EGJ, Da Rocha M,  
531 Rancurel C, Holroyd NE, Cotton JA, Szitenberg A, Grenier E, Montarry J, Mimee B,  
532 Duceppe MO, Boyes I, Marvin JMC, Jones LM, Yusup HB, Lafond-Lapalme J, Esquibet  
533 M, Sabeh M, Rott M, Overmars H, Finkers-Tomczak A, Smant G, Koutsovoulos G, Blok  
534 V, Mantelin S, Cock PJA, Phillips W, Henrissat B, Urwin PE, Blaxter M, Jones JT, 2016.

535 The genome of the yellow potato cyst nematode, *Globodera rostochiensis*, reveals insights  
536 into the basis of parasitism and virulence. *Genome Biology* **17**:124.

537 Excoffier L, Lischer HEL, 2010. ARLEQUIN suite ver 3.5: a new series of programs to perform  
538 population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**,  
539 564-7.

540 Faircloth BC, 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and  
541 automated, locus-specific primer design. *Molecular Ecology Resources* **8**, 92-4.

542 Fournet S, Kerlan MC, Renault L, Dantec JP, Rouaux C, Montarry J, 2013. Selection of  
543 nematodes by resistant plants has implications for local adaptation and cross-virulence.  
544 *Plant Pathology* **62**, 184–93.

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

547 Hockland S, Niere B, Grenier E, Blok V, Phillips M, den Nijs L, Anthoine G, Pickup J,  
548 Viaene N, 2012. An evaluation of the implications of virulence in non-European  
549 populations of *Globodera pallida* and *G. rostochiensis* for potato cultivation in Europe.  
550 *Nematology* **14**, 1-13.

551 Jones FGW, 1950. Observations on the beet eelworm and other cyst-forming species of  
552 *Heterodera*. *Annals of Applied Biology* **37**, 407–40.

553 Kauer MO, Dieringer D, Schlötterer C, 2003. A microsatellite variability screen for positive  
554 selection associated with the “Out of Africa” habitat expansion of *Drosophila*  
555 *melanogaster*. *Genetics* **165**, 1137-48.

556 Langella O, 2000. POPULATIONS 1.2: population genetic software, individuals or population  
557 distance, phylogenetic trees. [<http://bioinformatics.org/~tryphon/populations/>].

558 Lee S, Lee EJ, Yang EJ, Lee JE, Park AR, Song WH, Park OK, 2004. Proteomic  
559 identification of annexins, calcium-dependent membrane binding protein that mediate



560 osmotic stress and abscisic acid signal transduction in *Arabidopsis*. *Plant Cell* **16**, 1378–  
561 91.

562 Luikart G, England PR, Tallmon D, Jordan S, Taberlet P, 2003. The power and promise of  
563 population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**, 981–  
564 94.

565 Maynard-Smith J, Haigh J, 1974. Hitch-hiking effect of a favorable gene. *Genetics Research*  
566 **23**, 23-35.

567 Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF, 2010. QDD: a  
568 user-friendly program to select microsatellite markers and design primers from large  
569 sequencing projects. *Bioinformatics* **26**, 403–4.

570 Montarry J, Jan PL, Gracianne C, Overall ADJ, Bardou-Valette S, Olivier E, Fournet S,  
571 Grenier E, Petit EJ, 2015. Heterozygote deficits in cyst plant-parasitic nematodes: possible  
572 causes and consequences. *Molecular Ecology* **24**, 1654-77.

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

577 Nosil P, Egan SP, Funk DJ, 2008. Heterogeneous genomic differentiation between walking-  
578 stick ecotypes: “isolation by adaptation” and multiple roles for divergent selection.  
579 *Evolution* **62**, 316-36.

580 Palloix A, Ayme V, Moury B, 2009. Durability of plant major resistance genes to pathogens  
581 depends on the genetic background: experimental evidence and consequences for breeding  
582 strategies. *New Phytologist* **183**, 190-9.

583 Petersen TN, Brunak S, von Heijne G, Nielsen H, 2011. SIGNALP 4.0: discriminating signal  
584 peptides from transmembrane regions. *NatureMethods* **8**,785-6.

585 Przeworski M, Coop G, Wall JD, 2005. The signature of positive selection on standing  
586 genetic variation. *Evolution* **59**, 2312-23.

587 Rotmistrovsky K, Jang W, Schuler GD, 2004. A web server for performing electronic PCR.  
588 *Nucleic Acids Research* **32**, W108-W112.

589 Rouppe van der Voort JNAM, van der Vossen E, Bakker E, Overmars H, van Zandvoort P,  
590 Hutten R, Klein Lankhorst R, Bakker J, 2000. Two additive QTLs conferring broad-  
591 spectrum resistance in potato to *Globodera pallida* are localized on resistance gene  
592 clusters. *Theoretical and Applied Genetics* **101**, 1122-30.

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

597 Rousset F, 2008. GENEPOP'007: a complete reimplementation of the GENEPOP software for  
598 Windows and Linux. *Molecular Ecology Resources* **8**, 103-6.

599 Sacco MA, Koropacka K, Grenier E, Jaubert MJ, Blanchard A, Goverse A, Smant G, Moffett  
600 P, 2009. The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-  
601 dependent plant cell death. *PLoS Pathogens* **5**, e1000564.

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

605 Selkoe KA, Toonen RJ, 2006. Microsatellites for ecologists: a practical guide to using and  
606 evaluating microsatellite markers. *Ecology Letters* **9**, 615–29.

607 Storz JF, 2005. Using genome scans of DNA polymorphism to infer adaptive population  
608 divergence. *Molecular Ecology* **14**, 671-88.

609 Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM, 2011. Population-based  
610 resequencing of experimentally evolved populations reveals the genetic basis of body size  
611 variation in *Drosophila melanogaster*. PLoS Genetics 7, e1001336.

612 Vitalis R, 2012. DETSEL: an R-package to detect marker loci responding to selection. In:  
613 Pompanon F, Bonin A, eds. *Data Production and Analysis in Population Genomics:  
614 Methods and Protocols: Methods in Molecular Biology*. New York, NY, USA: Springer  
615 Science + Business Media, 277-93.

616 **SUPPORTING INFORMATION**

617 Additional Supporting Information may be found in the online version of this article at the  
618 publisher's website:

619 **Table S1** Polymorphic microsatellite loci used. The eight identified outliers are shown in  
620 bold.

621 **Table S2** Putative functions of the predicted genes located in a 240-kb window centred on the  
622 outlier microsatellite loci on the *G. pallida* genome assembly version Gpal.v1.0 (Cotton *et al.*  
623 2014). Genes likely to be involved in pathogenicity are shown in bold.

624

625 **Tables**

626

627 **Table 1** Microsatellite markers identified by the three outlier detection methods (DETSEL,  
 628 FDIST and lnRH) in NI/ND, SMI/SMD and SM360/SMD comparisons. \*: P < 0.05; \*\*: P <  
 629 0.01.

Outlier loci	Pairwise comparisons		
	NI/ND	SMI/SMD	SM360/SMD
Gp121	DETSEL**	DETSEL** + FDIST** + lnRH**	DETSEL** + FDIST** + lnRH
Gp101	DETSEL** + FDIST** + lnRH**		
Gp235		DETSEL**	DETSEL** + FDIST**
Gr69		DETSEL** + FDIST** + lnRH**	DETSEL** + FDIST* + lnRH*
Gp124		DETSEL** + FDIST** + lnRH**	
Gp249		DETSEL** + FDIST**	
Gp153	DETSEL** + FDIST* + lnRH**		
Gp135			DETSEL* + FDIST* + lnRH*

630

631 **Table 2** Genetic characteristics of the main *G. pallida* lineages studied. N: number of sampled  
 632 individuals. A<sub>R</sub>: allelic richness. H: Nei's (1978) expected unbiased heterozygosity. F<sub>IS</sub>:  
 633 fixation index (\*\*: P < 0.01). (53), (46), (7): estimates with data from 53 microsatellite  
 634 markers, 46 putatively neutral markers and 7 outlier markers, respectively.

Lineage	Locality	N	A <sub>R</sub>			H			F <sub>IS</sub>		
			(53)	(46)	(7)	(53)	(46)	(7)	(53)	(46)	(7)
SMD	Saint-Malo	48	2.55	2.57	2.38	0.386	0.397	0.318	0.125 **	0.085 **	0.451 **
SMI	Saint-Malo	48	2.49	2.54	2.16	0.396	0.412	0.292	0.136 **	0.118 **	0.308 **
ND	Noirmoutier	47	2.14	2.19	1.81	0.271	0.277	0.231	0.169 **	0.165 **	0.202 **
NI	Noirmoutier	47	2.20	2.23	1.99	0.266	0.278	0.188	0.132 **	0.120 **	0.243 **

635

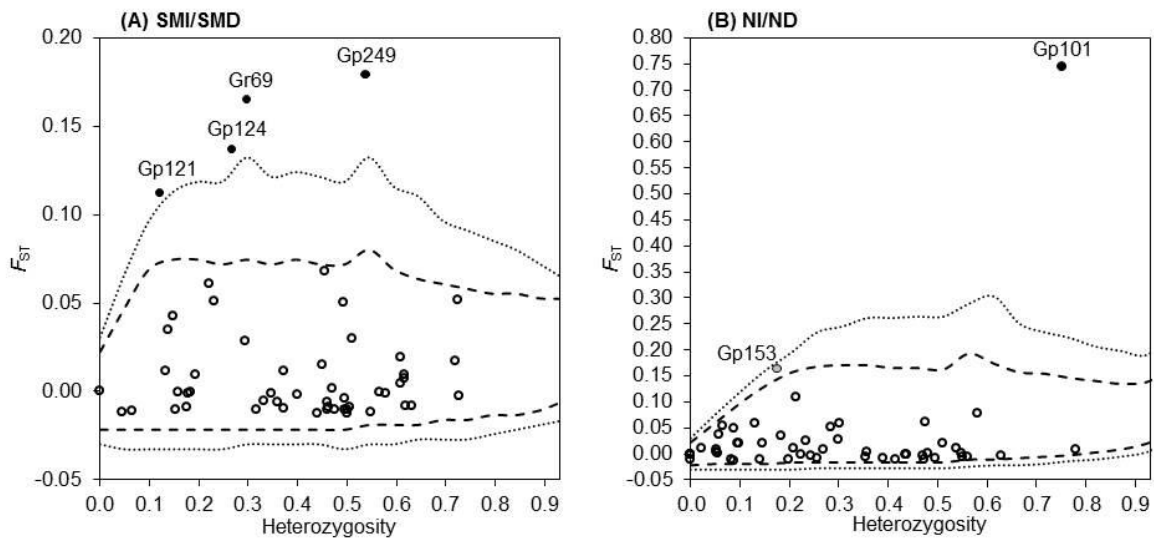
636

637 **Figures**

638

639

**Fig. 1** Distribution of empirical  $F_{ST}$  values as a function of heterozygosity for each of the 53 microsatellite loci in **(A)** SMI/SMD comparison and **(B)** NI/ND comparison. Upper and lower dotted lines: 99<sup>th</sup> and 1<sup>st</sup> quantiles of the neutral envelope; upper and lower dashed lines: 95<sup>th</sup> and 5<sup>th</sup> quantiles of the neutral envelope; black and grey dots: outlier loci at 0.01 and 0.05 thresholds.

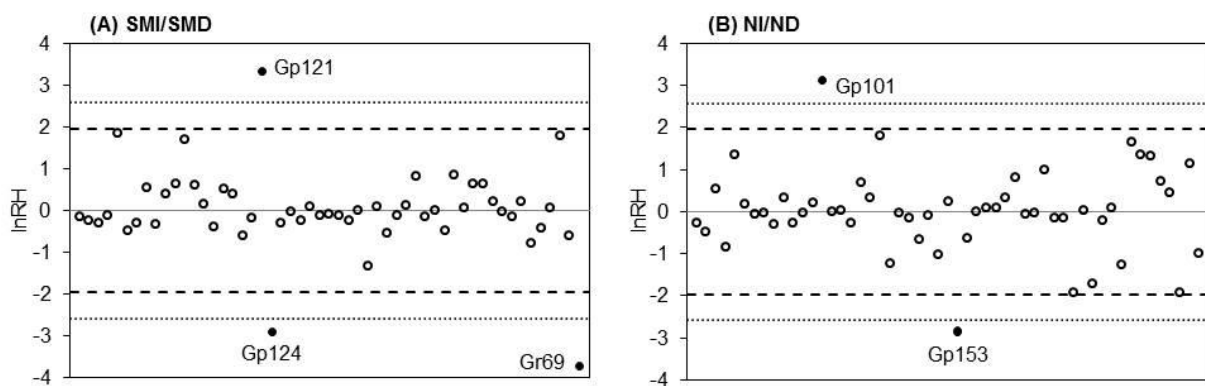


640

641

642

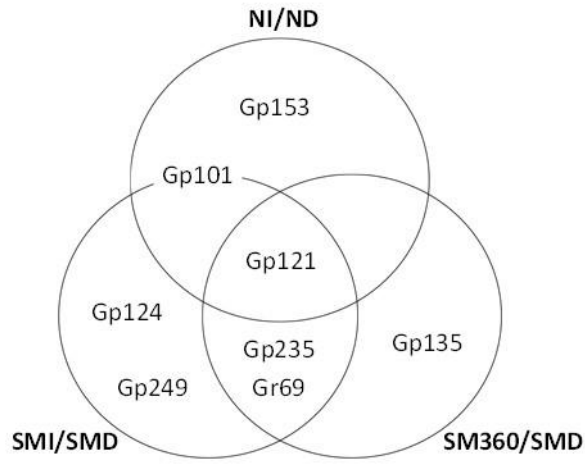
**Fig. 2** Standardized  $\ln RH$  values for each of the 53 microsatellite loci in **(A)** SMI/SMD comparison and **(B)** NI/ND comparison. Upper and lower dotted lines: 99% confidence interval (-2.58, +2.58); upper and lower dashed lines: 95% confidence interval (-1.96, 1.96); black dots: outlier loci at 0.01 threshold.



643

644

**Fig. 3** Venn diagram showing the microsatellite markers identified in one, two or three pairwise comparisons (NI/ND, SMI/SMD and SM360/SMD). Gp101 is on the line because it was detected in the first analysis in the NI/ND comparison and in the second analysis (after removing outlier loci) in the SMI/SMD comparison.



645

646