Microbiology (2007), **153**, 747–759 DOI 10.1099/mic.0.2006/002220-0

A novel non-protein-coding infection-specific gene family is clustered throughout the genome of *Phytophthora infestans*

Anna O. Avrova, Stephen C. Whisson, Leighton Pritchard, Eduard Venter,† Sergio De Luca, Ingo Hein and Paul R. J. Birch

Plant Pathology Programme, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

Correspondence
Paul R. J. Birch
pbirch@scri.sari.ac.uk

Received 6 September 2006
Revised 1 November 2006
Accepted 8 November 2006

Phytophthora infestans is the cause of late blight, a devastating and re-emerging disease of potato. Significant advances have been made in understanding the biology of *P. infestans*, and in the development of molecular tools to study this oomycete. Nevertheless, little is known about the molecular bases of the establishment or development of disease in this hemibiotrophic pathogen. Suppression subtractive hybridization (SSH) was used to generate cDNA enriched for sequences upregulated during potato infection. To identify pathogen-derived cDNAs, and eliminate host sequences from further study, SSH cDNA was hybridized to a *P. infestans* bacterial artificial chromosome library. A new gene family was identified called *Pinci*1, comprising more than 400 members arranged in clusters of up to nine copies throughout the *P. infestans* draft genome sequence. Real-time RT-PCR was used to quantify the expression of five classes of transcript within the family, relative to the constitutively expressed *Piact*A gene, and it revealed them to be significantly upregulated from 12 to 33 h post-inoculation, a period defining the biotrophic phase of infection. Computational analysis of sequences suggested that transcripts were non-protein coding, and this was confirmed by transient expression of FLAG-tagged ORFs in *P. infestans*.

INTRODUCTION

Plant disease caused by pathogenic microbes has had a significant impact on human health and economies, and none more so than the first epidemics of the plant disease late blight, caused by the oomycete *Phytophthora infestans*, which devastated the potato crops of north-west Europe in the 19th century. To this day, *P. infestans* remains the most serious constraint to potato production worldwide, with estimated losses due to disease and control measures exceeding $US 5 billion annually (Duncan, 1999). Despite its economic importance, the fundamental molecular processes underpinning *P. infestans* development and pathogenicity, and the factors that restrict its host range, are poorly understood (reviewed by Birch & Whisson, 2001; Kamoun 2003, 2006), and yet such knowledge could inform the development of novel strategies to combat this disease.

P. infestans has a filamentous growth habit, similar to many fungi, but belongs to the oomycetes, a group of filamentous organisms more closely related to the stramenopiles (Baldauf, 2003). Although outwardly similar to fungi, oomycetes exhibit clear differences to true fungi; these differences include: diploidy, cell walls containing cellulose, and differing gene promoter element structure and transcriptional initiation sites (reviewed by Kamoun 2003, 2006). Thus, mechanisms of fungal pathogenesis of plants may not apply to the oomycetes, which may have developed unique mechanisms to parasitize plant hosts.

A number of developmental stages are required to complete the *P. infestans* infection cycle, including the formation of zoospores, their encystment, production of a germ tube, and the development of appressoria, primary and secondary hyphae, haustoria and sporangiophores. *P. infestans* is a hemibiotroph, with an initial biotrophic phase of interaction with potato, followed by a necrotrophic phase. In the former phase, after cyst germination and appressorium formation, potato epidermal cells are penetrated at 16 h post-inoculation (hpi), and an infection vesicle is produced. At 22 hpi, one or two haustoria are produced in each cell encountered by ramifying hyphae. At 46 hpi, haustoria are

†Present address: Department of Botany and Plant Biotechnology, University of Johannesburg, PO Box 524, Auckland Park, South Africa.

Abbreviations: BAC, bacterial artificial chromosome; hpi, h post-inoculation; HSP, high-scoring pair; LCR, locus control region; miRNA, microRNA; ncRNA, non-coding RNA; SSH, suppression subtractive hybridization.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are EF091715–EF091740.

Supplementary Figs S1–S5 and Table S1 are available with the online version of this paper.

rarely seen at the infection site, and sporangiophores begin to emerge through stomata (e.g. Vleeshouwers *et al.*, 2000) as the pathogen progresses from biotrophy to necrotrophy. At 72 hpi, leaf necrosis and pathogen sporulation are clearly visible to the eye.

Many genes with a key role in pathogenicity, and certainly those involved in forming infection-stage-specific cell structures, will be upregulated during the potato–*P. infestans* association. However, during the early stages of infection, identifying pathogen gene transcripts from a mixture that contains predominantly host RNA species presents a significant challenge. That is, sequencing of cDNA clones from interaction libraries may yield only very low numbers of pathogen sequences, and these typically represent transcripts that are highly abundant and/or constitutively expressed. Several approaches exist for isolating such differentially expressed genes (reviewed by Birch *et al.*, 2003). A PCR-based method called suppression subtractive hybridization (SSH) has been used to isolate potato genes upregulated following *P. infestans* challenge (Birch *et al.*, 1999; Avrova *et al.*, 1999, 2004; Beyer *et al.*, 2001), and to identify *P. infestans* genes that are upregulated in the interaction with potato (Beyer *et al.*, 2002). However, despite the ability to enrich for pathogen transcripts, SSH cDNA libraries constructed from host–oomycete interactions still yield a majority proportion of host sequences (Bittner-Eddy *et al.*, 2003). To obviate the need to distinguish between *P. infestans* and potato gene sequences, Beyer *et al.* (2002) induced mycelium by contact with the host plant, and then separated it from the host tissue prior to SSH. In contrast, here we exposed potato leaf material to zoospores, and allowed infection to proceed before using SSH to generate cDNA enriched for sequences upregulated during the interaction. To separate pathogen sequences from those of the plant, subtracted cDNA was hybridized to a *P. infestans* bacterial artificial chromosome (BAC) library (Whisson *et al.*, 2001). A single gene family, with its members arranged in clusters throughout the *P. infestans* genome, was identified, and expression of members of this family was investigated prior to infection, and during biotrophic and necrotrophic phases of the interaction, using real-time RT-PCR. Evidence is presented that expressed members of the family produce novel, tightly regulated, non-coding RNAs (ncRNAs). Accordingly, this gene family has been named *Pinci*1, for *P. infestans* non-coding infection-specific family 1.

METHODS

Growth of *P. infestans*, potato plants and plant inoculation. Growth of potato plants and *P. infestans* isolate 88069, and plant inoculation, were carried out as described by Grenville-Briggs *et al.* (2005). Lifecycle stages of *P. infestans* (axenic cultured mycelium, sporangia, zoospores, germinated cysts without appressoria) were also prepared as described by Grenville-Briggs *et al.* (2005).

RNA isolation, SSH, and real-time RT-PCR expression analysis. RNA extraction was performed using the Qiagen RNeasy Plant Mini Kit, following the manufacturer's protocol. SSH, using the PCR-Select cDNA subtraction kit (Clontech), was performed to generate a cDNA library enriched for sequences upregulated during infection. cDNA generated from susceptible cv. Bintje leaves combined from the time points after inoculation with *P. infestans* was used as a tester. The driver was a 1 : 1 mixture of cDNA from uninoculated Bintje leaves, and *P. infestans* mycelium grown in rye broth for 10 days, to subtract constitutively expressed potato and *P. infestans* genes, respectively. A very stringent 1 : 1200 ratio of tester : driver was used for subtraction. For gene expression analysis, first-strand cDNA synthesis, and SYBR green real-time RT-PCR assays, were carried out as previously described (Avrova *et al.*, 2003).

Analyses of *P. infestans* BAC clones. Transfer of the *P. infestans* BAC library to nylon membranes, and colony and Southern hybridization, were as described by Whisson *et al.* (2001). To allow for DNA probe complexity, 100 ng SSH-derived cDNA was labelled with 100 μCi (3.7 MBq) [α-^{32}P]dCTP using the High Prime labelling kit (Roche), scaling up the manufacturer's protocol by a factor of two. Preparation of single-sequence hybridization probes used 25 ng probe DNA, 50 μCi (1.85 MBq) [^{32}P]dCTP, and a single reaction of the High Prime labelling kit. BAC clone plasmid preparations, and estimation of insert sizes, were as described by Whisson *et al.* (2001, 2005). For fingerprinting, BAC clones were restriction digested with *Hin*dIII, and analysed by 1 % agarose gel electrophoresis. Gel images were recorded digitally from the UV transilluminator, and data were analysed with GelCompar version 4.1 software (Applied Maths), using a 1 kb DNA ladder (New England Biolabs) as a reference marker to normalize tracks from different gels. The UPGMA clustering method was used to align different fingerprints. Insert DNA from BAC clones was subcloned into vector pGEM-3Z (digested with *Bam*HI and alkaline phosphatase; Promega), as described by Bell *et al.* (2002), following partial digestion with *Sau*3AI (Promega). A total of 1152 recombinant transformed clones were selected, and stored in 384-well microtitre plates in freezing medium (Whisson *et al.*, 2001) containing ampicillin, at −70 °C, until needed. BAC subclone plasmids were prepared using the Qiagen Plasmid Miniprep kit, and sequenced in both directions with SP6 and T7 primers using the Perkin Elmer ABI PRISM BigDye Terminator 3.1 cycle sequencing kit, manufacturer-recommended thermal cycling conditions, and ABI model 377 DNA sequencer. BAC ends were sequenced using the same primer, sequencing chemistry and apparatus, but with a higher (98 °C for 5 min) initial melt in the cycle sequencing program. To verify contigs, PCR from BAC ends with specific primers used the following thermal cycling conditions: initial melt at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s, with a final extension at 72 °C for 10 min. Each 20 μl PCR contained 0.5 U *Taq* polymerase (Promega), 1 × reaction buffer (reaction buffer B, containing $MgCl_2$; Promega), 250 μM deoxynucleotide triphosphates (Promega), 1 μM forward and reverse primers, and 1–5 ng BAC DNA.

Transcript length determination by RT-PCR, and cloning of RT-PCR products. Transcriptional termination was determined by 3′ RACE using an oligo dT reverse primer and four different forward primers (A4-23-10F, A4F17F2, GSPA4 and GSPA4N; see Supplementary Table S1) specific to different *Pinci*1 family members. The transcriptional start of *Pinci*1-1 and *Pinci*1-5 was determined by RT-PCR using oligonucleotide primers Nci1-4F, Nci1-1F, and A4-23-10F, which were 5′ of the predicted ORFs, in combination with either oligo dT or A4TAQR1 reverse primers (Supplementary Table S1, Fig. 2). RT-PCR conditions used cDNA from the *P.-infestans*–potato interaction 48 hpi as a template, and were as described for PCR from BAC clones. RT-PCR products were cloned into the pGEMTeasy vector (Promega), according to the manufacturer's protocol, and sequenced as described above.

DNA sequence analyses. DNA sequence trace editing and CAP3 (Huang & Madan, 1999) contig assembly were performed using the

Bioedit program (Hall, 1999). DNA sequence similarity searches were carried out using BLAST at NCBI (www.ncbi.nlm.nih.gov/blast/), the Phytophthora Functional Genomics Database (www.pfgd.org/), PhytophthoraDB (http://phytophthora.vbi.vt.edu/develop/blast_main.php), and locally with BLAST 2.2.10. ORF finder (www.ncbi.nlm.nih.gov/gorf/gorf.html), DNA Tools (http://biology.semo.edu/cgi-bin/dnatools.pl), and in-house Python scripts, were used to predict ORFs in the sequences obtained. The Testcode algorithm (Fickett, 1982; http://bioinformatics.org/sms/index.html) was used to identify potential protein-coding sequences. Secondary structures for RNA molecules were predicted using the Vienna RNA package RNAfold web interface (www.tbi.univie.ac.at/~ivo/RNA/), and the Sfold RNA secondary structure prediction program (http://sfold.wadsworth.org/srna.pl). In-house Python scripts were used to process BLAST output, identify co-located *Pinci*1 domains, classify *Pinci*1 clusters, associate *P. infestans* EST-derived unigenes with *Pinci*1 sites on the draft *P. infestans* supercontig sequences, and identify putative RXLR sequences and their locations on the supercontigs. Putative RXLR sequences were identified as ORFs of length greater than 100 aa, encoding a sequence containing a signal peptide as predicted by SignalP 3.0a (Bendtsen *et al.*, 2004), with an RXLR motif within 100 aa of the signal peptide cleavage site, as in Bhattacharjee *et al.* (2006). EST and *Pinci*1 features were marked on the draft genome sequence using GenomeDiagram (Pritchard *et al.*, 2006). Sequences of BAC clones were obtained from GenBank (BAC addresses in parentheses): AC172889 (25A6), AC172799 (25B1), AC173112 (25D1), AC173111 (25E8), AC172988 (25H7), AC154096 (31F6), AC173549 (60G5), AC172798 (60H9), AC147182 (9I20), AY830090 (11A5), AC146943 (14M19), AC146983 (14P22), AC147181 (25C5), AC147180 (26O7), AC147005 (42H10), AC146942 (61F2), AJ893356 (35J4), AJ893357 (49P21), AY497062 (21G17), AC147508 (19M21), AC175050 (60H1) and AC147544 (34A11). The draft *P. infestans* genome (assembly version 0.5) was downloaded from www.broad.mit.edu/annotation/genome/phytophthora_infestans/Downloads.html, and formatted for local BLAST searches with formatdb 2.2.10. The *P. infestans* EST-derived unigene set used was that described by Randall *et al.* (2005) in their supplementary information.

Cloning of *Pinci*1 and *Piact*A FLAG-tagged ORFs in *P. infestans*. The *Pinci*1 and *Piact*A (GenBank accession no. M59715) ORFs were PCR amplified with primers annealing to the putative start and stop codons of each gene (see Supplementary Table S1). Forward primers incorporated a *Cla*I restriction site, and reverse primers incorporated a sequence encoding the FLAG signal (DYKDDDDK) (Hopp *et al.*, 1988) prior to the stop codon. PCR amplicons were verified by gel electrophoresis, and used as templates in a secondary PCR with the same forward primers in combination with the *Not*IFLAG primer, incorporating a *Not*I restriction site. The amplicons were purified, restriction digested with *Not*I and *Cla*I, and cloned into the oomycete constitutive expression vector pTOR, described by Blanco & Judelson (2005), followed by electroporation into DH10B electrocompetent cells (Invitrogen). Insert integrity, and correct reading frame orientation of the cloned inserts, were verified by sequencing.

Transient transformation of *P. infestans*, and analysis of transformants. Microprojectile bombardment was used to transfer plasmid DNA for transient expression of FLAG-tagged *Pinci*1 and *Piact*A vectors in *P. infestans*. Sporangia were harvested from 10-day-old *P. infestans* cultures, and allowed to germinate on Cyclopore hydrophilic polycarbonate membranes (Whatman), which were placed on Rye agar for 24 h. Microprojectile bombardment conditions were as described by Cvitanich & Judelson (2003).

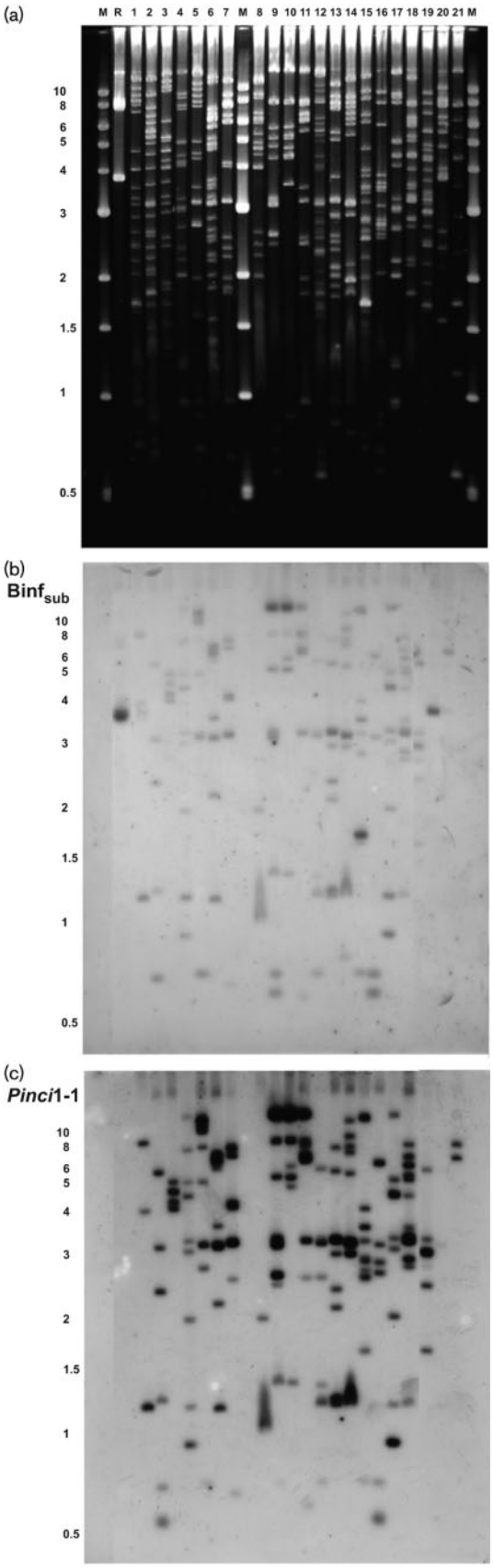
For analysis of transiently transformed *P. infestans* after microprojectile bombardment, RNA (as described above) and proteins (Latijnhouwers *et al.*, 2004) were prepared from the cultures that had been growing for 2 days on rye agar containing 5 μg geneticin ml^{-1} (Sigma). Synthesized first-strand cDNA was used for RT-PCR, with the different *Pinci*1 and *Piact*A ORF forward primers in combination with the *Not*IFLAG primer. RT-PCR conditions were as described for PCR from BAC end sequences. Protein concentrations were determined with Bradford Reagent (Bio-Rad). All protein samples were concentrated 10-fold using Microcon YM-3 centrifugal filters (Millipore), and equal concentrations were fractionated on Nu-PAGE 4–12 % Bis-Tris gel (Invitrogen), transferred to Hybond ECL nitrocellulose membranes (Amersham), and FLAG-tagged protein was detected with anti-FLAG M2 monoclonal antibody (Sigma-Aldrich).

RESULTS

Identification of BAC clones containing *P. infestans* sequences that are upregulated during infection

SSH was used to enrich for cDNAs derived from genes upregulated during the compatible interaction between *P. infestans* and potato cv. Bintje. Use of cDNA from axenic *P. infestans* mycelium as a radioactively labelled probe revealed strong hybridization to unsubtracted cDNA at 72 hpi, indicating many common transcripts, but there was little detectable hybridization to the SSH material; this demonstrated successful removal of most of these transcripts by the highly stringent subtraction conditions used (see Supplementary Fig. S1). The SSH cDNA can be predicted to be a mixture of cDNAs from both host and pathogen. Indeed, preliminary sequencing of cloned SSH cDNA fragments yielded only sequences of plant origin (data not shown). To specifically identify pathogen sequences, and effectively exclude host-derived cDNAs from further analyses, radioactively labelled $Binf_{sub}$ cDNA was hybridized to a BAC library of *P. infestans* genomic DNA (Whisson *et al.*, 2001). The $Binf_{sub}$ probe hybridized to 100 clones in the BAC library. Plasmid preparations of 21 representative BAC clones were restriction digested with *Hin*dIII to release insert DNA (Fig. 1A), Southern blotted, and hybridized with $Binf_{sub}$ cDNA. As expected, the probe hybridized to all of the clones. However, hybridization was observed to multiple restriction fragments, which ranged in size from 500 bp to more than 10 kb for each BAC clone (Fig. 1c), raising the question of whether there were multiple genes on each BAC clone that were co-ordinately upregulated during infection.

A BAC clone (12P14) was selected that contained nine *Hin*dIII fragments that hybridized to $Binf_{sub}$ cDNA. The insert of this BAC was subcloned and hybridized to $Binf_{sub}$ cDNA. Seventy-six hybridizing clones were sequenced and assembled to reveal three contigs containing a related DNA sequence. PCR primers A4-23-10F and A4TaqR1 (Supplementary Table S1, Fig. 2) were designed to amplify a portion of the DNA sequence that was conserved between these contigs, and this was hybridized to the Southern blot containing 21 restriction-digested BAC clones. Interestingly, the probe hybridized to all of the restriction fragments that had also hybridized to the $Binf_{sub}$ cDNA (Fig. 1c),



suggesting that a single large gene family was represented in the SSH cDNA. The probe did not hybridize to potato genomic DNA (results not shown), confirming that these sequences were derived from the pathogen, and that the host did not contain related sequences. Screening the BAC library by colony hybridization revealed that the probe hybridized to 936 BAC clones. The BAC library represents approximately 10-fold genome coverage (Whisson *et al.*, 2001), implying that related sequences are present in at least 90 regions of the *P. infestans* genome, and probably in multiple copies at each locus.

To deduce relationships between the 100 BAC clones hybridizing to Binf$_{sub}$, *Hin*dIII restriction digestion patterns of each were analysed using GelCompar, and they revealed that the clones fell into a number of restriction pattern groups (see Supplementary Fig. S2A). To test whether BAC clones containing overlapping regions of the *P. infestans* genome were clustered by the GelCompar analysis, the ends of BAC clones 56A9, 63J7 and 38J7 were sequenced, and primers (see Supplementary Table S1) were designed to seek, by PCR, equivalent sequences in all Binf$_{sub}$ hybridizing BAC clones. Primers derived from 63J7 amplified products from that BAC clone only. Primers derived from BAC 38J7 amplified products from 38J7 and the related BAC clone 44P17. The 56A9-derived primers PCR amplified a DNA fragment of expected size from 56A9, and seven of the nine BAC clones containing similar restriction patterns that clustered with 56A9 in the GelCompar analysis (see Supplementary Fig. S2A). Clone 56A9 and the nine related BACs were restriction digested with *Hin*dIII, *Pst*I and *Bam*HI, Southern blotted, and hybridized with the A4-23-10-A4TaqR1 probe. The hybridization pattern confirmed that these BAC clones spanned a common region of the genome (see Supplementary Fig. S2B).

Isolation of *Pinci*1 cDNA sequences

*Pinci*1-1 was amplified from cDNA prepared from cv. Bintje leaves at 72 hpi with *P. infestans* zoospores (B72), using A4-23-10F primer in combination with an oligo-dT primer (Fig. 2), and this confirmed that *Pinci*1 sequences were expressed and polyadenylated. The sequenced *Pinci*1-1 cDNA was 100 % identical to one of the original genomic sequences from BAC clone 12P14. BLASTN comparison to the 18 473 EST-derived *P. infestans* unigenes (Randall *et al.*, 2005) revealed a strongly significant match (E value, 10^{-108}) to contig 19259_1, and this independently verified the sequence to be transcribed and polyadenylated.

Fig. 1. *Hin*dIII restriction endonuclease profiling of 21 representative BAC clones hybridizing to the Binf$_{sub}$ probe (from left to right): 29D3, 44N17, 48M24, 44P17, 45L17, 53E7, 38D5, 42H21, 56A9, 56N16, 19J2, 22B13, 35H20, 16J1, 51D6, 12O18, 38J7, 12P14, 65P12, 63J7 and 5H6. M, restriction fragment markers (kb); R, ribosomal clone. (a) *Hin*dIII digests separated on an agarose gel. (b) *Hin*dIII digests hybridized to Binf$_{sub}$. (c) *Hin*dIII digests hybridized to *Pinci*1-1.

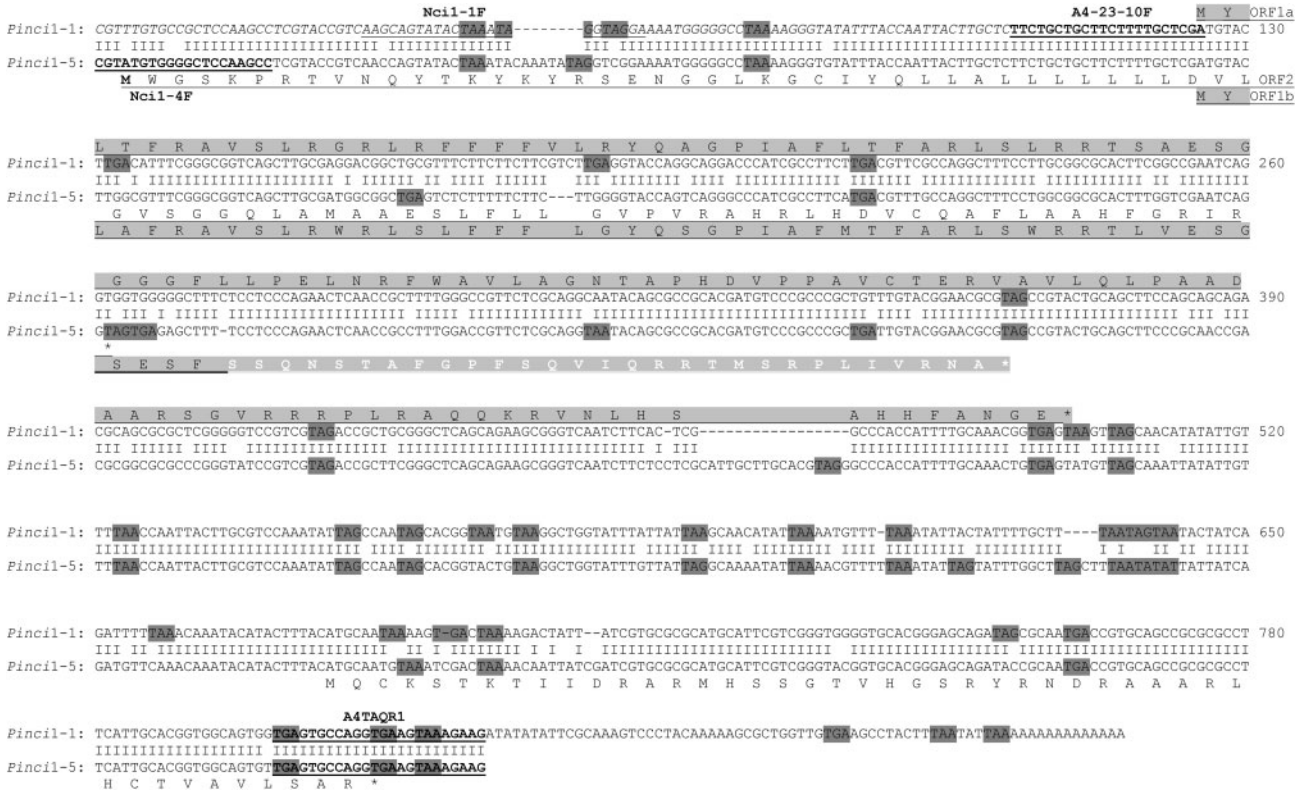


Fig. 2. DNA and translated protein alignment of *Pinci*1-1 and *Pinci*1-5. Stop codons are shown in dark grey boxes; similar ORFs are highlighted with the same shading; asterisks indicate the end of an ORF; and frameshifts are indicated by a change in the colouring of the amino acid letter. Underlined sequences represent the oligonucleotide primer sites (indicated in bold; sequences in Supplementary Table S1) used in RT-PCR for determination of transcript length.

*Pinci*1-2 and *Pinci*1-3 were obtained by cloning the PCR products amplified from B72 cDNA using A4-23-10F and A4-23-10R primers (Supplementary Table S1). To obtain the 3′ end of *Pinci*1-3 cDNA, we used 3′ RACE, and primer A4F17F2 in combination with the oligo dT primer. *Pinci*1-4 was amplified from B72 cDNA using GSPA4 primer in combination with the oligo-dT primer, followed by nested PCR with GSPA4N and oligo dT primers. *Pinci*1-5 to *Pinci*1-25 were amplified from B72 cDNA using primer Nci1-4F (see Supplementary Table S1), which anneals to the 5′ end of *Pinci*1-4, in combination with A4TAQR1 primer (see Supplementary Table S1) specific to the 3′ end of *Pinci*1-1. The *Pinci*1-5 to *Pinci*1-25 PCR products were obtained only from cDNA synthesized from B72 RNA, but not when B72 RNA alone was used as substrate in the PCR; this confirmed that each of these sequences was expressed, and not derived from genomic DNA contamination of RNA preparations.

Supplementary Fig. S3 provides an alignment of the nucleotide sequences of *Pinci*1-1 to *Pinci*1-25. Members of the *Pinci*1 family possessed scattered single-nucleotide polymorphisms, and insertions and deletions of different sizes, leading to frame shifts and the introduction of stop codons. Thus, although they are similar at the RNA level, they potentially code for two different types of short ORF with numerous amino acid substitutions, some of which are truncated to 29–50 aa, with the largest ORF encoding 158 aa. Supplementary Fig. S4 shows alignments of one of the ORFs present in 19 out of 25 *Pinci*1 cDNAs.

An alignment of *Pinci*1-1 and *Pinci*1-5 (Fig. 2) indicates the potential polypeptides encoded by each gene. A range of oligonucleotide primers 5′ to the predicted major ORFs were designed to anneal specifically to the *Pinci*1-1 genomic DNA sequence from BAC clone 12P14, and they were used in RT-PCR, in combination with an oligo dT or A4TAQR1 primer, to indicate the 5′ end of the *Pinci*1-1 transcript. Amplification of *Pinci*1-1 from interaction cDNA was detected with the A4-23-10F/oligo-dT primer combination, but not with the Nci1-1F/A4TAQR1 primer combination, revealing a transcript of at least 847 bp (Fig. 2). In contrast, transcription of the *Pinci*1-5 sequence was shown to start at least 100 bp further upstream relative to *Pinci*1-1, as indicated by amplification from cDNA using Nci1-4F in combination with A4TAQR1 (Fig. 2). No transcription start sequences similar to those commonly found 50–100 bp upstream of *P. infestans* genes (Kamoun, 2003) were evident in the *Pinci*1-1 5′ flanking sequence. *Pinci*1-1 contains a single ORF (ORF 1a) that is shared with *Pinci*1-5 (ORF 1b). However, in *Pinci*1-5, a frameshift would truncate the

polypeptide, and lead to an alternative C-terminal sequence (Fig. 2). *Pinci*1-5 contains a second ORF (ORF 2) that overlaps the first. Neither class of ORF possesses a translation start site similar to the eukaryotic consensus ACCATGA seen in *P. infestans* genes (Kamoun, 2003), and both result in 3′ untranslated regions longer (at least 375 bases) than observed in typical *P. infestans* mRNAs (Kamoun, 2003).

Genomic distribution and organization of *Pinci*1 sequences

A stringent BLASTN search with *Pinci*1-1 (Fig. 2) as the query identified 1474 high-scoring pairs (HSPs) to the draft *P. infestans* genome sequence on 243 supercontig sequences. HSPs with a length of less than 200 bp were excluded to leave a total of 517 individual matches on 135 supercontig sequences (see Supplementary Fig. S5). Most matches were to either a 5′ domain of 445 nt of the *Pinci*1 sequence, or a 3′ domain of the same sequence, from position 449 onwards (Fig. 3). Two-hundred and twenty of these individual domain matches corresponded to a correct orientation of 5′ and 3′ domains, separated by fewer than 400 bp on the parent supercontig sequences, and they were thus considered to be 110 full-length *Pinci*1 sequences, with an internal spacer of variable length between two conserved domains. The most common (47/110) spacer length was 23 bp, and the next most frequent (10/110) was 127 bp. Six full-length matches to the *Pinci*1 sequence not containing an internal spacer were found. Most BLAST hits (297/517) on the supercontig sequences were to unpaired 3′ or 5′ domains of the *Pinci*1 sequence (Fig. 3). A total of 407 *Pinci*1 (110 full-length plus 297 single-domain) sequences were thus identified within the draft genome sequence.

The BLAST hits to *Pinci*1 sequences on the *P. infestans* supercontigs were seen to occur frequently in clusters of up to nine hits, with a mean separation of approximately 4 kb between consecutive *Pinci*1 matches. Typically, such clusters contained a single match to *Pinci*1, comprising both 3′ and 5′ domains, and the remainder of the cluster was made up with matches to single domains, either 3′ or 5′, of the *Pinci*1 sequence. Most (224/407) *Pinci*1 matches occurred in clusters of at least two units, with separation of less than 5 kb on the shared supercontig sequence (see Supplementary Fig. S5).

*Pinci*1 members are upregulated during the biotrophic phase of infection

Real-time RT-PCR analysis was used to characterize expression of the *Pinci*1 sequences in *P. infestans* mycelium, sporangia, zoospores, germinating cysts, and during the infection of potato cv. Bintje at 12, 24, 33, 48, 56 and 72 hpi. Microscopic analysis (Fig. 4) indicated that the timing of *P. infestans* development during these infections closely followed the events described by Vleeshouwers *et al.* (2000). Thus, at 12 hpi, germinating cysts, appressoria and infection hyphae were clearly visible (Fig. 4a–c). At 24–33 hpi, infection hyphae (Fig. 4d) and numerous haustoria were visible (Fig. 4e), indicative of the biotrophic phase, whereas few haustoria were visible at the infection site at 48 hpi (Fig. 4f), indicating the transition from biotrophy to necrotrophy. At 72 hpi, the necrotrophic phase was well established, with sporulation and necrosis clearly visible (Fig. 4g).

Due to the extensive similarity between *Pinci*1 sequences, and the strict criteria for design of real-time RT-PCR primer pairs, it was not possible to design primers discriminating between each member of the *Pinci*1 family. However, primers (see Supplementary Table S1) were designed to discriminate between subclasses of *Pinci*1 sequences. Thus, primers were designed to anneal to both *Pinci*1-1 and *Pinci*1-5, but not to other *Pinci*1 sequences. Primers specific to *Pinci*1-3 alone, and primers that would anneal to both *Pinci*1-3 and *Pinci*1-4, but not to other *Pinci*1 sequences, were designed. Distinct primer sets were also designed

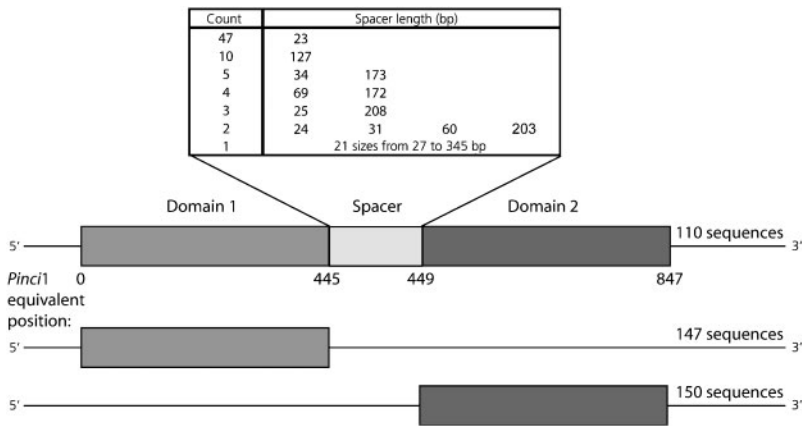


Fig. 3. Schematic representation of the proposed *Pinci*1 domain and spacer structure. Domain 1 (5′) covers bases 1–445 (approximately) of *Pinci*1, and domain 2 (3′) comprises bases 449–847 (approximately) of *Pinci*1. A total of 110 matches to both 5′ and 3′ domains in the correct orientation were observed on the *P. infestans* draft genome supercontigs. These paired matches were separated by a 'spacer' region of between 24 and 345 bp, as indicated in the table at the top of the figure, where a spacer length of 23 bp was most common, with 47 occurrences. A total of 147 unpaired matches to the 5′ domain of *Pinci*1, and 150 unpaired matches to the 3′ domain of *Pinci*1, were also observed on the *P. infestans* supercontigs.

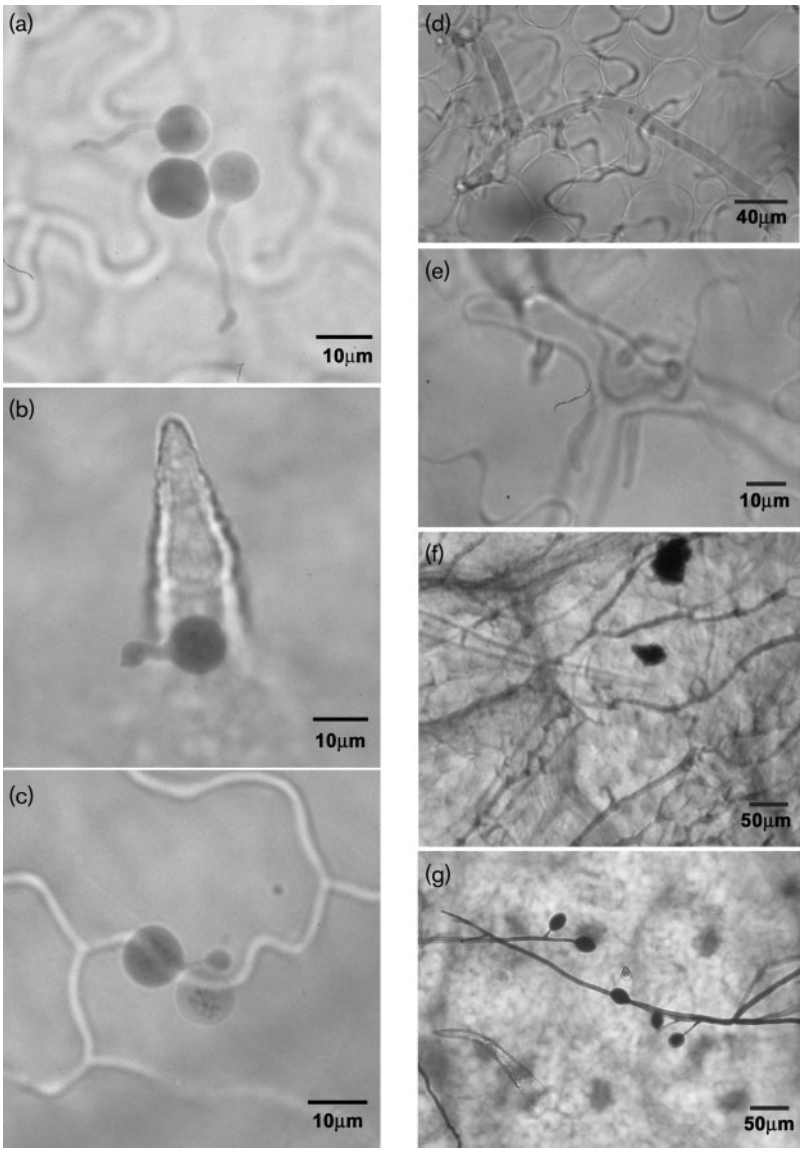


Fig. 4. Trypan blue staining and microscopic assessment of *P. infestans* infection stages on susceptible potato cv. Bintje. At 12 hpi, germinating cysts (a), appressoria (b), and infection vesicles (c) were visible. The large dark shape between the two germinating cysts in (a) is an artifact from trypan blue staining. At 24–33 hpi, infection hyphae (d) were observed spreading out from the infection point, producing numerous digit-like biotrophic haustoria in host cells (e). At 48–72 hpi, *P. infestans* hyphae were observed ramifying through host tissue (f), and emerging as sporangiophores (g).

specific to *Pinci*1-6, and specific to *Pinci*1-18. The *Piact*A gene from *P. infestans* was used as a constitutively expressed endogenous control, and the expression of each *Pinci*1 family member was determined relative to this gene (Fig. 5a–f). Expression of all *Pinci*1 sequences was compared with the corresponding level of expression in a calibrator sample, which was cDNA from non-sporulating mycelium grown in rye broth. The expression of each gene in the mycelium cDNA sample was assigned the value of 1.0.

The *P. infestans ipiO* gene, which is known to be upregulated in invading hyphae during infection (van West *et al.*, 1998), was used to confirm disease progression in inoculated leaf samples, and it yielded the expected expression profile (Fig. 5f). Unlike *ipiO*, which is upregulated in pre-infection stages and during the infection, all of the *Pinci*1 sequences tested by real-time RT-PCR were shown to be upregulated only during infection at 12–56 hpi, in comparison with their expression levels in mycelium. This expression peaked at 12–33 hpi (biotrophy), before tailing off by 72 hpi (Fig. 5a–e), confirming that the SSH had enriched for transcripts upregulated during the infection. Although the *Pinci*1 members were upregulated *in planta*, the abundance of *Pinci*1 transcripts was relatively low, and at the highest level (*Pinci*1-18) represented only 0.1-fold of the number of *Piact*A transcripts. Levels of mRNAs for other *Pinci*1 members were typically lower at, for example, 0.01-fold of *Piact*A (*Pinci*1-1 and 5 combined), or 0.03-fold *Piact*A (*Pinci*1-6).

*Pinci*1-1 and *Pinci*1-5 do not encode proteins

The Testcode algorithm (Fickett 1982) identifies potential protein coding sequences by measuring the non-randomness of the composition at every third base independently from the reading frames. The algorithm has been used to indicate probable ncRNA genes (e.g. Srikantan *et al.*, 2000), and analysis of the *Pinci*1-1 (Fig. 6a) and *Pinci*1-5 (Fig. 6b)

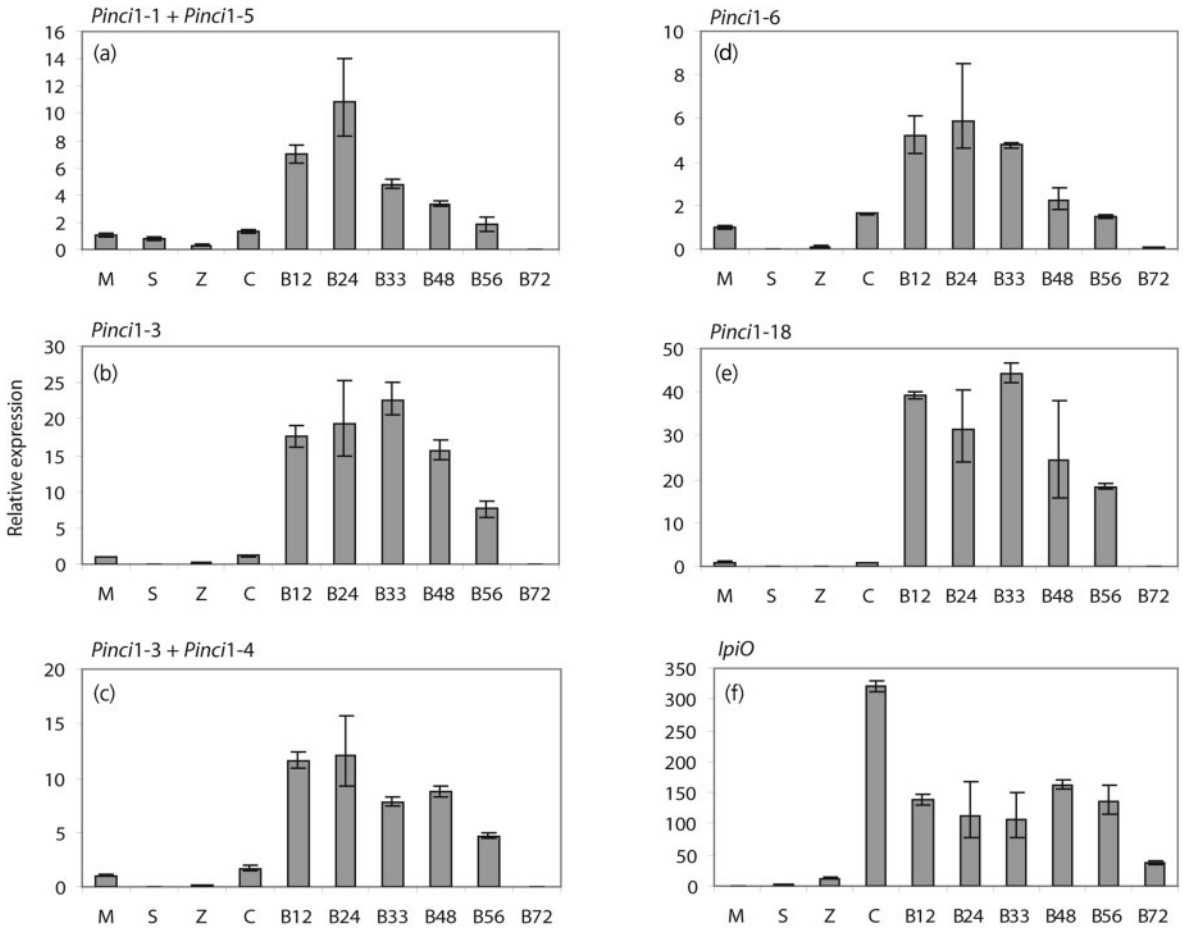


Fig. 5. Real-time RT-PCR expression profiles of five groups of *P. infestans Pinci*1 sequences (a–e) and *ipiO* (f) in pre-infection stages (S, sporangia; Z, zoospores; C, germinating cysts), and *in planta* 12 (B12), 24 (B24), 33 (B33), 48 (B48), 56 (B56) and 72 (B72) hpi of susceptible potato cv. Bintje. The expression values are relative to those for vegetative non-sporulating mycelium (M); for each group of sequences the value for M is relative to the expression of the constitutively expressed *Piact*A gene. Error bars represent confidence intervals calculated using three technical replicates for each sample within the RT-PCR assay.

cDNA sequences classified them as non-coding. In contrast, Testcode analysis of a number of known protein-coding *P. infestans* genes, including the *Piact*A gene (shown in Fig. 6c), classified the sequences as coding.

To further test the coding potential of ORFs in *Pinci*1-1 and *Pinci*1-5, ORFs 1a, 1b and 2 (Fig. 2) and, as a positive control, the *Piact*A gene ORF, were cloned into the oomycete constitutive expression vector pTOR, with the introduction of C-terminal FLAG tags, and they were transiently expressed in *P. infestans* following particle bombardment (Cvitanich & Judelson, 2003). RT-PCR, using primers specific to each ORF, in combination with a primer that anneals to the FLAG coding sequence, demonstrated that each ORF was transcribed in *P. infestans* (Fig. 6d). RT-PCR products were obtained only when cDNAs from *P. infestans* transformants, but not RNAs, were used as templates. In contrast, although a protein product of expected size was detected with FLAG antibody in protein extracted from *P. infestans* expressing the *Piact*A–FLAG fusion, no protein products were detected in *P. infestans* expressing *Pinci*1-ORF–FLAG fusions (Fig. 6e), suggesting that these RNAs are not translated.

Genomic distribution of *Pinci*1 in relation to other expressed sequences

A stringent BLASTN search was used to determine the locations of the 18 473 *P. infestans* EST-derived unigenes on the draft genome supercontigs. Matches to unigenes (E value, $<10^{-20}$) were found at 17 932 locations on 135 draft genome supercontigs that also contained matches to *Pinci*1 sequences. Visual inspection revealed a strong association between *Pinci*1 matches and *cis* downstream co-located (within 2.5 kb) matches to unigenes (see Supplementary Fig. S5). In total, only 12 different unigenes were found to have sequence matches immediately downstream of the 407 *Pinci*1 sequences. Of these, rpcy_7919.y1.abd occurs

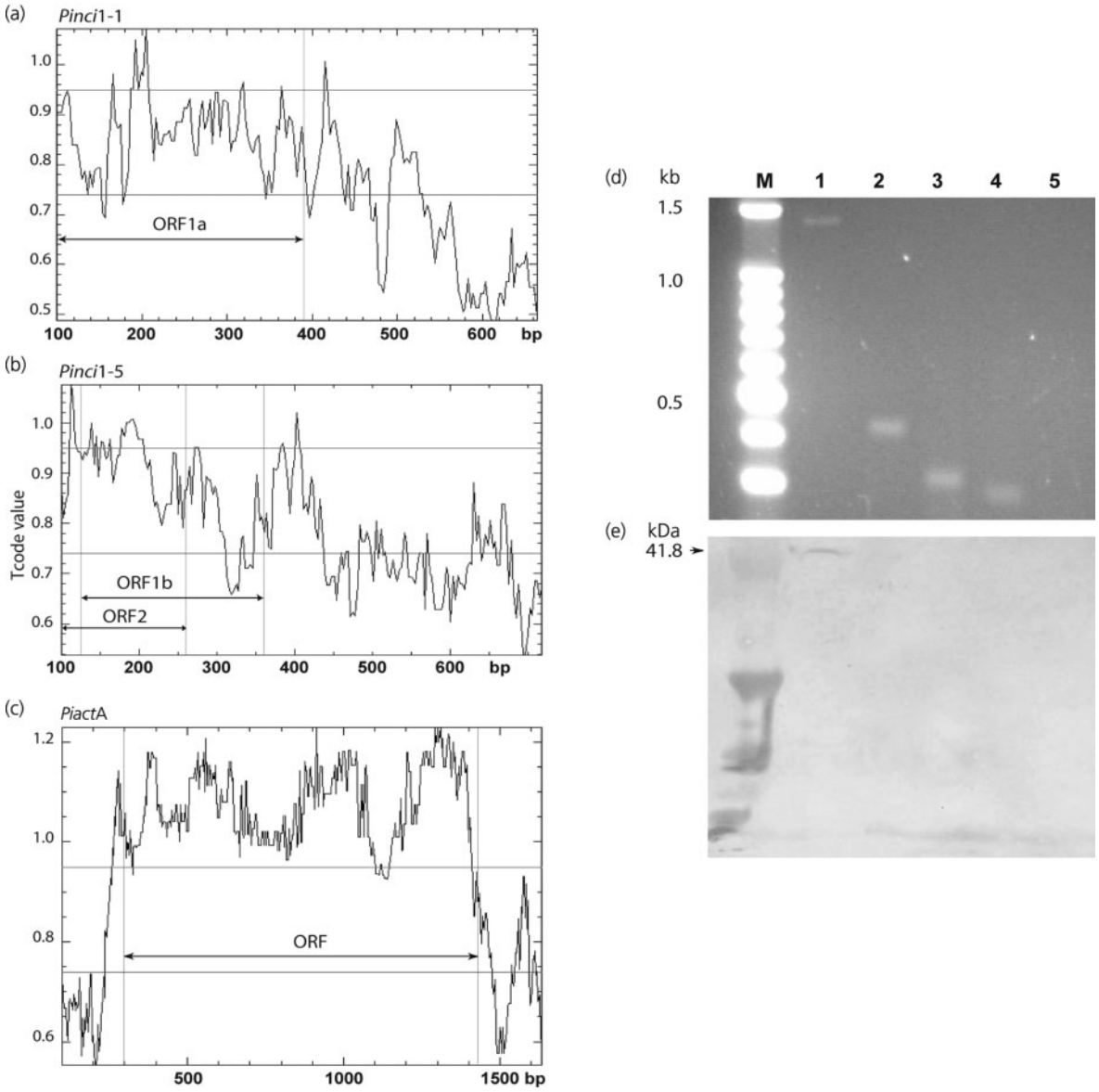


Fig. 6. Evaluation of the coding capacity of (a) *Pinci*1-1, (b) *Pinci*1-5, and (c) *P. infestans Piact*A (actin A), independently from the reading frame, by using the Testcode algorithm. The number of base pairs is indicated on the *x*-axis, the Testcode (Tcode) values are shown on the *y*-axis. Regions of longer than 200 bp above the upper line (at the 0.95 value) are considered coding; those under the lower line (at the 0.74 value) are considered non-coding at a confidence level of greater than 95 %. (d) Transient transcription and (e) translation of *Pinci*1 and *Piact*A FLAG-tagged ORFs in *P. infestans* following particle bombardment. M, marker; 1, *Piact*A gene ORF; 2, *Pinci*1-1 ORF 1a; 3, *Pinci*1-5 ORF 1b; 4, *Pinci*1-5 ORF 2; and 5, negative control. The PiactA protein band is visible at approximately 42 kDa; no protein bands were visible in the 10–20 kDa region predicted for FLAG-tagged *Pinci*1 following Western blotting with anti-FLAG antibody. Molecular size markers are indicted as kb (d) and kDa (e).

immediately downstream of a *Pinci*1 match 333 times (with modal separations of approximately either 90 or 1600 bp), and rpcy_7126.y1.abd 27 times (with a modal separation of approximately 270 bp). The remaining nine unigenes occur downstream of *Pinci*1 five times each, or less.

The unigene match most frequently co-located with *Pinci*1 (rpcy_7919.y1) was predicted to be non-coding by the Testcode algorithm (Fickett, 1982). Of the remaining 11 ESTs co-located with *Pinci*1, only two were predicted to be coding sequences by the Testcode algorithm (Table 1), suggestive of an association between *Pinci*1 and transcribed, but untranslated, sequences. Application of the Testcode algorithm to all 18 473 unigenes reported by Randall *et al.* (2005) classified 6919 as coding, 7995 as undecided, and 1430 as non-coding unigenes (Table 1), whilst 2129

Table 1. Frequency table describing the number of unigenes in the *P. infestans* draft genome sequence classified by the Testcode algorithm (Fickett, 1982) as coding, undecided or non-coding

The frequency counts of predicted classifications for unigenes with BLASTN hits found within 10 kb of *Pinci*1 matches, and those immediately downstream of *Pinci*1 matches.

Testcode classification	All unigenes	Unigenes within 10 kb of *Pinci*1	Unigenes immediately downstream of *Pinci*1
Coding	6919	173	2
Undecided	7995	209	6
Non-coding	1430	45	3

unigenes were either below the 200 bp length threshold for Testcode, or contained sequence ambiguities, and so could not be analysed by this algorithm. A total of 445 unigenes were located within 10 kb of *Pinci*1 sequences, and Testcode classified these as showing similar proportions of coding/undecided/non-coding sequences as generally observed amongst the entire unigene set (Table 1).

ORFs encoding putative RXLR sequences (Birch *et al.*, 2006; Kamoun 2006; Rehmany *et al.*, 2005) were identified using the combined procedures outlined by Bhattacharjee *et al.* (2006) and Torto *et al.* (2003). A total of 159 ORFs putatively encoding secreted RXLR proteins were identified on the 135 supercontig sequences containing matches to *Pinci*1. Of these, only 7 of 159 putative RXLR-class sequences, one of which is *Avr3a* (Armstrong *et al.*, 2005), lie within 10 kb of a *Pinci*1 sequence. It thus seems unlikely that there is any systematic co-location of putative RXLRs and *Pinci*1 sequences.

*Pinci*1 mRNA secondary structure, and potential for microRNA (miRNA) biogenesis

Prediction of the secondary structure of *Pinci*1 RNAs was performed using both the Vienna RNA package RNAfold web interface program, and the Sfold RNA secondary structure prediction program. Both programs predicted RNA structures with numerous stem–loops and other regions of self-complementarity. The longest regions of homology in stem–loop structures were less than 22 bp, allowing for up to five symmetrical sequence mismatches, and they were thus considered too short to represent substrates for miRNA generation.

MiRNA molecules are typically conserved among related species, and this property can be used to assist in miRNA identification (Bartel 2004; Zhang *et al.*, 2006). *Pinci*1 members were compared by BLASTN with the draft genome sequences of *Phytophthora sojae* and *Phytophthora ramorum*. Regions of similarity, when found, ranged from 18 to 20 bp, and were observed at a single locus in each genome. These short regions of similarity were not found to be within stem–loop structures in *Pinci*1 sequences, further suggesting that *Pinci*1 members do not yield miRNA molecules.

DISCUSSION

To the best of our knowledge, we report the first long non-protein-coding RNA from the stramenopile eukaryotic oomycete *P. infestans*. SSH was used to enrich for cDNAs from *P. infestans* that represent genes upregulated during potato infection, and the subtracted cDNA was hybridized to a *P. infestans* BAC library to identify 100 hybridizing BAC clones. These clones contained multiple copies of a related sequence called *Pinci*1. More than 800 further BAC clones that contain *Pinci*1 sequences were not initially identified in this screen, possibly reflecting the extremely low levels of *P. infestans*-derived cDNA in the mixed SSH cDNA used as a hybridization probe. Prior analysis of the genomic coverage of the BAC library (Whisson *et al.*, 2001), coupled with restriction digestion of *Pinci*1-hybridizing BAC clones, indicated that the related sequences reside at more than 90 locations in the *P. infestans* genome. Moreover, many of these regions contain clustered copies of the sequence, and thus we conservatively predicted that there are in excess of 300 copies of *Pinci*1 throughout the genome. This was confirmed by a whole-genome BLASTN search on the draft *P. infestans* genome that identified 407 *Pinci*1 matches arranged in clusters of up to nine copies. These matches frequently comprised only the 5′ or the 3′ domain of the complete *Pinci*1 sequence, but, at lengths of over 400 nt, they would still be expected to hybridize to the full-length *Pinci*1 probe.

More than 30 different repetitive DNA families, representing more than 50 % of the genome, have been reported in *P. infestans*, including those related to the *Gypsy* and *Copia* families of retrotransposons (Tooley & Garfinkel, 1996; Jiang *et al.*, 2005; Judelson & Randall, 1998; Judelson, 2002), a *hAT*-like transposon (Ah Fong & Judelson, 2004), and short interspersed elements (Whisson *et al.*, 2005). The *Pinci*1 sequences bear no resemblance to these or to any nucleotide sequences in public databases, and thus they represent a new type of transcribed repetitive element.

Transcribed *Pinci*1 sequences are polyadenylated like mRNAs, but Testcode analysis classified the sequences as non-coding. Moreover, ORFs identified within *Pinci*1 sequences lacked conserved eukaryotic translation start sites, and constructs containing FLAG-tagged ORFs from

*Pinci*1-1 and *Pinci*1-5, although transcribed, were not translated in *P. infestans*. The *Pinci*1 family thus apparently yields a novel class of ncRNA.

The paradigm that genetic information flows from DNA to RNA to protein is a central dogma in biology. Nevertheless, analyses of higher eukaryotic genomes have revealed large numbers of ncRNAs that lack substantial ORFs, and which may represent at least half of all transcripts (Claverie, 2005; Ota *et al.*, 2004; reviewed by Mattick, 2003). In the human genome, although only 1–2 % of the genome constitutes protein-coding sequence, as much as half of the genome may be transcribed, an estimate that has been supported by transcriptional analyses of chromosomes 21 and 22 (Cawley *et al.*, 2004), and analysis of tiling arrays (Cheng *et al.*, 2005). In mouse, annotation of 60 770 full-length cDNAs clustered into 33 409 transcription units, of which 15 815 appeared to be ncRNAs (Okazaki *et al.*, 2002). Similarly, Ota *et al.* (2004) identified 43 % of full-length human cDNA sequences as non-protein coding. Elucidating potential roles for these transcripts is a major current challenge. Analysis of the *P. infestans* unigene set using Testcode revealed that nearly 10 % of sequences are classified as non-protein coding (and nearly half as undecided); this is proportionally less than higher eukaryotes, such as mouse and human, but nevertheless a significant component of the transcriptome.

Non-coding RNAs can be classified as either housekeeping or regulatory (Morey & Avner, 2004). Housekeeping ncRNAs are constitutively expressed, and include infrastructural RNAs (rRNAs, tRNAs, snoRNAs, spliceosomal RNAs, etc.) that are directly or indirectly required for mRNA processing and translation. In contrast, regulatory ncRNAs are usually themselves tightly developmentally regulated (reviewed by Goodrich & Kugel, 2006; Hirsch *et al.*, 2006; Saha *et al.*, 2006), and they have been shown to be involved in diverse biological mechanisms, such as the control of chromosome architecture, transcriptional regulation, developmental timing of protein synthesis, and mRNA turnover, and they may also regulate alternative splicing (Ling *et al.*, 2005; reviewed by Goodrich & Kugel, 2006; Mattick, 2003; Morey & Avner, 2004). At least one non-coding RNA gene, *HAR1F* in humans, shows significant evolutionary acceleration, and is associated with the unique biology of that species in terms of brain development (Pollard *et al.*, 2006). The *Pinci*1 family encodes transcripts that are upregulated at 12–33 hpi, representing the biotrophic stage of *P. infestans* interaction with potato. The *Pinci1* sequence could simply be a repetitive element that, early in its evolution/distribution, acquired an infection stage-specific promoter element. However, it is also possible that the *Pinci1* sequences could constitute a regulatory ncRNA family.

Regulatory ncRNAs include miRNAs, which are short (21–25 nt) ncRNAs generated by post-transcriptional processing of larger ncRNAs containing double-stranded RNA stem–loop structures (reviewed by Bartel 2004). In plants, miRNAs play a major role in targeting protein-coding mRNAs for destruction and/or translational regulation (Llave *et al.*, 2002; reviewed by Mallory & Vaucheret, 2006; Zhang *et al.*, 2006), and in directing DNA methylation (Wassenegger, 2000). Predicted secondary structures of *Pinci*1 members did not identify stem–loop structures of sufficient size to generate miRNAs. Furthermore, sequences of the opposite strands of the stem–loops typically contained asymmetrical sequence mismatches, whereas miRNA sequence mismatches are typically symmetrical (Bartel, 2004). A further characteristic of miRNAs is their apparent conservation between related species and genera. Comparison of *Pinci*1 sequences with genome sequences of *P. sojae* and *P. ramorum* failed to identify any region of *Pinci*1 members that were conserved in these species, and that may thus act as a conserved miRNA.

A number of large ncRNAs have been implicated in the regulation of distant genes through modifications to alter the balance of heterochromatin and euchromatin (reviewed by Cook, 2003; Morey & Avner, 2004). In mammals, dosage compensation of X-linked gene products occurs through X-chromosome inactivation, initiated by the untranslated X-inactivation-specific transcript (*Xist*). *Xist* coats the entire length of one X chromosome, triggering heterochromatinization by H3K27 hypermethylation (reviewed by Morey & Avner, 2004). In contrast, in *Drosophila*, dosage compensation is achieved by twofold upregulation of genes on the single X chromosome. Intriguingly, this is also caused by ncRNAs *roX1* and *roX2* (Andersen & Panning, 2003). These examples illustrate the capacity for ncRNAs to spread in *cis* over long distances, where they act as binding sites for proteins that induce chemical or structural modifications that propagate down the chromosome fibre.

A number of silencers, enhancers and locus control regions (LCRs) constitute ncRNAs that act over thousands of base pairs to regulate adjacent genes, and transcribed barriers confine repressive heterochromatin to particular chromosomal regions (Cook, 2003). Using chromosome conformation capture, Dekker *et al.* (2002) demonstrated that an ncRNA LCR contacts and regulates the transcribed β-globin gene in mouse by formation of a chromatin loop. Ling *et al.* (2005) demonstrated that the β-globin HS2 enhancer initiated transcription of variable-length polyadenylated ncRNAs at multiple sites within, and downstream of, the enhancer, until the cognate promoter was reached. These ncRNAs were not capped, and remained in the nucleus. Thus, the enhancer-derived ncRNAs were probably a by-product of the process that acted to deliver the transcriptional components to the promoter. Intriguingly, *Pinci*1 transcripts were also of variable length, and were located on BAC clones that contain the *P. infestans in planta* upregulated genes *Avr*3a, *scr*91 and PYrpcy_0850, raising the possibility of *Pinci*1 involvement in regulation of these genes. From hybridization and PCR experiments in our laboratory, *Pinci*1 is also located on BACs containing the infection-upregulated genes *scr*74, *ipi*O and *ipi*B, and an

additional four predicted RXLR-class effectors (J. G. Morales & S. C. Whisson, unpublished). Moreover, a total of 445 unigenes match sequences within 10 kb of *Pinci*1 sequences. However, the wider genomic context of *Pinci*1 organization, with respect to other genes, will only become apparent with the completion and annotation of the *P. infestans* genome sequence (www.broad.mit.edu/annotation/genome/phytophthora_infestans/Home.html).

The large size of the *Pinci*1 gene family, and its distribution throughout the *P. infestans* genome, raise speculation that it may act to co-regulate genes required in the establishment of biotrophy, perhaps by acting as enhancers to assemble transcriptional components ahead of specific promoters, or to loop distant chromosomal regions into a 'transcription factory'; the latter hypothesis has been proposed by Cook (2003) to explain how transcribed regulators may act at a distance. The annotated genome of *P. infestans* will reveal genes closely linked to *Pinci*1 gene clusters, and transcriptional analyses will reveal whether these genes are co-ordinately upregulated with *Pinci*1 at 12–33 hpi.

ACKNOWLEDGEMENTS

This work was funded by the Scottish Executive Environment and Rural Affairs Department (SEERAD), and Biotechnology and Biological Sciences Research Council (BBSRC). Transformation of *P. infestans* was carried out under SEERAD licence GM/235/2005. We are grateful to the Broad Institute, Cambridge, MA, USA, for sequencing of the *P. infestans* genome, and timely release of the first draft assembly.

REFERENCES

Ah Fong, A. M. V. & Judelson, H. S. (2004). The *hAT*-like DNA transposon *DodoPi* resides in a cluster of retro- and DNA transposons in the stramenopile *Phytophthora infestans*. *Mol Genet Genom* **271**, 577–585.

Andersen, A. A. & Panning, B. (2003). Epigenetic gene regulation by noncoding RNAs. *Curr Opin Cell Biol* **15**, 281–289.

Armstrong, M. R., Whisson, S. C., Pritchard, L., Bos, J. I. B., Venter, E., Avrova, A. O., Rehmany, A. P., Böhme, U., Brooks, K. & other authors (2005). An ancestral oomycete locus contains late blight avirulence gene *Avr3a*, encoding a protein that is recognised in the host cytoplasm. *Proc Natl Acad Sci U S A* **102**, 7766–7771.

Avrova, A. O., Stewart, H. E., De Jong, W., Heilbronn, J., Lyon, G. D. & Birch, P. R. J. (1999). A cysteine protease gene is expressed early in resistant potato interactions with *Phytophthora infestans*. *Mol Plant Microbe Interact* **12**, 1114–1119.

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.

Avrova, A. O., Taleb, N., Rokka, V. M., Heilbronn, J., Hein, I., Gilroy, E. M., Cardle, L., Stewart, H. S., Loake, G. & other authors (2004). Potato oxysterol binding protein and cathepsin B are rapidly upregulated in independent defence pathways that distinguish *R* gene-mediated and field resistances to *Phytophthora infestans*. *Mol Plant Pathol* **5**, 45–56.

Baldauf, S. L. (2003). The deep roots of eukaryotes. *Science* **300**, 1703–1706.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297.

Bell, K. S., Avrova, A. O., Holeva, M. C., Cardle, L., Morris, W., De Jong, W., Toth, I. K., Waugh, R., Bryan, G. J. & Birch, P. R. J. (2002). Sample sequencing of a selected region of the genome of *Erwinia carotovora* subsp. *atroseptica* reveals candidate phytopathogenicity genes and allows comparison with *Escherichia coli*. *Microbiology* **148**, 1367–1378.

Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783–795.

Beyer, K., Binder, A., Boller, T. & Collinge, M. (2001). Identification of potato genes induced during colonization by *Phytophthora infestans*. *Mol Plant Pathol* **2**, 125–135.

Beyer, K., Jimenez Jimenez, S., Randall, T. A., Lam, S., Binder, A., Boller, T. & Collinge, M. (2002). Characterisation of *Phytophthora infestans* genes regulated during the interaction with potato. *Mol Plant Pathol* **3**, 473–485.

Bhattacharjee, S., Hiller, N. L., Liolios, K., Win, J., Kanneganti, T.-D., Young, C., Kamoun, S. & Haldar, K. (2006). The malarial host-targeting signal is conserved in the Irish potato famine pathogen. *PLoS Pathog* **2**, e50. doi:10.1371/journal.ppat.0020050.

Birch, P. R. J. & Whisson, S. C. (2001). *Phytophthora infestans* enters the genomics era. *Mol Plant Pathol* **2**, 257–263.

Birch, P. R. J., Avrova, A. O., Duncan, J. M., Lyon, G. D. & Toth, R. L. (1999). Isolation of potato genes that are induced during an early stage of the hypersensitive response to *Phytophthora infestans*. *Mol Plant Microbe Interact* **12**, 356–361.

Birch, P. R. J., Avrova, A. O., Armstrong, M., Venter, E., Taleb, N., Gilroy, E. M., Phillips, M. S. & Whisson, S. C. (2003). The potato–*Phytophthora infestans* interaction transcriptome. *Can J Plant Pathol* **25**, 226–231.

Birch, P. R. J., Rehmany, A. P., Pritchard, L., Kamoun, S., Beynon, J. L. (2006). Trafficking arms: Oomycete effectors enter host plant cells. *Trends Microbiol* **14**, 8–11.

Bittner-Eddy, P. D., Allen, R. L., Rehmany, A. P., Birch, P. R. J. & Beynon, J. L. (2003). Use of suppression subtractive hybridization to identify downy mildew genes expressed during infection of *Arabidopsis thaliana*. *Mol Plant Pathol* **4**, 501–508.

Blanco, F. A. & Judelson, H. S. (2005). A bZIP transcription factor from *Phytophthora* interacts with a protein kinase and is required for zoospore motility and plant infection. *Mol Microbiol* **56**, 638–648.

Cawley, S., Bekiranov, S., Ng, S. S., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J. & other authors (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H. & other authors (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149–1154.

Claverie, J. M. (2005). Fewer genes, more noncoding RNA. *Science* **309**, 1529–1530.

Cook, P. R. (2003). Nongenic transcription, gene regulation and action at a distance. *J Cell Sci* **116**, 4483–4491.

Cvitanich, C. & Judelson, H. S. (2003). Stable transformation of the oomycete, *Phytophthora infestans*, using microprojectile bombardment. *Curr Genet* **42**, 228–235.

Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. (2002). Capturing chromosome conformation. *Science* **295**, 1306–1311.

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

Fickett, J. W. (1982). Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res* **10**, 5303–5318.

Goodrich, J. A. & Kugel, J. F. (2006). Non-coding-RNA regulators of RNA polymerase II transcription. *Nat Rev Mol Cell Biol* **7**, 612–616.

Grenville-Briggs, L. J., Avrova, A. O., Bruce, C. R., Williams, A., Whisson, S. C., Birch, P. R. J. & van West, P. (2005). Elevated amino acid biosynthesis in *Phytophthora infestans* during appressorium formation and potato infection. *Fungal Genet Biol* **42**, 244–256.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.

Hirsch, J., Lefort, V., Vankersschaver, M., Boualem, A., Lucas, A., Thermes, C., d'Aubenton-Carafa, Y. & Crespi, M. (2006). Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the *MIR162a*-derived transcripts. *Plant Physiol* **140**, 1192–1204.

Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Cerretti, D. P., Urdal, D. L. & Conlon, P. J. (1988). A short polypeptide marker sequence useful for recombinant protein identification and purification. *Bio/Technology* **6**, 1204–1210.

Huang, X. & Madan, A. (1999). CAP3: *a* DNA sequence assembly program. *Genome Research* **9**, 868–877.

Jiang, R. H., Dawe, A. L., Weide, R., van Staveren, M., Peters, S., Nuss, D. L. & Govers, F. (2005). Elicitin genes in *Phytophthora infestans* are clustered and interspersed with various transposon-like elements. *Mol Genet Genom* **273**, 20–32.

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.

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

Kamoun, S. (2006). A Catalogue of the effector secretome of plant pathogenic oomycetes. *Annu Rev Phytopathol* **44**, 41–60.

Latijnhouwers, M., Ligterink, W., Vleeshouwers, V. G. A. A., Van West, P. & Govers, F. (2004). A Gα subunit controls zoospore motility and virulence in the potato late blight pathogen *Phytophthora infestans*. *Mol Microbiol* **51**, 925–936.

Ling, J., Baibakov, B., Pi, W., Emerson, B. M. & Tuan, D. (2005). The HS2 enhancer of the *β*-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a *cis*-linked globin promoter. *J Mol Biol* **350**, 883–896.

Llave, C., Kasschau, K. D., Rector, M. A. & Carrington, J. C. (2002). Endogenous and silencing-associated small RNAs in plants. *Plant Cell* **14**, 1605–1619.

Mallory, A. C. & Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. *Nat Genet* **38**, Suppl. S31–S36.

Mattick, J. S. (2003). Challenging the dogma: the hidden layers of non-protein-coding RNAs in complex organisms. *BioEssays* **25**, 930–939.

Morey, C. & Avner, P. (2004). Employment opportunities for non-coding RNAs. *FEBS Lett* **567**, 27–34.

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R. & other authors (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573.

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H. & other authors (2004). Complete sequencing and characterization of 21 243 full-length human cDNAs. *Nat Genet* **36**, 40–45.

Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J., Katzman, S., King, B., Onodera, C. & other authors (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172.

Pritchard, L., White, J. A., Birch, P. R. J. & Toth, I. K. (2006). GenomeDiagram: a Python package for the visualisation of large-scale genomic data. *Bioinformatics* **22**, 616–617.

Randall, T. A., Dwyer, R. A., Huitema, E., Beyer, K., Cvitanich, C., Kelkar, H., Ah Fong, A., Gates, K., Roberts, S. & other authors (2005). Large-scale gene discovery in the oomycete *Phytophthora infestans* reveals likely components of phytopathogenicity shared with true fungi. *Mol Plant Microbe Interact* **18**, 229–243.

Rehmany, A. P., Gordon, A., Rose, L. E., Allen, R. L., Armstrong, M. R., Whisson, S. C., Kamoun, S., Tyler, B. M., Birch, P. R. J. & Beynon, J. L. (2005). Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two *Arabidopsis* lines. *Plant Cell* **17**, 1839–1850.

Saha, S., Murthy, S. & Rangarajan, P. N. (2006). Identification and characterization of a virus-inducible non-coding RNA in mouse brain. *J Gen Virol* **87**, 1991–1995.

Srikantan, V., Zou, Z., Petrovics, G., Zu, L., Augustus, M., Davis, L., Livezey, J. R., Connell, T., Sesterhenn, I. A. & other authors (2000). *PCGEM1*, a prostate-specific gene, is overexpressed in prostate cancer. *Proc Natl Acad Sci U S A* **97**, 12216–12221.

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

Torto, T. A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N. A., van West, P. & Kamoun, S. (2003). EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Res* **13**, 1675–1685.

van West, P., de Jong, A. J., Judelson, H. S., Emons, A. M. C. & Govers, F. (1998). The *ipiO* gene of *Phytophthora infestans* is highly expressed in invading hyphae during infection. *Fungal Genet Biol* **23**, 126–138.

Vleeshouwers, V. G., van Dooijeweert, W., Govers, F., Kamoun, S. & Colon, L. T. (2000). The hypersensitive response is associated with host and nonhost resistance to *Phytophthora infestans*. *Planta* **210**, 853–864.

Wassenegger, M. (2000). RNA-directed DNA methylation. *Plant Mol Biol* **43**, 203–220.

Whisson, S. C., van der Lee, T., Bryan, G., Waugh, R., Govers, F. & Birch, P. R. J. (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.

Whisson, S. C., Avrova, A. O., Lavrova, O. & Pritchard, L. (2005). Families of short interspersed elements in the genome of the oomycete plant pathogen, *Phytophthora infestans*. *Fungal Genet Biol* **4**, 351–365.

Zhang, B., Pan, X., Cobb, G. P. & Anderson, T. A. (2006). Plant microRNA: a small regulatory molecule with big impact. *Dev Biol* **289**, 3–16.

Edited by: J. Alfano