

# Structure of the Glucanase Inhibitor Protein (GIP) Family from *Phytophthora* Species Suggests Coevolution with Plant Endo- $\beta$ -1,3-Glucanases

Cynthia M. B. Damasceno,<sup>1</sup> John G. Bishop,<sup>2</sup> Daniel R. Ripoll,<sup>3</sup> Joe Win,<sup>4</sup> Sophien Kamoun,<sup>4</sup> and Jocelyn K. C. Rose<sup>1</sup>

<sup>1</sup>Department of Plant Biology, Cornell University, Ithaca, NY 14853, U.S.A.; <sup>2</sup>School of Biological Sciences, Washington State University, Vancouver, WA 98686, U.S.A.; <sup>3</sup>Computational Biology Service Unit, Life Sciences Core Laboratories Center, Cornell University, Ithaca, NY 14853, U.S.A.; <sup>4</sup>Sainsbury Laboratory, Colney Lane, Norwich, NR4 7UH, U.K.

Submitted 4 September 2007. Accepted 8 February 2007.

During invasion of their plant hosts, species of the oomycete genus *Phytophthora* secrete glucanase inhibitor proteins (GIPs) into the plant apoplast, which bind and inhibit the activity of plant extracellular endo- $\beta$ -1,3-glucanases (EGases). GIPs show structural homology to the chymotrypsin class of serine proteases (SP) but lack proteolytic activity due to the absence of an intact catalytic triad and, thus, belong to a broader class of proteins called serine protease homologs (SPH). To study the evolutionary relationship between GIPs and functional SP, database searches were used to identify 48 GIP homologs in the *P. sojae*, *P. ramorum*, and *P. infestans* genomes, composing GIPs, SPH, and potentially functional SP. Analyses of *P. infestans*-inoculated tomato leaves showed that *P. infestans* GIPs and tomato EGases are present in the apoplast and form stable complexes in planta. Studies of the temporal expression of a four-membered GIP family from *P. infestans* (*PiGIP1* to *PiGIP4*) further revealed that the genes show distinctly different patterns during an infection timecourse. Codon evolution analyses of GIP homologs identified several positively selected peptide sites and structural modeling revealed them to be in close proximity to rapidly evolving EGase residues, suggesting that the interaction between GIPs and EGases has the hallmarks of a coevolving molecular arms race.

*Additional keywords:* glycosyl hydrolase family 17.

A notable characteristic of the interaction between plants and their microbial pathogens is the secretion by both partners of suites of proteins that are associated with attack, defense, and counterdefense (Kamoun 2006). A well-studied example is the secretion by the plant host of endo- $\beta$ -1,3-glucanases (EGases) of glycosyl hydrolase family 17 (CAZy website),

Corresponding author: Jocelyn K. C. Rose; 228 Plant Science Building, Cornell University; Telephone: +1.607.255.4781; Fax: +1.607.255.5407; E-mail: jr286@cornell.edu

Nucleotide sequence data is available in the GenBank database under the following accession numbers: *PiSPI.1*, EU443391; *PiGIP1*, EU443395; *PiGIP2*, EU443392; *PiGIP3*, EU443393; and *PiGIP4*, EU443394

\*The e-Xtra logo stands for “electronic extra” and indicates that one supplemental figure, three supplemental tables, and a video showing animation of the predicted EGaseA: PsGIP1 docking model with positively selected residues highlighted are published online.

which can digest the  $\beta$ -1,3 glucan polymers that are common constituents of fungal and oomycetous cell walls (Kauffmann et al. 1987; Stintzi et al. 1993). Such an activity represents a potentially important defense mechanism through both direct degradation of the pathogen mycelial wall and indirectly via the release of  $\beta$ -1,3 glucan oligosaccharide elicitors (van Loon et al. 2006).

Species of the oomycete genus *Phytophthora* employ a matching counterdefense system by secreting glucanase inhibitor proteins (GIPs) that specifically bind and inhibit the activity of plant EGases (Ham et al. 1997). The first cloning and sequencing of GIPs (Rose et al., 2002) from the soybean pathogen *P. sojae* revealed homology to serine proteases (SP) and, specifically, to the chymotrypsin (S1) subclass. However, GIPs can more accurately be classified as serine protease homologs (SPH), since they lack an intact His-Asp-Ser catalytic triad and, thus, do not have proteolytic activity (Rose et al. 2002; York et al. 2004). SP represent one of the best-studied enzyme families, and extensive computational modeling and structural and mutational analyses have provided detailed insights into the specific architectural features of the enzymes that influence substrate recognition and catalytic efficiency (Czapinska and Otlewski 1999; Gayathri et al. 2006; Krem et al. 2002; Radisky et al. 2006). SP exhibit a rapid rate of evolution (de Haën et al. 1975; Krem et al. 2000) and represent versatile protein modules that contribute to a broad spectrum of physiological processes (Krem et al. 2000). It has been suggested that their proteolytic and protein-binding properties evolve in a modular fashion and that the catalytic triad is an independently evolving motif that is associated with different binding sites and diverse functions (Iengar and Ramakrishnan 1999). This observation is supported by the identification of SPH from several taxa, including mammals (Højrup et al. 1985; Isackson and Bradshaw 1984; Kurosky et al. 1980), insects (Dimopoulos et al. 1997; Murugasu-Oei et al. 1995; Ross et al. 2003; Zou et al. 2006), crustaceans (Huang et al. 2000; Lin et al. 2006), and plants (Nakamura et al. 1989), which are associated with a broad range of biological processes.

Studies of GIPs to date have been limited to three genes from *P. sojae* (Ham et al. 1997; Rose et al. 2002). However, the recent release of the genome sequences of *P. sojae*, *P. ramorum*, and *P. infestans* (Tyler et al. 2006); C. Nusbaum, *personal communication*), has provided an opportunity to examine both the diversity and evolution of the GIP gene family and related SPH. In this current study, we present a survey of *Phytophthora* GIPs and homologous sequences as well as a

more detailed analysis of the four-member *GIP* family (*PiGIP1* to *PiGIP4*) from *P. infestans*.

The coexpression of families of secreted defense-related enzymes (e.g., EGases) and their corresponding inhibitor proteins (e.g., GIPs) in the plant apoplast raises the possibility that these protein families are continuously undergoing antagonistic coevolution in the form of a molecular arms race (Dawkins and Krebs 1979). Coevolving sets of genes encoding secreted proteins have previously been associated with plant-pathogen interactions, including plant resistance genes and pathogen avirulence genes as well as other examples of genes encoding cell wall-degrading enzymes and their respective inhibitor proteins (Chisholm et al. 2006; Stahl and Bishop 2000). Those genes experiencing frequent selective sweeps will exhibit greater non-synonymous substitution rates in functionally important sites, potentially including those that contribute to physical interactions between coevolving protein pairs. Examples of positive selection driving the evolution of an enzyme or its inhibitor in a plant-pathogen interaction have been provided by studies of endopolygalacturonases from fungal and oomycetous pathogens and cognate plant-derived polygalacturonase inhibitor proteins (Bishop 2005; Götesson et al. 2002; Stotz et al. 2000) as well as plant chitinases (Bishop et al. 2000; Tiffin and Moeller 2006).

We previously used phylogeny-based codon evolution analysis to obtain strong evidence for positive selection on EGases from the genus *Glycine* (EGaseA and EGaseB) as well as from a broad range of dicotyledons and used molecular modeling to predict sites on the surfaces of EGases and GIPs that may be involved in docking and high affinity binding (Bishop et al. 2005). Most of the positively selected residues and other rapidly evolving sites in EGaseA were located around the margins of the active site (the “ring of fire”) in close proximity to the bound *P. sojae* GIP (PsGIP1), which suggests the existence of a molecular arms race between GIPs and EGases (Bishop et al. 2005). In the present study, we extended this analysis by looking for positive selection on GIPs.

## RESULTS

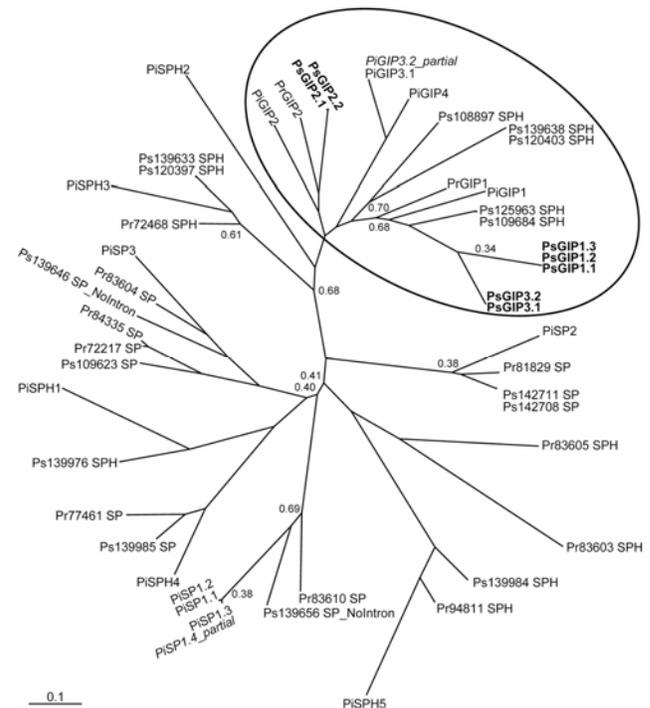
### Analysis of putative SP and SPH in *Phytophthora* species.

In order to examine the diversity of *GIP* gene families among *Phytophthora* species, a search of the *P. sojae*, *P. ramorum*, and *P. infestans* genome sequences was performed for homologs of the three previously described *P. sojae* GIPs (*PsGIP1*, *PsGIP2*, and *PsGIP3*; Rose et al. 2002). This analysis also provided the 5' sequence of *PsGIP3*, which was previously reported in a truncated form (Rose et al. 2002). A total of 48 homologous sequences (nontruncated versions) were identified, including the known *P. sojae* GIPs, and were categorized as either SP or SPH, according to the presence or absence, respectively, of an intact His-Asp-Ser catalytic triad, a feature essential for SP function (Perona and Craik 1995). Phylogenetic analysis identified a well-supported cluster of 18 *Phytophthora* GIPs (Fig. 1, circled) comprising the three known genes from *P. sojae* together with two homologs from *P. ramorum* (*PrGIP1* and *PrGIP2*), four from *P. infestans* (*PiGIP1*, *PiGIP2*, *PiGIP3*, and *PiGIP4*), and three additional *P. sojae* GIP-like sequences (*Ps120403\_SPH*, *Ps108897\_SPH*, and *Ps109684\_SPH*), along with identical or nearly identical copies of five of the six *P. sojae* genes. In addition, *PsGIP2* has a truncated version in the genome (*PsGIP2.3\_trunc*; not shown in phylogeny) and *PiGIP3* has a partial sequence in the genome (*PiGIP3.2\_partial*), due to an incomplete contig. The predicted GIP proteins translated from the genes in this clade showed 48 to 82% amino-acid identity. Thirteen other SPH and 17 SP with similarity to GIPs were also found. One group of five SPH was closely related to the group of 18 GIPs, and

the others were interspersed among diverse SP clades. While statistical support was strong for most groupings in the phylogeny, the basal groupings of the interspersed SP and SPH clades were poorly resolved (Fig. 1). The 48 translated proteins are predicted to have a N-terminal signal peptide (SignalP 3.0 software; Bendtsen et al. 2004), suggesting that they are all secreted.

The genus *Phytophthora* is one of only a few eukaryotic genera for which at least three complete genome sequences are available, thereby allowing an unusually broad view of gene-family evolution. The phylogeny provides eight examples of genes represented by at least one ortholog in each of the three species and two other cases in which an ortholog appears to have been duplicated, lost, or both in one of the three species (Supplementary Table 1). Examples include both GIPs (for example, *GIP1* has a full set of orthologs and paralogs while *GIP2* has orthologs but no paralogs) and apparently functional SP, such as *PiSP1*, which has a full set of orthologs. We estimate that there were at least five GIP-related SP and at least five GIP-related SPH prior to the diversification of *Phytophthora* spp.

The gene organization of GIPs, SPH, and SP was assessed by searching their respective locations in each of the genome sequences (Supplementary Fig. 1). For all three species, the GIP genes and some of the GIP-like genes were clustered in a scaffold, or supercontig, while the remaining genes were distributed among other supercontigs. All the *PiGIPs* are located in supercontig 29. *PiSP1*, a potential functional SP, has two gene copies flanking the 395-kb region (*PiSP1.1* and *PiSP1.2*), comprising four *PiGIPs*, *PiSPH2*, and *PiSPH3* and another putative SP, *PiSP3*. A fourth copy of *PiSP1* may be adjacent to *PiSP1* but, due to gaps in the supercontig, there is only a par-



**Fig. 1.** Unrooted Bayesian phylogeny of serine proteases (SP) and serine protease homologs (SPH) from *Phytophthora* species. Gene names are annotated with the species (Pr = *P. ramorum*; Ps = *P. sojae*; Pi = *P. infestans*) followed by protein ID number in the respective genome sequence or the names assigned in this paper. The circled clade contains the set of 18 sequences that include glucanase inhibitor protein (GIP) orthologs and other close related SPH. Cases of identical gene copies are represented by multiple labels at a single branch tip. *PsGIPs* are shown in bold. Partial gene sequences (due to incomplete contigs) are shown in italics. Only posterior probabilities < 0.90 are shown below nodes.

tial sequence (*PiSP1.4\_partial*) available. In the *P. sojae* genome, a region of approximately 318 kb in scaffold 76 shows 13 sequences comprised of GIPs, SPH, and SP, of which six are in a contiguous cluster. Similarly, most of the GIP-like genes in the *P. ramorum* genome are contained in an approximately 22-kb region of scaffold 92, including *PrGIP1* and *PrGIP2*, two SPH, and two SP.

### Members of the *P. infestans* GIP gene family are expressed in culture and in planta.

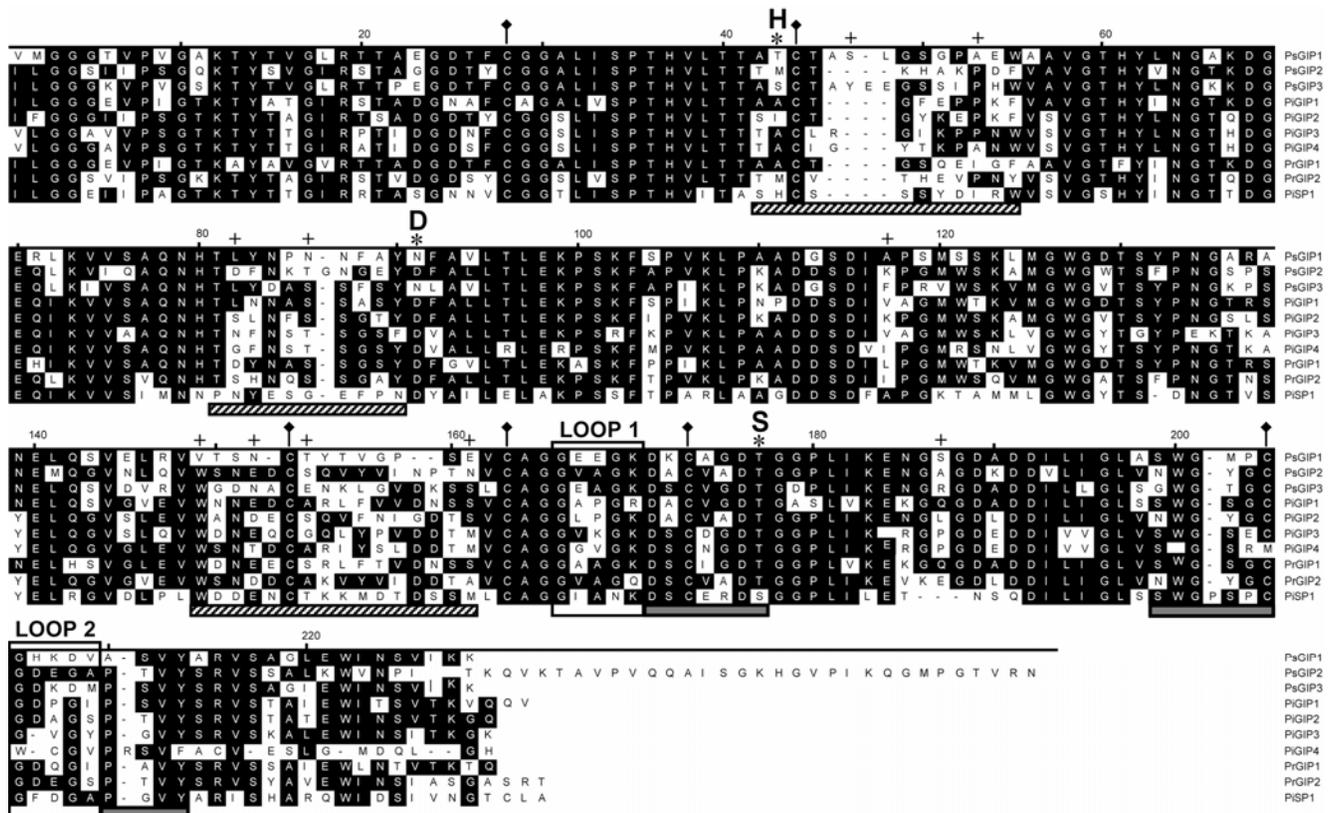
In addition to the broad survey of *Phytophthora* GIPs, particular attention was paid to those from *P. infestans*, and these were targeted in a subsequent study of expression during pathogenesis and interactions with EGases in planta. Four GIP homologs (*PiGIP1* to *PiGIP4*) were identified either by screening a genomic library of a tomato-specialized *P. infestans* isolate (US970001), using *P. sojae* GIP cDNAs as probes, and by searching the *P. infestans* draft genome sequence. *PiGIP2* was cloned and sequenced from a *P. infestans* (US970001) genomic library, while *PiGIP1* and *PiGIP4* were identified in the draft of the *P. infestans* (T30-4 strain) genome sequence. *PiGIP2* sequences from the genomic library and from the draft genome are identical at the nucleotide level. Although *PiGIP3* was originally found in a *P. infestans* expressed sequence tag (EST) collection, the sequence used in this study was derived from the *P. infestans* genome sequence. The PiGIP protein sequences, minus portions of the N-termini that also contain the signal peptides, were aligned (Fig. 2) with the previously reported PsGIPs from *P. sojae* (Rose et al. 2002), and the PrGIPs identified here, using the *P. ramorum* genome sequence. The predicted amino-acid

sequences for all putative GIPs share the SP features present in the PsGIP family, i.e., six conserved cysteine residues involved in disulfide bond formation, variable surface loops 1 and 2, and the S1 substrate binding pocket (Fig. 2; Rose et al. 2002). However, in the PiGIP4 sequence, one of the cysteines presumed to be involved in disulfide bond formation is shifted two positions downstream relative to the equivalent cysteine in other proteins. It is notable that the glycine residue at position 201 in Figure 2, (but referred to as Gly216, according to the chymotrypsin numbering system [Perona and Craik 1995]), which is important for catalysis and influencing substrate specificity, is conserved in all 18 GIP-like sequences.

Genomic Southern blot analysis using a pooled set of the PsGIP cDNAs (*PsGIP1*, *PsGIP2*, and *PsGIP3*) confirmed the existence of a multimer *P. infestans* GIP family (data not shown). Moreover, immunoblot analysis using an antiserum to PsGIP1 (Rose et al. 2002), identified a cross-reacting band in the growth-medium protein extract of *P. infestans* with a similar size to recombinant PsGIP1, used as a positive control, showing that PiGIPs are expressed and secreted in culture (Fig. 3A). When the same protein extract was subjected to two-dimensional electrophoresis followed by an immunoblot analysis, several spots were distinguishable, further suggesting the secretion of multiple PiGIP isoforms (data not shown).

### PiGIPs and tomato EGases are expressed and interact in the infected host leaf apoplast.

In order to determine whether PiGIPs and tomato EGases are present in the leaf apoplast during infection, a vacuum-infiltration and centrifugation method was used to isolate



**Fig. 2.** Sequence alignment of glucanase inhibitor protein (GIP) amino-acid sequences. The consensus amino acids are boxed in black, conserved Cys residues predicted to be involved in the formation of the three disulfide bonds are indicated by a diamond symbol, and the positions of the residues of the catalytic triad (H, D, and S) are shown with asterisks. Sites that are predicted to form serine protease surface loops 1 and 2 are boxed, and residues forming the walls of the substrate binding pocket are underlined with gray boxes. Predicted positively selected sites in PsGIP1 (S47, A53, L82, N86, A117, V149, N152, T154, E161, and S187) are denoted with a plus sign, and the hypervariable loops are underlined with gray hatched boxes. The N-termini are not shown, and the numbering follows a mature version of the PsGIP1 sequence (position 1 in the alignment corresponds to V29 in the full-length protein).

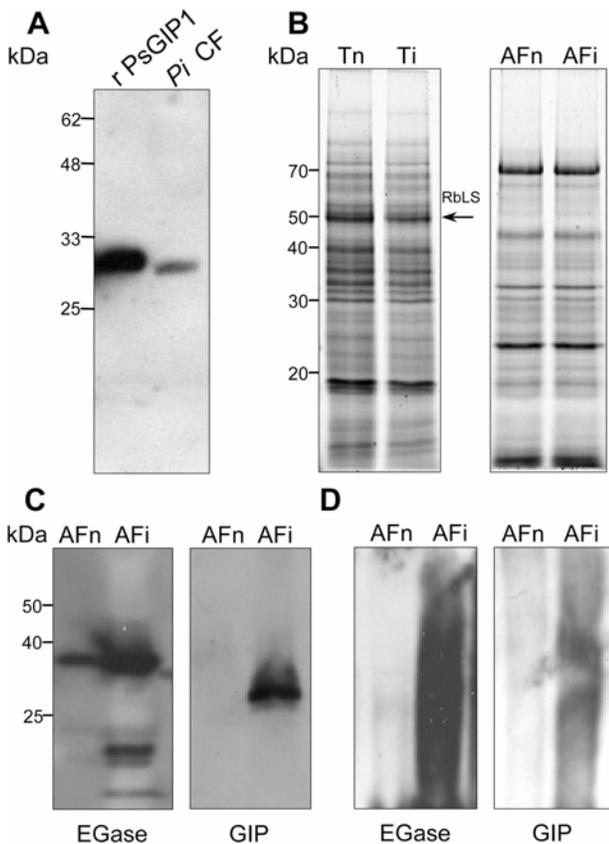
apoplastic fluid from inoculated and noninoculated tomato leaves during a timecourse of disease at 0 to 4 days postinoculation (dpi). A comparison of the proteins present in either the apoplastic extract or total leaf protein by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) revealed substantial differences in the banding patterns (Fig. 3B). Major bands corresponding to the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO; 52 kDa), the most abundant intracellular leaf protein, were detected in SYPRO Ruby-stained gels (Fig. 3B, left panel) and by immunoblot analysis using an anti-RuBisCO antiserum (data not shown). However, the equivalent protein bands were not detected in the apoplastic protein extracts (Fig. 3B, right panel), suggesting minimal contamination of the apoplastic fluid with intracellular proteins and, thus, a significant enrichment in the concentration of extracellular proteins.

No major differences were observed in the banding pattern of apoplastic proteins in the SYPRO Ruby-stained SDS-PAGE gels between inoculated and noninoculated leaves (Fig. 3B, right panel); however, clear differences in protein expression were apparent by immunoblot analysis, using antisera to GIPs and EGases (Figs. 3C and D). Hybridization with the EGase antiserum revealed a single band of approximately 37 kDa in

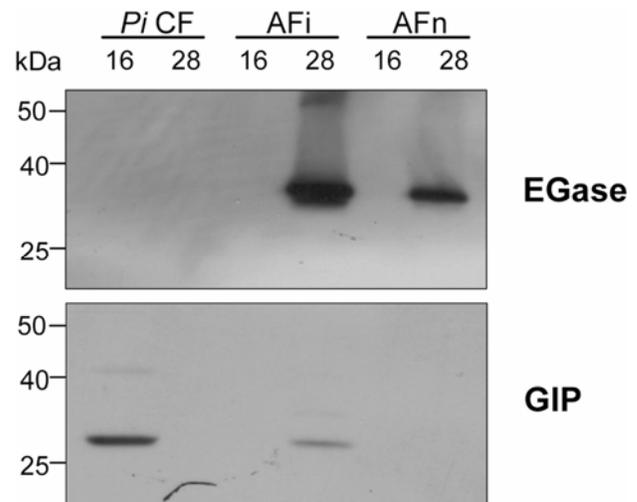
the noninoculated sample, but upon infection, the expression level was far greater and other cross-reacting bands of lower molecular weight were detected (Fig. 3C, left panel). In contrast, the PsGIP antiserum showed no cross-reaction with apoplastic proteins from noninoculated leaves but a strong cross-reacting band of approximately 30 kDa was observed in the inoculated samples (Fig. 3C, right panel). Finally, it was noted that the anti-GIP antibody did not cross-react with a total infected-leaf protein extract, underscoring the value of enriching for apoplastic proteins that would otherwise be below the detection limit.

To investigate whether PiGIPs and tomato EGases interact in the host leaf apoplast *in vivo*, apoplastic protein extracts from noninoculated and inoculated leaves (3 dpi) were subjected to native gel electrophoresis in a basic buffer. Under such conditions, basic extracellular EGases are not sufficiently charged to migrate into the gel but are present in the gel when complexed with acidic GIPs (Rose et al. 2002). The native gels were electroblotted onto membranes, which were then probed with PsGIP1 or EGase antisera. While no substantial cross-reaction was seen with either antiserum in the sample from noninoculated leaves, a strong smeared signal was detected with either antiserum in the equivalent inoculated sample (Fig. 3D), indicating that PiGIPs and EGases form an apoplastic complex in tomato leaves during *P. infestans* infection.

To confirm the interaction, gel filtration chromatography coupled with immunoblotting was used to examine the elution profiles of EGases, GIPs, and the putative complexes in the apoplastic protein extracts. Specifically, the assay involved comparing the migration patterns of PiGIPs that originated from either culture filtrates or apoplastic fluid in size-exclusion column fractions. Proteins from *P. infestans* culture filtrate and apoplastic fluid from 3-dpi and noninoculated samples were concentrated and loaded separately onto a Superdex 75 HR 10/30 size exclusion column, and aliquots from eluting fractions were dot-blotted onto membranes, which were then probed with the PsGIP or EGase antisera. Fractions that showed cross reactivity were then separated on SDS-PAGE gels and were similarly subjected to immunoblot analyses (Fig. 4). While PiGIPs that were concentrated from the culture filtrate consis-



**Fig. 3.** Native, and denaturing polyacrylamide gel electrophoresis (PAGE) blot analyses of glucanase inhibitor protein (GIP) and endo- $\beta$ -1,3-glucanase (EGase) expression in tomato leaves inoculated with *Phytophthora infestans*. **A**, Immunoblot analysis of proteins from *P. infestans* culture filtrates (PiCF) using the PsGIP1 polyclonal antisera. rPsGIP1 corresponds to recombinant PsGIP1, used as a positive control. Molecular-weight markers are indicated. **B**, Sodium dodecyl sulfate-PAGE showing the pattern of tomato leaf total protein or leaf apoplastic fluid extracts (Ti = total extract from inoculated leaves, Tn = total extract from noninoculated leaves, Afi = apoplastic fluid extracted from inoculated leaves, AFn = apoplastic fluid extracted from noninoculated leaves). Gels were stained with Sypro Ruby. RbLS = RuBisCO large subunit. **C**, Denaturing-protein gel blot or **D**, native-protein gel blot probed with EGase and GIP antisera. Molecular weight markers are shown.



**Fig. 4.** Immunoblot analysis of size-exclusion separation of glucanase inhibitor proteins (GIPs) and endo- $\beta$ -1,3-glucanases (EGases). Size-exclusion separation was performed in *Phytophthora infestans* culture filtrate (PiCF) and apoplastic fluid from inoculated (Afi) and noninoculated (AFn) leaves. Fractions eluted from the column at 16 and 28 min were fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and were subjected to immunoblot analysis using EGase and GIP antisera. Molecular weight markers are shown.

tently eluted from the column after 16 min, the migration pattern was substantially different in the sample from inoculated leaves, since immunoreactive bands with the same apparent molecular mass were detected only after 28 min (Fig. 4). Importantly, these cross-reacting proteins coeluted with the EGase proteins, as evidenced using the EGase antiserum.

#### ***PiGIP* genes are expressed during infection and *PiGIP1* and *PiGIP3* are induced in planta.**

Immunoblot analysis of a leaf infection timecourse indicated that *PiGIPs* and EGases were present in the tomato apoplast during infection and that their expression levels increased with disease progression (Fig. 5A). Furthermore, it was apparent that EGase expression was induced upon pathogen attack and that additional proteins that cross-reacted with the EGase antiserum were expressed at the same time that *PiGIPs* were first detected (2 dpi).

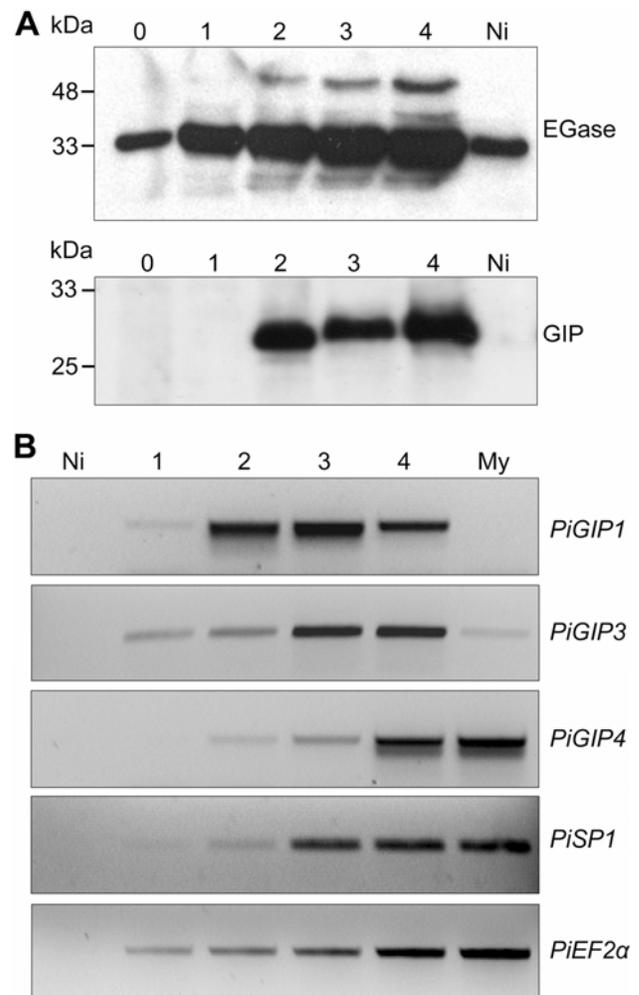
The temporal expression patterns of *PiGIP* genes in *P. infestans*-inoculated tomato leaves were monitored from inoculation to extensive sporulation of the pathogen (0, 1, 2, 3, and 4 dpi), using semiquantitative reverse transcription-polymerase chain reaction (RT-PCR) analysis. To assess the potential upregulation of *PiGIPs* in planta, mRNA extracts from axenically grown mycelia and from noninoculated leaves were used for comparison, and the increase in pathogen biomass was evaluated by quantifying the levels of a constitutively expressed *P. infestans* gene, elongation factor-2 $\alpha$  (*PiEF2 $\alpha$* ), as previously described (Torto et al. 2002). Other than *PiGIP2*, all the *PiGIP* transcripts were detected at 1 or 2 dpi (Fig. 5B). *PiGIP1* and *PiGIP3* expression was up-regulated in planta compared with mycelia grown in vitro. While *PiGIP1* expression peaked at 3 dpi and was not detected in the mycelial sample, *PiGIP3* showed a linear increase in expression, peaking at 4 dpi, which correlated with the increase in pathogen biomass. *PiGIP4* and *PiSP1* showed an expression pattern similar to the constitutively expressed *PiEF2 $\alpha$*  gene. When amplifying the *PiGIP4* fragment, a second faint lower band was detected at 4 dpi and in the mycelial sample, which likely represented an unrelated sequence. *PiGIP2* was only detected using gene-specific primers, at extremely low levels at 4 dpi and in the mycelial sample (data not shown). Since the *PiGIP2* genomic sequence and the amplified open reading frame (ORF) are the same size, it was not possible to definitively demonstrate that this corresponded to the amplified RT-PCR product. However, we note that total RNA extract was DNase-treated and a control reaction lacking reverse transcriptase was used to confirm the absence of genomic DNA contamination. It therefore remains to be seen whether *PiGIP2* is expressed at very low levels or only under circumstances not evaluated in our study.

#### **Evidence for positive selection in PsGIP1 residues in the vicinity of adaptive residues in the bound EgaseA.**

Codon evolution models and structural modeling were used to identify positively selected sites in GIPs and to assess whether they potentially interact with EGases. From the 48 sequences shown in Figure 1, six subsets of the 23 unique SPH sequences were used in the analysis. Significant evidence of positive selection was detected by the full-effects likelihood, single-likelihood ancestor counting (SLAC) or by modified maximum likelihood codon evolution (CODEML) models algorithms in each data set (Table 1). A total of 10 positively selected sites were identified (Table 1), of which six were identified in more than one data set and three of these six by more than one method. Of particular note are S47, V149, and N152, which were identified as positively selected in two independent segments of the phylogeny, and N86, which is identified in

three data sets and by two different methods. Figure 6 shows two views of the structural model of PsGIP1 docking to soybean EGaseA, an interaction that we previously presented, focusing primarily on the EGase (Bishop et al. 2005). An animation of the predicted PsGIP1-EGaseA docking is shown in Supplemental Video 1 (online). Analysis of the docking model shows that these positively selected PsGIP1 residues are on the molecule surface and that most are close to the catalytic cleft of EGaseA, and a region (the ring of fire) suggested by Bishop and associates (2005) to have several amino acids that are rapidly evolving in response to selection.

The docking model (Fig. 6) suggests that eight of the 10 positively selected peptide sites from PsGIP1 are members of hypervariable loops (Fig. 2) that extend downward to the surface of the bound EGase. These loops and possibly the positively selected sites themselves contact several positively selected EGase sites as well as many of the hypervariable EGase sites found on the margins of the active site (the precise location of these resi-



**Fig. 5. A**, Immunoblot analysis of apoplastic fluid from inoculated (lanes 0 to 4) and noninoculated (lane, Ni) tomato leaves. Leaves were collected from 0 to 4 days after inoculation. Immunoblots were probed with endo- $\beta$ -1,3-glucanase (EGase) and glucanase inhibitor protein (GIP) antisera. Molecular-weight markers are indicated. **B**, Reverse transcription-polymerase chain reaction (RT-PCR) analysis of *PiGIP* genes during the timecourse of the disease, using gene-specific primers for each *PiGIP* gene (*PiGIP1*, *PiGIP-3* and *PiGIP-4*) and *PiSP1*. RT-PCR was conducted under nonplateau PCR conditions. *Phytophthora infestans* elongation factor 2 $\alpha$  (*PiEF2 $\alpha$* ) was used as a constitutive expression control. Samples were labeled as follows. Ni = noninoculated tomato leaves, My = Mycelia grown in culture media. Lanes 0 to 4 represent inoculated tomato leaves 0 to 4 days postinoculation.

dues likely changes from gene to gene owing to numerous indels within each loop). For example, PsGIP1 residues V149 and N152 (part of loop V149 to E161) appear in close proximity to positively selected EGaseA residues L9 and Q289. These EGaseA residues are located on the margins of the catalytic cleft and are predicted to contact each other and the catalytic residue Y34 in the EGaseA protein tertiary structure. Of particular note are S47 and N86, which are part of two other descending loops and for which there was particularly strong evidence of positive selection (based on multiple data sets and methods). Site N86 (part of loop T81 to Y90) is inserted into the surface of EGaseA among a set of hypervariable EGaseA residues, including N62 and D58 of EGaseA, while S47 appears to make contact with EGaseA positively selected residues S150 and A155. Finally, two of these hypervariable loops (T81 to Y90 and A42 to W55) appear to clasp the loop in EGaseA containing E99 (P98 to F102), itself a hypervariable site in the EGaseA ring of fire. E99 and its adjacent residues protrude prominently above K97 (Fig. 6, red), the EGase residue predicted to bind in the SP K/R binding pocket of PsGIP1, in which changes in charge and conformation may guard it.

## DISCUSSION

GIPs show high homology to the S1A subfamily of SP or chymotrypsin proteases (Perona and Craik 1995) and constitute the first reported example of serine protease homologs (SPH) that function as enzyme inhibitors (Rose et al. 2002). SP often show modular evolution of their catalytic and substrate binding sites, and a taxonomically broad range of organisms have both SP and SPH, in which mutations within the catalytic site result in loss of proteolytic activity but substrate-binding capacity is retained. SPH, such as GIPs, are being

identified with increasing frequency, and a comprehensive search of the *Drosophila melanogaster* genome for SP-related genes identified 204 candidates, of which 28% were designated SPH due to lack of a complete catalytic triad (Ross et al. 2003). While 33 of these SPH sequences appear to be expressed, only one has been assigned a putative function. Indeed, while SPH from several organisms have been associated with a diverse range of biological processes, their mechanisms of action are generally not well understood. Among the SPH in Figure 1, EST have been identified for PiSPH1 and, among GIPs, for PiGIP3 and all PsGIPs. The functions of most of the proteins remain unknown. Nevertheless, as previously mentioned, other functionally important sites in SP are highly conserved in these SPH (Fig. 2), indicating retention of function.

Although the genomes of *P. sojae* and *P. ramorum* show a high level of colinearity and general orthology, in the case of the GIP family, there is no clear relationship between the family phylogeny (Fig. 1) and the gene locations on the supercontigs in each species, which suggests a complicated history. This should be clarified in the future when genes from each genome are assigned to a specific chromosome.

To date, information regarding the expression and interactions of GIPs has been limited to a description of a subset of the *P. sojae* GIP gene family (Rose et al. 2002), the identification of a GIP homolog in *P. parasitica* (Panabieres et al. 2005), and PiGIP3 in a computational analysis of *P. infestans* cDNAs (Win et al. 2006). Questions therefore remain about the expression patterns and potential roles of different GIPs during pathogenesis and their possible interaction with host EGases in the plant apoplast.

Based on genomic library screening and a *P. infestans* genome database search (Figs. 1 and 2), a four-member *P. infestans* GIP gene family was identified (PiGIP1 to PiGIP4).

**Table 1.** Results of codon evolution analyses of glucanase inhibitor proteins (GIPs) and serine protease homologs (SPH) homologs<sup>a</sup>

Data Set <sup>c</sup>	PiGIP1 residue <sup>d</sup>	Site <sup>d</sup>	Residue <sup>e</sup>	Nonsynonymous <sup>f</sup>	Synonymous <sup>f</sup>	dN <sup>g</sup>	dS <sup>g</sup>	P value <sup>b</sup>	
								Rating <sup>h</sup>	FEL <sup>g</sup>
7 right	ASL	47	4	5	0.0	0.69	0.00	NS	0.10
	IAP	117	5	6.5	0.5	1.4	0.16	NS	0.008
	VVT	149	5	6	0.0	0.94	0.00	NS	0.07
	CTY	154	4	7.3	2.7	2.02	0.98	0.74	NS
8 left	PNN	86	6	7	0.0	0.70	0.00	NS	0.03
	VVT	149	4	8	0.0	0.88	0.00	NS	0.07
	GSG	187	7	11.8	5.3	2.28	0.72	0.91	NS
8 bottom	ASL	47	5	7	0.0	0.51	0.00	NS	0.09
	SNC	152	5	10.7	3.3	2.56	0.44	0.62	NS
11 right	TLY	82	5	13	6	2.13	0.87	0.76	NS
	IAP	117	6	13	1	1.12	0.11	NS	0.02
	SNC	152	5	10.5	2.5	0.70	0.00	NS	0.10
	CTY	154	7	12.3	4.7	2.02	0.98	0.90	NS
	SEV	161	5	11	6	2.22	0.78	0.90	NS
15 top	PNN	86	7	13.5	1.5	0.78	0.14	0.09	0.06
	IAP	117	7	15.5	3.5	1.06	0.14	NS	0.06
	VVT	149	7	13	1	0.84	0.11	NS	0.06
23	ASL	47	10	26.5	1.5	0.74	0.14	0.10	0.09
	PAE	53	7	16.7	3.3	0.43	0.28	0.09	NS

<sup>a</sup> Results for putative serine proteases (SP) are all nonsignificant and are not shown.

<sup>b</sup> NS denotes nonsignificant ( $P > 0.1$ ).

<sup>c</sup> Number of sequences in data set. Sets "7 right" and "8 left" are combined into "15 top"; "15 top" and "8 bottom" are combined into the "23"; "11 right" includes "7 right" and half of "8 left".

<sup>d</sup> Numbering according to PsGIP1, where the first V in VMGGGTVP is site 1. The PiGIP1 residue is shown in bold with flanking residues shown to facilitate location.

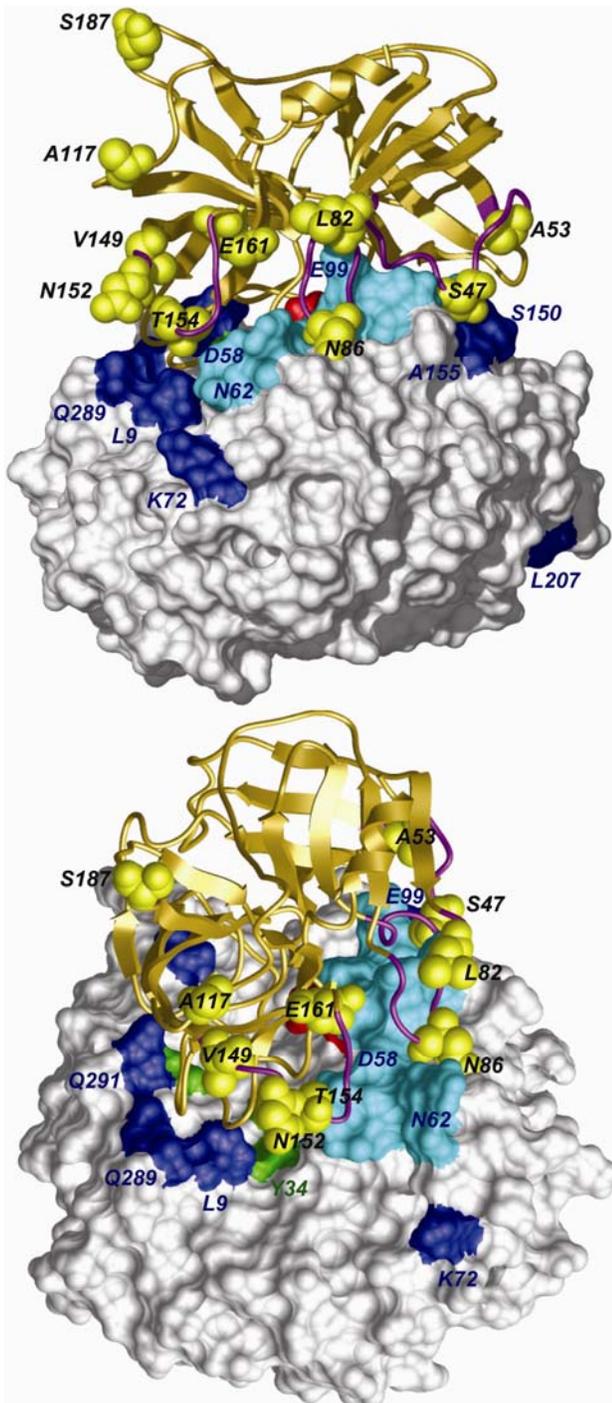
<sup>e</sup> Number of different residues at this position in the alignment.

<sup>f</sup> Number of nonsynonymous (Non) and synonymous (Syn) mutations were estimated using CODEML.

<sup>g</sup> Nonsynonymous (dN) and synonymous (dS) rates estimated by FEL (full effects likelihood model implemented in HyPhy) or by CODEML if identified by CODEML as positively selected.

<sup>h</sup> Rating according to single likelihood ancestor counting (SLAC) implemented in HyPhy or posterior probability estimated by CODEML (PAML). Though several of the posteriors are not high, they are included for illustrative purposes.

The only report to date of EGases and GIPs interacting in planta was provided by our earlier study of *P. sojae*-inoculated soybean roots; however, while that analysis confirmed in vitro binding studies, the possibility remained that this association



**Fig. 6.** Two views of a soybean endo- $\beta$ -1,3-glucanase–*Phytophthora sojae* GIP (EGaseA:PsGIP1) docking model and positively selected sites in PsGIP1. EGaseA structure is shown in white (surface with shadows) and PsGIP1 in copper (ribbon). The active-site residues of EGaseA (Y34, E238, and E295) are shown in green (E238 is not visible in these views). Sites that are predicted to have undergone positive selection, either in PsGIP1 or EGaseA, are shown in yellow and dark blue, respectively. PsGIP1 loops containing positively selected residues are shown in magenta (Loop 1, A42 to W55; Loop 2, T81 to Y90; and Loop 3, V149 to E161). Hypervariable EGaseA loops that are in close proximity to PsGIP1 magenta loops are shown in light blue (D58 to N62, P98 to F102 and D151 to Y152). K97 residue from EGaseA is shown in red.

was promoted as a consequence of homogenizing the inoculated plant material. To confirm that GIPs and EGases form complexes in the host plant apoplast, two different strategies were used in the present study to evaluate potential interactions in apoplastic fluid that was isolated nondestructively from *P. infestans*-inoculated tomato leaves. Immunoblot analysis of proteins that had been separated on gels under nondenaturing conditions indicated that PiGIPs and EGases are both present in the apoplast during infection and that they form stable complexes in vivo (Fig. 3D). This conclusion was further supported by examining the migration pattern of PiGIPs and EGases using size exclusion chromatography (Fig. 4), and it was shown that PiGIPs elute differentially, depending on whether they originated from culture filtrate or apoplastic fluid extracts. Interestingly, GIPs were observed to elute more rapidly from the size-exclusion column in the absence of potential EGases ligands (i.e., when derived from *P. infestans* culture filtrate) than in the presence of EGases (i.e., from inoculated tomato leaves). This might be considered counterintuitive, since a PiGIP-EGase complex of larger molecular weight would be expected to elute more rapidly than an unbound GIP monomer. However, it has previously been noted that, for currently unknown reasons, GIPs also exhibit anomalous migration behavior during electrophoresis (Rose et al. 2002). This is apparently not usual and several other such examples have been reported for different proteases (Nagamine-Natsuka et al. 1995). Importantly however, the immunoblot analyses indicated that the PiGIPs and EGases coeluted, further suggesting the formation of PiGIP:EGase complexes.

While it has been shown that GIPs are expressed in planta during infection (Rose et al. 2002), nothing has been reported to date regarding the number of different GIP isoforms that are expressed or differences in the temporal expression patterns that might provide insights into protein function. An examination of the expression of the four PiGIPs over an infection timecourse (Fig. 5B) revealed distinctly different patterns of expression, which may indicate functional specialization. PiGIP1 and PiGIP3 transcript levels were upregulated during infection, coincident with the elevated levels of host EGases in the apoplast (Fig. 5A), suggesting that the latter may be ligands for the induced PiGIPs. In contrast, PiGIP1 showed higher expression levels between 2 and 3 dpi, a timepoint before the development transition of *P. infestans* from a biotrophic to a necrotrophic phase (Kanneganti et al. 2006; van West et al. 1998). The temporal expression data support a model in which PiGIP3 and the constitutively expressed PiGIP4 defend the pathogen from host EGases that are constitutively expressed in the apoplast or that are secreted during early infection. Subsequent cell lysis then releases EGases from the vacuole that are targets for specific GIP isoforms, such as PiGIP1, which are expressed later during infection. The potentially functional SP PiSP1 showed an expression pattern similar to a control constitutively expressed gene, suggesting that it could also be important to the pathogen prior to infection. The apparently low abundance of GIPs in vivo thwarted attempts to identify the specific interacting GIP-EGase pairs, and repeated attempts to identify interactors using coimmunoprecipitation analyses were not successful. This question will be the target of future studies.

It is apparent that a complex mixture of EGases and GIPs are present in the host apoplast during pathogenesis, and we hypothesized that the presence of multiple isoforms from each protein family might have arisen through a process of duplication and coevolutionary specialization. Indeed, our evidence suggests that GIPs and their EGase ligands are examples of a relatively rapid coevolutionary interaction characteristic of an arms race. Bishop and associates (2005) showed that most

positively selected sites and rapidly evolving sites comprised a ring of fire around the margin of the EGaseA active site cleft in close proximity to PsGIP1. In addition, escalation is a crucial characteristic of an arms race (Stahl and Bishop 2000), and the presence of many different isoforms of the inhibitor (Fig. 1) suggests such an escalation of defenses against the attacking hydrolytic enzymes. However, it has been noted that the *P. sojae* and *P. ramorum* assemblies carry a number of artifactual duplications, which are usually quasi-identical in sequence. Therefore, we cannot rule out the possibility that the identical gene copies shown in Figure 1 are the result of assembly artifacts, which have been noted in these particular assemblies (S. Kamoun, *personal communication*).

In order to better support the arms race hypothesis, both members of an antagonistically interacting complex should display a characteristic signature of rapid adaptive evolution. To better understand GIP-EGase coevolution, and to investigate variation in GIP sequences for reciprocal evolutionary patterns, codon evolution models were applied to a set of GIPs and related SPH and SP of three species of genus *Phytophthora*. We found that GIPs show 10 positively selected sites, of which eight are adjacent to positively selected or rapidly evolving sites positioned on the margins of the EGaseA active site. These include two striking cases (S47 and N86) in which GIP residues are inserted into or pressed against the surface of EGaseA sites shown to be hypervariable or positively selected. Also of interest is the PsGIP1 N152 residue, which is positioned very closely to PsGIP1 D171, the key negatively charged residue in the K/R binding pocket of SP. Substitutions and indels at these sites may change the conformation of the pocket to match the target. Many positively selected and otherwise highly variable sites in both proteins interact along the margin of the active site above EGaseA K97, suggesting the possibility that rapidly changing sites in both GIPs and EGases are experiencing mutations that are favored because they determine whether K97 will enter the K/R binding pocket. On the other hand, it is likely that not only K97 but also additional residues bind GIPs to EGases, as suggested by the fact that there are rapidly evolving rims on both sides of the active site. Our present work on GIP function and evolution, together with our previous analysis of EGases positive selection (Bishop et al. 2005), reveals that rapidly evolving and positively selected sites in both GIPs and EGases are in close proximity to each other and strongly suggests an ongoing arms race between these two protein classes. Not surprisingly, there is a growing list of secreted proteins from *P. infestans* and other oomycetes showing evidence of adaptive selection, consistent with an arms race coevolution with the host (Liu et al. 2005; Win et al. 2007). As the field progresses and more plant pathogen genome sequences are available, this number will certainly increase.

## MATERIALS AND METHODS

### *P. infestans* growth conditions and tomato inoculation.

The tomato-specialized *P. infestans* isolate US970001 (ATCC number MYA-2350), a member of the US-17 clonal lineage (Smart et al. 2003), was cultured in rye B agar (Caten and Jinks 1968) or modified Plich liquid medium (Lee et al. 1997) at 15°C with 18 h of light per day. To generate zoospores for subsequent inoculation of tomato (*Solanum lycopersicum* cv. Rio Grande) leaves, detached tomato leaflets were inoculated with a *P. infestans* sporangial suspension and were incubated in moist chambers for 5 days (Smart et al. 1998). Sporangia were then collected in water, diluted to 20,000 sporangia ml<sup>-1</sup>, and were incubated at 4°C for at least 2 h to induce release of zoospores, which were used in inoculation experiments. Tomato plants were grown in a greenhouse (14-h-light and 10-h-dark

conditions) and 4- to 6-week-old plants were inoculated with *P. infestans* zoospore suspensions in a humid chamber at 15°C with 16 h of light and 12 h at 100% relative humidity (Smart et al. 2003). Symptoms were monitored daily, and leaves were harvested at 0, 1, 2, 3 and 4 days for apoplastic fluid (protein) or RNA extraction.

### Identification of GIP homologs, sequence alignments, and phylogenetic analysis.

The full-length cDNAs of three previously reported *P. sojae* GIP cDNAs (Rose et al. 2002) were used to probe a genomic library of *P. infestans* isolate US970001 (provided by C. Smart, as described in the Lambda Fix II library manual (Stratagene, La Jolla, CA, U.S.A.). The PsGIP ORF sequences were also used to search the National Center for Biotechnology Information nonredundant and EST databases, the *Phytophthora* Functional Genomics Database, and the *P. sojae*, *P. ramorum*, and *P. infestans* respective genome databases, using BLASTp and tBLASTn programs (Altschul et al. 1990). Initially, a threshold score of  $e^{-20}$  was used and over 70 sequences were selected. A preliminary DNA alignment was performed, and a group of 48 more closely related sequences was selected for further analysis. We named *P. sojae* and *P. ramorum* genes and proteins according to their protein IDs in their respective genome databases. Using the data in the *P. infestans* genome databases, we assigned *P. infestans* gene names as described in Supplementary Table 2. These GIP-related homologs (nucleotide sequences with predicted N-terminal signal peptide regions omitted) were then aligned using ClustalW software (Thompson et al. 1997) and were further manually aligned using BioEdit (Hall 1999) to minimize the number of substitutions. Sequences were categorized as SP or SPH, based on the presence or absence of the His-Asp-Ser catalytic triad (Perona and Craik 1995), respectively. The N-terminal portions, including the signal peptides, were removed so residue numbering is the same for Figures 2 and 6. The phylogenetic tree in Figure 1 was generated using Mr. Bayes 3.1 (Ronquist and Huelsenbeck 2003), using the site-specific rate model for coding sequences with four chains running 900,000 generations and discarding the first 25% as “burn-in.” Several alternative sequence alignments produced very similar trees.

### Detection of positively selected residues in GIPs.

The set of 48 diverse sequences presented in Figure 1 can be divided into 17 putative SP and 31 SPH, based on the absence or presence of substitutions at one or more of the three catalytic residues. We analyzed six semihierarchical subsets of 23 SPH (identical duplicates and several genes containing non-sense mutations were omitted) and separately analyzed the set of SP. The set of 23 SPH was divided into naturally clustered groups of 15 (Fig. 1, circled) and eight (Fig. 1 bottom). The group of 15 was arbitrarily subdivided into “left” and “right” groups of eight and seven sequences as well as an expanded group of 11 right sequences.

The sets were analyzed for evidence of selection using CODEML (PAML 3.14) with settings to account for codon bias (Yang et al. 2005). Phylogenies for use with these models were estimated by neighbor joining, using HyPhy. To test for positive selection acting on the sequence, we compared  $2 \cdot \Delta \ln(\text{likelihood})$  of the neutral (M7) and positive selection (M8) models to a  $\chi^2$  distribution with degrees of freedom = 1. The CODEML Naive Empirical Bayes test was used for selection at individual sites and present sites with posterior probabilities >0.6, because it is unclear what the threshold posterior probability should be. In addition, the fixed-effects likelihood (FEL) and SLAC models implemented in HyPhy were applied. The FEL model is a generalization of the models implemented

in CODEML but differs in that FEL explicitly models site-to-site synonymous rate variation and estimates a  $P$  value for the hypothesis  $dN > dS$  for each site (Pond and Frost 2005a and b; Pond et al. 2005). For FEL, HKY85 was used as the substitution model, which was identified as the best matrix by HyPhy's model selection algorithm. The SLAC model is a modification of the method proposed by Suzuki and Gojobori (1999). We regard all designations of positive selection by these methods as hypotheses and suggest that those with  $P > 0.10$  (SLAC and FEL analyses) or posterior  $P < 0.9$  (CODEML) should be treated with greater caution.

### RNA isolation and RT-PCR.

Total RNA extractions were carried out with the RNeasy mini kit (Qiagen, Valencia, CA, U.S.A.), according to the manufacturer's instructions. Following DNase I treatment, RNA was quantified and 2.5  $\mu\text{g}$  was converted to first-strand cDNA with oligo-dT primers using the SuperScriptIII cDNA synthesis kit (Stratagene). The gene-specific primers used for analysis of *PiGIP* gene expression are listed in Supplementary Table 3. For each sample, a minus RT negative control was included to allow assessment of genomic DNA contamination. A linear amplification range for each gene was tested, varying the number of amplification cycles (24 to 28 cycles), except for *PiGIP2*, which required 35 cycles for detection. The PCR conditions were 94°C for 3 min, 24 to 28 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 45 s, followed by 72°C for 5 min. The relative abundance of *PiGIP* transcripts was determined by comparison with the expression of the constitutively expressed *PiEF2 $\alpha$* , using the primers described by Torto and associates (2002). ORF of *PiGIP1* to *PiGIP4* and *PiSPI.1* genes were cloned from *P. infestans* isolate US-970001 genomic DNA, and the sequences were deposited in GenBank (accession numbers: *PiSPI.1*, EU443391; *PiGIP1*, EU443395; *PiGIP2*, EU443392, *PiGIP3*, EU443393; and *PiGIP4*, EU443394).

### Protein extraction, protein gel electrophoresis and Western analysis.

Proteins in tomato leaf apoplastic fluid were isolated by vacuum-infiltrating mature leaves with a solution of 500 mM NaCl and 0.01% Silwet L-77 and centrifuging at 1,350  $\times g$  for 10 min at 4°C (Lohaus et al. 2001). Total protein extracts from leaves were obtained as described by Rose and associates (2002). Proteins in *P. infestans* culture media were isolated by adjusting the culture filtrate to 90% ammonium sulfate, resuspending the pellet in 50 mM sodium acetate, 10 mM NaCl, pH 6.0, and dialyzing against 30 mM Tris, pH 8.0, using a molecular-weight cut off of 10,000. All protein concentrations were estimated using the Bradford assay (Bradford 1976) with bovine serum albumin as a standard. Protein samples were subjected to electrophoresis in 12% SDS-polyacrylamide gels under denaturing conditions described by Rose and associates (2002) or under native conditions by Tris-glycine native-PAGE (Novex 4 to 20% tris-glycine gel; NOVEX, San Diego, CA, U.S.A.), according to the manufacturer's instructions. For visualization, gels were stained with Sypro Ruby according the manufacturer's instructions (BioRad, Hercules, CA, U.S.A.). For immunoblot analysis, proteins were transferred to polyvinylidene difluoride (PVDF) membrane (Rose et al. 2000, 2002), and blots were incubated with PsGIP1 polyclonal antiserum (1:2,000 dilution) or tobacco PR-2c (EGase) polyclonal antiserum (1:5,000 dilution) as described by Ham and associates (1997), and hybridizing bands were visualized using the ECL kit (GE Healthcare, Piscataway, NJ, U.S.A.) according to the manufacturer's directions. For the positive control shown in Figure 3, recombinant PsGIP1 (rPsGIP1) was used, as previously described (Rose et al. 2002).

### Size-exclusion chromatography.

Protein extracts from culture filtrate of *P. infestans* grown in liquid Plich medium and apoplastic fluid from *P. infestans*-inoculated or uninoculated tomato leaves were fractionated by size-exclusion chromatography, using a Superdex 75 HR 10/30 column and an FPLC system (GE Healthcare). Protein samples were concentrated and buffer-exchanged using an Ultrafree 0.5 centrifugal filter device with a 5-kDa cutoff (Millipore, Bedford, MA, U.S.A.), with 200  $\mu\text{l}$  applied (0.5 ml  $\text{min}^{-1}$ ) to the column in running buffer (75 mM sodium acetate, 500 mM NaCl, pH 5.5). Eluting fractions (0.5 ml) were precipitated with a final concentration of 10% TCA (trichloroacetic acid), and duplicate samples were applied to a PVDF membrane, using a dot blot apparatus (GE Healthcare). Membranes were then probed with either PsGIP1 or EGase antisera. The size-exclusion separation was then repeated under the same conditions as described above, and the 0.5-ml fractions shown to contain GIPs or EGase, or both, were precipitated with TCA (adjacent fractions were pooled into 1-ml fractions prior to precipitation). Fractions corresponding to 16- and 28-min elution times were analyzed by SDS-PAGE, blotted onto PVDF membrane and probed with either PsGIP1 or EGase antisera.

### GIP and EGase molecular modeling.

The BIOINFO metasever (Ginalski et al. 2003) was used to identify the best templates for modeling the structure of PsGIP1. Pairwise alignments involving the protein  $\beta$ -acrosin from ram spermatozoa were assigned the highest scores by the "3D-Jury" tool used by the metasever. This structure was then selected as the optimal template, and its coordinates were retrieved from the Protein Data Bank (PDB) template (PDB code: 1FIW). Sequence alignments from the metasever and coordinates of the template structure were subsequently used as input of the program MODELLER (Sali and Blundell 1993; Sali et al. 1995; Sanchez and Sali 2000) to generate three-dimensional models (3-D) for PsGIP1. Similarly, 3-D models of soybean EGaseA were built, based on the X-ray structures of barley endo- $\beta$ -1,3-1,4-glucanase (PDB code: 1AQ0) as the 3-D structural templates, using the MODELLER program. The EGase model was compared with a new structure for a banana endo- $\beta$ -1,3-glucanase (PDB code: 2CYG) and was found to be extremely similar. The model of EGaseA docked to PsGIP1 was generated manually as described in Bishop and associates (2005). Assumptions made for generating the model are i) EGaseA inhibition is obtained by total or partial obstruction of the catalytic site by PsGIP1 and ii) PsGIP1 binds to EGaseA in a trypsin-like manner (i.e., by recognizing a lysine or an arginine residue on the substrate molecule).

### ACKNOWLEDGMENTS

The authors would like to thank C. Smart (Cornell University) for providing the *P. infestans* genomic library and for helpful discussion. We are also grateful for the assistance and advice provided by A. McLeod (University of Pretoria, South Africa) and to W. Fry and his lab members for providing *P. infestans* isolates and technical assistance (Cornell University). This work was supported by grants from the National Science Foundation's Plant Genome Program (award number DBI 0606595) and the New York state office of Science, Technology and Academic Research. Part of this work was carried out using the resources of the Computational Biology Service Unit from Cornell University, which is partially funded by the Microsoft Corporation. C. M. B. Damasceno was supported by a student fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

### LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

- Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340:783-795.
- Bishop, J. G. 2005. Directed mutagenesis confirms the functional importance of positively selected sites in polygalacturonase inhibitor protein. *Mol. Biol. Evol.* 22:1531-1534.
- Bishop, J. G., Dean, A. M., and Mitchell-Olds, T. 2000. Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. U.S.A.* 97:5322-5327.
- Bishop, J. G., Ripoll, D. R., Bashir, S., Damasceno, C. M., Seeds, J. D., and Rose, J. K. C. 2005. Selection on glycine beta-1,3-endoglucanase genes differentially inhibited by a *Phytophthora* glucanase inhibitor protein. *Genetics* 169:1009-1019.
- Bradford, M. M. 1976. Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Anal. Biochem.* 72:248-254.
- Caten, C. E., and Jinks, J. L. 1968. Spontaneous variability of single isolates of *Phytophthora infestans*. I. Cultural variation. *Can. J. Bot.* 46:329-348.
- Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. 2006. Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell* 124:803-814.
- Czapinska, H., and Otlewski, J. 1999. Structural and energetic determinants of the S1-site specificity in serine proteases. *Eur. J. Biochem.* 260:571-595.
- Dawkins, R., and Krebs, J. R. 1979. Arms races between and within species. *Proc. R. Soc. Lond. B. Biol. Sci.* 205:489-511.
- de Haën, C. D., Neurath, H., and Teller, D. C. 1975. Phylogeny of trypsin-related serine proteases and their zymogens—New methods for investigation of distant evolutionary relationships. *J. Mol. Biol.* 92:225-259.
- Dimopoulos, G., Richman, A., Muller, H. M., and Kafatos, F. C. 1997. Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. *Proc. Natl. Acad. Sci. U.S.A.* 94:11508-11513.
- Gayathri, P., Satheskumar, P. S., Prasad, K., Nair, S., Savithri, H. S., and Murthy, M. R. 2006. Crystal structure of the serine protease domain of *Sesbania mosaic virus* polyprotein and mutational analysis of residues forming the S1-binding pocket. *Virology* 346:440-451.
- Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. 2003. 3D-Jury: A simple approach to improve protein structure predictions. *Bioinformatics* 19:1015-1018.
- Göteborg, A., Marshall, J. S., Jones, D. A., and Hardham, A. R. 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.
- Hall, T. A. 1999. Bioedit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symp. Ser.* 41:95-98.
- Ham, K. S., Wu, S. C., Darvill, A. G., and Albersheim, P. 1997. Fungal pathogens secrete an inhibitor protein that distinguishes isoforms of plant pathogenesis-related endo-beta-1,3-glucanases. *Plant J.* 11:169-179.
- Højrup, P., Jensen, M. S., and Petersen, T. E. 1985. Amino acid sequence of bovine protein Z: A vitamin K-dependent serine protease homolog. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 184:333-338.
- Huang, T. S., Wang, H., Lee, S. Y., Johansson, M. W., Soderhall, K., and Cerenius, L. 2000. A cell adhesion protein from the crayfish *Pacifastacus leniusculus*, a serine proteinase homologue similar to *Drosophila* Masquerade. *J. Biol. Chem.* 275:9996-10001.
- Iengar, P., and Ramakrishnan, C. 1999. Knowledge-based modeling of the serine protease triad into non-proteases. *Protein Eng.* 12:649-656.
- Isackson, P. J., and Bradshaw, R. A. 1984. The alpha-subunit of mouse 7 S nerve growth factor is an inactive serine protease. *J. Biol. Chem.* 259:5380-5383.
- Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu. Rev. Phytopathol.* 44:41-60.
- Kanneganti, T.-D., Huitema, E., Cakir, C., and Kamoun, S. 2006. Synergistic interactions of the plant cell death pathways induced by *Phytophthora infestans* Nep1-like protein PiNPP1.1 and INF1 elicitor. *Mol. Plant-Microbe Interact.* 19:854-863.
- Kauffmann, S., Legrand, M., Geoffroy, P., and Fritig, B. 1987. Biological function of pathogenesis-related proteins—Four PR proteins of tobacco have 1,3-beta-glucanase activity. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:3209-3212.
- Krem, M. M., Rose, T., and Di Cera, E. 2000. Sequence determinants of function and evolution in serine proteases. *Trends Cardiovas. Med.* 10:171-176.
- Krem, M. M., Prasad, S., and Di Cera, E. 2002. Ser(214) is crucial for substrate binding to serine proteases. *J. Biol. Chem.* 277:40260-40264.
- Kurosky, A., Barnett, D. R., Lee, T. H., Touchstone, B., Hay, R. E., Arnott, M. S., Bowman, B. H., and Fitch, W. M. 1980. Covalent structure of human haptoglobin: A serine protease homolog. *Proc. Natl. Acad. Sci. U.S.A.* 77:3388-3392.
- Lee, V. D., Finstad, S. L., and Huang, B. 1997. Cloning and characterization of a gene encoding an actin-related protein in *Chlamydomonas*. *Gene* 197:153-159.
- Lin, C. Y., Hu, K. Y., Ho, S. H., and Song, Y. L. 2006. Cloning and characterization of a shrimp clip domain serine protease homolog (c-SPH) as a cell adhesion molecule. *Dev. Comp. Immunol.* 30:1132-1144.
- Liu, Z., Bos, J. I. B., Armstrong, M., Whisson, S. C., da Cunha, L., Torto-Alalibo, T., Win, J., Avrova, A. O., Wright, F., Birch P. R. J., and Kamoun, S. 2005. Patterns of diversifying selection in the phytoalexin-like *scr74* gene family of *Phytophthora infestans*. *Mol. Biol. Evol.* 22:659-672.
- Lohaus, G., Pennewiss, K., Sattelmacher, B., Hussmann, M., and Hermann Muehling, K. 2001. Is the infiltration-centrifugation technique appropriate for the isolation of apoplastic fluid? A critical evaluation with different plant species. *Physiol. Plant* 111:457-465.
- Murugasu-Oei, B., Rodrigues, V., Yang, X., and Chia, W. 1995. Masquerade: A novel secreted serine protease-like molecule is required for somatic muscle attachment in the *Drosophila* embryo. *Genes Dev.* 9:139-154.
- Nagamine-Natsuka, Y., Norioka, S., and Sakiyama, F. 1995. Molecular cloning, nucleotide sequence, and expression of the gene encoding a trypsin-like protease from *Streptomyces erythraeus*. *J. Biochem. (Tokyo)* 118:338-346.
- Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K., and Shimizu, S. 1989. Molecular cloning and expression of human hepatocyte growth factor. *Nature* 342:440-443.
- Panabieres, F., Amselem, J., Galiana, E., and Le Berre, J. Y. 2005. Gene identification in the oomycete pathogen *Phytophthora parasitica* during in vitro vegetative growth through expressed sequence tags. *Fungal Genet. Biol.* 42:611-623.
- Perona, J. J., and Craik, C. S. 1995. Structural basis of substrate-specificity in the serine proteases. *Protein Sci.* 4:337-360.
- Pond, S. L., and Frost, S. D. 2005a. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531-2533.
- Pond, S. L., and Frost, S. D. 2005b. A simple hierarchical approach to modeling distributions of substitution rates. *Mol. Biol. Evol.* 22:223-234.
- Pond, S. L., Frost, S. D., and Muse, S. V. 2005. Hyphy: Hypothesis testing using phylogenies. *Bioinformatics* 21:676-679.
- Radisky, E. S., Lee, J. M., Lu, C. J., and Koshland, D. E., Jr. 2006. Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc. Natl. Acad. Sci. U.S.A.* 103:6835-6840.
- Ronquist, F., and Huelsenbeck, J. D. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
- Rose, J. K. C., Cosgrove, D. J., Albersheim, P., Darvill, A. G., and Bennett, A. B. 2000. Detection of expansin proteins and activity during tomato fruit ontogeny. *Plant Physiol.* 123:1583-1592.
- Rose, J. K. C., Ham, K. S., Darvill, A. G., and Albersheim, P. 2002. Molecular cloning and characterization of glucanase inhibitor proteins: Coevolution of a counterdefense mechanism by plant pathogens. *Plant Cell* 14:1329-1345.
- Ross, J., Jiang, H., Kanost, M. R., and Wang, Y. 2003. Serine proteases and their homologs in the *Drosophila melanogaster* genome: An initial analysis of sequence conservation and phylogenetic relationships. *Gene* 304:117-131.
- Sali, A., and Blundell, T. L. 1993. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779-815.
- Sali, A., Potterton, L., Yuan, F., Vanlijmen, H., and Karplus, M. 1995. Evaluation of comparative protein modeling by modeler. *Proteins* 23:318-326.
- Sanchez, R., and Sali, A. 2000. Comparative protein structure modeling. Introduction and practical examples with Modeller. *Methods Mol. Biol.* 143:97-129.
- Smart, C. D., Myers, K. L., Restrepo, S., Martin, G. B., and Fry, W. E. 2003. Partial resistance of tomato to *Phytophthora infestans* is not dependent upon ethylene, jasmonic acid, or salicylic acid signaling pathways. *Mol. Plant-Microbe Interact.* 16:141-148.
- Smart, C. D., Willmann, M. R., Mayton, H., Mizubuti, E. S. G., Sandrock, R. W., Muldoon, A. E., and Fry, W. E. 1998. Self-fertility in two clonal lineages of *Phytophthora infestans*. *Fungal Genet. Biol.* 25:134-142.
- Stahl, E. A., and Bishop, J. G. 2000. Plant-pathogen arms races at the molecular level. *Curr. Opin. Plant Biol.* 3:299-304.
- Stintzi, A., Heitz, T., Prasad, V., Wiedemannmerdingoglou, S., Kauffmann, S., Geoffroy, P., Legrand, M., and Fritig, B. 1993. Plant pathogenesis-related proteins and their role in defense against pathogens. *Biochimie* 75:687-706.

- Stotz, H. U., Pittendrigh, B. R., Kroymann, J., Weniger, K., Fritsche, J., Bauke, A., and Mitchell-Olds, T. 2000. Induced plant defense responses against chewing insects. Ethylene signaling reduces resistance of Arabidopsis against egyptian cotton worm but not diamondback moth. *Plant Physiol.* 124:1007-1018.
- Suzuki, Y., and Gojobori, T. 1999. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* 16:1315-1328.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. 1997. The Clustal\_X Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876-4882.
- Tiffin, P., and Moeller, D. A. 2006. Molecular evolution of plant immune system genes. *Trends Genet.* 22:662-670.
- Torto, T. A., Rauser, L., and Kamoun, S. 2002. The *pigp1* gene of the oomycete *Phytophthora infestans* encodes a fungal-like endopolygalacturonase. *Curr. Genet.* 40:385-390.
- Tyler, B. M., Tripathy, S., Zhang, X. M., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M. B., Dorrance, A. E., Dou, D. L., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grunwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M. K., McDonald, W. H., Medina, M., Meijer, H. J. G., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K. C., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W. S., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z. Y., Zhang, H. B., Grigoriev, I. V., Rokhsar, D. S., and Boore, J. L. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313:1261-1266.
- van Loon, L. C., Rep, M., and Pieterse, C. M. J. 2006. Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.* 44:135-162.
- van West, P., de Jong, A. J., Judelson, H. S., Emons, A. M. C., and Govers, F. 1998. The *ipiO* gene of *Phytophthora infestans* is highly expressed in invading hyphae during infection. *Fungal Genet. Biol.* 23:126-138.
- Win, J., Kanneganti, T. D., Torto-Alalibo, T., and Kamoun, S. 2006. Computational and comparative analyses of 150 full-length cDNA sequences from the oomycete plant pathogen *Phytophthora infestans*. *Fungal Genet. Biol.* 43:20-33.
- Win, J., Morgan, W., Bos, J., Krasileva, K. V., Cano, L. M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B. J., and Kamoun, S. 2007. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* 19:2349-2369.
- Yang, Z., Wong, W. S. W., and Nielsen, R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22:1107-1118.
- York, W. S., Qin, Q., and Rose, J. K. 2004. Proteinaceous inhibitors of endo-beta-glucanases. *Biochim. Biophys. Acta* 1696:223-233.
- Zou, Z., Lopez, D. L., Kanost, M. R., Evans, J. D., and Jiang, H. B. 2006. Comparative analysis of serine protease-related genes in the honey bee genome: Possible involvement in embryonic development and innate immunity. *Insect Mol. Biol.* 15:603-614.

## AUTHOR RECOMMENDED INTERNET RESOURCES

- Carbohydrate-active enzymes (CAZy) database: [www.cazy.org](http://www.cazy.org)
- United States Department of Energy Joint Genome Institute (JGI) *P. sojae* genome database:  
[genome.jgi-psf.org/Physo1\\_1/Physo1\\_1.home.html](http://genome.jgi-psf.org/Physo1_1/Physo1_1.home.html)
- JGI *P. ramorum* genome database:  
[genome.jgi-psf.org/Phyra1\\_1/Phyra1\\_1.home.html](http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.home.html)
- P. infestans* genome database:  
[www.broad.mit.edu/annotation/genome/phytophthora\\_infestans/Home.html](http://www.broad.mit.edu/annotation/genome/phytophthora_infestans/Home.html)
- Phytophthora* Functional Genomics database: [www.pfgd.org](http://www.pfgd.org)