Ancient Origin of Elicitin Gene Clusters in *Phytophthora* Genomes

Rays H. Y. Jiang,* Brett M. Tyler,† Stephen C. Whisson,‡ Adrienne R. Hardham,§ and Francine Govers*

*Laboratory of Phytopathology, Wageningen University, Wageningen, and Graduate School Experimental Plant Sciences, The Netherlands; †Virginia Bioinformatics Institute, Virginia Polytechnic and State University; ‡Plant Pathogen Interactions Programme, Scottish Crop Research Institute, Invergowrie, Dundee, Scotland, United Kingdom; and §Plant Cell Biology Group, Research School of Biological Sciences, Australian National University, Canberra, Australia

The genus *Phytophthora* belongs to the oomycetes in the eukaryotic stramenopile lineage and is comprised of over 65 species that are all destructive plant pathogens on a wide range of dicotyledons. *Phytophthora* produces elicitins (ELIs), a group of extracellular elicitor proteins that cause a hypersensitive response in tobacco. Database mining revealed several new classes of elicitin-like (ELL) sequences with diverse elicitin domains in *Phytophthora infestans*, *Phytophthora sojae*, *Phytophthora brassicae*, and *Phytophthora ramorum*. ELIs and ELLs were shown to be unique to *Phytophthora* and *Pythium* species. They are ubiquitous among *Phytophthora* species and belong to one of the most highly conserved and complex protein families in the *Phytophthora* genus. Phylogeny construction with elicitin domains derived from 156 ELIs and ELLs showed that most of the diversified family members existed prior to divergence of *Phytophthora* species from a common ancestor. Analysis to discriminate diversifying and purifying selection showed that all 17 ELI and ELL clades are under purifying selection. Within highly similar ELI groups there was no evidence for positively selected amino acids suggesting that purifying selection contributes to the continued existence of this diverse protein family. Characteristic cysteine spacing patterns were found for each phylogenetic clade. Except for the canonical clade ELI-1, ELIs and ELLs possess C-terminal domains of variable length, many of which have a high threonine, serine, or proline content suggesting an association with the cell wall. In addition, some ELIs and ELLs have a predicted glycosylphosphatidylinositol site suggesting anchoring of the C-terminal domain to the cell membrane. The *eli* and *ell* genes belonging to different clades are clustered in the genomes. Overall, *eli* and *ell* genes are expressed at different levels and in different life cycle stages but those sharing the same phylogenetic clade appear to have similar expression patterns.

Introduction

The genus *Phytophthora* comprises over 65 phytopathogenic species that cause many economically important diseases and can have devastating effects on natural habitats (Erwin and Ribeiro 1996). *Phytophthora infestans*, also known as the notorious "Irish potato famine fungus," causes late blight disease on potato and tomato worldwide, and *Phytophthora sojae* is responsible for root and stem rot on soybean. Two recently discovered species are *Phytophthora ramorum*, the causal agent of "sudden oak death" (Werres et al. 2001; Rizzo, Garbelotto, and Hansen 2005), and *Phytophthora brassicae*, a pathogen on the model plant *Arabidopsis thaliana* (Roetschi et al. 2001; Man in 't Veld et al. 2002). *Phytophthora* belongs to the oomycetes, a diverse group of fungus-like eukaryotes that share phylogenetic similarity with brown algae and diatoms. In the tree of life oomycetes are grouped in the stramenopile lineage that is distant from the plant, animal, and fungal lineages (Margulis and Schwarts 2000; Baldauf 2003).

A common feature of many different types of plant pathogens is the secretion of a variety of extracellular effector molecules into the plant apoplast (van't Slot and Knogge 2002) that are presumed to promote infection of the host plant. Many of these proteins, called elicitors, elicit plant defense responses and, in particular, a form of programmed cell death called the hypersensitive response (HR). In most cases, the defense response benefits the plant, and the response is triggered by the detection of the elicitors by plant defense receptors. In some cases, however, elicitation of these responses promotes infection because the pathogen can thrive on the dying plant tissue. *Phytophthora* species ubiquitously secrete a unique class of highly conserved effector molecules named elicitins. Elicitins are widespread in *Phytophthora* species and closely related *Pythium* species (Panabieres et al. 1997) but are absent from any other organism studied so far. Hence, it is conceivable that they could be responsible for novel mechanisms of interaction with plants.

Molecular cloning and expressed sequence tag (EST) database analysis showed that elicitin genes form families in *Phytophthora* species such as *P. infestans* (Kamoun, Lindqvist, and Govers 1997; Kamoun et al. 1999), *P. sojae* (Mao and Tyler 1996; Qutob et al. 2003), *P. brassicae* (L. Belbahri and F. Mauch, personal communication), and *Phytophthora cinnamomi* (Duclos et al. 1998). Gene families are considered to arise from chromosomal duplications. In theory, only one copy must maintain the original function, whereas the other copies can undergo functional divergence. Remarkably, in *Phytophthora* most of the extracellular proteins described to date are encoded by multigene families (Gotesson et al. 2002; Qutob, Kamoun, and Gijzen 2002; Torto et al. 2003; Liu et al. 2005). Some of these gene families are *Phytophthora* specific whereas others have homologues in plant pathogens in other kingdoms such as fungi and bacteria.

Based on the phylogenetic distribution pattern, elicitor genes can roughly be divided into three groups (fig. 1). Group-A elicitors have homologues across kingdoms. The most prominent example is a family of necrosis-inducing proteins called NIP or NPP that belongs to the superfamily of Nep1-like proteins (NLPs) present in bacteria, fungi, and stramenopiles (Qutob, Kamoun, and Gijzen 2002; Pemberton and Salmond 2004). Elicitins are typical

Archaea
Bacteria
Eukaryota
...
| Euglenozoa
| Alveolates
| Plants
| Animals
| Fungi
| Stramenopiles
     Oomycetes
         | Pythium
         | Phytophthora
              | P. sojae
              | P. ramorum
              | P. infestans
              | P. brassicae
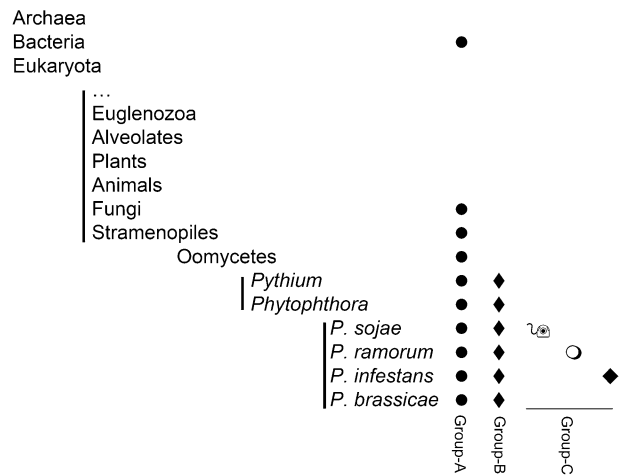
Group-A   Group-B   Group-C

FIG. 1.—Phylogenetic distribution of *Phytophthora* elicitors. Group-A elicitors are distributed across kingdoms, Group-B elicitors are restricted to oomycetes and conserved between species, and Group-C elicitors are species specific or highly divergent between species. A ● indicates that in one or more members of this group a homologue of the elicitor is found. Only the major eukaryotic groups from which one or more representatives have been fully sequenced are listed.

examples of Group-B elicitors which are conserved but only present in one or a few genera. Group-C elicitors are species specific or highly divergent such as IPI-O (Pieterse et al. 1994), AVR3a (Armstrong et al. 2005), and SCR74 (Liu et al. 2005) in *P. infestans* and AVR1b in *P. sojae* (Shan et al. 2004). The different phylogenetic distributions of elicitors may relate to the role they have in the interaction with host plants. Some of the highly divergent Group-C elicitors, for example, are race-specific elicitors (i.e., AVR3a and AVR1b of *P. infestans* and *P. sojae*, respectively) and are involved in highly specific gene-for-gene interactions with major nucleotide binding site-leucine rich repeat (NBS-LRR) resistance genes.

To address the question of how the elicitin gene family evolved and what kind of functions the different family members may have, a thorough investigation of the diversity of the family members is needed. Elicitins all share a conserved domain with a characteristic signature of six cysteine residues that form three distinct disulfide bonds (Fefeu et al. 1997). These features together with the small sizes of the proteins allow efficient mining of EST databases and genome sequences by PSI-Blast (Altschul et al. 1997). To investigate the diversity and evolutionary relationship within the elicitin gene family, we searched for new elicitin gene family members by making use of *Phytophthora* EST databases (Randall et al. 2005) and whole genome sequences of *P. sojae* and *P. ramorum* (http://genome.jgi-psf.org). Subsequently, we classified all family members based on sequence diversity and protein motifs and constructed a phylogenetic tree to reveal their evolutionary relationships. We also investigated the molecular evolution within clades by calculating the rate of nonsynonymous to synonymous substitutions (d_N/d_S or ω) and the spatial and temporal expression patterns of various family members by northern blot hybridization and transcript counting in EST databases. Finally, we analyzed and compared the genome organization of elicitin genes across *Phytophthora* species.

Materials and Methods
Genome Databases and EST Databases

The *P. infestans* and *P. sojae* EST databases are accessible at http://www.pfgd.org and http://staff.vbi.vt.edu/estap/, and most *Phytophthora* EST sequences are available through GenBank (Kamoun et al. 1999; Qutob et al. 2000). *Blumeria graminis*, *Magnaporthe grisea*, and *Cladosporium fulvum* EST databases were downloaded from Phytopathogenic Fungi and Oomycete EST Database Version 1.4 (Soanes et al. 2002) (http://cogeme.ex.ac.uk), and the *M. grisea* genome sequence was available at http://www.broad.mit.edu/annotation/fungi/magnaporthe/ (Dean et al. 2005). The genomic sequences and annotated protein sequences of *P. sojae*, *P. ramorum*, and of the diatom *Thalassiosira pseudonana* (Armbrust et al. 2004) were obtained from the Web site of the Department of Energy, Joint Genome Institute (http://www.jgi.doe.gov/genomes).

Nucleic Acid Manipulations

For the isolation of sporangia, zoospores, and germinating cysts, *P. infestans* strain NL-88069 was grown at 18°C in the dark on rye agar medium supplemented with 2% sucrose. Tissue collection, RNA isolation, and Northern hybridization were performed as previously described by van West et al. (1998).

Hybridization screening of the *P. infestans* bacterial artificial chromosome (BAC) library was carried out as described in Whisson et al. (2001). BAC contig building was performed as described by Jiang et al. (2005).

Database Mining

All known ELI and ELL sequences of different *Phytophthora* species and *Pythium* species were retrieved from GenBank. The elicitin domains of all ELIs and ELLs were used to construct an elicitin domain database, from which position-specific scoring matrices (PSSMs) were generated. Subsequently, PSI-Blast was performed by screening EST databases and annotated protein databases with the PSSMs to identify new ELI and ELL candidates with an E value cutoff of 0.05. PSI-Blast hits were placed in a candidate data set. BlastN of each individual candidate was performed against all identified elicitin nucleotide sequences. Hits with E value less than 10^{-30} were considered to be the known ELIs or ELLs and discarded, whereas hits with E value between 10^{-30} and 0.05 and the characteristic six cysteine signature were considered to be the new ELIs or ELLs and analyzed manually. These were then added to the screening process until no new ELIs or ELLs could be found.

Bioinformatics Tools

Sequences were analyzed in Vector NTI 8 package. For Blast searches we used the National Center for Biotechnology Information Blast program and the Standalone-Blast Version 2.2.3 (Altschul et al. 1997). Multiple sequence alignment was performed by ClustalX 1.8, and for phylogenetic tree construction Molecular Evolutionary Genetic Analysis 2.1 (MEGA) (Kumar et al. 2001) was used. Phylogeny reconstruction of ELI and ELL domains

was performed by Neighbor-Joining analysis. Poisson correction was chosen as the distance parameter as specified in the program MEGA. The inferred phylogeny was tested by 1,000 bootstrap replicates. Signal peptides were predicted by SignalP 2.0 (Krogh et al. 2001), and transmembrane domain prediction was performed with the program SOSUI (Hirokawa, Boon-Chieng, and Mitaku 1998). For glycosyl-phosphatidylinositol (GPI) anchor prediction, big-PI plant predictor (Eisenhaber et al. 2003) was used. Protein motifs were searched against the Prosite database (Bairoch 1991; Sigrist et al. 2002).

Global d_N/d_S ratios were calculated by the fast diversifying/purifying selection detection program with single likelihood ancestor counting (SLAC) analysis (Pond and Frost 2005). Tests for purifying or diversifying selection were performed with the codeml program in the Phylogenetic Analysis by Maximum Likelihood (PAML) v3.14 package (Yang 1997; Yang et al. 2000). Models M0, M1a, M2a, M7, and M8 were used for the analysis. Positively selected amino acid sites were assigned based on $P > 95\%$ with Bayes empirical Bayes statistics (Yang, Wong, and Nielsen 2005) in model M2a. Calculation scripts were written in Python 2.2 (http://www.python.org) and are available from the authors upon request.

Results
Elicitins and Elicitin-Like Proteins in *Phytophthora*

Previous studies in *P. infestans*, *P. sojae*, and *P. brassicae* showed that elicitins are encoded by complex gene families. In *P. infestans*, seven *inf* genes have been cloned by either low-stringency hybridization with heterologous probes, polymerase chain reaction amplification with degenerate primers, or random sequencing of cDNAs and were named *inf1*, *inf2a*, *inf2b*, *inf3*, *inf4*, *inf5*, *inf6*, and *inf7* (Kamoun, Lindqvist, and Govers 1997; Kamoun et al. 1997, 1999). In *P. sojae*, Qutob et al. (2003) identified an elicitin gene family comprised of *sojA*, *sojB*, *soj2*, *soj3*, *soj5*, and *soj6* and three family members with diverse sequences named *soj7*, *sojX*, and *sojY*. In *P. brassicae*, five members of the elicitin gene family have been described (L. Belbahri and F. Mauch, unpublished data). The proteins encoded by *bra1*, *bra2*, *bra5*, and *bra6* share the highly conserved 98-amino acid elicitin domain, while the elicitin domain in BRA7 is more diverse. In many *Phytophthora* species and a few *Pythium* species, one major elicitin has been identified as an abundantly secreted protein, and the protein sequences of several of these have been deposited into GenBank (supplementary material table S2, Supplementary Material online). Phylogenetic trees published previously distinguished different elicitin classes. However, proteins with a diverse elicitin domain were not classified and referred to as elicitin-like (Kamoun, Lindqvist, and Govers 1997; Qutob et al. 2003).

To name elicitins and elicitin-like proteins in a systematic and consistent way, we propose a novel classification system and three letter abbreviations for individual proteins. The elicitins sharing a highly conserved 98-amino acid domain with six cysteine residues and a typical elicitin type cysteine spacing pattern are classified as ELIs, and they are labeled with the first three or four letters of the species name followed by a number, such as INF1, SOJ1, BRA1, and RAM1. Elicitin-like proteins possessing shorter or longer elicitin domains that are more diverse at the sequence level than the conserved domains in ELIs are classified as ELLs. The ELLs of *P. infestans* are named INL, of *P. sojae* SOL, of *P. brassicae* BRL, and of *P. ramorum* RAL. Consequently, in this paper we renamed *P. sojae* SOJ7, SOJX, and SOJY, *P. infestans* INF7, and *P. brassicae* BRA7 into SOL1A, SOL6, SOL3A, INL1, and BRL1B, respectively.

Diverse Elicitin Gene Family Members in *P. infestans*, *P. sojae*, *P. brassicae*, and *P. ramorum*

Mining of 35,266 EST contig sequences comprised of transcripts derived from various developmental stages of *P. infestans* (Randall et al. 2005) revealed 11 new members of the elicitin gene family bringing the total number in *P. infestans* to 19. In a similar way, 13,234 *P. sojae* and 5,863 *P. brassicae* EST contig sequences were mined revealing 6 and 4 new members, respectively, and resulting in totals of 16 in *P. sojae* and 9 in *P. brassicae*. Except for SOJC and SOJ3B, all new members found in the EST libraries are ELLs, resulting in 21 new ELLs in total.

To analyze the diversity of elicitin gene families in *Phytophthora* at the whole genome level, we mined the assembled draft genome sequences of *P. sojae* and *P. ramorum*. In total, 18 SOJ and 39 SOL domains were found in *P. sojae* including the eight SOJs and seven SOLs extracted from the EST databases. In *P. ramorum*, 17 RAM and 31 RAL domains were found. As yet, there is no *P. ramorum* EST database available. A few *ell* genes (*sol2C*, *ral13A*, *sol13A*, and *sol13H*) found in the genome sequences encode repeated elicitin domains, but it is not known if these genes are active.

To investigate the presence of elicitins in other eukaryotic filamentous plant pathogens, 3,021 unique EST sequences of *B. graminis*, 513 unique EST sequences of *Cladosporium fulvum*, and 8,821 unique EST sequences and the draft genome sequence of *M. grisea* were analyzed using the same mining methods. In addition, the genome sequence of the marine diatom *T. pseudonana*, which as *Phytophthora*, belongs to the stramenopile lineage, was used for mining. No ELI or ELL sequences were found in any of these species. Other plant pathogenic oomycetes may produce elicitins but to our knowledge this has not been reported.

Phylogenetic Reconstruction of ELIs and ELLs

A total of 156 elicitin domains derived from 128 ELIs and ELLs from *P. brassicae*, *P. infestans*, *P. ramorum*, and *P. sojae* (supplementary material table S1, Supplementary Material online) and several additional ELIs identified in other *Phytophthora* species and two *Pythium* species (supplementary material table S2, Supplementary Material online) were used to construct a phylogenetic tree (fig. 2). Seventeen distinctive clades with high bootstrap values (>60) could be identified and most show bootstrap values higher than 80. Four are ELI clades that together form a distinct branch. The remaining 13 are ELL clades. ELL clades are typically more divergent than ELI clades.

Orthologues are homologues separated by a speciation process, for example, INF3 of *P. infestans* and SOJ3A of *P. sojae*. Paralogues are homologues generated by a gene duplication event, such as *P. sojae* SOJ3A and SOJ3B. From the tree it is clear that every clade containing more than two ELIs or ELLs is comprised of orthologues. All clades have ELIs or ELLs derived from *P. sojae* and *P. ramorum*, and in most cases genes from these two species are overrepresented. This is obviously due to the fact that the complete genome of these two species was sampled, including potential pseudogenes. All four ELI clades have a *P. infestans* member as have eight of the 13 ELL clades. The nine *P. brassicae* BRAs and BRLs are present in seven different clades. Of the 17 clades, 13 have members of three or more species and this strongly suggests that the diversity in the elicitin gene family existed before these species evolved.

In the phylogenetic tree shown in figure 2 the ELI-1 clade is the largest with 32 members. However, only 13 of these belong to the four species that we focus on in this paper. The other 19 ELI-1 elicitins were identified in various other *Phytophthora* species and in *Pythium vexans* (supplementary material table S2, Supplementary Material online). ELI-1 elicitins are easy to identify: the mature protein is just the 98-amino acid elicitin domain and they are the most abundantly secreted proteins in culture filtrates. The ELI-1 clade includes the previously identified class I-A, class I-B, class II, and class Py elicitins (Kamoun, Lindqvist, and Govers 1997; Qutob et al. 2003). This subclassification can be resolved when only the ELI-1 elicitins are used as input for phylogenetic tree construction (data not shown).

Interestingly, *P. infestans* INF4 has no apparent orthologue in *P. sojae* or *P. ramorum* and is thus not covered by any clade. Also OLI, an elicitin from *Pythium oligandrum* with a highly divergent elicitin domain, is not assigned to any specific clade. SOJ3X and RAM3X share sequence homology with ELI-2 and ELI-3 clades, but the apparent orthologues cannot be found. Therefore, neither SOJ3X nor RAM3X is covered by a clade. For similar reasons, no specific clade was assigned to SOL1E or SOL11F.

The 13 ELL clades show large sequence diversity within members of the same clade as well as between members of other ELL clades. The ELL-13 clade is the largest clade (18 members) and comprises ELLs with the most diverse elicitin domains.

ELI and ELL Clades Are Highly Conserved Across *Phytopththora* Species and Are Under Purifying Selection

The 17 ELI and ELL clades belong to the most conserved elicitors identified in *Phytophthora*. A Blast search of the *P. sojae* draft genome sequence with *P. infestans* ELIs and ELLs and a set of (putative) elicitors of *P. infestans* resulted in similarity matches with a broad range of *E* values. By plotting the Blast identity percentages the level of conservation of the genes between *P. infestans and P. sojae* can be visualized (fig. 3). Together with the highly conserved NLP NPP1 (a Group-A elicitor), ELIs and ELLs (Group-B elicitors) are more conserved than all other elicitors. Other Group-B elicitors, CRN1 and CRN2 (Torto et al. 2003), are undergoing expansion and gene loss in *Phytophthora* species (R. H. Y. Jiang and F. Govers, unpub-

lished data) and they show less sequence conservation than ELIs and ELLs. The Group-C elicitors (SCR74, AVR3a, and IPI-O) are highly divergent between *P. infestans* and *P. sojae* and thus belong to the least conserved elicitors. The ELI INL4A is the previously described mating associated factor M-25 (Fabritius, Cvitanich, and Judelson 2002) and in the analysis it appears to be the least conserved elicitin family member. Interestingly, family members of another mating associated secreted protein, M-96 (Fabritius, Cvitanich, and Judelson 2002), are highly divergent (J. H. Y. Jiang and F. Govers, unpublished data).

The survival of orthologues after speciation is due to selection pressure exerted on the genes. For a protein-coding gene, selection is estimated by comparing the rate of nonsynonymous nucleotide substitutions per nonsynonymous sites (d_N, amino acid replacing) and synonymous nucleotide substitution per synonymous sites (d_S, silent). The ratio d_N/d_S, denoted by ω, is used as a measure of selective pressure at the protein level and ω values of 1, <1 and >1 indicate neutral, purifying and diversifying selection, respectively. By using the fast diversifying/purifying selection detection program DataMonkey with SLAC analysis (Pond and Frost 2005), all 17 clades of *eli's* and *ell's* show overall ω values lower than 1 (table 1) which indicates purifying selection. The highly conserved *eli* clades were also analyzed with the codeml program of the PAML package developed by Yang (1997) and Yang et al. (2000). All *eli* clades show ω_0 value ranging from 0.02 to 0.09, and no positively selected sites could be detected (table 1). Moreover, in the highly similar *soj2*, *soj3*, *ram2*, and *ram3* groups within the *eli2* and *eli3* clades no positive selection was found. For comparison, we subjected *P. infestans scr74* sequences to the same analysis. In contrast to the conserved *eli's* and *ell's*, *scr74* is a Group-C elicitor with only weak homologues in other *Phytophthora* species. The *scr74* gene family was shown to be under diversifying selection in *P. infestans* (Liu et al. 2005), and eight positive selection sites identified by Liu et al. (2005) were also detected with the more stringent criteria that we used in this study (table 1). These results suggest that in the different *Phytophthora* species the complex ELI family is maintained by purifying selection.

Elicitin Domains Show Clade-Specific Cysteine Spacing Patterns

The elicitin domains of ELIs and ELLs are of variable length but they all contain six cysteine residues at conserved positions. The six cysteines form three disulfide bonds that stabilize the α-helix–folded protein (Boissy et al. 1996; Fefeu et al. 1997). Based on the cysteine spacing pattern, the ELIs and ELLs can be classified in distinct groups, and these groups coincide with the classification in clades based on the phylogenetic reconstruction (table 2).

All four ELI clades fall in one group and have a cysteine spacing pattern of C_1-23-C_2-23-C_3-4-C_4-14-C_5-23-C_6. In contrast, each ELL clade has a typical cysteine spacing pattern. The number of amino acid residues between C_3 and C_4 (i.e., four) is conserved in all groups, whereas the two most variable regions are between C_1 and C_2 and between C_5 and C_6. Thus, the cysteine spacing
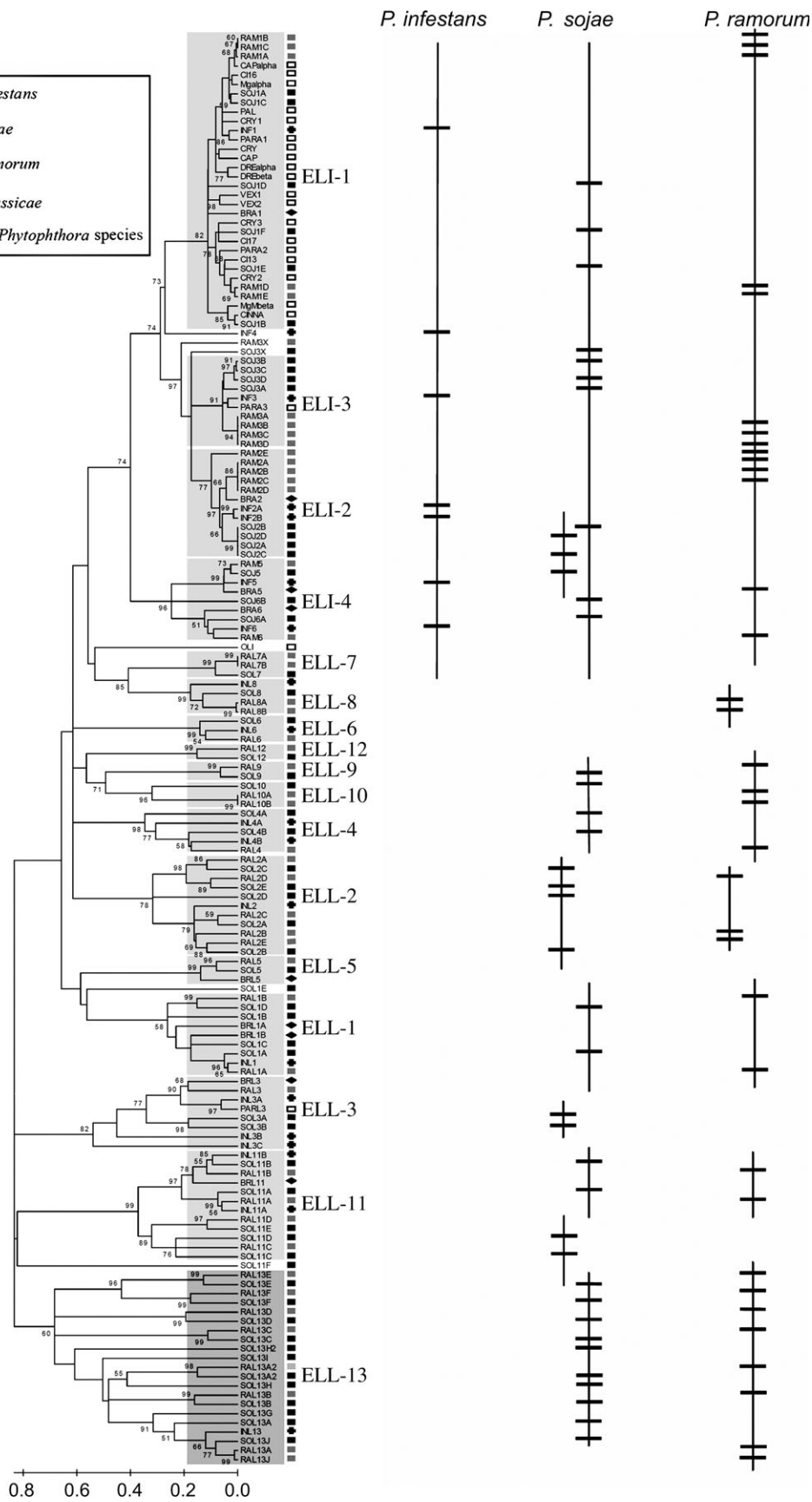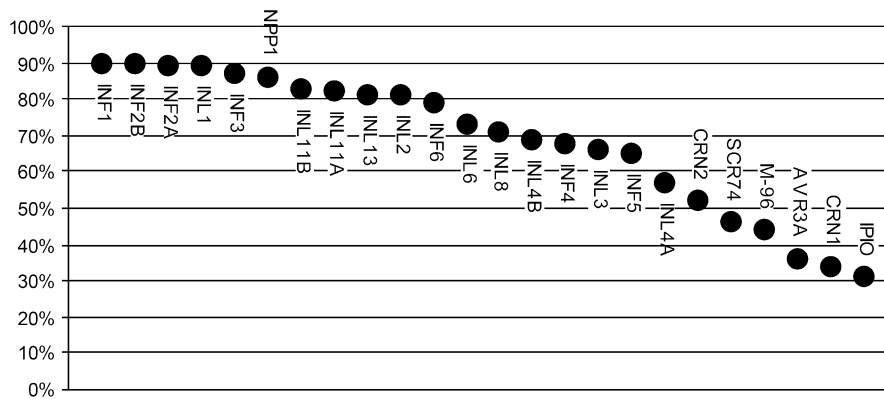
FIG. 3.—Conservation of ELIs and ELLs between *Phytophthora infestans* and *Phytophthora sojae*. The percent identity of *P. infestans* ELIs and ELLs to *P. sojae* is plotted and indicated on the *y* axis. The identity scores of several previously identified elicitor(-like) genes and gene family members are included for comparison: NPP1 (AAK25828), CRN2 (AAN31502), SCR74 (AAU21463), M-96 (AAN37691), AVR3a (CAI72254), CRN1 (AAN31500), and IPI-O (AAA21422). Signal peptides were omitted from the analysis. The labels of ELIs and ELLs are positioned below the data points and those of the other elicitor(-like) proteins above the data points.

pattern of the overall elicitin family can be visualized as C_1-variable-C_2-23-C_3-4-C_4-14-C_5-variable-C_6. The three disulfide bonds are formed between C_1 and C_5, C_2 and C_4, and C_3 and C_6. The two variable regions correspond to the two alignment gaps in a multiple sequence alignment as shown for *P. infestans* INFs and INLs in figure 4. In the sequence alignment the region between C_2 and C_4 is most conserved.

Most ELIs and ELLs Have C-Terminal Domains with Typical Repeat Structures and GPI Anchors

Most ELIs and ELLs are predicted to possess a signal peptide at the N-terminus in front of the conserved elicitin domain and an extended C-terminal domain following the conserved domain (supplementary material table S1, Supplementary Material online). ELIs and ELLs in 14 of the 17 clades have C-terminal domains ranging in length from 17 to 291 amino acids, whereas ELL-7, ELL-9, and ELL-10 members have shorter C-terminal domains of up to seven amino acids. The majority of the ELI-1 proteins lack a C-terminal domain. Only 11 out of the 32 known ELI-1 elicitins have a short C-terminal tail, and in the initial ELI-1 subclassification, these were classified as class II. Hence, most ELI-1 proteins are comprised solely of a signal peptide and the conserved 98-amino acid elicitin domain. In *P. sojae* and *P. ramorum*, the ELI-1 clade has several members but *P. infestans* has only one. In contrast, *P. infestans* has a second elicitin without a C-terminal domain, that is, INF4, which has no orthologues in the other species.

Many of the C-terminal domains appear to have a biased amino acid composition. They are particularly rich in threonine, serine, and proline residues, and quite often these residues are part of a repeat. Despite the fact that the phy-

logenetic tree of the elicitin family (fig. 2) was constructed with only the conserved elicitin domain, the C-terminal domains of ELIs and ELLs show clade-specific features, not only in the amino acid composition but also in the repeat structure. Table S1 (Supplementary Material online) summarizes the features of C-terminal domains of 129 ELIs and ELLs, and in figure S1 (Supplementary Material online) for each clade an example of a C-terminal domain is shown with the three most abundant amino acids highlighted. In the C-terminal domains of several of the ELI-2, ELI-3, ELI-4, ELL-1, ELL-2, and ELL-13, proteins more than 40% of the amino acid residues is comprised of threonine and serine. The C-terminal domains of ELI-4 and ELL-8 proteins are rich in proline, an amino acid that does not have a backbone proton and can easily form turns in the secondary protein structure (fig. S1, Supplementary Material online). In many of the C-terminal domains repeat units can be recognized such as the "APSAE" repeat unit in BRA5 and the "SA" repeats in INL2. They are comprised of two to five amino acids and can be repeated up to 15 times. The presence of several O-GalNAc-glycosylation sites as predicted by the program NetOGlyc 3.1 (Julenius et al. 2005) suggests that the C-terminal domains are glycosylated.

Several classes of ELLs such as ELL-1, ELL-2, and ELL-13 seem to possess hydrophobic regions at the extreme C-terminal end as predicted by the program SOSUI (Hirokawa, Boon-Chieng, and Mitaku 1998). These hydrophobic regions are part of the GPI anchor site predicted by the program big-PI plant predictor (Eisenhaber et al. 2003). In proteins carrying such a motif the hydrophobic C-terminal end is cleaved off from the mature protein and instead a GPI is added that will anchor the protein to the plasma membrane.

←

FIG. 2.—Phylogram of ELI and ELL amino acid sequences from *Phytophthora brassicae*, *Phytophthora infestans*, *Phytophthora ramorum*, *Phytophthora sojae*, and other *Phytophthora* spp. The conserved elicitin domains were used to construct the unrooted phylogram based on Neighbor-Joining analysis. Confidence of groupings was estimated by using 1,000 bootstrap replicates; numbers next to the branching point indicate the percentage of replicates supporting each branch. The shaded blocks show the ELIs or ELLs belonging to the same clade. Clade ELL-13 with the most diverse sequences is shaded with a grid pattern. On the right the physically linked *eli* and *ell* genes are schematically drawn. The length of the lines does not represent the physical distance. Lists of ELIs and ELLs are available as supplementary material (see tables S1 and S2 of Supplementary material online).

Table 1
Selection Test Based on Codeml Model M2a[a] of PAMLv3.14

Gene (sub) Family[b]	Number of Sequences[c]	Positively Selected Sites[d]	p_0	ω_0	p_1	ω_1	p_2	ω_2	κ^e	Overall ω^f
eli-1	13	None	0.92	0.08	0.08	1.00	0.00	—	2.08	0.22
eli-2	11	None	1.00	0.09	0.00	1.00	0.00	5.29	2.12	0.14
eli-3	9	None	0.98	0.03	0.00	1.00	0.02	1.83	3.81	0.26
eli-4a	4	None	0.94	0.04	0.04	1.00	0.02	16.76	2.94	0.11
eli-4b	4	None	0.85	0.06	0.15	1.00	0.00	—	2.00	0.16
scr74	20	8	0.73	0.15	0.00	1.00	0.27	9.19	3.35	2.28

[a] The site model allows the ω ratio to vary among sites (among codons or amino acids in the protein) (Yang et al. 2000). $\omega_0 < 1$ and $\omega_2 > 1$ are estimated from the data while $\omega_1 = 1$ is fixed.

[b] For the calculation only the sequences encoding the conserved elicitin domain and signal peptide were included.

[c] For *eli-1*, only the sequences derived from *Phytophthora brassicae*, *Phytophthora infestans*, *Phytophthora ramorum*, and *Phytophthora sojae* were used. The pseudogene *inf3* was not included in the *eli-3* sequences. *eli-4a* consists of *bra5*, *ram5*, *inf5*, and *soj5*. *eli-4b* consists of *bra6*, *inf6*, *ram6*, and *soj6*. A set of 20 randomly selected *scr74* genes (Liu et al. 2005) was used for comparison.

[d] Positively selected amino acid sites were assigned based on $P > 95\%$ with Bayes empirical Bayes statistics (Yang, Wong, and Nielsen 2005).

[e] κ is the estimated transition/transversion rate parameter.

[f] The overall ω value is based on SLAC analysis (Pond and Frost 2005).

Clustering of *eli* and *ell* Genes in the Genomes

In *P. infestans*, seven *eli* genes (*inf1*, *inf2a*, *inf2b*, *inf3*, *inf4*, *inf5*, and *inf6*) and one *ell* gene (*inl1*) were shown to be single copy genes by genomic Southern blot hybridization (data not shown). One physical contig of 250 kb spanning the seven *inf* genes was obtained by BAC library screening and contig building (fig. 5A). Sequencing and annotation of one of the BACs containing four *inf* genes showed that the average spacing between *inf* genes and other genes is 20 kb (Jiang et al. 2005).

Also in *P. sojae* and *P. ramorum*, *eli* genes were found to be clustered in the genome. In *P. sojae*, one contig of 115 kb containing 15 *eli* genes and one small contig with three *eli* genes could be identified (figs. 2 and 5A). In *P. ramorum*, a region of 59 kb with 14 *eli* genes was found (figs. 2 and 5A). The other *soj* and *ram* genes were found on other small scaffolds but it cannot be excluded that some genes are on different scaffolds because of gaps in the draft genome sequence.

Not only *eli* genes, but also *ell* genes were found to be clustered in the genomes of *P. sojae* and *P. ramorum*. For example, *sol4a*, *sol4b*, *sol9*, and *sol10* are located within a 123-kb sequence contig in *P. sojae*, while *ral4*, *ral10a*, *ral10b*, and *ral9* map within a 97-kb sequence contig in *P. ramorum* (figs. 2 and 5B). Another example is the clustering of genes encoding ELLs belonging to clade ELL-13. In *P. sojae* and *P. ramorum*, five *sol13* and six *ral13* genes, respectively, were found to be clustered in a 31-kb sequence contig in both species (figs. 2 and 5C). Other examples of *ell* gene clustering are shown in figure 2. In *P. sojae* and *P. ramorum*, often *ell*'s belonging to the same clade are clustered suggesting that these *ell* genes are paralogues that resulted from gene duplication.

The *eli* and *ell* Genes in the Same Clade Show Similar Expression Patterns

Members of gene families are often differentially expressed in space and time. To see whether that is also true for the different members of the complex elicitin gene family, we examined tissue-specific expression patterns as well as expression levels of different family members by transcript counting based on EST databases and by northern blot hybridization.

For *P. infestans* a large collection of over 75,000 ESTs is available (Randall et al. 2005), and for *P. sojae* a collection of 21,282 ESTs is available (B. M. Tyler, unpublished data; http://staff.vbi.vt.edu/estap). Transcript counting showed that elicitin genes are differentially expressed in zoospores, sporangia, and mycelia of *P. infestans* and *P. sojae* (table 3). These differential expression patterns are specific for a particular phylogenic clade of *eli* or *ell* genes. For example, ESTs from the *inf1* and *soj1* genes in the ELI-1 clade are strongly represented in mycelium libraries, as well as libraries from material that include substantial amounts of mycelia, such as mating cultures (*P. infestans*) and infected plant tissue (*P. sojae*). In contrast, ESTs from genes in the ELL-3 clade are primarily found in zoospore libraries of both *P. infestans* and *P. sojae*. A northern blot with RNA isolated from mycelia, sporangia, zoospores, cysts, and germinating cysts from *P. infestans* and hybridized with *inf1*, *inf2A*, *inf2B*, *inf4*, *inf5*, *inf6*, and *inl1* confirmed the findings of the transcript counting in *P. infestans* (fig. 6A). The differential expression of ELI-1 and ELL-3 genes was also observed experimentally in a third species, *Phytophthora parasitica*. Northern blot analysis showed that the ELI-1 *para1* is only expressed in mycelium, whereas *parl3*, which belongs to the ELL-3 clade, is expressed in zoospores and germinating cysts (fig. 6B). In *P. sojae*, several *eli* and *ell* genes are strongly expressed during infection of soybean, compared to their expression in mycelia, such as *soj1c*, *soj5*, and *soj6*, and especially *sol2a* and *sol2b*.

The expression levels of *eli* and *ell* genes also differ widely. For some, like *inl6*, only one transcript was found in the entire EST database, while for others, like for example, *inf6*, more than 100 transcripts were present. In *P. sojae*, many *soj* and *sol* genes predicted in the genome sequence have no ESTs at all. In *P. infestans*, *inf1*, *inf5*, and *inf6* are among the most abundantly expressed genes in
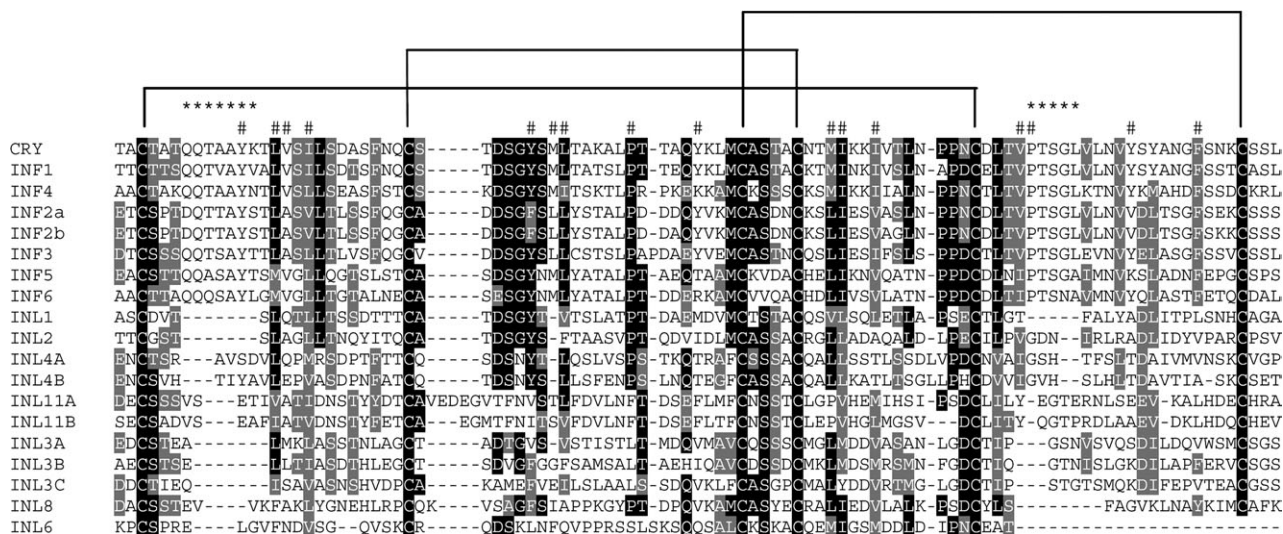
Table 2
The Spacing Pattern of the Six Cysteine Residues Present in the Elicitin Domain in ELIs and ELLs Belonging to Different Clades

Name[a]	Domain Size (amino acid)	Clades	Cysteine Spacing Pattern
BRA1, INF1, RAM1A, RAM1B, RAM1C, RAM1D, RAM1E, SOJ1A, SOJ1B, SOJ1C, SOJ1D, SOJ1E, SOJ1F	98	ELI-1	C-23-C-23-C-4-C-14-C-23-C CxxxxxxxxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxxxxxxxxxC
BRA2, INF2A, INF2B, RAM2A, RAM2B, RAM2C, RAM2D, RAM2E, SOJ2A, SOJ2B, SOJ2C, SOJ2D	98	ELI-2	C-23-C-23-C-4-C-14-C-23-C CxxxxxxxxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxxxxxxxxxC
INF3, RAM3A, RAM3B, RAM3C, RAM3D, SOJ3A, SOJ3B, SOJ3C, SOJ3D	98	ELI-3	C-23-C-23-C-4-C-14-C-23-C CxxxxxxxxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxxxxxxxxxC
BRA5, INF5, RAM5, SOJ5, BRA6, INF6, RAM6, SOJ6A, SOJ6B	98	ELI-4	C-23-C-23-C-4-C-14-C-23-C CxxxxxxxxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxxxxxxxxxC
BRL1A, BRL1B, INL1, RAL1A, RAL1B, SOL1A, SOL1B, SOL1C, SOL1D	85	ELL-1	C-16-C-22-C-4-C-14-C-18-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxx………xxxxC
INL2, RAL2A, RAL2B, RAL2C, RAL2D, RAL2E, SOL2A, SOL2B, SOL2C, SOL2D, SOL2E	88	ELL-2	C-16-C-22-C-4-C-14-C-21-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxx…….xxxxC
BRL3, INL3A, INL3B, INL3C, RAL3, SOL3A, SOL3B	87, 88	ELL-3	C-16-C-22/23-C-4-C-14-C-20-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxx…….xxxxC
INL4A, INL4B, RAL4, SOL4A, SOL4B	92	ELL-4	C-20-C-22-C-4-C-15-C-20-C Cxxxxx…xxxxxCxxxxxxxxxxxCxxCxxxxxxxxCxxxxx…….xxxxC
BRL5, RAL5, SOL5,	89	ELL-5	C-16-C-23-C-4-C-14-C-21-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxx…….xxxxC
INL6, RAL6, SOL6	91	ELL-6	C-17-C-24-C-4-C-14-C-21-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxx…….xxxxC
RAL7A, RAL7B, SOL7	92	ELL-7	C-19-C-23-C-4-C-14-C-21-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxx…….xxxxC
INL8, RAL8A, RAL8B, SOL8	87, 93	ELL-8	C-19-C-24-C-4-C-14-C-15/21-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxx………xxxxC
RAL9, SOL9	91	ELL-9	C-18-C-23-C-4-C-14-C-21-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxx…….xxxxC
RAL10A, RAL10B, SOL10	91	ELL-10	C-18-C-23-C-4-C-14-C-21-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxx…….xxxxC
BRL11, INL11A, INL11B, RAL11A, RAL11B, RAL11C, RAL11D, SOL11A, SOL11B, SOL11C, SOL11D, SOL11E	93, 98	ELL-11	C-20-C-25/28-C-4-C-12/14-C-21-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxx…….xxxxC
RAL12, SOL12	92	ELL-12	C-20-C-22-C-4-C-14-C-21-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxxxx….xxxxC
INL13, RAL13A, RAL13A2, RAL13B, RAL13C, RAL13D, RAL13E, RAL13F, RAL13J, SOL13A, SOL13A2, SOL13B, SOL13C, SOL13D, SOL13E, SOL13F, SOL13G, SOL13H, SOL13H2, SOL13I, SOL13J	75, 78, 80, 81, 82	ELL-13	C-20-C-(12-17)-C-4-C-13-C-17/18-C Cxxx……xxxxxCxxxxxxxxxxxCxxCxxxxxxxCxxx………xxxxC

[a] The shaded ELIs and ELLs represent ESTs.

FIG. 4.—Multiple sequence alignment of *Phytophthora infestans* ELIs and ELLs with the ELI-1 elicitin CRY from *Phytophthora cryptogea*. The sequence alignment was generated from the conserved elicitin domains. From INL6 only an incomplete gene sequence was present in the EST database. The three lines shown above the alignment connect the cysteine residues that form disulfide bonds in CRY. * indicates the residues in CRY that correspond to the gaps in the alignment. # indicates the residues that interact with ergosterol (Boissy et al. 1999).

mycelium (Kamoun et al. 1999). Overall, the expression levels of *eli* genes seem to be higher than those of *ell* genes with the exception of *sol3a* which is one of the most strongly expressed genes in *P. sojae* zoospores.

The difference in expression levels between *eli* and *ell* genes can also be related to the difference in the GC3 (third position codon) content in the coding regions. In a genome wide survey of *P. sojae* we found that highly expressed genes usually have a higher GC3 than lowly expressed genes (R. H. Y. Jiang and F. Govers, in preparation). In *P. brassicae*, *P. infestans*, *P. ramorum*, and *P. sojae*, the average GC3 of 46 *eli*'s is 86%, which is higher than the 75% of 82 *ell*'s. The higher GC3 percentage of *eli* genes agrees with the higher expression levels of *eli* genes as compared to *ell* genes.

Discussion
Elicitins Are Unique and Ubiquitous in *Phytophthora* and *Pythium*

The elicitin genes are present in all examined *Phytophthora* species. The encoded proteins are ubiquitous across the whole *Phytophthora* genus with ELI-1 elicitins as the most abundant component in *Phytophthora* culture filtrates. However, the elicitins seem to be limited to the oomycetes *Phytophthora* and *Pythium*. Searches in GenBank and Pfam protein domain database (Sonnhammer et al. 1998) did not reveal any other organism that has elicitin-like sequences. In this study, a thorough search was performed on EST databases from a few plant pathogenic ascomycete fungi and on the genome sequence of the rice blast fungus, and a marine diatom, but no elicitin-like sequences were found.

In plants there is a group of proteins called nsLTPs (nonspecific lipid transfer proteins) that in some aspects resemble the elicitins (Blein et al. 2002). Similar to elicitins nsLTPs are small secreted cysteine-rich proteins that interact with lipids. They facilitate the transfer of lipids between natural or artificial membranes (Kader 1996) and have been implicated to play roles in protection and defense (Buhot et al. 2001). However, the cysteine spacing patterns of these two classes of proteins are very different, and therefore, it is unlikely that elicitins and nsLTPs share a close phylogenetic relationship. Secreted proteins with a nsLTP-like cysteine spacing pattern cannot be found in the *P. infestans* EST database or the unigene sets of *P. sojae* and *P. ramorum*.

The Elicitin Gene Family Is an Ancient Family Within the *Phytophthora* Genus

In this study we focused on three *Phytophthora* species that differ from each other in various traits, such as host range (narrow or broad), sexual behavior (homo- or heterothallic), and genome size (ranging from 65 to 240 Mb). The internal transcribed spacer (rRNA)-based phylogenetic tree (Cooke et al. 2000) shows that *P. infestans* and *P. sojae* are located on widely divergent branches, indicating early species diversification during the evolution of the genus. In a phylogenic tree based on β-tubulin and EF-1α sequences *P. brassicae*, *P. infestans*, *P. sojae*, and *P. ramorum* also fall in four different clades (Kroon et al. 2004).

The examined ELIs and ELLs were derived from the whole genome sequence of *P. sojae* and *P. ramorum* and from a large *P. infestans* EST collection and a smaller set of *P. brassicae* EST's. Thus, the phylogenetic tree should cover all members present in *P. sojae* and *P. ramorum* and the majority of the ELIs and ELLs *in P. infestans*. From *P. brassicae* only a limited number of ELIs and ELLs were available for inclusion in the tree.

From the relationships represented by the phylogenic tree, it is evident that most clades possess family members from the three different species which are well-represented in the data set, indicating that the genes founding these
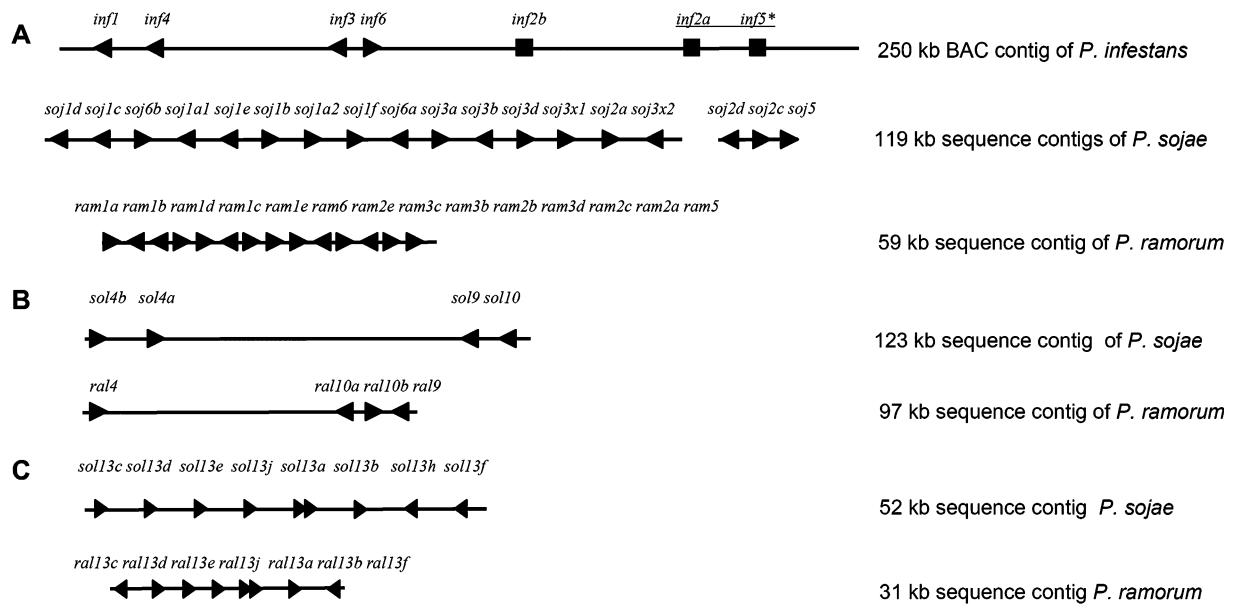
Fig. 5.—Clustering of elicitin genes in *Phytophthora ramorum*, *Phytophthora infestans*, and *Phytophthora sojae*. Black arrows indicate *eli* and *ell* genes and their orientations. Black squares indicate genes with unidentified orientation. Thin horizontal lines represent DNA contigs. * the order of *inf2A* and *inf5* has not been determined. (*A*) *eli* contigs. (*B*) *ell* contigs containing genes encoding ELL-4, ELL-9, and ELL-10 proteins. (*C*) *ell13* contigs. *sol13a* and *ral13a* have repeated elicitin domains.

clades had already diverged before the common ancestor gave rise to *P. infestans*, *P. sojae*, and *P. ramorum*. Because in most cases, all clades of ELIs and ELLs are preserved in the three species, we infer that the different clades of these proteins have distinct functions. This hypothesis is supported by the conserved tissue-specific expression of members of some clades, most notably the zoospore-specific expression of ELL-3 genes in three different species. Within individual clades, however, there has been extensive divergence of the genes in each species, and there is no evidence for diversifying selection.

In *P. infestans*, *P. cinnamomi*, and *Phytophthora cryptogea* the genes of the ELI-1 clade were reported to form a gene cluster within the genome (Panabieres et al. 1995; Duclos et al. 1998; Jiang et al. 2005). In this study we showed that also *eli* and *ell* genes belonging to other clades occur in clusters. Within the *eli* clusters few other genes are present (Jiang et al. 2005 and data not shown), and therefore the *eli* gene density can be compared. Interestingly, the gene density is different in the different species. In *P. infestans* the average density in this region is one *eli* per 22.5 kb. In *P. sojae*, the average gene density in a contig containing 15 *eli* genes is one *eli* per 7.7 kb, whereas in *P. ramorum* the gene density is even higher: one *eli* per 4.2 kb in a contig containing 14 *eli* genes. The species with the smallest genome size, that is, *P. ramorum* (65 Mb), shows a higher *eli* density, whereas in the much larger *P. infestans* genome (240 Mb) mobile elements fill the spaces in between *eli* genes (Jiang et al. 2005).

Assuming that the ELIs and ELLs are all members of a single gene family that arose from a series of duplication events, then the primary gene duplication should be an event that occurred in the common ancestor. Subsequent duplications of individual family members then occurred after speciation, for example, within the ELI-1 clade for *P. sojae* and

P. ramorum. In *P. infestans*, the clustered *eli* genes show similar expression patterns with a high level of expression in mycelium, thus it is conceivable that the clustering is important for simultaneous expression of these *eli* genes. The chromatin environment can be crucial for the establishment and maintenance of transcriptional activation or repression via histone and DNA modifications (Lusser 2002). A detailed analysis of the promoter regions and putative regulatory elements may reveal why several members of this ancient gene family are clustered in *Phytophthora* genomes.

The Intrinsic Function and Subcellular Location of ELIs and ELLs

The ELI-1 clade elicitins are well known for their HR-inducing activity on *Nicotiana* species. Silencing of the *inf1* gene in *P. infestans* resulted in strains that were able to colonize the nonhost *N. benthamiana*, and in this interaction INF1 protein apparently acts as a determinant of nonhost resistance (Kamoun et al. 1998). Also, members of the ELI-2 clade have HR-inducing activity as demonstrated for INF2A and INF2B (Huitema et al. 2005), and SOJ3 and SOJ6 (Qutob et al. 2003). From the ELL clades only a few members have been tested for HR-inducing activity (Qutob et al. 2003; J. S. Marshall, A. R. Hardham, and F. Govers, unpublished data) but so far no such activity was found. Similar to ELIs, ELLs have a structurally conserved six cysteine elicitin domain. However, the intervening amino acid sequences in ELLs exhibited more diversity than those in ELIs. The multiple sequence alignment of ELIs and ELLs from *P. infestans* showed that two regions are completely or partly deleted from ELLs as compared to ELIs (fig. 4) and this also results in the distinct cysteine spacing patterns of the various ELI and ELL clades as shown in table 2. In the 3D structure of cryptogein, an

Table 3
Expression Patterns of *eli* and *ell* Genes Represented by the Number of Transcripts Present in *Phytophthora infestans* and *Phytophthora sojae* EST Databases from Different Life Stages

Sequence[a]	Transcripts Per 10,000 in cDNA Libraries[b] from				
	Zoospores[c]	Sporangia	Mating	Mycelia	Infection
inf1	0.8	8.3	65.6	61.8	NA[d]
soj1a1; *1a2*[e]	0	NA	NA	7.8	18.2
soj1b	0	NA	NA	15.6	2.8
soj1c	0	NA	NA	6.7	11.2
inf2a	1.7	3.5	17.4	11.3	NA
inf2b	1.7	3.5	18	11.6	NA
soj2a; *2b*; *2d*[e]	0	NA	NA	3.4	1.4
soj2c1; *2c2*[e]	0	NA	NA	12.3	11.2
inf3	0	0	1.9	0.3	NA
soj3a	0	NA	NA	5.6	1.4
soj3c	0	NA	NA	3.4	1.4
inf4	0.8	0	1.9	2.7	NA
inf5	0	3.5	36	27.9	NA
soj5	0	NA	NA	3.4	12.6
inf6	1.7	2.4	96.5	58.8	NA
soj6a	1.9	NA	NA	7.8	14.0
inl1	0	4.7	2.6	4.8	NA
sol1a	1.9	NA	NA	1.1	1.4
inl2	0.8	1.2	0	0.9	NA
sol2a	0	NA	NA	0	11.2
sol2b	0	NA	NA	0	12.6
inl3a	1.7	0	0	0	NA
inl3b	1.7	1.2	0	0	NA
inl3c	6.7	1.2	0	0	NA
sol3a	34.7	NA	NA	0	2.8
sol3b	1.9	NA	NA	0	0
inl4a	0	0	0.6	0	NA
inl4b	0	1.2	0	0.6	NA
sol5	0	NA	NA	1.1	0
inl6	0	1.2	0	0	NA
sol6	1.9	NA	NA	0	0
inl8	0	0	0.6	0	NA
sol8	0	NA	NA	0	1.4
sol9	0	NA	NA	0	1.4
sol10	0	NA	NA	0	1.4
inl11a	0	0	0	0.3	NA
inl11b	1.7	0	0.6	2.1	NA
sol11a	0	NA	NA	2.2	0
sol11b	0	NA	NA	0	2.8
inl13	0	4.7	0	0.9	NA
sol13e	0	NA	NA	2.2	0

[a] An individual *eli* or *ell* sequence was searched against the databases using BlastN with E value cutoff of 1×10^{-100}. The following *P. sojae* genes, identified from the genome sequence, had no matches to ESTs in the database: *soj1d-f*; *soj3b*; *soj3d*; *soj3x*; *soj6b*; *sol1b-e*; *sol2c-e*; *sol4a-b*; *sol7*; *sol11c-f*; *sol12*; *sol13a-d*; *sol13f-j*.

[b] The *P. infestans* EST database contained 11,919 sequences from zoospores and germinated cysts, 8,470 from sporangia, 5,541 from mating cultures, and 33,651 from mycelia grown under various conditions. The infection library available for *P. infestans* was too small for a reliable counting. The *P. sojae* EST database contained 5,187 sequences from zoospores and germinated cysts, 8,949 from mycelia grown under various conditions, and 7,146 *P. sojae* sequences from infected soybean tissue (Qutob et al. 2000). No libraries from sporangia or mating cultures were available from *P. sojae*.

[c] Zoospore libraries from both species included cDNAs from both free-swimming zoospores and germinated cysts.

[d] NA = Not available.

[e] Genes *soj1a1* and *soj1a2* encode identical mRNA sequences, and so their ESTs were counted together. The same was true for the three genes *soj2a*, *soj2b*, and *soj2d*, as well as the two genes *soj2c1* and *soj2c2*.

ELI-1 from *P. cryptogea* (Mikes et al. 1997), residues of these variable regions (QQTAAY and PTSGL) are located at the surface. This suggests that the surface-exposed residues in ELIs may be variable or deleted due to selection pressure, possibly imposed by host plants.

Because the HR in plants blocks the growth of biotrophic pathogens, the HR-inducing activity is certainly not the primary function of ELIs in *Phytophthora*. Based on biochemical data it is now generally accepted that the intrinsic biological function of ELIs is related to lipid binding
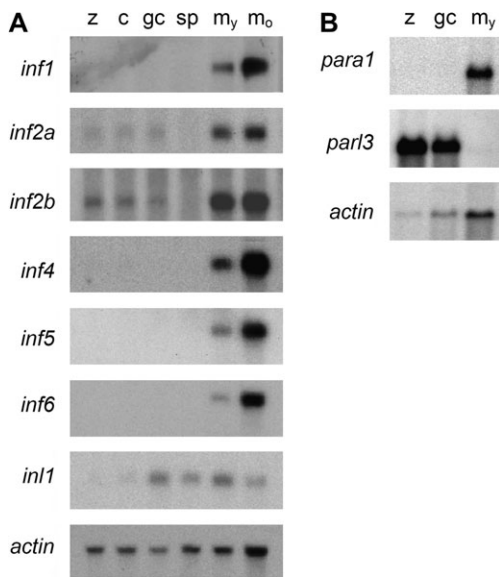
Fig. 6.—(*A*) Autoradiographs of a northern blot containing RNA isolated from *Phytophthora infestans* zoospores (z), cysts (c), cysts germinated in water for 2.5 h (gc), sporangia (s), a young mycelial culture started from sporangia that were allowed to germinate for 20 h at 18°C (m$_y$) and old mycelium (m$_o$) and hybridized with various *P. infestans inf* and *inl* probes and an actin probe as loading control. (*B*) Autoradiographs of a northern blot containing RNA isolated from *Phytophthora parasitica* zoospores (z), germinated cysts (gc) and young mycelium (m$_y$) and hybridized with *P. parasitica para1* and *parl3* probes, and an actin probe as loading control.

Fig. 7.—Schematic representation of ELIs and ELLs. (*A*) ELIs secreted in the culture filtrate. (*B*) ELIs and ELLs that are hypothesized to be linked to the cell wall by extensive glycosylation of the C-terminal domain. (*C*) ELLs hypothesized to be anchored to the cell membrane by GPI. (*D*) ELLs hypothesized to be anchored to the cell membrane of wall-less zoospores by GPI.

and/or processing. Cryptogein, for example, binds sterols, such as ergosterol, and functions as a sterol-carrier protein (Mikes et al. 1997; Vauthrin et al. 1999), whereas ELI-4 members in *P. capsici* were shown to have phospholipase activity (Nespoulous et al. 1999). *Phytophthora* species cannot synthesize sterols but still do require sterols for several physiological functions. Expression analysis showed that the ELI-1 members are highly expressed in mycelium, perhaps correlated with the requirement to acquire sterols from the environment during vegetative growth. Because other life cycle stages such as sporangia, zoospores, and cysts are short lived and primarily involved in pathogen dispersal and host invasion, the requirement for sterols may be less. The intrinsic functions of ELLs are unknown. The majority of the residues in cryptogein that are known to be involved in sterol binding are divergent in ELLs (fig. 4) thus bringing into question the sterol binding capacity of ELLs. One of the ELL-4 members, however, was postulated to function in relation to the sterol-like mating hormones in *Phytophthora*. INL4A is identical to the previously described mating associated factor M-25 and is encoded by a gene that is specifically expressed and highly induced in mating cultures during sexual development (Fabritius, Cvitanich, and Judelson 2002).

All ELIs and ELLs possess a signal peptide, and the majority is likely to be associated with the cell wall or anchored to the cell membrane. Figure 7 gives a schematic representation of the various ELIs and ELLs based on the presence or absence of a C-terminal domain, on typical features in the C-terminal domain and on stage-specific ex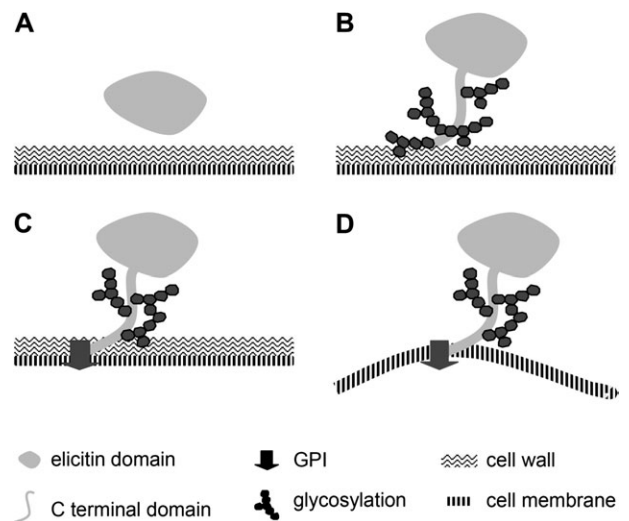pression patterns. The ELI-1 elicitins are extracellular proteins consisting solely of the elicitin domain, with occasionally a short tail, and are abundantly secreted during mycelial growth (fig. 7*A*). The overrepresentation of threonine and serine residues in the C-terminal domains of several ELIs end ELLs suggests extensive O-glycosylation and linkage to the cell wall (fig. 7*B*). In fact, INF2A was demonstrated experimentally to be a cell wall-associated protein (V. G. A. A. Vleeshouwers and F. Govers, in preparation). In plants, proline-rich and hydroxyproline-rich glycoproteins are amongst the most extensively characterized cell wall components (Cassab 1998). The C-terminal domains of ELI-4 and ELL-8 members are very proline rich suggesting that also these ELLs are associated with the cell wall. The occurrence of GPI sites in several ELL clades suggests that anchoring to the cell membrane is a common way of ELLs to be tethered to the exterior of the cell (fig. 7*C*). Also ELL-3 is predicted to be GPI anchored. Because of the specific expression of the ELI-3 genes in the zoospore stage in which a cell wall is lacking, the mature ELL-3 proteins could be tethered to the cell membrane of the mobile zoospores by the GPI anchor, and the putative O-linked glycosylation may be used to coat the zoospore surface with oligosaccharides (fig. 7*D*).

Supplementary Material

Tables S1 and S2 and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

Armbrust, E. V., J. A. Berges, C. Bowler et al. (45 co-authors). 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science **306**:79–86.

Armstrong, M. R., S. C. Whisson, L. Pritchard et al. (22 co-authors). 2005. An ancestral oomycete locus contains late blight avirulence gene *Avr3a*, encoding a protein that is recognized in the host cytoplasm. Proc. Natl. Acad. Sci. USA **102**:7766–7771.

Bairoch, A. 1991. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **19**(Suppl.):2241–2245.

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

Blein, J. P., P. Coutos-Thevenot, D. Marion, and M. Ponchet. 2002. From elicitins to lipid-transfer proteins: a new insight in cell signalling involved in plant defence mechanisms. Trends Plant Sci. **7**:293–296.

Boissy, G., E. deLaFortelle, R. Kahn, J. C. Huet, G. Bricogne, J. C. Pernollet, and S. Brunie. 1996. Crystal structure of a fungal elicitor secreted by *Phytophthora cryptogea*, a member of a novel class of plant necrotic proteins. Structure **4**:1429–1439.

Boissy, G., M. O'Donohue, O. Gaudemer, V. Perez, J. C. Pernollet, and S. Brunie. 1999. The 2.1 A structure of an elicitin-ergosterol complex: a recent addition to the sterol carrier protein family. Protein Sci. **8**:1191–1199.

Buhot, N., J. P. Douliez, A. Jacquemard et al. (11 co-authors). 2001. A lipid transfer protein binds to a receptor involved in the control of plant defence responses. FEBS Lett. **509**:27–30.

Cassab, G. I. 1998. Plant cell wall proteins. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49**:281–309.

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

Dean, R. A., N. J. Talbot, D. J. Ebbole et al. (35 co-authors). 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature **434**:980–986.

Duclos, J., A. Fauconnier, A. C. Coelho, A. Bollen, A. Cravador, and E. Godfroid. 1998. Identification of an elicitin gene cluster in *Phytophthora cinnamomi*. DNA Seq. **9**:231–237.

Eisenhaber, B., M. Wildpaner, C. J. Schultz, G. H. Borner, P. Dupree, and F. Eisenhaber. 2003. Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for *Arabidopsis* and rice. Plant Physiol. **133**:1691–1701.

Erwin, D. C., and O. K. Ribeiro. 1996. *Phytophthora* diseases worldwide. The American Phytopathological Society, St Paul, Minn.

Fabritius, A. L., C. Cvitanich, and H. S. Judelson. 2002. Stage-specific gene expression during sexual development in *Phytophthora infestans*. Mol. Microbiol. **45**:1057–1066.

Fefeu, S., S. Bouaziz, J. C. Huet, J. C. Pernollet, and E. Guittet. 1997. Three-dimensional solution structure of beta cryptogein, a beta elicitin secreted by a phytopathogenic fungus *Phytophthora cryptogea*. Protein Sci. **6**:2279–2284.

Gotesson, A., J. S. Marshall, D. A. Jones, and A. R. Hardham. 2002. Characterization and evolutionary analysis of a large polygalacturonase gene family in the oomycete plant pathogen *Phytophthora cinnamomi*. Mol. Plant-Microbe Interact. **15**:907–921.

Hirokawa, T., S. Boon-Chieng, and S. Mitaku. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics **14**:378–379.

Huitema, E., V. G. A. A. Vleeshouwers, C. Cakir, S. Kamoun, and F. Govers. 2005. Differences in intensity and specificity of hypersensitive response induction in *Nicotiana* spp. by INF1, INF2A, and INF2B of *Phytophthora infestans*. Mol. Plant. Microbe Interact. **18**:183–193.

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

Julenius, K., A. Molgaard, R. Gupta, and S. Brunak. 2005. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology **15**:153–164.

Kader, J. C. 1996. Lipid-transfer proteins in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47**:627–654.

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

Kamoun, S., H. Lindqvist, and F. Govers. 1997. A novel class of elicitin-like genes from *Phytophthora infestans*. Mol. Plant-Microbe Interact. **10**:1028–1030.

Kamoun, S., P. van West, A. J. de Jong, K. E. de Groot, V. Vleeshouwers, and F. Govers. 1997. A gene encoding a protein elicitor of *Phytophthora infestans* is down-regulated during infection of potato. Mol. Plant-Microbe Interact. **10**:13–20.

Kamoun, S., P. van West, V. Vleeshouwers, K. E. de Groot, and F. Govers. 1998. Resistance of *Nicotiana benthamiana* to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. Plant Cell **10**:1413–1425.

Krogh, A., B. Larsson, G. von Heijne, and E. L. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305**:567–580.

Kroon, L. P., F. T. Bakker, G. B. van den Bosch, P. J. Bonants, and W. G. Flier. 2004. Phylogenetic analysis of *Phytophthora* species based on mitochondrial and nuclear DNA sequences. Fungal Genet. Biol. **41**:766–782.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

Liu, Z., J. I. Bos, M. Armstrong et al. (11 co-authors). 2005. Patterns of diversifying selection in the phytotoxin-like *scr74* gene family of *Phytophthora infestans*. Mol. Biol. Evol. **22**:659–672.

Lusser, A. 2002. Acetylated, methylated, remodeled: chromatin states for gene regulation. Curr. Opin. Plant Biol. **5**:437–443.

Man in 't Veld, W. A., A. de Cock, E. Ilieva, and C. A. Levesque. 2002. Gene flow analysis of *Phytophthora porri* reveals a new species: *Phytophthora brassicae* sp nov. Eur. J. Plant Pathol. **108**:51–62.

Mao, Y., and B. M. Tyler. 1996. Cloning and sequence analysis of elicitin genes of *Phytophthora sojae*. Fungal Genet. Biol. **20**:169–172.

Margulis, L., and K. V. Schwarts. 2000. Five kingdoms: an illustrated guide to the phyla of life on earth. W.H. Freeman and Company, New York.

Mikes, V., M. L. Milat, M. Ponchet, P. Ricci, and J. P. Blein. 1997. The fungal elicitor cryptogein is a sterol carrier protein. FEBS Lett. **416**:190–192.

Nespoulous, C., O. Gaudemer, J. C. Huet, and J. C. Pernollet. 1999. Characterization of elicitin-like phospholipases isolated from *Phytophthora capsici* culture filtrate. FEBS Lett. **452**:400–406.

Panabieres, F., A. Marais, J. Y. LeBerre, I. Penot, D. Fournier, and P. Ricci. 1995. Characterization of a gene cluster of *Phytophthora cryptogea* which codes for elicitins, proteins inducing a hypersensitive- like response in tobacco Mol. Plant-Microbe Interact. **8**:996–1003.

Panabieres, F., M. Ponchet, V. Allasia, L. Cardin, and P. Ricci. 1997. Characterization of border species among *Pythiaceae*: several *Pythium* isolates produce elicitins, typical proteins from *Phytophthora* spp. Mycol. Res. **101**:1459–1468.

Pemberton, C. L., and G. P. C. Salmond. 2004. The Nep1-like proteins—a growing family of microbial elicitors of plant necrosis. Mol. Plant Pathol. **5**:353–359.

Pieterse, C. M., P. van West, H. M. Verbakel, P. W. Brasse, G. C. van den Berg-Velthuis, and F. Govers. 1994. Structure and genomic organization of the *ipiB* and *ipiO* gene clusters of *Phytophthora infestans*. Gene **138**:67–77.

Pond, S. L., and S. D. Frost. 2005. DataMonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics **21**:2531–2533.

Qutob, D., P. T. Hraber, B. W. S. Sobral, and M. Gijzen. 2000. Comparative analysis of expressed sequences in *Phytophthora sojae*. Plant Physiol. **123**:243–253.

Qutob, D., E. Huitema, M. Gijzen, and S. Kamoun. 2003. Variation in structure and activity among elicitins from *Phytophthora sojae*. Mol. Plant Pathol. **4**:119–124.

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

Randall, T. A., R. A. Dwyer, E. Huitema et al. (31 co-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.

Rizzo, D. M., M. Garbelotto, and E. M. Hansen. 2005. *Phytophthora ramorum*: integrative research and management of an emerging pathogen in California and Oregon forests. Annu. Rev. Phytopathol. **43**:309–335.

Roetschi, A., A. Si-Ammour, L. Belbahri, F. Mauch, and B. Mauch-Mani. 2001. Characterization of an *Arabidopsis-Phytophthora* pathosystem: resistance requires a functional PAD2 gene and is independent of salicylic acid, ethylene and jasmonic acid signalling. Plant J. **28**:293–305.

Shan, W., M. Cao, D. Leung, and B. M. Tyler. 2004. The *Avr1b* locus of *Phytophthora sojae* encodes an elicitor and a regulator required for avirulence on soybean plants carrying resistance gene *Rps1b*. Mol. Plant-Microbe Interact. **17**:394–403.

Sigrist, C. J., L. Cerutti, N. Hulo, A. Gattiker, L. Falquet, M. Pagni, A. Bairoch, and P. Bucher. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief. Bioinform. **3**:265–274.

Soanes, D. M., W. Skinner, J. Keon, J. Hargreaves, and N. J. Talbot. 2002. Genomics of phytopathogenic fungi and the development of bioinformatic resources. Mol. Plant-Microbe Interact. **15**:421–427.

Sonnhammer, E. L. L., S. R. Eddy, E. Birney, A. Bateman, and R. Durbin. 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res. **26**:320–322.

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

van't Slot, K. A. E., and W. Knogge. 2002. A dual role for microbial pathogen-derived effector proteins in plant disease and resistance. Crit. Rev. Plant Sci. **21**:229–271.

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

Vauthrin, S., V. Mikes, M. L. Milat, M. Ponchet, B. Maume, H. Osman, and J. P. Blein. 1999. Elicitins trap and transfer sterols from micelles, liposomes and plant plasma membranes. Biochim. Biophys. Acta **1419**:335–342.

Werres, S., R. Marwitz, W. Veld, A. De Cock, P. J. M. Bonants, M. De Weerdt, K. Themann, E. Ilieva, and R. P. Baayen. 2001. *Phytophthora ramorum* sp nov., a new pathogen on *Rhododendron* and *Viburnum*. Mycol. Res. **105**:1155–1165.

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

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**:555–556.

Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431–449.

Yang, Z., W. S. Wong, and R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. **22**:1107–1118.