MPMI Vol. 20, No. 7, 2007, pp. 781–793. doi:10.1094/MPMI-20-7-0781. © 2007 The American Phytopathological Society *e*-Xtra*

Expressed Sequence Tags from *Phytophthora sojae* Reveal Genes Specific to Development and Infection

Trudy A. Torto-Alalibo,[1] Sucheta Tripathy,[1] Brian M. Smith,[1] Felipe D. Arredondo,[1] Lecong Zhou,[1] Hua Li,[1] Marcus C. Chibucos,[1] Dinah Qutob,[2] Mark Gijzen,[2] Chunhong Mao,[1] Bruno W. S. Sobral,[1] Mark E. Waugh,[3] Thomas K. Mitchell,[4] Ralph A. Dean,[4] and Brett M. Tyler[1]

[1]Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, U.S.A.; [2]Agriculture and Agri-Food Canada, 1391 Sandford Street, London, Ontario N5V 4T3, Canada; [3]National Center for Genome Resources, Santa Fe, NM 87505, U.S.A.; [4]Center for Integrated Fungal Research, North Carolina State University, Raleigh, NC 27695, U.S.A.

Submitted 26 April 2006. Accepted 26 January 2007.

Six unique expressed sequence tag (EST) libraries were generated from four developmental stages of *Phytophthora sojae* P6497. RNA was extracted from mycelia, swimming zoospores, germinating cysts, and soybean (*Glycine max* (L.) Merr.) cv. Harosoy tissues heavily infected with *P. sojae*. Three libraries were created from mycelia growing on defined medium, complex medium, and nutrient-limited medium. The 26,943 high-quality sequences obtained clustered into 7,863 unigenes composed of 2,845 contigs and 5,018 singletons. The total number of *P. sojae* unigenes matching sequences in the genome assembly was 7,412 (94%). Of these unigenes, 7,088 (90%) matched gene models predicted from the *P. sojae* sequence assembly, but only 2,047 (26%) matched *P. ramorum* gene models. Analysis of EST frequency from different growth conditions and morphological stages revealed genes that were specific to or highly represented in particular growth conditions and life stages. Additionally, our results indicate that, during infection, the pathogen derives most of its carbon and energy via glycolysis of sugars in the plant. Sequences identified with putative roles in pathogenesis included avirulence homologs possessing the RxLR motif, elicitins, and hydrolytic enzymes. This large collection of *P. sojae* ESTs will serve as a valuable public genomic resource.

Additional keyword: effectors.

The most devastating oomycete phytopathogens include members of the genus *Phytophthora* (Erwin and Ribeiro 1996; Kamoun 2003). Although oomycetes exhibit fungal-like morphology, recent molecular and biochemical research have demonstrated that these organisms are actually more closely related to heterokont algae than they are to fungi (Baldauf et al. 2000). *Phytophthora sojae* (Kauf. & Gerd. (1958)) is the causal agent of stem and root rot in soybean. Annual yield losses were estimated to be 1,149,000 metric tons in the United States (Wrather et al. 2001).

P. sojae is a hemibiotroph with a life cycle consisting of several morphological stages. First, infection of soybean plants often is initiated by zoospores (or in some cases hyphae). Motile zoospores are chemotactically attracted to soybean roots (Carlile 1983; Morris et al. 1998) and, within 3 h of infection, zoospores exude a mucilaginous glycoprotein that enables them to adhere to root surfaces. Encysted zoospores germinate to produce hyphae that penetrate the host tissue. At 6 h postinfection, six to eight cell layers within the cortex are colonized by hyphae that produce haustoria to gain nutrients. By 24 h, hyphae saturate the deep cell layers within the roots, causing the death of many plant cells. At 48 h postinfection, root symptoms appear as large, water-soaked lesions that result in the maceration and collapse of host tissue—at this stage, the pathogen has progressed well into necrotrophy (Bhattacharyya and Ward 1986; Moy et al. 2004; Stossel et al. 1980). The plant's defense responses are activated during infection; however, in the case of a compatible interaction, this response is too weak or too late to restrict the ingress of the pathogen. Under wet conditions, hyphal tips can develop zoosporangia that release zoospores, which can swim through the soil to attack additional roots or new host plants.

In the past decade, genomic resources have been developed for many economically important phytopathogens, including bacterial, fungal, and oomycete species (Buell et al. 2003; Ebbole et al. 2004; Kamoun et al. 1999; Panabieres et al. 2005; Qutob et al. 2000; Randall et al. 2005). Expressed sequence tags (ESTs) are a relatively inexpensive means of gene discovery. Additionally, gene expression profiles can be obtained by analyzing ESTs generated under varying conditions, which makes ESTs resources important for comparative transcriptomics. Investigating the mechanisms plant pathogens use to infect and colonize their hosts can be facilitated by generation of EST from different life stages and especially from infection stages of a pathogen. The number of ESTs generated from various eukaryotic plant pathogens ranges from a few thousand (Keon et al. 2005; Panabieres et al. 2005; Thomas et al. 2001; Trail et al. 2002) to tens of thousands per species (Ebbole et al. 2004; Randall et al. 2005).

Corresponding author: B. M. Tyler; E-mail: bmtyler@vt.edu

Current address of M. E. Waugh: CIO Program Office, Los Alamos National Laboratory, Bikini Atoll Rd., SM 30, Los Alamos, NM 87545, U.S.A.

The sequence data presented in this study are available at GenBank (accession numbers CF839248 to CF865335) and in the Virginia Bioinformatics Institute Microbial Database.

*The *e*-Xtra logo stands for "electronic extra" and indicates that three supplemental tables not included in the print edition are available online.

We describe here a data set comprising 26,943 *P. sojae* EST representing 7,863 unique sequences from six different libraries. The ESTs represent four developmental stages of *P. sojae*, including swimming zoospores, germinating cysts, and mycelia, from infection sites and from axenically grown cultures. Analysis of these new EST in conjunction with the recently released draft genome of *P. sojae* (Tyler et al. 2006) offers insight into the repertoire of mechanisms employed by this pathogen during infection and colonization of its host.

RESULTS

EST library construction, sequencing, and assembly.

The ESTs were produced from six different libraries (Table 1). The libraries represent swimming zoospores, germinating cysts, mycelium grown in defined medium, nutrient-limited medium, complex medium, and *P. sojae* mycelia and other structures interacting with soybean. Of the 31,314 raw ESTs generated, including 3,035 reported by Qutob and associates (2000), 27,088 were considered high quality (99%) using Phred (Ewing et al. 1998), a base-calling program that assigns quality values to the bases. After further cleanup to remove vector sequence contaminants and chimeras, 26,943 ESTs remained. The combined EST set from all six libraries clustered into 2,845 contigs and 5,018 singletons, yielding a total set of 7,863 unigenes (Table 1).

Genes highly represented in or specific to particular libraries.

ESTs from individual libraries were assembled into contigs to determine the redundancy of tissue-specific transcripts (Table 2). Overall, the number of ESTs specific to particular libraries ranged from as low as 213 in the germinating cyst library to as high as 3,223 in the infection library. In addition, many ESTs appeared to be more abundant in some libraries than others. To evaluate the statistical significance of these abundance differences, we calculated log likelihood ratios (Stekel et al. 2000) and used false discovery rate control (FDR) for multiple test correction (Benjamini and Hochberg 1995). Using an FDR-adjusted *P* value of 0.05 as a cut-off, we identified four broad patterns representing genes that are significantly differentially represented among the six libraries (Fig. 1). The largest cluster, representing 16 genes, corresponded to ESTs that are abundant in the swimming zoospore library, and the second-largest cluster, with 10 genes, was most highly represented in the infection library. Two additional clusters, composed of seven and eight genes, were represented preferentially in mycelial tissues grown in limited and defined media, respectively. We further compared individual libraries with the rest of the libraries pooled together to identify unigenes that were found exclusively in individual libraries. The unigenes found exclusively in the swimming zoospores (27 genes) and infection libraries (46 genes) are listed in Tables 3 and 4, respectively.

Similarities of expression profiles of the libraries.

In order to obtain an overall estimate of the physiological similarity of the *P. sojae* tissues sampled for the different libraries, we carried out correlation analyses of the EST representations between pairs of libraries. One specific purpose was to determine which of the in vitro growth conditions most closely resembled growth inside the plant during infection. Correlation analyses were carried out for all ESTs for which

Table 1. Description of libraries

Library ID	Tissue	Total no. of ESTs[a]		Source
		Sublibrary[b]	Library	
psZO	Swimming zoospores	1,019	…	Qutob et al 2000
psZS	Swimming zoospores	3,484	…	This report
psZO/psZS	Total swimming zoospores	…	4,503	…
psZG	Germinating cyst	…	2,308	This report
psMY	Mycelium in defined medium	992	…	Qutob et al 2000
psMA	Mycelium in defined medium	4,034	…	This report
psMY/psMA	Total mycelium in defined medium	…	5,026	…
psML	Mycelium in nutrient-limited medium	…	2,947	This report
psMC	Mycelium in complex medium	…	2,915	This report
psHA	Infected soybean hypocotyl	789	…	Qutob et al 2000
psHB	Infected soybean hypocotyl	8,455	…	This report
psHA/psHB	Total infected soybean hypocotyl	…	9,244	…

[a] ESTs = expressed sequence tags.
[b] The psZS, psMA, and psHB ESTs were obtained from the same cDNA libraries from which the psZO, psMY, and psHA sequences, respectively, were generated,

Table 2. Distribution of expressed sequence tags (ESTs) representing sequences specific to libraries or groups of libraries

Library	Specific ESTs			Specific ESTs in each library (%)[d]	Specific unigenes[e]
	In singlets[a]	In contigs[b]	Total[c]		
psZO/psZS	1,256	782	2,038	45	1,508
psZG	191	22	213	9	202
psZO/psZS + psZG	1,447	804	2,251	33	1,710
psMY/psMA	1,035	307	1,342	27	1,161
psMC	324	54	378	12	349
psML	268	67	335	11	297
psMY/psMA + psMC + psML	1,627	428	2,055	19	1,807
psHA/psHB	1,910	1,313	3,223	35	2,350

[a] Singlet ESTs in each library or collection of libraries; by definition these are specific to each library.
[b] Number of ESTs from a particular library that form unigenes that do not overlap with ESTs from any other library in this study.
[c] Sum of specific ESTs in singlets and contigs defining the specific unigene set.
[d] Percentage is calculated as the total number of specific ESTs relative to the total number of ESTs in each library.
[e] Total number of unigenes, including singletons, found only in the given library or collection of libraries.

four or more sequences were obtained from any library. Both the pairwise correlations and a similarity diagram that summarizes the relationships of the libraries to one another based on their EST distributions are shown in Figure 2. Not surprisingly, the library from free-swimming zoospores (ZO/ZS) showed the lowest correlation to any other library: 0.50 to the infection library and 0.46 to the germinating cyst (ZG) library. The most closely related libraries were those from mycelia grown in complex medium (MC) or dilute complex medium (ML) and zoospores germinating in complex medium (ZG). The correla-

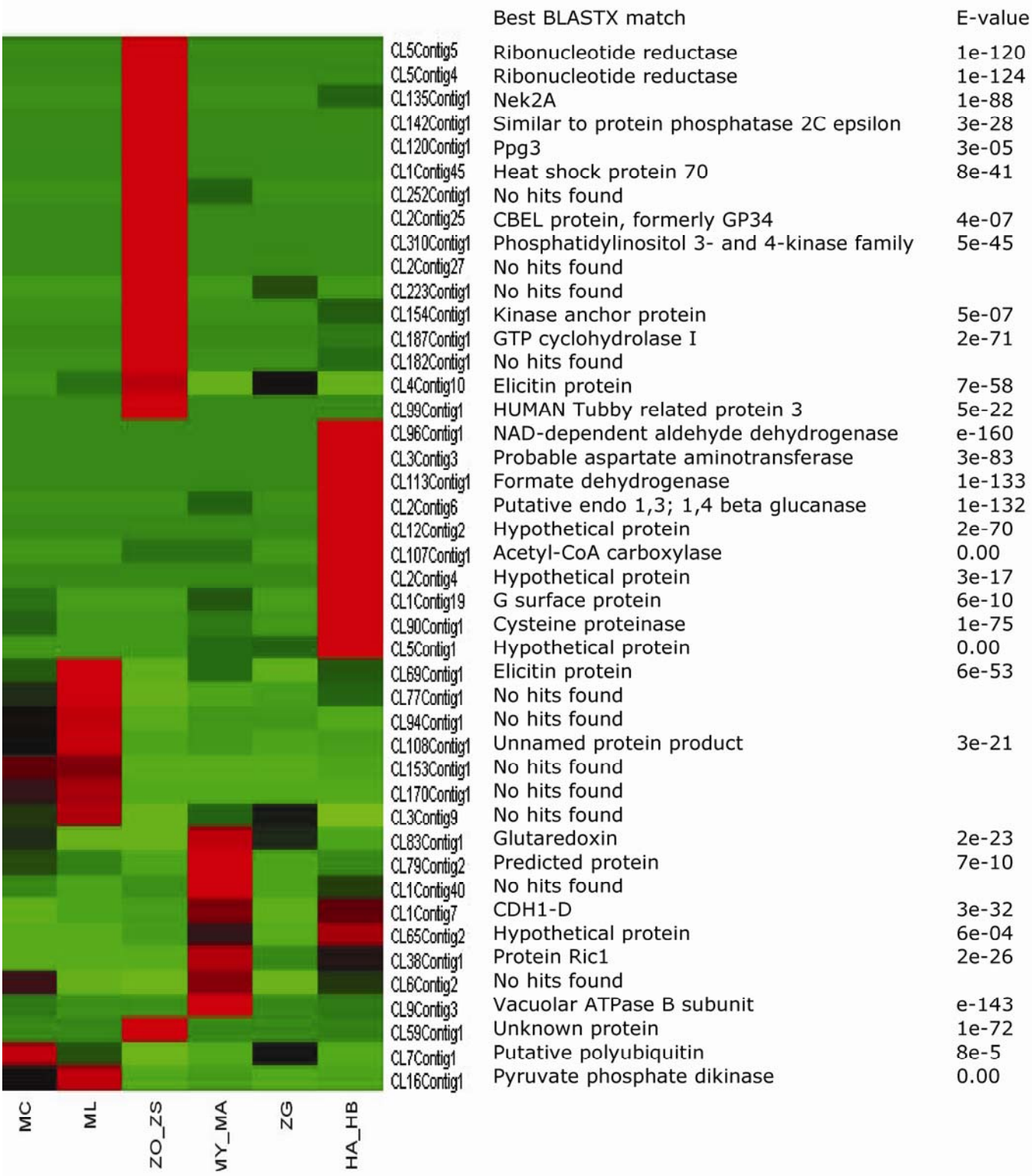

Fig. 1. Clustering analysis of differentially expressed transcripts. Contigs with five or more genes were selected and statistically proven to be differentially represented amongst the different libraries at an adjusted false discovery rate control *P* value of 0.05. We then used a distance matrix method incorporated in the heat map function of the Bioconductor package to generate the heat map. A frequency of zero is shown in bright green, and an increase in frequency is shown in a range from a darker shade of green to a bright shade of red. The description of the libraries is as follows: MC: mycelium in complex medium; ML: mycelium in limited medium, ZO_ZS: swimming zoospores, MY_MA: mycelium in defined medium, ZG: germinating zoospore, and HA_HB: infection library.

tions for the three pairs of libraries varied from 0.64 to 0.81. The infection library (HA/HB) and the library from mycelia grown in defined medium (MY/MA) also showed a relatively high correlation with each other, of 0.64, indicating that growth in planta resembled physiological growth on defined medium more closely than growth on complex medium derived from plant products (V8 juice, soybean broth, and so on). One specific example of the difference between growth on complex medium on the one hand and growth on defined medium or during infection on the other hand was the distribution of ESTs reflective of the flux of metabolites along the glycolysis/gluconeogenesis pathway (Table 5). In complex medium, ESTs representing the gluconeogenic enzymes pyruvate phosphate dikinase and phosphoenolpyruvate carboxykinase were elevated in the complex medium libraries and lowered in defined medium and infection libraries. In contrast, ESTs representing the glycolytic enzyme pyruvate dehydrogenase were elevated in defined medium and infection libraries but lowered in the complex medium libraries.

Representation of unigenes in defined functional categories.

Overall, 55% of the unigenes showed homology to previously described genes in GenBank using a BLASTX cut-off of $\leq 1e^{-4}$. CL2Contig19 with significant sequence similarity to translation elongation factor 1 α (EF-1 α) was identified as having the most ESTs (261) in one contig (Table 6). To group unigenes by putative functions, we used the gene ontology (GO) classification scheme (Ashburner et al. 2000; The Gene Ontology Consortium 2006). The GO provides a valuable resource that utilizes controlled vocabulary to describe molecular functions, biological processes, and cellular locations of gene products across numerous genomes. To assign GO terms to the unigenes, the program InterProScan (Zdobnov and Apweiler 2001) was used to identify protein domains that matched regions in the unigenes. Then, mappings of the InterPro domains to terms in the GO hierarchy were used to assign GO terms to the unigenes. Of the total number of unigenes (7,863), 1,947 were assigned protein domains after analysis with the InterPro tools. The assigned domains mapped to 2,277 terms in the GO biological process hierarchy (Fig. 3A). Furthermore, the protein domains of 973 unigenes mapped to 1,063 terms describing the cellular localization of the respective proteins (Fig. 3B). Disproportionate mappings result from protein products that may have more than one described function. Of the 2,277 terms mapped to the biological process ontology, 1,207 (54%) of the mapped terms were associated with metabolism, which included protein metabolism (17%), amino acid metabolism (6%), and carbohydrate metabolism (5%) (Fig. 3A). Terms matching to unigenes involved in transport formed the next largest category (24%) (Fig. 3A). In this category, 559 terms mapped to different forms of transport. This included 516 unigenes which mapped to single terms and 21 other unigenes which mapped to multiple terms. In all, 109 unigenes mapped to 116 terms representing terms associated with signal transduction category. This represents 5% of the total number of mappings to the biological process ontology. The cellular component mappings (1,063) were represented in eight categories (Fig. 3B), many of which were associated with membrane localization (37%).

Putative secreted proteins and pathogenicity genes.

The secretome forms an important part of the genome of most plant pathogens and has been shown to contain numerous

Table 3. Unigenes specific to the zoospore library and a selection significantly represented in *Phytophthora infestans* zoospore library

				Frequency per 1,000 clones[a]		
				P. sojae[b]	*P. infestans*[c]	
Unigene ID	Best BLASTX match[d]	Accession no.	Species	Zoospore	Zoospore	Others
CL608Contig1	Unknown protein	AAO24656	*P. sojae*	1.11	0.64	0.08
CL5Contig4	Ribonucleotide reductase M2	NP_575124	*Danio rerio*	3.33	3.34	0.24
CL5Contig5	Ribonucleotide reductase 2	AAL05057	*Aedes aegypti*	3.33	3.34	0.24
CL310Contig1	Phosphatidylinositol 3- and 4-kinase family	XP_483081	*Oryza sativum*	1.99	0.39	0.13
CL2Contig27	No hits found	...	...	1.99	2.06	0.12
CL505Contig1	Hypothetical protein	XP_637949	*Dictyostelium discoideum*	1.33	0.77	0.07
CL553Contig1	Hypothetical protein	XP_652966	*Entamoeba histolytica*	1.11	0.26	0.01
CL557Contig1	No hits found	...	...	1.11	0.39	0.01
CL120Contig1	Hypothetical protein	CAE75525	*Caenorhabditis briggsae*	5.10	0.00	0.01
CL361Contig1	Unknown protein	AAO24654	*P. sojae*	1.77	0.00	0.00
CL396Contig1	No hits found	...	...	1.55	0.00	0.00
CL482Contig1	TSP1 domain-containing protein	AAX84973	*P. cinnamomi*	1.33	0.00	0.01
CL99Contig1	Tubby related protein 3	O75386	*Homo sapiens*	6.44	0.13	0.03
CL1Contig45	Pleiotropic drug resistance transport	AAT85568	*P. sojae*	4.44	0.27	0.03
CL142Contig1	Similar to protein phosphatase 2C epsilon	XP_850909	*Canis familiaris*	3.77	0.13	0.01
CL2Contig25	CBEL protein, formerly GP34	CAA65843	*P. parasitica*	2.22	0.00	0.12
CL330Contig1	Putative mitochondrial import inner membrane transport	XP_570062	*Cryptococcus neoformans*	1.33	0.39	0.08
CL412Contig1	No hits found	...	...	1.55	0.00	0.07
CL407Contig1	Hypothetical protein	XP_663405	*Aspergilus nidulans*	1.55	0.13	0.00
CL485Contig1	Hypothetical protein	XP_957315	*Neurospora crassa*	1.33	0.00	0.01
CL503Contig1	Cyclin	CAB77269.1	*Pisum sativum*	1.33	0.00	0.01
CL617Contig1	OJ000114_01.8	XP_472605	*Oryza sativum*	1.11	0.12	0.00
CL663Contig1	Amine oxidase	YP_702571	*Rhodococcus sp.*	1.11	0.00	0.00
CL536Contig1	Histone H4	Q9U7D0	*Mastigamoeba balamuthi*	1.11	0.00	0.03
CL541Contig1	No hits found	...	...	1.11	0.00	0.00
CL559Contig1	No hits found	...	...	1.11	0.00	0.00
CL532Contig1	Hypothetical protein	CAJ20325	*Toxoplasma gondii*	1.11	0.00	0.01

[a] Frequency is the number of expressed sequence tags per 1,000 in the libraries. False discovery rate control-adjusted P value ≤ 0.05.
[b] All *P. sojae* unigenes exclusively represented in the *P. sojae* zoospore library.
[c] All specific *P. sojae* unigenes significantly represented (frequency of 0.26 to 3.34) in the *P. infestans* zoospore library compared with other *P. infestans* libraries.
[d] E value of best BlastX match is $\leq 1e^{-4}$.

pathogenicity effector proteins (Collmer et al. 2000; Dean et al. 2005; Kamoun 2006; Qutob et al. 2002; Torto et al. 2003). Analysis of the unigene set with the SignalP software (Bendtsen et al. 2004) indicated that, in all, 635 (8%) unigenes encode putative secreted proteins. Correspondingly, the number of ESTs represented in the putative set of unigenes with signal peptides is 1,669 (6%). Within this set of 635 unigenes is a subset of 51 genes (8%) with significant similarity to characterized pathogenicity genes from other organisms. Putative extracellular proteins encoded by the unigenes included hydrolases (Gotesson et al. 2002; McLeod et al. 2002; Torto et al. 2002), elicitins (Jiang et al. 2006; Kamoun et al. 1997), and crinkling and necrosis-inducing (*crn*; also called "crinkler") proteins (Torto et al. 2003).

Plant pathogens employ diverse strategies to infect their host. ESTs encoding potential pathogenicity proteins, placed into eight major categories according to a scheme modified from that of Idnurm and Howlett (2001) and Panabieres and associates (2005), are shown in Table 7. Identification of potential pathogenicity genes was based on significant similarities (BLASTX cut-off of $\leq 1e^{-4}$) to characterize pathogenicity genes from other organisms. Among the eight categories, effectors formed the largest group (n = 63) and, within this group, unigenes encoding crinklers formed the majority (n = 25). Also included in this category were 13 unigenes, most of which had no significant matches to sequences in public databases, but matched sequences recently identified as having RxLR and dEER motifs (Tyler et al. 2006). Genes encoding hydrolytic enzymes constituted the second-largest group (n = 43) and, within this group, glucanases formed the majority (n = 14). Unigenes associated with transport, including gene products potentially involved in detoxification, drug resistance, and metabolite transporters, formed the third-largest group (n = 30) and its most abundant unigene set was represented by ABC transporters (n = 16). The fourth-largest group is characterized by 27 unigenes representing proteins involved in protection against oxidative stress.

Gene families encoding potential pathogenicity proteins.

Several large gene families encoding pathogenicity genes have been identified in fungal and oomycete pathogens and are presumed to result from expansion and diversification of the genes in response to selection pressure from the defense systems of host plants. Such families include those encoding elicitins in *Phytophthora* spp., CRN proteins in *P. infestans*, and endopolygalcturonases in *Botrytis cinerea* and *P. cinnamomi*. In this study, a number of gene families were identified among the putative pathogenicity genes, ranging in membership from 2 to 25. A family encoding *crn*-like genes had the highest number of members and currently has been found only in *Phy-*

Table 4. Selected set of unigenes specific to the infection library and a subset highly represented in the *Phytophthora infestans* infection libraries

				Frequency per 1,000 clones[a]				
				P. sojae[b]	*P. infestans*[c]			
Unigene ID	Best BLASTX match[d]	Access. no.[e]	Species	Inf	Preinf	Inf	Starved	Others
CL96Contig1	Aldehyde dehydrogenase (NAD+)	ZP_00518574	*Crocosphaera watsonii*	2.16	0.00	3.04	0.29	0.07
CL585Contig1	Aldehyde dehydrogenase	ABE82378	*Medicago truncatula*	0.54	0.00	3.04	0.29	0.07
CL366Contig1	Putative ABC transporter	BAF01774	*Arabidopsis thaliana*	0.87	0.00	1.53	0.15	0.00
CL169Contig1	Acyl co-enzyme A dehydrogenase	XP_421806	*Gallus gallus*	1.62	0.25	1.52	0.87	0.66
CL633Contig1	Solute carrier family	NP_598656.1	*Mus musculus*	0.54	0.00	0.85	0.43	0.27
CL220Contig1	Hypothetical protein	XP_826856	*Trypanosoma brucei*	1.30	0.00	0.31	0.00	0.00
CL113Contig1	Formate dehydrogenase	BAB69476.1	*Mycobacterium vaccae*	2.27	1.85	0.00	0.87	0.66
CL452Contig1	Hypothetical protein	XP_326952.1	*Neurospora crassa*	0.64	0.24	0.00	0.58	0.00
CL2Contig4	Hypothetical protein	EAA16547.1	*Plasmodium yeolii*	2.90	0.25	0.00	0.44	0.00
CL203Contig2	No hits found	…	…	0.54	0.24	0.00	0.29	0.25
CL549Contig1	Cathepsin Z	NP_491023.1	*Caenorhabditis elegans*	0.54	0.24	0.00	0.29	0.25
CL12Contig2	Hypothetical protein	XP_329448.1	*Neurospora crassa*	2.49	0.00	0.00	0.15	1.53
CL203Contig1	No hits found	…	…	0.54	0.13	0.00	0.37	0.00
CL486Contig1	Acetyl/Propionyl-CoA carboxylase	ZP_00082145.1	*Geobacter metallireducens*	0.64	0.00	0.00	0.29	0.00
CL545Contig1	Subtilisin-like serine proteinase precursor	AAK39096.1	*Aphanomyces astaci*	0.54	0.00	0.00	0.15	0.00
CL670Contig1	CGI-77 protein	NP_057107.1	*Homo sapiens*	0.54	0.25	0.00	0.15	0.49
CL309Contig1	Dihydrodipicolinate synthase	AAN31469.1	*Phytophthora infestans*	0.97	0.13	0.00	0.15	0.25
CL341Contig1	Propionyl-Coenzyme A carboxylase,	NP_000273.1	*Homo sapiens*	0.87	0.00	0.00	0.15	0.00
CL254Contig1	Acetyl-coenzyme A synthetase	AAO50927.1	*Dictyostelium discoideum*	1.19	0.00	0.00	0.15	0.00
CL308Contig1	Acetyl-coA acetyltransferase	596686.1	*Schizosaccharomyces pombe*	0.97	0.00	0.00	0.15	0.00
CL430Contig1	No hits found	…	…	0.54	0.00	0.00	0.15	0.07
CL435Contig1	Hypohetical protein	NP_92752	*Photorhabdus luminescens*	0.64	0.00	0.00	0.14	0.00
CL473Contig1	Similar to chorein isoform	XP_343419.1	*Rattus norvegicus*	0.64	0.00	0.00	0.14	0.00
CL640Contig1	Cellular apoptosis susceptibility protein	AAN52370.1	*Oreochromis niloticus*	0.54	0.00	0.00	0.14	0.00
CL183Contig1	No hits found	…	…	1.51	0.00	0.00	0.12	0.00
CL427Contig1	No hits found	…	…	0.64	0.00	0.00	0.12	0.00
CL269Contig1	Elicitin-like INF7	AAL16014.1	*Phytophthora infestans*	1.08	0.13	0.00	0.00	0.44
CL205Contig1	DICDI RAC-family serine/threonine kinase homolog	P54644	*Dictyostelium discoideum*	1.19	0.00	0.00	0.00	0.07
CL2Contig10	OSJNBb0079B02.6	CAD41847.1	*Oryza sativa*	0.97	0.00	0.00	0.00	0.07
CL409Contig1	No hits found	…	…	0.76	0.00	0.00	0.00	0.00

[a] Frequency is the number of expressed sequence tags per 1,000 in the libraries. False discovery rate control-adjusted P value ≤ 0.05. Inf = infection and Preinf = preinfection.

[b] All unigenes exclusively represented in the *P. sojae* infection library.

[c] All specific *P. sojae* unigenes significantly represented (frequency of 0.31 to 3.04) in the *P. infestans* infection libraries compared with other *P. infestans* libraries.

[d] E value of best BlastX match is $\leq 1e^{-4}$.

[e] Accession number of best BlastX match.

tophthora spp. In comparison, 40 *crn*-like genes were reported in the *P. sojae* genome sequence (Tyler et al. 2006). The ABC transporters formed another family of genes with a membership of 22 unigenes, including PDR-like ABC transporters (6 unigenes). Unigenes encoding elicitins and elicitin-like genes (Jiang et al. 2006) were the next most common. Elicitins have been found so far only in *Phytophthora* and a few *Pythium* spp. and, thus, may be considered oomycete specific. Other gene families represented were cytochrome 450 ($n = 5$), mucin ($n = 5$), CBEL ($n = 5$), and cathepsin ($n = 2$) (Table 7).

Comparative analysis of developmentally specific unigenes in *P. sojae* and other *Phytophthora* sequences.

To ascertain whether unigenes identified exclusively in the swimming zoospore and infection libraries also are found in other *Phytophthora* spp., we performed a similarity search of the unigenes to the *P. ramorum* genome sequence and to the preliminary assemblies of *P. infestans* and *P. capsici* genomic sequences using the TBLASTN and BLASTX algorithms with a cut-off E-value of $1e^{-20}$. Of the 27 unigenes identified exclusively in the zoospore library, 6 had no significant matches to sequences in the *P. capsici* genome, 3 had no significant matches to the *P. infestans* genome sequences, and only 1 had no significant matches to the *P. ramorum* genome sequences. Of the 46 exclusive sequences found in the infection library, 2 were discounted from further study because they did not show clear matches to the *P. sojae* genome sequences. Of the 44 *P. sojae* infection-specific sequences used for the similarity searches, 2 did not significantly match sequences in the *P. capsici* genomic sequences and 2 did not match *P. infestans* sequences. The larger number of genes absent from the *P. infestans* and *P. capsici* assemblies likely reflects the incompleteness of these sequences at the current time.

The availability of a large EST dataset generated from a broad range of growth and stress conditions as well as developmental stages of *P. infestans* made it possible to compare the distribution of unigenes across libraries between *P. sojae* and *P. infestans*. We used the TBLASTN algorithm to find matches

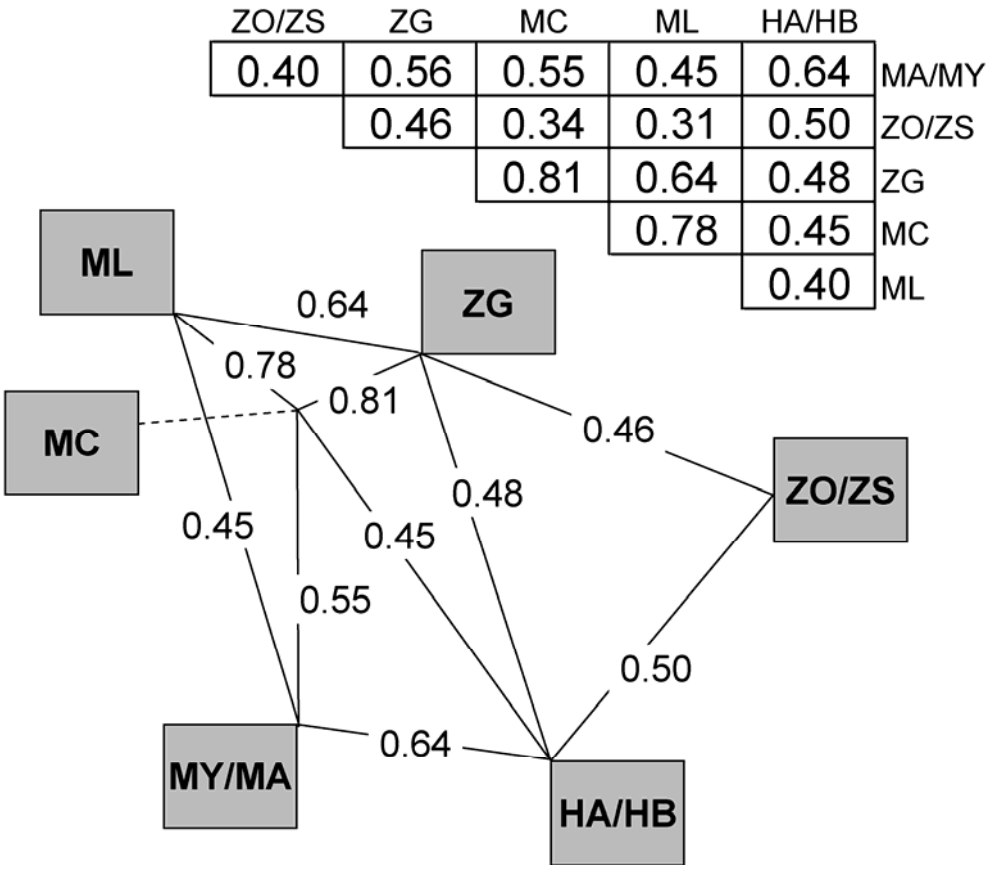

Fig. 2. Correlations of expressed sequence tag representations in the six libraries. Pearson's correlation coefficient *r* is shown. In the schematic illustration of library relationships, libraries were placed according to correlations with the two most similar libraries. Lengths of lines are proportional to $1 - r$.

Table 5. Representation of expressed sequence tags (ESTs) specific for glycolysis and gluconeogenesis[a]

	MY/MA		HA/HB		MC+ML	
	Defined medium		Infection		Complex medium	
Enzyme	**Number**	**No./1,000**	**Number**	**No./1,000**	**Number**	**No./1,000**
Pyruvate phosphate dikinase (gluconeogenesis)[b]	3	0.75	9	1.2	125	47
Phosphoenolpyruvate carboxykinase (gluconeogenesis)[c]	4	1.0	8	1.1	8	3.0
Pyruvate dehydrogenase (glycolysis)[d]	16	4.0	5.2	7.1	1	0.4

[a] Number = total number of ESTs, including those representing all enzyme subunits, and No./1,000 = number of ESTs per 1,000 in the libraries.
[b] EST contigs CL16Contig1, CL1993Contig1, CL1508Contig1, and CL398Contig1.
[c] EST contigs CL480Contig1, CL194Contig2, CL194Contig1, and CL65Contig1.
[d] EST contigs CL95Contig1, CL89Contig1, CL184Contig2, and CL666Contig1.

of the swimming zoospore and infection-specific unigenes from *P. sojae* in the *P. infestans* unigene set. We then looked at the EST distribution in the different libraries corresponding to each of the unique genes. Overall, 10 *P. sojae* unigenes specific to the swimming zoospore library were highly represented in the *P. infestans* zoospore library (Table 3). These included unigenes with sequence similarities to two ribonucleotide reductases, phosphatidylinositol 3- and 4-kinase, two hypothetical proteins, and three unknown proteins. However, only six *P. sojae* infection-specific unigenes were significantly represented in the *P. infestans* infection libraries. These included two distinct aldehyde dehydrogenases and a unigene encoding an ABC transporter. One of the *P. sojae* infection-specific unigenes encoding formate dehydrogenase was highly represented in the *P. infestans* preinfection (zoospore, sporangia, and germinating cyst) libraries (Table 4).

DISCUSSION

As part of our efforts to understand the molecular mechanisms underlying pathogenicity in *P. sojae,* we analyzed ESTs generated from different life stages of the pathogen. After cleaning and assembly of the EST data, we obtained a set of 7,863 unigenes. An earlier pilot-scale EST study of *P. sojae* generated approximately 3,000 ESTs and 2,189 unigenes (Qutob et al. 2000). Those libraries were derived from swimming zoospores, mycelia on defined medium, and soybean hypocotyls infected with *P. sojae*. The current study represents a much deeper EST sampling of this organism, including more ESTs from the three aforementioned cDNA libraries, as well as sequences from germinated cysts and from mycelia grown in limited and complex media (Table 1). The *P. sojae* ESTs add to the accumulating genomic resources available for *Phytophthora* spp. Presently, EST

Table 6. Expressed sequence tags (ESTs) occurring most frequently in *Phytophthora sojae* EST libraries[a]

Unigene ID	No. of ESTs	GenBank accession no.	Best BLASTX match	Species	E value
CL2Contig19	261	ABA00716	Translation elongation factor 1 α	*Phytophthora parasitica*	0
CL8Contig2	121	XP_750152	Ubiquitin fusion protein	*Aspergilus fumigatus*	5e-58
CL2Contig2	113	AAG01044	Actin	*Pythium splendens*	0
CL6Contig2	108	…	No hits found	…	…
CL16Contig1	98	AAK74148	Pyruvate phosphate dikinase	*Phytophthora cinnamomi*	0
CL38Contig1	86	Q9Y068	Protein Ric1	*P. infestans*	4e-26
CL9Contig3	76	…	No hits found	…	…
CL6Contig4	74	…	No hits found	…	…
CL45Contig1	73	AAM90674	Heat shock protein Hsp90	*Achlya ambisexualis*	0
CL1Contig7	67	AAL31950	CDH1-D (cadherin)	*Gallus gallus*	5e-32
CL1Contig14	61	ABC59142	Annexin	*Saprolegnia monoica*	5e-25
CL56Contig1	62	AAN31465	14-3-3-like protein	*P. infestans*	4e-138
CL37Contig2	61	CAE02348	Putative hydroxyproline rich glycoprotein	*Oryza sativa*	4e-24
CL59Contig1	61	AAO24658	Unknown protein	*P. sojae*	5e-69
CL57Contig1	60	…	No hits found	…	…

[a] This list excludes ribosomal proteins.

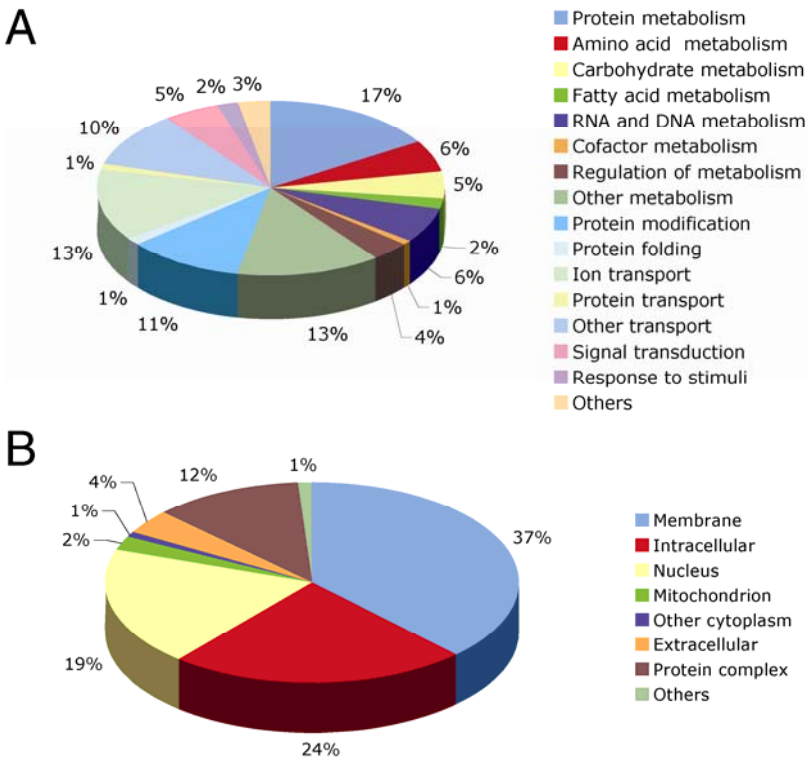

Fig. 3. Classification of unigenes based on gene ontology (GO) mappings for *Phytophthora sojae* unigenes obtained from InterProScan. **A,** Biological process: 1,947 unigenes of the total set of 7,863 unigenes from the *P. sojae* expressed sequence tag dataset were assigned protein domain which mapped to 2,277 terms in the GO biological process hierarchy. **B,** Cellular component: 973 unigenes mapped to 1,063 terms describing the cellular localization of the respective genes.

datasets have been generated for other *Phytophthora* spp., including *P. infestans* (18,256 unigenes) (Randall et al. 2005) and *P. parasitica* (2,269 unigenes) (Panabieres et al. 2005).

Many of the unigenes found (45%) did not show similarity to known genes in publicly accessible sequence databases. By comparison, 79% of the genes predicted from the genome sequences of *P. sojae* and *P. ramorum* lacked similarity to known genes (Tyler et al. 2006). Similarly, in a recent study of ESTs generated from three independent phases of the wheat leaf blotch pathogen, *Mycosphaerella graminicola,* 33% of these sequences had no known functions based on similarity searches done against GenBank and the phytopathogens EST database (Keon et al. 2005). These orphan sequences may well represent sequences unique to the pathogen and may be useful as targets for chemical control.

Table 7. Unigenes corresponding to different categories of putative pathogenicity genes[a]

Main category, putative pathogenicity	Number of[b] Unigenes	EST[c]
Adhesion	***10***	***66***
CBEL protein	5	43
Mucin-like protein	5	23
Hydrolases	***43***	***202***
Serine protease	3	8
Aspartyl proteinase	1	1
Cysteine protease/proteinase	5	41
Serine carboxypeptidase	1	1
Ubiquitin protease	1	1
Cutinase	1	2
Esterase	3	5
Lyases	4	6
Glucosidases	8	16
Glucanase	14	115
Polygalacturonase	2	4
Cellulase	1	4
Inhibitors of proteases	***6***	***8***
Protease inhibitor	6	8
Effectors	***63***	***404***
Unigenes matching sequences with RXLR motif	13	32
Elicitins and elicitin-like	16	210
Crinkling and necrosis-inducing protein	25	130
Necrosis-inducing peptide	1	1
Others	8	31
Small cysteine-rich proteins	***3***	***9***
Small cysteine-rich protein	3	9
Detoxification, drug resistance, and metabolite transport	***30***	***80***
Cytochrome P450	5	7
Putative ABC transporter	16	42
PDR-like ABC transporter	6	20
Bialaphos acetylhydrolase	1	4
Cathepsin Z	2	7
Protection against oxidative stress	***27***	***251***
Glutathione s-transferase	8	34
Peroxidases	9	57
Reductases	5	89
Superoxide dismutase	2	31
Glutaredoxin	2	39
Glutathione synthetase	1	1
Signal transduction and regulation	***7***	***48***
Mitogen-activated protein kinases	3	4
Myb-related transcription factor	3	6
G surface protein	1	38

[a] Defined categories modified from Idnurm and Howlett (2001) and Panabieres and associates (2005).
[b] The numbers in bold and italic are category totals.
[c] No. of expressed sequence tags (EST) represented in unigenes.

The controlled vocabulary proposed by the GO consortium to describe the function of a gene, the process in which it is involved, and to what cellular component it is targeted aid in comparative genome studies among species. The GO classification scheme also is useful for comparing global gene expression profiles in different organisms (Ashburner et al. 2000; Clark et al. 2005; The Gene Ontology Consortium 2006). We used GO to define categories into which functions of the genes represented in the *P. sojae* transcriptome were grouped. Unigenes involved in metabolism were the most represented in the functional classification. This corroborates the findings of EST analyses for other eukaryotic pathogens, including *Phytophthora* spp. (Panabieres et al. 2005; Randall et al. 2005).

Pathogenicity genes are of great importance in studying pathogens of economic importance because they can increase our overall knowledge of disease processes and can be targeted for use in disease control. Most bacterial phytopathogens manipulate their hosts by delivering effector proteins. These effectors may help to overcome host defenses and aid in the ingress of the pathogen into the plant; or, alternatively, the effectors may trigger defense responses in the plant and lead to host resistance (Collmer et al. 2000; Knogge 1996; Staskawicz et al. 2001). Secretion serves as an important mechanism for delivery of pathogenicity factors into host tissue. Thus, many as yet uncharacterized genes identified in the secretomes of pathogens may possess functions in virulence. There is growing evidence that phytopathogenic oomycetes produce secreted effector proteins to manipulate their hosts (Huitema et al. 2004; Kamoun 2006). Of the total number of unigenes in the *P. sojae* EST dataset, 635 (8%) encoded putative secreted proteins. This corresponds to 1,669 (6%) of all the ESTs. Among different libraries, the highest frequencies of ESTs encoding predicted secreted proteins was observed in the infection library (5.2%) and the library from mycelium grown on defined medium (5.1%); whereas, in libraries from mycelium grown on complex medium, the frequency was only 3.3%. Of the putative pathogenicity genes, 8% were determined to be putative secreted proteins.

Putative pathogenicity genes could be assigned to eight categories. Effectors formed the largest group. A subset of effectors represents sequences with significant similarity to avirulence homologs (Avh) possessing the RxLR and dEER motifs. Four avirulence genes have been identified from oomycetes in recent years (Allen et al. 2004; Armstrong et al. 2005; Rehmany et al. 2005; Shan et al. 2004). These avirulence gene products all possess an RxLR motif, and three contain a dEER motif in their sequences. A motif similar to the RxLR had been characterized previously in the malarial parasite *Plasmodium falcipurium* and was shown to be involved in the transport of proteins into the cytoplasm of human erythrocytes (Bhattacharjee et al. 2006; Hiller et al. 2004; Marti et al. 2004). Thus, it is speculated that the RxLR motifs in oomycete effectors may be involved in transporting these proteins into the host cytoplasm. The RxLR and dEER motifs also were identified in 700 sequences mined from the recently released *Phytophthora sojae* and *P. ramorum* genomic sequences and named *Avh* genes (Tyler et al. 2006). Thirteen unigenes had strong matches to *Avh* genes.

Multigene families are prevalent in *Phytophthora* spp., and their presence in plant pathogens such as *Phytophthora* spp. may reflect diversification of gene function driven by selection pressure from the plant defense responses (Jiang et al. 2006). ABC transporters have been found in many eukaryotic organisms, including plants, animals, and fungi (Bauer et al. 1999; Dassa and Bouige 2001; Van den Brule and Smart 2002), and their role in mediating multidrug resistance to chemotherapeutic reagents has been documented in bacterial, protozoan, and fungal pathogens of animals (Davidson and Chen 2004; Klokouzas

et al. 2003; Litman et al. 2001; Morschhauser 2002). Furthermore, in the interaction between fungal pathogens and their host, ABC transporters foster the protection of the pathogen from fungicides and other toxins produced by the plant (Del Sorbo et al. 2000; Fleissner et al. 2002; Hayashi et al. 2001, 2002). Thus, the abundance of these sequences in the *P. sojae* EST set (30 ESTs) and the existence of a large superfamily of ABC transporters (16 unigenes) and PDR ABC transporters in particular (6 unigenes) is noteworthy. By comparison, 134 sequences were identified in the *P. sojae* genome sequence as having significant similarity to ABC transporters in general (Tyler et al. 2006) and, in the *P. infestans* unigene set of approximately 18,000 reported recently (Randall et al. 2005), 73 were identified as putative ABC transporters based on BLASTX against the SwissProt database (Judelson and Senthil 2006). Our results show that approximately 23% of the predicted ABC transporter genes were expressed strongly enough in the tissues we sampled to yield EST sequences. One ABC transporter was expressed specifically in *P. sojae*-infected tissue, and the *P. infestans* homolog of this gene also was expressed specifically in infection. Therefore, this ABC transporter may have a strongly conserved role in infection.

Another defined group of pathogenicity factors are members of the crinkler superfamily. CRN1 and CRN2 first were identified as necrosis-inducing proteins in a study involving the *P. infestans–Nicotiana* spp. pathosystem (Torto et al. 2003). Currently, 16 crn homologs have been cloned and sequenced from *P. infestans* cDNA libraries. These proteins are relatively large and have been shown to be specific to *Phytophthora* spp. (Torto et al. 2003; Win et al. 2006). In the *P. sojae* EST dataset, 25 *crn*-like sequences were obtained, which makes this the largest gene family observed in this study. An additional 15 *crn*-like genes were reported in the *P. sojae* genome sequence. Preliminary data show that the CRN family is more expanded in the *P. sojae* genome compared with *P. ramorum* (Tyler et al. 2006; Torto-Alalibo T., S. Kamoun, and B. M. Tyler, *unpublished data*). It will be very important in the near future to determine the specific roles of members of this large superfamily. Elicitins and elicitin-like genes so far have been found in only *Phytophthora* and a few *Pythium* spp. and, thus, may be considered oomycete specific. In all, 17 elicitins and elicitin-like proteins were identified in this study compared with 57 elicitin-like gene products predicted from the *P. sojae* genome sequences (Jiang et al. 2006; Tyler et al. 2006). Tissue-specific sets of unigenes present a valuable resource for discovering genes uniquely expressed during particular morphological stages. The library generated from mRNA from swimming zoospores contained more stage-specific genes than any other library. Unigenes specific to the *P. sojae* swimming zoospore library and having homologs significantly highly represented in the *P. infestans* swimming zoospores library included ribonucleotide reductase, phosphatidylinositol kinase, and several with no known function. The swimming zoospore stage is important for the pathogen during infection and, thus, many of the unigenes identified may encode peptides and proteins that have special roles in the initiation of infection. In accord with this, the library most similar to the zoospore library by correlation analysis was the infection library. Also of great importance are the unigenes highly represented in the infection library because these genes are expressed during direct interaction with the host plant. Unigenes encoding effectors, hydrolases, and protease inhibitors were among the genes that were significantly highly represented in the infection library and also have been shown in previous studies in fungal and oomycete pathosystems to be expressed during infection (Gotesson et al. 2002; Kamoun et al. 1997; Qutob et al. 2002; Torto et al. 2002, 2003). A few of the genes specific to the *P. sojae* zoospore and infection libraries had homologs, which also were highly represented, in the *P. infestans* zoospore and infection libraries. This may suggest similar regulation of these transcripts in the two *Phytophthora* spp.

What are the physiological signals that are responsible for stimulating the expression of infection-specific genes? Comparisons of overall EST representation in the libraries by correlation analysis provided the unexpected insight that the physiological state of the pathogen during infection more closely resembled the state during growth on a defined medium of sucrose, asparagine, and salts rather than on any form of complex medium derived from unfractionated plant extracts. The results also indicate that, during infection, the pathogen derives most of its carbon and energy via glycolysis of sugars in the plant, presumably from degradation of plant cell wall material; whereas, in plant-derived complex medium, the pathogen carries out gluconeogenesis. Although many genes showed correlated expression during infection and growth in defined medium, there were many other genes expressed only during infection, indicating that there are further physiological signals that have not yet been discovered. The genes that are specifically expressed during infection will provide tools for identifying those signals.

In conclusion, this study provides a comprehensive insight into the transcriptome of *P. sojae* and the expression pattern of selected genes in the different life stages of the pathogen as well as stage-specific genes. The repertoire of these essential genes can be used as candidates for further functional analysis of pathogenicity and targets for chemical control.

MATERIALS AND METHODS

Culture and growth conditions for *P. sojae*.

P. sojae P6497, originally isolated in Arkansas (United States), was used in all cases. This strain has a race 2 phenotype and does not express the *Avr*1b-1 gene. *P. sojae* EST libraries were generated using the following growth conditions and treatments.

The psMY, psZO, and psHA ESTs were identical to the MY, ZO, and HA ESTs described by Qutob and associates (2000). The psMA, psZS, and psHB ESTs represent additional ESTs sequenced by us from the MY, ZO, and HA cDNA libraries, respectively. The cDNA library for the psZO and psZS ESTs was generated from mRNA from swimming zoospores as described by Qutob and associates (2000). The psZG library was generated from mRNA derived from germinating cysts described briefly as follows. Encysted zoospores were germinated in clarified V8 broth (Erwin and Ribeiro 1996) for 3 h or in soybean seed exudate for 6 h. Soybean seed exudate was produced by soaking 1 kg of organic, edible soybeans in 1 liter of sterile water overnight at room temperature and then filter sterilizing the liquid. RNA was extracted from each batch of germinating cysts, as described below, and equal amounts of RNA from each source was mixed for the cDNA library construction.

The cDNA library for the psHA and psHB ESTs was derived from mRNA extracted from soybean (*Glycine max* (L.) Merr.) cv. Harosoy hypocotyls inoculated with zoospores of *P. sojae* P6497. Etiolated Harosoy seedlings were placed in trays, and hypocotyls were inoculated with approximately 10^4 zoospores produced from 5- to 7-day-old mycelium. At 48 h post-infection, lesions were excised, snap frozen in liquid nitrogen, and stored at –80ºC (Qutob et al. 2000).

The cDNA library for the psMA and psMY ESTs was generated from mRNA derived from mycelia growing on defined media (Qutob et al. 2000). Briefly, mycelium discs cut from the growing edge of 7-day-old synthetic agar media were transferred to flasks containing 50 ml of liquid synthetic medium (2.4 g of sucrose, 0.27 g of asparagine, 10 mg of cholesterol,

10 mg of ascorbate, and 2 mg of thiamine, per liter, plus salts and minerals) and grown at 25ºC in the dark for 3 weeks. Liquid cultures were vacuum decanted onto filter paper and mycelium was collected, flash frozen in liquid nitrogen, lyophilized, and stored at –80ºC.

The psMC library was generated from cDNA derived from mycelium growing in complex medium. Mycelium was grown in V8 broth, lima bean broth (Erwin and Ribeiro 1996), or soybean broth (identical to lima bean broth but made using organic, edible soybeans instead of lima beans). RNA was extracted separately from each of the three batches of mycelia and then mixed in equal amounts for library construction.

The psML library was generated from cDNA derived from mycelium growing in nutrient-limited medium. Mycelium was grown first in clarified V8 broth and then washed well and transferred to water, 10% soybean broth, or 5% clarified V8 broth for 6 h. RNA was extracted separately from each of the three batches of mycelia and then mixed in equal amounts for library construction.

RNA isolation and construction of cDNA libraries.

Frozen zoospores, germinating cysts, mycelia, or infected plant tissues were ground to a fine powder in liquid nitrogen using a mortar and pestle. Total RNA was extracted from zoospores, germinating cysts, and mycelia with phenol-guanidinium isothiocyanate (TRIZOL; Life Technologies/Gibco-BRL, Cleveland, U.S.A.) according to instructions provided by the manufacturer. Total RNA was isolated from inoculated soybean hypocotyls following the procedure of Wang and Vodkin (1994). Construction of three libraries (psZS/psZO, psMY/psMA, and psHA/psHB) was described by Qutob and associates (2000). The psZG, psML, and psMC libraries were constructed by Invitrogen (Carlsbad, CA, U.S.A.) from frozen, total RNA provided after isolation. For the latter three libraries, double-stranded cDNA was synthesized from an oligo(dT) primer, followed by directional cloning into the *Sal*I (5′ end of cDNA) and *Not*I (3′ end of cDNA) sites of vector pCMV•SPORT6.1 (Invitrogen) using cloning adapters. Because *Not*I cleaved the double-stranded cDNA during this procedure, cDNAs with internal *Not*I sites would have been truncated at those sites. The ligated cDNAs were propagated in *Escherichia coli* D12s (Invitrogen).

DNA sequencing of cDNA clones.

Clones were sequenced at the North Carolina State University Genome Research Laboratory from the 5′ end using appropriate sequencing primers to minimize vector sequence (Foreman et al. 2003). Templates were prepared in 96-well format using a semiautomated alkaline lysis method, and sequencing reactions were performed following standard Big Dye (Applied Biosystems, Foster City, CA, U.S.A.) protocols for a 0.25× reaction. Cycle sequencing was performed over 35 cycles (96ºC for 10 s, 50ºC for 5 s, and 60ºC for 4 min) in an Applied Biosystems GenAmp 9700 thermocycler. DyeEx 96-well plates (Qiagen, Valencia, CA, U.S.A.) were used for dye-terminator removal. Samples were sequenced using an ABI 3700 capillary sequencer (Applied Biosystems). A PERL script was used to automatically generate a quality report to monitor sequencing success and read lengths. Successful sequences were defined as having 100 bases of Phred quality values ≥20 or an average quality ≥12. The sequencing success rate was 75 to 80% with an average read length of >450 bp (Phred 20). Trace files (and other needed information) from successful sequences were forwarded for bioinformatic analysis.

EST assembly.

Base calling of EST sequences was done using the Phred program with a cut-off value of 0.1 (Ewing et al. 1998). The sequences were quality trimmed using an in-house algorithm, and the quality values were averaged in a sliding window of 20 bases with the cut-off set at 25. If the average value descended below 25, the first base of the sequence was removed; in this fashion, the sliding window moved until no average value of less than 25 was found for 10 consecutive windows. Vector and adaptor sequences were removed using cross-match with settings at –minmatch 10 and –minscore 20 and, for adaptor, the –minmatch and –minscore were both set to 8. Clustering of cleaned EST sequences was performed by TGICL software (The Institute for Genomic Research, Rockville, MD, U.S.A.), which implements the clustering algorithm from MegaBlast. Clustered ESTs were assembled using CAP3 (Huang and Madan 1999) with the following parameters: a minimum percent identity for overlaps of 94, a minimum overlap length of 30, a maximum length of unmatched overhangs of 30, and a maximum number of sequences in a clustering slice of 1,000. To differentiate the soybean and *P. sojae* sequences within the infection libraries (psHA and psHB), the sequences were compared with soybean ESTs available at the Washington University Genome Sequencing Center, excluding EST sets that were derived from *P. sojae*-infected soybean tissue. Sequences were identified as being of soybean origin when there was greater than 95% identity over 100 bases nongapped high scoring pair. Sequences not identified as belonging to soybean using this method were compared at the same level of stringency with the pure *P. sojae* EST libraries: psMY, psMA, psMC, psML, psZO, psZS, and psZG. ESTs matching these libraries were categorized as *P. sojae* ESTs. After this step, ESTs as yet uncharacterized were analyzed with an in-house program to determine GC content and CG/TA dinucleotide ratio. ESTs with GC content greater than 50% and possessing a CG/TA dinucleotide frequency greater than 1.0 were categorized as originating from *P. sojae*. Sequences that were still of ambiguous origin then were examined manually; most of these proved to originate either from *P. sojae* rRNA or mtDNA or could be identified by strong matches to other plant sequences as originating from soybean sequences not present in the EST database. When the *P. sojae* genome sequence (Tyler et al. 2006) became available, all ESTs identified as of *P. sojae* origin were checked against that sequence. This was done using the BLAT alignment. The EST sequences from this project are available at the *Phytophthora* Functional Genomics Database (Gajendran et al. 2006) and the EST Analysis Pipeline (Mao et al. 2003); the most recent version is available in the *Phytophthora* EST Database, which is part of the Virginia Bioinformatics Institute (VBI) Microbial Database (Tripathy et al. 2006). To access the gene models corresponding to the unigenes used in this study, queries can be made with unigene labels in the *Phytophthora* EST database, which will open up a window listing, among others, the gene model or models corresponding to the unigene. Clicking on the gene model names will open up another window in the *Phytophthora* genome sequence database that showcases more features of the gene model in question.

EST sequence analysis and functional assignment.

To ascertain the frequency of occurrence of each EST within contigs, a data file was compiled that included contig ID, number of ESTs per contig, and distribution of EST across all libraries. Sequences that were represented in only one library were identified as library-specific sequences. Searches for similarity between datasets generated in this study and publicly available datasets (GenBank, National Center for Biotechnology Information [NCBI]) were performed using a locally installed BLASTX algorithm (Altschul et al. 1990) with a cut-off E value ≤ $1e^{-4}$. To identify putative pathogenicity genes, the categories published by Panabieres and associates (2005)

and Idnurm and Howlett (2001) were employed with some modifications. Annotations based on literature searches and best BLASTX matches facilitated the placement of unigenes into respective categories. To further assist assigning putative functions to unigenes, InterProScan (Zdobnov and Apweiler 2001) was run locally, and all sequences were electronically translated and then searched against all InterPro protein domains (Mulder et al. 2005). Unigenes were assigned GO terms based on their similarity to InterPro domains already mapped to the GO classification scheme of molecular function, biological process, and cellular component. Putative gene products containing signal peptides were assessed by analyzing batches of electronically translated sequences with SignalP 3.0 (Bendtsen et al. 2004). Selection of putative extracellular proteins was based on the criteria used by Win and associates (2006). Briefly, a gene product was identified as extracellular if the SignalP Hidden Markov Model algorithm predicted a signal peptide with a probability greater than 0.900 and if the SignalP Neural Network algorithm (Win et al. 2006) predicted a cleavage site between 10 and 40 amino acids.

Gene expression profiles.

To estimate differential gene expression among the different libraries, the number of ESTs contributing to each contig was determined. R statistics, the log of the likelihood ratio, was calculated to compare the gene expression level in different libraries (Stekel et al. 2000). *P* values were obtained from permutation tests with 5,000 permutations and were adjusted by the Benjamini-Hochberg false discovery rate (Benjamini and Hochberg 1995). Correlation analysis was performed using only those 2,845 unigenes that consisted of four or more ESTs. Pearson's correlation coefficient was calculated for individual pairs of libraries. Using the set of unigenes exclusively found in the zoospore and infection libraries, we searched the *P. ramorum* genome sequences (Tyler et al. 2006) and the preliminary assemblies of *P. capsici* and *P. infestans* genome sequences for matches to these stage-specific genes using BLASTX and TBLASTN with a cut-off E value of $1e^{-20}$. The distribution of the *P. sojae* zoospore and infection-specific genes also were compared with the *P. infection* EST libraries (Gajendran et al. 2006; Randall et al. 2005) to identify similar distributions.

ACKNOWLEDGMENTS

This work was supported by grants to B. M. Tyler from the National Research Initiative of the United States Department of Agriculture Cooperative State Research, Education, and Extension Service grant numbers 00-52100-9684 and 2002-35600-12747; the National Science Foundation grant number MCB-0242131; the Plant Genome Research Program of the National Science Foundation, cooperative agreement DBI-0211863; and by funds from the Virginia Bioinformatics Institute. B. W. S. Sobral was supported in part by grant number W911SR-04-C-0045 from the Department of Defense. The authors thank E. Bush and C. Volker for thorough editing of the manuscript.

LITERATURE CITED

Allen, R. L., Bittner-Eddy, P. D., Grenville-Brigs, L. J., Meitz, J. C., Rehmany, A. P., Rose, L. E., and Beynon, J. L. 2004. Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. Science 306:1957-1960.

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.

Armstrong, M. R., Whisson, S. C., Pritchard, L., Bos, J. I. B., Venter, E., Avrova, A. O., Rehmany, A. P., Bohme, U., Brooks, K., Cherevach, I., Hamlin, N., White, B., Frasers, A., Lord, A., Quail, M. A., Churcher, C., Hall, N., Berriman, M., Huang, S., Kamoun, S., Beynon, J. L., and Birch, P. R. J. 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. U.S.A. 102:7766-7771.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25:25-29.

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

Bauer, B. E., Wolfer, H., and Kuchler, K. 1999. Inventory and function of yeast ABC proteins: About sex, stress, pleiotropic drug and heavy metal resistance. Biochem. Biophys. Acta 1461:217-236.

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.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. 57 289-300.

Bhattacharyya, M. K., and Ward, E. W. B. 1986. Resistance, susceptibility, and accumulation of glyceollins I-III in soybean organs inoculated with *Phytophthora megasperma* f. sp. *glycinea*. Physiol. Mol. Plant Pathol. 29:227-237.

Bhattacharjee, S., Hiller, N. L., Liolios, K., Win, J., Kanneganti, T. D., Young, C., Kamoun, S., and Haldar, K. 2006. The malarial host-targeting signal is conserved in the Irish potato famine pathogen. PLoS Pathol. 2:453-465.

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., Dodson, R. J., Deboy, R. T., Durkin, A. S., Kolonay, J. F., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M. J., Haft, D. H., Nelson, W. C., Davidsen, D., Zafar, N., Zhou, L., Liu, J., Yuan, Q., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Utterback, T., Van Aken, S. E., Feldblyum, T. V., D'Ascenzo, M., Deng, W.-L., Ramos, A., Alfano, J. R., Cartinhour, S., Chatterjee, A. K., Delaney, T. P., Lazarowitz, S. G., Martin, G. B., Schneider, D. J., Tang, X., Bender, C. L., White, O., Fraser, C. M., and Collmer, A. 2003. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186.

Carlile, M. J. 1983. Motility, taxis, and tropism in *Phytophthora*. Pages 95-107 in: *Phytophthora*: Its Biology, Taxonomy, Ecology and Pathology. D. C. Erwin, S. Bartnicki-Garcia, and P. H. Tsao, eds. American Phytopathological Society Press, St. Paul, MN, U.S.A.

Clark, J. I., Brooksbank, C., and Lomax, J. 2005. Its all GO for plant scientists. Plant Physiol. 138:1268-1279.

Collmer, A., Badel, J. L., Charkowski, A. O., Deng, W. L., Fouts, D. E., Ramos, A. R., Rehm, A. H., Anderson, D. M., Schneewind, O., van Dijk, K., and Alfano, J. R. 2000. *Pseudomonas syringae* Hrp type III secretion system and effector proteins. Proc. Natl. Acad. Sci. U.S.A. 97:8770-8777.

Dassa, E., and Bouige, P. 2001. The ABC of ABCs: A phylogenetic and functional classification of ABC systems in living organisms. Res. Microbiol. 152:211-229.

Davidson, A. L., and Chen, J. 2004. ATP binding cassette transporters in bacteria. Ann. Rev. Biochem. 73:242-268.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., Thon, M., Kulkarni, R., Xu, J.-R., Pan, H., Read, N. D., Lee, Y.-H., Carbone, I., Brown, D., Oh, Y. Y., Donofrio, N., Jeong, J. S., Soanes, D. M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.-H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.-J., Nicol, R., Purcell, S., Nusbaum, C., Nusbaum, J. E., and Birren, B. W. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434:980-986.

Del Sorbo, G., Schoonbeek, H. J., and De Waard, M. A. 2000. Fungal transporters involved in efflux of natural toxic compounds and fungicides. Fungal Genet. Biol. 30:1-15.

Ebbole, D. J., Jin, Y., Thon, M., Pan, H., Bhattarai, E., Thomas, T., and Dean, R. A. 2004. Gene discovery and gene expression in the rice blast fungus, *Magnaporthe grisea*: Analysis of expressed sequence tags. Mol. Plant-Microbe Interact. 17:1337-1347.

Erwin, D. C., and Ribeiro, O. K. 1996. *Phytophthora* Diseases Worldwide. American Phytopathological Society, St. Paul, MN, U.S.A.

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175-185.

Fleissner, A., Sopalla, C., and Weltring, K. M. 2002. An ATP-binding cassette multidrug-resistance transporter is necessary for tolerance of *Gibberella pulicaris* to phytoalexins and virulence on potato tubers. Mol. Plant-Microbe Interact. 15:102-108.

Foreman, P. K., Brown, D., Dankmeyer, L., Dean, R. A., Diener, S., Dunn-

Coleman, N. S., Goedegebuur, F., Houfek, T. D., England, G. J., Kelley, A. S., Meerman, H. J., Mitchell, T. K., Mitchinson, C., Olivares, H. A., Teunissen, J. M., Yao, J., and Ward, M. 2003. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. J. Biol. Chem. 278:31988-31997.

Gajendran, K., Gonzales, M. D., Farmer, A., Archuleta, E., Win, J., Waugh, M. E., and Kamoun, S. 2006. Phytophthora functional genomics database (PFGD): Functional genomics of *Phytophthora*–plant interactions. Nucleic Acids Res. 34:D465-D470.

The Gene Ontology Consortium. 2006. The Gene Ontology (GO) project in 2006. Nucleic Acids Res. 34:D322-D326.

Gotesson, 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.

Hayashi, K., Schoonbeek, H. J., Sugiura, H., and De Waard, M. A. 2001. Multidrug resistance in *Botrytis cinerea* associated with decreased accumulation of the azole fungicide oxpoconazole and increased transcription of the ABC transporter gene BcatrD. Pestic. Biochem. Physiol. 70:168-179.

Hayashi, K., Schoonbeek, H. J., and De Waard, M. A. 2002. Bcmfs1, a novel major facilitator superfamily transporter from *Botrytis cinerea*, provides tolerance towards the natural toxic compounds camptothecin and cercosporin and towards fungicides. Appl. Environ. Microbiol. 68:4996-5004.

Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K. 2004. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science 306:1934-1937.

Huang, X. Q., and Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9:868-877.

Huitema, E., Bos, J. I. B., Tian, M., Win, J., Waugh, M. E., and Kamoun, S. 2004. Linking sequence to phenotype in *Phytophthora*–plant interactions. Trends Microbiol. 12:193-200.

Idnurm, A., and Howlett, B. J. 2001. Pathogenicity genes of phytopathogenic fungi. Mol. Plant Pathol. 12:241-255.

Jiang, R. H. Y., Tyler, B. M., Whisson, S. C., Hardham, A. R., and Govers, F. 2006. Ancient origin of elicitin gene clusters in *Phytophthora* genomes. Mol. Biol. Evol. 23:338-351.

Judelson, S. H., and Senthil, G. 2006. Investigating the role of ABC transporters in multifungicide insensitivity in *Phytophthora infestans*. Mol. Plant Pathol. 7:17-29.

Kamoun, S. 2003. Molecular genetics of pathogenic oomycetes. Eukaryotic Cell 2:191-199.

Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. Annu. Rev. Phytopathol. 44:2.1-2.20.

Kamoun, S., van-West, P., de Jong, A. J., Vleeshouwers, V. G. A. A., and Govers, F. 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., Hraber, P., Sobral, B., Nuss, D., and Govers, F. 1999. Initial assessment of gene diversity for the oomycete pathogen *Phytophthora infestans* based on expressed sequences. Fungal Genet. Biol. 28:94-106.

Kaufmann, M. J., and Gerdemann, J. W. 1958. Root and stem rot of soybeans caused by *Phytophthora sojae* n. sp. Phytopathology 48:201-208.

Keon, J., Antoniw, J., Rudd, J., Skinner, W., Hargreaves, J., and Hammond-Kosack, K. 2005. Analysis of expressed sequence tags from the wheat leaf blotch pathogen *Mycosphaerella graminicola* (anamorph *Septoria tritici*). Fungal Genet. Biol. 42:376-389.

Klokouzas, A., Shahi, S., Hladky, S. B., Barrand, M. A., and van Veen, H. W. 2003. ABC transporters and drug resistance in parasitic protozoa. Int. J. Antimicrob. Agents 22:301-317.

Knogge, W. 1996. Fungal infections of plants. Plant Cell 8:1711-1722.

Litman, T., Druley, T. E., Stein, W. D., and Bates, S. E. 2001. From MDR to MXR: New understanding of multidrug resistance systems, their properties and clinical significance. Cell Mol. Life Sci. 58:931-959.

Mao, C. H., Cushman, J. C., May, G. D., and Weller, J. W. 2003. ESTAP—an automated system for the analysis of EST data. Bioinformatics 19:1720-1722.

Marti, M., Good, R. T., Rug, M., Knuepfer, E., and Cowman, A. F. 2004. Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science 306:1930-1933.

McLeod, A., Smart, C. D., and Fry, W. E. 2002. Characterization of 1,3-beta-glucanase and 1,3;1,4-beta-glucanase genes from *Phytophthora infestans*. Fungal Genet. Biol. 38:250-263.

Morris, P. F., Bone, E., and Tyler, B. M. 1998. Chemotropic and contact responses of *Phytophthora sojae* hyphae to soybean isoflavonoids and artificial substrates. Plant Physiol. 117:1171-1178.

Morschhauser, J. 2002. The genetic basis of fluconazole resistance development in *Candida albicans*. Biochem. Biophys. Acta 1587:240-248.

Moy, P., Qutob, D., Chapman, B. P., Atkinson, I., and Gijzen, M. 2004. Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. Mol. Plant-Microbe Interact. 17:1051-1062.

Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Pagni, M., Pointing, C. P., Quevillon, E., Selengut, J., Sigrist, C. J. A., Silventoinen, V., Studholme, D. J., Vaughan, R., and Wu, C. H. 2005. InterPro, progress and status in 2005. Nucleic Acids Res. 33:D201-D205.

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.

Qutob, D., Hraber, P. T., Sobral, B. W., and Gijzen, M. 2000. Comparative analysis of expressed sequences in *Phytophthora sojae*. Plant Physiol. 123:243-254.

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

Randall, T. A., Dwyer, R. A., Huitema, E., Beyer, K., Cvitanich, C., Kelkar, H., Ah Fong, A. M. V., Gates, K., Roberts, S., Yatzkan, E., Gaffney, T., Law, M., Testa, A., Torto, T., Zhang, M., Zheng, L., Mueller, E., Windass, J., Binder, A., Birch, P. R. J., Gisi, U., Govers, F., Gow, N., Mauch, F., van West, P., Waugh, M., Yu, J., Boller, T., Kamoun, S., Lam, S. T., and Judelson, H. S. 2005. Large-scale gene discovery in the oomycete *Phytophthora infestans* reveals likely components of phytopathogenicity shared with true fungi. Mol. Plant-Microbe Interact. 18:229-243.

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

Shan, W., Cao, M., Leung, D., and Tyler, B. M. 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.

Staskawicz, B. J., Mudgett, M. B., Dangl, J. L., and Galan, J. E. 2001. Common and contrasting themes of plant and animal diseases. Science 292:2285-2289.

Stekel, D. J., Git, Y., and Falciani, F. 2000. The comparison of gene expression from multiple cDNA libraries. Genome Res. 10:2055-2061.

Stossel, P., Lazarovitis, G., and Ward, E. W. B. 1980. Penetration and growth of compatible and incompatible races of *Phytophthora megasperma* var. *sojae* in soybean hypocotyls tissues differing in age. Can. J. Bot. 58:2594-2601.

Thomas, S. W., Rasmussen, S. W., Glaring, M. A., Rouster, J. A., Christiansen, S. K., and Oliver, R. P. 2001. Gene identification in the obligate fungal pathogen *Blumeria graminis* by expressed sequence tag analysis. Fungal Genet. Biol. 33:195-211.

Torto, T., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N. A. R., van West, P., and Kamoun, S. 2003. EST mining and functional expression assays identify extracellular effector proteins from *Phytophthora*. Genome Res. 13:675-685.

Torto, T. A., Rauser, L., and Kamoun, S. 2002. The *pipg1* gene of the oomycete *Phytophthora infestans* encodes a fungal-like endopolygalacturonase. Curr. Genet. 40:385-390.

Trail, F., Xu, J.-R., Miguel, P. S., Halgren, R. G., and Kistler, H. C. 2002. Analysis of expressed sequence tags from *Gibberella zeae* (anamorph *Fusarium graminearum*). Fungal Genet. Biol. 38:187-197.

Tripathy, S., Pandey, V. N., Fang, B., Salas, F., and Tyler, B. M. 2006. VMD: A community annotation database for oomycetes and microbial genomes. Nucleic Acids Res. 34:D379-D381.

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F., Baxter, L., Bensasson, D., Beynon, J. L., Damasceno, C. M. B., Dickerman, A., Dorrance, A. E., Dou, D., Dubchak, I., Garbelotto, M., Gijzen, M., Gordon, S., Govers, F., Grunwald, N., Huang, W., Ivors, K., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K., Lee, M.-K., McDonald, W. H., Medina, M., Meijer, H. J. G., Nordberg, E., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P., Phuntumart, V., Putnam, N., Rash, S., Rose, J. K. C., Sakihama, Y., Salamov, A., Savidor, A., Scheuring, C., Smith, B., Sobral, B. W. S., Terry, A., Torto-Alalibo, T., Win, J., Xu, Z., Zhang, H., Grigoriev, I., Rokhsar, D., and Boore, J. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1197-1313.

Van den Brule, S., and Smart, C. C. 2002. The plant PDR family of ABC transporters. Planta 216:95-106.

Wang, C. S., and Vodkin, L. O. 1994. Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. Plant Mol. Biol. 12:132-145.
Win, J., Kanneganti, T.-D., Torto-Alalibo, T., and Kamoun, S. 2006. Computational and comparative analysis of 150 full-length cDNA sequences from the oomycete plant pathogen *Phytophthora infestans*. Fungal Genet. Biol. 43:20-33.
Wrather, J. A., Anderson, T. R., Arsyad, D. M., Tan, Y., Plopper, L. D., Porta-Puglia, A., Ram, H. H., and Yorinori, J. R. 2001. Soybean disease loss estimates for the top ten soybean-producing countries in 1998. Can. J. Plant Pathol. 23:115-121.
Zdobnov, E. M., and Apweiler, R. 2001. InterProScan-an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17:847-848.

AUTHOR-RECOMMENDED INTERNET RESOURCES

Broad Institute *Phytophthora infestans* genome sequence database: www.broad.mit.edu/annotation/genome/phytophthora_infestans/Home.html
EST Analysis Pipeline: www.vbi.vt.edu/~estap
NCBI GenBank database: www.ncbi.nlm.nih.gov/Genbank/index.html
Phytophthora functional genomics database: www.pfgd.org
Phytophthora soybean EST database: phytophthora.vbi.vt.edu/EST
SignalP 3.0: www.cbs.dtu.dk/services/SignalP/
Virgnia Bioinformatics Institute Microbial Database: phytophthora.vbi.vt.edu
Washington University Genome Sequencing Center: www.genome.wustl.edu/est/index.php