New Phytologist

Review

Research review

Intraspecific comparative genomics to identify avirulence genes from *Phytophthora*

Author for correspondence:
Sophien Kamoun
Tel: +1 330 263 3847
Fax: +1 330 263 3841
Email: kamoun.1@osu.edu

Received: *22 December 2002*
Accepted: *10 April 2003*

doi: 10.1046/j.1469-8137.2003.00801.x

Jorunn I. B. Bos[1], Miles Armstrong[2], Stephen C. Whisson[2], Trudy A. Torto[1], Mildred Ochwo[1], Paul R. J. Birch[2] and Sophien Kamoun[1]

[1]Department of Plant Pathology, The Ohio State University, Ohio Agricultural Research and Development Center, Wooster, OH 44691, USA; [2]Plant Pathogen Interaction Programme, Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, UK

Key words: functional and comparative genomics, plant–microbe interactions, avirulence, host-specificity, oomycetes, *Phytophthora*.

Summary

Members of the oomycete genus *Phytophthora* cause some of the most devastating plant diseases in the world and are arguably the most destructive pathogens of dicot plants. *Phytophthora* research has entered the genomics era. Current genomic resources include expressed sequence tags from a variety of developmental and infection stages, as well as sequences of selected regions of *Phytophthora* genomes. Genomics promise to impact upon our understanding of the molecular basis of infection by *Phytophthora*, for example, by facilitating the isolation of genes encoding effector molecules with a role in virulence and avirulence. Based on prevalent models of plant–pathogen coevolution, some of these effectors, notably those with avirulence functions, are predicted to exhibit significant sequence variation within populations of the pathogen. This and other features were used to identify candidate avirulence genes from sequence databases. Here, we describe a strategy that combines data mining with intraspecific comparative genomics and functional analyses for the identification of novel avirulence genes from *Phytophthora*. This approach provides a rapid and efficient alternative to classical positional cloning strategies for identifying avirulence genes that match known resistance genes. In addition, this approach has the potential to uncover 'orphan' avirulence genes for which corresponding resistance genes have not previously been characterized.

Introduction

Induced resistance in plant–microbe interactions is regulated by recognition of pathogen molecules by the plant. This is illustrated by the gene-for-gene concept, which implies that a ligand (elicitor) encoded by an avirulence (*Avr*) gene from the pathogen is recognized by a receptor encoded by a resistance (*R*) gene from the plant, resulting in recognition of the pathogen and activation of plant defense mechanisms such as the hypersensitive response (HR) (Staskawicz *et al.*, 1995; Hammond-Kosack & Jones, 1997; Dangl & Jones, 2001). In its simplest illustration, the biochemical basis of the gene-for-gene model consists of direct interaction between *Avr* and *R* gene products. However, recent studies indicate a more complex basis for recognition, in which perception of *Avr* products by R proteins is indirect and involves at least a third

component (Dixon *et al.*, 2000; Leister & Katagiri, 2000; Rivas *et al.*, 2002a,b). According to the 'Guard hypothesis' (Van der Biezen & Jones, 1998) this component is a virulence target (VT) that is recognized by the *Avr* product in both susceptible and resistant plants. The R protein thus acts as a 'guard', monitoring alterations in the VT mediated by the *Avr* product, and promoting further defense signaling. Supporting evidence for the Guard Hypothesis is accumulating with the identification of the VT in several gene-for-gene interactions (Dangl & Jones, 2001; Innes, 2001; Swiderski & Innes, 2001; Kim *et al.*, 2002; Kruger *et al.*, 2002; Rivas *et al.*, 2002a,b; Schneider, 2002).

Avirulence genes of eukaryotic plant pathogenic microbes exhibit a number of common structural and functional features (Lauge & De Wit, 1998; van't Slot & Knogge, 2002). For example, the majority of fungal *Avr* genes described to date encode extracellular proteins with a type II secretion peptide. Many eukaryotic *Avr* genes, such as *Cladosporium fulvum Avr2*, *Avr4*, and *Avr9*, *Rhynchosporium secalis nip1*, and *Phytophthora* elicitins, encode small secreted proteins with an even number of cysteine residues, that can induce defense responses when infiltrated into plant tissues (Lauge & De Wit, 1998; van't Slot & Knogge, 2002). Several of these common structural features, most notably secretion and the disulfide bridges formed by the pairs of cysteines are essential for HR induction and avirulence function (Joosten *et al.*, 1997; Kooman-Gersmann *et al.*, 1997; Lauge & De Wit, 1998; Kamoun *et al.*, 1999a; Luderer *et al.*, 2002a; van't Slot & Knogge, 2002). The disulfide bridges could enhance stability in the plant apoplast, which is known to be rich in degradative proteases (Joosten *et al.*, 1997; Luderer *et al.*, 2002a). However, despite the structural features of signal peptides for secretion and the presence of cysteine residues, there is no primary DNA sequence similarity between these *Avr* genes. The *Avr* genes *AVR-Pita* and *PWL* from the rice blast fungus *Magnaporthe grisea*, also encode extracellular proteins that are secreted via a type II signal peptide (Kang *et al.*, 1995; Sweigard *et al.*, 1995; Jia *et al.*, 2000; Orbach *et al.*, 2000). *AVR-Pita* encodes a metalloprotease that is thought to be delivered inside rice cells where it can interact with the Pi-ta R protein (Jia *et al.*, 2000; Orbach *et al.*, 2000). The *PWL* genes encode small proline and glycine rich proteins with unknown function (Kang *et al.*, 1995; Sweigard *et al.*, 1995). In addition to the described structural features, all fungal avirulence genes, including the *M. grisea AVR-Pita* gene, are actively expressed during infection of the host plant. Some are expressed exclusively during infection, whereas expression of other avirulence genes, such as *Avr9*, may also be induced by nutritional stress (Perez-Garcia *et al.*, 2001).

By definition, *Avr* genes that define cultivar-specific resistance exhibit significant sequence variation among races of the pathogen. For example, in virulent races, the *C. fulvum* avirulence gene *Avr4* occurs in various isoforms and the *Avr2* gene exhibits various mutations that typically result in truncated proteins (Joosten *et al.*, 1997; Luderer *et al.*, 2002b). The majority of the recessive virulence alleles of *Avr2* and *Avr4* carry single or simple nucleotide polymorphisms (SNPs) that result in unstable or nonfunctional HR elicitors. By contrast, the *Avr9* gene is deleted from races of *C. fulvum* that infect *Cf-9* plants (Lauge & De Wit, 1998). In the *R. secalis Avr* gene, *nip1*, single nucleotide changes, resulting in amino acid substitutions, were detected in the coding regions of *nip1* alleles from virulent races (Rohe *et al.*, 1995). As a consequence of these changes, interaction of *R. secalis* with barley is no longer controlled by *Rrs*1, indicating that recognition by the host plant can be circumvented by alteration of the primary structure of the NIP1 gene product (Rohe *et al.*, 1995). *AVR-Pita* is located in the unstable telomeric region of chromosome 3 of *M. grisea* and is frequently deleted in virulent strains (Orbach *et al.*, 2000). In some races of *M. grisea* that are virulent on *Pi-ta* plants, the *avr-Pita* alleles carry insertions (Kang *et al.*, 2001) or point mutations, one of which is within the predicted active site of the protease (Orbach *et al.*, 2000). The *PWL* genes are highly polymorphic among rice blast isolates from diverse grass species and geographic regions, and appear to contribute to species-specific avirulence (Sweigard *et al.*, 1995). This gene family appears to be highly dynamic and rapidly evolving (Kang *et al.*, 1995).

Oomycetes, such as *Phytophthora*, downy-mildews, and *Pythium*, form a unique branch of eukaryotic plant pathogens with an independent evolutionary history (Sogin & Silberman, 1998; Baldauf *et al.*, 2000; Margulis & Schwartz, 2000). Among the oomycetes, *Phytophthora* species cause some of the most destructive plant diseases in the world, and are arguably the most devastating pathogens of dicot plants (Erwin & Ribeiro, 1996). The most notable and best-studied oomycete is *Phytophthora infestans*, the Irish famine pathogen. *P. infestans* causes late blight, a devastating and re-emerging disease of potato and tomato (Fry & Goodwin, 1997a,b; Birch & Whisson, 2001; Schiermeier, 2001; Smart & Fry, 2001; Shattock, 2002). Despite their peculiar phylogenetic affinities and economic importance, oomycetes were chronically under-studied at the molecular level. This trend has dramatically reversed in recent years with significant technical developments, such as routine DNA transformation, use of reporter genes, genetic manipulation using gene silencing, and expanding genomic resources that promise to facilitate gene discovery and functional analyses (Kamoun, 2003).

Genetic analyses of *P. infestans* and *P. sojae* indicate that in many cases race/cultivar specificity follows the gene-for-gene model and involves segregating *Avr* genes. Several *Avr* genes have been targeted for positional cloning in *P. infestans* and *P. sojae* (van der Lee *et al.*, 2001a,b; Whisson *et al.*, 2001; MacGregor *et al.*, 2002; May *et al.*, 2002; Tyler, 2002). Nevertheless, so far, none of these race-specific *Avr* genes from *Phytophthora* have been described in the literature mainly because of the numerous difficulties encountered with positional cloning in *Phytophthora*, such as high levels of repetitive DNA and aberrant segregation at the target locus. To date, the

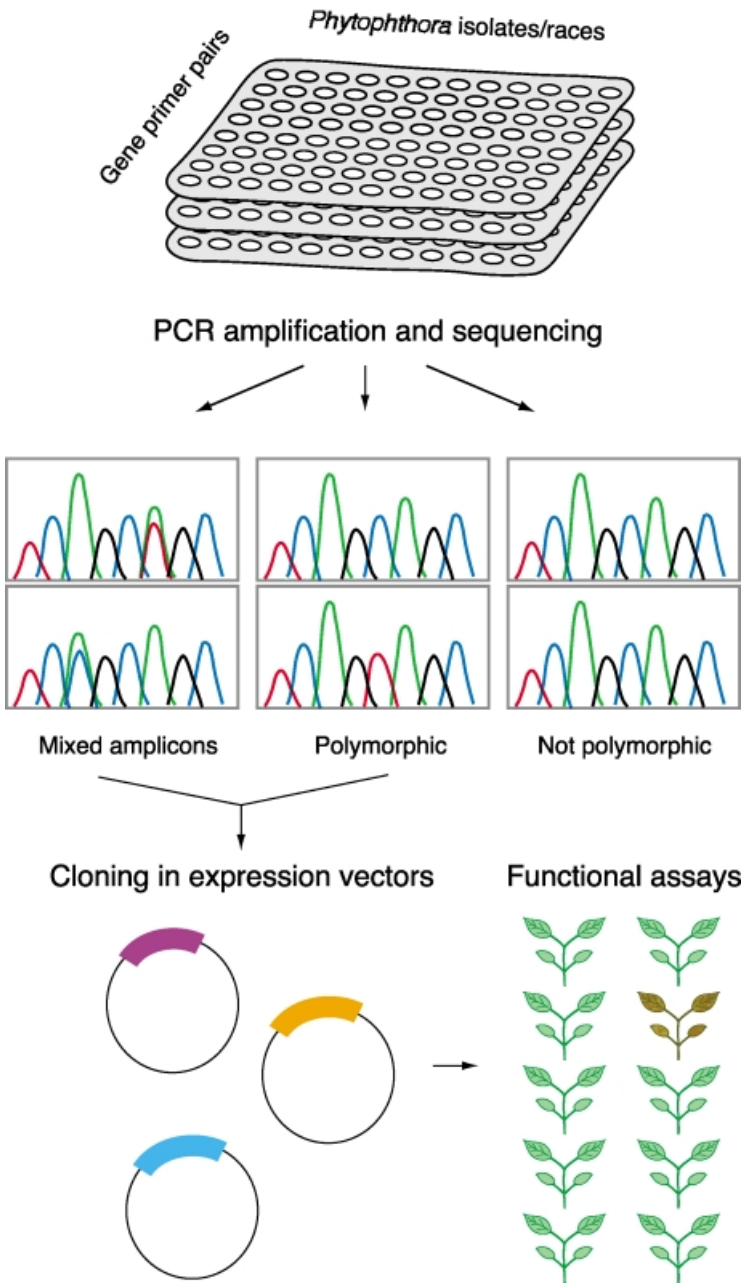

Fig. 1 Schematic illustration of strategy for identification of avirulence genes using intraspecific comparative genomics. Four steps are illustrated: (i) PCR amplification of candidate avirulence genes from panels of races of the target pathogen; (ii) DNA sequencing of amplicons; (iii) cloning in expression vectors; and (iv) functional assays on a collection of resistance gene containing plants. See text for details.

only *Avr* genes described from *Phytophthora* are members of the elicitin family, which are thought to condition avirulence to the nonhost plant tobacco (Kamoun *et al.*, 1998; Kamoun, 2001) and were first identified biochemically as elicitors of the HR (Ponchet *et al.*, 1999). In this paper, we describe an alternative approach to positional cloning for the identification of *Avr* genes in *Phytophthora*. This strategy consists of several steps (Fig. 1) and takes complete advantage of genome sequence data by combining expressed sequence tag (EST) data mining, intraspecific comparative genomics (association genetics, or linkage disequilibrium), and functional analyses. First, candidate genes exhibiting structural features indicative of known *Avr* genes are selected from sequence databases. Then, these genes are amplified and sequenced from a panel of phenotypically characterized *Phytophthora* races to identify polymorphic genes and establish associations between allele variation and avirulence phenotype. Finally, candidate *Avr* genes are validated using functional assays. Here, we describe the experimental steps required for this strategy as currently applied to the identification of race-specific *Avr* genes from *P. infestans* in our laboratories at the Ohio State University and the Scottish Crop Research Institute. We also present an update of our efforts to validate this strategy by providing preliminary analyses of polymorphisms for five candidate effector genes of *P. infestans*. Finally, we discuss the potential limitations of this approach and the outlook for future applications.

Experimental approach

Selection of candidate *Avr* genes from sequence databases

We have focused on *P. infestans* genes encoding proteins with structural features characteristic of known fungal and oomycete *Avr* and elicitor proteins. These features were used as a basis for data mining criteria to identify candidate *Avr* genes from expressed sequence tag (EST) or genome sequence databases (Kamoun *et al.*, 2002). Some of these criteria were discussed elsewhere (Qutob *et al.*, 2000; Birch & Whisson, 2001; Kamoun *et al.*, 2002; Torto *et al.*, 2003), and include genes encoding extracellular and small cysteine-rich proteins, as well as genes up-regulated during preinfection and infection stages. Computational tools and algorithms have been developed to facilitate the identification of these features from sequence databases. For example, PexFinder, an algorithm, that identifies genes encoding extracellular proteins from ESTs, has been used to select a set of 142 *Pex* (*Phytophthora ex*tracellular proteins) cDNAs, many of which are considered candidate effector genes (Torto *et al.*, 2003; http://www.oardc.ohio-state.edu/phytophthora/pexfinder). In addition, since ESTs from multiple races of *P. infestans* are available, comparative sequence analyses can be performed to identify allelic polymorphisms between the different races.

Primer design, PCR amplification, and high-throughput sequencing

To identify polymorphic genes, pairs of oligonucleotide primers were designed for the amplification of the entire open reading frame (ORF) from each selected candidate gene. Restriction enzyme sites were appended to the primers to allow facile cloning of the ORFs in plant or microbial expression vectors for functional assays. PCR amplifications were performed on genomic DNA from panels of up to 30 isolates of *P. infestans* that have been assessed for virulence on 11 (R1–R11) potato *R* gene differential genotypes. To increase the throughput of the PCR amplifications and subsequent DNA sequencing, all reactions were performed in panels of 8 primer pairs × 12 *P. infestans* isolates in 96-well microtiter plates. Amplicons were purified using 96-well PCR purification kits, checked on agarose gels, and their concentrations were measured as absorbance at 260 nm by spectrophotometry. Appropriately diluted samples were then submitted in 96-well microtiter plates for high-throughput DNA sequencing at our core sequencing facilities.

Sequence data analysis and interpretation

We used established bioinformatic platforms for sequence analysis and single nucleotide polymorphism (SNP) identification. Base calling, Quality Values (QV), and trimming were obtained with the Phred algorithm (Ewing & Green, 1998; Ewing *et al.*, 1998). Sequences were aligned and examined using Sequencher^{TM} 4.1 (Gene Codes Corp, Ann Arbor, MI, USA) or other programs that allow viewing of sequence calls, QV, and chromatogram data. Visual comparison of chromatogram data with the sequence alignment data was performed to evaluate whether differences in nucleotide sequences were likely to be polymorphisms or caused by base calling errors. Depending on the genes examined, several outcomes were observed. Occasionally, no amplification was obtained in some races suggesting that the gene might be missing or that a polymorphism corresponding to the primer sequences might interfere with PCR amplification. This was then tested by DNA blot hybridization and/or using different primer pairs. Some genes were not polymorphic within the set of *P. infestans* races examined. Other genes were polymorphic ranging from a few to a fairly large number of nucleotide substitutions. Considering that *P. infestans* is diploid and outbreeding, identified SNPs are frequently in a heterozygous state, leading to mixed sequence peaks at polymorphic nucleotides. Some genes resulted in more complex mixed DNA sequences suggesting that multiple alleles or paralogs were amplified to yield a mixture of amplicons. For these genes, a cloning step was introduced and was followed by sequencing the inserts of a set of randomly picked clones.

We catalogued all identified SNPs and examined the virulence typing data sets for correlations between polymorphisms that result in amino acid substitutions, other structural alterations of the encoded protein, and specific avirulence phenotypes. A simple statistical test was used to determine the extent of linkage disequilibrium between the polymorphic marker and avirulence phenotype (Lewontin, 1964):

$$D' = \frac{pABpab - pAbpaB}{pABpab + pAbpaB}$$

where *pAB* is the frequency of *AB* haplotypes, and similarly for *pAb*, *paB* and *pab*, as a measure of departure from linkage equilibrium. The value of *D* varies from 0 (linkage equilibrium) to 1 (complete association) and is taken as an absolute number due to the arbitrary assignment of symbols *A* and *B*. To date, all tested *P. infestans* avirulence loci have been determined to be dominant, although some loci can appear recessive in certain crosses. Consequently only those comparisons where a case could be made for avirulent individuals being heterozygous at a locus were considered.

Validation using functional assays

Polymorphic genes identified by comparative sequencing were selected for functional assays, using one or several of the methods available for *Phytophthora* genes (Kamoun *et al.*, 2002). For example, functional assays can be performed through ectopic expression of candidate *Avr* genes in plants.

Table 1 List of candidate *Phytophthora infestans scr* (small cysteine rich) and *pex* (*Phytophthora* extracellular) genes analyzed for single nucleotide polymorphisms

Gene	GenBank accession	Length of protein[a]	Length of signal peptide[a]	Number of cysteines[b]	Best BLASTP hit[c]	*E*-value
scr50	AAN31504	50	23	4	No significant hits	NA
scr58	AAN31505	58	18	2	No significant hits	NA
scr76	AAN31495	76	29	4	No significant hits	NA
scr91	BE775988	91	21	6	AAK63068 phytotoxic protein PcF precursor (*Phytophthora cactorum*)	1e-08
pex208	AAN31499	208	18	18	No significant hits	NA

[a]Predicted sequence features are based on Torto *et al*. (2003), length of sequences in amino-acids. [b]Number of cysteines in predicted mature protein. [c]Best BLASTP hit to non *P. infestans* sequences deposited in GenBank nonredundant database (April 2003). *E*-values larger than 1e-02 were not considered significant.

To this end, the candidate genes were cloned in plant expression vectors, such as the binary vector pGR106 that allows agroinfection of potato virus X (PVX) in solanaceous plants (Jones *et al*., 1999) or standard binary expression cassettes that allow transient expression by agroinfiltration (Van der Hoorn *et al*., 2000). Both of these vectors allow assaying for HR-like symptoms and have been successfully used with several *Phytophthora* genes (Kamoun *et al*., 1999a; Kamoun *et al*., 2002; Qutob *et al*., 2002; Torto *et al*., 2003). All of the genes tested for linkage disequilibrium with avirulence have previously been assayed for HR-eliciting activity in the nonhost tobacco, using the PVX expression system. None were shown to exhibit any HR-eliciting activity in this plant.

Alternatively, candidate genes can be cloned in vectors for bacterial or yeast expression. Recombinant proteins purified from *E. coli* or yeast can then be injected into plant leaves to assay for induction of the HR. For all these experiments, a panel of resistant plant genotypes, including varieties of tomato and potato known to carry *R* genes, as well as wild relatives of these host plants, will be selected for the functional assays.

A standard complementation approach can also be taken for functional analysis of candidate *Avr* genes. Currently, candidate genes are being cloned in oomycete transformation vectors and used to transform virulent strains that lack the particular allele. Transformed strains expressing the candidate gene can then be assessed for avirulence on the appropriate plant genotype to validate the *Avr* function. A similar gain of function experiment has been reported in which functional expression of the *Phytophthora cryptogea* gene encoding the basic elicitin cryptogein in *P. infestans* resulted in altered interaction with tobacco plants (Panabieres *et al*., 1998).

SNPs in Phytophthora extracellular proteins (Pex) and association with avirulence

Current work in our laboratories aims at validating the described strategy and applying it to the discovery of *Avr* genes from the late blight pathogen *P. infestans*. Here, we present preliminary sequence polymorphism data for five candidate effector genes identified by Torto *et al*. (2003) as encoding secreted and cysteine-rich proteins. The general characteristics of the five selected genes, *scr50*, *scr58*, *scr76*, *scr91*, and *pex208* are described in Table 1. All five genes encode proteins with a predicted signal peptide ranging from 18 to 29 amino acids in length. The predcited mature peptides for *scr50*, *scr58*, *scr76*, and *scr91* are relatively small ranging from 27 to 70 amino acids, and contain an even number of cysteines (range 2–6). *Pex208* encodes a predicted 190 amino acid protein with a total of 18 cysteines. All five genes appear restricted to *Phytophthora* and do not show significant similarity to sequences from other organisms based on BLASTP searches against GenBank nonredundant database (*E*-value cutoff 1e-02). *Scr91* shows similarity to PcF, a 52 amino acid extracellular protein from *Phytophthora cactorum*, that causes necrosis in strawberry and tomato (Orsomando *et al*., 2001).

We used PCR amplification with primers flanking the ORFs of the five selected genes and DNA sequencing to identify SNPs. A minimum of 16 isolates of *P. infestans* were examined and scored for SNPs in the coding region. A total of nine polymorphic sites corresponding to seven SNPs were identified (Table 2). All five genes displayed at least one SNP. Five SNPs consisted of single nucleotide substitutions, whereas two SNPs in *scr58* and *scr91* consisted of two-nucleotide changes. Two SNPs consisted of more than two polymorphic nucleotides. The SNP in *scr50* occurred as a C, G, or T, whereas the second SNP of *pex208* occurred as a T, G, or C. Six of the seven SNPs were nonsynonymous resulting in amino acid replacements. The distribution frequency of the identified SNPs was examined. Heterozygocity was assumed when two SNPs were identified in the same isolate and was detected at variable frequency for all seven SNPs (range 0.06–0.75). However, we cannot exclude at this stage that some of these genes occur as gene families and the two SNPs were amplified from two paralogs and not alleles.

Table 2 Catalog of single nucleotide polymorphisms identified in examined *Phytophthora infestans* genes

Gene[a]	Single nucleotide polymorphism (SNPs)[b]		Frequency[c]		
	Nucleotide (codons)	Amino acid	Homozygous (reference)	Heterozygous	Homozygous (mutant)
scr50	CTC to GTC	Leu^{8} to Val^{8}	0.88	0.06	0
	CTC to TTC	Leu^{8} to Phe^{8}	–	0.06	0
scr58	GCT to GTC	Ala^{12} to Val^{12}	0.31	0.69	0
scr76	TGT to TGC	Cys^{59} to Cys^{59} (silent)	0.38	0.56	0.06
scr91 (SNP #1)	GAA to GTA	Glu^{52} to Val^{52}	0.86	0.14	0
scr91 (SNP #2)	GCT CCC to GCG TCC	Ala^{76} Pro^{77} to Ala^{76} Ser^{77}	0	0.19	0.81
pex208 (SNP #1)	ACG to ATG	Thr^{29} to Met^{29}	0.19	0.75	0.06
pex208 (SNP #2)	ATC to AGC	Ile^{83} to Ser^{83}	0.13	0.75	0
	ATC to ACC	Ile^{83} to Thr^{83}	–	0.06	0
	AGC to ACC	Ser^{83} to Thr^{83}	–	0.06	–

[a]GenBank sequence accessions listed in Table 1. [b]The SNPs are underlined. Reference sequences correspond to the first codon listed. [c]A minimum of 16 *P. infestans* isolates were examined for each SNP. Heterozygous was assumed when two versions of a SNP were detected in one isolate. However, we cannot exclude at this stage that the two SNPs were amplified from paralogs and not alleles.

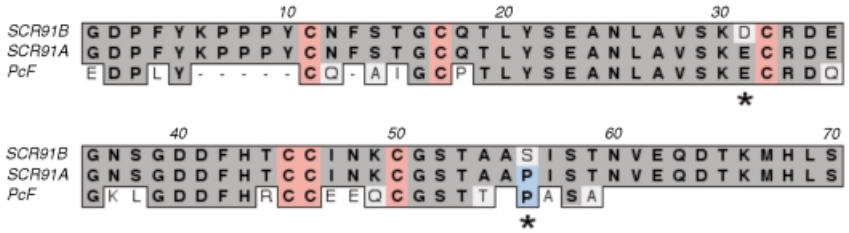

Fig. 2 Alignment of *Phytophthora cactorum* phytotoxic PcF protein with two *Phytophthora infestans* SCR91 predicted proteins. Identical amino acids are shaded in dark gray and similar amino acids shaded in light gray. Residue numbers are indicated above the sequences. Only the mature sequence of the proteins was included. The conserved cysteine residues are in pink and the proline residue that is hydroxylated in *P. cactorum* is highlighted in blue. Residues polymorphic in *P. infestans* are underscored with an asterisk.

We used the test for linkage disequilibrium described above to assess whether any of the seven SNPs is associated with one of the 11 known *Avr* genes of *P. infestans*. The C/T SNP detected at base 54 in the *scr58* gene resulted in an alanine to valine replacement (Table 2) and exhibited complete association ($D = 1$) with avirulence on the potato *R1* differential genotype. However, no reliable conclusion can be reached at this point since D is sensitive to low allele frequencies and small sample sizes. In this case the complete association may have resulted from the absence of one of the four possible haplotypes. Importantly, the majority of associations were between virulence and heterozygous SNPs, a result at variance with the genetics of avirulence towards the *R1* genotype.

The *pex208* gene exhibited a C/T SNP (SNP #1, Table 2) which showed stronger evidence for linkage disequilibrium with *Avr11* ($D = 0.43$). The strength of the association is not enough to warrant selecting this gene as a candidate *Avr11*, but the principle of combining SNP discovery with evaluation using test statistics such as D was validated.

Other tested candidate genes typically yielded only SNPs dissociated with the avirulence phenotypes tested. This was particularly of interest for *scr91*, which exhibits significant similarity to PcF, a necrosis-inducing protein from *P. cactorum* (Fig. 2; Orsomando *et al.*, 2001). *Scr91* is an attractive candidate *Avr* gene for two reasons. Firstly, *scr91* was shown in our laboratories to be strongly up-regulated during host plant infection by *P. infestans* (unpublished results), and, secondly, the Pro to Ser substitution caused by SNP #2 is in a potentially key residue. In *P. cactorum*, the specific activity of this protein may involve a unique 4-hydroxyproline residue near the C-terminal end of the protein (Orsomando *et al.*, 2001). In *scr91*, there is also a proline at this position, although it is not known if it is hydroxylated or not (Fig. 2). SNP #2 that leads to a proline-serine (P/S) substitution was detected at high frequency in the examined *P. infestans* isolates and could affect the biological activity of this protein. However, no association of this polymorphism with any of the 11 characterized avirulence phenotypes was observed. Nevertheless, it is worthy to note that the proline substitution was not detected in a homozygous form and was always observed in older populations of the US-1 clonal lineage of *P. infestans*, as well as in several race 0 isolates that cannot infect potato

differential lines that carry at least one of the 11 known *R* genes.

Promising polymorphism-to-phenotype associations will be further tested across a wider range of *P. infestans* isolates and across virulence-tested progeny from appropriate mapping populations. This will allow rigorous statistical evaluation of the significance of the associations. In addition, functional assays are under way to determine whether the examined polymorphic genes induce HR-like responses in late blight resistant *Solanum*.

Potential limitations and considerations

Selection criteria

The series of selection criteria used to identify candidate *Avr* genes form the central assumption behind this strategy. It is possible that *P. infestans* or other *Phytophthora* spp. carry some *Avr* genes with novel or unexpected structural features that can not be presently identified by data mining. For instance, the *Avr* elicitor may not be the direct product of an *Avr* gene but could be a secondary metabolite produced by a biochemical pathway, making it difficult to predict the genes involved. Nevertheless, it is reasonable to expect that many *Avr* genes will exhibit some of the obvious structural features, such as the presence of a secretion signal. Ultimately, as technologies for sequencing and SNP identification continue to improve, it will be possible to loosen the criteria and assay larger numbers of candidate genes.

Noncoding and complex polymorphisms

It is possible that in some races, the virulence allele of the *Avr* gene does not carry polymorphisms that result in structural alterations of the protein but rather mutations that affect gene expression and regulation. These polymorphisms might be harder to identify and interpret, but nevertheless they can still be used in classical and association genetic analyses. In fungal *Avr* genes, such as those cloned from *M. grisea*, the change in interaction phenotype from avirulent to virulent may result from many different types of polymorphisms in the *Avr* gene, or its expression. Therefore, for any candidate *Avr* gene it is essential to relate any polymorphism (presence/absence, insertions/deletions, or SNPs) within the gene, as a whole, to phenotype. In the *P. infestans* genes assayed thus far, only SNPs have been identified. In the future, differences in gene expression can also be assessed for correlation with avirulence.

Gene families

The genomes of many oomycete pathogens are large and complex. Of these, *P. infestans* is one of the largest at 240 Mb and contains high levels of repetitive DNA. Many of the *P. infestans* genes that have been determined to be up-regulated *in planta* are also members of gene families. Some candidate *Avr* genes occur as complex gene families with several closely related paralogs identified in one individual isolate. This creates technical difficulties in assessing the particular genotype of a given strain since PCR amplifications occasionally result in mixed amplicons that cannot be sequenced directly. Additional steps are then needed for cloning and sequencing the corresponding genomic regions in order to identify the various haplotypes and design specific markers for genotyping.

Genetic hitchiking effect

In identifying polymorphisms associated with avirulence, the genetic hitchhiking effect must be considered (Barton, 2000). For instance, genetic loci conditioning avirulence/virulence can be considered as potentially under strong selective pressure by deployment of resistance genes in potato and tomato. An *R* gene selective sweep through a given *P. infestans* population may not only increase the prevalence of the virulence allele, but would also increase the prevalence of linked but selectively neutral loci. *P. infestans* can reproduce both sexually and asexually, depending on whether both mating types are present in a population. In sexual populations, the hitchhiking effect breaks down with physical distance due to sexual recombination over many generations. In asexually reproducing populations, which are effectively clonal, the hitchhiking effect would act across the entire selected genotype. However, an *R* gene selective sweep would select against an *Avr* allele and any hitchhiking effect in the surrounding genomic regions would be due to population founder effects. To test whether a particular association between a SNP and an *Avr* allele is due to the hitchiking effect, the association should be tested in multiple *P. infestans* populations that have evolved in relative isolation from each other. If a single gene is responsible for the avirulence phenotype, the evolution to virulence should have occurred independently in separate populations. Ultimately, functional assays for avirulence are needed to validate whether the identified gene is an *Avr* gene.

Orphan *Avr* genes

In determining the extent of linkage disequilibrium between alleles of *P. infestans* genes and interaction phenotypes, the number of host genotypes considered and the robustness of the correlation must be taken into account. For instance, *P. infestans* isolates are routinely tested against potato genotypes carrying the resistance alleles *R1*–*R11*, and occasionally against differential tomato genotypes. However, many species of *Solanum* exhibit both HR-based resistance and susceptibility to *P. infestans* infection and are likely to bear a fairly large number of uncharacterized *R* genes (Kamoun

et al., 1999b; Vleeshouwers *et al.*, 2000). Such species include, among others, accessions of *S. papita*, *S. chacoense*, *S. verrucosum*, and *S. bulbocastanum*. Interestingly, the approach we describe in this article has the potential to identify *P. infestans Avr* genes that do not correspond to the known potato or tomato *R* genes, but correspond to unknown *R* genes that have not been genetically characterized. Such *Avr* genes are still expected to be polymorphic but may not associate with the known races of *P. infestans*. These genes might be termed 'orphan' *Avr* genes until screening of plant germplasm uncovers the matching *R* genes. Extensive infection surveys of host genotypes with a diverse collection of *P. infestans* isolates will permit orphan *Avr* gene–SNP associations to be tested in parallel with the typical *R1*/*Avr1*–*R11*/*Avr11* interactions.

Outlook

Compared to bacteria, the number of *Avr* genes cloned from fungal and oomycete plant pathogens remains relatively low (van't Slot & Knogge, 2002). This is mainly caused by the relatively large size of the genomes of phytopathogenic eukaryotes, and the lack of large scale transformation procedures that hinders the use of the functional complementation approach that has proven successful in the cloning of many bacterial *Avr* genes. In the oomycete *Phytophthora*, positional cloning has been the technique of choice for the cloning of race-specific *Avr* genes. However, even though projects for positional cloning of *Avr* genes from *P. infestans* and *P. sojae* have been running for 5–10 yr in several laboratories, to date no race-specific *Avr* genes from *Phytophthora* have been described in the literature.

In this paper, we describe an alternative strategy for cloning *Phytophthora Avr* genes. This approach exploits emerging sequence data for *Phytophthora* and combines intraspecific comparative genomics with data mining and functional analyses. In a collaborative effort between our two groups, we have been applying this strategy to the cloning of *Avr* genes from the late blight pathogen *P. infestans* and have obtained promising preliminary results. Eventually, this work will lead to the identification of novel genes from *P. infestans* that condition avirulence to resistant potato and tomato and will help unravel the molecular basis of host specificity in this important pathosystem. Interestingly, this approach has also the potential to identify *P. infestans* 'orphan' *Avr* genes that do not correspond to the known potato or tomato *R* genes, but correspond to uncharacterized *R* genes. A promising 'orphan' *Avr* gene candidate is *scr91* (Fig. 2). This gene encodes a secreted small cysteine-rich protein with similarity to a necrosis-inducing protein, is strongly up-regulated during infection of host plants, and is polymorphic in a key amino acid residue. Future work, inlcuding functional assays, will help confirm that *scr91* encodes an avirulence product and will unravel the nature of the matching *R* gene.

Acknowledgements

This research is part of a consortium of oomycete–plant interaction laboratories that includes the groups of Jim Beynon, Horticulture Research International, Wellesbourne, Warwick, UK; Vivianne Vleeshouwers, Wageningen University, Wageningen, The Netherlands; and Pieter van West, The University of Aberdeen, Aberdeen, Scotland, UK. We are grateful to these colleagues and their coworkers for useful discussions and access to unpublished information. Research at OSU is supported by NSF Plant Genome grant DBI-0211659, Syngenta Biotechnology, and State and Federal Funds appropriated to the Ohio Agricultural Research and Development Center, the Ohio State University. The SCRI group are grateful to the Scottish Executive Environment and Rural Affairs Department (SEERAD) for continuing financial support. JIBB, MA, and SCW contributed equally to this work and should be considered as cofirst authors.

References

Baldauf SL, RogerAJ, Wenk-Siefert I, Doolittle WF. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**: 972–977.

Barton NH. 2000. Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London: Biological Sciences* **355**: 1553–1562.

Birch PRJ, Whisson S. 2001. *Phytophthora infestans* enters the genomics era. *Molecular Plant Pathology* **2**: 257–263.

Dangl JL, Jones JD. 2001. Plant pathogens and integrated defence responses to infection. *Nature* **411**: 826–833.

Dixon MS, Golstein C, Thomas CM, van Der Biezen EA, Jones JD. 2000. Genetic complexity of pathogen perception by plants: the example of *Rcr3*, a tomato gene required specifically by *Cf-2*. *Proceedings of the National Academy of Sciences, USA* **97**: 8807–8814.

Erwin DC, Ribeiro OK. 1996. *Phytophthora diseases worldwide*. St. Paul, MN, USA: APS Press.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8**: 186–194.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* **8**: 175–185.

Fry WE, Goodwin SB. 1997a. Re-emergence of potato and tomato late blight in the United States. *Plant Disease* **81**: 1349–1357.

Fry WE, Goodwin SB. 1997b. Resurgence of the Irish potato famine fungus. *Bioscience* **47**: 363–371.

Hammond-Kosack KE, Jones JDG. 1997. Plant disease resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**: 575–607.

Innes RW. 2001. Targeting the targets of Type III effector proteins secreted by phytopathogenic bacteria. *Molecular Plant Pathology* **2**: 109–115.

Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B. 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO Journal* **19**: 4004–4014.

Jones L, Hamilton AJ, Voinnet O, Thomas CL, Maule AJ, Baulcombe DC. 1999. RNA–DNA interactions and DNA methylation in post-transcriptional gene silencing. *Plant Cell* **11**: 2291–2301.

Joosten MHAJ, Vogelsang R, Cozijnsen TJ, Verberne MC, de Wit PJGM. 1997. The biotrophic fungus *Cladosporium fulvum* circumvents Cf-4 mediated resistance by producing unstable AVR4 elicitors. *Plant Cell* **9**: 367–379.

Kamoun S. 2001. Nonhost resistance to *Phytophthora*: novel prospects for a classical problem. *Current Opinion in Plant Biology* **4**: 295–300.

Kamoun S. 2003. Molecular genetics of pathogenic oomycetes. *Eukaryotic Cell* **2**: 191–199.

Kamoun S, Dong S, Hamada W, Huitema E, Kinney D, Morgan WR, Styer A, Testa A, Torto T. 2002. From sequence to phenotype: Functional genomics of *Phytophthora*. *Canadian Journal of Plant Pathology* **24**: 6–9.

Kamoun S, Honee G, Weide R, Lauge R, Kooman-Gersmann M, de Groot K, Govers F, de Wit PJGM. 1999a. The fungal gene *Avr9* and the oomycete gene *inf1* confer avirulence to potato virus X on tobacco. *Molecular Plant–Microbe Interactions* **12**: 459–462.

Kamoun S, Huitema E, Vleeshouwers VGAA. 1999b. Resistance to oomycetes: a general role for the hypersensitive response? *Trends in Plant Sciences* **4**: 196–200.

Kamoun S, van West P, Vleeshouwers VG, de Groot KE, Govers F. 1998. Resistance of *Nicotiana benthamiana* to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. *Plant Cell* **10**: 1413–1426.

Kang S, Lebrun MH, Farrall L, Valent B. 2001. Gain of virulence caused by insertion of a Pot-3 transposon in a *Magnaporthe grisea* avirulence gene. *Molecular Plant–Microbe Interactions* **14**: 671–674.

Kang S, Sweigard JA, Valent B. 1995. The PWL host-specificity gene family in the blast fungus Magnaporthe grisea. *Molecular Plant–Microbe Interactions* **8**: 939–948.

Kim YJ, Lin NC, Martin GB. 2002. Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity. *Cell* **109**: 589–598.

Kooman-Gersmann M, Vogelsang R, Hoogendijk ECM, de Wit PJGM. 1997. Assignment of amino acid residues of the Avr9 peptide of *Cladosporium fulvum* that determine elicitor activity. *Molecular Plant–Microbe Interactions* **10**: 821–829.

Kruger J, Thomas CM, Golstein C, Dixon MS, Smoker M, Tang S, Mulder L, Jones JD. 2002. A tomato cysteine protease required for *Cf-2*-dependent disease resistance and suppression of autonecrosis. *Science* **296**: 744–747.

Lauge R, De Wit PJ. 1998. Fungal avirulence genes: structure and possible functions. *Fungal Genetics and Biology* **24**: 285–297.

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

van der Lee T, Testa A, van't Klooster JW, van den Berg-Velthuis G, Govers F. 2001b. Chromosomal deletion in isolates of *Phytophthora infestans* correlates with virulence on R3, R10, and R11 potato lines. *Molecular Plant–Microbe Interactions* **14**: 1444–1452.

Leister RT, Katagiri F. 2000. A resistance gene product of the nucleotide binding site – leucine rich repeats class can form a complex with bacterial avirulence proteins in vivo. *Plant Journal* **22**: 345–354.

Lewontin RC. 1964. The interaction of selection and linkage. I. General consideration of heterotic models. *Genetics* **49**: 49–67.

Luderer R, de Kock MJD, Dees RHL, de Wit PJM, Joosten MH. 2002a. Functional analysis of cysteine residues of ECP elicitor proteins of the fungal tomato pathogen *Cladosporium fulvum*. *Molecular Plant Pathology* **2**: 91–95.

Luderer R, Takken FL, de Wit PJ, Joosten MH. 2002b. *Cladosporium fulvum* overcomes *Cf-2*-mediated resistance by producing truncated AVR2 elicitor proteins. *Molecular Microbiology* **45**: 875–884.

MacGregor T, Bhattacharyya M, Tyler B, Bhat R, Schmitthenner AF, Gijzen M. 2002. Genetic and physical mapping of *Avr1a*. *Phytophthora sojae*. *Genetics* **160**: 949–959.

Margulis L, Schwartz KV. 2000. Five Kingdoms: *An illustrated guide to the phyla of life on Earth*. New York, USA: W.H. Freeman.

May K, Whisson S, Zwart R, Searle I, Irwin J, Maclean D, Carroll B, Drenth A. 2002. Inheritance and mapping of 11 avirulence genes in *Phytophthora sojae*. *Fungal Genetics and Biology* **37**: 1–17.

Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B. 2000. A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. *Plant Cell* **12**: 2019–2032.

Orsomando G, Lorenzi M, Raffaelli N, Rizza MD, Mezzetti B, Ruggieri S. 2001. Phytotoxic protein PcF, purification, characterization, and cDNA sequencing of a novel hydoxyproline-containing factor secreted by the strawberry pathogen *Phytophthora cactorum*. *Journal of Biological Chemistry* **276**: 21578–21584.

Panabieres F, Birch PR, Unkles SE, Ponchet M, Lacourt I, Venard P, Keller H, Allasia V, Ricci P, Duncan JM. 1998. Heterologous expression of a basic elicitin from *Phytophthora cryptogea*. *Phytophthora infestans* increases its ability to cause leaf necrosis in tobacco. *Microbiology* **144**: 3343–3349.

Perez-Garcia A, Snoeijers SS, Joosten MH, Goosen T, De Wit PJ. 2001. Expression of the avirulence gene *Avr9* of the fungal tomato pathogen *Cladosporium fulvum* is regulated by the global nitrogen response factor NRF1. *Molecular Plant–Microbe Interactions* **14**: 316–325.

Ponchet M, Panabieres F, Milat M-L, Mikes V, Montillet J-L, Suty L, Triantaphylides C, Tirilly Y, Blein J-P. 1999. Are elicitins cryptograms in plant-Oomycete communications? *Cellular and Molecular Life Sciences* **56**: 1020–1047.

Qutob D, Hraber PT, Sobral BW, Gijzen M. 2000. Comparative analysis of expressed sequences in *Phytophthora sojae*. *Plant Physiology* **123**: 243–254.

Qutob D, Kamoun S, Gijzen M. 2002. Expression of a *Phytophthora sojae* necrosis-inducing protein occurs during transition from biotrophy to necrotrophy. *Plant Journal* **32**: 361–373.

Rivas S, Mucyn T, van den Burg HA, Vervoort J, Jones JD. 2002a. An approximately 400 kDA membrane associated complex that contains one molecule of the resistance protein *Cf-4*. *Plant Journal* **29**: 783–796.

Rivas S, Romeis T, Jones JD. 2002b. The *Cf-9* resistance protein is present in an approximately 420-kilodalton heteromultimeric membrane-associated complex at one molecule per complex. *Plant Cell* **14**: 689–702.

Rohe M, Gierlich A, Hermann H, Hahn M, Schmidt B, Rosahl S, Knooge W. 1995. The race-specific elicitor, NIP1, from the barley pathogen, *Rhynchosporium secalis*, determines avirulence on host plants of the *Rrs1* resistance genotype. *EMBO Journal* **14**: 4168–4177.

Schiermeier Q. 2001. Russia needs help to fend off potato famine, researchers warn. *Nature* **410**: 1011.

Schneider DS. 2002. Plant immunity and film Noir: what gumshoe detectives can teach us about plant–pathogen interactions. *Cell* **109**: 537–540.

Shattock RC. 2002. *Phytophthora infestans*: populations, pathogenicity and phenylamides. *Pest Management Science* **58**: 944–950.

van't Slot KAE, Knogge W. 2002. A dual role for microbial pathogen-derived effector proteins in plant disease and resistance. *Critical Reviews in Plant Science* **21**: 229–271.

Smart CD, Fry WE. 2001. Invasions by the late blight pathogen: renewed sex and enhanced fitness. *Biological Invasions* **3**: 235–243.

Sogin ML, Silberman JD. 1998. Evolution of the protists and protistan parasites from the perspective of molecular systematics. *International Journal of Parasitology* **28**: 11–20.

Staskawicz BJ, Ausubel FM, Baker BJ, Ellis JG, Jones JDG. 1995. Molecular genetics of plant disease resistance. *Science* **268**: 661–667.

Sweigard JA, Caroll AM, Kang S, Farrall L, Chumley FG, Valent B. 1995. Identification, cloning, and characterization of *PWL2*, a gene for host species specificity in the rice blast fungus. *Plant Cell* **7**: 1221–1233.

Swiderski MR, Innes RW. 2001. The *Arabidopsis* PBS1 resistance gene encodes a member of a novel protein kinase subfamily. *Plant Journal* **26**: 101–112.

Torto TA, Li S, Styer A, Huitema E, Testa A, Gow NAR, van West P, Kamoun S. 2003. EST mining and functional expression assays identify effector proteins from the plant pathogen *Phytophthora*. *Genome Research* (In press.)

Tyler BM. 2002. Molecular basis of recognition between *Phytophthora* pathogens and their hosts. *Annual Review of Phytopathology* **40**: 137–167.

Van der Biezen EA, Jones JD. 1998. Plant disease-resistance proteins and the gene-for-gene concept. *Trends in Biochemal Sciences* **23**: 454–456.

Van der Hoorn RA, Laurent F, Roth R, De Wit PJ. 2000. Agroinfiltration is a versatile tool that facilitates comparative analyses of *Avr9*/*Cf-9*-induced and *Avr4*/*Cf-4*-induced necrosis. *Molecular Plant–Microbe Interactions* **13**: 439–446.

Vleeshouwers VGAA, van Dooijeweert W, Govers F, Kamoun S, Colon LT. 2000. The hypersensitive response is associated with host and nonhost resistance to *Phytophthora infestans*. *Planta* **210**: 853–864.

Whisson SC, van der Lee T, Bryan GJ, Waugh R, Govers F, Birch PR. 2001. Physical mapping across an avirulence locus of *Phytophthora infestans* using a highly representative, large-insert bacterial artificial chromosome library. *Molecular Genetics and Genomics* **266**: 289–295.