MPMI Vol. 18, No. 3, 2005, pp. 229–243. DOI: 10.1094/MPMI-18-0229. © 2005 The American Phytopathological Society

e-Xtra*

Large-Scale Gene Discovery in the Oomycete *Phytophthora infestans* Reveals Likely Components of Phytopathogenicity Shared with True Fungi

Thomas A. Randall,[1] Rex A. Dwyer,[1] Edgar Huitema,[2] Katinka Beyer,[3] Cristina Cvitanich,[4] Hemant Kelkar,[1] Audrey M. V. Ah Fong,[4] Krista Gates,[1] Samuel Roberts,[4] Einat Yatzkan,[1] Thomas Gaffney,[1] Marcus Law,[1] Antonino Testa,[2] Trudy Torto-Alalibo,[2] Meng Zhang,[5] Li Zheng,[1] Elisabeth Mueller,[6] John Windass,[6] Andres Binder,[7] Paul R. J. Birch,[8] Ulrich Gisi,[7] Francine Govers,[9] Neil A. Gow,[10] Felix Mauch,[11] Pieter van West,[10] Mark E. Waugh,[12] Jun Yu,[5] Thomas Boller,[3] Sophien Kamoun,[2] Stephen T. Lam,[1] and Howard S. Judelson[4]

[1]Syngenta Biotechnology Inc., 3054 Cornwallis Road, Research Triangle Park, NC 27709, U.S.A.; [2]Department of Plant Pathology, The Ohio State University, Ohio Agricultural Research and Development Center, Wooster 44691, U.S.A.; [3]Friedrich Miescher Institute, P. O. Box 2543, CH-4002 Basel, Switzerland; [4]Department of Plant Pathology and Center for Plant Cell Biology, University of California, Riverside 92521, U.S.A.; [5]Beijing Genomics Institute, Institute of Genetics, and Graduate School, Chinese Academy of Sciences, Beijing, China; [6]Syngenta Limited, Jealott's Hill International Research Station, Bracknell, Berks RG42 6EY, U.K.; [7]Syngenta Crop Protection AG, Werk Stein, Schaffhauserstrasse, CH-4332 Stein, Switzerland; [8]Plant Pathogen Interactions Program, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland; [9]Laboratory of Phytopathology, Wageningen University, 6709 PD Wageningen, The Netherlands; [10]Department of Molecular and Cell Biology, Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, Scotland, U.K.; [11]Department of Biology, University of Fribourg, CH-1700 Fribourg, Switzerland; [12]National Center for Genome Resources, 1800 Old Pecos Trail, Santa Fe, NM 87505 U.S.A.

Submitted 19 July 2004. Accepted 7 November 2004.

To overview the gene content of the important pathogen *Phytophthora infestans,* large-scale cDNA and genomic sequencing was performed. A set of 75,757 high-quality expressed sequence tags (ESTs) from *P. infestans* was obtained from 20 cDNA libraries representing a broad range of growth conditions, stress responses, and developmental stages. These included libraries from *P. infestans*–potato and –tomato interactions, from which 963 pathogen ESTs were identified. To complement the ESTs, onefold coverage of the *P. infestans* genome was obtained and regions of coding potential identified. A unigene set of 18,256 sequences was derived from the EST and genomic data and characterized for potential functions, stage-specific patterns of expression, and codon bias. Cluster analysis of ESTs revealed major differences between the expressed gene content of mycelial and spore-related stages, and affinities between some growth conditions. Comparisons with databases of fungal pathogenicity genes revealed conserved elements of pathogenicity, such as class III pectate lyases, despite the considerable evolutionary distance between oomycetes and fungi. Thirty-seven genes encoding components of flagella also were identified. Several genes not anticipated to occur in oomycetes were detected, including chitin synthases, phosphagen kinases, and a bacterial-type FtsZ cell-division protein. The sequence data described are available in a searchable public database.

Corresponding author: Howard S. Judelson, Telephone: 951-827-4199; Fax: 951-827-4294; E-mail: howard.judelson@ucr.edu

This is a publication of the Syngenta Phytophthora Consortium, a confederation of Syngenta and academic laboratories at Friedrich Miescher Institute, Ohio State University, Scottish Crop Research Institute, University of Aberdeen, University of California, University of Fribourg, and Wageningen University.

Nucleotide sequence data of individual ESTs are available in the dbEST division of GenBank as accession numbers CV893097 to CV970752. Additional sequences predicted from genomic sequencing to be of high coding potential are available in the GSS division of GenBank as accession numbers CW803167 to CW822241.

*The *e*-Xtra logo stands for "electronic extra" and indicates the HTML abstract available on-line contains supplemental material with the sequences of the unigenes predicted from *P. infestans* and a comparison of the frequencies of Pfam domains in predicted proteins from *P. infestans, Neurospora crassa,* and *Arabidopsis thaliana*.

The eukaryotic microbes known as oomycetes include several groups of important plant pathogens. All members of the genus *Phytophthora* infect plants, although some exhibit a broad host range while others infect only a few species (Erwin and Ribeiro 1996). In the latter class is *Phytophthora infestans*, the cause of late blight of potato and tomato. This pathogen has posed a serious constraint to crop production ever since the time of the Irish potato famine, causing billions of dollars of losses per year (Fry and Goodwin 1997). Numerous other *Phytophthora* spp. also have a history of causing important diseases, and new diseases continue to emerge. For example, a recently discovered European alder disease is caused by an interspecific *Phytophthora* hybrid (Brasier et al. 1999), and sudden oak death is caused by a newly defined species, *P. ramorum* (Rizzo et al. 2002; Werres et al. 2001).

The growth habit of oomycetes resembles that of true fungi, and the two groups use similar strategies to colonize

plants (Latijnhouwers et al. 2003). However, the two groups are only distantly related because oomycetes are phylogenetically closest to diatoms and golden-brown algae (Margulis and Schwartz 2000; Sogin and Silberman 1998). Therefore, true fungi provide poor models for understanding oomycete biology. In recent years, tools and resources for molecular and genetic analyses of oomycetes, particularly *P. infestans*, have been developed and now are used routinely by several laboratories. These include methods for efficiently performing genetic crosses (Judelson et al. 1995), a detailed genetic map (van der Lee et al. 2004), BAC (bacterial artificial chromosome) libraries (Randall and Judelson 1999; Whisson et al. 2001), and a range of transformation procedures and reporter genes (Kamoun 2003). Gene silencing also has facilitated studies of genes involved in growth, development, and pathogenicity (Ah Fong and Judelson 2003; Kamoun et al. 1998; Latijnhouwers and Govers 2003; Latijnhouwers et al. 2004).

To complement these tools and expand our understanding of oomycete biology, we report here the results of a major cDNA and genomic DNA sequencing project for *P. infestans*. First, an expressed sequence tag (EST) approach (Adams et al. 1991) was used to identify transcribed sequences from *P. infestans*. This appeared to represent the most efficient means of gaining information on gene content of *P. infestans* because the genome of this organism is relatively large (237 Mb) (Tooley and Therrien 1987) and contains 50% or more repetitive DNA (Judelson and Randall 1998). A pilot EST project, which examined 1,000 clones from a single library of *P. infestans,* also had proved informative (Kamoun et al. 1999). In the present study, cDNA libraries from 19 conditions of growth and development, including four plant interaction treatments, were constructed and sequenced to generate 75,757 high-quality ESTs from *P. infestans*. Genomic DNA from *P. infestans* also was sequenced to onefold coverage to further add to the discovered gene set. In total, a unigene set of 18,256 was assembled from the ESTs and coding regions predicted from the genome sequence. The unigenes were functionally annotated, stage-specific genes were predicted, and putative pathogenicity genes were compared with those from true fungi to gain insight into the mechanisms of infection by these distantly related microbial pathogens.

RESULTS

cDNA libraries and cDNA sequencing.

The cDNA libraries examined and the number of ESTs obtained are shown in Table 1. The libraries represented hyphae grown under eight nutritional and stress conditions, sporulating hyphae, mating cultures, ungerminated and directly germinating sporangia, germinating cysts (including many forming appressorium-like structures), and interactions of *P. infestans* with plants. The latter included young and older lesions from potato and tomato as well as hyphae exposed to potato leaf exudates.

An iterative process of sequencing, data analysis, and further sequencing maximized the number of informative ESTs obtained. Initially, 500 to 2,000 ESTs were sampled from each library. These were examined for redundancy and for the fraction of novel sequences based on dissimilarity to data in GenBank or the other cDNA libraries. Each library was sequenced until the yield of new data became unacceptably low, generally when the fraction of new genes fell to 5 to 10% based on BLASTN analysis with a cutoff of $E = 1 \times 10^{-30}$.

At the conclusion of sequencing, 74,789 high-quality ESTs (phred score >20) were obtained from *P. infestans* libraries plus 5,236 from mixed *P. infestans*–plant libraries. Of the latter, 968 were predicted to be of *P. infestans* origin. This was based on their having a GC content >50%, which helps distinguish *P. infestans* from host transcripts (Huitema et al. 2003; Ronning et al. 2002), and lacking strong matches against potato or tomato sequences in GenBank (using a threshold BLASTN *E* value of 10^{-45}).

Genomic sequencing of *P. infestans*.

To complement the EST analysis, genomic DNA was sequenced using libraries constructed from DNA that was *Hin*dIII or *Eco*RI digested, or sheared randomly. A total of 390,944 kb (phred score >20) was generated from 588,501 sequence reads from these libraries. The fraction of the genome sampled was estimated by comparisons with 4,082 consensus sequences assembled from a preliminary set of 35,000 ESTs. Using BLASTN with a cutoff *E* value of 10^{-30}, 62% of the genomic reads matched the EST assembly. This suggested that the genome sequence represented approximately onefold coverage

Table 1. Description of cDNA libraries

Library	Strain of *P. infestans*	Library description	No. of ESTs[b]	Library-specific[a]	
				Number	Percent
PA	88069	Mycelium, non-sporulating growth	4,579	337	8.2
PB	88069	Mycelium, sporulating growth	4,041	523	12.9
PD	88069	Mycelium, nitrogen starvation	8,469	1,177	13.9
PE	88069	Mycelium, carbon starvation	2,730	449	16.0
PK	88069	Mycelium, heat treated	765	60	8.0
PL	88069	Mycelium, H_2O_2 treated	1,182	96	8.0
PY	88069	Mycelium, Plich medium	6,864	1,084	15.8
PX	88069	Mycelium, starved in water	6,934	665	9.6
PW	88069	Mycelium, infection mimic	352	19	5.0
PU	88069	Mycelium, subtracted infection mimic	2,355	220	9.3
PM	88069, 618	Mating of 88069 (A1) and 618 (A2)	14,834	2,281	15.0
PJ	88069	Sporangia, purified	4,029	462	11.5
PF	88069	Sporangia, cleaving	4,174	622	15.0
PG	88096	Sporangia, germinating	1,622	147	9.0
PV	88069	Zoospores, purified	7,786	896	11.5
PH	88069	Cysts, germinating	7,287	1,414	19.0
PC	90128	Infected tomato, center of lesion 3 days after infection	3,213	184[c]	…
PI	90128	Infected tomato, outside lesion 3 days after infection	710	184	…
PN	88069	Infected potato, center of lesion 6 days after infection	658	184	…
PO	88069	Infected potato, outside lesion 6 days after infection	655	184	…

[a] Not present in other libraries at a BLASTN cutoff of $E = 10^{-30}$.
[b] ESTs = expressed sequence tags.
[c] Unique content for libraries PC, PI, PN, and PO determined collectively to be 184.

based on the Lander and Waterman formula (Lander and Waterman 1988).

Assembly of *P. infestans* unigene set.

An 18,256-member unigene set was assembled from the EST and genomic data using the Paracel Transcript Assembler (v. 3.6.1; Paracel, Pasadena, CA U.S.A.). The term "unigene" is used here to describe a sequence that is predicted to represent a single gene. The assembly included the ESTs from the *P. infestans* libraries, ESTs from the *P. infestans*–plant interaction libraries predicted to be of *P. infestans* origin, 215 previously identified genes, and genomic reads with coding potential based on a codon bias-based algorithm (Dwyer 2002). Among the unigenes, 10,332 were consensus sequences assembled from multiple sequence reads and 7,924 were singlets. Of the latter, 2,330 represented coding regions predicted from genome data that failed to match an EST.

The unigene sequences ranged from 59 to 3,650 bp, with an average of 682 bp and 1.67 sequences per cluster. Overall, 43% of unigenes had a match in the SwissProt protein database using a BLASTX cutoff of $E = 1 \times 10^{-5}$.

Codon usage within the *P. infestans* transcriptome.

To assess codon usage, 400 full-length open reading frames (ORFs) representing 94,719 codons were predicted from the assembly and analyzed (Table 2). Overall, coding regions exhibited 57.9% G+C with the first, second, and third positions of each triplet being 57.7, 45.5, and 70.5% G+C. A moderately biased pattern of codon choice was revealed by calculating relative synonymous codon usage (RSCU) values, which measures nonrandomness in codons encoding each amino acid (Table 2) (Sharp et al. 1986). The bias was largely due to a preference for guanine or cytosine in the last position, as exemplified by the case of phenylalanine, where UUC was used four times more than UUU. However, this was not the only contributor to the skewed choice of codons. For example, GGC was employed eight times more than GGG to encode glycine, whereas UCC was used half as often as UCG for serine. Substantial variation in codon usage was noted between different genes, because effective number of codons (N_c) values ranged between 28 and 61; this metric scores bias ranged from the maximum ($N_c = 20$) to the minimum ($N_c = 61$) (Wright 1990). The strongest source of intergenic N_c variation was the G+C content at the third position of the codon, with a correlation coefficient calculated at –0.87.

This analysis involved genes predicted from ESTs; therefore, the data shown in Table 2 should be considered preliminary. Efforts were made to minimize the bias of the dataset by selecting genes of diverse function (metabolic, regulatory, no BLAST hit, and so on) and expression patterns (high and low abundance, and stage-specific and constitutive), and no striking differences in codon usage were observed between these subgroups. However, small genes were over-represented in the dataset due to the challenge of constructing larger ORFs from ESTs. Consequently, the mean ORF analyzed was 710 nucleotides compared with approximately 1.1 kb for average genes. Gene length, as well as expression pattern, have been shown to influence codon usage in other organisms (Duret and Mouchiroud 1999).

Functional annotation of the unigene set.

The 18,256-unigene set was annotated using comparisons with the SwissProt database and the Pfam database of protein domains, using *E* value cutoffs of 10^{-5}. The unigenes were then placed in Munich Information Center for Protein Sequences (MIPS) categories (Mewes et al. 2002). Genes involved in metabolism were predominant, followed by those in cellular communication and signal transduction (Fig. 1).

The 40 Pfam domains detected most often in the *P. infestans* sequences are listed in Table 3, and the full set is provided as supplementary data. Of the 5 Pfam domains most common in *P. infestans*, 4 also were among the 20 most common in other

Table 2. Codon usage in *Phytophthora infestans*[a]

Amino acid	Codon	Frequency/1,000	RSCU
Phe	UUU	7.7	0.44
	UUC	27.7	1.57
Leu	UUA	1.7	0.13
	UUG	10	0.76
Tyr	UAU	5.6	0.36
	UAC	25.9	1.64
ter	UAA	1.8	1.32
ter	UAG	0.9	0.62
Leu	CUU	11	0.84
	CUC	19.2	1.48
	CUA	5.5	0.42
	CUG	30.7	2.36
His	CAU	4.7	0.42
	CAC	17.6	1.57
Gln	CAA	8.9	0.52
	CAG	25.3	1.48
Ile	AUU	13.4	0.98
	AUC	26.5	1.93
	AUA	1.3	0.1
Met	AUG	23.8	1
Asn	AAU	7.7	0.41
	AAC	29.9	1.59
Lys	AAA	8	0.29
	AAG	47.1	1.71
Val	GUU	8.8	0.51
	GUC	19.7	1.15
	GUA	4.8	0.28
	GUG	35.2	2.05
Asp	GAU	13.7	0.51
	GAC	40.5	1.5
Glu	GAA	15.2	0.49
	GAG	46.5	1.51
Ser	UCU	8.3	0.66
	UCC	12.1	0.96
	UCA	6.2	0.49
	UCG	25	1.98
Cys	UGU	5.4	0.61
	UGC	12.2	1.38
ter	UGA	1.5	1.09
Trp	UGG	11.8	1
Pro	CCU	11.9	0.9
	CCC	13.6	1.03
	CCA	8.7	0.66
	CCG	18.8	1.42
Arg	CGU	16.9	1.92
	CGC	19	2.16
	CGA	6.7	0.76
	CGG	3.7	0.42
Thr	ACU	12.1	0.69
	ACC	22.1	1.25
	ACA	9	0.51
	ACG	27.3	1.55
Ser	AGU	7.8	0.62
	AGC	16.2	1.29
Arg	AGA	3.4	0.38
	AGG	3.2	0.36
Ala	GCU	29.7	1.23
	GCC	34.5	1.43
	GCA	13	0.54
	GCG	19.3	0.8
Gly	GGU	22.1	1.19
	GGC	35	1.89
	GGA	12.6	0.68
	GGG	4.5	0.24

[a] Codon frequencies and relative synonymous codon usage (RSCU) values were calculated from a dataset of 94,719 codons from 400 genes. RSCU equals the ratio of the observed number of occurrences of a codon divided by the number expected based on random selection of synonymous codons.

species (PF0069, PF00400, PF00005, and PF00078). For example, kinase domains were the most prevalent in *P. infestans,* as in many eukaryotic genomes, as were RNA-dependent DNA polymerases (which were derived mostly from *copia*-like retrotransposons).

Genes involved in pathogenesis and secondary metabolism.

Many genes in true fungi are known to participate in plant infection, based on the criterion that their disruption reduces pathogenesis. Such genes are involved in cell wall degradation, secondary metabolism, signal transduction, and a variety of other functions (Idnurm and Howlett 2001). To examine whether such genes are also in *P. infestans,* a database of 104 fungal pathogenicity (FP) genes was constructed based on the genes cited by Idnurm and Howlett (2001) and compared with the *P. infestans* dataset. In all, 71 FP genes were similar to those in the *P. infestans* unigene set, based on a TBLASTX search with a cutoff of $E = 1 \times 10^{-5}$. Accounting for redundancy in the FP database (multiple MAP kinases are known pathogenicity factors, for example) (Xu 2000), the 71 unique FP genes matched 41 different *P. infestans* sequences (Table 4).

Ascertaining how many of the predicted *P. infestans* pathogenicity factors are authentic homologues of the FP proteins was complicated because low BLAST *E* values often were obtained. This might simply indicate that oomycete proteins are highly diverged from their fungal homologues, or that the BLAST results are misleading. To resolve this, *P. infestans* genes in Table 4 that yielded TBLASTX *E* values $>10^{-20}$ against the FP database were checked against all proteins in GenBank. The results of this analysis are listed as a footnote to Table 4. Most of the *P. infestans* genes appeared to be homologues of FP genes, except those matching the pathogenicity factors involved in secondary metabolism. One example where this analysis supported the role of a *P. infestans* gene as a pathogenicity factor is E7.8809.C1, which weakly matched a *Fusarium* cutinase ($E = 3 \times 10^{-10}$) but strongly matched a bacterial cutinase ($E = 3 \times 10^{-47}$). Another example is PMrpcm4404, which weakly matched a *Magnaporthe* trehelase ($E = 2 \times 10^{-8}$) but strongly matched other trehalases ($E = 4 \times 10^{-35}$). A case where the *P. infestans* gene appeared to not match the pathogenicity factor was E7.10101.C1; although this weakly matched a polyketide synthase ($E = 2 \times 10^{-10}$), it had a stronger match against a protein with a very different function (ornithine decarboxylase; $E = 8 \times 10^{-31}$).

Despite the weak matches against secondary metabolism genes, we tested directly whether *P. infestans* encodes homologues of such genes, in light of the importance they play in the pathogenicity of many true fungi (Keller and Hohn 1997). This entailed searching for relatives of genes involved in the synthesis of fumonisin in *Gibberella moniliformis* (GenBank accession AF155773) (Proctor et al. 2003), sterigmatocystin in *Aspergillus nidulans* (U34740) (Brown et al. 2001), indole-diterpenes in *Penicillium paxilli* (AF279808) (Young et al. 2001), and trichothecene in *Fusarium sporotrichioides* (AF359360 and AF359361) (Brown et al. 2001), as well as seven predicted polyketide synthases and two nonribosomal peptide synthetases from *Neurospora crassa* (Galagan et al. 2003). Each of these genes matched a *P. infestans* sequence in TBLASTX analyses at $E < 10^{-3}$; however, most of the matches were weak. Therefore, evidence for the conservation of major components of these toxin-producing pathways in *P. infestans* was not compelling, in contrast to the conservation of other pathogenicity factors such as cell-wall-degrading enzymes.

Pectate lyases from *P. infestans*.

The genes predicted to encode pectate lyases were examined in more detail. These were selected not simply due to their role in the colonization of plants by bacteria, fungi, and nematodes (Popeijus et al. 2000), but also because the structure, function, and classification of most such enzymes are poorly understood. Of the five (or six; Shevchik et al. 1997) classes of pectate lyases, the four *P. infestans* proteins had the strongest affinity to members of class III. The best BLASTP matches of each *P. infestans* protein were against pectate lyases from *A. nidulans* (GenBank accessions EAA64647 or EAA67075), with 55 to 68% overall amino acid similarity. Slightly weaker matches were observed against the pectate lyases from *F. solani* f. sp. *pisi* that were noted in Table 4. Close affinity to class III pectate lyases also was demonstrated by tree-building exercises using members of the six pectate lyase classes (Fig. 2A). This analysis included pectate lyases from the potato-cyst nematode *Globodera rostochiensis,* the potato ring rot agent *Clavibacter michiganensis* subsp. *sepedonicus,* and the potato soft rot pathogen *Erwinia carotovora* pv. *carotovora;* none of these tightly clustered with a *P. infestans* protein, suggesting that horizontal transfer had not occurred between these potato pathogens.

An amino acid alignment indicated that the *P. infestans* proteins contained the four highly conserved domains characteristic

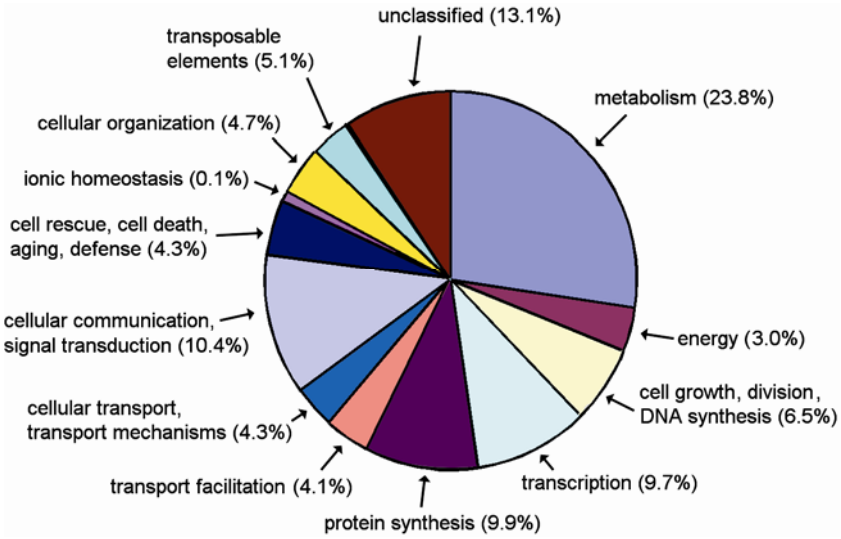

Fig. 1. Functional classification of genes in *Phytophthora infestans*. Unigenes with significant matches ($E < 10^{-5}$) to known sequences based on Pfam and SwissProt annotations were classified according to the Munich Information Center for Protein Sequences (MIPS) system.

of class III pectate lyases (Fig. 2B) (Popeijus et al. 2000). Also, the predicted proteins were cysteine-rich like pectate lyases from other species, with E7.7113.C1, Contig1494, and PXrpxc4412 predicted to encode 10, 8, and 10 cysteines, respectively. It should be noted that the latter appears to be missing 10 to 20 amino acids because a full-length EST was not identified; however, the absent region is not conserved in other pectate lyases and does not impact for the analyses presented here.

Flagellar proteins.

Unigenes predicted to encode components of flagella were identified by searching for matches against such proteins from the green alga *Chlamydomonas reinhardtii.* Of 101 sequences annotated as composing the flagellar proteome in release 2 of its genome sequence (available on-line at the U.S. Department of Energy-Joint Genome Institute), putative homologues were detected against 37 *P. infestans* unigenes (Table 5). In general, this list does not include *P. infestans* sequences which had similarity to *C. reinhardtii* flagellar proteins, but which had stronger matches against distinct proteins in GenBank or those of generic function. An example of the latter would be a Ca^{+2}-ATPase which, although part of the algal flagellar apparatus, may be found in many cellular compartments. The exception was dynein, which exists as both the major constituent of flagella and in cytoplasmic forms.

Cluster analysis based on EST frequency.

The ESTs primarily were generated from non-normalized cDNA libraries; therefore, similarities in expression profiles during the developmental stages and physiological treatments applied to *P. infestans* could be examined by comparing the frequency of ESTs within each library. A hierarchical tree of relatedness was calculated by complete linkage cluster analysis applied to parts-per-million values, calculated from a compiled and normalized data set (Fig. 3).

Many clusters in the tree were consistent with the known biology of *P. infestans,* supporting the validity of the analysis. For example, one cluster grouped libraries of purified zoospores (PV), cleaving sporangia (PF), ungerminated sporangia (PJ), and germinating cysts with appressoria-like structures (PH); because these represent sporelike stages, they would be expected to have overlapping expression profiles. However, the library from directly germinated sporangia (PG) instead clustered with mycelial libraries (e.g., sporulating hyphae [PB]), which suggests that the spectrum of mRNA in sporangia rapidly transitions to that of hyphae soon after germination. Unexpectedly, the mating (PM) library was most similar to the three starvation-related treatments (control in in vitro potato interactions [PX], nitrogen-limited growth [PD], and plich medium [PY]). Other stress- and starvation-related treatments (carbon-limited growth [PE], H_2O_2-stressed mycelium [PL], and heat-stressed

Table 3. Forty most-common domains within proteins predicted from *Phytophthora infestans* unigene set

Pfam	Domain[a]	No. of unigenes[b]	Percent
PF00069[c]	Protein kinase	224	4.7
PF00665	Integrase core	86	1.8
PF00400[c]	WD domain, G-β repeat	73	1.5
PF00005[c]	ABC transporter	70	1.5
PF00078[c]	Reverse transcriptase	63	1.3
PF00169	Pleckstrin homology	56	1.2
PF00076	RNA recognition motif	52	1.1
PF00501	AMP binding enzyme	44	0.9
PF00153	Mitochondrial carrier protein	43	0.9
PF00004	ATPase family	41	0.9
PF00023[c]	Ankyrin repeat	40	0.8
PF00107	Zinc binding dehydrogenase	39	0.8
PF00106	Short chain dehydrogenase	38	0.8
PF00083	Sugar (and other) transporter	36	0.8
PF00226	DnaJ domain	35	0.7
PF00063	Myosin head (motor domain)	33	0.7
PF00097	Zinc finger, C3HC4 type (RING finger)	32	0.7
PF00271	Helicase conserved C terminal domain	32	0.7
PF00270	DEAD/DEAH box helicase	31	0.7
PF01363	FYVE zinc finger	31	0.7
PF03184	DDE superfamily endonuclease	28	0.6
PF00071	Ras family	27	0.6
PF00249	Myb-like DNA binding domain	26	0.5
PF00515	Tetratrico peptide repeat (TPR)	25	0.5
PF00561	α/β Hydrolase fold	25	0.5
PF00481	Protein phosphatase 2C	25	0.5
PF00225	Kinesin motor domain	22	0.5
PF00168	C2 Ca^{2+}-dep. membrane-targeting domain	21	0.4
PF01490	Transmembrane amino acid transporter	21	0.4
PF00171	Aldehyde dehydrogenase family	20	0.4
PF00628	PHD zinc finger-like	20	0.4
PF00856	SET (lysine methyltransferase) domain	19	0.4
PF03028	Dynein heavy chain	19	0.4
PF00098	Zinc knuckle	18	0.4
PF00070	Pyridine nucleotide-disulphide oxidase	18	0.4
PF00632	HECT domain (ubiquitin transferase)	18	0.4
PF00085	Thioredoxin	18	0.4
PF00118	TCP1/cpn60 chaperonin family	18	0.4
PF00012	Hsp70	17	0.4
PF00664	ABC transporter transmembrane receptor	17	0.4

[a] PF designation is the Pfam database designation, not our expressed sequence tag library designation PF.
[b] Number of unigenes matching Pfam domain at $E < 10^{-5}$.
[c] One of the top 20 most common domains in the Pfam database.

mycelium [PK]) clustered elsewhere, suggesting that *P. infestans* responds variably to different stresses.

Highly abundant and stage-specific ESTs.

A suggestion of relative expression levels was provided by the extent of EST redundancy per library. Most of the genes predicted to be abundantly expressed are involved in translation, encoding ribosomal proteins, 16S ribosomal RNA, and translation elongation factor 1-α (Table 6). Genes known to be highly expressed in *P. infestans*, such as the elicitin genes *inf1, inf5,* and *inf6* (Kamoun et al. 1999), and xylitol dehydrogenase (Kim and Judelson 2003) also were represented highly among the ESTs. The GenBank matches, excluding those best-matching oomycete genes, included 15 against genes from animals, 10 from plants, 2 from fungi, and 2 from the glaucocystophytic protistic alga, *Cyanophora.*

Library-specific ESTs were identified by comparing the ESTs from each library with the remaining *P. infestans* ESTs using

Table 4. *Phytophthora infestans* genes resembling known fungal pathogenesis-related genes

***P. infestans* sequence**[a]	**Representative match**[b]	**BLAST *E***
Signal transduction		
E7.4944.C1	*Ustilago maydis* cAMP-dependent protein kinase, AF025290	4e-94
E7.5424.C1	*Cochliobolus carbonum* serine/threonine kinase, AF159253	8e-66
E7.4509.C2	*U. maydis Ukc1* serine/threonine kinase, AF041843	1e-60
E7.1195.C2	*Magnaporthe grisea MPS1* MAP kinase, AF020316	3e-56
E7.7756.C1	*Botryotinia fuckeliana BMP1* MAP kinase, AF205375	9e-54
E7.4312.C1	*M. grisea* cAMP-dependent protein kinase, AF024633	7e-50
E7.3372.C1	*Colletotrichum lindemuthianum clk1* serine/threonine kinase, AF000309	4e-32
E7.4528.C1	*Cryphonectria parasitica cpgb-1* G protein beta subunit, U95139	5e-26
E7.8972.C1	*Glomerella cingulata EMK1* MAP kinase kinase, AF169644	1e-24
Contig1266	*U. maydis* adenylate cyclase, L33918	2e-23
Pigpa1	*Botrytis cinerea bcg1*, G protein alpha subunit, Y18436	7e-12[c]
Primary metabolism		
MY-05-B-09	*Fusarium oxysporum ARG1* arginosuccinate lyase, AB045736	1e-126
E7.4391.C1	*Candida albicans MLS1* malate synthase, AF222907	1e-116
E7.3362.C1	*M. grisea PTH3* imidazole glycerol phosphate dehydratase, AF027980	3e-52
E7.7863.C1	*M. grisea PTH2* carnitine acetyl transferase, AF027979	1e-32
PFrpvb3771	*Cladosporium fulvum aox1* alcohol oxidase, AF275346	6e-21
E7.4304.C1	*M. grisea* polyhydroxynaphthalene reductase, L22309	1e-21
E7.5655.C1	*Colletotrichum lagenarium* trihydroxynaphthalene reductase, P87025	1e-19[c]
PMrpcm4404	*M. grisea PTH9* trehalase, AF027981	2e-08[c]
Cell wall biology		
Contig568	*Aspergillus flavus pecA* endopolygalacturonase, U05015	2e-53
E7.7113.C1	*F. solani* f. sp. *pisi pelD* pectate lyase, U13050	7e-46
Contig1494	*F. solani* f. sp. *pisi pelA* pectate lyase, M94692	1e-38
E7.161.C1	*A. flavus pecA* endopolygalacturonase, U05015	3e-24
E7.9981.C1	*F. solani* f. sp. *pisi pelD* pectate lyase, U13050	2e-20
PXrpxc4112	*F. solani* f. sp. *pisi pelA* pectate lyase, M94692	4e-16[c]
E7.8809.C1	*F. solani* cutinase, M29759	3e-10[c]
Secondary metabolism		
Contig1340	*Cochliobolus carbonum TOXC* putative fatty acid synthase beta, U73650	1e-83
E7.54.C3	*Cladosporium fulvum* pSI-10, Y14556	2e-24
E7.1804.C1	*Gaeumannomyces graminis* var. *avenae* avenacinase, U35463	2e-19[c]
Contig218	*C. fulvum* pSI-9, Y14555	7e-17[c]
E7.7016.C3	*Nectria haematococca* mpVI T9 pisatin demethylase, L20976	4e-16[c]
E7.1478.C1	*Septoria lycopersici* tomatinase, U35462	7e-15[c]
E7.3081.C2	*Cochliobolus carbonum TOXF* branched chain amino acid aminotransferase, AF157629	5e-12[c]
E7.7248.C1	*N. haematococca Pda6-1* pisatin demethylase, X73145	2e-11[c]
E7.10101.C1	*Colletotrichum lagenarium* polyketide synthase, D83643	3e-10[c]
E7.953.C1	*Cochliobolus heterostrophus PKS1* polyketide synthase, U68040	2e-10[c]
Protein fate		
MY-03-C-08	*Glomerella cingulata UBC1* ubiquitin conjugating enzyme, AF030296	2e-73
Transport		
E7.934.C2	*Botryotinia fuckeliana atrB* ABC transporter, AJ006217	6e-82
E7.1999.C1	*Uromyces fabae PIG2* putative permease, U81794	1e-07[c]
Unknown		
E7.3145.C1	*Alternaria alternata* Akt1, AB015351	3e-22
E7.7988.C1	*A. alternata* Akt2, AB015352	1e-09[c]
E7.2246.C1	*G. cingulata* hard-surface inducible protein, AF149296	9e-09[c]

[a] *P. infestans* sequences are named as follows: those starting with "Contig" represent coding regions predicted from genomic DNA, sequences beginning with "E7" represent a consensus derived from the assembly of multiple expressed sequenced tags (ESTs), Pigpa1 represents a previously cloned gene (Latijnhouwers et al. 2004), and the other names represent EST singlets. Of the latter, those beginning with "MY" come from the pilot EST project (Kamoun et al. 1999) while the rest represent singlet ESTs from the current project.

[b] Shown is the member of the fungal pathogenicity (FP) database that provided the best BLASTX match against the indicated *P. infestans* sequence.

[c] *P. infestans* sequences matching proteins in the fungal pathogenicity database with BLASTX *E* values above 1e-20 were searched against GenBank, giving the following matches: PMrpcm4404, *Arabidopsis thaliana* trehalase AAF22127.1, 4e-35; Pigpa1, *Halocynthia roretzi* G-protein α BAC67544.1. 9e-15; E7.5655.C1, rat peroxisomal β-oxidation protein S74209, 9e-69; PXrpxc4112, *Aspergillus nidulans* pectate lyase EAA64647, 5e-24; E7.8809.C1, *Kineococcus radiotolerans* cutinase ZP_00227628, 3e-47; E7.3081.C2, *Ovis aries* amino acid aminotransferase AAG16994.1, 2e-79; E7.1478.C1, *Cytophaga hutchinsonii* β-glucosidase-related protein ZP_00308266.1, 9e-33; E7.10101.C1, *U. maydis* ornithine decarboxylase CAA61274.1, 8e-31; E7.7016.C3, *A. thaliana* cytochrome P450 enzyme NP_171666.1, 8e-40; E7.1804.C1, *C. hutchinsonii* β-glucosidase-related protein, ZP_00309695.1, 7e-40; E7.953.C1, *Nostoc* sp. oxidoreductase BAB76707.1, 2e-65; E7.7248.C1, *A. thaliana* probable cytochrome P450 protein B96662, 7e-66; Contig 218, *Crocosphaera watsonii* aldehyde dehydrogenase ZP_00174906.1, 4e-57; E7.1999.C1, *Dictyostelium discoideum* amino acid permease AAB69390.1, 5e-30; E7.7988.C1, *M. grisea* hypothetical protein XP_368211.1, 3e-09; E7.2246.C1, no other hit.

BLASTN with a cutoff E of 10^{-30} (Table 1). In total, 5,119 putative stage-specific unigenes representing 10,432 ESTs were identified. For example, 1,414 unigenes were specific to germinated cysts with appressoria-like structures, which is a stage important for plant infection. Only 20% of these unigenes matched sequences in the Pfam or SwissProt databases, which is less than the value (43%) obtained for the total unigene set.

In planta-induced genes.

As described above, 963 ESTs from *P. infestans*–plant interaction libraries were predicted to be of *P. infestans* origin. These were compared to a fungus–host interaction (FHI) database made using public ESTs from in vivo and in vitro fungus–plant interactions and appressorial libraries (containing 12,070 ESTs or contigs of ESTs, depending on the source database, constructed as described below). Among the 963 ESTs, 232 resembled sequences in the FHI database by TBLASTN using a cutoff of $E = 1 \times 10^{-5}$.

Of the 963 *P. infestans* ESTs from interaction libraries, 150 were not detected in the cDNA libraries prepared from *P. infestans* grown in vitro (BLASTN cutoff of $E = 1 \times 10^{-30}$) and, therefore, might correspond to infection-specific genes. Of these, 28 had matches in the SwissProt database (Table 7). Of those sequences, 4 were among the 232 showing similarity to sequences in the FHI database (underlined in Table 6) and possibly represent components conserved between fungal and oomycete pathosystems. Among the remaining 122 sequences, 1 EST recently was annotated as a Kazal-like serine protease inhibitor, EPI1, that inhibits and interacts with the tomato subtilisin-like protease P69B (Tian et al. 2004).

Lysine biosynthetic pathways.

One of the early data indicating that oomycetes were distinct from true fungi involved studies of lysine biosynthesis pathways and, consequently, we assessed whether related genes were present in the unigene dataset. Although true fungi synthesize lysine using the α-aminoadipate (AAA) pathway (Zabriskie and Jackson 2000), oomycetes use the diamino-

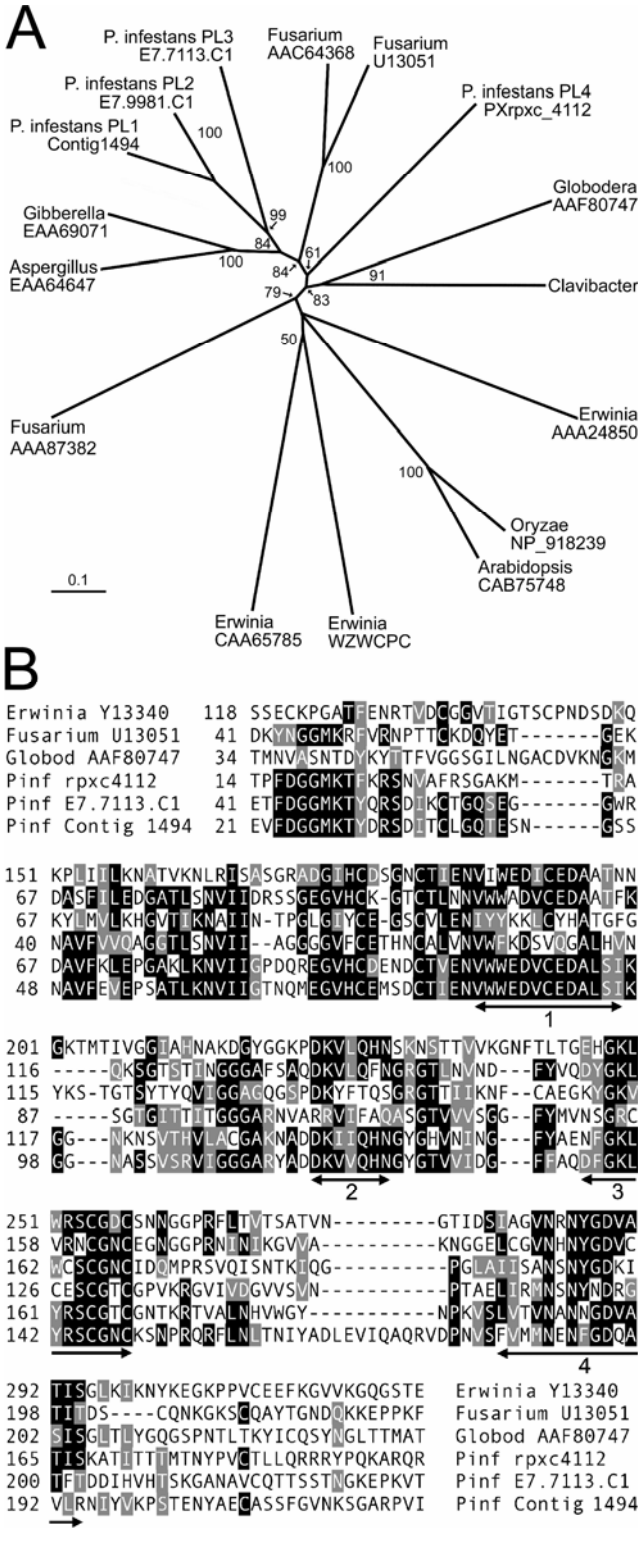

Fig. 2. Phylogram of pectate lyases. **A,** Consensus neighbor-joining tree of predicted *Phytophthora infestans* pectate lyases and other representatives of the superfamily, with numbers at nodes denoting their percent occurrence in 1,000 bootstrap replicates and the scale equaling 0.1 BLOSUM units. Shown in addition to the four *P. infestans* proteins (labeled PL1 to PL4, and with their unigene designator) are homologues from other genera. Except for the *Clavibacter michiganensis* subsp. *sepedonicus* protein, which was derived from data available on-line from the Sanger Institute (anchored by sequence clav-361d12.q1k), the proteins are labeled with their GenBank accession numbers. The latter include representatives of pectate lyase family 1 (from *Arabidopsis thaliana* and *Oryzae sativa*), family 2 (*Erwinia carotovora* subsp. *carotovora* WZWCPC), family 3 (*Globodera rostochiensis, Aspergillus nidulans, Gibberella zeae,* and *Fusarium solani* f. sp. *pisi* U13051 and AAC64368), family 4 (*E. chrysanthemi* AAA24850, known as PelX), family 5 (*E. chrysanthemi* AA65785, known as PelZ), and a unique pectate lyase from *E. chrysanthemi* (CAA73784, known as PelI). **B,** ClustalW alignment of pectate lysases from *P. infestans, E. chrysanthemi, F. solani* f. sp. *pisi*, and *G. rostochiensis*. Black and gray boxes represent identical and semi-conserved blocks of amino acids, respectively. Also indicated are the four conserved domains typically observed in class III pectate lyases.

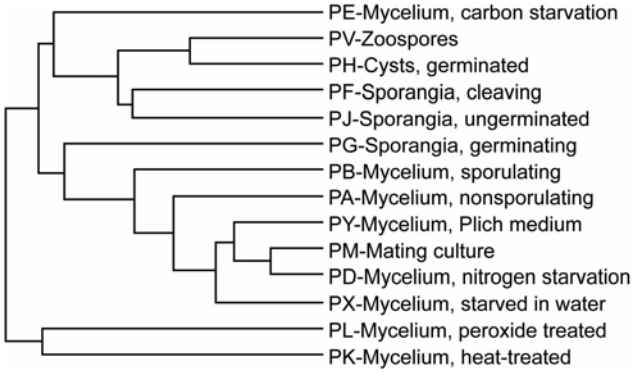

Fig. 3. Hierarchical cluster analysis of developmental stages and physiological treatments. Part-per-million values corresponding to expressed sequence tag frequencies were calculated for each contig across 14 libraries and used for cluster analysis. Library letter codes are followed by a brief description of each library.

pimelate (DAP) pathway like most bacteria and plants (Born and Blanchard 1999). A search of the dataset initially suggested that genes for both pathways might be present, based on similarity searches against genes in the DAP and AAA pathways from *Escherichia coli* and *Saccharomyces cerevisiae*, respectively (Table 8).

However, a more detailed examination suggested that hits against AAA pathway genes generally were misleading. An essential limitation in such analyses is that many of these enzymes belong to large and diverse groups that have similar structural and catalytic domains. For example, although E7.3038.C1 exhibited a fairly strong blast hit ($E = 2 \times 10^{-30}$) against an aminoadipate aminotransferase (E.C.2.6.1.39), its strongest hits were against other aminotransferases such as a putative *Deinococcus radians* valine-pyruvate transaminase ($E = 10^{-64}$; accession NP_294754.1; EC 2.6.1.66). Similarly, although E7.648.C1 matched the *S. cerevisiae* homoisocitrate dehydrogenase (EC.1.1.1.155) at $E = 2 \times 10^{-41}$, a better match was obtained against isopropylmalate dehydrogenase from *Clostridium thermocellum* ($E = 10^{-91}$; ZP_00313248.1; EC 1.1.1.85).

In contrast, although the average matches against the *E. coli* DAP pathway enzymes were weaker than against the AAA pathway proteins, these generally proved to be better predictors of the presence of the pathway. For example, whereas E7.3609.C1 yielded only a fair BLASTX E of 3×10^{-37} against a lysine pathway aspartokinase from *E. coli,* its strongest hits in GenBank were against such enzymes from several plants including *Arabidopsis thaliana* ($E = 4 \times 10^{-95}$; AAF24602.1). Similarly, the putative diaminopimelate decarboxylase MY-10-B-11 best matched a homologue from the archaebacterium *Methanosarcina* ($E = 10^{-104}$; AAM04166.1). Matches were not detected against some genes in the DAP pathway, such as diaminopimelate decarboxylase, but this simply may reflect their absence from the unigene database and not from *P. infestans*.

Data dissemination.

Sequences generated by this project are available through GenBank and the *Phytophthora* Functional Genomics Database (PFGD), available at the National Center for Genome Resources (upon acceptance of manuscript). Data can be queried through

Table 5. *Phytophthora infestans* unigenes matching components of eukaryotic flagella[a]

Matching protein	GenBank accession of *Chlamydomonas reinhardtii* sequence	*P. infestans* sequence	TBLASTN E
Axoneme central apparatus protein PF16	AAC49169.1	E7.4297.C1	3e-19
Calcium-binding protein RIB72	AAM44303.1	E7.6883.C1	2e-25
Centrin (caltractin)	CAA31163.1	E7.3159.C1	4e-64
	P54213	Pi015812	2e-18
Dynein 1-α heavy chain, flagellar inner arm	AJ243806	PF054D2	3e-21
	Q9SMH3	PD034A10	9e-54
	Q9SMH3	rpcy_13321	2e-41
Dynein 1-β heavy chain, flagellar inner arm	CAB99316.1	E7.1801.C1	4e-44
	CAB99316.1	PJ015F6	5e-21
	Q39565	PM040F6	7e-33
	Q39565	Contig6816	4e-44
	Q9MBF8	Contig1646	2e-20
	Q9MBF8	Contig3511	5e-45
	Q9MBF8	Contig7586	5e-42
	Q9MBF8	E7.1137.C1	5e-48
	Q9MBF8	E7.1535.C2	1e-104
	Q9MBF8	E7.4644.C1	5e-27
	Q9MBF8	PDpcd_1271	1e-12
	Q9MBF8	rpcd_1044	6e-29
	Q9MBF8	rpcd_1044	6e-29
Dynein 14-kDa light chain, flagellar outer arm	Q39591	Contig3694	1e-10
Dynein γ chain, flagellar outer arm	Q39575	PYrpcy_8813	4e-37
	Q39575	E7.8627.C1	2e-26
	Q39575	PJ044B2	2e-08
Dynein intermediate chain (IC140), flagellar inner arm	AAD45352.1	E7.8070.C1	8e-18
Dynein intermediate chain 2	Q16959	rpcm_4965	1e-60
Dynein light chain 1	Q22799	PM059D04	3e-10
Dynein light chain LC6, flagellar outer arm	O02414	E7.4320.C2	2e-44
Dynein light chain Tctex1 protein	T07930	E7.1978.C1	8e-25
Dynein light chain, flagellar outer arm	AAD45881.1	E7.6496.C1	4e-20
Dynein regulatory complex, axonemal protein PF2	AAP57169.1	E7.9347.C1	2e-21
Dynein, 70-kDa intermediate chain, flagellar outer arm (IC70)	P27766	E7.1294.C1	7e-07
	P27766	PM042C1	1e-40
Flagellar assembly protein fliH.	P15934	E7.738.C1	1e-19
Flagellar motor switch protein	P24072	rpch_14001	2e-21
Flagellar WD-repeat protein PF20	P93107	rpcd_12981	4e-21
Intraflagellar transport protein IFT88	AAG37228.1	PVrpvb_3675	1e-37
Intraflagellar transport 172 protein (IFT172)	C_170190	E7.9276.C1	3e-73
Kinesin light chain (KLC)	P46825	E7.5111.C1	4e-08
Kinesin-like protein 1 (KLP1)	P46870	PM035C1	5e-09
Kinesin-like protein 2 (KLP2)	P46864	Contig6588	1e-19
Kinesin-like protein 3A (KIF3A)	P28741	Contig3590	5e-17
Microtubule-associated protein EB1	AAO62368.1	E7.7853.C1	2e-36
Osm-6-like intraflagellar transport protein 52	AAK92457.1	PMrpcm_1288	6e-61

[a] *P. infestans* sequences were searched by TBLASTN with proteins annotated as being components of *C. reinhardtii* flagella, using data from version two of its genome sequence which are available on-line from the United States Department of Energy–Joint Genome Institute (DOE-JGI). *P. infestans* sequences that yielded matches only at $E > 10^{-5}$, or which yielded more compelling matches against nonflagellar proteins in GenBank, are not shown. All accession numbers are from GenBank, except for IFT172, which is only identified at the DOE-JGI site.

BLAST searches, and by text searches of the results of BLAST, BLOCKS, and InterPro searches and Gene Ontology (GO) annotations.

DISCUSSION

This work greatly expands the genomics resources available for *P. infestans,* building upon prior studies of this species and its relatives. Previously published EST projects include a study of 1,000 ESTs from *P. infestans* (Kamoun et al. 1999) and 2,500 of *P. sojae* origin (Qutob et al. 2000). The current study is more expansive, not only in terms of the number of cDNA clones sequenced and the companion analysis of genomic DNA but also due to the diversity of developmental stages and growth conditions analyzed. Twenty cDNA libraries representing 18 diverse growth conditions were examined, compared with the prior study of *P. infestans* which involved only 1 growth condition (corresponding to library PY) and the *P. sojae* study which examined only hyphae, zoospores, and infected plant tissue.

Assembly of the EST and genomic sequences generated a set of 18,256 unigenes, which we predict represents a significant proportion of the total from *P. infestans.* However, this number should be treated as preliminary because some weakly expressed genes may have escaped detection as ESTs and because several partial-length unigenes may, in fact, represent the same gene. Also, not all coding regions predicted from genomic DNA may be authentic; microarray studies using probes from tissues similar to those used for the cDNA libraries detected transcripts for only 58% of the 2,330 predicted unigenes (*unpublished results*). Despite these caveats, the extensive nature of the current data set enables improved analyses of the *P. infestans* transcriptome.

The true number of genes in *P. infestans* may be further estimated based on precedents from other major EST projects. For example, when ESTs from *A. thaliana* were compared with its completed genome, EST-derived unigene sets were found to overestimate the actual number of genes by 35% (Arabidopsis Genome Initiative 2000). Furthermore, in large tomato and *A. thaliana* EST projects, only 40 to 80% of genes were shown to be sampled (Arabidopsis Genome Initiative 2000; Van der Hoeven et al. 2002). Guided by these values, *P. infestans* may be predicted to contain (18,256/[100 – 35]) × (100/40 to 100/80) or 22,500 ± 7,500 genes. By comparison, 16,066 genes were predicted in of the genome sequence of *P. ramorum (*annotation 1.0, available on-line). This indicates that oomycetes have a significantly more complex complement of genes than plant-pathogenic true fungi, such as *Magnaporthe grisea,* for which 11,109 genes are reported in its latest genome assembly (version 2.3).

In common with other eukaryotes, the most abundant ESTs were involved in translation. Also like other eukaryotes, the

Table 6. Forty most abundant expressed sequence tags (ESTs) in *Phytophthora infestans*

P. infestans unigene	No. of ESTs	Best non-*P. infestans* BLAST match[a]	*E* value
E8.4149.CB1	500	Xylitol dehydrogenase, AF072541	1e-111
E7.2430.CB1	376	No hit	…
E8.4141.CB1	345	Elicitin-like INF6, AF419843	1e-142
E8.4077.CB1	336	Host-specific elicitor INF1, U50844	9e-94
E8.2637.CB1	299	Ribosomal protein L5, L78668	1e-100
E8.4082.CB1	267	S-phase-specific ribosomal protein, AY062500	1e-102
E8.4702.C2	249	Ribosomal protein L10, AY114542	2e-76
E7.3394.CB2	242	Actin A, M59715	0.0
E8.4203.C1	230	Ribosomal protein L12, NM_070141	4e-66
E8.4156.CB2	226	Ribosomal protein S12, AJ011717	1e-47
E8.4274.CB1	224	Ribosomal protein L15, AF051244	2e-80
E7.1377.CB2	218	Elongation factor 1-alpha, P17786	0.0
E8.4285.CB1	214	Ribosomal protein S23, AF400219	8e-69
E8.4232.C1	210	Ribosomal protein L23, AF401577	2e-78
E7.2570.CB1	208	Heat-shock cognate protein 80, P36181	0.0
E8.4112.CB1	202	Ribosomal protein S27, AF070668	4e-38
E7.2014.CB2	199	Heat-shock 70-kDa protein, P16394	0.0
E8.4146.CB1	197	Ribosomal protein L30, AF063243	4e-30
E7.4075.C4	196	No hit	…
E8.4130.C1	195	Ribosomal protein L11, AJ295006	2e-84
E8.4140.CB1	195	Ribosomal protein L32 (RP49), U66458	8e-44
E8.4195.CB1	193	Ribosomal protein S21, D12633	3e-24
E8.4159.CB1	187	Ribosomal protein L7, AF401559	3e-81
E8.4191.CB1	180	Ribosomal protein S6, XM_125406	3e-98
E8.3990.CB1	179	Ribosomal protein L37a, AF401594	2e-38
E8.4142.C2	178	Ribosomal protein L12, X02633	8e-07
E7.4229.CB1	174	Guanine nucleotide-binding protein, P25387	1e-137
E7.4249.C1	170	No hit	…
E7.4207.C3	168	Ribosomal protein S16, P62249	7e-62
E8.4248.CB1	166	Ribosomal protein L31, AJ005204	3e-32
E8.4287.C1	166	Ribosomal protein L8, NM_119780	1e-134
E7.4105.CB1	165	Ribosomal protein L9, Q963B7	8e-52
E7.2790.CB2	163	Glyceraldehyde 3-phosphate dehydrogenase, P00355	1e-132
E8.4479.C1	160	Ribosomal protein L17 (PETRP), AF307336	6e-58
E8.4123.CB1	154	Acidic ribosomal protein P0, L28704	2e-88
E8.4326.CB1	152	Ribosomal protein L30, XM_122462	8e-46
E8.4268.CB1	149	Ribosomal protein L6, AY062622	7e-34
E8.4384.C1	149	Ribosomal protein S17, NM_126548	1e-42
E8.4213.CB1	145	Elicitin-like INF5, AF419842	1e-105
E7.4076.CB1	145	Ribosomal protein S2, P31009	1e-100

[a] Indicated is the number of ESTs per unigene and their most significant BLASTX hit (with SwissProt or GenBank accession number), using a criterion of $E < 10^{-5}$.

most common Pfam motif was the protein kinase domain (4.7% of *P. infestans* unigenes, versus 2.1% in *N. crassa* and 1.5% in *A. thaliana*). However, several highly abundant ESTs showed no similarity to genes or motifs in public databases and, thus, may encode novel oomycete-specific products. The distribution of common Pfam motifs also diverged between *P. infestans* and other species; only 5 of the 20 Pfam domains most common in other species were in the top 20 Pfam hits in *P. infestans*, likely reflecting the evolutionary distance between *P. infestans* and intensely studied organisms. For example, FYVE zinc fingers were encoded by 0.7% of *P. infestans* unigenes versus 0.1% in both *N. crassa* and *A. thaliana,* while bHLH transcription factors were relatively rare in *P. infestans* at 0.01% versus 0.2 and 0.6% in *N. crassa* and *A. thaliana*, respectively. Similar deviations are observed in comparisons between other eukaryotic kingdoms, especially for transcription factors. For example, MADS box transcription factors form an expanded gene family in *A. thaliana* whereas *Hox* factors are expanded in vertebrates (De Bodt et al. 2003; Duboule 1994).

Consistent with their wide distribution within *P. infestans* (Judelson and Randall 1998), features common to retrotransposons were highly represented because integrase and reverse transcriptases composed 3.3% of identified domains. Of this amount, 1.9% came from EST-based unigenes and 1.4% from coding regions predicted from genomic DNA. However, because 87% of unigenes were derived from ESTs, some predicted retroelements probably were not transcribed. A recent study examining retrotransposons in *Phytophthora* spp. showed drastic interspecific differences in their abundance, implying that such elements have the potential to shape oomycete genomes (Judelson 2002). However, whether the transcribed elements are mobile in *P. infestans* remains to be determined. In contrast to retroelements, only one unigene (E7.8106.C1) with a significant match to a DNA transposon domain was detected. This is consistent with observations that such elements are relatively uncommon in *P. infestans* and largely inactive (Ah Fong and Judelson 2004).

Some features of the transcriptome were consistent with the pathogenic character of *P. infestans*. For example, 70 ABC transporters were detected and the corresponding proteins may help protect *P. infestans* from plant defense molecules (De Waard 1997). ABC transporters represented 1.5% of detected Pfam domains in *P. infestans,* a greater number than in *N. crassa* (0.8%) or *A. thaliana* (0.7%). Also, 39 *P. infestans* sequences were similar to genes known to be essential in the interaction between true fungi and plants, including enzymes that degrade plant cell walls such as pectate lyase, cutinase, and polygalacturonase. All top hits for these enzymes were to fungal sequences. This supports earlier observations that, despite the taxonomic dissimilarity between oomycetes and true fungi, several oomycete enzymes have a closer-than-expected similarity to fungal genes (Gotesson et al. 2002; McLeod et al. 2003; Torto et al. 2002). It remains to be proved whether this represents convergent evolution or horizontal transfer of such genes into or from oomycetes. It was suggested that horizontal transfer explains the presence of related pectate lyases in plant-pathogenic bacteria, fungi, and nematodes (Bird and Koltai 2000; Popeijus et al. 2000). The comparative analysis in the present study, however, suggests that horizontal transfer between different potato pathogens is unlikely.

The availability of the sequences of the *P. infestans* pectate lyases also is useful for addressing structure-function issues within this enzyme superfamily. Pectate lyases are placed into at least five groups: class I contains several bacterial and the plant enzymes; class II includes the periplasmic bacterial pectate lyases; class III contains several bacterial enzymes, and those characterized from nematodes, true fungi, and now *P. in-*

Table 7. *Phytophthora infestans* infection-specific expressed sequence tags (ESTs) with putative homologs[a]

EST name	Best BLAST match	*E* value
PN004A9	DNA topoisomerase III, P13099	8e-49
PC028C5	Probable histone deacetylase, O42227	3e-36
PN003A9	Tyrosine aminotransferase, P04694	5e-34
PN001H3	Fatty acid synthase subunit α, Q10289	8e-30
PC004D3	Isoleucine-tRNA synthetase, Q21926	5e-28
PC028D5	Hypothetical protein At2g18220, Q9ZPV5	1e-24
PC059H4	Uroporphyrinogen decarboxylase, Q9PTS2	3e-22
PN008D11	Poly(A) polymerase β, Q9NRJ5	7e-19
PN009E12	Nonsense-mediated mRNA decay protein, O13824	1e-19
PC058F10	GTP-binding protein TypA/BipA homolog, O25225	9e-19
PN009E12	Nonsense-mediated mRNA decay protein, O13824	1e-19
PC009C6	Photosystem I assembly protein ycf4, P12207	2e-16
PC025G11	Nuclear architecture related protein, P23503	1e-15
PC004D9	NADP-dependent malic enzyme, P37222	2e-14
PN003C12	Glucosamine-fructose-6-phosphate aminotransferase, Q8KG38	3e-14
PN007E12	Copper-transporting P-type ATPase, Q9X5X3	6e-12
PC020F3	Calcium-transporting ATPase 1, O43108	7e-11
PC062F5	Tubulin tyrosine ligase-like protein, O95922	8e-11
PN005A1	β-Galactosidase, P00722	2e-10
PO002A7	Collagen α 1(I) chain precursor, P02457	1e-10
PC023D7	Nonreceptor tyrosine kinase, P18160	3e-09
PO010A7	Internalin A precursor, P25146	3e-09
PC002D3	Cleavage and polyadenylation specificity factor, Q9FGR0	4e-08
PC059G3	Carboxypeptidase KEX1 precursor, P09620	4e-08
PN005F10	Ras-related protein Rab18A, Q05976	3e-08
PO001G5	Chlorophyll A-B binding protein 3C, P07369	8e-08
PC020C8	Myosin VI, Q64331	1e-07
PN010E2	Kinesin-like protein KIF3A, P28741	9e-07
PN002C2	Hypothetical WD-repeat protein Alr3, Q8YRI1	7e-06

[a] Sequences from interaction libraries were filtered to exclude those matching known potato or tomato genes using a cutoff of $E = 10^{-45}$, and then screened to eliminate those present in *P. infestans*-only libraries. Sequences matching records in SwissProt with BLASTX *E* values of less than 10^{-5} are shown; the entire set is presented as supplementary material. Underlined *P. infestans* sequences also matched sequences in the fungal host interaction (FHI) database. The first two letters of each sequence indicate its source library. PO001G5 appears to be pathogen-derived based on its GC content and the absence of a close match from plants, even though its best hit is against a plant gene.

festans; and classes IV and V include other bacterial proteins (Shevchik et al. 1997). Detailed information is available on the structural requirements for activity of class I and class II enzymes; however, such data are limited for members of class III. It was proposed that the catalytic activity of class III pectate lyases depends on one or more amino acids bearing charged side chains within the conserved domains of the proteins (Popeijus et al. 2000). Analyses of previously characterized class III pectate lyases suggested that this role might be played by aspartates in domains 2 and 4, lysines in domains 2 and 3, or both (Fig. 2). However, both lysines are absent in the lyase predicted by PXrpxc4112, implying that catalysis involves the aspartates. Such a conclusion assumes that the PXrpxc4112 product is a functional pectate lyase; although testing this is beyond the scope of this study, the data illustrates the role that oomycete enzymes may play in dissecting the mechanisms of their catalysis.

Other types of genes also provided interesting insights into the evolutionary history of oomycetes. For example, despite the minimal presence of chitin in oomycete cell walls versus that of true fungi (Bartnicki-Garcia and Wang 1983), members of two potential classes of chitin synthases were detected in the EST database (Table 8) based on similarity searches against representatives of each of the classes of fungal chitin synthase genes (Munro and Gow 2001; Roncero 2002). This included an apparent homologue (E7.3733.C1) of a previously identified oomycete chitin synthase (Mort-Bontemps et al. 1997) and a new gene (*rpvb10183*).

One surprising feature relevant to oomycete evolution was the presence of an apparent *FtsZ* homologue in *P. infestans* (unigene E7.758.C1; best match against *Mallomonas splendens* FtsZ AF120116, BLASTX $E = 10^{-35}$). The protein is a key component of the mechanism of division of bacterial cells and probably also of chloroplasts in plants, but not of mitochondria in animals, plants, or true fungi (Erickson 1997; Osteryoung et al. 1998). After finding the FtsZ-like protein in *P. infestans,* homologues also were reported from the draft genome sequence of two oomycete relatives, the diatom *Thalassiosira pseudonana* and the red alga *Mallomonas* (Kiefel et al. 2004). This suggests that these species may employ an ancient mechanism for the division of their mitochondria.

Another unexpected discovery was a family of phosphagen kinases (creatine kinases; Table 8), which are known for their roles in buffering ATP levels or metabolic channeling in animal muscle (Ellington 2001). Until the recent discovery of a phosphagen kinase in trypanosomes (Pereira et al. 2000), such proteins were believed to be restricted to metazoans. Flagellated life-stages are common to oomycetes, animals, and trypanosomes; therefore, it is tempting to speculate that flagella and phosphagen kinases have coevolved, possibly to maintain ATP concentrations in their highly active motile stages. Such kinases may reside in the cytoplasm or mitochondria, or be integral to the flagellar apparatus.

Other likely constituents of flagella also were identified (Table 5). The structure of eukaryotic flagella is highly conserved (Dick 1997). Consequently, it was not surprising that strong matches were obtained against proteins from the green alga *C. reinhardtii* despite the evolutionary distance of oomycetes from plants. Also conserved was the existence of multigene families encoding several flagellar components. For example, in *C. reinhardtii,* 11 genes encode the dynein heavy chain protein of the flagellar inner arm (Porter et al. 1996), and 12 unigenes were detected within *P. infestans.* Kinesins also exist as multigene families in both species. However, some of these may act on cytoplasmic microtubular components instead of, or in addition to, flagella.

Beyond characterizing the protein-coding potential of the *P. infestans* genome, another aspect of this study involved estimating the expression patterns of its genes. This was enabled by the use of primarily non-normalized cDNA libraries to gen-

Table 8. *Phytophthora infestans* genes from notable metabolic pathways

Pathway	*P. infestans* sequences[a]	*E* value
Diaminopimelic acid (DAP) pathway[b]		
dapB, dihydrodipicolinate reductase	E7.3038.C1	2e-10
dapA, dihydrodipicolinate synthase	MY-04-D-03	5e-08
dapD, tetrahydrodipicolinate succinylase	No hit	
dapE, succinyl-diaminopimelate desuccinylase	PMrpcm0801	8e-09
asd, Aspartate semialdehyde dehydrogenase	E7.7375.C1	1e-10
*da*pF, diaminopimelate epimerase	No hit	
lysA, diaminopimelate decarboxylase	MY-10-B-11	1e-38
lysC, lysine-sensitive aspartokinase III	E7.2609.C1	3e-37
α-Aminoadipic acid (AAA) pathway[c]		
LYS20, homocitrate synthase	E7.805.C1	3e-27
LYS4, homoaconitate hydratase	No hit	
LYS2, α aminoadipate reductase	E7.6125.C1	5e-29
LYS9, saccharopine (NAD) dehydrogenase	PDrpcd12884	3e-51
LYS1, sacharopine (NADP+) dehydrogenase	No hit	
LYS12, putative homoisocitrate dehydrogenase	E7.648.C1	2e-41
NM016228 *Homo sapiens* aminoadipate aminotransferase	E7.3038.C1	2e-30
Chitin synthase		
Aspergillus nidulans chsC gene for chitin synthase class I, AB023911	E7.3733.C1	2e-08
A. nidulans chsA gene for chitin synthase class II, D21268	E7.3733.C1	4e-35
A. nidulans chsB gene for chitin synthase class III, D21269	E7.3733.C1	5e-39
A. nidulans chsD gene for chitin synthase class IV, D83246	E7.3733.C1	1e-12
A. nidulans csmA gene for chitin synthase class V, AB000125	PFrpvb11935	5e-17
Saprolegnia monoica chitin synthase, U19946	E7.3733.C1	7e-51
Phytophthora capsici chitin synthase, U42304	PFrpvb10183	2e-08
Achlya ambisexualis chitin synthase, U55044	PFrpvb10183	2e-10
Phosphagen kinases		
Ictalurus punctatus creatine kinase, AAO25755	E7.5355.C1	8e-85
Danio rerio creatine kinase, AAH49529	E7.3128.C3	1e-107

[a] All unigenes in this table are considered homologs based on having at least one Pfam domain in common with the top match shown.
[b] Genes from the *Escherichia coli* genome database hosted by SRI International (Menlo Park, CA, U.S.A.).
[c] From the *Saccharomyces* Genome Database at Stanford University (Palo Alto, CA, U.S.A.), except for NM_016228.

erate ESTs from diverse physiological and developmental conditions. Cluster analysis of EST frequency data, although potentially limited by the small sizes of some data sets, revealed clear and predicted distinctions between various mycelial treatments and spore-related developmental stages, such as sporangia, zoospores, and germinating cysts (Fig. 3). These analyses also indicated that asexual spores are among the most differentiated developmental stages in *Phytophthora* spp. Interestingly, sporangia producing germ tubes through direct germination did not cluster near germinating zoospore cysts. This may reflect either the rapid transition of germinating sporangia to the hyphal state or the presence of genes related to appressoria in the germinating cyst stage. Carbon-starved, peroxide-treated, and heat-stressed hyphae also were markedly distinct from other mycelial stages and different from each other, indicating that the transcriptome exhibits variable responses to different sources of stress.

EST distributions also were analyzed to reveal library-specific ESTs indicative of putative stage-specific genes. In all, 5,119 such consensus sequences were identified, including 1,682 from stages of the zoospore pathway which is important for plant infection. The proportion of putative stage-specific *P. infestans* proteins with matches in the Pfam or SwissProt databases was lower than the overall unigene assembly (20 versus 43%); therefore, such proteins may participate in novel aspects of oomycete development. Stage-specific genes were most prevalent in the starvation, mating, and cyst germination libraries, suggesting that such tissues will be most informative in future oomycete EST projects. The EST distributions also suggested which genes are developmentally regulated. Because this has been an in silico analysis, the datasets only represent a first screen for truly stage-specific genes. However, the differential expression of several predicted spore-specific genes has been confirmed (Ah Fong and Judelson 2003; Judelson and Roberts 2002).

The *P. infestans* EST set described here is the largest yet described for a plant-pathogenic microorganism. Because it encompasses a wide range of growth, development, and infection stages, it will be of value in furthering the understanding of this economically and historically important organism and its relatives. Other sizable EST datasets also have been generated from plant-pathogenic true fungi (Ebbole et al. in press; Soanes et al. 2002). Comparisons between the *P. infestans* and true fungal datasets revealed overlap, including in many genes implicated in pathogenesis and in 20% of the *P. infestans* genes expressed in planta. Areas of little similarity also were detected, such as the absence from *P. infestans* of genes with high similarity to those used by fungi to synthesize toxins (Kroken et al. 2003). In the future, analyses of the complete genome sequence of *P. infestans*, comparative genomics analysis, and functional studies will reveal additional common and nonconserved mechanisms used by these diverse filamentous eukaryotes to colonize their hosts.

MATERIALS AND METHODS

Strains and growth conditions used for cDNA libraries.

Unless otherwise specified, cultures were grown in the dark at 18°C using either rye A agar (Caten and Jinks 1968) or rye broth. The latter was prepared by clarifying liquid rye A by centrifugation for 10 min at 7,500 × *g*. The strain used for genomic libraries was T30-4, an F_1 of 80029 (A1, The Netherlands) and 88133 (A2, The Netherlands). Strain 88069 (A1, The Netherlands) was used for cDNA library construction, except for mating cultures where 88069 was paired with 618 (A2, Mexico). The name of each library and growth conditions are listed below and are summarized in Table 1.

PA (nonsporulating hyphae)—clarified rye broth was inoculated with asexual sporangia and grown for 6 days to a presporulation stage.

PB (sporulating hyphae)—a polycarbonate membrane placed on top of rye agar was spread with sporangia and incubated for 11 days, at which time profuse sporulation was present.

PC, PI (tomato interactions)—strain 90128, which establishes a biotrophic interaction with tomato, was inoculated on cv. OH7814 using a zoospore suspension. After 3 days, tissue was isolated from the center of the lesion for library PC, or from tissue surrounding the lesion for library PI.

PD (nitrogen-starved hyphae)—a culture was initiated in defined liquid medium (Xu 1982) and then grown for 2 days in the same medium with 5% of the normal level of nitrogen. This reduced growth by two-thirds compared with complete defined medium.

PE (carbon-starved hyphae)—a culture was initiated in defined liquid medium (Xu 1982) and then grown for 2 days in the same medium with 5% of the normal level of carbon. This reduced growth by two-thirds compared with complete defined medium.

PF (cleaving sporangia)—sporangia were harvested from 13- to 15-day-old rye agar plates using ice-cold sterile water, filtered through a 50-µm nylon mesh to remove hyphal fragments, and then incubated on ice for 2 h until most sporangia were in the cleaving stage (i.e., early zoosporogenesis).

PG (germinating sporangia)—sporangia were harvested from rye agar plates as for library PF, diluted 1:1 with ALBA medium (Bruck et al. 1980) to 1.5×10^5 per ml, and incubated for 48 h at 18°C. The resulting germinating sporangia were collected on 50-µm mesh.

PH (germinating cysts)—zoospores were induced by placing sporangia at 4°C, encysted by vortexing, and allowed to germinate in water for 16 h.

PJ (sporangia)—sporangia were obtained as described for library PF, but were quickly frozen to prevent differentiation into zoospores.

PK (heat-stressed hyphae)—a nonsporulating culture similar to that used for library PA was incubated at 26°C for 24 h. At this temperature, growth was reduced by approximately 60% relative to the optimal temperature of 20°C.

PL (peroxide-treated hyphae)—two broth cultures as used for library PA were supplemented with 0.01 and 0.001% H_2O_2, respectively, and pooled after 4 h.

PM (mating)—two strips of inoculum from strains 88069 and 618 were placed, at a distance of 2.5 cm, on opposite sides of a 100-mm petri dish containing rye agar overlaid with a polycarbonate membrane. Tissue from the mating zone was harvested after 6 days.

PN, PO (unsubtracted potato interactions)—as described (Beyer et al. 2001), zoospores were spot-inoculated onto leaves from cv. Bintje and, after 6 days, tissue was harvested from within and outside the lesion for libraries PN and PO, respectively.

PU, PW, PX (in vitro potato interactions)—as described (Beyer et al. 2001), mycelia were grown in rye broth for 6 days, washed with water, and placed in water or on a potato leaf floating in water to generate control and tester samples, respectively. Both control and tester samples were pooled from 4-, 8-, and 24-h time-points. A subtracted library (PU) then was constructed from the control and tester RNA using the polymerase chain reaction-select cDNA subtraction kit (BD Biosciences, Palo Alto, CA U.S.A.), and nonsubtracted libraries were made from the tester (PW) and control (PX).

PV (zoospores)—a 13-day rye agar culture was flooded with ice-cold water and incubated at 4°C until zoospore release was maximal. The zoospores were purified by passage through a 10-μm mesh and pelleted at 5,000 × *g* for 5 min.

PY (Pilch medium hyphae)—a nonsporulating culture in pilch medium was harvested after 4 weeks; this resembles conditions used for a pilot EST project (Kamoun et al. 1999) except strain 88069 was used here.

Cloning and DNA sequencing.

For all cDNA libraries except PU, PW, and PX, polyT-primed cDNAs were directionally cloned in pSPORT1 using the Superscript system (Invitrogen, Carlsbad, CA, U.S.A.). Libraries PU, PW, and PX employed pGEM3+ (Promega Corp., Madison, WI U.S.A.). Average inserts in each unsubtracted library were 1.0 to 1.2 kb. Sequencing was done by Lark Technologies (Houston, TX, U.S.A.) or the Beijing Genomics Institute (BGI) (Beijing, China). All clones were initially sequenced from their 5′ ends, and a subset was also sequenced from their 3′ ends, yielding 833 high-quality 3′ reads.

For genomic sequencing, DNA from strain T30-4 (Whisson et al. 2001) was cloned into pUC18 and sequenced at BGI. Inserts for these libraries included DNA sheared by sonication and size selected on an agarose gel to 1.0- to 1.5- and 1.5- to 3.0-kb fractions (for libraries ping and pinx, respectively), which were repaired with T4 DNA polymerase, T4 kinase, and Klenow DNA polymerase. Other inserts included 1.0- to 1.5-kb and 3-kb fractions of DNA digested with *Hin*dIII (libraries pind and pine) or 1.0- to 1.5-kb and 3-kb fractions of DNA digested with *Eco*RI (libraries pina and pinb).

EST assembly and identification of genomic coding regions.

Sequences were base-called using the Phred program (Ewing et al. 1998) and assembled with Paracel Transcript Assembler (v. 3.6.1). Input included approximately 114,500 ESTs plus genomic sequences predicted to have high coding potential; the latter were identified as described (Dwyer 2002) by screening the 588,801 genomic reads using a statistical model of codon bias developed using a training set of 79 known *P. infestans* genes, followed by trimming to minimize the possibility of including introns. The default settings for CAP4 were used and all ESTs were trimmed to have >200 bp of sequence with Phred score >20. Repetitive sequences (usually <50 bp and >5,000 copies) escaping initial vector masking were identified during early iterations of the assembly and masked in subsequent assemblies.

Codon usage analysis.

Codon usage frequencies and RSCU values were calculated from 400 predicted full-length ORFs using the general codon usage analysis (GCUA) program (McInerney 1998). Effective N_c calculations (Wright 1990) were determined using the CodonW program, which is available online from the Institut Pasteur.

Database searches.

Datasets were compared with a locally installed NCBI-BLAST2.1 algorithm (Altschul et al. 1990). Comparisons between *P. infestans* and non-*P. infestans* data were performed at the amino-acid level using the appropriate BLAST program, usually with an *E* value cut-off of 10^{-5}. The only exception involved searches of interaction libraries for plant sequences, which were performed at the nucleotide level using an *E*-value threshold of 10^{-45}.

SwissProt, GenBank, and Pfam databases were searched at either Syngenta Biotechnology, Inc, Research Triangle Park, NC, U.S.A.; Syngenta Jealots Hill International Research Centre, Bracknell, Berkshire, UK; or the University of California–Riverside, U.S.A. The FHI database was complied from 4,838 ESTs from wheat spikes infected with *F. graminarium* (BM134271-138704, BM140307-140614, and BM259010-259098) (Kruger et al. 2002), 4,908 ESTs from germinating conidia of *Blumeria graminis* and barley infected with that species (Thomas et al. 2001); 1,168 ESTs from rice infected with *M. grisea* (AW154962-155632 and AW069892-070183) (Kim et al. 2001; Rauyaree et al. 2001), 1,148 ESTs from pine infected with *Heterobasidium annosum* (BM346837-347258) (Karlsson et al. 2003), and 1,156 ESTs from wheat infected with *Mycosphaerella graminicola* (AW067761-067765 and AW179955-181107). The FP database of validated genes involved in fungus–plant pathogen interaction was compiled from the literature (Idnurm and Howlett 2001).

Cluster analysis based on EST frequency.

To determine the frequency of ESTs, a file containing contig identifiers was constructed that indicated the number of ESTs per contig and their distribution across libraries. A cut-off of ≥10 ESTs per contig was applied to select moderately expressed sequences across all libraries. Next, focusing on *P. infestans*-only libraries and those with good representation (>1,000 ESTs), the numbers of ESTs per library were normalized by converting them into parts-per-million values using the formula $(10^6/N_a) \times E_a$, where N_a = number of ESTs sequenced per library "a" and E_a = number of unique ESTs identified in EST dataset "a". Cluster analysis then was performed using the Cluster/Treeview package (Eisen et al. 1998). To obtain normal distributions within the data set, log transformations were performed on parts-per-million values, and log-transformed data were subjected to self-organizing maps (SOM) analysis. Output files of SOM analysis were used to order the contigs as well as the libraries, and the ordered files then were used as input files for hierarchical clustering. Complete linkage clustering analysis was performed on the rearranged files and the resulting output files were evaluated.

Comparisons of pectate lyases.

Alignments and distance-based trees were generated using CLUSTALW and Phylip (Felsenstein 1989), available on-line from the DNA Databank of Japan, using 1,000 bootstrap replicates and the BLOSUM option. Alignments were formatted using BOXSHADE, available on-line from the Swiss Institute of Bioinformatics.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support provided by Syngenta and its predecessor, the agribusinesses of Novartis; and Adam Burkholder (University of North Carolina, Chapel Hill) and Thomas Girke (University of California–Riverside) for assistance with data management.

LITERATURE CITED

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnik, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu., A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. 1991. Complementary DNA sequencing: expressed sequence tags and the human genome project. Science 252:1651-1656.

Ah Fong, A., and Judelson, H. S. 2003. Cell cycle regulator *cdc14* is expressed during sporulation but not hyphal growth in the fungus-like oomycete *Phytophthora infestans*. Mol. Microbiol. 50:487-494.

Ah Fong, A., and Judelson, H. S. 2004. The *hat*-like DNA transposon *DodoPi* resides in a cluster of retro and DNA transposons in the stramenopile *Phytophthora infestans*. Mol. Gen. Genomics 271:577-585.

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.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of

the flowering plant *Arabidopsis thaliana*. Nature 408:796-815.
Bartnicki-Garcia, S., and Wang, M. C. 1983. Biochemical aspects of morphogenesis in *Phytophthora*. Pages 121-137 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.
Beyer, K., Binder, A., Boller, T., and Collinge, M. 2001. Identification of potato genes induced during colonization by *Phytophthora infestans.* Mol. Plant Pathol. 2:125-134.
Bird, D. M., and Koltai, H. 2000. Plant parasitic nematodes: Habitats, hormones, and horizontally-acquired genes. J. Plant. Growth Regul. 19:183-194.
Born, T. L., and Blanchard, J. S. 1999. Structure/function studies on enzymes in the diaminopimelate pathway of bacterial cell wall biosynthesis. Curr. Opin. Chem. Biol. 3:607-613.
Brasier, C. M., Cooke, D. E. L., and Duncan, J. M. 1999. Origin of a new *Phytophthora* pathogen through interspecific hybridization. Proc. Natl. Acad. Sci. U.S.A. 96:5878-5883.
Brown, D. W., McCormick, S. P., Alexander, N. J., Proctor, R. H., and Desjardins, A. E. 2001. A genetic and biochemical approach to study trichothecene diversity in *Fusarium sporotrichioides* and *Fusarium graminearum*. Fungal Genet. Biol. 32:121-133.
Bruck, R. I., Fry, W. E., and Apple, A. E. 1980. Effect of metalaxyl an acyl alanine fungicide on developmental stages of *Phytophthora infestans.* Phytopathology 70:597-601.
Caten, C. E., and Jinks, J. L. 1968. Spontaneous variability in isolates of *Phytophthora infestans*. I. Cultural variation. Can. J. Bot. 46:329-348.
De Bodt, S., Raes, J., Van de Peer, Y., and Theissen, G. 2003. And then there were many: MADS goes genomic. Trends Plant Sci. 8:475-483.
De Waard, M. A. 1997. Significance of ABC transporters in fungicide sensitivity and resistance. Pestic. Sci. 51:271-275.
Dick, M. W. 1997. Fungi, flagella and phylogeny. Mycol. Res. 101:385-394.
Duboule, D. 1994. Guidebook to the homeobox genes. Oxford University Press, Oxford.
Duret, L., and Mouchiroud, D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis, Drosophila, and Arabidopsis*. Proc. Natl. Acad. Sci. U.S.A. 96:4482-4487.
Dwyer, R. 2002. Genomic Perl. Cambridge University Press, Cambridge.
Ebbole, D. J., Jin, Y., Thon, M., Pan, H., Bhatterai, E., Thomas, T., and Dean, R. Gene discovery and gene expression in the rice blast fungus *Magnaporthe grisea:* Analysis of expressed sequence tags. Mol. Plant-Microbe Interact. 12: 1337-1347.
Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. 95:14863-14868.
Ellington, W. R. 2001. Evolution and physiological roles of phosphagen systems. Annu. Rev. Physiol. 63:289-325.
Erickson, H. P. 1997. Ftsz, a tubulin homologue in prokaryotic cell division. Trends Cell Biol. 7:362-367.
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.
Felsenstein, J. 19889. Phylip-phylogeny inference package (version3.2) Cadistics 5:164-166.
Fry, W. E., and Goodwin, S. B. 1997. Re-emergence of potato and tomato late blight in the United States. Plant Dis. 81:1349-1357.
Galagan, J. E., and others. 2003. The genome sequence of the filamentous fungus *Neurospora crassa.* Nature 422:859-868.
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.
Huitema, E., Torto, T. A., Styer, A., and Kamoun, S. 2003. Combined ESTs from plant-microbe interactions: Using GC counting to determine the species of origin. Pages 79 to 83 in: Plant Functional Genomics: Methods and Protocols. E. Grotewold, ed. Humana Press, Totowa, NJ, U.S.A.
Idnurm, A., and Howlett, B. J. 2001. Pathogenicity genes of phytopathogenic fungi. Mol. Plant Pathol. 2:241-255.
Judelson, H. S. 2002. Sequence variation and genomic amplification of a family of *gypsy*-like elements in the oomycete genus *Phytophthora*. Mol. Biol. Evol. 19:1313-1322.
Judelson, H. S., and Randall, T. A. 1998. Families of repeated DNA in the oomycete *Phytophthora infestans* and their distribution within the genus. Genome 41:605-615.
Judelson, H. S., and Roberts, S. 2002. Novel protein kinase induced during sporangial cleavage in the oomycete *Phytophthora infestans*. Eukaryot. Cell 1:687-695.
Judelson, H. S., Spielman, L. J., and Shattock, R. C. 1995. Genetic mapping and non-Mendelian segregation of mating type loci in the oomycete, *Phytophthora infestans*. Genetics 141:503-512.
Kamoun, S. 2003. Molecular genetics of pathogenic oomycetes. Eukaryot. Cell 2:191–199.
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.
Kamoun, S., Van West, P., Vleeshouwers, V. G. A. A., De Groot, K. E., and Govers, F. 1998. Resistance of *Nicotiana benthamiana* to *Phytophthora infestans* is mediated by the recognition of the elicitor protein INF1. Plant Cell 10:1413-1425.
Karlsson, M., Olson, A., and Stenlid, J. 2003. Expressed sequences from the basidiomycete tree pathogen *Heterobasidium annosum* during early infection of Scots pine. Fungal Genet. Biol. 39:51-59.
Keller, N. P., and Hohn, T. M. 1997. Metabolic pathway gene clusters in filamentous fungi. Fungal Genet. Biol. 21:17-29.
Kiefel, B. R., Gilson, P. R., and Beech, P. L. 2004. Diverse eukaryotes have retained mitochondrial homologues of the bacterial division protein Ftsz. Protist 155:105-115.
Kim, K. S., and Judelson, H. S. 2003. Sporangia-specific gene expression in the oomyceteous phytopathogen *Phytophthora infestans*. Eukaryot. Cell 2:1376-1385.
Kim, S., Ahn, I.-P., and Lee, Y.-H. 2001. Analysis of genes expressed during rice-*Magnaporthe grisea* interactions. Mol. Plant-Microbe Interact. 14:1340-1346.
Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C., and Turgeon, B. G. 2003. Phylogenomic analysis of type i polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc. Natl. Acad. Sci. U.S.A. 100:15670-15675.
Kruger, W. M., Pritsch, C., Chao, S., and Muelbauer, G. J. 2002. Functional and comparative bioinformatics analysis of expressed genes from wheat spikes infected with *Fusarium graminearum*. Mol. Plant-Microbe Interact. 15:445-455.
Lander, E. S., and Waterman, M. S. 1988. Genomic mapping by fingerprinting random clones: A mathematical analysis. Genomics 2:231-239.
Latijnhouwers, M., de Wit, P. J. G. M., and Govers, F. 2003. Oomycetes and fungi: Similar weaponry to attack plants. Trends Microbiol. 11:462-469.
Latijnhouwers, M., and Govers, F. 2003. A *Phytophthora infestans* G-protein β subunit is involved in sporangium formation. Eukaryot. Cell 2:971-977.
Latijnhouwers, M., Ligterink, W., Vleeshouwers, V. G. A. A., van West, P., and Govers, F. 2004. A G-α subunit controls zoospore motility and virulence in the potato late blight pathogen *Phytophthora infestans*. Mol. Microbiol. 51:925-936.
Margulis, L., and Schwartz, K. V. 2000. Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth. W. H. Freeman and Company, New York.
McInerney, J. O. 1998. Gcua (general codon usage analysis). Bioinformatics 14:372-373.
McLeod, A., Smart, C. D., and Fry, W. E. 2003. Characterization of 1,3-β-glucanase and 1,3;1,4-β-glucanase genes from *Phytophthora infestans.* Fungal Genet. Biol. 38:250-263.
Mewes, H. W., Frishman, D., Guldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Munsterkotter, M., Rudd, S., and Weil, B. 2002. MIPS: A database for genomes and protein sequences. Nucleic Acids Res. 30:31-34.
Mort-Bontemps, M., Gay, L., and Fevre, M. 1997. *Chs2,* a chitin synthase gene from the oomycete *Saprolegnia monoica.* Microbiology 143:2003-2020.
Munro, C. A., and Gow, N. A. R. 2001. Chitin synthesis in human pathogenic fungi. Med. Mycol. 39S:41-53.
Osteryoung, K. W., Stokes, K. D., Rutherford, S. M., Percival, A. L., and Lee, W. Y. 1998. Chloroplast division in higher plants requires members of two functionally divergent gene families with homology to bacterial Ftsz. Plant Cell 10:1991–2004.
Pereira, C. A., Alonso, G. D., Paveto, M. C., Iribarren, A., Cabanas, M. L., Torres, H., N., and Flawia, M. M. 2000. *Trypanosoma cruzi* arginine kinase characterization and cloning: A novel energetic pathway in protozoan parasites. J. Biol. Chem. 275:1495-1501.
Popeijus, H., Overmars, H., Jones, J., Blok, V., Goverse, A., Helder, J., Schots, A., Bakker, J., and Smant, G. 2000. Degradation of plant cell walls by a nematode. Nature 406:36-37.
Porter, M. E., Knott, J. A., Myster, S. H., and Farlow, S. J. 1996. The dynein gene family in *Chlamydomonas reinhardtii*. Genetics 144:569-585.
Proctor, R. H., Brown, D. W., Plattner, R. D., and Desjardins, A. E. 2003. Co-expression of 15 contiguous genes delineates a fumonisin biosynthetic gene cluster in *Gibberella moniliformis*. Fungal Genet. Biol. 38:237-249.

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

Randall, T. A., and Judelson, H. S. 1999. Construction of a bacterial artificial chromosome library of *Phytophthora infestans* and transformation of clones into *P. infestans*. Fungal Genet. Biol. 28:160-170.

Rauyaree, P., Choi, W., Fang, E., Blackmon, B., and Dean, R. A. 2001. Genes expressed during early stages of rice infection with the rice blast fungus *Magnaporthe grisea*. Mol. Plant Pathol. 2:347-354.

Rizzo, D. M., Garbelotto, M., Davidson, J. M., Slaughter, G. W., and Koike, S. T. 2002. *Phytophthora ramorum* as the cause of extensive mortality of *Quercus* spp. and *Lithocarpus densiflorus* in California. Plant Dis. 86:205-214.

Roncero, C. 2002. The genetic complexity of chitin synthesis in fungi. Curr. Genet. 41:367-378.

Ronning, C. M., Stegalkina, S. S., Ascenzi, R. A., Bougri, O., Hart, A. L., Utterbach, T. R., Vanaken, S. E., Riedmuller, S. B., White, J. A., Cho, J., Pertea, G. M., Lee, Y., Karamycheva, S., Sultana, R., Tsai, J., Quackenbush, J., Griffiths, H. M., Restrepo, S., Smart, C. D., Fry, W. E., Van Der Hoeven, R., Tanksley, S., Zhang, P., Jin, H., Yamamoto, M. L., Baker, B. J., and Buell, C. R. 2002. Comparative analyses of potato expressed sequence tag libraries. Plant Physiol. 131:419-429.

Sharp, P. M., Tuohy, T. M. F., and Mosurski, K. R. 1986. Cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14:5125-5143.

Shevchik, V. E., Robert-Baudouy, J., and Hugouvieux-Cotte-Pattat, N. 1997. Pectate byase *Pel1* of *Erwinia chrysanthemi* 3937 belongs to a new family. J. Bacterial. 179:7321-7330.

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

Sogin, M. L., and Silberman, J. D. 1998. Evolution of the protists and protistan parasites from the perspective of molecular systematics. Int. J. Parasitol. 28:11-20.

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.

Tian, M., Huitema, E., da Cunha, L., Torto-Alalibo, T., and Kamoun, S. 2004. A kazal-like extracellular serine protease inhibitor from *Phytophthora infestans targets* the *tomato pathogenesis*-related protease p69b. J. Biol. Chem. 279:26370-26377.

Tooley, P. W., and Therrien, C. D. 1987. Cytophotometric determination of the nuclear DNA content of 23 Mexican and 18 non-Mexican isolates of *Phytophthora infestans*. Exp. Mycol. 11:19-26.

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.

Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G., and Tanksley, S. 2002. Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell 14:1441-1456.

van der Lee, T., Testa, A., Robold, A., van 't Klooster, J. W., and Govers, F. 2004. High density genetic linkage maps of *Phytophthora infestans* reveal trisomic progeny and chromosomal rearrangements. Genetics 167:1643-1661.

Werres, S., Marwitz, R., Man In'T Veld, W. A., e Cock, A. W. A. M., Bonants, P. J. M., De Weerdt, M., Themann, K., Ilieva, E., and Baayen, R. P. 2001. *Phytophthora ramorum* sp. Nov., a new pathogen on rhododendron and viburnum. Mycol. Res. 105:1155-1165.

Whisson, S. C., van der Lee, T., Bryan, G. J., Waugh, R., Govers, F., and Birch, P. R. J. 2001. Physical mapping across an avirulence locus of *Phytophthora infestans* using a highly representative, large-insert bacterial artificial chromosome library. Mol. Gen. Genomics 266:289-295.

Wright, F. 1990. The 'effective number of codons' used in a gene. Gene 87:23-29.

Xu, J.-R. 2000. Map kinases in fungal pathogens. Fungal Genet. Biol. 31:137-152.

Xu, R. 1982. A defined media for *Phytophthora*. Acta Mycol. Sin. 1:40-47.

Young, C., McMillan, L., Teifer, E., and Scott, B. 2001. Molecular cloning and genetic analysis of an indole-diterpene gene cluster from *Penicillium paxilli*. Mol. Microbiol. 39:754-764.

Zabriskie, T. M., and Jackson, M. D. 2000. Lysine biosynthesis and metabolism in fungi. Nat. Prod. Rep. 17:85-97.

AUTHOR-RECOMMENDED INTERNET RESOURCES

U.S. Department of Energy-Joint Genome Institute genome portal website: genome.jgi-psf.org

National Center for Genome Resources *Phytophthora* functional genomics database: www.pfgd.org