Available online at www.sciencedirect.com
SCIENCE DIRECT®
Fungal Genetics and Biology 42 (2005) 351–365
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

Families of short interspersed elements in the genome of the oomycete plant pathogen, *Phytophthora infestans*☆

Stephen C. Whisson[a,*], Anna O. Avrova[a], Olga Lavrova[b], Leighton Pritchard[a]

[a] *Plant-Pathogen Interactions Program, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK*
[b] *Moscow Lomonosov State University, 1-12 Vorob'evy gory, Moscow 119992 GSP-2, Russian Federation*

Received 1 July 2004; accepted 6 January 2005

Abstract

The first known families of tRNA-related short interspersed elements (SINEs) in the oomycetes were identified by exploiting the genomic DNA sequence resources for the potato late blight pathogen, *Phytophthora infestans*. Fifteen families of tRNA-related SINEs, as well as predicted tRNAs, and other possible RNA polymerase III-transcribed sequences were identified. The size of individual elements ranges from 101 to 392 bp, representing sequences present from low (1) to highly abundant (over 2000) copy number in the *P. infestans* genome, based on quantitative PCR analysis. Putative short direct repeat sequences (6–14 bp) flanking the elements were also identified for eight of the SINEs. Predicted SINEs were named in a series prefixed *infSINE* (for *infestans*-SINE). Two SINEs were apparently present as multimers of tRNA-related units; four copies of a related unit for *infSINEr*, and two unrelated units for *infSINEz*. Two SINEs, *infSINEh* and *infSINEi*, were typically located within 400 bp of each other. These were also the only two elements identified as being actively transcribed in the mycelial stage of *P. infestans* by RT-PCR. It is possible that *infSINEh* and *infSINEi* represent active retrotransposons in *P. infestans*. Based on the quantitative PCR estimates of copy number for all of the elements identified, tRNA-related SINEs were estimated to comprise 0.3% of the 250 Mb *P. infestans* genome. *InfSINE*-related sequences were found to occur in species throughout the genus *Phytophthora*. However, seven elements were shown to be exclusive to *P. infestans*.

Keywords: *Phytophthora infestans*; Late blight; Oomycete; SINE; Retrotransposon

1. Introduction

All eukaryotic genomes contain repetitive DNA sequences; many of these are derived from transposable elements, which shape the genomic landscape during the organism's evolutionary history. Retrotransposons are the most abundant class of these mobile elements to be found in eukaryotic genomes, relocating via reverse transcription and reinsertion, thereby increasing their copy number. These elements may be subdivided into classes: long terminal repeats (LTR), non-LTR (or long interspersed elements; LINEs), and short interspersed elements (SINEs) (Deininger et al., 2003; Kumar and Bennetzen, 1999). LTR and LINE retrotransposons encode reverse transcriptase and endonucleases for retroposition. SINEs do not possess any discernible open reading frames and are thought to be reliant on the activity of LINEs for retroposition (Deininger et al., 2003; Weiner, 2002). SINEs also differ from other retrotransposable elements in that they possess an internal RNA polymerase III promoter for transcription. Most SINEs have been shown to exhibit properties similar to RNA polymerase III-transcribed genes. In the human genome for example, the most abundant SINE is the *Alu*

☆ DNA sequences described in this paper have been deposited in GenBank as Accession Nos. AY623609–AY623621 and AY848952–AY848954.

* Corresponding author. Fax: +44 1382 562426.
E-mail address: swhiss@scri.sari.ac.uk (S.C. Whisson).

1087-1845/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2005.01.004

family of repeats which resemble the 7SL RNA, part of the signal recognition particle (Ullu and Tschudi, 1984). However, the most frequently encountered SINEs in many other organisms are related to tRNAs. Structurally, these elements possess both internal promoter A and B boxes for RNA polymerase III, a cloverleaf tRNA secondary structure at the 5′ end, and an adenosine-rich 3′ tail of variable length that finishes at an oligothymidine tract.

SINEs have been identified in the genomes of eukaryotes as diverse as fungi, fish, plants, mammals, reptiles, and insects (Borodulina and Kramerov, 1999; Feschotte et al., 2001; Kachroo et al., 1995; Myouga et al., 2001; Ogiwara et al., 2002; Schmitz and Zischler, 2003). Copy numbers for the diverse SINE families vary enormously from species to species; *Mg-SINE* in the rice blast fungus, *Magnaporthe grisea*, is present in approximately 100 copies (Kachroo et al., 1995), but in tobacco the SINE TS is present in over 50,000 copies per haploid genome (Yoshioka et al., 1993). In the human genome, the *Alu* SINE family is present in over a million copies and comprises 10% of the genome (International Human Genome Sequencing Consortium, 2001).

SINEs and other transposons have frequently been described as 'junk' DNA or genomic parasites without a functional role. However, recent reports suggest these elements may play a major role in the genomic evolution of many organisms (Makalowski, 2003). For example, insertional inactivation of genes may occur when a SINE inserts either into the coding region of a gene or an intron (leading to RNA splicing abnormalities). SINEs may also promote genomic rearrangements, or stimulate translation under conditions of cellular stress (Deininger et al., 2003; Rubin et al., 2002). These observations imply that some SINEs may be under stabilizing selection pressure and have a functional role within the cell.

Phytophthora infestans is the causative agent of potato and tomato late blight, a devastating disease of crops worldwide, and the cause of the Irish potato famine in the 1840s. *P. infestans* belongs to the oomycetes, a class of fungus-like organisms that are more closely related to the heterokont brown algae (Stramenopiles) than the true fungi (Baldauf et al., 2000; Sogin and Silberman, 1998). Compared with the true fungi, plants, and animals, the oomycetes have been relatively unstudied until the recent development of molecular biological tools and resources for them. *P. infestans* is one of the most intensively studied oomycete species, and genomics resources have been developed in the form of large expressed sequence tag (EST) datasets (Kamoun et al., 1999; Lam, 2001; Waugh et al., 2000), partial genomic DNA sequencing (Garelik, 2002), large insert DNA libraries (Randall and Judelson, 1999; Whisson et al., 2001), and molecular genetic linkage maps (van der Lee et al., 1997, 2001). *P. infestans* is notorious for its ability to adapt rapidly to overcome resistance genes deployed in its host plants, and to develop resistance to control chemicals (reviewed in: Duncan, 1999; Fry and Goodwin, 1997; Shattock, 2002); these observations are suggestive of a highly flexible and adaptable genome. The genome of *P. infestans* is estimated to be 250 Mb, large compared with the other oomycete species characterized to date (Tooley and Therrein, 1987; Voglmayr and Greilhuber, 1998). Over 30 different repetitive DNA families, representing more than 50% of the genome, have been reported from *P. infestans*, including those related to the *Gypsy* and *Copia* families of retrotransposons (Judelson, 2002; Judelson and Randall, 1998; Tooley and Garfinkel, 1996). Retrotransposons in *P. infestans*, if active, may in part explain the extraordinary adaptability of this plant pathogen. However, it is as yet unknown whether other classes of retrotransposons, such as LINEs or SINEs, exist and to what extent they have proliferated in the genome of *P. infestans*.

This paper reports the identification and characterization of tRNA-related SINEs in the *P. infestans* genome by the exploitation of existing partial genomic DNA sequence, EST datasets, and a large insert bacterial artificial chromosome (BAC) library. The diversity of individual SINE families, together with their distribution within the genus *Phytophthora*, suggests that SINEs have been active throughout the evolution of *P. infestans*.

2. Materials and methods

2.1. Phytophthora isolates

Phytophthora infestans isolates T30-4, 88133, 80029, 88069, and CY29 were selected from the culture collection of the Scottish Crop Research Institute (SCRI; Invergowrie, Dundee, UK) and used to isolate and characterize candidate SINEs. Additional *Phytophthora* species *P. cactorum* (IMI 296524; clade 1a), *P. nicotianae* (IMI 268688; 1b), *P. citricola* (cit2; 2), *P. citrophthora* (IMI332632; 2), *P. palmivora* (Palm P222; 4), *P. megasperma* (IMI 333317; 6), *P. cinnamomi* (CBS 342.72; 7a), *P. fragariae* var. *rubi* (IMI 355974; 7a), *P. cryptogea* (IMI 045168; 8a), *P. drechsleri* (P532; 8a), *P. porri* (P1720; 8b), and *P. brassicae* (HH; 8b) were also selected from the SCRI culture collection. From sequencing of ribosomal DNA internal transcribed spacer (rDNA-ITS), *P. infestans* is most closely related to *P. nicotianae* and *P. cactorum*, and most distantly related to *P. porri* and *P. brassicae* (Cooke et al., 2000). DNA from *P. brassicae* (HH) was provided by N. Williams and D. Cooke, SCRI. Cultures were maintained on Rye A agar (Ribeiro, 1978) at 20 °C. *Phytophthora* mycelium for nucleic acid extraction was grown at 20 °C in pea broth (125 g/L frozen fresh peas, boiled for 1 h, filtered, and autoclaved) for 3 days, harvested by filtration, frozen in liquid nitrogen, and stored at −70 °C prior to use.

2.2. Nucleic acid manipulations

Genomic DNA was extracted from frozen mycelium using the DNeasy plant mini kit (Qiagen, UK), following the manufacturer's protocol. DNA from BAC clones was isolated by alkaline lysis (Sambrook et al., 1989). All DNA samples were stored at −20 °C until used.

Genomic DNA samples (5 μg) used in Southern hybridizations were digested to completion with either *Hin*dIII or *Pst*I (New England Biolabs) according to the conditions recommended by the manufacturer. For Southern hybridizations, BAC DNA was digested with either *Hin*dIII or *Pst*I. For subcloning, BAC DNA was partially digested with *Sau*3AI (New England Biolabs).

Genomic and BAC DNA restriction fragments were electrophoresed on 1% w/v agarose gels in 1× TBE buffer. PCR products (10 μl) were electrophoresed on 1.5% agarose gels in 1× TBE buffer. Gels for Southern hybridization were treated and DNA transferred to Hybond N+ nylon membrane (Amersham Biosciences, UK) using standard capillary transfer methods (Sambrook et al., 1989), followed by UV crosslinking of the DNA to the membrane.

DNA probes (25 ng) were labeled with [α-^{32}P]dCTP using the High Prime labeling system (Roche Diagnostics Gmbh, Germany). Prehybridization, hybridization, and autoradiography were carried out as described by Whisson et al. (2001).

Products from PCR were cloned into the pGEM-T Easy vector (Promega) using the manufacturer's protocol. BAC DNA, partially digested with *Sau*3AI, was cloned into vector pGEM-3Z (Promega) that had been previously digested with *Bam*HI. Cloned DNA was electroporated into *Escherichia coli* strain DH10B (Invitrogen, UK) as described in Whisson et al. (2001). Plasmid DNA for sequence determination was isolated from bacterial clones by alkaline lysis. Sequences of cloned DNA were determined using SP6 and T7 promoter primers, and BigDye v2.0 or v3.1 sequencing chemistry (Applied Biosystems, UK).

Total RNA was extracted from frozen *P. infestans* mycelium using the RNeasy plant mini kit (Qiagen, UK), following the manufacturer's protocol, and RNA samples stored at −70 °C until used for cDNA synthesis.

RNA samples were treated with DNase (DNA Free kit, Ambion Europe, UK) prior to cDNA synthesis to remove DNA contamination. First-strand cDNA was synthesized from total RNA by both random and oligo(dT) priming using the first-strand cDNA synthesis kit (Amersham Biosciences, UK) and following the manufacturer's protocol. Synthesized cDNA was stored at −20 °C until used in RT-PCR.

2.3. PCR amplification of SINEs

Each 20 μl PCR contained 0.5 U *Taq* polymerase (Promega, UK), 1× reaction buffer (reaction buffer B, containing $MgCl_2$; Promega), 250 μM deoxynucleotide triphosphates (Promega), 1 μM forward and reverse primers, and 10 ng genomic DNA. The same reagent concentrations were used for RT-PCR, except that undiluted (0.5 μl) and diluted (1:10) in water first-strand cDNA were used.

Stringency of PCR conditions was achieved through altering the primer annealing temperature. Low stringency conditions used for A and B box primers (TGGCTCAGTGG and GGGATCGAACC, respectively; Borodulina and Kramerov, 1999) were as follows: 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 45 °C for 30 s, and 72 °C for 30 s. Higher stringency conditions were used for PCR from *P. infestans* genomic DNA. The reaction conditions were as described above, with the exception that the annealing temperature used was 60 °C. Similarly, PCR of SINEs from various *Phytophthora* species used an annealing temperature of 50 °C. Sequences for oligonucleotide primers used in PCR of *P. infestans* SINEs are listed in Table 1.

2.4. High-density colony filters of BAC clones

The *P. infestans* BAC library used for hybridization to SINEs was described by Whisson et al. (2001). Clones from the BAC library were transferred to Performa positively charged nylon membranes (Genetix, New Milton, UK) using a Genetix Q-Bot laboratory robot. The entire BAC library was represented on three filters. Each clone was double spotted in a unique pattern and the position of the clones in the array was designed to be diagnostic of the plate from which the clones originated. Membranes containing arrayed clones were grown for 24 h at 37 °C on LB–chloramphenicol agar before the bacterial colonies were lysed with alkali to bind the DNA to the membrane (Sambrook et al., 1989).

2.5. DNA sequence datasets and search strategy

Phytophthora infestans genomic DNA sequence reads (not assembled) were available to us through the *Phytophthora* Genomics Consortium (Garelik, 2002; Lam, 2001). This sequence resource consisted of 530801 individual single pass sequence reads, typically of 400–600 bases in length, in FASTA format, trimmed of poor quality and vector sequence as described by Waugh et al. (2000). Sequences were generated from *P. infestans* genomic DNA libraries of 3 kb *Hin*dIII fragments, 1–1.5 kb *Hin*dIII fragments, 1.5–3 kb sonicated fragments, 1–1.5 kb sonicated fragments, 1–1.5 kb *Eco*RI fragments, and 3 kb *Eco*RI fragments.

A consensus sequence to search for *Phytophthora* RNA polymerase III promoters was generated as follows. Sequences for RNA polymerase III promoters

Table 1
Oligonucleotide primers used for PCR amplification of *P. infestans* SINEs from cDNA of *P. infestans*, genomic DNA of *Phytophthora* species, and for quantitative PCR from *P. infestans* genomic DNA

SINE	Primer	Primer sequence	Quantitative PCR primer	Quantitative PCR primer sequence
infSINEb	infSINEbF	GTGGCTCAGATGGCTACAGT	infSINEbTMf	CAGTGGCACATCATTGCTTTG
	infSINEbR	GTACGAATGGGACTGCTCT	infSINEbTMr	TGCTCTTTTCGCCGACTGA
infSINEc	SINE4F	GAATTGGCGCAGTGGAAAA	SINE4F	GAATTGGCGCAGTGGAAAA
	SINE4R1	TAAAACGCGAGCACCGCCC	SINE4R	GCGGAGGGACTCGAACTCA
infSINEd	InfSINEdF	CAGTGGTTGGCGCAATGGGA	infSINEdTMf	TGTCGATCCTCGCGTTAAACT
	infSINEdR	CGATGCTGTGTTGTTCGGT	infSINEdTMr	TGGCACCCACGCAGAAA
infSINEh	infSINEhF	ACGTCGTGTTCGAGCGGGT	infSINEhTMf	GTGTCCGAGCGGGTAACG
	infSINEhR	CACTTGCTCGTGCACTTGTCG	infSINEhTMr	TCTCCGTGGGAATCGAACTC
infSINEi	infSINEiF	GCTTACTCACCGAGTGGCT	infSINEiTMf	AGCGGTACAGTGCCGAGTTC
	infSINEiR	AACTGGAGTTGACGCCGGGA	infSINEiTMr	TGATTAAGAAGAAAACTGGAGCTGAC
infSINEj	SINE7F1	ACGGCATTGGTGGAGAGGT	SINE7F	GCAATATCGCGGGTTCG
	SINE7R	GCAGACCCCTAAGGGATCCT	SINE7R	GCAGACCCCTAAGGGATCCT
infSINEl	infSINElF	CCGACTTCGTGGCGCAGTGGT	infSINElTMf	TCGCCGCTGGTAATCGA
	infSINElR	CACTCCCGTGTGTGAGCGGA	infSINElTMr	GGTGACCAGTGGGTTGTCAAC
infSINEm	SINE10F	GGTAAAATGGAGGCGGTACGT	SINE10F	GGTAAAATGGAGGCGGTACGT
	SINE10R1	TTCGCCCTTCTTGCCCTTC	SINE10R	GGCTCCGACCGGCGTTC
infSINEo	infSINEoF	GCGGTGGCGGAGGGGGTGGAT	infSINEoTMf	GATTGGGTTCGAATTCTGCATAC
	infSINEoR	CTGAATAGTCCGGATGATGCGA	infSINEoTMr	AGCGACTAATTCGCTCATTGG
infSINEp	infSINEpF	CCAACGCCCCCGTAGCTCA	infSINEpTMf	CGGTAGAGCGGGAGACTACGT
	infSINEpR	CCCACACAGTACAGCTTACA	infSINEpTMr	CCCACACAGTACAGCTTACAACTATCT
infSINEq	infSINEqF	CGAATTCTCGTTAGTATAGT	infSINEqTMf	CGAGTTGTGGTTGTACGAATTCTC
	infSINEqR	CTGGCAGTATGGGCGCAGCA	infSINEqTMr	CGTCCGGGAATTGAACCC
infSINEr	infSINErF	CCCTTATAGCTCAATGGT	infSINErTMf	CGGCCTGTCAAGCCAAAG
	infSINErR1	GTAACAGCGAAAGTGGTGTCC	infSINErTMr	TCGATGAGCATGATCCAAATG
	infSINErR2	GTTCATGGAACGAGAGGGCT		
infSINEt	infSINEtF	CGACAGCTGGCCGAGTGGT	infSINEtTMf	GAGTGGTTAAGGCGATAGACTAAAATC
	infSINEtR	CAAGGCAATATGGATCCGA	infSINEtTMr	GCAAGGCAATATGGATCCGA
infSINEy	infSINEyF	GCGTGACTGGCGCAATGGA	infSINEyTMf	CGCAATGGATAGCATAATATGCTATT
	infSINEyR	CGCGACGCAAGACTCGA	infSINEyTMr	CAAATTTACGCGACGCAAGA
infSINEz	infSINEzF1	GGGATGTAGCTTAAATGGT	infSINEzTMf	GCTTCTCCAGGCGTGTGTAGT
	infSINEzR1	CACACGCCTGGAGAAGCA	infSINEzTMr	CGGATTCGAACCAGGGATTT
	infSINEzF2	CGTGTGTAGTATAGTGGTCA		
	infSINEzR2	GCATAATGAATATAAGAGTGCGCGT		

from *Arabidopsis thaliana* and rice were examined (Cheng et al., 2003; Myouga et al., 2001) and these sequences aligned with a *Phytophthora* sequence identified through low stringency PCR, BAC library screening, and subcloning. The range of base separations for A and B box sequences (25–50 bp) was also derived from observations in the tRNA database of Lowe and Eddy (1997) (http://rna.wustl.edu/GtRDB/). Plant sequences were used, as opposed to those from fungi or animals, since the phylogenetic affinities of the oomycetes lie closer to plants than other phyla.

The genomic sequence was searched for degenerate RNA polymerase III promoter A and B box sequences using the in-house program FindSeq3. This program finds occurrences of a query sequence (which may include IUPAC ambiguity symbols) and its reverse complement that are separated by a prescribed maximum and minimum number of bases within a larger nucleotide sequence. FindSeq3 is available as a Python script and a Windows executable at http://bioinf.scri.sari.ac.uk/lp/findseq.shtml, or from the authors on request (lpritc@scri.sari.ac.uk).

Sequence reads yielding a positive match were tested for presence of tRNAs and tRNA-related elements using the tRNAscan-SE program with the default settings (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/; Lowe and Eddy, 1997). Sequences yielding a positive tRNA identification were further inspected for the presence of an oligothymidine tract at the 3′ end of the predicted tRNA. Sequences predicted to encode tRNAs were catalogued, grouped into amino acid classes, and aligned using the MultAlign program (http://prodes.toulouse.inra.fr/multalin/multalin.html; Corpet, 1988).

Sequences encoding a predicted tRNA structure but with an extended 3′ tRNA-unrelated sequence before an oligothymidine tract were considered as candidate SINEs for further analysis, and were named in an alphabetical series from *infSINEa* to *infSINEz* (where *infSINE* signifies *infestans SINE*; Table 2). Additional candidate SINEs were obtained from those sequences found to contain matches to both A and B boxes, but where no apparent 5′ tRNA-related region was predicted by tRNAscan-SE. These sequences were then resubmitted to tRNAscan-SE using only the EuFindtRNA search algorithm

Table 2
Examples of genomic DNA sequences and copy number estimates for *Phytophthora infestans* SINEs

SINE	Sample DNA sequence [a]	Copy no. estimate—quantitative PCR (confidence interval)	Hybridizing BAC clones per haploid genome equivalent	Presence in other *Phytophthora* spp.
infSINEb	TTTTGATGCTGCCAGTGGCTCAGATGGCTACAGTGGCACATCATTGCTTT GACGTTCGTTGGCGTGGGTTCGAGTCCTCAGTCGGCGAAAAGAGCAGTCC CATTCGTACTTAGGATTTGTAAATTCAATTTGTAGCATACTTCGAAGCCC GATGTGTCTGTACAATTAGACTAAGGATCGGGTCAAAATGTGTGCAATAG TGGCAAAGCGTGTCAGGGTGTCAGTACTGGGTGATTCAGATAAAAGCTCA ATTAGTCTCAGAAACTGCCAATCGACATGCGCACATTCTTTTGATTG	1 (0.8–1.2)	1	All species tested
infSINEc	GGGCATTCGAATTGGCGCAGTGGAAAAGACTGGCCCGGTTTACACTCC TGGCGTGAGTTCGAGTCCCTCCGCAGCAACGGGCGGTGCTCGCGTTTTAC TGGTGATGCAATAACAGGTCATT	21 (20.1–21.5)	89	Not detected
infSINEd	TGGTTGGCGCAATGGGAAGTAGATGGGGAACCAACTGAACATGCCATCCG AGTGCGTGTCTGGTTCGATACTCAGCGTCACGTCACCGAACAACACAGCA TCGAAGTAACCCGCCAGTCCATGGACGACAGCGCTCTCTGCAGCGATAAA CCGAAGCTCTGCTTGTCTGCGGTTGTCGATCCTCGCGTTAAACTTCGGGT GCGTGAACCGGAAGCATTTCTGCGTGGGTGCCAAGGCGTAAAAGTTGTGA AGTCGCACCACATAGGGCGTCTCGAAGCTCTTGAGCGAGTCGTGCACTTT GACTTCGTTCCAGAGCTTGGAGGACATTACAGGTGTCACGAACG	1 (0.8 – 1.2)	1	*P. sojae* and *P. ramorum*
infSINEh	CTGCGACGTCGTGTCCGAGCGGGTAAAAGAGCATTCTCACGCGCAGTACG CGTGAGTTCGATTCCCACGAGGACCAAACTCCGCGACAAGTGCACGTGCA AGTGTTTTTGCAAGGAAACGCTGGCACGATGATCACCACAGACGCAAAAA AAGCGCCGACTGGAGACTTCCTTGTGCTTGGGGATATGCTCATCGTGGAA CGTTCACGATGTTCTAGCAGACACGAGGATCACGAAGACAGCGCACGAGG TACTTGAGACTGCGAATACGGGGGCGACAACCACTTCACTCAGTTGACTG AGCGACGATGTGCGACTGCACAGCATACGAAGAAGTATCTTGTAGACCCA GGCGACTGCGACGTCGGAATATAGGTGACCTGCGAGCTGGGGTTTCACTC GTCGTCGTGGAACCTCGTGTCACTCACGCCGCACTCACTCACTCACTCGC	35 (25–46)	41	Not detected
infSINEi	TCTCACTCACTCACTTACTTACTCACTTGTTCACTCGGCTTACTCACCGA GTGGCTCAGCGGTACAGTGCCGAGTTCTCGTATAGCTGGTCACCGGTTCG ATCCCGGCGTCAACTCCAGTTTTCTTCTTAATCATCATGGCGAGCGCCTC CGTGTACCCGCGCAGGCTGGTCGCCTTCTTCACCG	42 (30–57)	37	All species tested except *P. sojae*, *P. ramorum*, *P. citricola*, and *P. citrophthora*
infSINEj	ACGACGGCATTGGTGGAGAGGTTTAGCCGCGTCGTTCCACGCGCAATATC GCGGGTTCGAGTCCGCGGGCAGCGATAACACAGAAGGATCCCTTAGGGGT CTGCAACTGCCATGTAAGTCCTTTTGGAAAAAGTAGAAAAATATATTTAT GCATTATGAATACAAACAATGTACCACTATTTATTTGCAGACCTGGCAGC TTCCTTGTATCCGTCGTAAAAGCGTCAGGATACATCGCTATAATCTCTTG AAGCTCTCTACTGAGCTGCTCGACTTTCCAGACACCGTCACGAGGAAGAT CCAGGCCGCACTGAGCCATGGACTTATTTGCCCAATTTTCTCCGTTTTTG TCTTTCTCTAAGTTGACGAGGC	10 (10.1–10.2)	26	Not detected
infSINEl	CTCACGAAGCCGACTTCGTGGCGCAGTGGTCAATGCATGGGGCCGCGAGT CAGAACGCCAGGGTTCGATTCCCACCAGGCGCATCCGCTCACACACGGGA GTGTTGTATTTGCTGCCAGAACGGAACGGCTTCGAACATCTCAAAGACAT GAGACTTGTCGACTGGAAGTCAAGCACTTTAGCAACCCAAGTGAGCGGCA CGTTGTCACCACTGGTGTCACCAAATTTAGGTACTTGTTGACCACTGTGT CGCCGCTGGTAATCGAAGTCATTTCGGCGCAGTCGTAGCGTGTCGAGTTG ACAACCCACTGGTCACCGGTTCGATGTCGCCGGTGTTTATTTGCTTTCAT TATTGCACTGACAAGTTGTGTCAGTGTTCGCCGCGGCTCACG	86 (65–114)	35	Not detected

(*continued on next page*)

Table 2 (*continued*)

SINE	Sample DNA sequence[a]	Copy no. estimate—quantitative PCR (confidence interval)	Hybridizing BAC clones per haploid genome equivalent	Presence in other *Phytophthora* spp.
infSINEm	**GTCACTTGGTAAAATGG**AGGCGGTACGTGTAGCTGAAACCACCTCGGAGT TCGAGTCC**GAACGCCGGTC**GGAGCCCCCACCTGAGCCTCCGTCACCACCA GAAATCGTGGCGATTGCAGACCTACTTGACGAGGTATACGCGGAGCCTAA ACAGCGTCAGATGCATTATACAATGAAGAAGAAACGACATGCTCTTGAAG GGCAAGAAGGGCGAATATACTGCGTATCAGTCTCTACAAAGGTTGCTACG ACGTTTT	14 (11.9–17.3)	8	Not detected
infSINEo	**TGGCGGTGGCGGAGGGCGTGGATTTTCAACTGGCTCCCTTTCTTCGATTG** GGTTCGAATTC**TGCATACCTTC**GAGCAGCACTCGTAGCAACGGGCGCCCA GCAATAATTCCGACCAATGAGCGAATTAGTCGCTCAGCATCAGATTGAGA TAATGTCGCATCATCCGGACTATTCAGTCTTTCGACCAATCCCATGGAAT CTTCACCTTGATTCGTCAAAACTTCATCTTGTGCGTTTT	47 (35–61)	41	Not detected
infSINEp	**CCCCCGTAGCTCAGCGGTAGAGCGGGAGACTACGTCTCCAGGCCGCTGGT** TCGAATCC**AGCCGGGGACA**GATAGTTGTAAGCTGTACTGTGTGGGATGTG TTTGTGTTTT	21 (16–28)	26	*P. sojae* (tRNA only), *P.cactorum*, *P.cryptogea*, *P. brassicae*
infSINEq	**TCTCGTTAGTATAGTGGTTAGTATACCCGCCTGTCACGCGGGTGACCCGG** GTTCAATTCC**CGGACGGGAAG**TCTGCTGCGCCCATACTGCCAGCTTGTTTT	776 (630–955)	6	*P. sojae* and *P.ramorum* (tRNA only), not detected in other species
infSINEr	CAACA**CCCTTTTAGCTCAATGGTTGAGCGCTGGCTTTGAGAGCCACAGGT** **CGGT**GGTTCGAGCCC**ACCAGGGCA**CAGCCCTGTCGTTCCATGAACGATGT TCACGCTTGGACAAGCTTAAGAGCAATC**CGCTTTTAGGTCAGTGGTCGAG** **CGTTGGCTTTGGAAGCCACAGGTCGGT**GGTTCGAACCC**G**CCAGGGCACAG CCCTCTCGTTCCATGAACTATTGTCCTTGAATGGACAAGCTTAAGGGCAA **CGCCCTTTCAGCTCAGTGGTCGAGCGCTGGCTTTGAAAGCCACAGGTCGG** **T**GGTTCAAATCC**ACCCGGGCGCA**GCCCTCTCATACCATGAGTAGTCCGTT CTCGGACTGACGTACGGGCATCG**CCCTTTCTAGCTCAATGGTAGAGCTCCG** **GCCTGTCAAGCCAAAGGCCAGT**GGTTCGATCCC**CGTGCGGGAC**ACCACTT TCGCTGTTACAGCGCATCCACATTTGGATCATGCTCATCGAGCAAGTTGT TT*CAAAACAACTATGCGTCTATCGAGACGCTTGTGCCTGAACCTTGTTCA* *GCACAACACGAAGCACCTGCTTCCTTCGGGACCATGTCCCTAACACTACG* *AGGTCTCTCCACCTCCTGCTCCCGC*	562 (457–692)	155	Not detected
infSINEt	**GACAGCTGGCCGAGTGGTTAAGGCGATAGACTAAAATCTATGGGGTAACC** **CGCGTG**AGTTCAAATCT**CACAGCTGTCG**GATCCATATTGCCTTGCGTTTTT	157 (119–206)	1	*P. sojae* and *P.ramorum* (tRNA only), not detected in other species
infSINEy	CAAGTTTTGTTGGAGCAAATGTATAAAACATG**CGTGACTGGCGCAATGGA** **TAGCATAATATGCTATTATCAGACTTCGAATCTGAAGGTTGCG**GGTTCGA GTCT**TGCGTCGCGTA**AATTTGTTTGAATATCCCGAAAACCACAATTTTGT	8 (6–11)	3	*P. sojae* (5'truncated), *P.cactorum*
infSINEz	**GGGATGTAGCTTAAATGGTAGAGCGCTCGCTTAGCATGCGAGAGGTACAG** GGATCGATACC**CTGCTTCTCCA**GG**CGTGTGTAGTATAGTGGTCAGTACCT** **CACCTTGTGGCGGTGAAATCCCT**GGTTCGAATCC**GGGCACGCGCA**CTCTT ATATTCATTATGCTTCTTCTACCTTTGCTTTTT	541 (410–714)	14	*P. sojae*, *P.ramorum*, *P.cinnamomi*, *P.megasperma*, and *P. citricola*

[a]The 5′ tRNA-related regions are shown in bold type, A (left) and B box (right) promoter elements are shown as black and grey shaded boxes, respectively, and possible insertion site direct repeats (where present) are underlined. The region of *infSINEr* shown in italics is similar to a *P. infestans Gypsy*-like retrotransposon.

operating on 'relaxed' parameters. Sequences that still yielded no support for a 5′ tRNA-related region were not considered further, as these may encode other RNA polymerase III-transcribed sequences such as 7SL RNA, 5S rRNA, or some snRNAs. All candidate SINEs were aligned together to identify further related SINE family members. The SINE discovery process to this stage was automated in later analyses to accelerate SINE discovery and the script for this analysis pipeline can also be found at http://bioinf. scri.sari.ac.uk/programs.html. Potential ancestral tRNAs for predicted SINEs were identified by pairwise sequence alignment, using MultAlign, with *P. infestans* tRNAs predicted by tRNAscan-SE. Insertion site sequence duplications flanking the SINEs were searched for using the 30 bp 5′ from the A box, and sequences of variable length 3′ of the predicted RNA polymerase III transcriptional stop. When searching for insertion sequence duplication, direct repeat sequences were searched for behind each possible transcriptional stop using MultAlign.

2.6. Quantitative SYBR green real-time PCR assays

Primer pairs for quantitative PCR specific to each *P. infestans* SINE (Table 1) were designed and optimized as described by Avrova et al. (2003). Quantitative PCR primers for *infSINEr* were targeted to one of the four monomers. Genomic DNA from *P. infestans* isolate T30-4 was used in all optimization procedures. Primers were also designed and optimized for the *ActB* (5′-CCG TACCCCTGAGGTGCTATT-3′ and 5′-GGAAAGCG CACTCGTGAACT-3′), *Inf*1 (5′-CCTCTCGGACAC GTCGTTTAA-3′ and 5′-ACGCGCACATGAGCTTG TAC-3′), and *piCdc*14 (5′-TCGGCGGGCATCGA-3′ and 5′-CGGCATAGGCGCATTTG-3′) genes, as known single copy genes. The amplification efficiencies of *ActB*, *Inf*1, *piCdc*14, and all SINE amplicons were determined to be equivalent over a wide range of genomic DNA concentrations (from 1:10 to 1:160 dilution), allowing the use of the comparative C_t method for relative quantification of target sequence copy number (Avrova et al., 2003). Detection of real-time PCR products was by binding of the fluorescent DNA dye SYBR green (SYBR green PCR master mix; Applied Biosystems, USA) to PCR products. All assays were carried out in triplicate, and no template controls were included. PCR conditions were as described by Avrova et al. (2003).

3. Results

3.1. Development of RNA polymerase III promoter A and B box sequences to search P. infestans DNA sequence databases

An initial strategy to isolate SINEs using PCR primers designed to general RNA polymerase III promoter A and B boxes (Borodulina and Kramerov, 1999) yielded a mixture of PCR products between 50 and 100 bp. These DNA fragments, when cloned and sequenced, all showed greatest homology by BLASTN search to an aspartic acid tRNA from *P. parasitica*.

To determine the full sequence of the tRNA-like sequence, the BAC library was screened by hybridization with the PCR product, and positive clones identified. Subclone libraries of two BACs, 31P22 and 27I22, were made, screened with the PCR product, and positive subclones sequenced. This yielded only a single sequence type, and contained a predicted aspartic acid tRNA with a typical oligothymidine transcriptional stop for RNA polymerase III at the 3′ end of the tRNA sequence. Comparison of the *P. infestans* A and B box promoter elements with those from *A. thaliana* and rice yielded degenerate A and B box sequences of TRKYNNARNGG or TRKYNNARWNGG (A box) and RGTTCRANHYY (B box). Two versions of the A box consensus were used in further sequence searches, as it was observed that tRNAs from *A. thaliana* frequently possessed an additional base at position 9.

3.2. Prediction of tRNAs and SINEs from P. infestans

All available sequence reads were scanned using the FindSeq3 program, which was set to search for the degenerate A and B box sequences. FindSeq3 identified a total of 4414 sequences as containing the A and B box sequences described in the previous section. From these sequences, 277 were predicted by tRNAscan-SE and/or EuFindtRNA not to possess a tRNA-related sequence, and were not considered further as they may encode other RNA polymerase III transcripts such as 5S rRNA, 7 SL RNA, or snRNAs.

The remaining 4137 sequences were examined further using tRNAscan-SE or EuFindtRNA. Using tRNAscan-SE set to default parameters, the remaining 4137 sequences were examined for presence of a 5′ tRNA-related region. A tRNA-related region was not always observed in potential SINE-like elements, despite the presence of A and B box sequences. Relaxing the stringency of the tRNA search parameters using only the EuFindtRNA algorithm identified possible 5′ tRNA-like structures in several sequences. In total, 12 different sequence groups were identified that contained RNA polymerase III A and B box motifs, a 3′ tRNA-unrelated region, and where a tRNA was predicted under the relaxed parameters of EuFindtRNA. Representative sequences of all potential tRNA-related SINEs are shown in Table 2. SINE sequences described have been deposited in GenBank as Accession Nos. AY623609 to AY623621, and AY848952–AY848954.

Our search strategy also identified numerous predicted tRNA sequences, thus validating the A and B box sequences used to search the *P. infestans* genomic

DNA database. Based on the predicted tRNA sequences, the RNA polymerase III A and B box promoter sequences for *P. infestans* can be further refined to TRGYNBAR(T)YGG (A box) and RGTTCRADYYY (B box). The T in parentheses for the A box sequences denotes an extra base at this position in phenylalanine tRNAs. Sequence diversity among the tRNA sequence group members was typically very low at less than 1% within the tRNA coding region, implying that there is stabilizing selection pressure to maintain tRNA integrity.

Two SINEs are composed of multiple copies of tRNA-related units. *InfSINEr* is composed of at least four similar copies of a tRNA-related repeat element. Each element possesses a tRNA-related sequence at the 5′ end, and a 56–68 bp 3′ sequence that overruns the tRNA region of the next unit. That is, the final thymidine of the transcriptional stop for the first element forms the 5′ thymidine of the A box motif for the second element. In some copies of this element, the oligothymidine tract marking the end of one unit and the start of the next was not obvious (less than T_4; Table 2; Fig. 1B). A further element named *infSINEz* was identified that comprised two different tRNAs separated by only two base pairs. No oligothymidine tract was identified between the two tRNAs, and in the absence of contradictory evidence, this element was considered as a potential SINE or dicistronic tRNA.

The full complement of tRNA gene sequences for *P. infestans* is not known and this prevents unambiguous assignation of the tRNA background for each of the SINEs identified. However, *infSINEq* is most probably derived from aspartic acid tRNAs, as this SINE exhibits near identity in its tRNA-related region to the aspartic acid tRNA from *P. parasitica* and *P. infestans* (Fig. 1A), and the 3′ sequence of *infSINEq* exhibits similarity to the region immediately 3′ of aspartic acid tRNA sequences. *InfSINEq* differs from the aspartic acid tRNA due to a 10 bp deletion containing the oligothymidine tract for the aspartic acid tRNA resulting in a 5′ aspartic acid tRNA followed by a 30 base 3′ tRNA-unrelated sequence. Each of the four tRNA-related SINE units that comprise *infSINEr* are homologous to predicted threonine tRNAs (72–79% identity; Fig. 1B). *InfSINEj* exhibited 68% identity to predicted glycine tRNAs. The two tRNAs that form *infSINEz* are predicted by tRNAscan-SE to be alanine and histidine tRNAs; the alanine tRNA is similar to alanine tRNAs from other organisms. Other predicted SINEs do not show significant similarity to known or predicted tRNAs.

3.3. Copy number and genomic organization of predicted SINEs

Although the genome of *P. infestans* has not been fully sequenced and assembled, an approximate range of copy number for each putative SINE could be estimated from the number of BLASTN matches of each SINE sequence to the existing genomic DNA sequence reads. However, there is the possibility of sequence over- or under-representation in our genomic DNA sequence data, and so a more accurate estimate of SINE copy numbers was made by real-time quantitative PCR. The copy numbers obtained from quantitative PCR are expected to be a minimum copy number estimate rather than the exact number present within the genome as primer pairs may fail to anneal and amplify all copies of the SINEs, due to sequence mismatch, insertions or deletions (Fig. 2 for example). For example, over 63 different *infSINEh* sequences were identified from the single pass

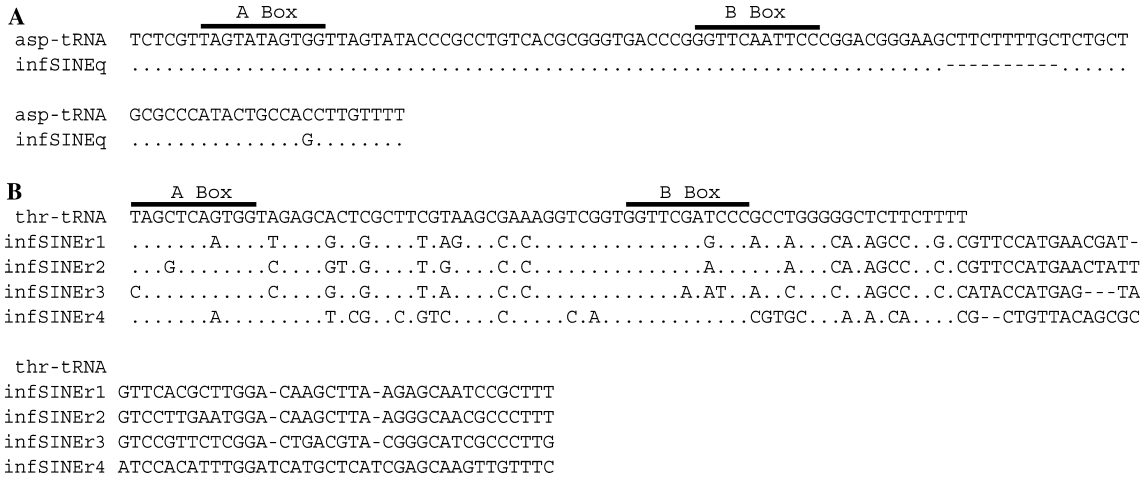

Fig. 1. Alignment of genomic DNA sequence of *infSINEq* with a predicted *P. infestans* aspartic acid tRNA (A). The sequence of *infSINEq* is almost identical to that of the aspartic acid tRNA, with the difference that *infSINEq* has gained a short 3′ tRNA-unrelated sequence by the deletion of the oligothymidine tract from the aspartic acid tRNA. The sequence of *infSINEr* from Table 2 has been separated into its component 5′ tRNA-related and 3′ tRNA-unrelated monomers (*infSINEr*1 to *infSINEr*4) to enable alignment (B). All monomers exhibit sequence similarity to predicted *P. infestans* threonine tRNAs.

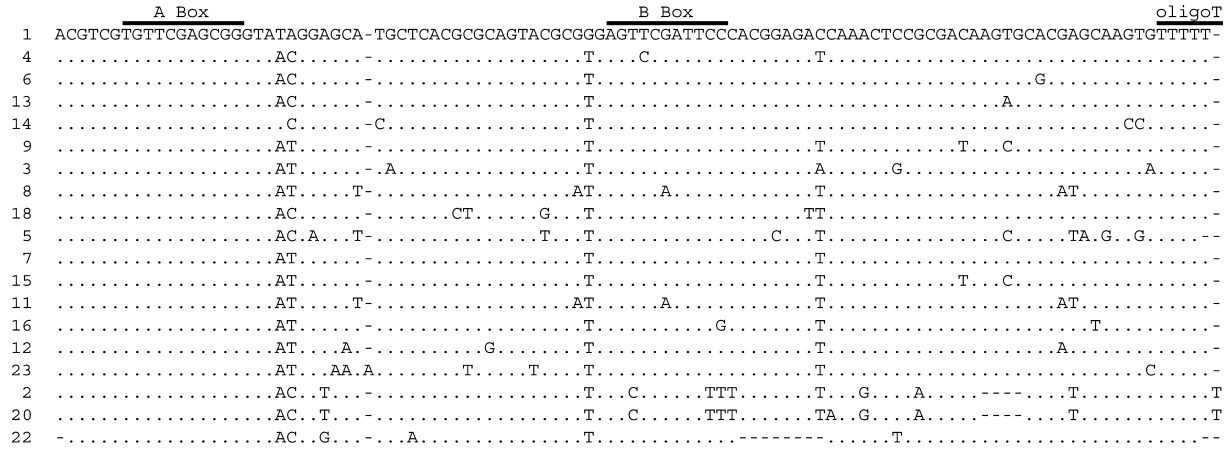

Fig. 2. Alignment of 19 different copies of *infSINEh* from the genomic DNA of *P. infestans*. Sequences were derived from cloned PCR products of *infSINEh*. The predicted transcribed region consisting of a 5′ tRNA-related region (A–B box), 3′ tRNA-unrelated region, and 3′ oligothymidine tract ($T_{4–6}$) is shown. The sequence conservation at the 5′ end of the sequence is due to the binding site of the PCR primer used to amplify the sequence.

genomic sequences, almost double the total estimate (35) by quantitative PCR. However, only 26 of these 63 sequences support complete sequence homology with the primers used. Minimum copy numbers estimates by quantitative PCR can be found in Table 2.

To gain another estimate of SINE copy number in the *P. infestans* genome, each SINE sequence was also hybridized to the existing *P. infestans* BAC library on high density filters. With the exception of *infSINEc*, *q*, *r*, *t*, and *z*, good agreement was obtained between the copy number estimates from quantitative PCR and BAC library hybridization (Table 2). For *infSINEq*, 776 combined copies were estimated by quantitative PCR, but only 64 hybridizing BACs, corresponding to 6–7 copies per haploid genome equivalent. Primers designed for *infSINEq* did not discriminate between *infSINEq* and aspartic acid tRNA sequences due to their sequence similarity, and so the copy number estimate is combined for these sequences. Hybridization of an *infSINEq* probe to BACs yielded more intense signals on autoradiographs than those observed for most other SINEs. One possible explanation for this may be that all of the copies of *infSINEq* are clustered at a few genomic loci. This interpretation is supported by the observation that PCR of this element from genomic DNA of *P. infestans* yielded a ladder of multiple DNA fragments of increasing size approximately every 500 bp, indicative of an array of *infSINEq* and/or aspartic acid tRNA sequences. It was not possible to determine if arrays of one sequence or the other were present in particular genomic regions, or if *infSINEq* and aspartic acid tRNA sequences were mixed within an array. Similar observations to those for *infSINEq* were also made for *infSINEt* and *z* (Table 2), and were verified for *infSINEq* and *z* by Southern analysis of separate *Hin*dIII and *Pst*I-digested BAC clones (not shown).

The most abundant SINEs encountered in the search of *P. infestans* genome sequence reads were *infSINEh* and *infSINEi*. Both elements were typically encountered together, separated by approximately 350–400 bp. The region between the two SINEs is characterized by variable numbers of tetranucleotide repeats (ACTC) close to *infSINEi* (Table 2). The number of positively hybridizing BAC clones was in agreement with the quantitative PCR copy number estimates for *infSINEh* and *i* of 35 and 42 copies, respectively. It was expected that all BAC clones that hybridized to *infSINEh* would also hybridize to *infSINEi*. However, numerous BAC clones were identified that hybridized to *infSINEh*, but not *infSINEi*, and vice versa. The presence/absence of *infSINEi* and/or *h* on individual BAC clones was also verified by Southern hybridization of *Hin*dIII-digested BAC DNA (Fig. 3). This suggests that not all of the genomic sites where these elements can occur have been occupied by *infSINEh* and *i*. This is not due to a *Hin*dIII site between *infSINEh* and *i* (the BAC library predominantly used *Hin*dIII as a cloning enzyme) separating the two SINEs in different BAC clones. Southern hybridization of selected BAC clones to *infSINEh* or *i* (Fig. 3) also revealed that over half of the BAC clones tested contained multiple copies of these SINEs (similar results were also obtained for *infSINEo*, *r*, and *z*). Substantial sequence diversity was identified among different copies of *infSINEh*, mostly as single base pair substitutions, but at least two indels also occur in the 3′ tRNA-unrelated sequence (Fig. 2). The sequence diversity observed in the single pass genomic sequences was verified by cloning and sequencing 22 PCR-amplified copies of *infSINEh* in both directions (Fig. 2). The level of sequence diversity observed was similar to that in the single pass sequences for *infSINEh*.

From the sequences and copy number estimates presented in Table 2, the predicted SINEs comprise 0.3% of the 250 Mb genome of *P. infestans*.

For each family of SINEs, the immediate 5′ and 3′ regions (outside A–B box) were found to be highly similar

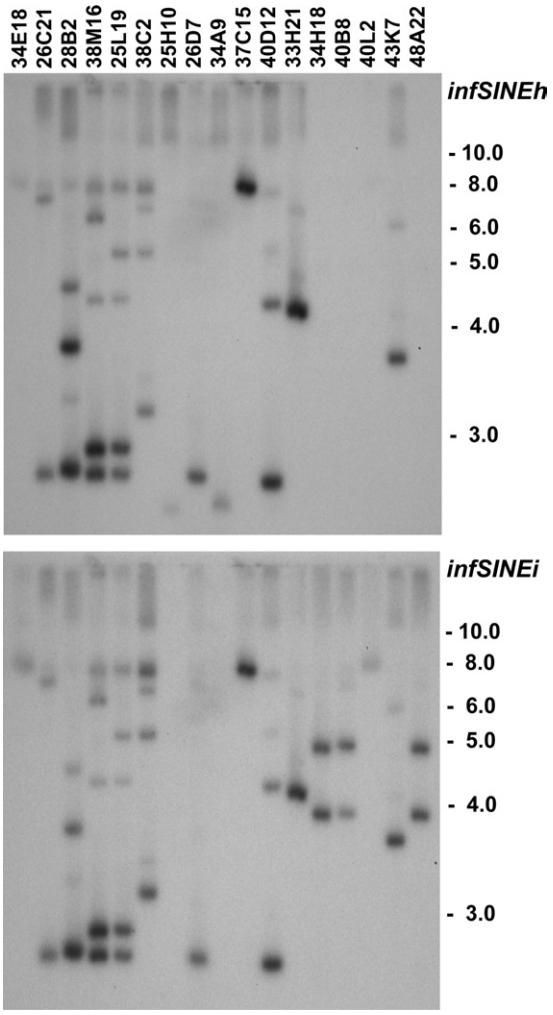

Fig. 3. Southern hybridization of *infSINEh* and *infSINEi* to *Hin*dIII-digested *P. infestans* BAC DNA. The same BAC clones in the same order are shown in both panels. Most BAC clones contain multiple copies of both SINEs. Although *infSINEh* and *i* are typically found together, BAC clones 25H10 and 34A9 contain *infSINEh* but not *infSINEi*, and clones 34H18, 40B8, 40L2, and 48A22 contain *infSINEi* but not *infSINEh*. Molecular size markers (kb; 2-log ladder; New England Biolabs) are shown at right.

for each copy of that element. For example, the regions of sequence directly surrounding the various copies of *infSINEh* are all very similar to each other, as are the sequences surrounding copies of *infSINEq* (but distinct to those surrounding *infSINEh*). Without a fully assembled genome sequence or fully sequenced BAC clones containing these elements, it is difficult to draw any conclusions as to the broader genomic environment surrounding SINEs. However, since the surrounding DNA for each family of SINEs is highly similar, it is probable that *P. infestans* SINEs may typically form part of larger, interspersed repeat units. Evidence for this comes from the observation that some elements discovered in this report are physically located adjacent to other repeat sequences described by Judelson and Randall (1998) and Judelson (2002). For instance, *infSINEo* is located next to a segment of repeat AA2 (Judelson and Randall, 1998); repeat AA2 shares homology with a reverse transcriptase sequence. The final 3′ tRNA-unrelated sequence of the multimeric *infSINEr* marks a junction with a *P. infestans Gypsy*-like retrotransposon (Judelson, 2002).

A characteristic of SINE insertion into genomic DNA is the direct duplication of short sequences flanking the insertion site. For *infSINEd*, *m*, *o*, *p*, *q*, and *t*, no obvious direct repeats could be identified. Short direct repeats of 6 bp were identified for *infSINEl* and *r*, and a direct repeat of 7 bp for *infSINEb*, *c*, and *j*. Evidence for longer direct repeats was found for *infSINEh* (11 bp), *i* (10–14 bp), and *y* (9 bp). However, these repeats were typically partially degraded, although the length of the repeats was preserved (Table 2). For *infSINEh*, an 11 bp direct repeat was located adjacent to the A box sequence, and its cognate sequence was located 238–250 bp after the oligothymidine and approximately 60 bp before the ACTC repeats prior to *infSINEi* (Table 2).

3.4. Transcription of *P. infestans* SINEs

The transcriptional activity of the predicted SINEs was tested by RT-PCR. Random-primed cDNA synthesis was used since none of the identified elements were predicted to have a substantial polyadenylated 3′ tail. To confirm the lack of polyadenylation, cDNAs were also synthesized in an oligo(dT)-primed reverse transcription reaction. RT-PCR was carried out using both sets of cDNA samples, undiluted and diluted 1:10, on cDNAs synthesized using total RNA from the mycelial growth stage. Only *infSINEh* and *infSINEi* were found to be actively transcribed in *P. infestans* mycelium. No RT-PCR amplification was observed from any other predicted SINEs. This also indicated that the cDNA synthesis was not contaminated with genomic DNA. The two transcribed elements were detected only in undiluted random-primed cDNA samples, suggesting that they are transcribed at a relatively low level and are not polyadenylated. Typically, a 1:10 dilution of synthesized first-strand cDNA is adequate for RT-PCR detection of most *P. infestans* genes transcribed in mycelium (A.O. Avrova, personal communication).

SINEs composed of multiple units (*infSINEr* and *z*) were tested as individual units. That is, each tRNA-related unit was tested separately for its transcription. For both elements, no transcription was detected either as an intact multimeric element, or as individual tRNA-related units. The significance of this for *infSINEz* is that it shows that this element is not responsible for producing a tRNA dicistron that is processed later to yield two tRNAs.

To gain some indication of any SINE expression not detected in our RT-PCR assay, available *P. infestans* EST databases (Lam, 2001; http://www.pfgd.org/pfgd/filter/filter.html) were searched by BLASTN for the presence of any SINEs, either as individual ESTs, or as insertions into expressed genes. However, no tRNA-related SINE insertions into 5′ UTRs, 3′ UTRs, or coding sequences of any ESTs were identified.

3.5. SINEs detect restriction fragment length polymorphisms among P. infestans isolates

A panel of five *P. infestans* isolates and two restriction endonucleases (*Hin*dIII and *Pst*I) were used to assess whether restriction fragment length polymorphisms (RFLPs) could be detected by six SINEs (*infSINEc*, *h*, *i*, *o*, *r*, and *z*). Complete digestion of the genomic DNA was verified by Southern hybridization to a probe of known copy number (A.O. Avrova; unpublished). Of the isolates used, 80029 and 88133 are known to exhibit numerous polymorphisms with other multi-locus DNA markers (van der Lee et al., 1997, 2001). All of the SINEs used as hybridization probes detected RFLPs. However, the high copy numbers of all of the SINEs tested prevented clear interpretation of hybridization profiles; RFLPs were only visible over a background smear of comigrating fragments (Fig. 4).

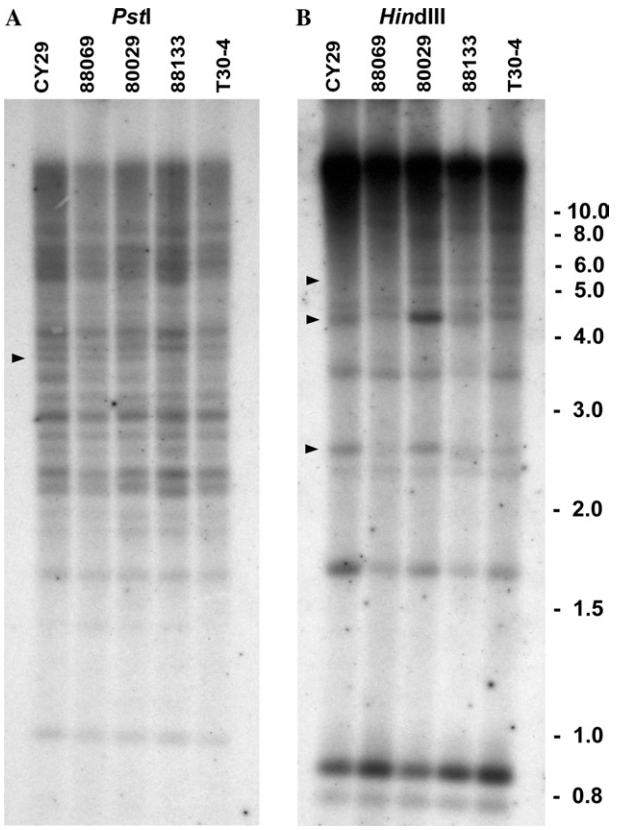

Fig. 4. Southern hybridization of *infSINEh* (A) and *infSINEz* (B) to *P. infestans* genomic DNA. Genomic DNA was digested with either *Pst*I (A) or *Hin*dIII (B) restriction endonucleases. Due to the high copy numbers of these sequences, polymorphisms (arrowed) are observed over a background smear of comigrating fragments. Molecular size markers (kb; 2-log ladder; New England Biolabs) are shown at right.

3.6. Distribution of P. infestans SINEs within the genus Phytophthora

Homologs of *P. infestans* SINEs were identified in other species of the genus *Phytophthora* in two ways. First, the *P. sojae* and *P. ramorum* genome sequences (http://genome.jgi-psf.org/) were searched by BLASTN to identify homologs of *P. infestans* SINEs. *P. sojae* and *P. ramorum* are not close relatives of *P. infestans* and are placed in clades 7b and 8a, respectively, of the *Phytophthora* phylogram (Cooke et al., 2000); *P. infestans* is placed in clade 1c. In *P. sojae*, these comparisons yielded sequences with significant similarity to *infSINEb*, *d*, *p*, *q*, *t*, *y*, and *z*. For *infSINEp*, *q*, *t*, *y*, and *z*, similarities to *P. sojae* sequences were limited to the A–B box region only. All of these elements (*infSINEp*, *q*, *t*, *y*, and *z*) possess a predicted tRNA structure in this region. In *P. ramorum*, homologs of *infSINEb*, *d*, *q*, *t*, and *z* were identified. As for *P. sojae*, the region of similarity to *P. ramorum* was typically restricted to the tRNA region of the *P. infestans* SINEs. Other elements, notably *infSINEb* and *d*, exhibited similarity to *P. sojae* and *P. ramorum*, either across the entire element (*infSINEd*) or only the 3′ tRNA-unrelated sequence (*infSINEb*). *InfSINEz* yielded a full-length sequence match with *P. sojae* and *P. ramorum* genome sequences, indicating that the organization of this element is conserved in distantly related *Phytophthora* species.

The occurrence of *P. infestans* SINEs within the genus was further investigated by PCR amplification of related sequences from a panel of 12 *Phytophthora* species representing the major clades within the genus. Although it is likely that primer binding in PCR will be prevented by sequence divergence of SINE sequences between species, it can serve as an indicator of SINEs that are evolving at a slower rate within the genus. Low stringency PCR from the selected 12 *Phytophthora* species yielded no products for *infSINEc*, *d*, *h*, *j*, *l*, *m*, *q* or *r*. Of these elements, *infSINEd* and *q* both have homologs in the *P. sojae* genome, suggesting that their sequences in other species are sufficiently divergent to prevent binding of primers during PCR. Other elements giving no PCR products for the other species tested may be restricted to *P. infestans* or other very closely related species such as *P. mirabilis* or *P. phaseoli*, both of which were unavailable for testing in this instance. Primers for *infSINEo* and *y* both generated poorly amplified, larger than expected fragments. Successful amplification of *infSINEb*, *i*, *p*, and *z* was observed for various species tested (summarized in Table 2).

To determine if the PCR-amplified products from the various *Phytophthora* species were genuinely representative of the existence of a SINE homolog in these species, the agarose gels from the species PCRs were Southern blotted and hybridized to probes of the appropriate SINEs. This strategy confirmed the amplification results observed for *infSINEb*, *i*, *p*, and *z*, and also detected hybridization to PCR products from some species not identified as harboring the sequence only by ethidium bromide staining of agarose gels (Table 2). For *infSINEo* and *y*, which generated PCR products that were larger than expected, it was shown that these were typically not homologous to the SINE probes, and were most likely to be randomly amplified products. An exception to this was the weak hybridization of *infSINEy* to a PCR product from *P. cactorum*.

4. Discussion

We have described 15 tRNA-related SINEs from the *P. infestans* genome, and their possible distribution within the genus *Phytophthora*. This represents the first report of SINEs from the oomycetes, although many other repetitive DNA sequences from *Phytophthora* species have been described previously. The genome of *P. infestans* contains many different families of repetitive DNA that together comprise over 50% of the haploid genome (Judelson and Randall, 1998). Our data suggest that only 0.3% of the 250 Mb *P. infestans* genome is composed of SINEs. The exact percentage is likely to be higher than this estimate, as our copy number estimates represent minimum copy numbers. The predicted SINEs each possess a predicted 5′ tRNA-related region containing the A and B box internal promoter elements for RNA polymerase III, and a 3′ tRNA-unrelated sequence. Eight of the SINEs also possess flanking short direct repeats.

Oomycetes exhibit marked differences to other eukaryotes at the gene organization and transcription level. For example, *Phytophthora* gene sequences rarely contain introns, a unique transcriptional start site sequence is frequently present 50–200 nt upstream of the ATG codon, and promoter consensus sequences do not always conform to those observed for other eukaryotes (reviewed in Kamoun, 2003; McLeod et al., 2004). Due to these differences, we aimed to first identify if *P. infestans* conformed to a general model for RNA polymerase III A and B box internal promoter elements, by using a low stringency PCR with 'universal primers.' Borodulina and Kramerov (1999) used this strategy to directly identify SINEs from the genomes of numerous organisms. For *P. infestans*, this approach did not directly identify any SINEs, but was successful in identifying an aspartic acid tRNA gene. This confirmed that general RNA polymerase III A and B box consensus sequences could be adapted and used to search our *P. infestans* genomic DNA sequence database for SINEs.

Copy number of the predicted SINEs was highly variable, ranging from low copy numbers (*infSINEb*, *d*, and *y*), moderately repetitive (*infSINEc*, *h*, *i*, *j*, *l*, *m*, *o*, *p*, and *t*) to highly repetitive (*infSINEq*, *r*, and *z*). With the exception of *infSINEq*, *t*, and *z*, which are clustered in arrays at several genomic sites, all other SINEs are dispersed throughout the genome. While most of the identified SINEs exist in many copies within the *P. infestans* genome, the genomic DNA sequences immediately flanking the SINEs were observed to be similar for each copy of each element. One possible explanation for this observation is that each SINE has inserted into a different, larger repetitive element, either in a burst of proliferation targeted to the repeat sequence, or as a single insertion into a genomic sequence that has subsequently amplified by other means. Such a situation has also been observed in the rice blast fungus, *M. grisea*, where *MgSINE* was discovered within the *Pot1* retrotransposon (Kachroo et al., 1995). Similarly, the *A. thaliana AtSN2-1* SINE was also found to be located within a much larger repeat unit of 12 kb (Myouga et al., 2001). The observation that the *P. infestans* SINEs all appear to be located within other repeat sequences, and not within the 3′ or 5′ UTRs, or coding regions, of available ESTs, suggests that gene disruption or modification by SINE insertion is not leading to major phenotypic changes in this organism. However, the wider genomic context of SINE insertion in *P. infestans*, potentially leading to genome rearrangements, is presently unknown.

In other organisms, expression of some SINEs is implicated in stress responses (Kimura et al., 2001; Rubin et al., 2002). The *P. infestans* lifecycle progresses through several different cell types or stages, and it is possible that some of the identified elements may be specifically transcribed in these stages or in planta during infection. In previous work, we have identified stress-related genes in cell types formed prior to host infection, and in planta (Avrova et al., 2003). Both *infSINEi* and *h* were transcribed in *P. infestans* mycelium growing in a nutrient-rich medium, and not an environment of cellular stress. It is uncertain if there is any significance in the simultaneous transcription of these SINEs since they are physically located only 350–400 bp apart. The detection of transcripts for *infSINEh* and *i* implies that both may be active retroposons within the *P. infestans* genome. This is also supported by the observation that not all BAC clones that hybridize to *infSINEh* also hybridize to *infSINEi*, and vice versa, suggesting that not all available genomic target sites for *infSINEh* and *i* retroposition have been occupied. It is uncertain which of these elements is older within the *P. infestans* genome. However, *infSINEh* is not found in any other analyzed *Phytophthora* species, but *infSINEi* was apparently present in all species tested except *P. citricola*,

P. citrophthora, *P. ramorum*, and *P. sojae*. This suggests that *infSINEi* is the older element, and *infSINEh* more recent, possibly having proliferated late in the speciation process among very few, closely related *Phytophthora* species. The presence of direct repeats for both of these SINEs can also provide some indication of the age of these elements. That is, direct repeats surrounding *infSINEi* were typically partially degraded, although the length of the repeat was preserved, as may occur through nucleotide substitution over time. In comparison, *infSINEh* frequently possessed intact direct repeats of 11 bp, as may be expected if its occurrence within the genome was more recent.

Sequence similarity between the predicted *P. infestans* SINEs was not obvious, and all of the SINEs appear to have arisen from independent tRNA-related sequences. The observed diversity of tRNA-related sequences from the predicted *P. infestans* SINEs indicates independent bursts of SINE retroposition activity in the evolutionary history of *P. infestans* and other *Phytophthora* species. This process may be ongoing, as demonstrated by the transcriptional activity of *infSINEh* and *i*.

SINE proliferation has been hypothesized to involve utilization of the retropositional process provided by LINEs. The 3′ sequence of some SINEs shares sequence homology with the 3′ sequence of LINEs and this is thought to aid retroposition of the SINE (Ogiwara et al., 2002; Weiner, 2002). Although the aim of this research was not to identify LINEs, sequences that may be characteristic of LINEs, such as reverse transcriptase and endonuclease, can be found in the *P. infestans* EST databases. However, none of the 3′ sequences from any of the SINEs identified for *P. infestans* exhibits any similarity to these sequences. In primates, the *Alu* SINE superfamily is considered to rely on LINE *L1* for retroposition, although there is no obvious primary sequence similarity between these elements, and many copies of *Alu* possess a polyadenosine (poly(A)) tail (reviewed in Deininger et al., 2003). None of the *P. infestans* SINEs possesses a poly(A) tract longer than A_5 in the 3′ region for reverse transcription. The lack of poly(A) tracts in the 3′ tRNA-unrelated region of the *P. infestans* may be partly explained by the apparent lack of transcription for all but *infSINEh* and *i*. That is, once a SINE is no longer transcribed, any poly(A) tract that did exist will be lost through random nucleotide substitution through generations and time. For active SINEs *infSINEh* and *i* that lack a poly(A) tract, the mechanism by which retroposition occurs is therefore not obvious, but may involve trans-mobilization by other retrotransposons.

Multimeric SINEs have been reported from several organisms, for example the *Alu* SINEs from primates (Quentin, 1992), *twin* from the mosquito *Culex pipiens* (Feschotte et al., 2001), and *CYN* from dermopterans (Schmitz and Zischler, 2003). The process by which multimerization of the SINE monomers occurred for each of these SINEs is unknown. Feschotte et al. (2001) proposed a model for the generation of *twin* in *C. pipiens* where the SINE arose through an unprocessed dimeric tRNA precursor. Alternatively, multimeric SINEs could arise through retroposition of a tRNA-related SINE near a tRNA, or preferential SINE insertion into existing SINE insertions (Piskurek et al., 2003). Dicistronic tRNAs that are processed post-transcription have been described in the yeast, *Saccharomyces cerevisiae* (Schmidt et al., 1980). It could not be ascertained if the two distinct *P. infestans* tRNAs in *infSINEz* were transcribed as a dicistron, and processed post-transcription into two functional tRNAs; transcription of the entire tRNA dicistron or the individual tRNAs could not be shown. This does not necessarily signify that *infSINEz* represents a SINE, but may be two tRNA pseudogenes. The dicistron tRNA structure was also preserved in other species of *Phytophthora*, such as *P. sojae*, *P. cinnamomi*, *P. citricola*, and *P. megasperma*.

A multimeric repeat of up to four similar tRNA-related SINEs is characteristic of *infSINEr*, in which each copy of the tRNA-related monomer is apparently self-contained. A curious aspect of this element is the similarity of the genomic region beyond the 3′ tRNA-unrelated sequence of the final monomer to a *Gypsy*-like retroelement previously described by Judelson (2002). In this scenario, it is possible that the tRNA-related SINEs may have used the *Gypsy*-like element to proliferate in the *P. infestans* genome. This could have occurred if *infSINEr* exhibited any insertion site specificity, targeting it to *Gypsy*-like sequences. Alternatively, *infSINEr* may also have been carried with *Gypsy*-like retroelements when they have been active in *P. infestans*.

This study has presented the first directed search for SINEs in oomycete genomic DNA. Future experiments will be required to more fully assess their distribution among *Phytophthora* species, their present activity (if any), and assess their organization in context of the entire genome. SINEs that were widely distributed among the different *Phytophthora* species (*infSINEz* for example), and therefore evolving slowly, could be used to examine speciation within the genus. Potentially active SINEs, such as *infSINEh* and *i*, could be further developed and used as multi-locus tools in population genetics of *P. infestans* since it is probable that insertional polymorphisms (presence/absence) do exist for these SINEs. Furthermore, any future demonstration that *P. infestans* SINEs function to accelerate genome evolution may in part yield insights into the notorious variability of this plant pathogen.

Acknowledgments

Financial support for S.W., A.A., and L.P. was provided by the Scottish Executive Environment and

Rural Affairs Department (SEERAD). O.L. was supported at SCRI by a National Science Foundation (USA) Research Collaboration Network grant *Phytophthora* Training Internship (B. Tyler, Virginia Bioinformatics Institute, Virginia Polytechnic and State University, USA). S.W. thanks Amar Kumar at SCRI for helpful discussions, and critical reading of the manuscript. This research was made possible through the *Phytophthora* Genomics Consortium funded by Syngenta.

References

Avrova, A.O., Venter, E., Birch, P.R.J., Whisson, S.C., 2003. Profiling and quantifying differential gene transcription in *Phytophthora infestans* prior to and during the early stages of potato infection. Fungal Genet. Biol. 40, 4–14.

Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., Doolittle, W.F., 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290, 972–977.

Borodulina, O.R., Kramerov, D.A., 1999. Wide distribution of short interspersed elements among eukaryotic genomes. FEBS Lett. 457, 409–413.

Cheng, C., Motohashi, R., Tsuchimoto, S., Fukuta, Y., Ohtsubo, H., Ohtsubo, E., 2003. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Mol. Biol. Evol. 20, 67–75.

Cooke, D.E., Drenth, A., Duncan, J.M., Wagels, G., Brasier, C.M., 2000. A molecular phylogeny of *Phytophthora* and related oomycetes. Fungal Genet. Biol. 30, 17–32.

Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16, 10881–10890.

Deininger, P.L., Moran, J.V., Batzer, M.A., Kazazian, H.H., 2003. Mobile elements and mammalian genome evolution. Curr. Opin. Genet. Dev. 13, 651–658.

Duncan, J.M., 1999. *Phytophthora*—an abiding threat to our crops. Microbiol. Today 26, 114–116.

Feschotte, C., Fourrier, N., Desmons, I., Mouches, C., 2001. Birth of a retroposon: the *Twin* SINE family from the vector mosquito *Culex pipiens* may have originated from a dimeric tRNA precursor. Mol. Biol. Evol. 18, 74–84.

Fry, W.E., Goodwin, S.B., 1997. Re-emergence of potato and tomato late blight in the United States. Plant Dis. 81, 1349–1357.

Garelik, G., 2002. Taking the bite out of potato blight. Science 298, 1702–1704.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Judelson, H.S., 2002. Sequence variation and genomic amplification of a family of *Gypsy*-like elements in the oomycete genus *Phytophthora*. Mol. Biol. Evol. 19, 1313–1322.

Judelson, H.S., Randall, T.A., 1998. Families of repeated DNA in the oomycete *Phytophthora infestans* and their distribution within the genus. Genome 41, 605–615.

Kachroo, P., Leong, S.A., Chattoo, B.B., 1995. *Mg-SINE*: a short interspersed nuclear element from the rice blast fungus, *Magnaporthe grisea*. Proc. Natl. Acad. Sci. USA 92, 11125–11129.

Kamoun, S., 2003. Molecular genetics of pathogenic oomycetes. Eukaryot. Cell 2, 191–199.

Kamoun, S., Hraber, P., Sobral, B., Nuss, D., Govers, F., 1999. Initial assessment of gene diversity for the oomycete pathogen *Phytophthora infestans* based on expressed sequences. Fungal Genet. Biol. 28, 94–106.

Kimura, R.H., Choudary, P.V., Stone, K.K., Schmid, C.W., 2001. Stress induction of *Bm1* RNA in silkworm larvae: SINEs, an unusual class of stress genes. Cell Stress Chaperones 6, 263–272.

Kumar, A., Bennetzen, J.L., 1999. Plant retrotransposons. Annu. Rev. Genet. 33, 479–532.

Lam, S.T., 2001. *Phytophthora* genomics consortium. Phytopathology 91, S158.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Makalowski, W., 2003. Not junk after all. Science 300, 1246–1247.

McLeod, A., Smart, C.D., Fry, W.E., 2004. Core promoter structure in the oomycete *Phytophthora infestans*. Eukaryot. Cell 3, 91–99.

Myouga, F., Tsuchimoto, S., Noma, K., Ohtsubo, H., Ohtsubo, E., 2001. Identification and structural analysis of SINE elements in the *Arabidopsis thaliana* genome. Genes Genet. Syst. 76, 169–179.

Ogiwara, I., Miya, M., Ohshima, K., Okada, N., 2002. V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. Genome Res. 12, 316–324.

Piskurek, O., Nikaido, M., Boeadi, Baba, M., Okada, N., 2003. Unique mammalian tRNA-derived repetitive elements in dermopterans: the t-SINE family and its retrotransposition through multiple sources. Mol. Biol. Evol. 20, 1659–1668.

Quentin, Y., 1992. Origin of the *Alu* family: a family of *Alu*-like monomers gave birth to the left and the right arms of the *Alu* elements. Nucleic Acids Res. 20, 3397–3401.

Randall, T.A., Judelson, H.S., 1999. Construction of a bacterial artificial chromosome library of *Phytophthora infestans* and transformation of clones into *P. infestans*. Fungal Genet. Biol. 28, 160–170.

Ribeiro, O.K., 1978. A Source Book of the Genus *Phytophthora*. Cramer, Vaduz, Liechtenstein.

Rubin, C.M., Kimura, R.H., Schmid, C.W., 2002. Selective stimulation of translational expression by *Alu* RNA. Nucleic Acids Res. 30, 3253–3261.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, second ed. Cold Spring Harbour Press, USA.

Schmidt, O., Mao, J., Ogden, R., Beckmann, J., Sakano, H., Abelson, J., Soll, D., 1980. Dimeric tRNA precursors in yeast. Nature 287, 750–752.

Schmitz, J., Zischler, H., 2003. A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates. Mol. Phylogenet. Evol. 28, 341–349.

Shattock, R.C., 2002. *Phytophthora infestans*: populations, pathogenicity and phenylamides. Pest Manag. Sci. 58, 944–950.

Sogin, M.L., Silberman, J.D., 1998. Evolution of the protists and protistan parasites from the perspective of molecular systematics. Int. J. Parasitol. 28, 11–20.

Tooley, P.W., Garfinkel, D.J., 1996. Presence of *Ty1-copia* group retrotransposon sequences in the potato late blight pathogen *Phytophthora infestans*. Mol. Plant Microbe Interact. 9, 305–309.

Tooley, P.W., Therrein, C.D., 1987. Cytophotometric determination of the nuclear DNA content of 23 Mexican and 18 non-Mexican isolates of *Phytophthora infestans*. Exp. Mycol. 11, 19–26.

Ullu, E., Tschudi, C., 1984. *Alu* sequences are processed 7SL RNA genes. Nature 312, 171–172.

van der Lee, T., de Witte, I., Drenth, A., Alfonso, C., Govers, F., 1997. AFLP linkage map of the oomycete *Phytophthora infestans*. Fungal Genet. Biol. 21, 278–291.

van der Lee, T., Robold, A., Testa, A., van 't Klooster, J.W., Govers, F., 2001. Mapping of avirulence genes in *Phytophthora infestans* with amplified fragment length polymorphism markers selected by bulked segregant analysis. Genetics 157, 949–956.

Voglmayr, H., Greilhuber, J., 1998. Genome size determination in peronosporales (Oomycota) by Feulgen image analysis. Fungal Genet. Biol. 25, 181–195.

Waugh, M., Hraber, P., Weller, J., Wu, Y., Chen, G., Inman, J., Kiphart, D., Sobral, B., 2000. The *Phytophthora* genome initiative database: informatics and analysis for distributed pathogenomic research. Nucleic Acids Res. 28, 87–90.

Weiner, A.M., 2002. SINEs and LINEs: the art of biting the hand that feeds you. Curr. Opin. Cell Biol. 14, 343–350.

Whisson, S.C., van der Lee, T., Bryan, G.J., Waugh, R., Govers, F., Birch, P.R., 2001. Physical mapping across an avirulence locus of *Phytophthora infestans* using a highly representative, large-insert bacterial artificial chromosome library. Mol. Genet. Genomics 266, 289–295.

Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N., Machida, Y., 1993. Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc. Natl. Acad. Sci. USA 90, 6562–6566.