The evolution of the Gp-Rbp-1 gene in Globodera pallida includes multiple selective replacements

JEAN CARPENTIER1,2,3, MAGALI ESQUIBET1, DIDIER FOUVILLE1, MARIA J. MANZANARES-DAULEUX3, MARIE-CLAIRE KERLAN2 AND ERIC GRENIER1,*

1INRA, UMR 1099 Bio3P INRA/Agrocampus Ouest/Université Rennes 1, F-35653 Le Rheu, France
2INRA, UMR 118 APBV INRA/Agrocampus Ouest/Université Rennes 1, Keraïber F-29260 Ploudaniel, France
3INRA, UMR 1099 Bio3P INRA/Agrocampus Ouest/Université Rennes 1, F-35653 Le Rheu, France

SUMMARY

The Globodera pallida SPRYSEC Gp-Rbp-1 gene encodes a secreted protein which induces effector-triggered immunity (ETI) mediated by the Solanum tuberosum disease resistance gene Gpa2. Nonetheless, it is not known how the Andes orogeny, the richness in Solanum species found along the Cordillera or the introduction of the nematode into Europe have affected the diversity of Gp-Rbp-1 and its recognition by Gpa2. We generated a dataset of 157 highly polymorphic Gp-Rbp-1 sequences and identified three Gp-Rbp-1 evolutionary pathways: the ‘Northern Peru’, ‘Peru clade I/European’ and ‘Chilean’ paths. These may have been shaped by passive dispersion of the nematode and by climatic variations that have influenced the nature and diversity of wild host species. We also confirmed that, by an analysis of the selection pressures acting on Gp-Rbp-1, this gene has evolved under positive/diversifying selection, but differently among the three evolutionary pathways described. Using this extended sequence dataset, we were able to detect eight sites under positive selection. Six sites appear to be of particular interest because of their predicted localization to the extended loops of the B30.2 domain and/or support by several computational methods. The P/S 187 position was previously identified for its effect on the interaction with GPA2. The functional importance of the other five amino acid polymorphisms observed was investigated using Agrobacterium transient transformation assays. None of these new residues, however, appears to be directly involved in Gpa2-mediated plant defence mechanisms. Thus, the P/S polymorphism observed at position 187 remains the sole variation sufficient to explain the recognition of Gp-Rbp-1 by Gpa2.

INTRODUCTION

Plant–parasite interactions are often analysed in the context of the gene-for-gene model described by Flor (1955). In this model, both plant disease resistance and avirulence genes evolve under diversifying selection, the former to improve the recognition of pathogen effectors and the latter to escape recognition by resistance proteins (Flor, 1955; Jones and Dangl, 2006). In cyst nematodes, the role of certain secreted proteins in the host–pathogen interaction has been well documented (Davis et al., 2008). One of these secreted proteins, Gp-RBP-1 (Blanchard et al., 2005), belongs to the SPRYSEC multigene family and is described as the avirulence gene whose product is recognized by the resistance protein GPA2 (Sacco et al., 2009).

The SPRYSEC gene family encodes nematode secreted proteins specifically expressed in the dorsal oesophageal gland of the nematode early in the parasitic cycle (Rehman et al., 2009). Apart from the presence of a signal peptide (SP), SPRYSEC genes contain a single B30.2 domain, which is an extended domain structure comprising PRY and SPRY units known to be involved in protein–protein interactions (Woo et al., 2006). The subcellular expression in plants of several members of this gene family from the potato cyst nematode, Globodera pallida, has been characterized (Jones et al., 2009). Although the Gp-RBP-1 protein appears to be localized in the cytoplasm, other SPRYSEC members show different subcellular localization patterns (nucleus and nucleolus), suggesting that SPRYSEC proteins can target a range of host proteins and act as suppressors of host defences (Jones et al., 2009; Rehman et al., 2009).

Among the G. pallida SPRYSECs, the Gp-RBP-1 protein has been shown to be able to trigger a hypersensitive response (HR) in the presence of the coiled coil–nucleotide binding–leucine–rich repeat (CC-NB-LRR) protein Gpa2. Transient expression assays co-expressing Gpa2 with variants of Gp-Rbp-1 in tobacco leaves have also suggested that a proline residue at position 187 (P/S polymorphism) in the Gp-RBP-1 protein sequence is required for recognition by Gpa2 (Sacco et al., 2009). Despite all this functional information, the evolution of Gp-Rbp-1 has not been investigated to date and, in particular, it is unknown how the phylogeographical history of G. pallida may have had an impact on the diversity of Gp-Rbp-1. Moreover, the Gp-Rbp-1 polymorphism assessed in Sacco’s study was mostly based on sequences obtained from European G. pallida populations. Considering the diversity observed in indigenous South American G. pallida populations, it is not clear whether the P-187 residue alone is still sufficient for Gpa2 recognition.
Indeed, for millions of years, *Solanum* species and nematodes have co-evolved and have gradually spread throughout most of the South American continent, with the Andes orogeny playing a dominant role in the South American conquest by *G. pallida* (Grenier et al., 2010). Nearly 200 wild potato species can be found in South America. Many of these wild species are present in Argentina, Bolivia and Peru (where 93 species have been recorded), whereas, in other countries, such as Chile or Brazil, the spectrum of *Solanum* species is much poorer (Hijmans and Spooner, 2001). Genetic studies on Peruvian populations of *G. pallida*, conducted using neutral markers, have shown a clear south-to-north phylogeographical pattern with five well-supported clades (Picard et al., 2007). All cyst nematodes sampled in Europe have been shown to be derived from populations belonging to Peru clade I, the southernmost Peruvian clade, which also shows the greatest genetic diversity (Picard et al., 2007).

In this study, we broadly explored the variability of *Gp-Rbp-1* by sequencing it in various novel *G. pallida* populations, especially populations from Peru (clades I–V) and Chile. We analysed the genetic structure of more than 150 sequences from 22 *G. pallida* populations to study the effects of ecological changes linked to the Andes orogeny or the accidental introduction of *G. pallida* in Europe. Our findings suggest that the evolution of *Gp-Rbp-1* has been shaped by three distinct evolutionary pathways. We then investigated the presence of selective pressures considering these three evolutionary pathways and were able to identify novel sites under positive selection. Finally, we used transient expression assays in tobacco leaves to examine the effect of these polymorphisms on the outcome of the interaction between S-187 *Gp-Rbp-1* variants and *Gpa2*.

**RESULTS**

**Structure and variability of *Gp-Rbp-1***

The previously reported *Gp-Rbp-1* sequences vary in length, but all are characterized by the presence of SP and B30.2 domains divided into a 129-amino-acid SPRY subdomain and two PRY subdomains. We amplified a total of 149 sequences from 19 populations with an average of eight different sequences per population. We completed our dataset with 33 sequences available in GenBank/EMBL databases, including sequences from the three populations D383, Rookmaker and Guiclan. Of these 182 sequences, 13 contain a SPRY subdomain of greater or less than 129 amino acids and, for this reason, these were removed from the final dataset. In the 169 remaining sequences, very few identical sequences were observed. Extensive variations in the coding DNA sequence length were observed, which ranged from 720 to 870 bp. These variations in length are a result of large indels affecting the B30.2 domain and the number of PRY subdomains (Fig. 1). Thus, 12 further sequences corresponding to variants showing large indels in the B30.2 domain were also removed (Fig. 1). As such, the final sequence dataset used for all further analyses included 157 sequences: 30 from GenBank and 127 from our study.

Polymorphism analysis conducted using DnaSP software revealed 250 polymorphic sites dispersed along the entire *Gp-Rbp-1* sequence: 137 singletons and 113 parsimony-informative sites. Among these 250 mutations, 72% are nonsynonymous and represent, on average, one possible amino acid replacement every seven amino acids along the entire protein sequence. In contrast with the findings reported following analysis with neutral markers (Picard et al., 2007), we found no clear evidence of a reduction in variability from south to north Peru (Fig. 2). The maximum genetic diversity was observed around latitude 13°S, which corresponds to the geographical region of the *G. pallida* clades II and III. Nonetheless, strong local effects were also observed. For example, the populations P297 and P298 showed different levels of nucleotide diversity, even though they are geographically very close.

Results from factorial analysis based on the *Gp-Rbp-1* dissimilarity matrix are shown in Fig. 3: axes 1 and 2 explain 27.4% and 13.6% of the total observed variability, respectively. The third calculated axis, which explained 6.9% of the variability, was not considered in our study. No clear structure was observed in the five clades from the Peruvian populations. However, sequences issued from populations of the same geographical region (Peru vs. Chile vs. Europe) tended to cluster together. Two exceptions to this rule are noteworthy: (i) the relationship between the European and Peru clade I; and (ii) the relationship between Chilean and Peru P309 *Gp-Rbp-1* sequences. The factorial analysis (Fig. 3) showed a central group (composed essentially of *Gp-Rbp-1* sequences from European, clade I, clade II and clade III populations), which itself is split into three distinct groups. We interpreted these results in terms of three distinct evolutionary pathways: (i) the ‘Northern Peru path’, including sequences from populations of clades II–V; (ii) the ‘Peru clade I/European path’, grouping sequences from European and Peru clade I populations; and (iii) the ‘Chilean path’, encompassing sequences from P309 and Tierra del Fuego populations. Based on analysis of molecular variance (*AMOVA*), the consideration of eight geographical *G. pallida* groups (five Peruvian clades + the Chilean population + the P309 population + European populations) could only explain 11.9% of the total variability (Table 1A). By contrast, 28.2% of the variability was explained by the three evolutionary pathways (Table 1B), thus confirming the observations made in the factorial analysis.

*Gp-Rbp-1* has been subject to positive selection

Site-specific models for the detection of positive selection were used to align the 157 *Gp-Rbp-1* sequences obtained. We compared four evolutionary models implemented in the CODEML
Fig. 1  Structure and sequence size variation in the Gp-Rbp-1 sequence dataset. Insertions and deletions are indicated on the gene structure by an arrow. Indel sizes and positions are boxed above the arrows (numbering was made on the basis of the FJ392677.1 sequence used here as a reference). The size of the sequence in nucleotides and the number of sequences concerned are indicated on the right-hand side of each structure. Information available (Sacco et al., 2009) on the strength of the Gpa2-dependent hypersensitive response (HR) induction is indicated as follows: ++, complete collapse and rapid desiccation of the infiltration patch within 2 days; +++, complete collapse of the infiltration patch by 3 days post-infiltration; +, slow and incomplete collapse with residual live cells; NT, not previously tested. Sequences that were kept in the final sequence dataset and used for further analysis are highlighted in bold. nt, nucleotide.

| Structure          | coding DNA size (nucleotides) | number of sequences concerned | Strength of the recognition by Gpa2 |
|--------------------|-------------------------------|-------------------------------|-----------------------------------|
| SP PRY PRY SPRY    | 795                           | 133                           | ++                                |
| -3nt (91 to 93)    | 795                           | 17                            | +++/++++                          |
| SP PRY PRY SPRY    | 798                           | 2                             | NT                                |
| -3nt (88 to 90)    | 792                           | 4                             | ++                                |
| SP PRY PRY SPRY    | 870                           | 1                             | NT                                |
| +75nt (243 to 244) | 840                           | 2                             | +/++                              |
| SP PRY PRY SPRY    | 723                           | 7                             | NT                                |
| -72nt (244 to 316) | 720                           | 2                             | NT                                |

Fig. 2  Variation in Gp-Rbp-1 nucleotide diversity ($\rho$) among the 12 Peruvian populations distributed from southern to northern Peru. The locations of the 12 Peruvian nematode populations studied are indicated on the Peru map. A colour code was used to indicate the affiliation of each population to its phylogeographical clade. The nucleotide diversity ($\rho$) calculated from the Gp-Rbp-1 sequences for each of these populations and its standard deviation are shown on the right-hand side of the map and according to the geographical position of the population.
program of PAML: M1vsM2 and M7vsM8. The M2 and M8 models of positive selection appeared to be significantly ($P < 0.001$) better adapted to our dataset (Table 2). Both the M2 and M8 models identified eight sites under positive selection (C/G/R/S 37, H/N/Y 100, K/E 102, A/G 103, K/Q/R 119, F/I/V 184, P/S 187 and K/E 248) with a posterior probability greater than 98%. Residues 103, 119 and 184 were confirmed by single-likelihood ancestor counting (SLAC) or fixed-effects likelihood (FEL) methods, and residue 187 was the only one found by all methods. Of the eight sites predicted to be under positive selection by PAML, seven were found in the B30.2 domain and, more specifically, five (100, 102, 103, 184 and 187) mapped to predicted loops that shape surface A of the SPRY domain, based on comparison with SPRYSEC 19 (Murrin and Talbot, 2007) (Fig. 4).

Table 1 Analysis of molecular variance (AMOVA) of the Gp-Rbp-1 dataset under two different partitionings.

| Source of variation | Degree of freedom | Sum of squares | Variance components | Percentage of variation |
|---------------------|-------------------|----------------|---------------------|------------------------|
| (A)                 |                   |                |                     |                        |
| Among groups        | 7                 | 736.93         | 1.98                | 11.94                  |
| Among populations within groups | 14 | 931.50         | 8.75                | 52.68                  |
| Within populations  | 135               | 793.53         | 5.88                | 35.38                  |
| Total               | 156               | 2461.96        | 16.61               |                        |
| (B)                 |                   |                |                     |                        |
| Among evolutionary pathways | 2  | 294.25         | 2.94                | 28.20                  |
| Among populations within evolutionary pathways | 19  | 432.96         | 2.51                | 24.11                  |
| Within populations  | 135               | 671.28         | 4.97                | 47.69                  |
| Total               | 156               | 1398.49        | 10.42               |                        |

AMOVA considering: (A) eight geographical groups (five Peruvian clades, one Chilean group, one P309 group and one European group); or (B) three evolutionary pathways (‘Northern Peru path’, ‘Peru clade I/European path’ and ‘Chilean path’). In each case, the variance was calculated among the different groups or pathways, among the 22 populations within each group or pathway and within the 22 populations. Sequences obtained from Peruvian populations of clades II–V were assigned to the ‘Northern Peru path’, sequences obtained from European populations and from Peru clade I populations were assigned to the ‘Peru clade I/European path’, and sequences from the P309 and Tierra del Fuego populations were included in the ‘Chilean path’.
We identified 33 haplotypes (Table S1, see Supporting Information) corresponding to all observed combinations of the eight sites found under positive selection. The eight most frequent haplotypes (haplotypes 1–8) are represented by 9–23 sequences and include two-thirds of the sequences. Among these most frequent haplotypes, it should be noted that: (i) all P309 and Tierra del Fuego Gp-Rbp-1 sequences correspond to the same haplotype (haplotype 7), which was only found in these two populations; and (ii) haplotype 8 seems to be restricted to Peru clade I populations. Loss of haplotype 8 in Europe and North Peru may have been caused by genetic drift or by a founder effect. At least two to seven haplotypes were identified in most of the studied populations (mean number of haplotypes per population, 3). Populations D383, Guiclan, P235, P308, P309 and Tierra del Fuego were represented by only one haplotype, which may have been a result in some cases of the small number of sequences retained in the final sequence dataset. In addition to the D383 population, we recovered only avirulent variants of Gp-Rbp-1 (i.e. P-187 sequences) in five other populations: Luffness, Pukekoe, P205, P212 and P235. Nearly all the polymorphisms observed for the eight sites under positive selection were found in the ‘Northern Peru path’ and in the ‘Peru clade I/European path’ (Fig. 5). The exceptions were R-37, which is missing in the ‘Northern Peru path’, and R-119, which is missing in the ‘Peru clade I/European path’. Residues H-100, A-103 and I-184 were predominant in each evolutionary pathway (present in more than 58% of the sequences). Residues G-103, F-184 and K-248 were found in less than 20% of the sequences belonging to the ‘Peru clade I/European path’ and the ‘Northern Peru path’, and were missing in the ‘Chilean path’.

The P-187 position plays a predominant role in the recognition of Gp-Rbp-1 by Gpa2

Among the 33 haplotypes identified previously, 19 have a serine amino acid in position 187, but show variability for the seven other sites under positive selection. When only six of the sites, that appeared to be supported by several methods and/or were predicted to be localized to the loops of the B30.2 domain, were considered, 12 haplotypes still contained the most frequent S-187 residue. We tested whether, among these 12 haplotypes (haplotypes 4, 5, 7, 8, 11, 14, 16, 22, 23, 24, 28 and 29), sequence variation at the sites under positive selection had an effect on the interaction with GPA2. For this, 10 Gp-Rbp-1 variants were transiently expressed with Gpa2 in tobacco leaves (haplotypes 4 and

| Table 2 | Choice of the best evolutionary model fitting the Gp-Rbp-1 dataset. |
|---------|---------------------------------------------------------------|
| Models  | Null model | Positive selection model | Null model | Positive selection model |
| ln(likelihood) | -4056.08 | -4009.35 | -4059.88 | -4009.62 |
| 2\(\Delta\)l | 93.46 | 100.52 | 9.21 | 9.21 |
| \(\chi^2\) (\(n = 2, P = 0.01\)) | 9.21 | 9.21 |

To identify the model that best fitted our data, we used PAML (Phylogenetic Analysis by Maximum Likelihood). Null models (M1 and M7) are compared with positive selection models (M2 and M8) using a likelihood score for each model. The significance of likelihood score difference was estimated by comparing the null model and positive selection model (2\(\Delta\)l) with a \(\chi^2\) table (likelihood ratio test, LRT). In the two comparisons (M1vsM2 and M7vsM8), 2\(\Delta\)l > 9.210 indicated that the positive selection models were a better fit than the null models with our dataset.

Fig. 4 Distribution of the dN/dS ratio along the Gp-Rbp-1 amino acid sequence. Analyses were conducted using the CODEML module of PAML on the Gp-Rbp-1 sequences. Amino acid variants found to be subjected to positive selection with a posterior probability of >95% are indicated above each site. Positions under positive selection that were supported by PAML and at least one other method (single-likelihood ancestor counting (SLAC) or fixed-effects likelihood (FEL)) are indicated with a star. Sequence portions corresponding to the SPRYSEC extended loops in the B30.2 protein structure are shaded in grey, and amino acids under positive selection found in theses portions are boxed. The positions of the predicted different domains (SP, PRY and SPRY) are indicated below the graph.

© 2011 THE AUTHORS
MOLECULAR PLANT PATHOLOGY © 2011 BSPP AND BLACKWELL PUBLISHING LTD

MOLeCUlAR PLaNT PaTHOlOgy (2012) 13(6), 546–555
22 were tested previously in Sacco et al. (2009). Four days after agro-infiltration, for each experimental replicate, we observed an HR with our positive control, but not with the 10 Gp-Rbp-1 variants expressed with Gpa2 (Fig. S1, see Supporting Information), as observed previously for haplotypes 4 and 22 in Sacco et al. (2009). Although it was not possible to test all the haplotype combinations and despite the detection of novel sites under positive selection, our results support the previously suggested idea that the recognition of Gp-Rbp-1 by Gpa2 is fully controlled by the proline residue at position 187.

**DISCUSSION**

One of the main objectives of this study was to broadly explore the variability of the Gp-Rbp-1 gene in G. pallida populations. Our results revealed a considerable sequence variability, which could be assigned to three evolutionary pathways. Furthermore, Gp-Rbp-1 appears to have been subject to positive selection and this selection appears to have mostly affected sites within the encoded protein.

All variants of Gp-Rbp-1 in our dataset were amplified from G. pallida cDNA with a single set of specific primers, and showed best BLAST hits with Gp-Rbp-1. However, as Gp-Rbp-1 belongs to the SPRYSEC multigene family, we cannot rule out the possibility that some of the sequences obtained are not true Gp-Rbp-1 variants, but actually represent other members of the SPRYSEC family. To limit this risk, we conducted our analysis on a reduced sequence dataset in which only sequences containing a SPRY subdomain length of 129 amino acids were retained. We also removed 12 sequences that had large indels in the B30.2 domain. However, these sequences may represent Gp-Rbp-1 variants obtained through alternative splicing. Alternative splicing events represent one of the main sources of proteomic diversity in multicellular eukaryotes (Nilsen and Graveley, 2010). Some cases of alternative splicing of putative effector proteins have been described in plant parasitic nematodes (Yu et al., 2010; Zahler, 2005). It is known that Gp-Rbp-1 contains six introns, including intron 3, which is located between positions 243 and 244 of the most commonly identified variant of the Gp-Rbp-1 cDNA (Blanchard et al., 2005). It is therefore probable that the large indels observed in some of the sequences correspond to intron 3 retention (+45 nucleotides) or exon 4 skipping (~72 nucleotides). Based on the previously published results by Sacco et al. (2009) and our agro-infiltration experiments (data not shown), it appears that both types of splice variant are still able to trigger Gpa2 recognition as long as they encode a proline at residue 187, although the timing and intensity of the response varied.

As shown by AMOVA, a substantial part (one-third to nearly one-half) of the variability remains at the intrapopulation level. This variation represents the interindividual variation that can be observed in cysts (cysts contain hundreds of larvae), Globodera pallida reproduces by obligate outcrossing and intrapopulation variation can also be increased when a single female is copulated by several males. A high level of intrapopulation variability has been reported previously for the nematode pectate lyase gene,
Globodera pallida has co-evolved for a long time with wild plants rather than cultivated plants. We can thus hypothesize that the considerable intrapopulation sequence variation observed in the genes involved in the plant–nematode interaction is related to this history of co-evolution with the many wild Solanum species that can be found in South America, the native area of this nematode.

Gp-Rbp-1 evolution appears to have occurred independently in the five phylogeographical clades described for the G. pallida Peruvian populations (Picard et al., 2007). In particular, populations from clade I were not the most polymorphic for Gp-Rbp-1 and we did not observe a reduction in diversity from south to north Peru. The maximum amount of diversity was observed in G. pallida populations sampled around latitude 13°S (geographical region of clades II and III), where Hijmans and Spooner (2001) previously observed the largest number of wild potato species (around 25 species in 250 km²). Thus, the diversity of Gp-Rbp-1 sequences appears to be linked to host plant diversity. Moreover, our results suggest that Gp-Rbp-1 was influenced by three evolutionary pathways that correlated with major dispersion events which represented drastic changes in terms of host richness and climate. The first evolutionary pathway, which we called the ‘Northern Peru path’, affects most of the Gp-Rbp-1 sequences of G. pallida populations from Peruvian clades II–V. This evolutionary pathway may be linked to the Andean uplift which impacted the landscape, climate and ecosystem during the pre-Quaternary (Hoorn et al., 2010). The second evolutionary pathway, called the ‘Peru clade I/European path’, affects nearly all Gp-Rbp-1 sequences from Peruvian clade I and Europe. It can be linked to the recent introduction of G. pallida from a Peruvian clade I population into Europe during the 19th century. It has been established that all the European populations of G. pallida come from a single restricted area on the north shore of Lake Titicaca (Plantard et al., 2008). Therefore, it is not surprising that the most frequent European Gp-Rbp-1 haplotypes are also very common in clade I Peruvian populations. Such a new European environment with only one host species (S. tuberosum spp. tuberosum), long photoperiod and different climatic conditions could explain the emergence of another evolutionary pathway that has affected Gp-Rbp-1 evolution. Furthermore, five rare Gp-Rbp-1 haplotypes (haplotypes 14, 21, 22, 23 and 32) found in Europe appear to be absent from South America. The last evolutionary pathway, the ‘Chilean path’, includes all Gp-Rbp-1 sequences obtained for P309 and Tierra del Fuego. These sequences have diverged more than the European sequences from the clade I sequences. Thus, it is probable that this evolutionary pathway, like the ‘Northern Peru path’, occurred during geological time. In a similar manner, the Andes orogeny also occurred southward from Lake Titicaca (towards Tierra del Fuego), and our results reveal that the evolution of Gp-Rbp-1 in Chile does not mirror the pattern observed in northern Peru. It seems that a distinct evolutionary trajectory occurred in Chile, probably because of the presence of more contrasted ecosystems and/or a more fragmented distribution of wild potatoes in this area (Hijmans and Spooner, 2001). This Chilean evolutionary pathway needs to be confirmed with other Chilean or Argentinean populations and sequences. Surprisingly, Gp-Rbp-1 sequences isolated in P309, a population sampled in the south of Peru, appear to be genetically closer to the Chilean Gp-Rbp-1 sequences than to Peruvian clade I Gp-Rbp-1 sequences. It is thus highly likely that P309 represents an anthropogenic introduction event to a Chilean population in the south of Peru, rather than a native Peruvian G. pallida population.

We have shown that the evolution of Gp-Rbp-1 has not strictly followed—at least in Peru—the phylogeographical history of the nematode population. Gp-Rbp-1 diversity seems rather to have been influenced by passive and anthropogenic dispersion of G. pallida in the American and European continents as well as climate variations that have influenced the nature and richness of wild host species. This evolutionary pattern is not surprising considering that SPRYSEC proteins are thought to be able to interact with a range of host proteins (Jones et al., 2009; Rehman et al., 2009). Thus, we can assume that the evolution of other such effectors was also strongly impacted by environmental changes in terms of host richness, as observed for the evolution of the Gp-Rbp-1 gene.

Our analysis of the selection pressures acting on Gp-Rbp-1 sequences have revealed eight sites under positive selection: six are described for the first time in this study (the remaining two were found previously by Sacco et al., 2009). Clearly, we were able to detect more sites under positive selection because we used more divergent populations, and thus circumvented one of the major limitations of the dN/dS approaches that have been developed and optimized to study interspecies rather than intraspecies’ divergences (Goldman and Yang, 1994; Kimura, 1977; Muse and Gaut, 1994). Five of the eight sites predicted to be under positive selection were found in the B30.2 domain and mapped to predicted loops that shape surface A of the SPRY domain based on a comparison with SPRYSEC 19 (Murrin and Talbot, 2007). Similar results were obtained in a SPRYSEC protein of Globodera rostochiensis, where seven sites under positive selection were also found in these extended loops (Rehman et al., 2009). In Phytophthora sp., it has been shown that most effector genes reveal hallmarks of positive selection and that this kind of selection is implicated in adaptation to the host (Raffaele et al., 2010). Thus, we hypothesized that the eight sites in Gp-Rbp-1 and, more specifically, the five sites in the loops shaping the hypervariable surface A represent key positions that may influence the outcome of the interaction with host proteins, such as the resistance protein GPA2. This view is supported by the fact that P/S187, which was identified by all the methods used, has been shown previously to be an important variation allowing the recognition or not of Gp-RBP-1 by GPA2 in tobacco. However, none of these other sites appear to confer the ability to trigger an
HR to the S-187 Gp-Rbp-1 variants when expressed with Gpa2. The P/S polymorphism observed at position 187 therefore remains the only variation which plays a significant and direct role in the Gp-Rbp-1 – Gpa2 interaction. It is still possible that the seven other sites identified in this study quantitatively affect the Gpa2 – P-187 Gp-Rbp-1 interaction, but they may also simply be involved in interactions with other plant host proteins.

Interestingly, as for the D383 population, only avirulent Gp-Rbp-1 variants (P-187) were recovered from the European population Luffness and the New Zealand population Pukekeo. This result is surprising because potatoes harbouring the Gpa2 resistance gene are susceptible to these G. pallida populations (E. Grenier, unpublished results). Thus, Gpa2-mediated resistance in potato to G. pallida nematodes is not only explained by the nature of the amino acid site at position 187 in Gp-Rbp-1. The absence of virulent Gp-Rbp-1 variants (S-187) in the Luffness and Pukekeo populations in our dataset could be explained by: (i) the number of sequences, which may not have been sufficiently large to detect the full extent of sequence mixtures within each of these two populations; or (ii) the presence of another effector in the Luffness and Pukekeo populations (but which is absent from D383) which suppresses the effector-triggered immunity induced when Gpa2 recognizes an avirulent Gp-Rbp-1 variant.

**EXPERIMENTAL PROCEDURES**

**Nematode populations**

Twenty-two G. pallida populations were analysed: eight from Europe (Duddingston, Pukekeo, Rookmaker, Guiclan, Luffness, Ouessant, Chavor, Glasgow, UK). One microlitre of 3 G. pallida in potato to Gp-Rbp-1 sequences were amplified by PCR using the RT reactions as a template. PCRs were carried out in a final volume of 25 μL using 3 μL of the RT reaction, 1 μM of the IC5-4 forward primer (5′-TTTTTTGGCTCAAAATGGCGC-3′), 1 μM of the IC5-11 reverse primer (5′-ACAGCAAAACCATCAAATTCTC-3′), 0.2 μM of each dNTP, 1 unit of Taq polymerase (Promega, Madison, WI, USA), 2 mM MgCl2 and 5 μL of the provided 10 × Ex Taq buffer. The PCR protocol was 5 min at 98 °C, followed by 35 cycles of 10 s at 98 °C, 30 s at 51 °C and 1 min and 45 s at 72 °C, and a final step of 10 min at 72 °C.

**Gp-Rbp-1 cloning, sequencing and sequence alignments**

PCR products were cloned using the Stratagene™ PCR cloning kit (Agilent Technologies, Palo Alto, CA, USA). Eight to twelve transformed colonies per cloning were picked and used to amplify the insert. PCRs were performed in a final volume of 30 μL containing 0.5 μM of M13 forward and reverse primers, 0.2 mM of each dNTP, 0.75 unit of Taq polymerase (Promega), 1.5 mM MgCl2 and 6 μL of the provided buffer (Promega). The PCR protocol started with 5 min at 95 °C and the 30-cycle amplification profiles were as follows: 1 min at 92 °C, 1 min at 55 °C and 3 min at 72 °C, and a final step of 5 min at 72 °C.

PCR products were sequenced by Genoscreen using the M13-20 forward primer as sequencing primer. When possible, ambiguous nucleotides were manually corrected based on the chromatograms. New Gp-Rbp-1 sequences analysed in this study were deposited in GenBank/EMBL (GenBank: JF933771–JF933897, Table S1). We completed our dataset with 33 sequences already available in GenBank/EMBL databases: 12 sequences from the three populations D383, Rookmaker and Guiclan [GenBank: EF423893–EF423896 (D383), EF423897–EF423902 (Rookmaker), FJ882986.1 and EU982199.1 (Guiclan)] and 21 sequences from seven populations already represented in our dataset [AM491352–AM491355 (Chavornay), FJ392678.1 and FJ392677.1 (Chavornay), FJ882983.1–FJ882985.1 (Pukekeo), EU982198.1 (Duddingston), FJ882983, FJ882997 and FJ882998 (Humachuco), FJ882987.1–FJ882989.1 (Pukekeo), FJ882990.1–FJ882992.1 (Pukekeo), FJ882993.1 and FJ882995.1 (Ouessant)]. As such, the final dataset comprised DNA sequences from 22 G. pallida populations. DNA sequences, once translated into protein sequences, were aligned using CLUSTAL.W. The alignment of the different length repeats in the PRY motif was manually corrected. Motif detection in translated DNA sequences was performed using SMART (Simple Modular Architecture Research Tool) (Schultz et al., 1998) and InterProscan (Zdobnov and Apweiler, 2001).

**Phylogenetic and evolutionary analyses**

Phylogenetic analysis was performed using the DARwin program (Perrier and Jacquemoud-Collet, 2006). A dissimilarity matrix was calculated using the Kimura model (Kimura, 1980), 80% minimal proportion of valid data required for each unit pair and 1000-replicate bootstrapping. A factorial analysis was carried out based on the previously calculated dissimilarity matrix.

The diversity indicator ρ and its standard deviation for each G. pallida Peruvian population were estimated with DnaSP v5 software (Librado
and Rozas, 2009). This indicator, corresponding to the average number of nucleotide differences per site between two sequences, was used to compare nematode population diversity following their latitude. Evaluation of the intragroup and intergroup/clade variability was performed by an AMOVA using Arlequin 3.1 software (Excoffier et al., 2005). The distance matrix between the different clades/groups was estimated with MEGA version 4 software (Tamura et al., 2007) using a Kimura two-parameter model (Kimura, 1980) with a gamma parameter of 0.45. Nucleotide sites with more than 20% gap/missing data were not considered in this calculation.

To evaluate the selective pressures on Gp-Rbp-1, we used the ratio of nonsynonymous to synonymous substitution rates per site ($d_N/d_S$), estimated by site-specific models implemented in the PAML package version 3.14 (Yang, 1997). The CODEML program of PAML assigns a likelihood score to models for selection. A likelihood score for a model incorporating positive selection (M2 or M8), which is higher than that for a null model without positive selection (M1 or M7), is evidence for positive selection. We estimated the significance of the differences by comparing the null model and positive selection model ($w_\text{LRT}$; Table 2). To identify the sites under positive selection, we used different methods, such as Bayes empirical Bayes implemented in CODEML, which calculates the posterior probabilities that each site falls into a different $\omega$ class, SLAC (Kosakovsky Pond and Frost, 2005) and FEL (Kosakovsky Pond and Frost, 2005), both available through the DataMonkey web interface (Delport et al., 2010).

Plasmid construction and transient expression

For the generation of Gp-Rbp-1 expression clones, the different inserts were ligated into the 5′ XbaI and 3′ BamHI sites of the pBIN61 binary vector series. For this, Gp-Rbp-1 clones Otuz-1-JC (JF933857), P309-5-JC (JF933824), P308-1-JC (JF933893), Choc-10-JC (JF933850), Guic-4-DF (EU982199.1), Choc-7-JC (JF933852), Oues-11-JC (JF933791), Huaem-16-DF (FJ882966.1) and Choc-9-JC (JF933854) were amplified with the primers GpRBPMPMexba (5′-CTTCTAGACCATGGAGTCGCCAAAACCAAAC-3′) and Rbp1stopRev (5′-GGATCCGCAAACCCATCATAAATTCTCG-3′) in order to remove the SP domain and to add XbaI and BamHI restriction sites at the 5′ and 3′ ends of the PCR product, respectively. After purification on agarose gel, PCR products were first ligated into a pSC-A-amp/kan vector (Strataclone PCR Cloning kit), digested with XbaI and BamHI restriction enzymes and then ligated into a pBIN61 vector.

Nicotiana benthamiana plants were germinated and grown in a growth chamber maintained at 20 °C. For the expression of proteins, the plants were infiltrated by syringe with Agrobacterium tumefaciens strain C58C1 carrying the virulence plasmid pCH32 and the appropriate pBIN61 binary expression vector. Agrobacterium cultures were diluted to an optical density at 600 nm (OD$_{600}$) = 1 and co-infiltrated at a final OD$_{600}$ = 0.5. Plants were transferred to a growth chamber maintained with 16-h light and 8-h darkness at 20 °C for 3–5 days. All experiments were repeated twice on at least two leaves of two different plants.

ACKNOWLEDGEMENTS

The authors would like to thank COST 872 for fruitful discussions and advice concerning the analyses. We thank Josselin Montarry and Peter Moffett for critical reading of the manuscript. We also thank Peter Moffett for providing us with several expression vectors and Bruno Faverey for the Agrobacterium strains. JC is supported by the Région de Bretagne through a 3-year PhD grant (nb. 08005871). We also thank the INRA department ‘Santé des Plantes et Environnement’ for financial support.

REFERENCES

Blanchard, A., Esquibel, M., Fouville, D. and Grenier, E. (2005) Ranbpm homologue genes characterised in the cyst nematode Globodera pallida and Globodera mexicana. Physiol. Mol. Plant Pathol. 67, 15–22.

Davis, E.L., Hussey, R.S., Mitchum, M.G. and Baum, T.J. (2008) Parasitism proteins in nematode-plant interactions. Curr. Opin. Plant Biol. 11, 360–366.

Delport, W., Poon, A.F., Frost, S.D. and Kosakovsky Pond, S.L. (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics, 26, 2455–2457.

Excoffier, L., Laval, G. and Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. Online, 1, 47–50.

Flor, H.H. (1955) Host–parasite interaction in flax rust—its genetics and other implications. Phytopathology, 45, 680–685.

Geric Stare, B., Fouville, D., Sirca, S., Gallot, A., Urek, G. and Grenier, E. (2011) Molecular variability and evolution of the lectate lyase (pel-2) parasitism gene in cyst nematodes parasitizing different solanaceous plants. J. Mol. Evol. 72, 169–181.

Goldman, N. and Yang, Z.H. (1994) Codon-based model of nucleotide substitution for protein-coding DNA-sequences. Mol. Biol. Evol. 11, 725–736.

Grenier, E., Fourent, S., Petel, I. and Anthoine, G. (2010) A cyst nematode ‘species factory’ called the Andes. Nematology, 12, 163–169.

Hijmans, R.J. and Spooner, D.M. (2001) Geographic distribution of wild potato species. Am. J. Bot. 88, 2101–2112.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sannart, I., Sanchez-Meseguer, A., Anderson, C.L., Figueredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghmiestra, H., Lundberg, J., Stadler, T., Sarkinen, T. and Antonelli, A. (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science, 330, 927–931.

Jones, J.D.G. and Dangl J.L. (2006) The plant immune system. Nature, 444, 323–329.

Jones, J.T., Amar, K., Pylypenko, L.A., Amarnath, T., Castelli, L., Chapman, S., Cock, P.J.A., Grenier, E., Lilley, C.J., Phillips, M.S. and Blok, V.C. (2009) Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode Globodera pallida. Mol. Plant Pathol. 10, 815–828.

Kosakovsky Pond, S.L. and Frost, S.D. (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22, 1208–1222.

Librado, P. and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25, 1451–1452.

Murrin, L.C. and Talbot, J.N. (2007) RanBP1, a scaffolding protein in the immune and nervous systems. J. Neuroimmune Pharmacol. 2, 290–295.

Muse, S.V. and Gaut, B.S. (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. 11, 715–724.

Nilsen, T.W. and Graveley, B.R. (2007) Expansion of the eukaryotic proteome by alternative splicing. Nature, 463, 457–463.

Perrier, X. and Jacquemaud-Collet, J.P. (2006) DARwin software. Available at http://darwin.crdi.fr/darwin.

Picard, D., Sempre, T. and Plantard, O. (2007) A northward colonisation of the Andes by the potato cyst nematode during geological times suggests multiple host-shifts from wild to cultivated potatoes. Mol. Phylogenet. Evol. 42, 308–316.

Plantard, O., Picard, D., Valette, S., Scarratt, M., Grenier, E. and Mugniery, D. (2008) Origin and genetic diversity of Western European populations of the potato cyst nematode (Globodera pallida) inferred from mitochondrial sequences and microsatellite loci. Mol. Ecol. 17, 2209–2218.

Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H.Y., Zody, M.C., Kunjeti, S.G., Donofrio, N.M., Meyers, B.C., Nusbaum,
C. and Kamoun, S. (2010) Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science, 330, 1540–1543.

Rehman, S., Postma, W., Tytgat, T., Prins, P., Qin, L., Overmars, H., Vossen, J., Spiridon, I.N., Petrescu, A.J., Govers, A., Bakker, J. and Smant, G. (2009) A secreted SPRY domain-containing protein (SPRYSEC) from the plant-parasitic nematode Globodera rostochiensis interacts with a CC-NB-LRR protein from a susceptible tomato. Mol. Plant-Microbe Interact. 22, 330–340.

Sacco, M.A., Koropacka, K., Grenier, E., Jaubert, M.J., Blanchard, A., Govers, A., Smant, G. and Moffett, P. (2009) The cyst nematode SPRYSEC protein RBP-1 elicits Gpa2- and RanGAP2-dependent plant cell death. PLoS Pathog. 5, e1000564.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA, 95, 5857–5864.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Woo, J.-S., Imm, J.-H., Min, C.-K., Kim, K.-J., Cha, S.-S. and Oh, B.-H. (2006) Structural and functional insights into the B30.2/SPRY domain. EMBO J. 25, 1353–1363.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yu, H., Chronis, D., Lu, S. and Wang, X. (2010) Chorismate mutase: an alternatively spliced parasitism gene and a diagnostic marker for three important Globodera nematode species. Eur. J. Plant Pathol. 129, 89–102.

Zahler, A.M. (2005) Alternative splicing in C. elegans. WormBook, Sep 26, 1–13.

Zdobnov, E.M. and Apweiler, R. (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics, 17, 847–848.

Fig. S1 Agrobacterium transient expression assay using nine S-187 Gp-Rbp-1 haplotypes. Gp-Rbp-1 variants cloned into pBIN61 were transiently expressed via agro-infiltration in Nicotiana benthamiana leaves in combination with either the empty vector or the Gpa2 potato resistance gene. Hypersensitive response (HR) phenotypes obtained 4 days after agro-infiltration are shown. The Gp-Rbp-1 variant Rook-6 (EF423902.1), co-expressed with Gpa2, was used as a positive control. Numbers 0–9 represent the Gp-Rbp-1 variants Rook-6, Otuz-1-JC, Choc-10-JC, P309-5-JC, P308-1-JC, Guic-4-DF, Choc-7-JC, Choc-9-JC, Huam-16-DF and Oues-11-JC, respectively; (A) and (EV) indicate whether the Gpa2 potato resistance or the empty vector gene, respectively, was co-infiltrated.

Table S1 Haplotypes and GenBank accession numbers of the 157 Gp-Rbp-1 sequences. Haplotypes were defined solely on the basis of the amino acid combination at the eight sites found to be under positive selection. Each line corresponds to one sequence with a specific amino acid at positions 37, 100, 102, 103, 119, 184, 187 and 248. Haplotype 1 was the most frequent haplotype and haplotype 33 was the rarest haplotype.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.