Journal of Nematology 44(1):7–17. 2012.

A Comparison of Three Molecular Markers for the Identification of Populations of *Globodera pallida*

Angelique H. Hoolahan,[1] Vivian C. Blok,[2] Tracey Gibson,[1] Mark Dowton[1]

Abstract: Potato cyst nematodes cost the potato industry substantial financial losses annually. Through the use of molecular markers, the distribution and infestation routes of these nematodes can be better elucidated, permitting the development of more effective preventative methods. Here we assess the ability of three molecular markers to resolve multiple representatives of five *Globodera pallida* populations as monophyletic groups. Molecular markers included a region of the *rbp-1* gene (an effector), a non-coding nuclear DNA region (the ITS region), and a novel marker for *G. pallida*, a ~3.4 kb non-coding mitochondrial DNA (mtDNA) region. Multiple phylogenetic analysis methods were performed on the three DNA regions separately, and on a data set of these three regions combined. The analyses of the combined data set were similar to that of the sole mtDNA marker; resolving more populations as monophyletic groups, relative to that of the ITS region and *rbp-1* gene region. This suggests that individual markers may be inadequate for distinguishing populations of *G. pallida*. The use of this new non-coding mtDNA marker may provide further insights into the historical distribution of *G. pallida*, as well as enable the development of more sensitive diagnostic methods.

Key words: Diagnostics, genetics, genetic distance, *Globodera pallida*, internal transcribed spacer, mitochondrial DNA, phylogenetics, potato cyst nematode, *rbp-1*, systematics, effector.

The potato cyst nematode (PCN) *Globodera pallida* is a significant agricultural pest. Annually, *G. pallida* and *Globodera rostochiensis* are responsible for the loss of over 12% of potato yields worldwide (Bates et al., 2002), with associated costs estimated at £43 million in the UK (Haydock and Evans, 1998) and £300 million in Europe (Ryan et al., 2000). Infestations of PCN are most extensive in Europe, South America, and New Zealand, but are also reported for India and Africa (Evans and Stone, 1977). However, in North America, PCN are rare, and in Australia, only *G. rostochiensis* has been reported (Hafez and Sundararaj, 2007). In many countries there are strict quarantine regulations aimed at preventing the introduction and spread of *G. rostochiensis* and *G. pallida* as it is predicted that an infestation of *G. pallida* in Australia, for example, will cost up to AUD$20 million annually (Hodda and Cook, 2009). Thus, there is a need to further understand the global migration of PCN, and to establish definitive diagnostic approaches. This also applies to the various populations of PCN which are recognized as pathotype variants and in particular for *G. pallida* for which single major gene resistance has not been found for use in potato breeding (Phillips and Trudgill, 1998).

There are many divergent populations of *G. pallida*, which most likely originated in Peru (Canto Saenz and De Scurrah, 1977; Evans and Stone, 1977; Picard et al., 2007), the origin of its main host, *Solanum tuberosum* (Hijmans and Spooner, 2001). The initial migration of *G. pallida* out of South America is suggested to have occurred during exportation of infected *S. tuberosum* to Europe or with guano, with global distribution likely radiating from there (Evans and Stone, 1977; Stone, 1985). As *G. pallida* has few distinguishing morphological characteristics, various molecular markers have been used to attempt to differentiate between these populations and elucidate their evolutionary histories. These include the nuclear internal transcribed spacer (ITS) region between (and including part of) the 18S and 28S ribosomal RNA genes (Blok et al., 1998; Subbotin et al., 2000; Pylypenko et al., 2005). The ITS region has also been used in conjunction with other molecular markers, including satellites (Grenier et al., 2001), and the mitochondrial DNA (mtDNA) gene, Cytochrome b (*Cytb*) (Madani et al., 2010). Satellites and *Cytb* have also been used as independent markers (e.g. Picard et al., 2004; Picard and Plantard, 2006; Pylypenko et al., 2008), and in conjunction with one another (Picard et al., 2007; Plantard et al., 2008). Although there are two copies of *Cytb* in the mtDNA on different mtDNA molecules in the atypical mitochondrial genome of *G. pallida*, they are 99% identical (Gibson, et al., 2007), permitting its use as a phylogenetic marker (Madani et al., 2010). This genome consists of several unique small circular mtDNA (scmtDNA) molecules, which have an overlapping genetic organisation, such that large multigenic regions are represented on several scmtDNAs (Armstrong et al., 2000; Gibson et al., 2007). Other, potential genetic markers include effectors, also called pathogenicity factors (Geric Stare et al., 2011). Several schemes have been proposed to group PCN based on their ability to reproduce on different hosts (Canto Saenz and De Scurrah, 1977; Kort et al., 1977). Thus, effector may be suitable as molecular markers to discriminate these pathotypes, and potentially different populations. This has only been assessed using a pectate lyase effector, with no

Received for publication October 27, 2011.
[1]School of Biological Sciences, University of Wollongong, Wollongong, NSW, 2522, Australia.
[2]Cell and Molecular Sciences, James Hutton Institute, Invergowrie, Dundee, DD2 5DA, UK.
The authors are grateful for the technical support and assistance of M. Phillips, A. Holt and A. Paterson is gratefully acknowledged. This work was supported by the Australian Research Council [DP0556520]; the University of Wollongong Internationalisation Committee [UIC2007]; the Australian Academy of Science [RI 146], and the Scottish Executive's Rural and Environment Research and Analysis Directorate Workpackage 1.5.4. The authors declare that they have no competing interests. Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession numbers HQ670282-HQ670316 and HQ730090- HQ730117.
E-mail: angeliqu@uow.edu.au
This paper was edited by Nancy Kokalis-Burelle.

correlation observed between pathotypes of PCN and the effector (Geric Stare et al., 2011). Another potential effector, the *rbp-1* gene, may be a useful genetic marker. The product of this gene, rbp-1, contains a SPRY domain homologous to that of *RanBPM* proteins (*Ras-related nuclear protein binding protein to microtubules*), and is suggested to be a member of secreted SPRY domain-containing (SPRYSEC) proteins (Rehman et al., 2009). Analysis of the cDNA of the rbp-1 showed substantial sequence variation for multiple *G. pallida* populations (Qin et al., 2000; Blanchard et al., 2005; Sacco et al., 2009). However, its suitability as a molecular marker is yet to be determined.

Here we assess the ability of three molecular markers to resolve various populations of *G. pallida* as reciprocally monophyletic groups. Populations from South America (P5A and P4A) and the UK (Luffness, Gourdie and Lindley) were analysed targeting a region of the *rbp-1* gene, the ITS region and a novel marker for *G. pallida*, a ~3.4 kb mtDNA region. This region is primarily non-coding, except for the ~0.5 kb of the 5' end coding for 16S ribosomal RNA. This region is specific to one scmtDNA molecule (Armstrong et al., 2000; Armstrong et al., 2007), and thus should be single-copy.

We also compared the phylogenetic relationships resolved between these populations using these three markers with that of published studies. Relative to the populations assessed in the present study, previous phylogenetic analyses have indicated that the South American populations P4A and P5A are the most ancestral, consistent with the hypothesis that *G. pallida* originated in South America (Canto Saenz and De Scurrah, 1977; Evans and Stone, 1977; Picard et al., 2007). Phylogenetic analysis of *Cytb* by Pylypenko et al. (2008) shows Luffness as one of the closest European/UK relatives to the South American populations. Madani et al. (2010) expanded on this study, adding additional *Cytb* sequences from other *G. pallida* populations to the analysis. The resulting phylogeny continued to indicate that Luffness has a more ancestral relationship with South American populations than did most other European/UK populations. Thus, it may be that one/several South American populations, possibly from a P4A type or hybrid population, gave rise to UK populations that are closely related to Luffness. Using multiple molecular approaches, Grenier et al. (2001) concluded that the South American populations were not likely direct ancestors of the European populations studied. Interestingly, 2-dimensional gel electrophoresis analysis showed the population Luffness as the most closely related European/UK population to the South American populations. A study by Plantard et al. (2008) reported that the origin of European/UK populations could be traced to a small region in the south of Peru. However, in the *Cytb* analysis (Madani et al., 2010), Luffness *Cytb* groups with South American populations that are from the southern region that Plantard et al. (2008) reported as the origin of European/UK populations (Sicuani and Arapa, Peru). Further studies are needed to determine whether a single or multiple introductions of *G. pallida* into Europe/UK gave rise to the current populations, and thus also whether the Luffness population is more ancestral to this/these introduced populations relative to other European/UK populations. Phylogenetically informative markers will improve our understanding of the historical distribution of *G. pallida*, enabling the prediction of infestation routes and the development of more sensitive diagnostic methods.

Materials and Methods

Retrieval of ITS and non-coding scmtDNA IV DNA sequences: Sequences of the ITS region (965 bp) from five populations of *G. pallida* were retrieved from GenBank (see Table 1). The populations included P5A and P4A from South America, and Luffness, Gourdie and Lindley from the UK. In total 24 sequences were retrieved, with accession numbers from HQ670253 to HQ670264 each representing two identical clones, and HQ670263 representing three identical sequences (i.e. 28 *G. pallida* ITS sequences in total were used for analysis). Sequences of the scmtDNA IV non-coding region (~3.4 kb) from population P4A were also retrieved from GenBank (see Table 1), totalling 10 clones. ITS and scmtDNA IV region sequences were generated by randomly sequencing multiple clones from amplifications of genomic DNA extracted from multiple pooled cysts; 5-8 cysts for the

Table 1. Accession numbers (inclusive) of sequences used for phylogenetic analysis of populations of *Globodera pallida*.

	Retrieved from GenBank	Generated in the present study	
Species and population	ITS region	*rbp-1*	scmtDNA IV region
G. pallida			
P5A	HQ670263-HQ670265[b]	HQ730111-HQ730116	HQ670308-HQ670316
P4A	HQ670257-HQ670262	HQ730106-HQ730110	DQ288929-DQ288939[a]
Luffness	HQ670253-HQ670256[b]	HQ730101-HQ730105[b]	HQ670299-HQ670307
Gourdie	HQ670247-HQ670252	HQ730096-HQ730100	HQ670290-HQ670298
Lindley	HQ670242-HQ670246	HQ730090-HQ730095	HQ670282-HQ670289
G. rostochiensis	GQ294521	HQ730117	EF193005

[a] The scmtDNA IV region of P4A was characterised by Armstrong et al. (2007).
[b] Some accession numbers are representative of clones with identical sequences; all sequences, including duplicates were used for analysis.

ITS region and 200 cysts for the scmtDNA IV region of P4A (Armstrong et al., 2007; Hoolahan et al., 2011). These studies used cysts from the population stocks at the James Hutton Institute. The ITS and non-coding mtDNA regions of *G. rostochiensis* were also retrieved from GenBank (see Table 1), which were used as outgroup sequences for phylogenetic analyses.

Sequencing of scmtDNA IV and rbp-1 sequences: The scmtDNA IV region was characterised in the four remaining populations; P5A, Luffness, Gourdie and Lindley, using the same approach described for characterising the ITS and mtDNA sequences; by randomly sequencing multiple clones from amplifications of genomic DNA extracted from 5-8 cysts for each of these five populations from the nematode collection at the James Hutton Institute. Sequences of the *rbp-1* gene from all five *G. pallida* populations were also characterised in this way. DNA was extracted by the method of Sunnucks and Hales (1996). Amplification of the scmtDNA IV non-coding region used the primers s86F and x222F (Armstrong et al. 2007), and the BIO-X-ACT^{TM} Long DNA Polymerase Kit (Bioline). Cycling conditions were 94°C for 2 min, 40 cycles of 94°C for 30 s, 57°C for 30 sec and 68°C for 5-10 min, followed by 68°C for 10 min. Amplification of the *rbp-1* gene region used the primers RBP-F (5'- GCTCTGTCTTCGCTGTTGAG-3') and RBP-R (5'- CCCGGAGCATACACCGTTA-3'), and either *Taq* (Promega) or BIOTAQ (Bioline) DNA polymerases. Cycling conditions were 94°C for 2 min, 40 cycles of 94°C for 30 s, 55° for 15 s, and 72°C for 1 min 30 s, with a final elongation of 72°C for 4 min.

PCR products were purified using either the MinElute® Gel Extraction Kit (QIAGEN) or the Wizard® SV Gel and PCR Clean-Up System (Promega), and ligated into the pGEM®-T Easy Vector System (Promega). Ligated products were transformed into either chemically competent JM109 *Escherichia coli* cells (Promega) or α-Select Electrocompetent cells (Bioline). Plasmid DNA was purified using either the GeneJET^{TM} Plasmid Miniprep Kit (Fermentas) or the Wizard® Plus SV Minipreps DNA Purification System (Promega). Sequencing reactions were performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), and sequences edited using ChromasPro (Technelysium Ltd., Tewantin) and BioEdit (Hall, 1999). Protein-coding sequences were conceptually translated using Transeq (http://www.ebi.ac.uk/Tools/emboss/transeq/index.html).

For use as an outgroup, *G. rostochiensis* genomic DNA was also extracted and a region of the *rbp-1* gene amplified and sequenced. DNA was extracted from a single cyst by the method of Sunnucks and Hales (1996). Amplification used the primers Rbp-F-Gro (5'-CGCTATGTGTTTGCTAAACAG-3') and Rbp-R-Gro (5'-CCAGGCATGAACATCGAAA-3'), and the BIOTAQ DNA polymerase as described above for *rbp-1* from *G. pallida*. The PCR product was directly sequenced following treatment with ExoSAP-IT® (USB Corporation).

Accession numbers of the sequences generated in the present study are shown in Table 1. For the scmtDNA IV region characterised in four *G. pallida* populations, a total of 120 clones were sequenced. For the *rbp-1* gene characterised in five *G. pallida* populations, a total of 28 clones were sequenced, with accession number HQ730104 representing two identical clones.

Sequence and phylogenetic analysis: Sequence data were aligned using MAFFT (multiple sequence alignment based on fast Fourier transform) (Katoh et al., 2002) using the generic mode for globally-alignable sequences (G-INS-I), as this approach was found to be most reliable in a comparison of six popular alignment approaches (Golubchik et al., 2007). Visual inspection of the alignments did not identify misalignments. cDNA sequences of the *rbp-1* gene of *G. pallida* (Sacco et al., 2009) and *G. rostochiensis* (Qin et al., 2000) were used to identify intron and exon boundaries. Coding and non-coding regions were aligned independently, and aligned sequences then concatenated to restore the data sets. For the non-coding mtDNA region of *G. rostochiensis*, a region homologous to the retrieved *G. pallida* sequences was identified (from 5504-8531 bp of accession number EF193005). Terminal gaps in the alignments which were known to be due to missing data were specified as a '?'. An additional data set was created containing all three DNA regions; the *rbp-1* gene region, the ITS region and the ~3.4 kb mtDNA region. Due to the different number of clones that had been sequenced from each gene and population, five randomly selected clones of each population (i.e. 25 sequences) and the outgroup from each DNA region were individually aligned, and subsequently concatenated. This number was chosen to minimise the amount of missing data in the combined dataset. The alignment of this combined data set has been submitted to TreeBase (http://purl.org/phylo/treebase/phylows/study/TB2:S11163). Thus, seven data sets were used for phylogenetic analysis; complete and 'reduced' data sets for each of the three individual DNA regions, and a combined data set including the three reduced data sets.

To quantify the divergence between sequences, genetic distances were calculated in MEGA4 (Tamura et al., 2007), specifying the Kimura-2-parameter (k2p) substitution model for DNA sequences (Kimura, 1980), and the p-distance model for amino acid sequences, with the pairwise deletion option specified for the treatment of gaps/missing data for both DNA and amino acid sequences (as recommended by Dwivedi and Gadagkar, 2009).

Neighbor-Joining (NJ) and Bayesian Inference (BI) methods of phylogenetic analysis were performed on each data set. NJ is a distance-based method in which a specified standard model of evolution is applied for the entire dataset. NJ was implemented in MEGA4 (Tamura et al., 2007), with the k2p model of substitution specified (Kimura, 1980), and the pairwise deletion option applied for gap treatment (as recommended by Dwivedi and Gadagkar, 2009). The reliability of the various nodes

resolved was assessed using 1000 bootstrap replications. The other phylogenetic method applied, BI, is a likelihood-based method that uses a Markov chain Monte Carlo algorithm, which runs multiple simultaneous analyses or 'chains', considering the average branch lengths, tree topologies, as well as evolutionary parameters of each, and predicts the phylogeny based on the distribution of these (Huelsenbeck et al., 2001). BI also allows the user to specify different models of evolution for different regions in a data set. BI was implemented in MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003), following prediction of the optimal model of evolution using MrModeltest, based on the Akaike information criterion (Nylander, 2004). The protein-coding gene, *rbp-1*, was modelled both in its entirety, and as a partitioned dataset where the first, second and third codon positions and introns were independently modelled. A comparison of the predictive likelihoods of these two models (the Bayes factor) was used to evaluate the two modelling approaches (Kass and Raftery, 1995). This indicated that the less complex model was least favoured, thus the partitioned model was used. However, Bayesian analyses of partitioned data sets are susceptible to becoming trapped in local optima (Marshall, 2010). To counter this, an additional command was employed [prset applyto=(all) brlenspr=Unconstrained: Exponential (100)]. BI analyses ran four independent Markov chain Monte Carlo chains for 1000000 generations each, sampling every 100th tree. Analysis of these independent chains indicated that stationarity (or "burnin") was reached at no later than 50,000 generations; thus, the first 500 trees were discarded from each analysis as the burnin. The remaining trees were used to generate a majority consensus tree. Trees were visualised using TreeView v1.6.6 (Page, 1996).

Results

Sequence divergence: The ability of three phylogenetic markers to resolve populations of *G. pallida* was assessed using multiple representatives of five populations; P5A and P4A from South America, and Luffness, Gourdie and Lindley from the UK. *G. rostochiensis* was used as an outgroup. The most suitable population marker will have the greatest difference between intra- (within) and inter- (between) population genetic variation, i.e. the genetic variation within populations will not mask the variation between populations. To assess this, the average intra- and inter-population k2p genetic distances were calculated for all representative of each DNA region (Table 2, and summarized in Table 3). The inter-species genetic distances, between *G. pallida* and *G. rostochiensis*, were also calculated. From these analyses, the *rbp-1* gene region appeared to be least suitable as a population marker,

Table 2. Average genetic distances between sequences of three DNA regions and a polypeptide, including representatives of the *G. pallida* populations from the UK, Lindley, Gourdie and Luffness, and from populations from South America, P4A and P5A. Intra-population variation is indicated in bold. Inter-species variation is also shown relative to *G. rostochiensis*.

		Globodera pallida				
	Lindley	Gourdie	Luffiness	P4A	P5A	*Globodera rostochiensis*
DNA sequences						
ITS region						
Lindley	**0.00689**	0.00746	0.00879	0.01012	0.01620	0.02929
Gourdie		**0.00333**	0.00554	0.00811	0.01804	0.03136
Luffness			**0.00775**	0.00837	0.01585	0.03059
P4A				**0.01018**	0.01258	0.02971
P5A					**0.00187**	0.02895
scIV non-coding						
Lindley	**0.00246**	0.00193	0.06821	0.13440	0.18209	0.55634
Gourdie		**0.00139**	0.06665	0.13259	0.18009	0.55619
Luffness			**0.00240**	0.14588	0.19198	0.55738
P4A				**0.10420**	0.08066	0.53251
P5A					**0.00249**	0.52674
rbp-1						
Lindley	**0.01250**	0.02727	0.01495	0.02538	0.01001	0.37245
Gourdie		**0.02223**	0.02465	0.02272	0.02603	0.37432
Luffness			**0.01657**	0.02237	0.01345	0.36918
P4A				**0.02281**	0.02354	0.37954
P5A					**0.00969**	0.37287
Amino acid sequences						
rbp-1						
Lindley	**0.0144**	0.0398	0.0203	0.0412	0.0116	0.4192
Gourdie		**0.0192**	0.0298	0.0361	0.0331	0.4181
Luffness			**0.0239**	0.0354	0.0180	0.4188
P4A				**0.0465**	0.0380	0.4112
P5A					**0.0103**	0.4232

with the least difference between intra- and inter-population genetic distances (0.0043). The *rbp-1* sequences were 851-884 bp, of which ~512 bp were from introns. The difference between intra- and inter-population genetic distance was slightly greater for the ITS region (0.0051), and a substantially greater difference for the scmtDNA IV region (0.0959). Thus, the scmtDNA IV region appears to be the most suitable marker. However, as the scmtDNA IV region from the population P4A was amplified from genomic DNA extracted from up to 40 times as many representative cysts as the other populations, these results may have been biased. To assess this, the analysis was repeated, but excluding the scmtDNA IV data from population P4A. The impact of this additional data appeared to be minimal, as the analysis did not substantially change the difference between intra- or inter-population genetic distances (Table 3).

To perform a preliminary assessment on the utility of the *rbp-1* protein sequences as a suitable marker, the exons were conceptually translated, aligned (Fig. 1), and pairwise p-distances calculated (Table 2, and summarised in Table 3). Inspection of the alignment indicated that sequence variation was not localised to specific populations. This was further supported by a small difference between intra- and inter-population genetic divergences (0.0075), suggesting that this would not be a more suitable marker than the three DNA markers analysed. Phylogenetic analysis was not performed on this protein sequence.

Phylogenetic analysis of individual data sets: NJ and BI phylogenetic analyses were preformed for each individual DNA region on both the complete (29-48 sequences) and reduced (26 sequences) data sets. Although NJ (k2p) analyses are the standard for species and population level analyses (e.g. Hebert et al., 2003), inadequate or oversimplified models can fail to resolve reciprocally monophyletic groups (Nelson et al., 2007). For each DNA region, similar phylogenies were generated using either the NJ or BI phylogenetic method, regardless of whether the complete or reduced data set was used. As the results from the reduced data set are most directly comparable with that of the combined data set (the same sequences were used), the results will hereafter focus on the phylogenies generated from these reduced data sets.

In analysing the utility of each phylogenetic marker, we expect the phylogenetic analyses to resolve sequences from the same population as a monophyletic group. NJ and BI analyses of the ITS region and *rbp-1* gene region did not resolve these relationships (NJ: Fig. 2a and 2b; BI: Fig. 3a and 3b). Sequences from the same population were infrequently resolved together, except for population P5A which was monophyletic in the ITS phylogeny. However, in the phylogeny of the complete ITS data set (data not shown), an additional sequence from population P4A resolves within the P5A clade, making it paraphyletic. The phylogeny of the *rbp-1* gene region was not improved by separate NJ analysis of intron and exon regions of the *rbp-1* gene (data not shown). NJ and BI analyses of the ~3.4 kb scmtDNA IV region (Fig. 2c and Fig. 3c, respectively) showed a marked improvement in population resolution, with the populations P5A and Luffness resolving as monophyletic groups. This result was true for analyses of the complete and reduced data sets.

As *G. pallida* is thought to have originated in South America (Canto Saenz and De Scurrah, 1977; Evans and Stone, 1977; Picard et al., 2007), we could also anticipate segregation between the South American (P5A and P4A) and UK (Luffness, Gourdie and Lindley) populations. However, this may not be the case, as the history of introduction/s of *G. pallida* into the UK are not firmly established. The phylogenies of the ITS region and *rbp-1* gene region did not demonstrate any segregation of South American and UK representatives. The phylogeny of the scmtDNA IV region had four predominant clades; I – all P5A representatives and two P4A sequences; II – two P4A representatives, III – all Luffness representative and a P4A sequence; and IV – all Gourdie and Lindley representatives (Fig. 2c and Fig. 3c). Thus, although most South American and UK representatives were segregated, several P4A sequences appear more closely related to the UK population Luffness. BI analysis resolved Clade II closer to other South American populations in comparison to NJ analysis, which resolved Clade II closer to UK populations.

The posterior probability and bootstrap support values indicate few nodes were well-supported in all analyses of the *rbp-1* gene region or ITS region. In

TABLE 3. Average genetic distances between and within sequences representative of 5 populations of *Globodera pallida*, and between these and *Globodera rostochiensis* based on three DNA regions and a polypeptide.

Target region	Average genetic distance			Difference between intra- and inter-population variation
	Inter-specific	Inter-population	Intra-population	
DNA sequences				
ITS region	0.0300	0.0111	0.0060	0.0051
scmtDNA IV non-coding region	0.5458	0.1185	0.0226	0.0959
Excluding P4A[a]	0.5492	0.1102	0.0022	0.1081
rbp-1	0.3737	0.0210	0.0168	0.0043
Amino acid sequences				
rbp-1	0.4181	0.0303	0.0229	0.0075

[a] Genomic DNA used for amplifying the scmtDNA IV fragment of population P4A had 25 times as many population representatives as any other population.

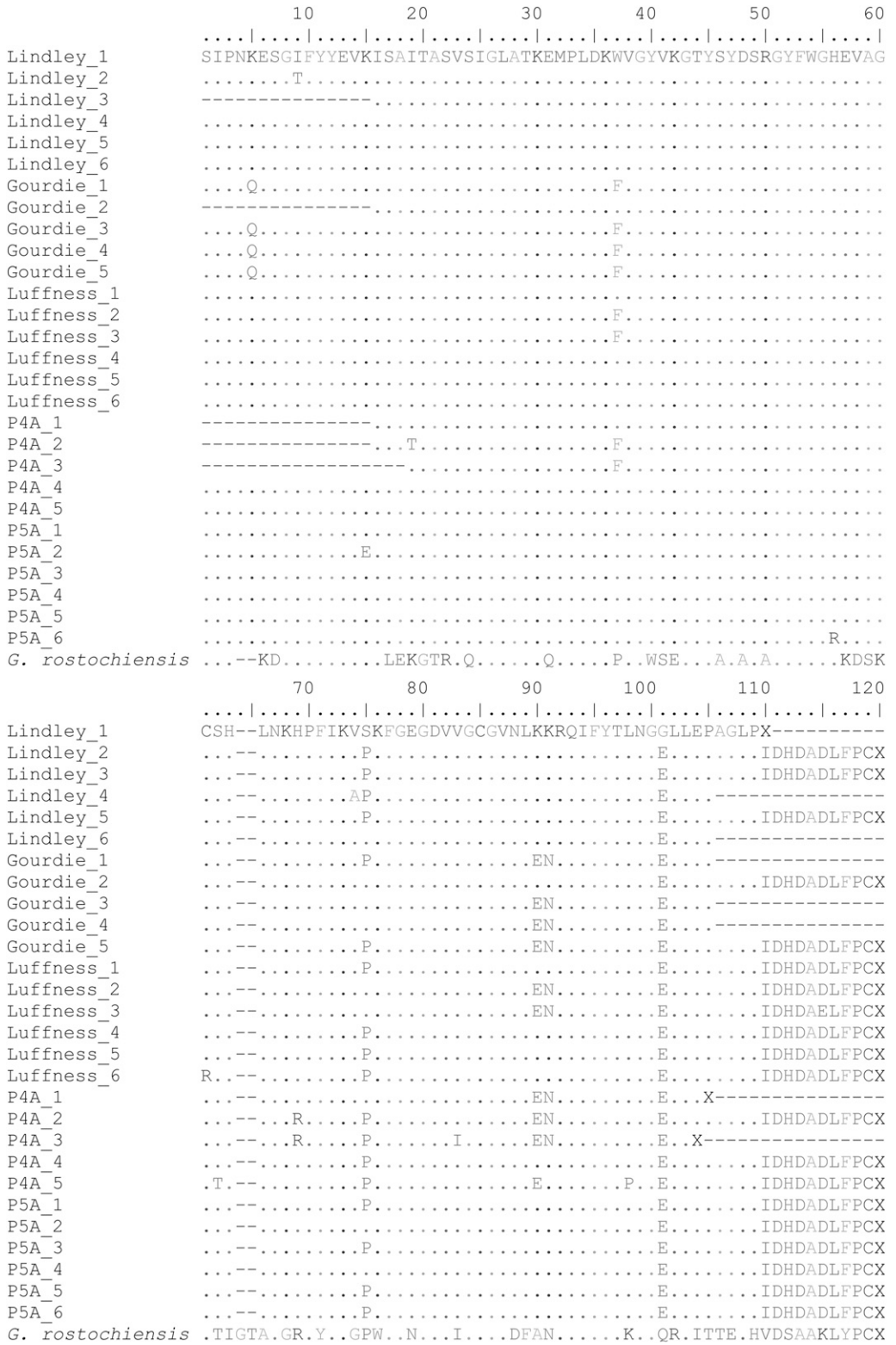

FIG. 1. Amino acid alignment of the translated region of the *rbp-1* gene sequenced from *Globodera pallida* populations and *Globodera rostochiensis*.

the phylogeny from the data sets of the scmtDNA IV region, the statistical support values for the monophyletic clades were considerably higher than that for the intra-population relationships.

Phylogenetic analysis of the combined data set: NJ and BI analyses were also performed on the combined data set, which contained the reduced data sets from the *rbp-1* gene region, the ITS region and the scmtDNA IV region

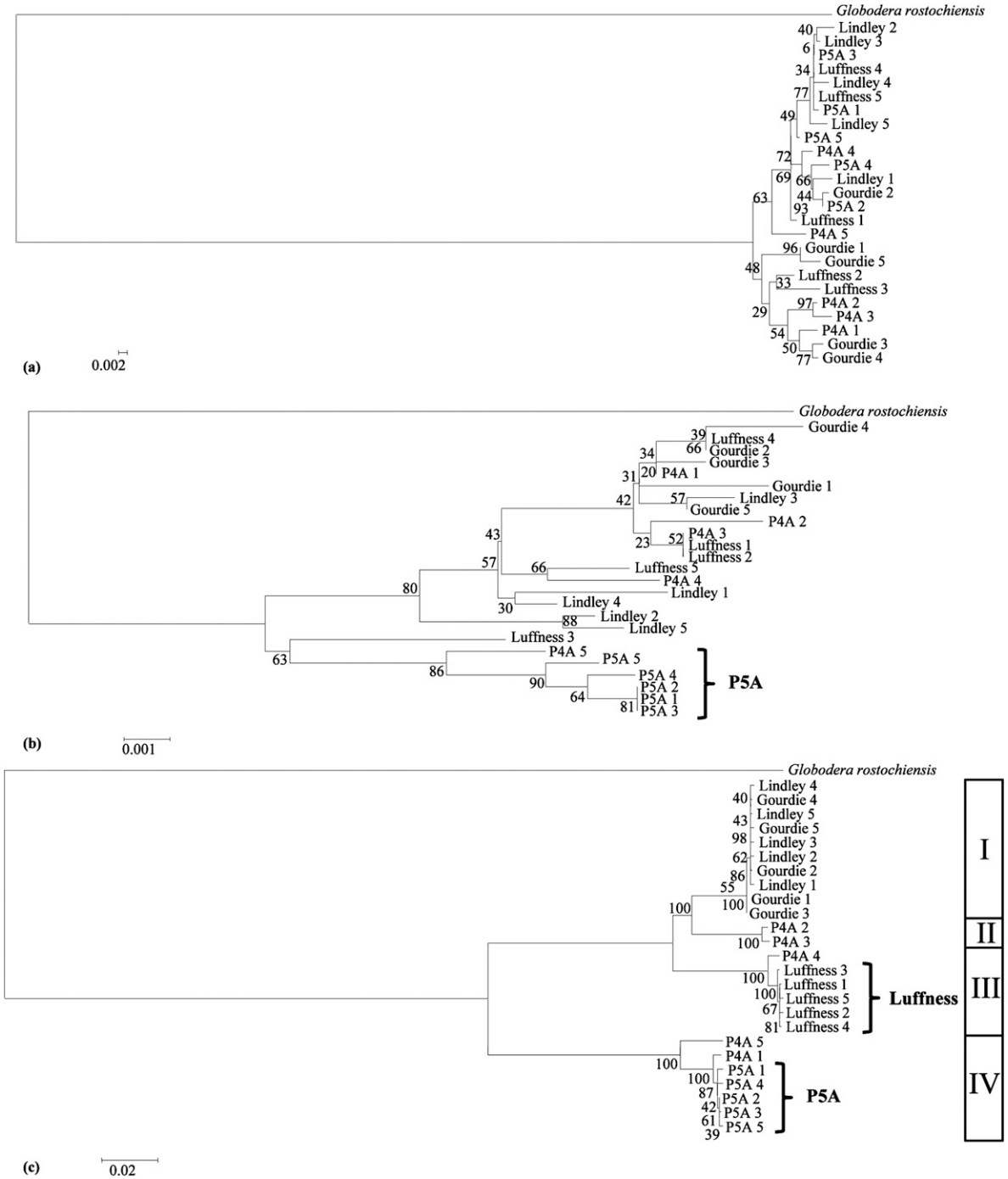

FIG. 2. Neighbor-Joining phylogenetic analysis of *Globodera pallida* populations. (a) Analysis of part of a protein-coding nuclear gene, *rbp-1*. (b) Analysis of the nuclear non-coding ITS region. (c) Analysis of a non-coding mtDNA region specific to scmtDNA IV. Numbers adjacent to nodes (generally positioned hierarchically relative to the outgroup, from left to right) indicate bootstrap proportions.

(Fig. 4a and 4b). These analyses were almost identical to that of the scmtDNA IV data set, resolving the same four major clades (I-IV). However, analysis of the combined data set resolved most representatives of the populations Gourdie and Lindley separately, although still not as monophyletic groups. The posterior probability and bootstrap support were generally highest for the combined data set; with all evolutionary relationships well supported.

DISCUSSION

Analyses of the ITS region and *rbp-1* gene region of *G. pallida* failed to resolve any populations as

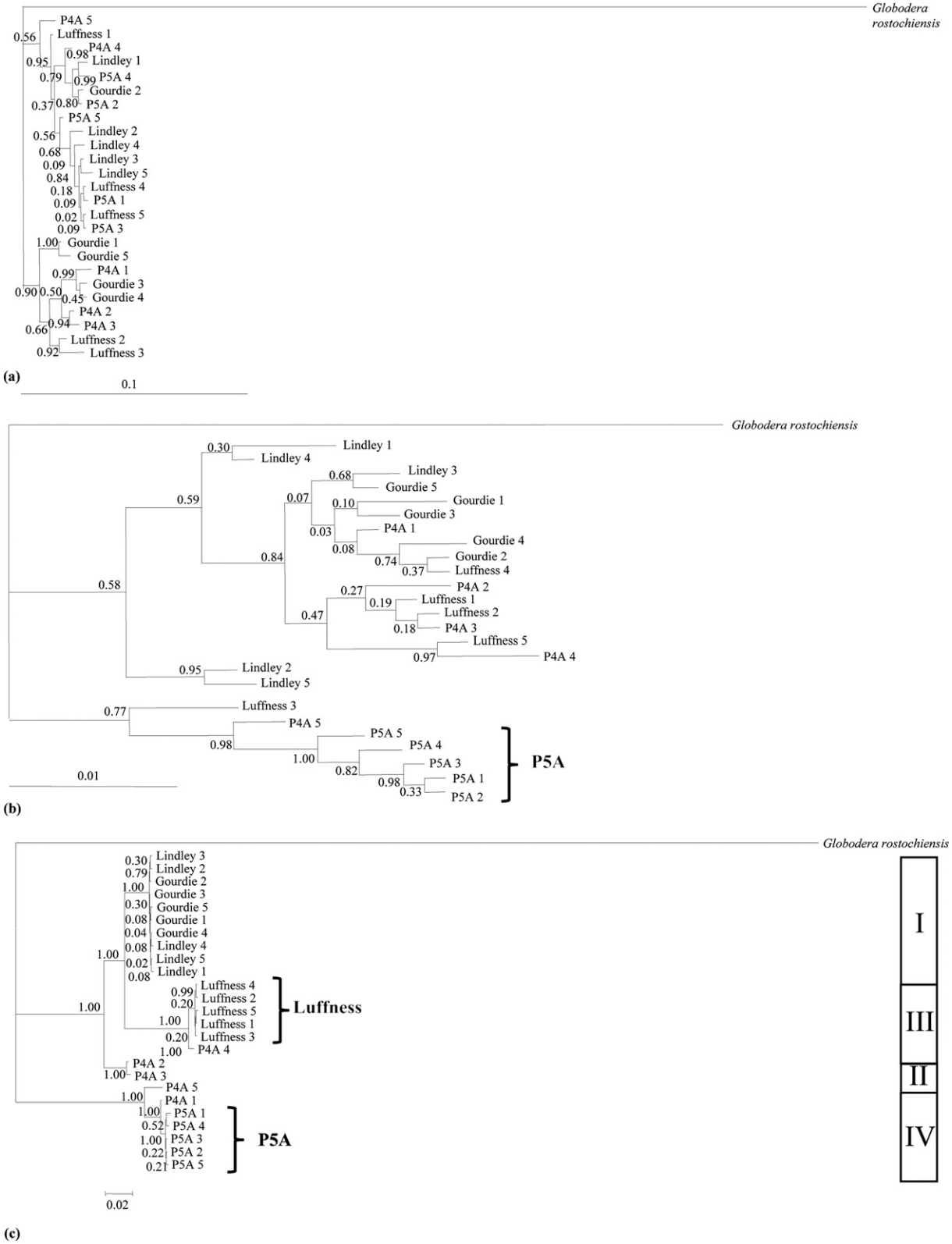

FIG. 3. Bayesian Inference analysis of *Globodera pallida* populations. (a) Analysis of part of a protein-coding nuclear gene region, *rbp-1*. (b) Analysis of the nuclear non-coding ITS region. (c) Analysis of a non-coding mtDNA region specific to scmtDNA IV. Numbers adjacent to nodes (generally positioned hierarchically relative to the outgroup, from left to right) indicate posterior probabilities.

monophyletic, except for analysis of the ITS region of population P5A. These results indicated that the intra-population sequence identity within these regions was high, providing limited data for the analysis of population histories. Thus, *G. pallida* analyses that include the ITS region should include additional molecular

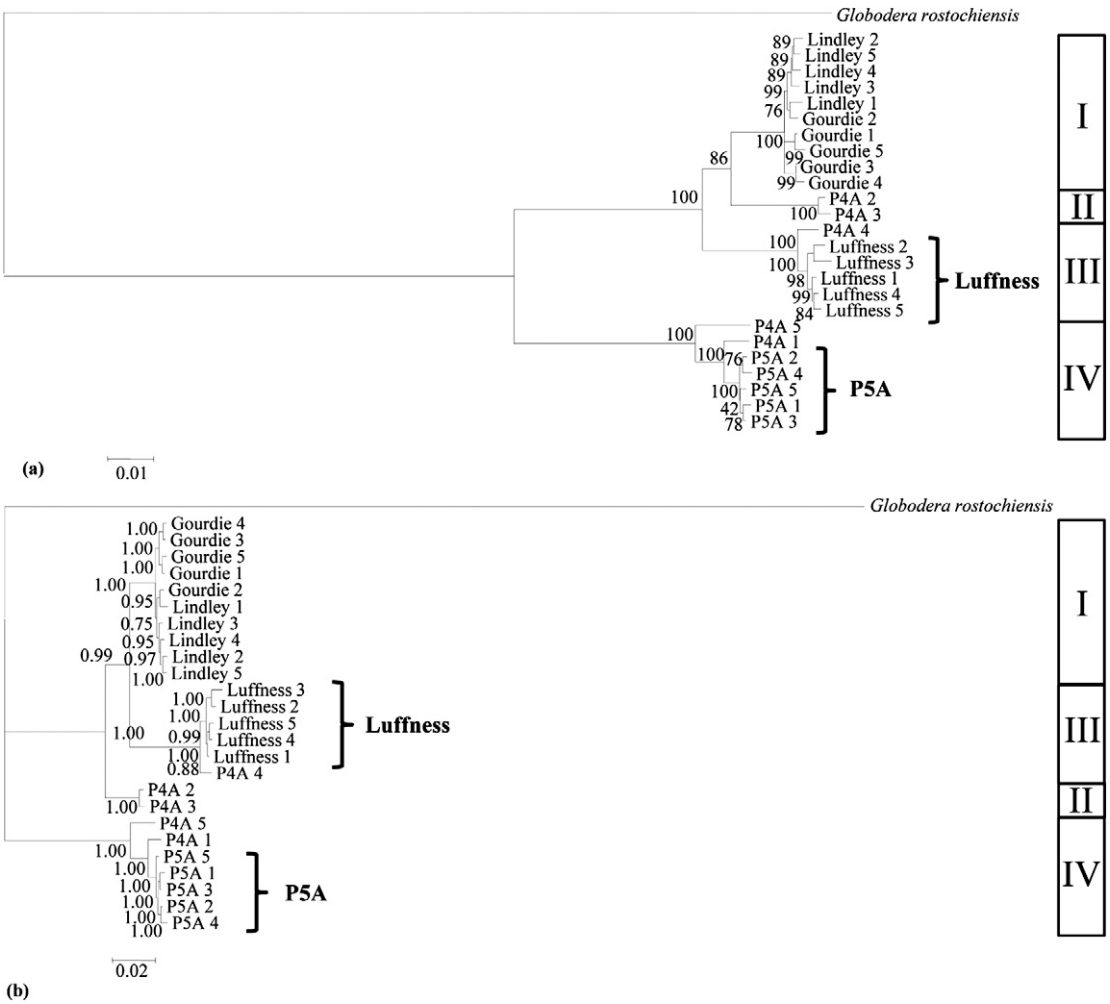

FIG. 4. Phylogenetic analysis of *Globodera pallida* populations using a combined data set of a protein-coding nuclear gene region (*rbp-1*), a non-coding nuclear region (the ITS region) and a non-coding mtDNA region. (a) Neighbor-Joining analysis. (b) Bayesian Inference analysis. Numbers adjacent to nodes (generally positioned hierarchically relative to the outgroup, from left to right) indicate bootstrap proportions in (a), or posterior probabilities in (b).

markers, as well as multiple representatives of each population to ensure that the genetic variability of that population is captured. Further, there was no clear *rbp-1* genotype associated with any population, despite this protein being a potential effector (Blanchard et al., 2005; Sacco et al., 2009), and different populations having different pathotypes (Phillips and Trudgill, 1998). Thus, the *rbp-1* gene region does not appear to be a suitable molecular marker for *G. pallida*. The rbp-1 protein also appeared to be unsuitable also as a population marker, with many mutations within and between populations.

Comparison of the phylogenetic analyses of the non-coding scmtDNA IV data set and of the combined data set of all three DNA regions indicated that both analyses resolved the populations P5A and Luffness as reciprocally monophyletic groups, and resolved the same four major clades. Analysis of these phylogenies indicated that the more ancestral evolutionary relationships were generally between representatives of the populations P5A, and between the population P5A and some representatives of P4A, supporting the hypothesis that *G. pallida* is derived from South America (Canto Saenz and De Scurrah, 1977; Evans and Stone, 1977; Picard et al., 2007). These analyses also suggest that several representatives of the population P4A are more genetically similar to Luffness than to the other UK and South American populations studied. Although the number of South American populations compared in this study was small, this is consistent with previous phylogenetic analyses of *G. pallida*, which also suggest that P4A and Luffness have a more ancestral relationship than that of P4A and other European populations (Pylypenko et al., 2008; Madani et al., 2010). The population P4A demonstrates considerable genetic variability, with representatives resolving within and between the UK and South American populations. We consider it unlikely that contamination from samples of other populations caused this variation. In another study (unpublished), we performed the same mtDNA sequence analysis on an individual P4A cyst, observing the same genetic

variants of the non-coding mtDNA region in these representative P4A sequences. Alternately, it has been suggested that the genetic variation observed in population P4A may be a result of it being of hybrid origin (Armstrong et al., 2007). Another confounding factor is the origin of 'P4A'. Although 'P4A' is used as a common grouping for phylogenetic analyses, 'P4A' more specifically refers to the pathotype, and the precise composition of that population in Peru is unknown.

The ability of the scmtDNA IV region to resolve more populations as reciprocally monophyletic than the other potential population markers suggests that this DNA region may be a useful molecular marker for *G. pallida*, either for phylogenetic or diagnostic applications, or for further understanding introductions of these and possibly other *G. pallida* populations. For characterising this DNA region, the approach described for the sequences used in the present study would be the most efficient. This involved incorporating the taxonomic sampling of populations into the cloning step by amplifying from DNA extracted from multiple individuals. The efficiency is also enhanced by the use of internal sequencing primers that anneal to multiple populations (see Table 4). However, a 3.4 kb region which requires cloning prior to sequencing may be impractical for some applications. Interestingly, most of the variation between populations is in a central ~1 kb region of this fragment. We suggest that a smaller region within this fragment may provide a more efficient molecular marker for *G. pallida*, but advise that a preliminary analysis is needed, as part of the non-coding mtDNA of *G. pallida* is homologous on different scmtDNA molecules (Gibson et al., 2007), and thus multiple copies of this internal region may be present. Further, the use of *Cytb* in conjunction with the non-coding scmtDNA IV region should be investigated. Phylogenetic analyses of *Cytb* in conjunction with other markers have previously demonstrated that they are able to resolve most South American and European populations of *G. pallida* independently (Picard et al., 2007; Plantard et al., 2008; Madani et al., 2010).

TABLE 4. Sequences for primers that anneal to the ~3.4 kb non-coding scmtDNA IV region from Globodera pallida populations P5A, P4A, Luffness, Gourdie and Lindley.

Primer	Sequence (5'-3')
scmt4-F1	GGAAAATTTGTACGAGAGGATCG
scmt4-F2	CGCTGCTCTGTACCTGGAG
scmt4-F3	ATTAGACCGATAAGTTTACACCTTG
scmt4-F4	ATGAGTTTTAAGAGTGTGTGAAAGG
scmt4-F5	TGATATAGGTTCAGATTGATGGAC
scmt4-F6	GCGATAGCGTACAAGATAAAATAGG
scmt4-F7	CACGAACCAGGAGAACAGAG
scmt4-R1	AGGCAGGAAACCTGGATCG
scmt4-R2	TCTTCTACGCTTCGCTACCAC
scmt4-R3	TGCCTACGTCAAATAGATAAGGA
scmt4-R4	TGCGTTACGAGCCAGTATAAG
scmt4-R5	CTTACAAGGTGTAAACTTATCGGTCT
scmt4-R6	CTCCAAAACTATTTGACTACGTCCT
scmt4-R8	GACTAGGTCCATCAATCTGAACC

LITERATURE CITED

Armstrong, M. R., Blok, V. C., and Phillips, M. S. 2000. A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. Genetics 154:181–192.

Armstrong, M. R., Husmeier, D., Phillips, M. S., and Blok, V. C. 2007. Segregation and recombination of a multipartite mitochondrial DNA in populations of the potato cyst nematode *Globodera pallida*. Journal of Molecular Evolution 64:689–701.

Bates, J. A., Taylor, E. J. A., Gans, P. T., and Thomas, J. E. 2002. Determination of relative proportions of *Globodera* species in mixed populations of potato cyst nematodes using PCR product melting peak analysis. Molecular Plant Pathology 3:153–161.

Blanchard, A., Esquibet, M., Fouville, D., and Grenier, E. 2005. *Ranbpm* homologue genes characterised in the cyst nematodes *Globodera pallida* and *Globodera 'mexicana'*. Physiological and Molecular Plant Pathology 67:15–22.

Blok, V. C., Malloch, G., Harrower, B., Phillips, M. S., and Vrain, T. C. 1998. Intraspecific variation in ribosomal DNA in populations of the potato cyst nematode *Globodera pallida*. Journal of Nematology 30:262–74.

Canto Saenz, M., and De Scurrah, M. M. 1977. Races of the potato cyst nematode in the Andean region and a new system of classification. Nematologica 23:340–349.

Dwivedi, B., and Gadagkar, S. 2009. Phylogenetic inference under varying proportions of indel-induced alignment gaps. BMC Evolutionary Biology 9:211.

Evans, K., and Stone, A. R. 1977. A review of the distribution and biology of the potato cyst-nematodes *Globodera rostochiensis* and *G. pallida*. Tropical Pest Management 23:178–189.

Geric Stare, B., Fouville, D., Širca, S., Gallot, A., Urek, G., and Grenier, E. 2011. Molecular variability and evolution of the pectate lyase (*pel-2*) parasitism gene in cyst nematodes parasitizing different *Solanaceous* plants. Journal of Molecular Evolution 72:169–181.

Gibson, T., Blok, V. C., Phillips, M. S., Hong, G., Kumarasinghe, D., Riley, I. T., and Dowton, M. 2007. The mitochondrial subgenomes of the nematode *Globodera pallida* are mosaics: Evidence of recombination in an animal mitochondrial genome. Journal of Molecular Evolution 64:463–71.

Golubchik, T., Wise, M. J., Easteal, S., and Jermiin, L. S. 2007. Mind the gaps: Evidence of bias in estimates of multiple sequence alignments. Molecular Biology and Evolution 24:2433–42.

Grenier, E., Bossis, M., Didier, F., Renault, L., and Didier, M. 2001. Molecular approaches to the taxonomic position of Peruvian potato cyst nematodes and gene pool similarities in indigenous and imported populations of *Globodera*. Heredity 86:277–290.

Hafez, S. L., and Sundararaj, P. 2007. First report of the pale cyst nematode, *Globodera pallida*, in the United States. Plant Disease 91:325–325.

Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95–98.

Haydock, P. P. J., and Evans, K. 1998. Management of potato cyst nematodes in the UK: An integrated approach? Outlook on Agriculture 27:253–260.

Hebert, P. D. N., Cywinska, A., Ball, S. L., and J. R., d. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series B-Biological Sciences 270:313–321.

Hijmans, R. J., and Spooner, D. M. 2001. Geographic distribution of wild potato species. American Journal of Botany 88:2101–2112.

Hodda, M., and Cook, D. C. 2009. Economic impact from unrestricted spread of potato cyst nematodes in Australia. Phytopathology 99:1387–1393.

Hoolahan, A. H., Blok, V. C., Gibson, T., and Dowton, M. 2011. Paternal leakage of mitochondrial DNA in experimental crosses of populations of the potato cyst nematode *Globodera pallida*. Genetica 139:1509–1519.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.

Kass, R. E., and Raftery, A. E. 1995. Bayes factors. Journal of American Statistics Association 90:773–795.

Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059–3066.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111–120.

Kort, J., Ross, H., Rumpenhorst, H. J., and Stone, A. R. 1977. An international scheme for identifying and classifying pathotypes of potato cyst nematodes *Globodera rostochiensis* and *G. pallida*. Nematologica 23:333–339.

Madani, M., Subbotin, S. A., Ward, L. J., Li, X., and DeBoer, S. H. 2010. Molecular characterization of Canadian populations of potato cyst nematodes, *Globodera rostochiensis* and *G. pallida* using ribosomal nuclear RNA and cytochrome *b* genes. Canadian Journal of Plant Pathology 32:252–263.

Marshall, D. C. 2010. Cryptic failure of partitioned Bayesian phylogenetic analyses; Lost in the land of the long trees. Systems Biology 59:108–117.

Nelson, L. A., Wallman, J. F., and Dowton, M. 2007. Using COI barcodes to identify forensically and medically important blowflies. Medical and Veterinary Entomology 21:44–52.

Nylander, J. A. A. 2004. MrModeltest v2. In: Program distributed by the author, Evolutionary Biology Centre, Uppsala University, p.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in Biosciences 12:357–358.

Phillips, M. S., and Trudgill, D. L. 1998. Variation of virulence, in terms of quantitative reproduction of *Globodera pallida* populations, from Europe and South America, in relation to resistance from *Solanum vernei* and *S. tuberosum ssp.* Andigena Cpc 2802. Nematologica 44:409–423.

Picard, D., and Plantard, O. 2006. What constitutes a population for the plant parasitic nematode *Globodera pallida* in its native area (Peru)? International Journal for Parasitology 36:115–122.

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

Picard, D., Sempere, 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. Molecular Phylogenetics and Evolution 42:308–316.

Plantard, O., Picard, D., Valette, S., Scurrah, M., Grenier, E., and Mugniéry, D. 2008. Origin and genetic diversity of Western European populations of the potato cyst nematode (*Globodera pallida*) inferred from mitochondrial sequences and microsatellite loci. Molecular Ecology 17:2208–2218.

Pylypenko, L. A., Phillips, M. S., and Blok, V. C. 2008. Characterisation of two Ukrainian populations of *Globodera pallida* in terms of their virulence and mtDNA, and the biological assessment of a new resistant cultivar Vales Everest. Nematology 10:585–590.

Pylypenko, L. A., Uehara, T., Phillips, M. S., Sigareva, D. D., and Blok, V. C. 2005. Identification of *Globodera rostochiensis* and *G. pallida* in the Ukraine by PCR. European Journal of Plant Pathology 111:39–46.

Qin, L., Overmars, H., Helder, J., Popeijus, H., van der Voort, J. R., Groenink, W., van Koert, P., Schots, A., Bakker, J., and Smant, G. 2000. An efficient cDNA-AFLP-based strategy for the identification of putative pathogenicity factors from the potato cyst nematode *Globodera rostochiensis*. Molecular Plant-Microbe Interactions 13:830–836.

Rehman, S., Postma, W., Tytgat, T., Prins, P., Qin, L., Overmars, H., Vossen, J., Spiridon, L., Petrescu, A., Goverse, 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. Molecular Plant-Microbe Interactions 22:330–340.

Ronquist, F., and Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Ryan, N. A., Duffy, E. M., Cassells, A. C., and Jones, P. W. 2000. The effect of mycorrhizal fungi on the hatch of potato cyst nematodes. Applied Soil Ecology 15:233–240.

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

Stone, A. R. 1985. Co-evolution of potato cyst nematodes and their hosts: Implications for pathotypes and resistance. EPPO Bulletin 15:131–137.

Subbotin, S. A., Halford, P. D., Warry, A., and Perry, R. N. 2000. Variations in ribosomal DNA sequences and phylogeny of *Globodera* parasitising solanaceous plants. Nematology 2:591–604.

Sunnucks, P., and Hales, D. F. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). Molecular Biology and Evolution 13:510–524.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596–1599.