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

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

D. EOCHE-BOSY, J. GAUTHIER[1], A.S. JUHEL[2], M. ESQUIBET, S. FOURNET, E. GRENIER and J. MONTARRY*

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

Correspondence: Josselin Montarry, josselin.montarry@inra.fr

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

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

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

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

ABSTRACT

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

INTRODUCTION

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

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

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

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

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

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

MATERIALS AND METHODS

Selection of virulent and avirulent *G. pallida* lineages

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

DNA extraction and microsatellite genotyping

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

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

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

PCR buffer at 1× final concentration, 2 mM of MgCl2, 0.2 mM of each dNTP (all products from Promega, Madison, WI, USA), 0.25 µM of forward and reverse primer for each loci, and 0.25 µM of a universal primer (M13) labelled with either 6-FAM™, VIC™, NED™, or PET™ fluorescent dyes (Dye-GTGCTGCAACATTTTGCTG, Applied Biosystems, Foster City, CA, USA). Amplification steps were an initial denaturation at 94 °C for 5 min, followed by 20 cycles consisting of a denaturation step at 94 °C for 30 s, an annealing step at 57 °C for 90 s and an elongation step at 72 °C for 30 s, then 20 cycles consisting of a denaturation step 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, and a final elongation at 60°C for 30 min. Amplifications were conducted on C1000™ Thermal Cyclers (Bio-Rad Laboratories, Hercules, CA, USA). After amplification, the PCR products from eight loci – four pairs of loci labelled each with a different dye and loci of the same pair labelled with the same dye but showing nonoverlapping amplicon sizes (2 µL of PCR products with 6-FAM™, VIC™ and NED™ fluorescent dyes and 4 µL of PCR products with PET™ fluorescent dyes) - were added to 5 µL H_2O. Two microliters of this dilution were then added to 4.95 µL of HiDi™ Formamide (Applied Biosystems) and 0.05 µL of GeneScan™ 500 LIZ® Size Standard (Applied Biosystems). Electrophoresis of amplified fragments was carried out in a capillary sequencer ABI PRISM® 3130xl (Applied Biosystems). For each marker, peaks of the different runs were automatically binned using GENEMAPPER® v4.1 (Applied Biosystems) and the allele assignation was additionally proofread manually.

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

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

Detection of outlier loci

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

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

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

Genic environment of outlier loci

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

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

Genetic diversity and differentiation

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

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

RESULTS

Detection of outlier loci

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

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

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

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

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

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

Genic environment of outlier loci

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

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

Genetic diversity and differentiation

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

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

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

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

DISCUSSION

In this study, we tested if a genome scan approach on a biological material coming from a short experimental evolution is efficient to target genomic footprints of selection due to host adaptation in the potato cyst nematode *Globodera pallida*. The genome scan performed on *G. 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 adaption 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

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

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

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-

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

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

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

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

ACKNOWLEDGMENTS

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

REFERENCES

Aguileta G, Refrégier G, Yockteng R, Fournier E, Giraud T, 2009. Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution* **9**, 656-70.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G, 2008. LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics* **9**, 323.

Bakker J, 2002. Durability of resistance against potato cyst nematodes. *Euphytica* **124**, 157-62.

Beaumont MA, Nichols RA, 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological sciences* **263**, 1619-26.

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

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F, 1996-2004. GENETIX 4.05, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

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.

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

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

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

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**, 184–93.

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

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

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

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**, 1137-48.

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

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

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 varation. *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.

SUPPORTING INFORMATION

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

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

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

Tables

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

Table 2 Genetic characteristics of the main *G. pallida* lineages studied. N: number of sampled individuals. A_R: allelic richness. H: Nei's (1978) expected unbiased heterozygosity. F_{IS}: fixation index (**: P < 0.01). (53), (46), (7): estimates with data from 53 microsatellite markers, 46 putatively neutral markers and 7 outlier markers, respectively.

Lineage	Locality	N	A_R			H			F_{IS}		
			(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 **

Figures

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: 99th and 1st quantiles of the neutral envelope; upper and lower dashed lines: 95th and 5th quantiles of the neutral envelope; black and grey dots: outlier loci at 0.01 and 0.05 thresholds.

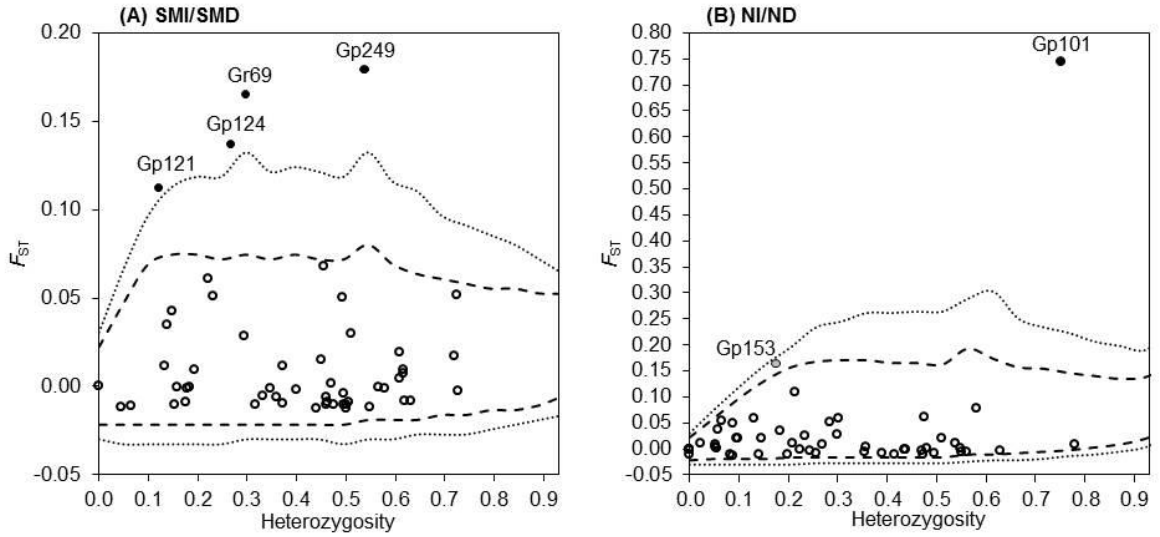


Fig. 2 Standardized lnRH values for each of the 53 microsatellite loci in **(A)** SMI/SMD comparison and **(B)** NI/ND comparison. Upper and lower dotted lines: 99% confidence interval (-2.58, +2.58); upper and lower dashed lines: 95% confidence interval (-1.96, 1.96); black dots: outlier loci at 0.01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

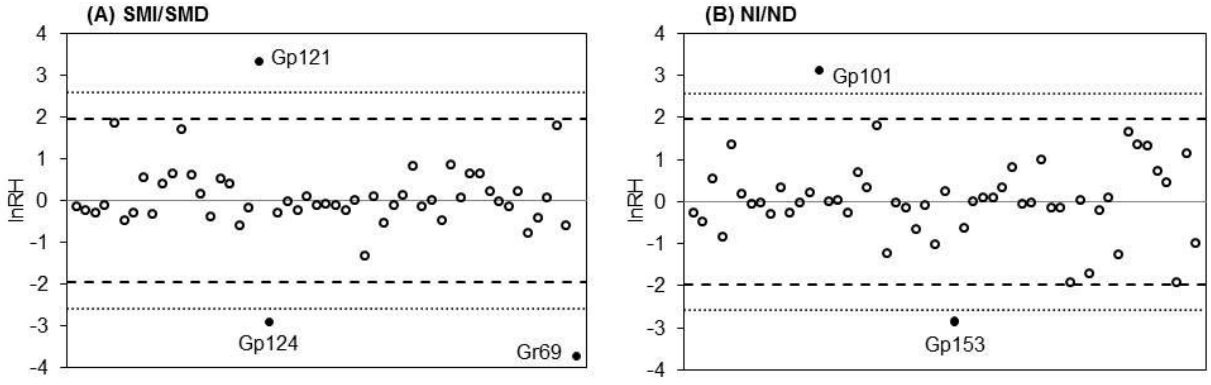
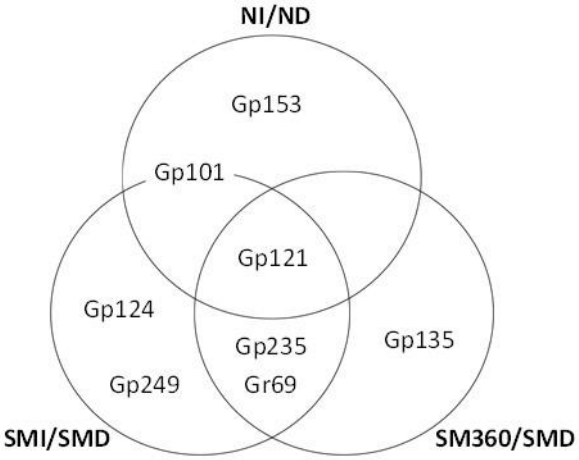


Fig. 3 Venn diagram showing the microsatellite markers identified in one, two or three pairwise comparisons (NI/ND, SMI/SMD and SM360/SMD). Gp101 is on the line because it was detected in the first analysis in the NI/ND comparison and in the second analysis (after removing outlier loci) in the SMI/SMD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



645

646