Plant Pathology (2015) **64**, 729–737 Doi: 10.1111/ppa.12300

The origin and genetic diversity of the causal agent of Asian soybean rust, *Phakopsora pachyrhizi*, in South America

V. R. Jorge[a], M. R. Silva[a], E. A. Guillin[b], M. C. M. Freire[a], I. Schuster[c], A. M. R. Almeida[d] and L. O. Oliveira[a]*

[a]*Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, 36570-000 Viçosa, MG, Brazil;* [b]*Instituto de Genética Ewald A. Favret, Centro Nacional de Investigaciones Agropecuarias, Instituto Nacional de Tecnologia Agropecuaria, INTA Castelar, Buenos Aires, Argentina;* [c]*Cooperativa Central de Pesquisa Agrícola, BR 467 – km 98, 85813-450 Cascavel; and* [d]*Empresa Brasileira de Pesquisa Agropecuária, Embrapa-Soja, 86001-970 Londrina, PR, Brazil*

A sequence-based approach was used to investigate molecular genetic variations in *Phakopsora pachyrhizi*, an obligate biotrophic pathogen that causes Asian soybean rust. In Argentina, the samples came from uredinium-bearing leaves taken from 11 soybean fields; in Brazil, the samples comprised urediniospores from leaves of 10 soybean genotypes that had been grown in three experimental stations during two growing seasons. PCR-based cloning techniques were used to generate DNA sequences for two gene regions and alignments were supplemented with data from GenBank. A total of 575 sequences for the internal transcribed spacer region (18 ribotypes) and 160 partial sequences for a housekeeping gene encoding ADP-ribosylation factor (10 haplotypes) were obtained. Ribotype accumulation curves predicted that about 20 bacterial clones would recover 5–6 ribotypes (*c.* 70–80% of the total molecular variation) per locality. The samples from the three experimental stations in Brazil displayed most (14 out of 16) ribotypes found worldwide; the lack of genetic structure and differentiation at a diverse geographic scale suggests that both local and distant sources provide airborne inoculum during disease establishment. Soybean genotypes with resistance genes for the Asian soybean rust did not decrease the molecular genetic variation of fungal populations.

Keywords: Asian soybean rust, bacterial cloning, genetic diversity, phylogeography, *Rpp* gene

Introduction

Phakopsora pachyrhizi, the causal agent of Asian soybean rust (ASR), is an obligate biotrophic pathogen; for growth and reproduction, *P. pachyrhizi* depends completely on living hosts (Bromfield, 1984). Globally, ASR poses a serious threat to soybean production, with the potential to cause yield losses from 10 to 80% (Bromfield, 1984; Pandey *et al.*, 2011). This rust is of Asian origin and was historically confined to areas of Asia and Australia (Ono *et al.*, 1992). However, in recent years, ASR has spread beyond its traditional range. Intercontinental dispersal, most probably through wind-based mechanisms (Isard *et al.*, 2005), allowed *P. pachyrhizi* to reach Paraguay (2001), Brazil and Argentina (2002), Bolivia (2003), and the USA (2004) (Schneider *et al.*, 2005). *Phakopsora pachyrhizi* produces only two types of spores: urediniospores (the repeating, asexual spores) and teliospores (spores that do not infect plants, but are the survival stage of the life cycle); germination of teliospores, which is crucial for entering the sexual cycle, has never been observed in the wild (Bromfield, 1984). After successful infection of the host plant, the pathogen produces a massive number of wind-dispersed urediniospores, and multiple infection cycles can take place within a single growing season. Asexual production of urediniospores is suggestive of low genetic diversity in the field owing to the presence of a reduced number of clonal lineages and recurrent founder events (Twizeyimana *et al.*, 2011). In soybeans, seven loci (also referred to as the *Rpp* genes) are known to confer resistance to certain isolates of *P. pachyrhizi* (Pandey *et al.*, 2011). The best-characterized *Rpp* genes are *Rpp1*, *Rpp2*, *Rpp3*, and *Rpp4*. The effect of host resistance on molecular diversity of *P. pachyrhizi* is yet to be investigated.

Previous studies began to shed light into the population structure of *P. pachyrhizi*. In general, molecular data from distinct markers – microsatellites (Twizeyimana *et al.*, 2011) and DNA sequencing analyses (Freire *et al.*, 2008, 2012; Zhang *et al.*, 2012) – were congruent in showing that genetic diversity in *P. pachyrhizi* displayed little structure over large geographic areas; most of the variation resided within locations. These findings suggested the existence of efficient mechanisms for long-distance dispersal and large gene flow of *P. pachyrhizi* amongst soybean fields.

It is not yet possible to obtain *in vitro* cultures of *P. pachyrhizi*; thus, field-based studies should use environmental samples as sources of genomic DNA. Hereafter, a sample of *P. pachyrhizi* taken directly from host

*E-mail: luiz.ufv@hotmail.com

Published online 5 October 2014

Plant Pathology

tissues is referred to as an 'environmental sample'; an environmental sample may come from one to several soybean leaves and consists of a blend of urediniospores and host parts bearing fungal tissues. In some circumstances, an environmental sample yields DNA extract that contains genomic DNA from a single fungal genotype; single-urediniospore isolates, for example, were used to investigate molecular variation in rusts through microsatellite marker analyses (see Ordoñez & Kolmer, 2007; Kubisiak *et al.*, 2011; Twizeyimana *et al.*, 2011). This strategy imposed requirements for specialized laboratory resources and greenhouse facilities to handle the collection free of cross-contamination. In other circumstances, environmental samples were taken directly from field-collected leaves bearing uredinia and contained a pool of individuals of *P. pachyrhizi* with an unknown level of genetic relatedness (see Freire *et al.*, 2008, 2012; Zhang *et al.*, 2012). When used as template for PCR, the pool of mixed DNA genotypes gave rise to DNA amplicons of distinct sizes. Under these circumstances, direct sequencing of PCR products yielded electropherograms bearing overlapping peaks and unreadable results. The strategy required bacterial cloning to obtain high-quality electropherograms free of double peaks; each clone recovered a unique sequence from the pool of distinct PCR amplicons (Freire *et al.*, 2008, 2012). Thus far, genetic diversity studies in *P. pachyrhizi* have been carried out using small sampling sizes per location. These small sampling sizes were directly related to the difficulties associated with working with obligate biotrophs outside of their living hosts. Exploiting large-scale sampling to uncover molecular genetic variation in environmental samples of an obligate biotrophic pathogen, such as *P. pachyrhizi*, is labour-intensive. The assembly of large molecular data sets could contribute to better understanding of the fundamental aspects of genetic diversity of *P. pachyrhizi*. Increasing the sampling size per location, for example, allows for the capture of rare genotypes in addition to the most prevalent genotypes. Another added value of sampling intensively within locations is the possibility of investigating whether the presence of a resistance gene in the host is associated with levels of molecular variation of the rust at the local scale.

For this study, data sets using DNA sequences from two gene regions of the nuclear genome were assembled: the entire internal transcribed spacer (ITS) region (ITS1–5·8S gene–ITS2) and a partial sequence of a housekeeping gene encoding ADP-ribosylation factor (ARF). Although such sequence-based approaches are commonly used to explore systematic questions, previous studies showed that these two gene regions were polymorphic within soybean fields (Freire *et al.*, 2008, 2012; Zhang *et al.*, 2012); they could be, therefore, suitable to unravel phylogeographic questions. In the present investigation, at the local level, the sequences came from three soybean fields from Brazil; at the continental level, a collection of sequences from distinct countries was available. A strategy developed previously (Freire *et al.*, 2012) was used to identify and remove any sequence variation artificially generated from the molecular methods. A phylogeographic approach compared patterns of molecular genetic variation at the local level with the continental level and made inferences about the influence of a host resistance gene on the molecular variation of the rust. Specifically, the following three questions were addressed: (i) how many bacterial clones need to be sampled to capture most of the molecular variation in environmental samples of *P. pachyrhizi* at the local level? (ii) Given that efficient mechanisms for long-distance dispersal *of P. pachyrhizi* exist, is molecular variation at the local level as diverse as molecular variation at the regional level? (iii) To what extent would the presence of an *Rpp* gene in a given soybean genotype contribute to a decrease in the molecular variation of environmental samples of *P. pachyrhizi* taken from that genotype?

Materials and methods

Sampling strategy, DNA extraction and PCR amplification

The environmental samples analysed were from Argentina and Brazil (Table S1). Tissue samples collected in Argentina were taken from leaves bearing uredinia that had been obtained between 2003 and 2007 from 11 fields spread throughout the main areas of soybean cultivation. In Brazil, experiments were undertaken at three experimental stations (Cascavel, Londrina, and Viçosa) during the growing seasons of 2009/2010 (hereafter referred to as season 2010) and 2011/2012 (season 2012). Cascavel and Londrina are located within a core area of soybean production, whilst Viçosa is an outlier located at least 250 km from the nearest area where soybeans are intensively farmed (Fig. S1).

The experiments involved 10 soybean genotypes; four of these genotypes each have a distinct gene that confers resistance to *P. pachyrhizi*: PI 200492 (*Rpp1*), PI 230970 (*Rpp2*), PI 462312 (*Rpp3*), and PI 459025 (*Rpp4*). The remaining six soybean genotypes (BRS 134, FT 17, FT 2, IAC 1, BRS MS Bacuri, and CD 208) have no genes known to confer resistance to *P. pachyrhizi*. The 10 genotypes were grown in plots with two replicates. In each plot, seeds of a given genotype were planted in rows 4 m long and spaced 0·5 m apart. The plants were grown using standard agronomic procedures, but fungicide was not applied. When symptoms of ASR were clearly visible, infected leaves were harvested and stored in paper bags until processing.

Total genomic DNA was extracted from urediniospores or soybean leaves bearing uredinia using a previously described protocol (Freire *et al.*, 2008). The two regions of the nuclear genome (ITS and ARF) were amplified by PCR. Amplification of the ITS region was carried out using the *P. pachyrhizi*-specific primers ITSPP5 and ITSPP3 in accordance with a previously described PCR protocol (Freire *et al.*, 2008). Amplification of a 700 bp region within the *ARF* gene used primers DN739857F2 (5′-AGCAGACCCAACCCATCGATAT-3′) and DN73985R5b (5′-AGTCCACCACAAAGATGATTCCCT-3′), following the protocol of Zhang *et al.* (2012). This stretch of genomic DNA is subsequently referred to as the *ARF* region. For sequences of Brazilian origin, Phusion High-Fidelity DNA polymerase (New

England Biolabs) was used during PCR amplifications; the only exception was the amplification of ITS from Londrina (season 2010), for which Platinum *Taq* DNA polymerase (Invitrogen Life Science Technologies) was used. For sequences of samples collected in Argentina, GoTaq DNA polymerase (Promega) was used. Prior to bacterial cloning, amplification products were cleaned using the Wizard SV and PCR Clean-UP System (Promega).

Bacterial cloning, sequencing and sequencing alignments

Cloning was carried out using the pGEM-T Easy kit (Promega), and plasmid DNA was extracted using a Wizard Minipreps DNA Purification System (Promega). Sequencing was performed by Macrogen Inc., South Korea (www.macrogen.com) with Sanger sequencing technology and the plasmid primer M13 pUCF (5′-GTTTTCCCAGTCACGAC-3′). About 8–10 clones of ITS were submitted for sequencing for each of the 10 soybean genotypes grown in experimental stations, for each location and growing season; four to six clones of ITS were submitted for sequencing for each of the 11 samples from Argentina. Usable data was received from 486 clones (Brazil, $n = 437$; Argentina, $n = 49$). For the *ARF* region of *P. pachyrhizi*, clones from Cascavel and Viçosa for the year 2012 only were submitted; usable data from 92 clones was received. Sequences were imported into the program SEQUENCHER v. 4.8 (Gene Codes) for alignment and editing. Complete sequence alignments were performed with the introduction of gaps to compensate for the presence of insertions or deletions (indels).

Data set assemblies and structural work

Additional sequences were obtained from GenBank. For the ITS region, there were 12 sequences from Frederick *et al.* (2002), 21 sequences from Barnes *et al.* (2009), and 56 sequences from Zhang *et al.* (2012). Partial sequences for ITS from the authors' previous study (Freire *et al.*, 2008) were not used. For the *ARF* region, there were 68 sequences from Zhang *et al.* (2012). Although most of the sequences obtained from GenBank have a North American origin, there were also sequences from Asia, Africa, and South America. For network construction, the GenBank accessions from the newly established ranges in the Americas and Africa are collectively referred to as 'Americas+Africa', and those accessions from the traditional ranges in Asia and Australia as 'Asia+Australia'. For the remaining analyses, the GenBank sequences are referred to as 'GenBank'. Hereafter, each of these sets is treated as a distinct 'location' in subsequent statistical analyses.

After all of the sequences were aligned, their ends were trimmed to eliminate fragments that could not be obtained for all sequences. Finally, the sequences were assembled into two preliminary data sets. Subsequently, each alignment was inspected for the presence of autapomorphies (the presence of a derived character in a single sequence of the data set). Misincorporation errors that take place during bacterial cloning of PCR fragments appear as autapomorphies; their position along the ITS region displayed no association with the occurrence of conserved secondary structures or conserved sequence motifs, in which mutations were supposed to be either absent or of very low incidence (see Queiroz *et al.*, 2011; Freire *et al.*, 2012). Sites that contained autapomorphies were excluded from subsequent analyses.

For the subset of ITS sequences from which autapomorphies were excluded, the site where the autapomorphy had occurred was recorded using HYPERMUT v. 2.0 software (Rose & Korber, 2000). A web-based interface for HYPERMUT is available (http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html). After the exclusion of autapomorphies, the final data set for the ITS region (data set A) comprised 575 sequences of 565 bp each (437 from Brazil, 49 from Argentina, and 89 from GenBank). The final data set for the *ARF* region (data set B) contained 160 sequences of 684 bp each (92 from Brazil and 68 from GenBank). Sequences obtained for this study were deposited in GenBank with the following accession numbers: ITS region, KJ852028–KJ852513; *ARF* region, KJ852514–KJ852605.

The limits of the ITS1, 5·8S gene, and ITS2 sequences in *P. pachyrhizi* were defined and depict the location of polymorphic sites along the sequences in accordance with previously published strategies (Garcia *et al.*, 2011; Freire *et al.*, 2012). To define the gene structure in the *ARF* region, several BLASTX searches (Altschul *et al.*, 1990) were run using translated nucleotide sequences of the *ARF* gene as queries against public databases at the National Centre for Biotechnological Information. The searches retrieved the complete mRNA sequence that encodes the ADP-ribosylation factor of *Puccinia graminis* (GenBank accession XM_003307028.2). The deduced protein of the *ARF* gene of *P. graminis* displays complete sequence identity with that of the *ARF* gene of *P. pachyrhizi*; the molecular characterization of the *ARF* gene of *P. graminis* was used to delimit exon and intron regions in the homologous *ARF* gene of *P. pachyrhizi*.

Statistical analysis

Gene genealogies for each of the two loci (data set A, for ITS; data set B, for *ARF*) were inferred independently using the median-joining (MJ) network method (Bandelt *et al.*, 1999) as implemented in NETWORK v. 4.5.0.2 software (Fluxus Technology Ltd). This software was run such that indels were considered to be a fifth character state and coded such that each indel, regardless of its size, was considered to be a single mutation. Indels that were bordered by mononucleotide repeats (e.g. polyT) were rejected as a source of information for these statistical analyses. Haplotype accumulation and haplotype estimation curves were calculated with ESTIMATES v. 8.2 software (Colwell *et al.*, 2012). These analyses allowed the estimation of haplotype diversity and the quantification of the effects of sampling effort on haplotype diversity. They were also useful to predict the eventual asymptote in haplotype diversity for a particular location during a given growing season. An online tool (Preacher, 2001) was used to perform chi-square tests to compare the observed frequencies of haplotypes in resistant genotypes with those in susceptible genotypes, using the null hypothesis of no association between haplotype frequency and the presence of genes for ASR resistance. Population differentiation was estimated by analysis of molecular variance (AMOVA; Excoffier *et al.*, 1992) using ARLEQUIN v. 3.0 software (Excoffier *et al.*, 2005) and the data set A. AMOVA considered the different levels of genetic structure: among locations (Cascavel, Londrina, Viçosa, Argentina, and 'GenBank'), between seasons (2010 and 2012, for data sampled in Brazil only), and between groups of genotypes (resistant and susceptible, for data sampled in Brazil only). The significance of the genetic differentiation was tested with 1000 permutations, where P denotes the probability of having a more extreme variance component than the values observed by chance alone.

Results

Treatment of autapomorphies and polymorphic sites

The pattern of distribution of autapomorphies along the spacer regions (ITS1, $n = 34$; ITS2, $n = 25$) and along the otherwise conserved 5·8S gene ($n = 29$) (Fig. S2) provided strong support for these 88 mutations not being naturally occurring polymorphisms (see Queiroz *et al.*, 2011; Freire *et al.*, 2012) and justifies their removal from the final ITS data set. There were five autapomorphies in clones of the *ARF* region; these were removed from subsequent analyses. The use of a high-fidelity enzyme during PCR amplification (Phusion DNA polymerase) yielded clones (of both the ITS and *ARF* regions) that displayed very low rates of autapomorphy. The use of DNA polymerases of lower fidelity (Platinum *Taq* DNA polymerase for Londrina season 2010, and GoTaq DNA polymerase for Argentina) resulted in clones that harboured high rates of autapomorphy (data not shown). Among the three DNA polymerases that were used for PCR amplification, Phusion DNA polymerase possessed the highest fidelity; its error rate was about 50-fold lower than that of *Taq* DNA polymerase (Frey & Suppmann, 1995).

The ITS1 region displayed six polymorphisms. Among them, there was one polymorphic site that presented size variation within a mononucleotide repeat (polyT) and one polymorphic site that exhibited size variation within a dinucleotide repeat (polyAT). Although these two polymorphisms were not autapomorphies, they were rejected as a source of information during statistical analyses. This rejection was intended to reduce the effects of experimental error, which could arise from misincorporation errors caused by misfunction of DNA polymerase during either PCR or bacterial cloning (Queiroz *et al.*, 2011) or evolutionary lability associated with these types of indel (Mast *et al.*, 2001). The four usable polymorphic sites in the ITS1 region (Fig. S3a) were three substitutions (at positions 52, 120 and 185) and one 1-bp indel (at position 105). The 5·8S gene was highly conserved; it contained a single substitution (at position 227). Similarly to the ITS1 region, the ITS2 region harboured six polymorphisms, two of which exhibited size variation within mononucleotide repeats (polyT); these two mononucleotide repeats were excluded from further analyses. The four usable polymorphic sites of the ITS2 region (Fig. S3a) were one substitution (at position 382) and three indels (at positions 378–380, 389 and 550). Overall, the eight polymorphic sites of the ITS region were located within single-stranded structures, such as internal loops and bulges. The *ARF* region exhibited seven polymorphic sites: three indels (at positions 242–247, 309–312 and 317) and four substitutions (at positions 44, 403, 418 and 615) (Fig. S3b). Polymorphisms within the coding sequence were restricted to two synonymous substitutions at sites 403 (A↔T) and 418 (A↔G).

A given haplotype of the ITS region is referred to as a 'ribotype', given that it was observed in a nuclear ribosomal DNA sequence; the use of the term 'haplotype' will be used only to refer exclusively to haplotypes of the *ARF* region. The sequence-based approach was effective. After excluding false polymorphisms, the levels of sequence variation found (18 ribotypes for ITS region and 10 haplotypes for the *ARF* gene) were unusually high; these levels of molecular variation allowed the analysis to be carried out.

Effort to capture molecular variation

Most of the accumulation curves depicted a similar pattern, with a horizontal asymptote in diversity towards eight ribotypes. Londrina (season 2010) was the only exception; no plateau was reached after 65 bacterial clones. The most likely explanation for the higher than expected ribotype diversity in Londrina (season 2010) was false positive results for polymorphisms owing to the use (inadvertently) of a low-fidelity enzyme (Platinum *Taq* DNA polymerase) for PCR amplification prior to bacterial cloning. It seems that the strategies used could not eliminate these inaccurate results, causing them to persist in the data sets. Visual inspection of the accumulation curves (Fig. 1) suggests that the expected, maximum number of ribotypes in each of the three experimental stations was about eight; about 20 bacterial clones would recover 5–6 ribotypes, which is about 70–80% of the total genetic diversity of a given locality.

Molecular variation at local level and regional level

Median-joining networks were constructed for ITS and *ARF* data sets independently; these networks provided insights about how molecular variation at the local level was associated with molecular variation at the regional level. For reference purposes, the networks were colour-coded according to geographic location and growing season.

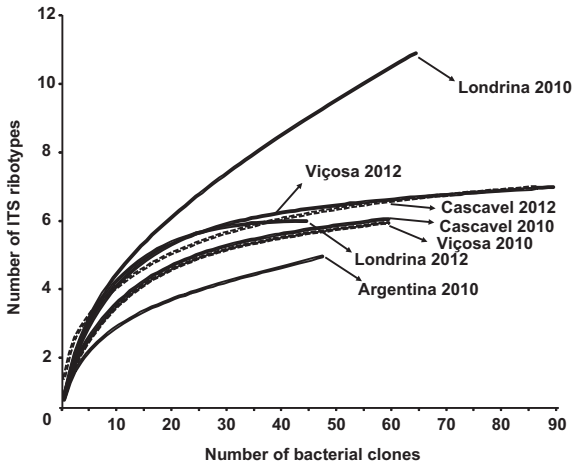

Figure 1 Rarefaction curves for ITS of *Phakopsora pachyrhizi* recovered through bacterial cloning. Locations and years are as indicated.

At the local level, evidence for the lack of spatial structure was the fact that the three experimental stations in Brazil shared the six most common ribotypes of ITS (Fig. 2a); these locations also shared the two most common haplotypes of *ARF* (Fig. 3). These results suggest that gene flow among the three experimental stations was high.

Further evidence of the lack of spatial structure came when considering a broader geographic scale. The three experimental stations in Brazil displayed 14 out of the 16 ribotypes found worldwide; these ribotypes are present in GenBank accessions – 'Americas+Africa' and 'Asia+Australia' (Fig. 2b); the most common ribotype in Brazil (ribotype 1) was also the most common in Argentina, 'Americas+Africa', and 'Asia+Australia' (Fig. 2b). Ribotype 16 (Fig. 2b), for example, was found over a wide geographic area; it was uncovered from samples collected in Brazil, Vietnam, and the USA (Barnes *et al.*, 2009; Zhang *et al.*, 2012).

The temporal genetic structure appeared to be very weak, if present at all. At a short temporal scale (growing seasons 2010 and 2012), it was found that the networks obtained for sequences from the three experimental stations displayed almost identical topologies (Fig. 2c). At a wider temporal scale (1972 to the present time), there were ribotypes (such as ribotype 4, for example) that were recovered from samples obtained from 1972 to 1980 in Asia and Australia (Frederick *et al.*, 2002) and from samples obtained from Argentina (2003) and Brazil (seasons 2010 and 2012). Results of AMOVA corroborated the lack of temporal and spatial genetic structure in *P. pachyrhizi*. It revealed that only 6·3% of the total molecular variance can be attributed to differences between locations, whereas 93·7% is apportioned within locations (Table 1).

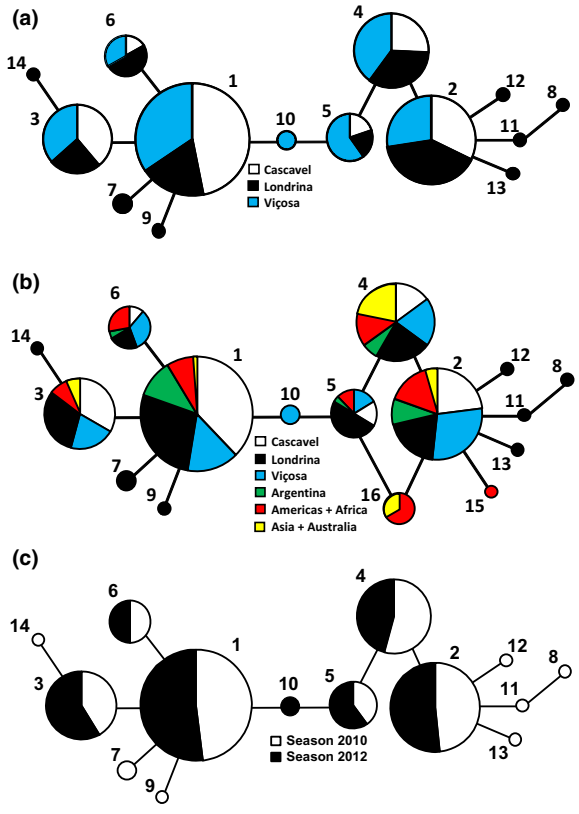

Figure 2 Median-joining networks for the ITS ribotypes of *Phakopsora pachyrhizi*. (a) Network uncovered for ribotypes found in the three experimental stations in Brazil. (b) Network for complete ITS data set. (c) Network for the ITS region according to season. In each network, a circle represents a ribotype (coded with numbers). Circle size is proportional to the relative frequency. Each line represents a mutation step. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

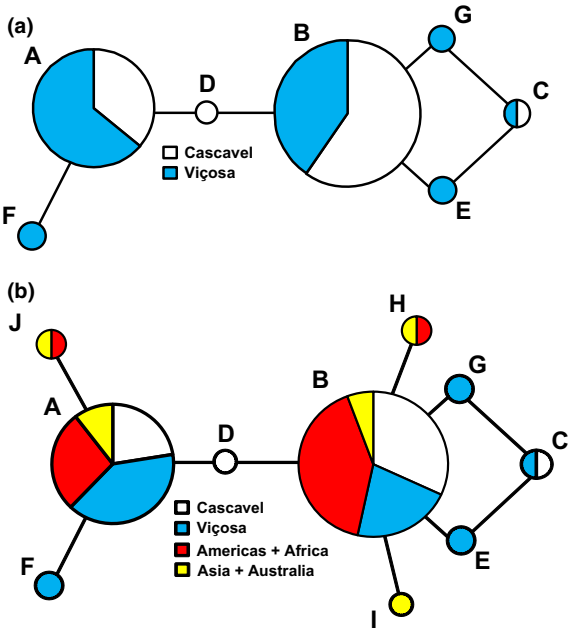

Figure 3 Median-joining networks for the haplotypes of the ADP-ribosylation factor gene of *Phakopsora pachyrhizi*. (a) Network uncovered for haplotypes found in two experimental stations in Brazil. (b) Complete network. In each network, a circle represents a haplotype (coded with letters). Circle size is proportional to the relative frequency. Each line represents a mutation step. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

The *ARF* region displayed a signature that also indicated a lack of genetic structure; the two most common haplotypes in Brazil (haplotypes A and B) were also the most prevalent worldwide (Fig. 3). For both gene regions, only the rare ribotypes (for the ITS region) and haplotypes (for the *ARF* region) were characteristic of a particular geographic location.

Rpp genes do not lead to a decrease in molecular variation

Median-joining networks also provided insights about the lack of a clear relationship between the presence of resistance genes and variation in levels of molecular

Table 1 Analysis of molecular variance (AMOVA) for the ITS region of *Phakopsora pachyrhizi*

Source of variation	d.f.	Sum of squares	Variance components	% Total variance	*P* value[a]
Among locations[b]	3	11·49	0·053	6·31	<0·001
Within locations	482	454·47	0·797	93·69	
Total	485	481·45	0·850		
Between seasons[c]	1	0·16	−0·002	−0·36	0·856
Within seasons	435	343·07	0·788	100·36	
Total	436	343·24	0·785		
Between groups[d]	1	0·22	−0·0026	−0·33	0·755
Within groups	435	343·24	0·79	100·33	
Total	436	343·01	0·78		

[a]Significance test is based on 1023 permutations.
[b]Locations: Cascavel, Londrina and Viçosa, Brazil; Argentina.
[c]Growing seasons: 2010 and 2012.
[d]Groups of genotypes: group harbouring no known resistance genes and group harbouring resistance genes.

variation of the ITS and *ARF* regions (Fig. 4). For the ITS data set, there were no significant differences between groups of genotypes (resistant versus susceptible) for both the occurrence of ribotypes and the occurrence of common ribotypes (Table S2). Analysis of the *ARF* data set provided results (Table S3) similar to those obtained for the ITS data set. When AMOVA considered groups of genotypes according to the presence of resistance genes, 100% of the total variance was attributed to differences within seasons or within groups of genotypes, respectively (Table 1).

Discussion

Experimental samples of *P. pachyrhizi* usually consist of urediniospores or fungal tissue harvested directly from the host. Upon extraction, the samples yield genomic DNA in which both pathogen and host genomes are present. In addition, the fungal genomic DNA mostly probably corresponds to a fungal population of individuals whose genetic relatedness is unknown. Under these circumstances, the results of direct sequencing of products obtained by PCR are frequently uninterpretable (Freire *et al.*, 2008, 2012). Fortunately, cloning of discrete amplicons from the pool of PCR products generates high-quality sequence reads. This strategy requires bacterial clones at a number that captures the molecular diversity within the fungal population under investigation.

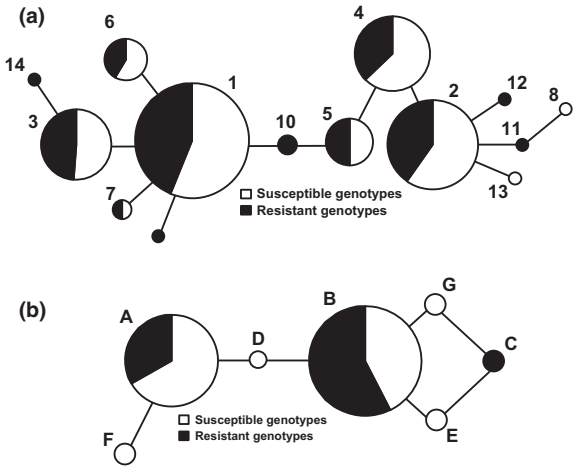

Figure 4 Median-joining networks uncovered for *Phakopsora pachyrhizi* in the three experimental stations in Brazil according to the presence of an *Rpp* gene in the host genotype. (a) Network for the ITS region. (b) Network for the ADP-ribosylation factor gene. In each network, a circle represents a ribotype (coded with numbers) or haplotype (coded with letters). Circle size is proportional to the relative frequency. Each line represents a mutation step.

Given that diversity often increases with sampling effort, it is difficult to decide the number of clones necessary to capture a given fraction of the molecular variation present in environmental samples of *P. pachyrhizi*. In the present study, ribotype accumulation and ribotype estimation curves (Colwell *et al.*, 2012) indicated that about 20 bacterial clones would uncover about 80% of the total genetic diversity, with the likely number of ribotypes found within a given locality being eight. Most probably owing to widespread dispersal of inoculum, the maximum number of ribotypes expected (eight ribotypes, in this study) remained relatively unchanged at the distinct geographic scales under consideration. That is, horizontal asymptotes in haplotype diversity at both local scales (the three experimental stations in Brazil) and large scales (Argentina) are suggestive of saturation in the recovery of molecular variation through bacterial cloning. It is probable that additional rounds of cloning would recover only rare sequences; when added to the network analyses, such rare sequences would appear at tip positions and would not cause significant changes in network topologies.

Next, the sequence-based approach was used to investigate the phylogeography and population structure of *P. pachyrhizi* at the local and regional levels. Although microsatellite markers are the ideal for high-resolution population genetic studies, the implementation of this approach to investigate an obligate biotrophic pathogen, such as *P. pachyrhizi*, is not straightforward. Obtaining

single-urediniospore isolates and maintaining these isolates in living hosts free of cross-contamination are activities that require specialized laboratory resources and greenhouse facilities; these resources are not widely available.

A high level of sequence variation in the ITS region has been documented in other fungi (Blaalid *et al.*, 2013), including *Basidiomycota* (Nilsson *et al.*, 2008). Overall, there are three major potential causes of unusually high levels of variation of *P. pachyrhizi*: intragenomic variation of ITS, the dikaryotic nature of the urediniospores, and among-individual variation. Incomplete concerted evolution is one of the most significant mechanisms that trigger intragenomic variation of ITS; the lack of complete homogenization allows for the coexistence tandem array of the nuclear ribosomal DNA harbouring distinct ribotypes (Elder & Turner, 1995). The dikaryotic nature of the urediniospores of *P. pachyrhizi* promote the maintenance of sequence variation of both ITS and *ARF* gene regions; each nucleus may preserve distinct repeat motifs within a single urediniospore. Germ tube and hyphal anastomosis and the possibility of a parasexual cycle (Vittal *et al.*, 2012) may increase further the chances for high genetic diversity in *P. pachyrhizi*. The high level of sequence variation found for these two gene regions in the present investigation suggests that mutation rates in *P. pachyrhizi* are high; therefore, breeders should not overlook the likelihood of genotype conversion from avirulence to virulence in *P. pachyrhizi*.

The results of this study confirmed the *a priori* expectation of weak genetic structuring for *P. pachyrhizi*. Levels of genetic variation at the local scale (the three experimental stations) exhibited no decrease when compared with the levels found at larger geographic scales (Argentina, 'Americas+Africa', and 'Asia+Australia'). This study was congruent with previous investigations that showed that *P. pachyrhizi* lacked genetic structure at the regional level (Freire *et al.*, 2008, 2012; Twizeyimana *et al.*, 2011; Zhang *et al.*, 2012). The present investigation showed that the molecular variation of *P. pachyrhizi* at a local level was not less diverse than at a regional level. The failure of pathogen populations to carry signatures of genetic drift suggests that the capacity for dispersal plays a chief role in recolonization; the high capability for long-distance dispersal enables virulent genotypes of *P. pachyrhizi* to move quickly among different populations.

The tendency of *P. pachyrhizi* to display weak genetic structure over a large geographic area may not be exclusive to fungal populations sampled in Brazil or Argentina. In Nigeria, *P. pachyrhizi* displayed about 90% of its genetic diversity within soybean fields, with very little genetic diversity (about 6%) distributed among fields (Twizeyimana *et al.*, 2011). The range of dispersal distances is potentially large for urediniospores of *P. pachyrhizi* given that the winds produced by storms constitute effective dispersal mechanisms (Isard *et al.*, 2005). In tropical and subtropical regions, substantial sources of airborne urediniospores may exist year-round owing to the large array of secondary hosts for *P. pachyrhizi*. In contrast to most rust species, which have a narrow range of host species, *P. pachyrhizi* infects more than 42 species in 19 genera of legumes (Bromfield, 1984; Ono *et al.*, 1992). In Brazil, the alternative host range of *P. pachyrhizi* encompasses many legume species, including kudzu (*Pueraria lobata*), perennial soybean (*Neonotonia wightii*), lima bean (*Phaseolus lunatus*), common bean (*Phaseolus vulgaris*) and least snoutbean (*Rhynchosia minima*) (Kato *et al.*, 2007). These alternative hosts, in addition to volunteer soybean plants, may provide a significant supply of fresh urediniospores as inoculum to perpetuate the infection. The lack of genetic structure and differentiation in *P. pachyrhizi* at the diverse geographic scale suggests that the regular re-establishment of the disease is unlikely to be the result of inoculum derived from a few, possibly local, sources. More likely, both local and distant sources of airborne inoculum are important for both the re-establishment of ASR and the maintenance of minimum levels of fluctuation in the genetic composition of *P. pachyrhizi* populations from year to year.

The dense sampling of molecular variation at the local level enabled the investigation of whether the presence of *Rpp* genes (*Rpp1*, *Rpp2*, *Rpp3* or *Rpp4*) could shape the level of molecular variation of environmental samples of *P. pachyrhizi*. In this study, the presence of *Rpp* genes in the hosts did not lead to a reduction of molecular variation in the pathogen population. A probable explanation for this result is the possibility that asexual reproduction may not be the only mode of reproduction of *P. pachyrhizi*; the pathogen may undergo parasexual processes to a certain extent, even though its life cycle in the wild remains incompletely understood. Hyphal anastomosis among germ tubes of urediniospores has been observed for *P. pachyrhizi* (Vittal *et al.*, 2012). In the case that fusion of hyphae is followed by heterokaryosis, nuclear fusion, recombination or chromosome reassortments (Vittal *et al.*, 2012), reshuffling of genes may take place from generation to generation, which would disentangle the expected resilient linkage between an avirulence gene and a marker gene (either ITS or *ARF*) in a given fungal clone. High frequency of virulence genes in field populations of ASR would also account for the lack of association between molecular variation and resistance genes as observed in the present study. Recently, field studies showed that *Rpp1–Rpp4* genes performed poorly against *P. pachyrhizi* in Brazil and that there was no association of pathogenicity with the geographic location of the populations (Akamatsu *et al.*, 2013). These findings have implications for breeding programmes. The deployment of resistant soybean cultivars has been hampered by the lack of durable sources of resistance (Hartman *et al.*, 2005). The present results strongly suggest that the local deployment of soybean varieties harbouring single ASR resistance genes will expose these varieties to pathogen populations that may carry the full range of genetic variation that is present at the regional level. Pathogens such as *P. pachyrhizi*, which display long-distance dispersal

capabilities of urediniospores, pose a great risk of epidemics given that each urediniospore contains a linkage block of co-adapted alleles (McDonald & Linde, 2002). The lack of geographically substructured populations in *P. pachyrhizi* may be one of the reasons why this pathosystem has the ability to overcome single-gene resistance so rapidly.

Acknowledgements

The authors would like to express their sincere gratitude to the field agronomists who provided assistance during trips and samplings. This study forms part of the MS dissertation of V. R. J. L. O. O. received a fellowship from CNPq (PQ 304153/2012-5) and grants from FAPEMIG (PPM-00291-13) and CNPq (473970/2011-2).

References

Akamatsu H, Yamanaka N, Yamaoka Y *et al.*, 2013. Pathogenic diversity of soybean rust in Argentina, Brazil, and Paraguay. *Journal of General Plant Pathology* **79**, 28–40.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–10.

Bandelt HJ, Forster P, Rohl A, 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48.

Barnes CW, Szabo LJ, Bowersox VC, 2009. Identifying and quantifying *Phakopsora pachyrhizi* spores in rain. *Phytopathology* **99**, 328–38.

Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H, 2013. ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources* **13**, 218–24.

Bromfield KR, 1984. *Soybean Rust*. Monograph No.11. St Paul, MN, USA: APS Press.

Colwell RK, Chao A, Gotelli NJ *et al.*, 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology* **5**, 3–21.

Elder JF, Turner BJ, 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Quaternary Review of Biology* **70**, 297–320.

Excoffier L, Smouse PE, Quatro JM, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics* **131**, 479–91.

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

Frederick RD, Snyder CL, Peterson GL, Bonde MR, 2002. Polymerase chain reaction assays for the detection and discrimination of the soybean rust pathogens *Phakopsora pachyrhizi* and *P. meibomiae*. *Phytopathology* **92**, 217–22.

Freire MCM, Oliveira LO, Almeida AMR *et al.*, 2008. Evolutionary history of *Phakopsora pachyrhizi* (the Asian soybean rust) in Brazil based on nucleotide sequences of the internal transcribed spacer region of the nuclear ribosomal DNA. *Genetics and Molecular Biology* **31**, 920–31.

Freire MCM, Silva MR, Zhang XC, Almeida AMR, Stacey G, Oliveira LO, 2012. Nucleotide polymorphism in the 5·8S nrDNA gene and internal transcribed spacers in *Phakopsora pachyrhizi* viewed from structural models. *Fungal Genetics and Biology* **49**, 95–100.

Frey B, Suppmann B, 1995. Demonstration of the expand PCR system's greater fidelity and higher yields with a *lacI*-based PCR fidelity assay. *Biochemica* **2**, 34–5.

Garcia MG, Silva RS, Carniello MA, Veldman JW, Rossi AAB, Oliveira LO, 2011. Molecular evidence of cryptic speciation, historical range expansion, and recent intraspecific hybridization in the Neotropical seasonal forest tree *Cedrela fissilis* (Meliaceae). *Molecular Phylogenetics and Evolution* **61**, 639–49.

Hartman GL, Miles MR, Frederick RD, 2005. Breeding for resistance to soybean rust. *Plant Disease* **8**, 664–6.

Isard SA, Gage SH, Comtjois P, Russo JM, 2005. Principles of the atmospheric pathway for invasive species applied to soybean rust. *BioScience* **55**, 851–61.

Kato M, Yorinori JT, Paiva WM, Yamaoka Y, Almeida AMR, 2007. Epidemiological studies of soybean rust in South America: host susceptibility and pathogenic races. In: Suenaga K, Kudo H, Oshio S, eds. *Comprehensive Studies on the Development of Sustainable Soybean Production Technology in South America*. Tsukuba, Japan: JIRCAS, 35–46.

Kubisiak TL, Anderson CL, Amerson HV, Smith JA, Davis JM, Nelson CD, 2011. A genomic map enriched for markers linked to *Avr1* in *Cronartium quercuum* f. sp. *fusiforme*. *Fungal Genetics and Biology* **48**, 266–74.

Mast AR, Kelso S, Richards AJ, Lang DJ, Feller DMS, Conti E, 2001. Phylogenetic relationship in *Primula* L. and related genera (Primulaceae) based on noncoding chloroplast DNA. *International Journal of Plant Science* **162**, 1381–400.

McDonald BA, Linde C, 2002. Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**, 349–79.

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson K-H, 2008. Intraspecific ITS variability in the kingdom fungi as expressed in the International Sequence Databases and its implications for molecular species identification. *Evolutionary Bioinformatics Online* **4**, 193–201.

Ono Y, Buritica P, Hennen JF, 1992. Delimitation of *Phakospora*, *Physopella* and *Cerotelium* and their species on Leguminosae. *Mycological Research* **96**, 825–50.

Ordoñez ME, Kolmer JA, 2007. Simple sequence repeat diversity of a worldwide collection of *Puccinia triticina* from durum wheat. *Phytopathology* **95**, 574–83.

Pandey AK, Yang C, Zhang C *et al.*, 2011. Functional analysis of the Asian soybean rust resistance pathway mediated by *Rpp2*. *Molecular Plant–Microbe Interactions* **24**, 194–206.

Preacher KJ, 2001. Calculation for the chi-square test: an interactive calculation tool for chi-square tests of goodness of fit and independence. [http://www.quantpsy.org/chisq/chisq.htm]. Accessed 16 September 2014.

Queiroz CS, Batista FRC, Oliveira LO, 2011. Evolution of the 5·8S nrDNA gene and internal transcribed spacers in *Carapichea ipecacuanha* (Rubiaceae) within a phylogeographic context. *Molecular Phylogenetics and Evolution* **59**, 293–302.

Rose PP, Korber BT, 2000. Detecting hypermutations in viral sequences with an emphasis on G→A hypermutation. *Bioinformatics* **16**, 400–1.

Schneider RW, Hollier CA, Whitman HK *et al.*, 2005. First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States. *Plant Disease* **89**, 773–9.

Twizeyimana M, Ojiambo PS, Haudenshield JS *et al.*, 2011. Genetic structure and diversity of *Phakopsora pachyrhizi* isolates from soyabean. *Plant Pathology* **60**, 719–29.

Vittal R, Yang HC, Hartmann GL, 2012. Anastomosis of germ tubes and migration of nuclei in germ tube networks of the soybean rust pathogen, *Phakopsora pachyrhizi*. *European Journal of Plant Pathology* **132**, 163–7.

Zhang XC, Freire MCM, Le MHG *et al.*, 2012. Genetic diversity and origins of *Phakopsora pachyrhizi* isolates in the United States. *Asian Journal of Plant Pathology* **6**, 52–65.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Figure S1 Map that shows the municipalities in which soybeans are produced (shaded grey) in Brazil and the sites where the three experimental stations are located.

Figure S2 Distribution of autapomorphies among ITS ribotypes obtained from bacterial clones of *Phakopsora pachyrhizi*.

Figure S3 Sequence alignment of the variable sites in two gene regions studied in *Phakopsora pachyrhizi*.

Table S1 Sampling locations of *Phakopsora pachyrhizi* in Argentina and Brazil.

Table S2 Chi-square tests for the association of the number of sequences and the number of ribotypes of the ITS region of *Phakopsora pachyrhizi*.

Table S3 Chi-square tests for the association of the number of sequences and the number of haplotypes of the ADP-ribosylation factor gene of *Phakopsora pachyrhizi*.