Curr Genet (2009) 55:521–536
DOI 10.1007/s00294-009-0263-9

RESEARCH ARTICLE

Identification and occurrence of the LTR-*Copia*-like retrotransposon, *PSCR* and other *Copia*-like elements in the genome of *Phytophthora sojae*

Shiromi Basnayake · Donald J. Maclean · Stephen C. Whisson · Andre Drenth

Received: 16 February 2009 / Revised: 22 June 2009 / Accepted: 9 July 2009 / Published online: 30 July 2009

Abstract Sequence analysis of the genomic region of *Phytophthora sojae* close to the *Avr*4/6 locus specifying virulence on soybean identified a *Ty1/Copia*-like retrotransposon that we have named *Phytophthora sojae Copia*-like retrotransposon (*PSCR*). Twelve near-complete homologs of *PSCR* were found in the published *P. sojae* genome sequence, none of which encoded a full-length polyprotein characteristic of *Copia*-like retrotransposons, or appears to exhibit transcriptional activity or show evidence of recent movement, suggesting they are non-functional and unlikely to have caused pathogenic variability. However, reconstructed consensus *PSCR* sequence encoding a full-length polyprotein resembles a functional, ancestral retroelement within *P. sojae*. Homologs were also found in sequence databases of other *Phytophthora* species. Database searches found other families of *Copia*-like elements in genomes of *P. sojae*, *P. ramorum* and *P. infestans* that were different from members of the PSCR family and from *Copia*-like elements reported in other organisms. It is possible that the various families of *Copia*-like retroelements identified in this study represent introgressions into the genome of ancient ancestor(s) of current *Phytophthora* species, where they have evolved and diverged considerably during the speciation. Some *Copia*-like families are transcriptionally active with the potential to transpose and contribute to pathogenic variation in current populations of *P. sojae*.

Keywords Linkage maps · Oömycetes · Transposable elements · Homologs · *Phytophthora infestans* · *Phytophthora ramorum*

Communicated by H. Osiewacz.

S. Basnayake (✉)
School of Biological Sciences, The University of Queensland, Brisbane, QLD 4072, Australia
e-mail: s.basnayake@uq.edu.au

S. C. Whisson
Plant Pathology Programme, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK

A. Drenth
Department of Primary Industries and Fisheries, Tree Pathology Centre, The University of Queensland, Brisbane, QLD 4068, Australia

S. Basnayake · D. J. Maclean
School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia

Introduction

The oömycete plant pathogen *Phytophthora sojae*, which causes root and stem rot of soybean (*Glycine max*), is diploid and homothallic and was long considered a poor model for genetic manipulation. However, this organism was found to be capable of outcrossing at a low frequency, which enabled the construction of chromosomal linkage maps and the genetic analysis of virulence/avirulence genes (Tyler et al. 1995; Whisson et al. 1994, 1995; May et al. 2002). Genetic approaches have revealed that *P. sojae* follows a gene-for-gene interaction with its host plant soybean and is hypothesized to have coevolved with its host (Whisson et al. 1995). Control of the disease mainly relies on the introduction of resistance genes (*Rps* genes) into commercial soybean cultivars through conventional plant breeding. The ability of *P. sojae* to mutate and recombine in the field to generate new races that overcome the resistance genes has reduced the effectiveness of this disease management approach (Dorrance and Schmitthenner 2000; Förster et al. 1994; Ryley et al.

1998). A better understanding of mechanisms that contribute to mutation at loci controlling virulence has the potential to assist the development of improved strategies for deployment of resistance genes in soybean.

Molecular causes of phenotypic variation in virulence and other characters are poorly understood for most phytopathogens, and have often been attributed to genetic instability caused by general mechanisms with the potential to cause rapid mutation, such as transposable elements (Shaw 1988). Active transposable elements can potentially knock out the function of any gene, including genes controlling general pathogenicity characters, specific avirulence genes in field isolates, or genes controlling cultural characteristics in the laboratory.

Because retrotransposons move via replication rather than excision, each copy is integrated permanently into the genome as evidenced by a unique flanking sequence at each point of insertion. Hence retrotransposon-induced mutations are relatively stable. It has been suggested that there might be a cause-and-effect relationship between retrotransposon family copy number in the genome and the propensity to insert into genes (SanMiguel et al. 1996). One group of retrotransposons, that includes the eponymous *Copia* from *Drosophila melanogaster* (Mount and Rubin 1985), *Ty1* from *Saccharomyces cerevisiae* (Clare and Farabaugh 1985) and *Tnt1* from *Nicotiana tabacum* (Grandbastien et al. 1989), is generally referred to as *Copia*-like (or *Ty1/Copia* -like) retroelements. All *Copia*-like elements present the same general structure including a long terminal repeat (LTR), and encode a polyprotein that includes major domains with highly conserved amino acid sequences such as gag and pol, and the integrase domain located at the 5′ end of the reverse transcriptase domain (Grandbastien 1998; Flavell et al. 1997). Another LTR-containing class is the *Ty3/Gypsy* group where the integrase is located at the 3′ end of the RNaseH domain.

We have previously characterized the genomic region containing two co-segregating genes that confer avirulence against soybean lines carrying the *Rps*4 and *Rps*6 genes for resistance (the *Avr*4/6 locus, Whisson et al. 2004). In this work a chromosome walk was carried out and three overlapping cosmids (SC183, SC248 and SC480) were isolated that encompassed the *Avr*4/6 genomic region (Whisson et al. 2004). Analysis of a segregating population indicated that this region is recombination-rich, and some F_2 recombinants gave anomalous segregation ratios for polymorphic cDNA and phenotypic markers dispersed across this region. Construction of a physical map of the 50-kb insert of cosmid SC248 enabled us to identify F_2 progeny that gave segregation patterns which fitted the order of these cDNA markers in the genome, and identified a 24-kb region within cosmid SC248 that is closely associated with the *Avr*4/6 locus (Whisson et al. 2004).

In the current work, when sequencing this recombination-rich 24 kb region we found elements of retroelement motifs including a near-complete copy of a *Copia*-like element that we have named *Phytophthora sojae Copia*-like retrotransposon (*PSCR*). It was considered possible that a recent *PSCR* insertion was responsible for the mutation that generated virulence at the *Avr*4/6 locus. Furthermore, because relatively few *Copia*-like sequences had previously been reported in the genus *Phytophthora* (Jiang et al. 2005; Jiang and Govers 2006; Tooley and Garfinkel 1996), it was of interest to identify other *Copia*-like sequences in the *P. sojae* genome and explore homologies with cognate sequences in *P. infestans* and *P. ramorum*. Therefore the aim of the research described in this paper was to investigate the frequency and nature of homologs of *PSCR* and related *Ty1/Copia*-like retrotransposable elements in *P. sojae* and determine if related elements are present in other oömycetes of interest. We also sought to determine whether retroelements such as *PSCR* are likely to be still active in *P. sojae* and other oömycetes, and hence might contribute to the phenotypic variation observed at the *Avr*4/6 locus and other avirulence loci.

Materials and methods

Phytophthora sojae cultures and recombinant DNA techniques

Phytophthora sojae accessions UQ1200 (avirulent for both *Avr*4 and *Avr*6), UQ2990 (US7, virulent for both *avr*4 and *avr*6), UQ408, UQ416, R50, R17 (a Canadian isolate) and four recombinant isolates representing F_2 progeny of the cross UQ1200 × UQ2990 (F_2-81, F_2-104, F_2-115, F_2-203) used for this study were maintained and grown for DNA extraction as described by Whisson et al. (1994). Cosmid SC248 was derived from accession UQ1200 (Whisson et al. 2004). Cloning, subcloning, purification of plasmid DNA, sequencing, chromosome walking of cosmid SC248 via PCR, and contig assembly were done as described by Whisson et al. (2004), using the primers listed in Table 1 to sequence the *PSCR* element. To check the actual length of the *PSCR* element in cosmid SC248, two primers; *PSCR*-F from the region upstream to 5′LTR and *PSCR*-REV from the region downstream to 3′LTR, were designed (Table 1). PCR amplification of the full length of the retroelement in cosmid SC248 was carried out using *PfuTurbo* DNA Polymerase (Stratagene, USA) according to the manufacturer's instructions. The resulting approximately 5 kb amplification product was in agreement with the structural organization determined previously by sequencing within cosmid SC248 and genomic DNA (Whisson et al. 2004 and data not shown). The above (Table 1) and other primers were

Table 1 Oligonucleotides used in primer walking and amplifying the full-length *PSCR*

Oligonucleotide	Location	Sequence 5′–3′
PSCR-F	204 bp upstream 5′ LTR	ATCACATCTCGAACCTCCGAC
PSCR-REV	219 bp downstream 3′ LTR	AGCGCTTTGAATCGATGAGAC
Primer I	LTR	CGAAACTACCGATGCATA
Primer II	LTR/gag	CACGAGTTCGAGCATGCT
Primer III	gag	CAACTCGAACTCAGCACGA
Primer IV	gag/Protease	ACAACTAGCCACAAGAG
Primer V	Protease/Int	ACCTGATTTCAGTGGCTCAAC
Primer VI	Integrase core domain	TTGAGCGGCAATACGACAC
Primer VII	INT/RT	TCGACGCAGACAATAGC
Primer VIII	Reverse transcriptase	GCATTGACTTCGCAGAGACC
Primer IX	RT/RNaseH	AAGTGCGGTATGGAGAATAGC
Primer X	RNaseH	GCGATCAAGAACATGGA
Primer XI	LTR	ACTGTTGGAGATCGCAAG

Reverse sequences of Primers I–XI were used as primer sequences to sequence the complimentary strand to confirm the sequence obtained

used with genomic DNA of accession UQ2990 as template for PCR to obtain a partial sequence of the 3′-end of the *PSCR*-12A homolog in a genetically diverse race that differs in virulence from UQ1200. Sequencing was continued into the 3′ flank to ensure the correct *PSCR* homolog close to the *avr*4/6 locus of accession UQ2990 had been sequenced.

Restriction enzyme (New England Biolab, USA) digestions, gel electrophoresis and non-isotopic labeling of DNA with digoxygenin-dUTP (DIG-dUTP) by amplification with random primers (Boehringer Mannheim) were performed following the manufacturer's instructions. For Southern analysis, genomic DNA from ten *P. sojae* isolates was digested with *Eco*RI, agarose gels were prepared (Sambrook et al. 1989) and the DNA was transferred to nitrocellulose membranes (Sigma, USA) by standard procedures. Blots were probed by hybridization with a DNA fragment (1,270 bp) derived from a region of ORF1 (Fig. 1) which did not contain restriction sites for *Eco*RI; *Eco*RI site is upstream of the probe. The probe was labeled with digoxygenin (Boehringer Mannheim digoxygenin labeling kit) prior to hybridization (Sambrook et al. 1989). Hybridized blots were washed at high stringency [sequential washes in 1 × SSC and 0.1% sodium dodecyl sulfate (SDS) for 15 min at 65°C, 0.5 × SSC and 0.1% SDS for 15 min at 65°C, 0.1 × SSC and 0.1% SDS for 15 min at 65°C, Southern 1975] and visualized as described by the Boehringer Mannheim digoxygenin labeling kit.

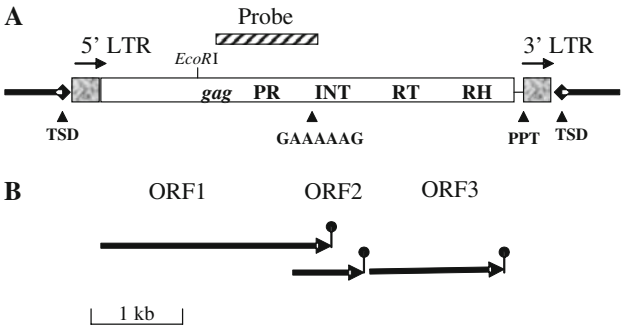

Fig. 1 Organization of *PSCR*: *Phytophthora sojae Copia*- like Retrotransposon. **a** Analysis of *PSCR* homologs in the *P. sojae* genome (Table 2) enabled the construction of a consensus sequence with a single ORF. The *open box* shows the single ORF obtained from the consensus sequence of *PSCR*, with the approximate positions of the domains encoding *gag*, protease (*PR*), integrase (*INT*), reverse transcriptase (*RT*) and RNase H (*RH*). TSD is the target site duplication (gcttg and gtttg for the *PSCR* sequence near the *Avr*4/6 locus, see Sect. "Results"). The heptanucleotide sequence GAAAAAG at position 2686 includes the most likely frameshifting site, and PPT is the polypurine tract (CGAGGAGGAC), representing key sequence regions necessary for replication. The complete consensus sequence possesses a single *Eco*RI site prior to the *gag* domain, as indicated. The original *PSCR* sequence close to the *Avr*4/6 locus in *P. sojae* had a polyprotein coding sequence that was split among three ORFs (see Fig. 1b below). **b** *Arrows* represent the open reading frames (*ORFs*) obtained from the initial *PSCR* sequence found in scaffold 12 near the *Avr*4/6 locus (this study). Compared to the consensus sequence in **a**, a 2-bp deletion and a single nucleotide substitution created two premature stop codons resulting in the three shortened ORFs (indicated by *pins*) shown in the figure

Screening of genomic and EST databases

Phytophthora sojae and *P. ramorum* genomic databases, produced by the DOE Joint Genome Institute (US Department of Energy, Washington), were accessed at http://genome.jgi-psf.org/. *P. infestans* and *P. sojae* EST databases were accessed at https://xgi.ncgr.org/spc (Syngenta *Phytophthora* Consortium (SPC) EST sequence data bases), http://www.pfgd.org (*Phytophthora* Functional Genomics Database) (Qutob et al. 2000), and National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). *P. infestans* genome sequence data, produced by the Broad Institute of MIT, Harvard and Cambridge, were accessed at http://www.broad.mit.edu/annotation/

genome/phytophthora_infestans. Additional oömycete genome sequence data for *Hyaloperonospora parasitica* and *P. capsici*, and also *P. sojae* and *P. ramorum*, were accessed through the VBI Microbial Database (http://phytophthora.vbi.vt.edu/phytophthora/develop/index.html; Tripathy et al. 2006). Initially the genomic sequence databases of *P. sojae*, *P. ramorum*, *P. infestans, P. capsici* and *H. parasitica* noted above were screened using the complete *PSCR* sequence as a query on the TBLASTN or BLASTP platforms with default parameters (1×10^{-5} for expectation value, 0 for word size and BLOSUM 62 for matrix). EST databases were screened as described in Table 4.

To identify further, more divergent *Copia*-like sequences in the *P. sojae* or *P. ramorum* genomes (http://genome.jgi-psf.org/), amino acid sequences representing the conserved region of the RT domain (as defined by Xiong and Eickbush 1990) from *Copia*-like elements of interest as described in the text were used as queries for screening, using the BLASTP platform with default parameters (1×10^{-5} for expectation value, 3 for word size and BLOSUM 62 for matrix). The deduced amino acid sequences representing the RT domain from the best 10–30 hits for each query were aligned to identify their relationship with other *Copia*-like retroelements.

Additional partial or complete *Copia*-like retroelement accessions of interest were obtained from the following sources: three *Copia*-like sequences from *P. infestans*, AY830098 (*CopiaPi-1*), AY830099 (*CopiaPi-2*) and AY 830100 (*CopiaPi-3*) (Jiang and Govers, 2006); *Copia* from *D. melanogaster*, X02599; *Tvv1* from *Vitis vinifera*, AF116598; *Tnt1* from *N. tabacum*, X13777; *Ty1* from *Saccharomyces cerevisiae* (Clare and Farabaugh 1985); *Elsa* from *Stagonospora nodorum*, AJ277966 (Rawson 2000); *Tgmr* from *G. max*, U96748 (Bhattacharyya et al. 1997). The *Copia*-like sequences were aligned to the following *Gypsy*-like sequences: *Gypsy* from *D. melanogaster*, M12927; *Ty3* from *S. cerevisiae*, M23367; *Skippy* from *Fusarium oxysporum*, L34658; *CfT-1* from *Cladosporium fulvum*, Z11866. *Skippy* and *CfT-1* were used to optimize the alignment of *Gypsy* and *Ty3* with the *Copia*-like elements, and were removed from the final alignment used to produce Fig. 4. Where necessary, frameshifts were corrected to obtain full deduced amino acid sequences of retroelement domains being compared.

DNA and protein sequence analysis

Computer-assisted analysis of the sequence data was performed using multiple sequence alignment with the Clustal series of programs (Clustal X2 http://www.clustal.org/) (Chenna et al. 2003). Alignments were adjusted manually with the Protein Family Alignment Annotation Tool (PFAAT) program (http://pfaat.sourceforge.net/); where necessary, alignments were subjected to successive iterations of Clustal X2 and manual adjustment to obtain the best fit of the less conserved regions between conserved motifs. Initial phylogenetic analysis used the Neighbor Joining programs in either Clustal X2 or PFAAT. For more rigorous phylogenetic analysis, the proml, seqboot and consense programs of Phylip3.67 (http://evolution.genetics.washington.edu/phylip.html) were used with default settings and the Jones–Taylor–Thornton probability model, Global Rearrangements, Jumble (1) and multiple data sets (500 bootstrap) options as specified. For comparison of reverse transcriptase homologies, sequences encompassing the core conserved regions reported by Xiong and Eickbush (1990) were selected prior to alignment and phylogenetic analysis.

Results

Identification of a retroelement sequence from the *Avr*4/6 genomic region of *P. sojae*

The full 50 kb sequence of the recombination-rich genomic region described by Whisson et al. (2004) as encompassing the *Avr*4/6 locus, was obtained by PCR "primer walking" using cosmid SC248 derived from accession UQ1200 as template. This enabled us to identify a 24-kb region flanked by cDNA markers cDNA71 and cDNA100 that was considered by Whisson et al. (2004) as likely to contain the *Avr*4/6 locus. Within this 24 kb region, a contiguous 4,980 bp sequence was identified that comprised a retrotransposon-like element flanked by two almost identical LTRs in the same orientation (Fig. 1). This element was named *PSCR* as defined in the Introduction, for reasons detailed below. The *PSCR* nucleotide sequence appears in the GenBank database under the accession number DQ485162. A search of the *P. sojae* genomic database identified 12 close homologs discussed in more detail later (Table 2). The general organization of this element was very similar in structure to *Copia*-like retrotransposons previously reported in other organisms (Flavell et al. 1997), as it includes domains encoding the gag, protease (PR), integrase (INT), reverse transcriptase (RT) and RNaseH (RH) proteins in the same order as reported for other *Copia*-like retrotransposons (Fig. 1a).

Internal organization of *PSCR*

The *PSCR* retroelement identified within the *Avr*4/6 region had the following features that are characteristic of *Copia*-like retrotransposons. It had LTRs at each end that were identical apart from one G > A transition (LTR total length 228 bp, 99.6% nucleotide identity between LTRs). Because the insertion process creates a pair of identical LTRs that are

Table 2 Near full copy homologs of *PSCR* showing structural heterogeneity

Scaffold[a]	Variations compared to consensus sequence		Notes
	Transitions/ transversions[d]	Deletions and rearrangements etc.[e]	
12-A (*PSCR*)[b]	24/14	(−2)	Complete sequence data (located within *Avr*4/6 region)
12-B[c]	30/13	(−1) (−1) (−53) (−1) (−11) (+9) (−5) (−14)	Complete sequence data
2	15/8	(5′LTR) (−1) (−10)	Sequence starts at no. 834[h]
4	74/30	No deletions	Complete sequence data
19	42/19	(−1) (−44) (−1), large insertion[f]	Sequence starts at no. 633[h]
21	19/11	No deletions	No. 1311–2322[h] data absent
24	6/3	No deletions	5′ and 3′ LTR data absent (no. 960–4133 available)[h]
29	14/8	(−1) (−10) (−68)	No. 948–2234[h] absent[g]
33	21/15	(−315) (−2)	Complete sequence data
52	16/11	No deletions	Complete sequence data
69	45/33	(−4) (−15[g]) (−1) 15 bp replaced by G_8	Complete sequence data
99	20/13	(−1) (−2)	Complete sequence data
102	45/22	(−11) (−1) (−12) (−1) (−1)	Complete sequence data

Homologs at different loci are indicated by their scaffolds in the genomic sequence database for *P. sojae*. Homologs are compared to a consensus sequence of 4,980 bp obtained through the Clustal X2 program and corrected manually

[a] Incomplete homologs from the database of less than 2,500 bp are not shown

[b] Sequenced during this investigation by primer-walking

[c] In complementary strand of Scaffold 12 compared to *PSCR* homolog 12A

[d] Transitions and transversions identified by comparison to consensus sequence

[e] Deletions shown as number of base-pairs lost compared to consensus sequence, and are listed sequentially from 5′ end

[f] ~10 kb Insertion near 5′ end of element; insertion is homologous to *Gypsy*-like retrotransposons

[g] Replaced by other (unrelated) sequence

[h] Incomplete sequence information in database

subsequently subject to random mutation (Jiang et al. 2005), the presence of a single nucleotide substitution suggests that *PSCR* moved to this site during relatively recent evolution, e.g., during the speciation of *P. sojae* or generation of intraspecific genetic diversity. The LTR terminated in short inverted repeats, 5′…TG-3′ and 5′…CA-3′, which are characteristic of retroviral and most retrotransposon LTRs and are required for integration (Varmus and Brown 1989). There was an imperfect 5 bp target site duplication (TSD), (gcttg and gtttg), characteristic of the duplications that are generated during the insertion stage of retrotransposition (Fig. 1). These TSDs together with the two LTRs delimited the extremities of the repeated sequence. These LTRs do not encode any known proteins, but they do contain motifs resembling the promoters and terminators associated with the transcription of other LTR retrotransposons (McHale et al. 1992). For example, five pairs of short direct repeats were observed in each LTR (data not shown) followed by a region similar to the polyomavirus enhancer core sequence, 5′-AACCACA-3′ (Nuchprayoon et al. 1994).

Analysis of deduced amino acid sequences showed that the *PSCR* internal structure contained a minimal leader region of 4 bp between the 5′ LTR and the first open reading frame (ORF1). No tRNA primer binding site for initiation of translation was found, implying that an internal tRNA fragment rather than an intact tRNA is used for priming (Kikuchi et al. 1986). In *PSCR* a polypurine tract, CGAGGAGGAC, was present just upstream of the 3′ LTR (Fig. 1), consistent with its involvement in second strand priming during replication (Wilhelm and Wilhelm 2001).

Translation of the DNA sequence revealed three open reading frames (ORF1, ORF2, ORF3) of 2,469 bp (823 aa), 462 bp (154 aa) and 1,536 bp (512 aa) respectively (Fig. 1b). The deduced amino acid sequence of these ORFs revealed functional domains homologous to the protease, integrase, reverse transcriptase (RT) and RNaseH proteins from retroelements of diverse organisms. The RT domain showed higher homologies to *Copia*-like sequences found in other organisms than to *Ty3/Gypsy*-like retrotransposons (Fig. 4). The functional domains were arranged in an order characteristic of active *Copia*-like elements that contain a single ORF encoding a polyprotein (Pelsy and Merdinoglu 2002), and in a different order to *Gypsy*-like retrotransposons which also possess LTRs. The three ORFs could be a consequence of mutations in a single ORF from an ancestral retroelement that generated two stop codons including one

frameshift mutation (Fig. 1; Table 2). As suggested below, it is possible that an ancestral *PSCR* encoded one long ORF (1,492 aa) and the *PSCR* element in the *Avr*4/6 region may be non-functional due to the presence of stop codons.

Copy number and distribution of *PSCR* homologs in the *P. sojae* genome

Southern analysis was used to determine if the *PSCR* sequence described above represents an active mobile element within the *P. sojae* genome. Genomic DNA samples from ten isolates representing genetically-distinct field isolates and F2 progeny of *P. sojae* were digested with *Eco*RI (which has one restriction site within the complete *PSCR* sequence and also cleaves outside of the probe target sequence), and probed with a 1.27-kb fragment derived from the *PSCR* sequence (Fig. 1). High stringency conditions were used to visualize only sequences closely related to *PSCR*. At least eight distinct bands were seen on Southern blots of each isolate (Fig. 2), indicating that *PSCR*-like sequences are moderately repeated in the *P. sojae* genome. Banding patterns were identical among nine of the ten isolates, with the Canadian isolate R17 showing two polymorphic bands, suggesting that most of the transpositional or mutational events that generated them have not happened recently. All accessions except Canadian isolate R17 also gave a ~ 550-bp fragment, and the largest fragment from R17 was polymorphic and smaller than the largest fragment in the other isolates tested.

By searching the genomic sequence database for *P. sojae* (http://genome.jgi-psf.org) using the full *PSCR* nucleotide sequence as query, we identified 12 full or nearly full length *PSCR* elements sharing greater than 96% homology (Table 2), and six incomplete elements representing portions of this conserved *PSCR* sequence (data not shown). The initial *PSCR* sequence (12-A) within the *Avr*4/6 genomic region is represented in scaffold 12 of the draft *P. sojae* genome sequence. Another *PSCR* homolog (12-B) is present in scaffold 12 in an inverted orientation (Table 2). Alignment of these 12 homologs identified scattered point mutations, insertions and deletions that were mostly unique to each homolog together with some mutations shared by subsets of homologs. This suggests that most homologs were created by transpositional events prior to generation of the genetic diversity among *PSCR* homologs seen in current genotypes of *P. sojae* such as the isolate used for obtaining the genome sequence (Zhang et al. 2006). From the alignment it was possible to construct a consensus *PSCR* sequence based on the most frequent base at each nucleotide site among the 12 homologs (the consensus sequence and an alignment of the 12 homologs are available from s.basnayake@uq.edu.au). The consensus *PSCR* sequence and most of the homologs listed in Table 2 possess a common *Eco*RI site prior to the gag domain (Fig. 1a). However the homolog in scaffold 99 of the genome sequence has an additional *Eco*RI site, 542 bp downstream of the common *Eco*RI site in the consensus *PSCR* sequence, and is probably the source of the ~ 550 bp *Eco*RI restriction fragment seen after Southern analysis of all isolates of *P. sojae* except R17 (Fig. 2; note position of the probe in Fig. 1a). The apparent absence of the 550-bp band in R17 (Fig. 2) could be explained if R17 lacks the *PSCR* homolog in scaffold 99 of the isolate used to obtain the genome sequence, or if R17 possesses this homolog but lacks the mutation that generated the additional *Eco*RI site within it.

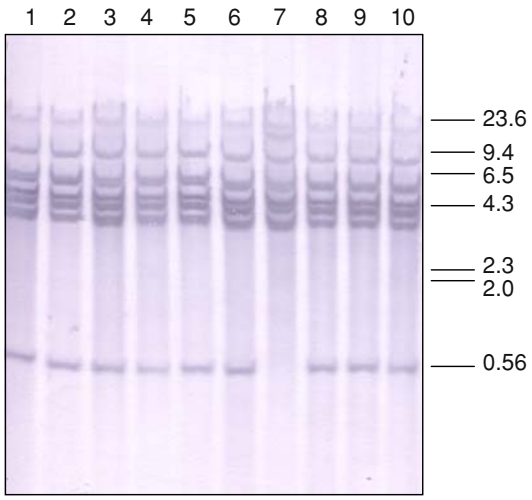

Fig. 2 Southern hybridization of genomic DNA from *P. sojae* isolates digested with *Eco*RI and probed with a 1.2-kb fragment from ORF1 of *PSCR*. *Lanes 1–10* respectively: *1* UQ2990, *2* UQ1200, *3* F2-81, *4* F2-104, *5* F2-115, *6* F2-203, *7* R17 (Canadian isolate), *8* R50, *9* UQ408, *10* UQ416. *Numbers to the right* are DNA molecular weight markers (kb)

The nucleotide sequence of the initial *PSCR* homolog close to the *Avr*4/6 locus (*PSCR*-12A) identified in accession UQ1200, was compared to the 3′ end of its allelic equivalent in accession UQ2990 (Whisson et al. 1994, 1995). Three transitions and one transversion were observed at the following positions numbered from the 5′ end of the 4,980 bp retroelement (UQ1200 → UQ2990): 3,974 (C → A, ORF3), 4,004 (T → C, ORF3), 4,717 (G → A, ORF3), 4,843 (T → C, LTR). This mutation frequency is similar to polymorphisms between UQ1200 and UQ2990 observed at other locations across the wider *Avr*4/6 region (Whisson et al. 2004). This suggests that the transpositional insertion of *PSCR*-12A occurred in an ancestral population of *P. sojae* prior to its divergence into the genetically distinct UQ1200 and UQ2990 genotypes.

The full or nearly full-length *PSCR* homologs found in the genome sequence show the consensus end motifs 5′-TG…CA-3′ expected for *Copia*-like retrotransposons, where the LTR sequences have inserted into the flanking

Table 3 5′ and 3′ Flanking regions of *PSCR* homologs, showing 5 bp target site duplications (TSD) (in bold)

Scaffold	Upstream of 5′ LTR	*PSCR* homolog	Downstream of 3′ LTR	Remarks
12-A (*PSCR*)	TTTCACCTCTTTCACTGGGC**GCTTG**	……//……	.**GTTTG**GCATTGACTTGTCGCGACGC	This study, Imperfect TSD
12-B	CGCCAAGATTGCGATCTGAT**TGATT**	……//……	.**TGATT**CGCCACCTCGCCACCCCCAC	Perfect TSD
2	NA	……//……	.AATAACAGTCAAGACATTCTCACGG	5′ LTR absent
4	TCAATGCTAAAACAATAATC**GTTTA**	……//……	.**GTTTA**GGCACAAGGTAAGGTTCCAC	Perfect TSD
19	AATGTAGGCAGGCCGACTTA**CAGAA**	……//……	.**CAGAA**TATGATGGGCATGATGACTT	Perfect TSD
21	TTCAACAACGCTCCCTTGCC**GATAC**	……//……	.**GATAC**GTCCCGACGACAATACGACT	Perfect TSD
29	GAGGCGTCTTAATTTCCACT**GTGTA**	……//……	.**GTGTA**GACCCCATCCACTACCGAGC	Perfect TSD
33	GCACCGATCATACCCGGGAA**CTCAG**	……//……	.**CTCAG**TACATTAACCCACTCACTAT	Perfect TSD
52	GTGCGTGTAACCCCGTTACA**AAGAA**	……//……	.**AAGAA**GGACAATCACGATGATGATC	Perfect TSD
69	GTACCTAGTACTGTCATGCGgagct	……//……	.TGCCTGCTTGGTACTGCATGCACAT	5′ of left LTR absent
99	TTTTGATTTGGTGGTCCCAG**CGTGG**	……//……	.**CGTGG**TAGTTTGCGAGATTTTCAGT	Perfect TSD
102	GGTCTCAAGTGGCTGCGAAA**GTGTA**	……//……	.**GTGTA**CGATTGGACCCGCTCGGTTG	Perfect TSD

The sequences of homologs were obtained from the JGI *P. sojae* genome database

NA not present or insufficient data available

genomic DNA (Varmus and Brown 1989). The 5 bp TSDs that immediately flank these 5′-TG…CA-3′ ends differ among the homologs (Table 3), consistent with them representing independent insertions in different locations of the *P. sojae* genome. An alignment of the homologs revealed that the LTR sequence of *PSCR* is highly conserved (data not shown). *Copia*-like elements reported in other organisms typically contain a single long ORF that extends almost to the inside end of the flanking LTRs (McHale et al. 1992) and the full-length consensus *PSCR* sequence contains a single long ORF within the LTRs. However, all homologs of *PSCR* present in the *P. sojae* genome contained multiple ORFs due to premature stop codons caused by nucleotide substitutions (multiple transitions and transversions in all homologs), deletions (all homologs except those in scaffolds 4, 21, 24, 52), as well as insertions or sequence rearrangements (homologs in scaffold 12B, 69) (Table 2). For example, Fig. 1a shows the position of stop codons generated by frameshifts and point mutations in the initial *PSCR* homolog found close to the *Avr*4/6 region and located in scaffold 12 (homolog 12A). The consensus sequence lacks the deleterious mutations that gave rise to premature stop codons within the polyprotein coding sequence of these 12 homologs, and presumably resembles the ancestral *Copia*-like *PSCR* element that spread across the *P. sojae* genome.

Codon usage of *PSCR*

To evaluate possible differences in codon usage in *PSCR* compared to its host *P. sojae* genome, we first calculated the percentage of each base in the third position of codons (GC3) in the deduced ORF of the consensus *PSCR* polyprotein and showed it to be GC-rich (GC3: 29% G, 33% C, total GC3 61%), compared to the overall GC content of the coding region of *PSCR* (54%). The GC3 value for *PSCR* was appreciably lower than that determined from 10,000 deduced ORFs in the *P. sojae* genome: GC3 = 76%; overall GC content = 60% (Jiang and Govers 2006), and lower than in six cDNAs described by Whisson et al. (2004) that are located close to the *Avr*4/6 locus (GC3 66%, overall GC content 58%). These data and a preliminary analysis of the frequency of individual codons (data not shown) suggest that the *PSCR* polyprotein differs in codon usage from ORFs in the host *P. sojae* genome (Jiang and Govers 2006).

Comparison of *PSCR* with conserved motifs in other *Copia*-like retrotransposons

The *PSCR* sequence contains a number of motifs characteristic of *Copia*-like retrotransposons, perhaps the most striking being centered on a potential zinc-finger domain,

Element	Zinc Finger Domain	Protease
	* * ** *	***
PSCR	CHVCGKPGHKIFTC	VDSGATHHL
CopiaPi-2	CLYCLKSGRFKSDC	VDTGAGRAI
Copia	CHHCGREGHIKKDC	LDSGASDHL
Tvv1	CFYCHEAGHTKKNC	IDSGATDHM
Tnt*1-94*	CYNCNQPGHFKRDC	VDTAASHHA
Tgmr	CAYCRKLGHTIDVC	LDSGATDHV
Skippy	CYNCGKKGHYEREC	SDSGYDTRS
CfT-1	CYGCGKPGHIARDC	IDSGASGNF

Element	Integrase core Domain
	* * * * **
PSCR	IRSDGGGEFGSTRLGRFLRNRGILHQETEAGTSSSNGKAER
CopiaPi-2	RRTDGGTEFINTEVSKICDKLGLQFESSNVESPEENGSAER
Copia	LYIDNGREYLSNEMRQFCVKKGISYHLTVPHTPQLNGVSER
Tvv1	LRSDNGKEYVSNSFQNYMSHNGILHQTSCVDTPSQNGVAER
Tnt*1-94*	LRSDNGGEYTSREFEEYCSSHGIRHEKTVPGTPQHNGVAER
Tgmr	LQSDNGAEFLMHDF---YARKGIIHQTTCVETPEQNGIAER
Skippy	ILSDRGPTFAATFWQSLMARLGLNHRLTTAFRPQVDGQTER
CfT-1	FITDRDKLFTSNYWKTLMGTIGIKHKLSTAYHPETDGQTER

Element	Reverse Transcriptase Domain 1	Domain 2	Domain 3
	* **	* *	* **
PSCR	QGDVPNAYLRA <6aa>	YMRAPVGLQLP <54aa>	LGLYVDDV
CopiaPi-2	QMDVSTAFLNG <6aa>	YMRQPMGFRKG <59aa>	VCVYVDDL
Copia	QMDVKT-FLHG <6aa>	YMRLPQGISCN <58aa>	VLLYVDDV
Tvv1	QLDIKN-FLHG <6aa>	YLEOPPGFVAQ <56aa>	LVVYVDDI
Tnt*1-94*	QLDVKTAFLHG <6aa>	YMEOPEGFEVA <57aa>	LLLYVDDM
Tgmr	QLDVNNAFLHG <6aa>	YMKLPPGLVVD <56aa>	ILVYVDDI
Skippy	EADEWKTAFRT <6aa>	YLVMPFGLTNA <19aa>	VVCYLDDI
CfT-1	EGEEWKTAFRT <6aa>	FLVMPMGLTNA <19aa>	VVAYMDDI

Element	RNaseH Domain 1	Domain 2
	* ***	** **
PSCR	RMKHIN <24aa>	MKADMFTKPLGATLHGRNLAMIK
CopiaPi-2	STKHID <24aa>	MIADALTKPLS
Copia	RAKHID <24aa>	QLADIFTKPLPAARFVELRDKLG
Tvv1	RTKHIE <24aa>	QLGDIFTKALNGTRVEYFCNKLG
Tnt*1-94*	RTKHID <24aa>	NPADMLTKVVPRNKFELCKELVG

Fig. 3 Alignment of portions of the putative gag, pol, reverse transcriptase and RNaseH domains for various retroelement protein sequences: *PSCR,* homolog in scaffold 12A; *CopiaPi-2*, *Copia*-like element from *P. infestans* (AY830099)*; Copia* from *Drosophila melanogaster; Skippy* from *Fusarium oxysporum*; *CfT-1* from *Cladosporium fulvum*; *Tvv1* from *Vitis vinifera; Tnt1-94* from *Nicotiana tabacum;* and *Tgmr* from *Glycine max. Skippy* and *CfT-1* represent *Gypsy*-like elements, the other elements are all *Copia*-like. *Asterisk* indicate highly conserved amino acids

Cys-X2-Cys-X9-Cys (Fig. 3) (Leong et al. 1994). This zinc finger domain is also found in eukaryotic proteins involved in RNA binding or single strand DNA binding. In the first 450 amino acid region of the polyprotein encoded by the consensus *PSCR* sequence, the conserved amino acid DSG motif was identified (Fig. 3), which presumably acts as the active site of the acidic protease involved in the processing of the polyprotein (Fig. 1) (Katoh et al. 1987). The putative integrase (endonuclease) lies within a 633-790 amino acid region of the polyprotein, and shows all the highly conserved amino acids known from other *Copia*-like retroelements (Fig. 3). Several other short stretches of amino acids, shown to be essential for the catalytic activity of retroelement-encoded proteins, were found in the ORFs of the *PSCR* located in the *Avr*4/6 region (Fig. 3) and in the consensus *PSCR*. In the reverse transcriptase (RT) gene, the first two conserved domains in *PSCR* differ from retroelements in true fungi, and domain 3 has the YXDD motif which is highly conserved in *Copia*-like retroelements from diverse organisms including plants (Flavell et al. 1992) (Fig. 3). Downstream of the putative RT domain lies a 296 amino acid sequence with similarities to the RNaseH domains of the other *Copia*-like retroelements.

The full-length consensus nucleotide sequence of *PSCR* was used to search the genomic sequence databases of the oömycetes *P. ramorum* and *P. infestans*, and the EST databases of *P. infestans* and *P. sojae*. A series of sequences of varying homology to *PSCR* were found in each of these organisms. Most interestingly, BLASTN searches of the Syngenta *Phytophthora* Consortium database identified

three short EST clones, EST rpvb_10634 (158 nt), EST rpcd_6530 (84 nt) and EST rpvb_3051 (112 nt) from *P. infestans* that were very closely related to the *PSCR* sequence from *P. sojae*. These ESTs showed 82.9, 83.3 and 85.7% nucleotide homology respectively to the equivalent regions of the *PSCR* sequence. The deduced amino acid sequence of these ESTs aligned to the integrase central catalytic domain INT (EST rpvb_10634), RT domain (EST rpcd_6530) and RNaseH domain (EST rpvb_3051), where they showed 96.1, 89.3 and 94.4% identity respectively to *PSCR*. A BLASTN search of the *P. infestans* genomic sequence released by the Broad Institute (www.broad.mit.edu/annotation/genome/phytophthora_infestans) revealed a genomic homolog with 100% identity to all three of these ESTs in Supercontig 111 and other genomic homologs with near-100% identity in Supercontigs 44, 95 and 21. The full-length retroelement sequence was 4,948 bp in length (data not shown), and was named *PICR-1* (*P. infestans Copia*-like retrotransposon 1, EU567069). Many other genomic sequences that were homologous (but non-identical) to the three ESTs were found in the Broad Institute genomic database, suggesting that *PICR-1* represents a family of retroelements that has moved around the genome during the evolution of *P. infestans*, similarly to *PSCR* in *P. sojae*. A further BLASTN search of the *P. infestans* genomic sequence using *PSCR* as query revealed a related full-length *Copia*-like ORF of 5,040 bp in length in Supercontig 32 that we have named *PICR-2*, EU567070). *PICR-1* encoded a single long polyprotein sequence characteristic of *Copia*-like retrotransposons, whereas the deduced *PICR-2* polyprotein was interrupted by a premature stop codon suggesting that it was no longer functional. Neighbor-joining analysis using nucleotide sequences from all three of the above EST clones of *P. infestans*, indicated that *PICR-1* and *PICR-2* are about as divergent from each other as from *PSCR* (data not shown), similarly to an analysis based on deduced amino acid sequences from the conserved domain of the RT gene (Fig. 4). This suggests that *PICR-1* and *PICR-2* in *P. infestans* and *PSCR* in *P. sojae* diverged from a common group of ancestral retroelements during speciation of the *Phytophthora* complex or its progenitors. A BLASTN search of the *P. ramorum* genome using the full length nucleotide sequence of *PSCR* as query also revealed a related *Copia*-like nucleotide sequence of 4973 bp in length that we have named *PRCR*, EU567071. The *P. ramorum* nucleotide sequence was more divergent from *PSCR* than its homologs in *P. infestans* (Fig. 4), suggesting it could be derived from a lineage representing an earlier speciation event. Other sequences homologous to *PSCR* were also identified in searches of unassembled sequencing reads of the *Hyaloperonospora parasitica* (Fig. 4) and *P. capsici* (data not shown) genomes.

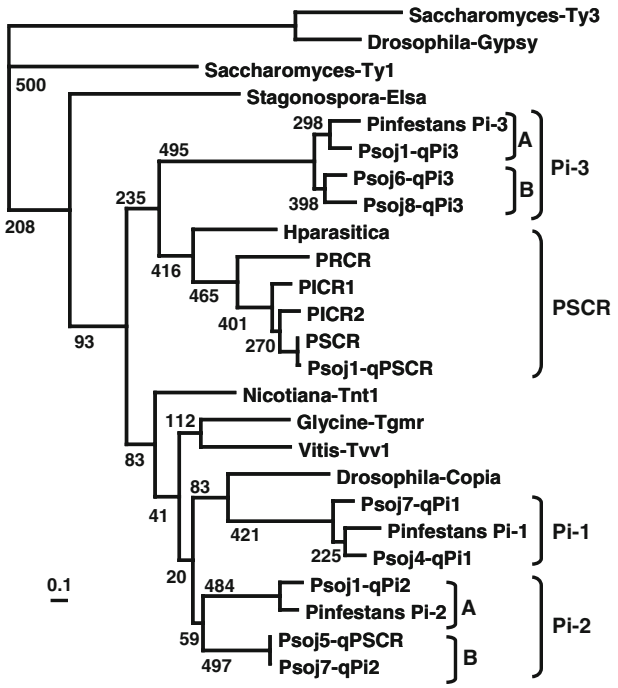

Fig. 4 Phylogram of the conserved region of the reverse transcriptase gene (Xiong and Eickbush 1990) from diverse *Copia*-like retrotransposon elements in the genome of *Phytophthora sojae* compared with various LTR-containing elements reported in other organisms, as described in the Sect. "Materials and methods." The proml, seqboot, and consense programs of PHYLIP3.67 (http://evolution.genetics.washington.edu/phylip.html) were used to generate a consensus maximum likelihood tree after 500 bootstrap iterations: *the numbers on the branches* indicate the number of times the two sets of sequences separated by the branch occurred among the 500 trees. The RT region of the following *Copia*-like sequences were retrieved from the *Phytophthora sojae* genomic database (http://genome.jgi-psf.org/Physo1_1/Physo1_1.home.html) using BLASTP searches of final predicted proteins with the query (q) sequence indicated in the name, as described in the Sect. "Materials and methods" and Sect. "Results" (protein ID number is given in parentheses): **Psoj1-qPSCR** (109011); **Psoj5-qPSCR** (145342); **Psoj4-qPi1** (145441); **Psoj7-qPi1** (140554); **Psoj1-qPi2** (109999); **Psoj7-qPi2** (132617); **Psoj1-qPi3** (132346); **Psoj6-qPi3** (140795); **Psoj8-qPi3** (135319); **PSCR** (consensus sequence described in the Sect. "Results"). *Copia*-like elements from other organisms (described more fully in the Sects. "Materials and methods" and Sect. "Results") were: **PRCR**, identified as a deduced protein in scaffold 17 of the *Phytophthora ramorum* genomic database (http://genome.jgi-psf.org/, protein ID76292); **PICR-1** and **PICR-2** were identified in supercontigs 111 (EU567069) and 32 (EU567070) respectively of *Phytophthora infestans* (http://www.broad.mit.edu/annotation/genome/phytophthora_infestans). Accessions of the following *Copia*-like elements are indicated in the Sect. "Materials and methods" or Sect. "Results": **Hparasitica**, *Hyaloperonospora parasitica*; **Ty1**, *Saccharomyces cerevisiae*; **Elsa**, *Stagonospora nodorum*; **Tnt1**, *Nicotiana tabacum*; **Tgmr**, *Glycine max*; **Tvv1**, *Vitis vinifera*; **Copia**, *Drosophila melanogaster*; **Pinfestans Pi-1**, *CopiaPi-1* (AY830098) from *P. infestans*; **Pinfestans Pi-2**, *CopiaPi-2* (AY830099) from *P. infestans*; **Pinfestans Pi-3**, *CopiaPi-3* (AY830100) from *P. infestans*; *Gypsy*-like elements were: **Gypsy**, *Drosophila melanogaster*; **Ty3**, *Saccharomyces cerevisiae*. *Scale* indicates branch length (0.1 = 10% non-identity). The alignment used to generate this phylogram is available from s.basnayake@uq.edu.au

Identification and analysis of further, more divergent *Copia*-like retroelements in the *P. sojae* genome

Jiang et al. (2005) reported three *Ty1/Copia*-like retrotransposons in *P. infestans* [AY830098 (=*CopiaPi-1*), AY830099 (=*CopiaPi-2*) and AY830100 (=*CopiaPi-3*)] that show less than 50% homology at the nucleotide level to the *PSCR* sequence. These three sequences were difficult to align at the nucleotide level, and their identity as *Copia*-like retrotransposons was confirmed by alignment of conserved domains of their translated sequences. We chose the deduced amino acid sequence of the core conserved domain of the reverse transcriptase (RT) gene (Xiong and Eickbush 1990) of *PSCR* and these three *Phytophthora* retroelements as queries to search for further, more divergent *Copia*-like elements in the genomes of *P. sojae* and *P. ramorum*. For example, using the core RT sequence of *PSCR* (representing 214 amino acid residues) as query for BLASTP searches, we found 297 hits better than $1.0e^{-05}$ in the *P. sojae* genomic database, most of which showed considerable sequence divergence from members of the *PSCR* family described above. Additional BLASTP searches using the core RT sequence of the three *Copia*-like sequences from *P. infestans* as queries revealed further large sets of *Copia*-like elements in the *P. sojae* and *P. ramorum* genomes that showed considerable sequence divergence from members of the *PSCR* family. Examples of members of the further *Copia*-like RT sequence families found in the genome of *P. sojae* are presented in Fig. 4, which depicts a PHYLIP-generated phylogram that compares their relationship to *PSCR* and other retroelements of interest.

Figure 4 shows that *PSCR* and the further *Copia*-like RT sequences found in the *P. sojae* genome fell into four clades named after the query sequences used for the database searches (PSCR, Pi-1, Pi-2 and Pi-3 in Fig. 4). Individual sequences are named after the genome in which they were found and the query sequence, e.g., *Psoj6-qPi3* in clade Pi-3 represents the sixth best hit in the *P. sojae* genomic database using the core RT sequence from *CopiaPi-3* (from *P. infestans*) as query for the BLASTP search. Some sequences presented in Fig. 4 represent BLASTP hits to large deduced ORFs (>1,000 bp) that have many close homologs in the *P. sojae* genome (e.g., *Psoj5-qPSCR* and *Psoj7-qPi2* in clade Pi-2(B); and *Psoj7-qPi1* in clade Pi-1). However, other sequences represent less abundant and smaller ORFs, some of which (e.g., *Psoj1-qPi2* in clade Pi-2(A)) required frameshift corrections to align with the full deduced amino acid sequence of the core RT domain. *Psoj4-qPi1* in clade Pi-1 also represented a subgroup of low abundance sequences with smaller ORFs.

BLASTP searches of the *P. ramorum* genomic database revealed *Copia*-like RT sequences that also fell within each of the above four clades shown in Fig. 4, e.g., *PRCR* from *P. ramorum* fell within the PSCR clade (Fig. 4; data not presented for the *P. ramorum* homologs of the other clades). Because the four clades labeled as PSCR, Pi-1, Pi-2 and Pi-3 in Fig. 4 each contained sequences found in the *P. sojae*, *P. ramorum* and *P. infestans* genomes, each of these clades represents a unique family of *Copia*-like elements common to at least these members of the genus *Phytophthora*. Some clades were split into distinct subclades, e.g., A and B in Pi-2 and Pi-3, that include sequences which appear to have evolved into coherent *Copia*-like families within *P. sojae*.

Comparison with *Copia*-like RT sequences from other organisms showed that the four Phytophthora-derived clades described above were about as different from each other as from *Copia*-like RT sequences derived from higher plants (*Nicotiana*, *Glycine*, *Vitis*) and insects (*Drosophila*) (Fig. 4). *Copia*-like sequences from true fungi (*Stagonospora* and *Saccharomyces*) were possibly more divergent than those from *Phytophthora* (Fig. 4). RT sequences from *Gypsy*-like retrotransposons (derived from *Drosophila* and *Saccharomyces*) formed a clear out-group that was very different from all of the *Copia*-like sequences shown in Fig. 4.

Database search for ESTs expressed from diverse *Copia*-like families

Copia-like elements from *P. sojae* and *P. infestans* included in Fig. 4 were used to search the EST database at NCBI to explore whether they represent *Copia*-like families that are still active. Strong hits were found only to elements from clades PSCR, Pi-2A and Pi2B of Fig. 4 (Table 4). Hence there is no evidence that elements from the other families represented in clades Pi-1 and Pi-3 actively express transcripts in either *P. sojae* or *P. infestans*, and these elements are provisionally assigned as "inactive." This agrees with the database search by Jiang et al. (2005) who found a strong EST hit to *P. infestans* using *CopiaPi-2* but no hits to *CopiaPi-1* or *CopiaPi-3* as queries.

The active elements from clades PSCR, Pi-2A and Pi-2B appear to be species-specific as indicated in Table 4, and ESTs gave strong (but not perfect) matches indicating they were expressed from genomic elements closely related to the query sequence. Within clade PSCR, only *PICR-1* from *P. infestans* gave evidence of transcriptional activity; within clade Pi-2A, only *CopiaPi-2* from *P. infestans* (=Pinfestans Pi-2 in Fig. 4) gave evidence of activity; and within clade Pi-2B, only elements from *P. sojae* gave evidence of transcriptional activity. The five ESTs from *P. infestans* hit by *PICR-1* (Table 4) are additional to the three small ESTs from the Syngenta *Phytophthora* Consortium database reported previously.

Table 4 Search of NCBI database for ESTs homologous to the *Copia*-like elements from *P. sojae* and *P. infestans* shown in Fig. 4

Clade	Query sequence	Source of query and EST	ESTs hit by query (GenBank accession number)[c]	
			Strong hits	Moderate hits
			$\leq e-95$ (or $\geq$95% I)	$\leq e-60$ (or $\geq$60% I)
PSCR	*PSCR*	*P. sojae*	–	
	PICR-1	***P. infestans***	CV946731.1, CV949879.1, CV909146.1, CV947462.1, CV952596.1	
	PICR-2	*P. infestans*	–	
Pi-1	*CopiaPi-1*	*P. infestans*	–	
	Psoj4-qPi1[a]	*P. sojae*	–	GE632036.1
Pi-2A	***CopiaPi-2***	***P. infestans***	CV897306.1	
	Psoj1-qPi2	*P. sojae*	–	
Pi-2B	***Psoj7-qPi2***	***P. sojae***	GE632064.1, CF840723.1	
	Psoj8-qPSCR[b]	***P. sojae***	GE632001.1	
Pi-3A	*CopiaPi-3*	*P. infestans*	–	BE776643.1, CV911344.1, CV924283.1, CV907774.1
	Psoj1-qPi3	*P. sojae*	–	CF862909.1
Pi-3B	*Psoj6-qPi3*	*P. sojae*	–	CF864888.1
Pi-3B	*Psoj8-qPi3*	*P. sojae*	–	CF841137.1

ESTs from other oömycetes showing moderately strong hits to queries from *P. sojae* and *P. infestans*

Clade	Query	Source of EST	$\leq e-80$ (or $\geq$80% I)
PSCR	*PSCR*	*Phytophthora capsici*	FG017534.1, FG041977.1
Pi-2B	*Psoj7-qPi2*	*Pythium oligandrum*	EV243817.1, EV243508.1, EV243713.1, EL738132.1
Pi-3A	*CopiaPi-3*	*Phytophthora brassicae*	ES288687.1, ES290568.1
Pi-3B	*Psoj6-qPi3*	*Phytophthora brassicae*	ES287194.1

Query sequences that showed strong hits to ESTs are indicated in bold

e Denotes expectation value of best ORF fragment in EST found by query, *I* denotes amino acid identity of best ORF fragment in EST found by query, – denotes query found no strong hits

[a] *Psoj7-qPi1* from clade Pi-1 as query gave similar results to *Psoj4-qPi1*

[b] Psoj8-qPSCR is another member of Pi-2B subfamily in *P. sojae* that has a longer deduced ORF than Psoj5-qPSCR and Psoj7-qPi2 shown in Fig. 4, and hence was more likely to find ESTs from the other end of transcripts

[c] TBLASTX or TBLASTN searches used the full nucleotide or deduced ORF sequence of each element (as found by the genomic analysis) to identify protein homologies. When an EST accession was hit by multiple queries, the query showing highest homology is listed in the table

Table 4 shows many hits of moderate homology to query sequences, representing ESTs expressed from *Copia*-like families that show some degree of divergence from the query and hence constitute distinct but related families of elements. For example, although the families represented in clade Pi-3 of Fig. 4 appear to be inactive, hits of "moderate" homology were found to ESTs from both *P. infestans* and *P. sojae*, indicating the presence of further *Copia*-like families of elements related to Pi-3 that actively express transcripts in these species. Among other oömycetes, transcriptionally-active families of elements were identified by ESTs that are related to clade Pi-3 (*Phytophthora brassicae*), clade Pi-2B (*Pythium oligandrum*) and clade PSCR (*Phytophthora brassicae*) (Table 4). Other hits of lower homology than shown in Table 4 were found to ESTs from *P. sojae*, *P. infestans* and other oömycetes (data not shown), indicating that further *Copia*-like families with greater degrees of divergence to those noted above are still active in the genomes of these species.

Discussion

Here we have described a new low copy number family of retrotransposons, PSCR, which is represented by about 12 near-full-length copies in the genome of the oömycete,

P. sojae. *PSCR* is flanked by a LTR at each end, and has a general internal organization similar to that found in other *Copia*-like retroelements that distinguishes it from *Gypsy*-like retroelements which also have a pair of LTRs. Despite a relatively low degree of nucleotide sequence similarity between *PSCR* and the *Copia*-like retroelements from evolutionary diverse organisms, there is relatively high amino acid sequence homology in conserved domains of these elements. Before the advent of genome sequencing projects, most studies of retrotransposon phylogeny were based on sequence comparison of PCR products corresponding to highly conserved RT domains (Doolittle et al. 1989; Tooley et al. 1996; Judelson 2002), and further comparisons can now be made with the larger set of *Copia*-like sequences found by database searches of genome sequences. In the current project, database analysis revealed a greater amino acid identity of the *PSCR* sequence with the cognate regions corresponding to the RT (Fig. 4), RNaseH and other conserved domains (data not shown) of other *Copia*-like retrotransposons, than to the cognate regions of other groups of retroelements such as *Gypsy*-like elements. Indeed, the level of similarity shown by *PSCR* sequences to conserved domains of the putative polyproteins of the *Copia*-like retrotransposons in other oömycetes and organisms, strongly supports the assignment of *PSCR* to the group of *Copia*-like retrotransposons. A large portion of the RT domain (representing 214 amino acid residues in *PSCR*) was chosen for further analysis of homologs of *PSCR* because it contains a mixture of highly and less conserved regions that readily distinguish *Copia*-like retroelements from other groups of retroelements including *Gypsy*-like retrotransposons (Xiong and Eickbush 1990). For example, Fig. 4 shows that *Gypsy* from *D. melanogaster* and *Ty3* from *Saccharomyces cerevisiae* are very different from a selection of *Copia*-like retroelements from diverse organisms. Hence diverse genomic homologs that align in the RT domain with previously characterized *Copia*-like sequences can confidently be assigned to this group of retroelements.

Relatively close homologs of the *PSCR* sequence were found in the databases of other oömycetes whose genomes have been sequenced to date, with the closest being two distinct homologs (*PICR-1* and *PICR-2*) found in *P. infestans* followed by a homolog (*PRCR*) in *P. ramorum* and a homolog in *H. parasitica* (Fig. 4); together these elements represent a "PSCR-family" of *Copia*-like retrotransposons that may well be specific to the oömycetes. For example, using the core conserved domain of the RT gene (Xiong and Eickbush 1990) for comparison, members of the PSCR family showed much greater homology to each other than to *Copia*-like sequences found in other groups of organisms represented by higher plants, insects, fungi and other oömycetes (Fig. 4). Fungal *Copia*-like retroelements represented by *Ty1* in *S. cerevisiae* and *Elsa* in *Stagonospora nodorum* were the most divergent *Copia*-like lineages to the PSCR family.

Among the oömycetes, the three *Copia*-like elements previously reported in *P. infestans* by Jiang et al. (2005), *CopiaPi-1*, *CopiaPi-2* and *CopiaPi-3*, were about as divergent from the PSCR family as *Copia*-like elements reported in higher plants and insects (Fig. 4). Multiple homologs of *CopiaPi-1*, *CopiaPi-2* and *CopiaPi-3* were found in the *P. sojae* genomic database and contribute to further *Copia*-like families shown in clades Pi-1, Pi-2 and Pi-3 in Fig. 4; Pi-2 is split into two deeply-divided lineages A and B. The number of copies of each family was not investigated exhaustively. High bootstrap values for branches representing each of these families (Fig. 4) suggest that members within each clade represent distinct, independently-evolving lineages. Low bootstrap values among branches after Maximum Likelihood analysis (Fig. 4) suggest that four of these *Copia*-like clades, PSCR, Pi-1, Pi-2A and Pi2B, represent ancient lineages that are approximately as divergent from each other as from *Copia*-like families represented in the higher plants and insects. Tooley and Garfinkel (1996) reported a high degree of divergence among *Copia*-like elements found in *P. infestans* using PCR. The Pi-3 family possibly shows somewhat greater homology to the PSCR family than to the other families, but this requires confirmation by analysis of other domains within the general *Ty1/Copia* structure. Orthologous members of each of the *Copia*-like families present in the genomes of *P. sojae* and *P. infestans* as listed above, were also found in the genomic sequence of *P. ramorum* by using BLASTP searches (data not shown). The BLASTP search of the *P. ramorum* genome also revealed a further distinct family of *Copia*-like sequences, that was used to find an equivalent *Copia*-like family in the *P. sojae* genome that had not been identified previously (data not shown). Searches of EST databases confirm that *P. infestans* and *P. sojae* host further, divergent *Copia*-like families that were not identified by the BLASTP screens of their genomes used in this investigation.

Consistent with the above observations it is suggested that after the divergence of *Copia*-like elements from other groups of retroelements such as the *Ty3/Gypsy* group (Xiong and Eickbush 1990), fungal *Copia*-like elements first branched off into separate lineages, and another lineage subsequently gave rise to a common ancestor of the different families of *Copia*-like elements currently found in the oömycetes, higher plants and insects. Certainly the *Copia*-like families found within each of the clades displayed in Fig. 4 appear to have entered an ancestral oömycete genome prior to speciation into *P. infestans*, *P. sojae* and *P. ramorum*, and further analysis of other conserved domains in *Copia*-like elements in the genomes of other species may assist the inference of speciation events among

the *Copia*-like elements present in current oömycetes. Similarly, Jiang and Govers (2006) found that most of the *Copia*- and *Gypsy*-like retrotransposons that they identified in *P. infestans* had homologs in both the *P. sojae* and *P. ramorum* genomes.

Internal sequence heterogeneity of homologs of *PSCR* in the *P. sojae* genome

The relatively high nucleotide homologies (Table 2) and conserved genomic locations of *PSCR* homologs in isolates from diverse genetic backgrounds (Fig. 2), suggest that an ancestral retroelement multiplied during a period of active transposition prior to the diversification of *P. sojae* into current genotypes isolated from the field. The reverse transcriptase from *Copia*-like retroelements in other organisms is known to lack proof-reading activity during replication (Bhattacharyya et al. 1997), and such an error-prone transposition mechanism by *PSCR* may have contributed to the divergence observed among the copies found in the *P. sojae* genome (Table 2). In the initial *PSCR* sequence (12A) found in the *Avr*4/6 region of the *P. sojae* genome, point mutations and deletions in the integrase co-domain region caused premature termination (Fig. 1; Table 2). Most of the other homologs also have at least point mutations in this region whereas the homolog in scaffold 19, for example, showed a major rearrangement (Table 2). Scaffold 12 contains two homologs of *PSCR* where one (12B in Table 2) shows a complex rearrangement causing duplication in the polyprotein region. Internal sequence heterogeneity can also arise from recombination between or within copies of a retrotransposon and its host genome, which may result in large deletions, insertions, duplications and inversions (Table 2) (Vershinin and Ellis 1999). The presence of various rearrangements and multiple point mutations among the near-full-length copies of *PSCR* (Table 2), and the lack of *PSCR* sequences in EST databases of *P. sojae*, is consistent with it being transcriptionally silent or expressed at low levels in current populations of this organism. This suggestion is consistent with the near-uniform restriction digestion patterns observed in Fig. 2, which provide no evidence for new transpositional events during evolution of the genetic diversity observed in current races of this pathogen (Förster et al. 1994; Drenth et al. 1996). Furthermore, major differences in the untranslated leader region resulting from recombination or mutation could have an effect on the retrotransposition capacity of these elements. Although none of the *PSCR* homologs present in the genome encodes a full-length functional polyprotein, the consensus sequence of *PSCR* assembled from these homologs encodes a single, long polyprotein, characteristic of functional *Copia*-like retrotransposons in other organisms.

Phytophthora sojae Copia-like retrotransposon was discovered when sequencing the genomic region containing the *Avr*4/6 locus. This region also contains other repeated sequences with similarities to retroelements (S. Basnayake, unpublished). For example, the genetic marker cDNA71, which flanks a 24-kb region close to the *Avr*4/6 locus (Whisson et al. 2004), shows homology to integrases. Because none of the copies of *PSCR* contain a single ORF encoding a continuous polyprotein, it appears that none of the homologs found to date in the current *P. sojae* genome represents the initial *PSCR*-like sequence that transposed and multiplied during the speciation of *P. sojae*. However some of the mutations noted in Table 2 are shared by subsets of these homologs, and it is possible that imperfect homologs may collectively express all of the genes required for transposition, and hence by trans-complementation enable successful movement of imperfect homologs (Kalendar et al. 2004). Current populations of a related species, *P. infestans*, harbor another member of the *PSCR* family, *PICR-1*, that actively expresses transcripts. *PICR-1* is represented in the *P. infestans* genome by intact, full-length copies with a single long polyprotein-encoding ORF, and thus appears to be functional. Hence it is feasible that *PSCR* has lost function during the speciation of *P. sojae* from a common ancestor with *P. infestans*.

Origin and mode of transfer of *PSCR*

Several phylogenic studies of retrotransposons based on their conserved RT domains have revealed that the distribution of closely related retrotransposons does not always follow the phylogenic relationship of their host species (Flavell et al. 1997), supporting the notion that *Copia*-like LTR-retrotransposons may have entered ancestral oömycetes by horizontal transfer. As well as suggesting that apparently unique *Gypsy*-like LTR-retrotransposon lineages may have entered particular *Phytophthora* species by recent horizontal movement, Judelson (2002) observed that other *Gypsy*-like elements in the genus *Phytophthora* clearly exist as ancient distinct lineages that evolved by vertical radiation. Supporting the notion of vertical radiation of *Copia*-like elements within *Phytophthora*, moderately homologous sequences to *PSCR* in *P. sojae* have been found in related species such as *P. infestans* and *P. ramorum*. Because *PICR-1* and *PICR-2* in the genome of *P. infestans* together with *PRCR* from *P. ramorum* and the *PSCR* homolog from *H. parasitica* formed a distinct clade in Fig. 4, it is possible that this *PSCR*-related group of *Copia*-like elements may have a common horizontal entry point into an ancestral oömycete that predates the separation of oömycetes into current species of *Phytophthora*. Furthermore, this investigation has demonstrated that a series of other, more divergent *Copia*-like families related to

elements first identified in *P. infestans*, are each represented in other *Phytophthora* species such as *P. sojae* and *P. ramorum* and other oömycetes such as *P. brassicae*, *P. capsici* and *Pythium oligandrum*. These observations are consistent with horizontal transfer of *Copia*-like retroelements into oömycetes being mostly ancient and rare rather than from recent events. The simplest hypothesis is that many or most of the divergent families of retroelements present in current species of *Phytophthora* entered ancient ancestors of oömycetes and other eukaryotic organisms from a common pool, and their divergent progeny are represented in current populations of higher plants, insects and oömycetes. Genome-wide screening of other more divergent oömycetes is necessary to test this hypothesis. Alternatively, an environmental source that cohabited with speciating oömycetes may have harbored diverging ancestors of the PSCR family, and may have allowed multiple entry points of this PSCR-related group of *Copia*-like elements to the *Phytophthora* genus during its speciation.

Biased codon usage has been suggested as evidence for recent horizontal transfer of retroelements into the genomes of eukaryotes (Nakayashiki et al. 1999; Springer et al. 1995). It has recently been shown that the third position bases in codons from *Phytophthora* ORFs are dominated by G and C to give a higher GC% for the third position (GC3) than the overall GC content in *Phytophthora* ORFs (Qutob et al. 2000; Huitema et al. 2003; Jiang et al. 2005, Jiang and Govers 2006), and that differences in GC3 among ORFs indicate differences in codon bias (Jiang and Govers 2006). The ORF encoding the 1,492 amino acid polyprotein deduced from the *PSCR* consensus sequence showed appreciably lower values for GC3 and overall GC content (61 and 54% respectively, a difference of 7%) than 10,000 ORFs found in the *P. sojae* genomic sequence database (76 and 60%, a difference of 16%, Jiang and Govers 2006). This suggests that codon bias does exist in *PSCR* but is less than in most genes that encode housekeeping proteins in the *P. sojae* genome. Jiang and Govers (2006) compared GC3 and other measures of codon bias in a number of retrotransposons in the genomes of *P. sojae*, *P. ramorum* and *P. infestans*, and found high GC3 and similar codon usage to host genes only in some high copy number retroelements such as two *Gypsy* families in *P. infestans*, suggesting that such retroelements are subject to the same selection pressures contributing codon bias as other host genes. Contrary to the *Gypsy* retroelements, Jiang and Govers (2006) found that members of a *Copia*-like family in *P. sojae* homologous to *CopiaPi-2* (AY830099) from *P. infestans* (cf. clade Pi-2A in Fig. 4) showed no evidence of high GC3 or codon bias compared to host genes. However, another *Copia*-like family in *P. ramorum* showed a high GC3 content of 80%. The values for GC3 in *PSCR* and other measures such as RSCU (relative synonymous codon usage) for individual codons (data not shown) indicate that *PSCR* may differ in codon usage to host genes, but it is difficult to quantify the effect of host selection pressures leading to high GC3. Hence information on the magnitude of the bias is unlikely to assist us to deduce the time line of introgression of the ancestors of *PSCR* into the *P. sojae* genome. Although independent horizontal transfer events could have contributed to the diverse *Copia*-like families including *PSCR* observed in current populations of *P. sojae* and other *oömycetes*, we suggest that a detailed examination of vestigial genomic fragments related to the *Copia*-like families identified by this investigation is more likely to provide useful information on the origin of *Copia*-like retroelements among the oömycetes, as discussed below.

Consistent with the hypothesis that the introgression of *Copia*-like elements is ancient and rare, database searches of the *P. sojae* genome using the conserved core RT domain as query found very few intact hits for some *Copia*-like families such as those represented by *Psoj1-qPi2* (clade 2A) and to a lesser extent *Psoj4-qPi1* (clade Pi-1) (Fig. 4). Database searches found no ESTs homologous to members of clades 2A and Pi-1 from *P. sojae* (Table 4), consistent with loss of transpositional activity and subsequent mutational degradation of the retroelement families they represent in this species. The PSCR clade (Fig. 4) is particularly interesting as homologs of variable copy number and intactness were found in the three *Phytophthora* species investigated. The *P. infestans* genome included two PSCR families, *PICR-1* and *PICR-2*; the former being represented by multiple copies including members with a full polyprotein ORF that was expressed as transcripts detected via cDNA libraries. The *P. sojae* genome harbored 12 near-full-length copies of the PSCR family, only four of which were detected among the best 10 hits using the conserved RT domain as query for BLASTP searches, and other evidence presented herein suggests that this family is no longer functional in *P. sojae*. In the *P. ramorum* genome we found only one member of the PSCR clade, *PRCR*, and this member required frameshift corrections to fit the *PRCR* ORF to the full conserved RT domain of *PSCR* used as the search query. Jiang et al. (2005) were unable to find homologs of *CopiaPi-2* (from *P. infestans*) in the genomic and EST databases of *P. sojae*, presumably because such homologs, e.g., *Psoj1-qPi2* found in the current study, have fragmented ORFs that have long been inactive in *P. sojae*. We conclude that some *Copia*-like families such as those belonging to the PSCR, Pi-1 and Pi-2 clades in *P. sojae*, appear to be losing function and are in the process of being eroded into smaller ORFs and fragments in the genome. Future studies may need to detect such fragments at multiple domains across the *Copia* genome in order to trace their evolution among the oömycetes. Replication and divergence of an ancestral *Copia* into the different clades

observed in current *Phytophthora* species would be expected to result in the accumulation of phylogenetically intermediate sequences within the genome, and the putative intermediates might not be detected readily by BLASTP searches due to loss of function and erosion during evolution. Conversely, other *Copia*-like families are likely to include members that retain transpositional activity, such as the families found to exist as multiple copies with long deduced ORFs in the genome and are strongly homologous to expressed transcripts found in EST databases. Examples of families which contain currently active members are represented by *PICR-1* in *P. infestans* (clade PSCR); *Psoj5-qPSCR* and *Psoj7-qPi2* in *P. sojae* (clade 2B); and *CopiaPi-2* in *P. infestans* (clade Pi-2A). *Psoj7-qPi1* from clade Pi-1 is represented by multiple homologs in the *P. sojae* genome but no transcripts of this family have yet been detected in current EST databases.

Do *Copia*-like elements play a role in mutation to virulence in *P. sojae*?

From searches of the JGI *P. sojae* database and Southern hybridization to genomic DNA as noted above, we present evidence that a series of genetically diverse races of *P. sojae* that differ in virulence on soybean all appear to possess the same set of *PSCR* homologs inserted at the same loci in the genome. This suggests that the PSCR family has not been active at transposition in the genome of *P. sojae* during recent evolution leading to the genetic diversity observed among current races isolated from soybean. Analysis of mutational events near the *Avr*4/6 region of the *P. sojae* genome confirms this conclusion. The *PSCR* homolog close to the *Avr*4/6 locus of an avirulent accession (*PSCR*-12A in UQ1200) showed a single nucleotide substitution between the 228 bp LTR sequences at each end. Furthermore, comparison of *PSCR*-12A in UQ1200 with the 3′end of its allelic equivalent in a genetically-diverse, virulent accession (race UQ2990), revealed four nucleotide substitutions (three transitions and one transversion) one of which was located in the 3′-LTR, indicating that the insertional event was unlikely to have occurred during recent agriculture when new races have arisen during the deployment of resistant cultivars of soybean. Hence we conclude that the mutations which generated virulence alleles at the *Avr*4/6 locus can not be attributed to transposable activity of the *PSCR* homolog (*PSCR*-12A) close to this locus. Further research is necessary to determine whether the other families of *Copia*-like elements identified by this study contain members that have caused mutation at other avirulence loci. Other *Copia*-like families identified by this study appear to contain transcriptionally active members with the potential to cause further mutation to virulence by transposition in current populations of *P. sojae*.

Acknowledgments This research was funded by the Australian Research Council and the Cooperative Research Centre for Tropical Plant Protection. Stephen C. Whisson is funded by the Scottish Government Rural and Environment Research and Analysis Directorate (RERAD).

References

Bhattacharyya MK, Gonzales RA, Kraft M, Buzzell RI (1997) A *Copia*-like retrotransposon Tgm*r* closely linked to the *Rps1*-k allele that confers race-specific resistance of soybean to *Phytophthora sojae*. Plant Mol Biol 34:255–264

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31:3497–3500

Clare J, Farabaugh P (1985) Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc Natl Acad Sci USA 82:2829–2833

Doolittle RF, Feng DF, Johnson MS, McClure MA (1989) Origins and the evolutionary relationships of retroviruses. Q Rev Biol 64:1–30

Dorrance AE, Schmitthenner AF (2000) New sources of resistance to *Phytophthora sojae* in the soybean plant introductions. Plant Dis 84:1303–1308

Drenth A, Whisson SC, Maclean DJ, Irwin JAG, Obst NR, Ryley MJ (1996) The evolution of races of *Phytophthora sojae* in Australia. Phytopathology 86:163–169

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) *Ty1-Copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 20:3639–3644

Flavell AJ, Pearce SR, Heslop-Harrison JS, Kumar A (1997) The evolution of *Ty1-Copia* group retrotransposons in eukaryote genomes. Genetica 100:185–195

Förster H, Tyler BM, Coffey MD (1994) *Phytophthora sojae* races have arisen by clonal evolution and rare outcrosses. Mol Plant Microbe Interact 7:780–791

Grandbastien MA (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci 3:181–187

Grandbastien MA, Spielmann A, Caboche M (1989) Tnt1: a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337:376–380

Huitema E, Vleeshouwers VGAA, Francis DM, Kamoun S (2003) Active defence responses associated with non-host resistance of *Arabidopsis thaliana* to the oömycete pathogen *Phytophthora infestans*. Mol Plant Pathol 4:487–500

Jiang RHY, Govers F (2006) Nonneutral GC3 and retroelement codon mimicry in *Phytophthora*. J Mol Evol 63:458–472

Jiang RHY, Dawe AL, Weide R, van Staveren M, Peters S, Nuss DL, Govers F (2005) Elicitin genes in *Phytophthora infestans* are clustered and interspersed with various transposon-like elements. Mol Genet Genomics 273:20–32

Judelson HS (2002) Sequence variation and genomic amplification of a family of *Gypsy*-like elements in the oömycete genus *Phytophthora*. Mol Biol Evol 19:1313–1322

Kalendar R, Vicient CM, Peleg O, Anamthawat-Jonsson K, Bolshoy AH, Schulman AH (2004) Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes. Genetics 166:1437–1450

Katoh IT, Yasunaga T, Ikawa Y, Yoshinaga Y (1987) Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor. Nature 329:654–656

Kikuchi Y, Ando Y, Shiba T (1986) Unusual priming mechanism of RNA-directed DNA synthesis in *Copia* retrovirus-like particles of *Drosophila*. Nature 323:824–826

Leong SA, Farman ML, Smith J, Budde A, Tosa Y, Nitta N (1994) Molecular genetic approach to the study of cultivar specificity in the rice blast fungus. In: Zeigler RS, Leong SA, Teng PS (eds) Rice blast disease. CAB International, Wallingford, pp 87–110

May KJ, Whisson SC, Zwart RS, Searle IR, Irwin JAG, Maclean DJ, Carrol BJ, Drenth A (2002) Inheritance and mapping of eleven avirulence genes in *Phytophthora sojae*. Fungal Genet Biol 37:1–12

McHale MT, Roberts IN, Noble SM, Beaumont C, Whitehead MP, Seth D, Oliver RP (1992) CfT-1: an LTR-retrotransposon in *Cladosporium fulvum*, a fungal pathogen of tomato. Mol Gen Genet 233:337–347

Mount SM, Rubin GM (1985) Complete nucleotide sequence of the *Drosophila* transposable element *Copia*: homology between *Copia* and retroviral proteins. Mol Cell Biol 5:1630–1638

Nakayashiki H, Kiyotomi K, Tosa Y, Mayama S (1999) Transposition of the retrotransposon Maggy in heterologous species of filamentous fungi. Genetics 153:693–703

Nuchprayoon I, Meyers S, Scott L, Suzow J, Hiebert S, Friedman A (1994) Pebp2/Cbf, the murine homolog of the human myeloid Aml1 and Pebp2-Beta/Cbf-Beta Proto-Oncoproteins, regulates the murine myeloperoxidase and neutrophil elastase genes in immature myeloid cells. Mol Cell Biol 14:5558–5568

Pelsy F, Merdinoglu D (2002) Complete sequence of Tvv1, a family of *Ty1 Copia*-like retrotransposons of *Vitis vinifera* L., reconstituted by chromosome walking. Theor Appl Genet 105:614–621

Qutob D, Hraber PT, Sobral BWS, Gijzen M (2000) Comparative analysis of expressed sequences in *Phytophthora sojae*. Plant Physiol 123:243–253

Rawson JM (2000) Transposable elements in the phytopathogenic fungus *Stagonospora nodorum*. Ph.D. thesis, School of Biosciences, The University of Birmingham, Edgbaston

Ryley MJ, Obst NR, Irwin JAG, Drenth A (1998) Changes in the racial composition *Phytophthora sojae* in Australia between 1979 and 1996. Plant Dis 82:1048–1054

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Shaw DS (1988) The *Phytophthora* Species. In: Sidhu GS (ed) Advances in plant pathology: genetics of plant pathogenic fungi, vol 6. Academic, London, pp 27–51

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Springer MS, Tusneem NA, Davidson EH, Britten RJ (1995) Phylogeny, rates of evolution, and patterns of codon usage among Sea-Urchin retroviral-like elements, with implications for the recognition of horizontal transfer. Mol Biol Evol 12:219–230

Tooley PW, Garfinkel DJ (1996) Presence of *Ty1-Copia* group retrotransposon sequences in the potato late blight pathogen *Phytophthora infestans*. Mol Plant Microbe Interact 9:305–309

Tooley PW, Carras MM, Falkenstein KF (1996) Relationships among group IV *Phytophthora* species inferred by restriction analysis of the ITS2 region. J Phytopathol 144:363–369

Tripathy S, Pandey VN, Fang B, Salas F, Tyler BM (2006) VMD: a community annotation database for microbial genomes. Nucleic Acids Res 34:379–381

Tyler BM, Förster H, Coffey MD (1995) Inheritance of avirulence factors and RFLP markers in outcrosses of the oömycete *Phytophthora sojae*. Mol Plant Microbe Interact 8:515–523

Varmus H, Brown P (1989) Retroviruses. In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington DC, pp 53–108

Vershinin AV, Ellis THN (1999) Heterogeneity of the internal structure of PDR1, a family of *Ty1*/*Copia*-like retrotransposons. Mol Gen Genet 262:703–713

Whisson SC, Drenth A, Maclean DJ, Irwin JAG (1994) Evidence for outcrossing in *Phytophthora sojae* and linkage of a DNA marker to two avirulence genes. Curr Genet 27:77–82

Whisson SC, Drenth A, Maclean DJ, Irwin JAG (1995) *Phytophthora sojae* avirulence genes, RAPD and RFLP markers used to construct a detailed genetic linkage map. Mol Plant Microbe Interact 8:988–995

Whisson SC, Basnayake S, Maclean DJ, Irwin JAG, Drenth A (2004) *Phytophthora sojae* avirulence genes *Avr4* and *Avr6* are located in a 24 kb, recombination-rich region of genomic DNA. Fungal Genet Biol 41:62–74

Wilhelm M, Wilhelm FX (2001) Reverse transcription of retroviruses and LTR retrotransposons. Cell Mol Life Sci 58:1246–1262

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based on their reverse transcriptase sequences. EMBO J 9:3353–3362

Zhang X, Scheuring C, Tripathy S, Xu Z, Wu C, Ko A, Ken Tian S, Arredondo F, Lee MK, Santos FA, Jiang RHY, Zhang HB, Tyler BM (2006) An Integrated BAC and genome sequence physical map of *Phytophthora sojae*. Mol Plant Microbe Interact 19:1302–1310