Curr Genet (2007) 52:115–123
DOI 10.1007/s00294-007-0144-z

RESEARCH ARTICLE

Alternate intron processing of family 5 endoglucanase transcripts from the genus *Phytophthora*

Stefano Costanzo · Manuel D. Ospina-Giraldo · Kenneth L. Deahl · C. Jacyn Baker · Richard W. Jones

Received: 18 April 2007 / Revised: 3 July 2007 / Accepted: 9 July 2007 / Published online: 28 July 2007

Abstract Twenty-one homologs of family 5 endo-(1–4)-β-glucanase genes (EGLs) were identified and characterized in the oomycete plant pathogens *Phytophthora infestans*, *P. sojae*, and *P. ramorum*, providing the first comprehensive analysis of this family in *Phytophthora*. Phylogenetic analysis revealed that these genes constitute a unique eukaryotic group, with closest similarity to bacterial endoglucanases. Many of the identified EGL copies were clustered in a few genomic regions, and contained from zero to three introns. Using reverse transcription PCR to study in vitro and in planta gene expression levels of *P. sojae*, we detected partially processed RNA transcripts retaining one or more of their introns. In some cases, the positions of intron/exon splicing sites were also found to be variable. The relative proportions of these transcripts remain apparently unchanged under various growing conditions, but differ among orthogolous copies of the three *Phytophthora* species. The alternate processing of introns in this group of EGLs generates both coding and non-coding RNA isoforms. This is the first report on *Phytophthora* family 5 endoglucanases, and the first record for alternative intron processing of oomycete transcripts.

Keywords Endoglucanase · Intron splicing · Gene duplication · *Phytophthora infestans* · *Phytophthora sojae* · *Phytophthora ramorum*

Communicated by D. Ebbole.

Electronic supplementary material The online version of this article (doi:10.1007/s00294-007-0144-z) contains supplementary material, which is available to authorized users.

R. W. Jones (✉)
USDA/ARS, 10300 Baltimore Avenue,
Bldg. 010A, Room 311, Beltsville, MD 20705, USA
e-mail: Richard.Jones@ars.usda.gov

S. Costanzo
Department of Plant Science and Landscape Architecture,
University of Maryland, College Park, MD 20742, USA

M. D. Ospina-Giraldo
Department of Biology, Lafayette College,
Easton, PA 18042, USA

K. L. Deahl
Vegetable Laboratory, USDA-ARS,
10300 Baltimore Ave., Beltsville, MD 20705, USA

C. J. Baker
Molecular Plant Pathology Laboratory, USDA-ARS,
10300 Baltimore Ave., Beltsville, MD 20705, USA

Introduction

Family 5 endoglucanases from the phytopathogenic genus *Phytophthora* have not yet been characterized. Endoglucanases within the family 5 glycoside hydrolases show hydrolytic activity toward glucans containing β-(1,4) linkages, such as those found in cellulose. Various members of this family have economically important industrial applications in food and fiber processing, and for this reason, they have been isolated and characterized from a wide range of organisms. They are also of interest due to their presence in most, if not all, plant pathogenic fungi, as well as plant endoparasitic nematodes, highlighting their potential role in pathogenesis.

The *Phytophthora* family 5 endoglucanase ESTs are detected during pathogenesis, having appeared in microarray studies of *P. sojae*-infected soybean tissue, where it

shows strong up-regulation in the early stages of infection (Moy et al. 2004). To date, the only other β-(1,4) endoglucanase known to be strongly up-regulated in the soybean pathosystem belonged to the family 12 endoglucanases that were recently characterized (Costanzo et al. 2006). Cell wall modifying enzymes are required for invasive growth by oomycetes, as hyphal forces are below the level of tensile strength of the plant cell wall (Money et al. 2004). Very likely, endoglucanases play a role in invasive growth by *Phytophthora*, by helping to loosen the microfibril structure contributing to tensile strength.

Using the family 5 EST that was up-regulated during soybean infection, we surveyed the draft genome sequences for *P. sojae* and *P. ramorum* (Tyler et al. 2006) and the raw sequence trace files from the whole genome shotgun project (WGS) of *Phytophthora infestans*, to identify new members of EGL5 in *Phytophthora*, and provide proper annotation.

In this study we describe the cloning, characterization, and genomic organization of 21 putative family 5 endoglucanase genes from *P. infestans, P. ramorum*, and *P. sojae*. These EGL5 genes differed with respect to the number of introns possessed and their genomic localization. We utilized this information to provide more specific characterization of gene expression in vitro and in planta. While analyzing the expression level of *P. sojae* EGL5 genes, we identified various alternatively processed transcripts among homologous members containing one or more introns.

Materials and methods

Culturing of *Phytophthora* spp

Three *Phytophthora* species were used in this study: *P. sojae*, *P. infestans*, and *P. ramorum*. The *P. sojae* isolate (race 25) was kindly provided by Dr. S. Li at the National Soybean Pathogen Collection Center, Department of Crop Sciences, University of Illinois. A *P. infestans* isolate was obtained from the collection maintained at the Vegetable Laboratory, Beltsville Agricultural Research Center. Genomic DNA and total RNA from *P. ramorum* was kindly provided by Dr. M. Palm and Dr. M. C. Aime in the Systematic Mycology and Botany Laboratory at the Beltsville Agricultural Research Center. All isolates were routinely cultured on rye agar medium supplemented with 2% sucrose (Caten and Jinks 1968).

Nucleic acid manipulation

Genomic DNA and total RNA were extracted from *Phytophthora* mycelium grown in pea-broth (filtrate from 120 g of autoclaved frozen peas, supplemented with 2 g of calcium carbonate and 0.05 g of β-sitosterol per liter of medium) for ~2 weeks at 22°C in the dark before harvesting. Total RNA was obtained, in the sole case of *P. sojae*, from cultures grown on modified Plich liquid medium (Van der Lee et al. 1997). Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA) following the protocol supplied by the manufacturer. Genomic DNA was extracted using the GenElute Plant Genomic DNA Miniprep Kit (Sigma, St. Louis, MO), following the instructions provided by the supplier.

RT-PCR

All RT-PCR reactions were conducted using 5 µg of DNase-treated (TURBO DNA-free, Ambion, Austin, TX) RNA, and reverse transcription (SuperScript III One-Step RT-PCR Kit, Invitrogen, Carlsbad, CA) was performed using gene-specific primers. A control reaction (minus RTase) was used to exclude spurious amplifications due to potential presence of contaminating DNA. Transcript levels of *P. sojae* actin (*ActA*), based on JGI's predicted gene model (estExt_fgenesh1_pm.C_490003), were monitored during infection assays using the specific primers 5′-CG ACCAGGAGATGAAGAC-3′ and 5′-CCAGATTCTAA GACGAGTTG-3′ (forward and reverse, respectively), and 56.7°C for PCR annealing temperature. The expected amplicon size for the actin target is 497 bp. An additional internal control was obtained by designing intron-spanning primers, based on the gene model of a putative cellulose synthase-like glycosyltransferase identified in the three *Phytophthora* spp. genomes. Forward and reverse primer sequences for the internal control are 5′-CTCGGGT TCCTCTACTAC-3′ and 5′-TCGTTGTCCAGAATGA TG-3′ in *P. sojae*, 5′-TCGTGGTGCTCGTTATCTG-3′ and 5′-GCCGTGGTAGTGGTAGTC-3′ in *P. infestans*, and 5′-CCGCAGTGGTATCATTGG-3′ and 5′-GCCGTG GTAGTGGTAGTC-3′ in *P. ramorum*. The expected amplicon sizes are, respectively, 702, 518, and 664 bp including the intron and 608, 442, and 600 bp without it. Annealing temperature for these primer sets is 57.5°C. Primers were designed, for *P. sojae* and *P. ramorum*, based on sequences extracted from gene models estExt_ fgenesh1_pg.C_840015 and fgenesh1_pg.C_scaffold_8300 0038 from assembly 1.1 of their respective genome projects. For the putative homologue gene in *P. infestans* primers were designed based on sequence from the supercontig 47: 317874–318413 from assembly 1 of its genome project.

Primers were designed using the Primer3 web site (Rozen and Skaletsky 2000), for the direct amplification of each individual EGL open reading frame (ORF). For those copies with ORFs interrupted by introns, specific sets of primers were also designed targeting a region including the

intron and, when multiple introns were present, to amplify the entire region across all of the introns. Amplification products were analyzed using 1.5% agarose gel electrophoresis. RT-PCR products were excised from the gel or directly purified using PureLink PCR Purification Kit (Invitrogen) following the manufacturer's protocol. The purified DNA amplicons were cloned into pCR4-TOPO vector (Invitrogen-Life Technologies). Transformed *Escherichia coli* colonies were screened by PCR and positive clones were sequenced.

Sequence analysis

Phytophthora sojae and *P. ramorum* sequences of putative endoglucanases were identified by searching for conserved domains of EGL5 β-(1,4) endoglucanase protein using BlastP analysis (*E*-value $1e^{-5}$), against the Version 1.0 of their complete genome data produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/). *P. infestans* copies were subsequently identified by searching the NCBI Trace Archive Databases with Discontiguous MegaBLAST, and comparing our annotated *P. sojae* and *P. ramorum* EGL5 nucleotide sequences against the raw sequence trace files available at NCBI, http://www.ncbi.nlm.nih.gov/blast/tracemb.shtml, from the *P. infestans* WGS by the Broad Institute. The matching sequence trace files were visually analyzed to remove low quality sections and assembled into contigs using Vector NTI Advance 10. The presence of introns was determined by aligning and comparing genomic and cDNA sequences for those genes where both sequences were available. Moreover, the automated annotation of the *P. infestans* genome was recently released by The Broad Institute (March 2007), allowing to BLAST search and view annotated features of our manually assembled EGL5 sequences (http://www.broad.mit.edu/annotation/genome/phytophthora_infestans).

Alignments, phylogenetic and molecular evolutionary analyses were conducted using *MEGA* Version 3.0 (Kumar et al. 2004). Alignments were formatted using *BOXSHADE* Version 3.21, available on-line from the Swiss Institute of Bioinformatics. Motif and signal peptide predictions were carried out utilizing software developed by the Center for Biological Sequence Analysis at the Technical University of Denmark DTU (http://www.cbs.dtu.dk/).

Infection assay

Inoculation of 2-week-old soybean [*Glycine max* (L.) Merr.] cv. Williams seedlings was carried out as previously described (Costanzo et al. 2006). Excised tissue was flash frozen in liquid nitrogen and stored at −80°C for subsequent total RNA extraction.

Results

Genomic organization and sequence analysis

Sequence analysis of *P. sojae*, *P. ramorum*, and *P. infestans* revealed the presence of family 5 β-(1,4) endoglucanase (EGL5)-coding sequences in multiple copy number (Fig. 1). The relative location and organization of the EGL5 genes were conserved within the *P. sojae* and *P. ramorum* genomes. In *P. sojae* a total of nine predicted functional EGL5 copies (named PsEGL-V.1 up to PsEGL-V.9) were identified, while *P. ramorum* had eight (named PrEGL-V.1 up to PrEGL-V.5, PrEGL-V.7 up to PrEGL-V.9) and *P. infestans* had six (named PiEGL-V.1 up to PiEGL-V.3, PiEGL-V.5, PiEGL-V.9, and PiEGL-V.10).

Analysis of deduced amino acid sequence

The predicted ORFs for all the identified EGL gene copies were confirmed using BioEdit software (Hall 1999). Based on those predictions, nucleotide sequences of each individual EGL copy were translated into amino acid sequences. These genes encode proteins varying in length from 428 to 646 amino acid residues. Predicted protein sequence analysis from all EGL gene copies, identified in the three genomes, revealed a high percentage of identical amino acid residues, with sequence variation mostly occurring at the N-terminus region of the molecule (Supplementary Fig. 1).

Implementing a combination of prediction models from the SignalP 3.0 server (Nielsen et al. 1997; Bendtsen et al. 2004), presence of a signal peptide cleavage site was not supported in any of the analyzed sequences. However, when subjected to screening using the Secretome P2.0 algorithm (Bendtsen et al. 2005), which is used to identify non-classical protein secretion patterns in bacteria, highly significant scores (ranging from 0.84 to 0.94 for each sequence) were obtained. The multiple sequence alignment identifies three members (PiEGL-V.9, PrEGL-V.9, and PsEGL-V.9) as being clearly distinct from the others, particularly by having a 70 amino acid long serine/threonine-rich region at their 3′ ends. In addition, PsEGL-V.4 and PsEGL-V.6 appear to be unusually short copies with respect to all other copies of EGL5, and could potentially represent two different pseudogenes in this group.

RNA processing

Eleven of the EGL5 genes identified in *Phytophthora* spp. contained one or more introns. Electrophoretic analysis of RT-PCR samples, obtained using primers specific for the individual intron-encompassing regions, revealed in all cases the presence of two main amplicons (data not shown). These products were subsequently gel purified and cloned.

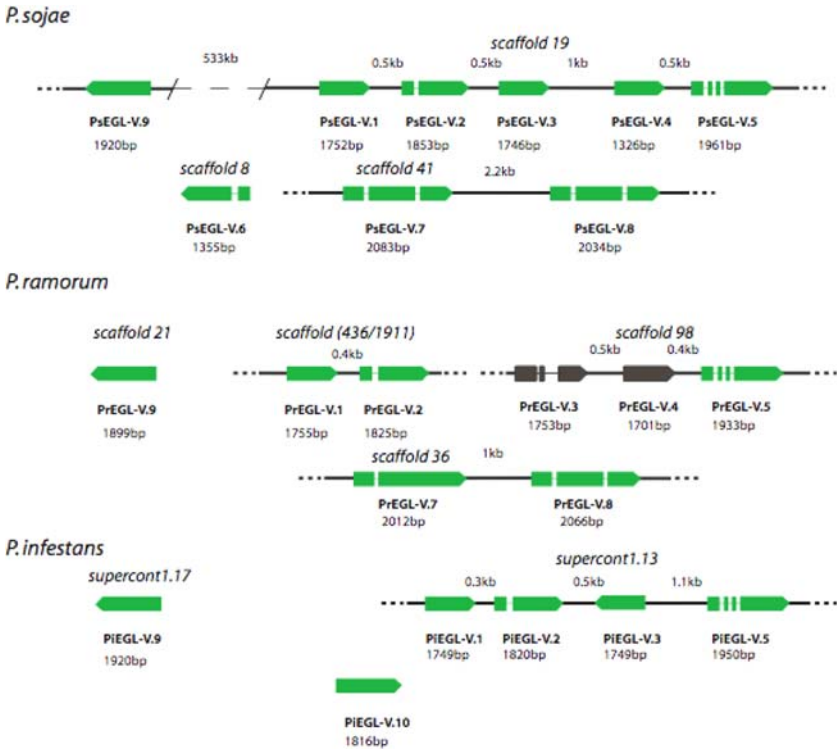

Fig. 1 Diagram of *Phytophthora sojae*, *P. ramorum*, and *P. infestans* family 5 endoglucanases. Relative distances of adjacent paralogs are indicated above the *solid line* representing a scaffold section. A *horizontal thin line joining filled green boxes* representing exons indicates the presence of introns. *Brown boxes* represent computationally predicted EGL members for which position and sequence could not be confirmed by the present study. The individual sequence lengths are reported below the EGL copy name

Sequence analysis from these clones confirmed the presence of both fully and non-processed cDNAs corresponding to each intron-containing EGL copy. To further investigate the presence of splice variants and their potential effect on transcripts from this group of endoglucanases, we selected a specific member present in all three *Phytophthora* spp. in this study (PsEGL-V.5, PrEGL-V.5, and PiEGL-V.5) characterized by the presence of three introns. Using primers specifically designed to target regions harboring multiple introns, various products were amplified and separated by electrophoresis in a 2.0% agarose gel (Fig. 2). After gel extraction of the region containing the bands, the purified products were cloned into pSC-A StrataClone (Stratagene) to make three separate cDNA clone libraries. Ninety-six colonies from each cDNA library were randomly selected and screened by colony PCR on the basis of their product size. Sequence analysis of these cloned cDNA fragments revealed transcripts retaining one or more of their introns, and intron splicing sites were found in one or more alternate positions. Alternate splice variants were found to be produced at both the 5′ and the 3′ ends of the exon/intron junctions (see positions termed α and β in Fig. 3). The 5′ exon/intron junction variant (β) was present in transcripts for both *P. sojae* and *P. infestans*. Based on these findings, we analyzed the effect on the predicted ORFs caused by the presence or absence of every possible combination of the three intronic regions and the alternative position for the donor/acceptor splicing sites. In

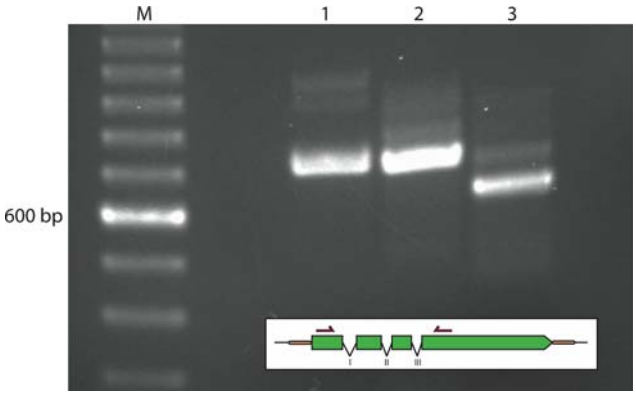

Fig. 2 RT-PCR analysis of EGL-V.5 genes from the three *Phytophthora* species, using specific primers across the complete intron-containing region as represented in the gene model diagram. The prominent band in each reaction lane represents completely processed transcript. *M*: 100 bp ladder; 1: PsEGL-V.5; 2: PrEGL-V.5; 3: PiEGL-V.5

the majority of cases the detected splice variants caused a premature termination of their transcripts. Interestingly, ORFs of various lengths, still retaining the complete conserved region of family 5 EGLs, could be found downstream from those early termination codons. Sequence analysis of the cloned fragments from the intron containing regions did not provide evidence for the existence of all theoretically possible combinations of transcripts.

Fig. 3 Partial 5′ nucleotide sequence of *P. sojae* EGL-V.5. *Italicized lowercase letters* indicate predicted intron sequence. α: alternative position of the 3′ intron/exon boundary. β: alternative position of the 5′ exon/intron boundary. α and β variants were identified in separate transcripts

Transcriptional and phylogenetic analysis of *Phytophthora* spp. EGL5

Expression studies of the various members of EGL5 genes focused on *P. sojae*, and consisted of experiments using reverse transcription-PCR with gene-specific primers, targeting all nine EGL5 genes in vitro and in planta (Fig. 4). Among these nine copies, only seven were actively transcribed and clearly detectable in total RNA extracted from 2-week-old mycelium grown in vitro. Expression of two members, characterized by a relative short sequence, PsEGL-V.4 and PsEGL-V.6, were not detected. Studies with infected soybean tissue revealed high expression of PsEGL-V.1, PsEGL-V.2 PsEGL-V.3, and PsEGL-V.8 at 72 h after inoculation, while PsEGL-V.5, PsEGL-V.7, and PsEGL-V.9 had weak expression, relative to actin levels. Again, expression of PsEGL-V.4 and PsEGL-V.6 was not detected using various primer combinations.

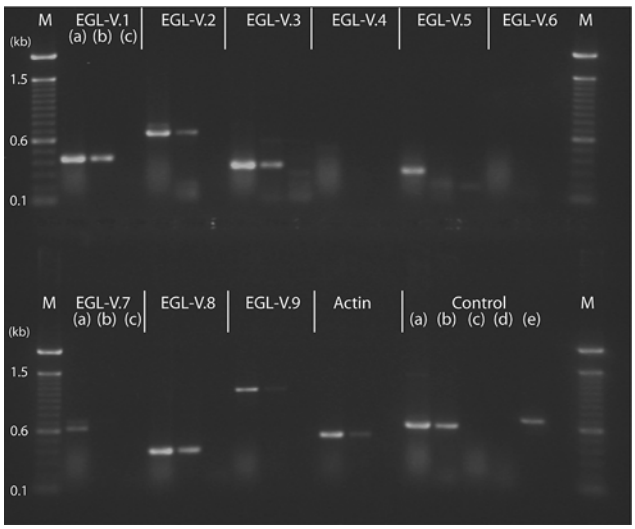

Fig. 4 RT-PCR analysis using gene-specific primers of *P. sojae* EGL-V.1 to 9 and actin on total RNA from: **a** *P. sojae* culture mycelium; **b** *P. sojae*-infected soybean tissue; **c** mock-inoculated soybean tissue. Primer sets used for PsEGL-V.2 and 7 were located across a single intron region, revealing the alternate processing. *M*: 100 bp ladder; Controls utilized primers spanning an intron region of a *P. sojae* glycosyltransferase: Control (**a**) *P. sojae* mycelial RNA, treated with DNase (RTase + Taq); **b** *P. sojae*-infected soybean tissue RNA, treated with DNase (Rtase + Taq); **c** mock-infected soybean tissue RNA, treated with DNase (Rtase + Taq); **d** *P. sojae* mycelial RNA, treated with DNase (+Taq); **e** *P. sojae* DNA (+Taq)

A RT-PCR experiment was also conducted to directly monitor the amplification of transcript regions encompassing each individual intron present in *P. sojae* EGL-V.5 in three different growing conditions (Fig. 5). Total RNA was extracted from *P. sojae* mycelium grown on liquid pea broth and Plich medium, as well as from infected soybean tissue, at 72 h after inoculation. The results of this experiment showed ∼20% of amplicon products still retaining their intron and no significant changes, among the three regions, in the intron processing due to changes in the growing conditions.

To investigate the relationship between *Phytophthora* spp. EGL5 and different homologs from other taxa, we performed standard protein–protein BLAST (blastp) searches against NCBI Reference Sequence (RefSeq) collection. Matches with an Expect (*E*) value smaller than $3e^{-10}$ were selected for further studies. A phylogenetic analysis was conducted using the Neighbor-joining distance matrix method with a 1,000 bootstrap replications to examine the reliability of the interior branches (Felsenstein 1985). Our analysis indicated that the *Phytophthora* spp. endoglucanase family 5 genes could be grouped into four major branches (Fig. 6). The branch most closely related to bacterial cellulases contained the homologous members PsEGL-V9, PrEGL-V9, and PiEGL-V9. The second branch included PsEGL-V8, PrEGL-V7, and PiEGL-V8, which are characterized by the presence of two introns located closer to the 5′ and 3′ ends of their transcripts, and PrEGL-V7 which contains a single intron closer to the 5′ end. A third branch separated members containing three introns (PsEGL-V5 PrEGL-V5 PiEGL-V5) from the remaining copies. The eukaryotic sequences most closely related to *Phytophthora's* EGL5 were from

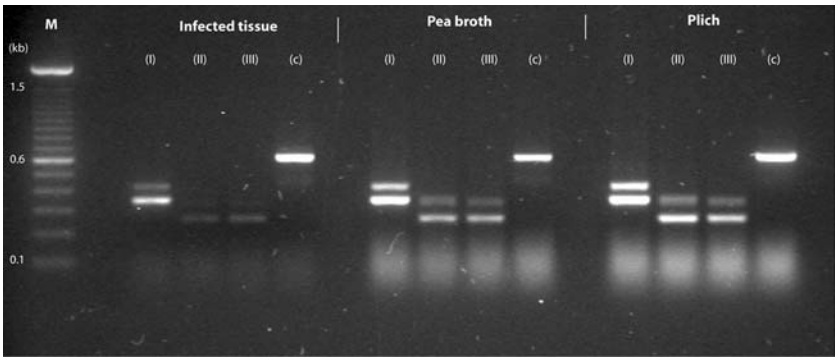

Fig. 5 RT-PCR analysis of individual introns of PsEGL-V.5 from infected soybean tissue, Pea broth and Plich liquid culture, using specific primers across each individual intron-containing region (I–III). Controls utilized primers spanning an intron region of a *P. sojae* glycosyltransferase (**c**). *M*: 100 bp ladder

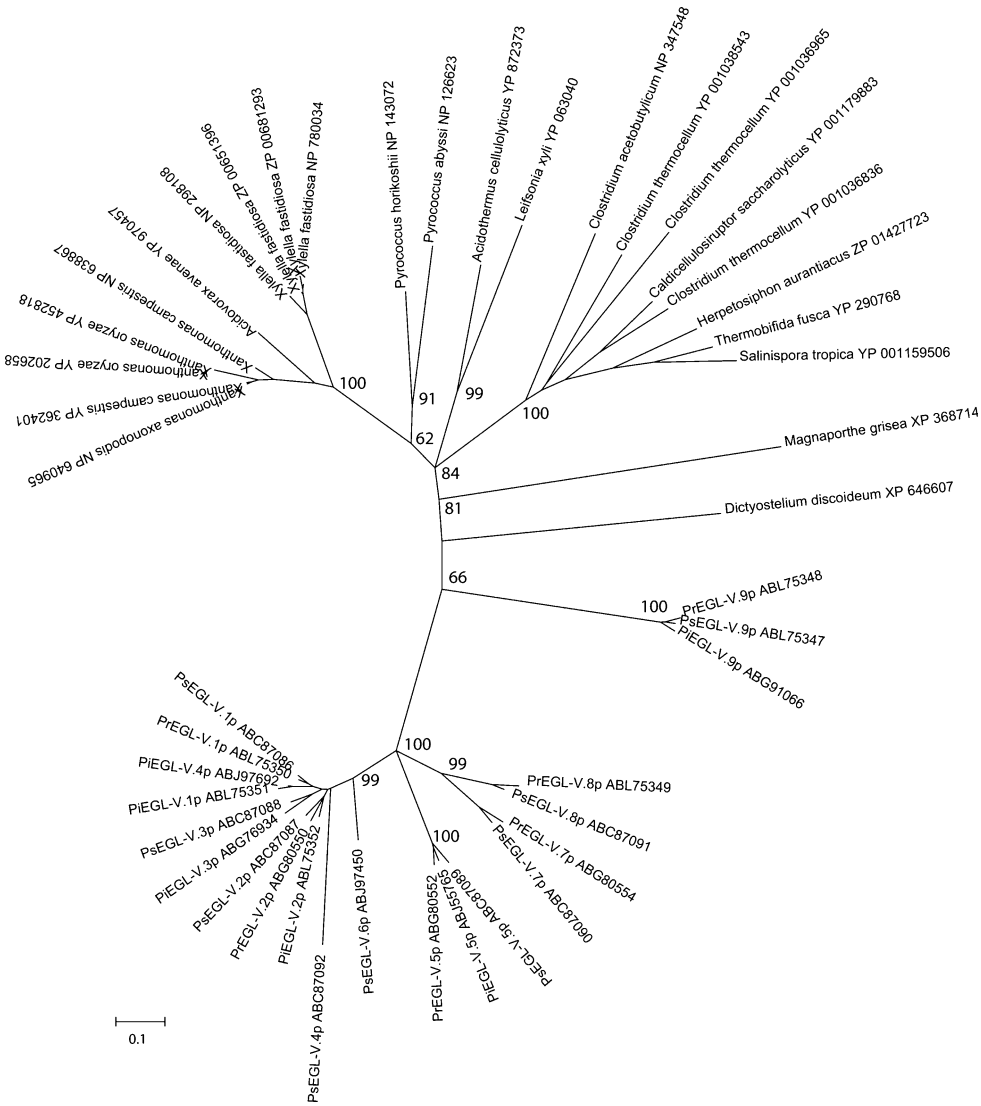

Fig. 6 Unrooted Neighbor-joining phylogenetic tree of EGL5 protein members identified in *Phytophthora* and other species. The GenBank accession numbers of reported sequences follow the organism name. The *scale bar*, indicating 10% sequence variation per branch length, is shown under the tree, and bootstrap values (1,000 replicates) are shown for selected nodes

Dictyostelium discoideum [GenBank accession no. XP646607 (*E*-value $7e^{-22}$)] and *Magnaporthe grisea* [GenBank accession XP368714 (*E*-value $4e^{-11}$)]. The most closely related bacterial endoglucanases were those from the genera *Mycobacterium*, *Acidothermus*, *Leitsonia*, and *Pyrococcus*.

While alternate intron processing was commonly observed for some of the intron-containing EGL5 genes, preliminary results from 11 other disparate intron-containing genes from *Phytophthora* showed efficient and consistent intron processing, leaving a single transcript of expected size and translational product (Fig. 7).

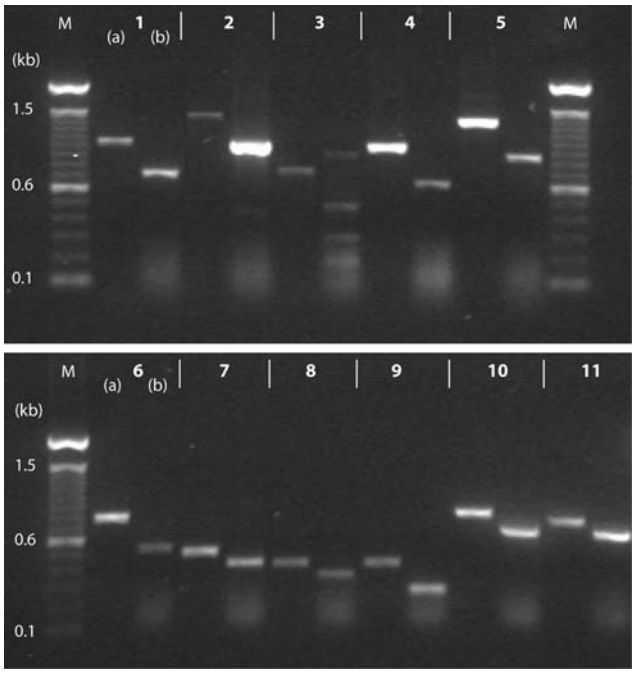

Fig. 7 Analysis of *P. sojae* intron-containing genes. Primers were chosen that amplify the region surrounding an intron(s). PCR amplicons from genomic DNA template loaded in lane (**a**) and from cDNA template in lane (**b**). No alternate processing was detected in the randomly selected genes. Gene primer sequences are listed in supplementary material. *Lane 1*: Isopenicillin *N* synthase; 2: Serine/threonine protein kinase; 3: Activator of heat shock 90 kDa protein ATPase homolog; 4: Ribosomal protein S13; 5: Aldehyde dehydrogenase; 6: Di-*trans*-poly-*cis*-decaprenylcistransferase; 7: Nucleotidyl transferase; 8: Chitin synthase; 9: Small cysteine rich protein similar to SCR76; 10: Protein phosphatase 2C subfamily; 11: Pleckstrin homology-type. *M*: 100 bp ladder

Discussion

Family 5 glycosyl hydrolases comprise a range of hydrolytic activities; however, in this study we have focused on one specific group of family 5 endoglucanases that has been well characterized in other organisms and shown to be active toward β-1,4 glycosyl linkages (Gilkes et al. 1991a, b; Henrissat 1991a, b; Sakon et al. 1996; Henrissat and Davies 2000; Davies et al. 2005). Proteins encoded by the genes identified in this study would be classified as high molecular weight family 5 endoglucanases (50–60 kD range). Structurally, the proteins have the conserved catalytic domain typical for EGL5, and lack a cellulose-binding domain. The lack of a cellulose-binding domain was noted in the first EGL5 cloned from a phytopathogenic fungus (Wang and Jones 1995a, b, 1997, 1999) and remains the structural paradigm for phytopathogen-encoded endoglucanases.

The multiplicity of *Phytophthora* family 5 endoglucanase genes is unusual for phytopathogens. This may be a reflection of the high copy numbers for various gene families encoding cell wall modifying enzymes in *Phytophthora*, more than the specific amplification of these endoglucanase genes. While genes previously characterized from *Phytophthora* have generally shown homology to plant or fungal genes, the group reported in this study is clearly homologous to bacterial genes (Fig. 6). Gene transfer from bacterial sources seems unlikely due to the presence of introns in half of the *Phytophthora* EGL5 genes. Moreover, we find multiple copies of EGL5, unlike the very limited copy number of specific endoglucanase genes generally seen in other organisms. This extensive copy number follows the pattern previously identified for family 12 endoglucanases from *Phytophthora* spp. (Costanzo et al. 2006).

A significant attribute of the *Phytophthora* EGL 5 proteins encoded by the genes in this study is the lack of a clear signal peptide for extracellular secretion. However, the high values for non-classical protein secretion suggest that the proteins are in fact extracellular, as would be expected for endoglucanases. Whether they are membrane-bound or secreted into the environment remains to be tested. The finding that *Phytophthora* EGL5 proteins are closest to bacterial endoglucanases may help explain the differences in protein secretion mechanisms, and brings a new perspective to protein secretion studies in *Phytophthora*, which could result in an incomplete picture when standard algorithms were used to detect the presence of a signal peptide sequence: *Phytophthora* proteins scored as not having a signal peptide, may indeed be extracellular. Further research in immunolocalization of the EGL 5 proteins will provide additional insight into this issue.

An interesting feature separating the multiplicity of EGL5 and EGL12 genes is the presence of introns in certain members of the EGL5 clusters. None of the 19 EGL12 paralogs from *P. sojae*, *P. ramorum*, and *P. infestans* (GenBank accession no. DQ782379) previously identified contain introns (Costanzo et al. 2006). Among the 21 EGL5 paralogs in this study, five contain one intron, three contain two introns, and three contain three introns (Fig. 1). Intron conservation is seen between the adjacent genes PsEGL-V.7 and PsEGL-V.8, while PsEGL-V.2 (one intron) and PsEGL-V.5 (three introns) are interspersed in a cluster of three other EGL5 genes lacking introns. Thus, while the copy number is high, many of the EGL5 copies do not seem to have recently arisen from duplication. Given the synteny and individual gene intron conservation between EGL5 genes from *P. sojae*, *P. ramorum*, and *P. infestans*, the endoglucanase origins would precede speciation, suggesting that introns have remained stable within these genes during this time. A clear example in support of this hypothesis comes from the direct comparison of PsEGL-V.5, PrEGL-V.5, and PiEGL-V.5 genomic and transcript sequences.

Introns may have been lost prior to speciation, considering the positive 5′ bias of current introns. Using intragenic

models, Roy and Gilbert (2005) identified a bias toward loss of introns at the 3′ end of genomic copies, lending support to the idea of reverse transcriptase editing of these introns. The favorable bias toward 5′ introns becomes more pronounced in eukaryotes that have few introns in their genome (Mourier and Jeffares 2003). Intron loss has been supported, while intron gain has not. In a study of three divergent eukaryotes (*Drosophila*, *Caenorhabditis*, and *Arabidopsis*) there were no examples of intron transposition into non-homologous genes (Fedorov et al. 2003). This supports, once again, the idea that introns found in *Phytophthora* EGL5 genes would have preceded speciation. A recent report of intron-rich genes in the marine annelid *Platynereis*, equivalent to the level found in homologous human genes, provides a highly unexpected example of intron-retention (Raible et al. 2005).

Intron-containing genes had been considered rare within *Phytophthora* species. Current genome information suggests that they may be more common than previously thought, although still present in a minority of genes (Tyler et al. 2006). The true number of introns in the *Phytophthora* genomes may not be known for many years unless efforts are made to test their validity. Studies on the authenticity of predicted introns in the genome of *Saccharomyces cerevisiae*, revealed cases of misidentified and undetected introns, as well as examples of alternative splicing that was meiotic-specific (Davis et al. 2000). An unusual feature of the EGL5 introns, across the three *Phytophthora* species studied, is the alternative intron processing observed in certain transcripts. With PsEGL-V.5, PrEGL-V.5, and PiEGL-V.5, each containing three introns, various combinations of processed transcripts are observed. Some variants lead to truncated, non-functional translational products, and, at least in one case, to a larger than expected product (PsEGL-V.5 with intermediate intron retention). Notably, the majority of identified splice variants had ORFs with complete EGL5 conserved domains, downstream from the early termination codons, theoretically preserving their enzymatic functionality if translation could still take place.

Surprisingly, there are no previous published reports of intron processing in oomycetes. One might expect that *Phytophthora* would have a mechanism of intron processing, however, none of the extensive publications on ESTs have revealed this feature. While alternate intron processing is common in higher eukaryotes, reports of alternative intron processing in lower eukaryotes are extremely limited, possibly due in part to the paucity of introns usually found. In these few cases, the retained introns are generally at the 5′ or 3′ ends of the mRNA. Interestingly, many of these examples occur in carbohydrate modifying enzymes, within the carbohydrate binding regions. Specific examples include an ascomycete (*Aspergillus niger*) glucoamylase gene (Boel et al. 1984) and basidiomycete (*Phanerochaete chrysosporium*) exocellulases (Birch et al. 1995; Larrondo et al. 2004; Macarena et al. 2005).

Another example, where a carbohydrate-binding region is modified, involves not intron processing, but the slightly more common alternative exon splicing seen in fungi. In the zygomycete fungus *Mucor circinelloides*, alternative exon splicing with loss of a 5′ proximal exon, resulted in an endoglucanase lacking one of the two carbohydrate binding regions (Baba et al. 2005). An interesting combination of splicing events was reported in the ascomycete *M. grisea* (Ebbole et al. 2004). A series of transcripts corresponding to the coding region of a hypothetical protein of unknown function were identified in an EST library. The multiple, alternate transcript forms were believed to arise from intron retention events as well as alternative exon splicing. All three exons of the gene were retained, but 5′ splice sites varied. Beyond fungi, another example of alternative mRNAs giving rise to altered terminal regions of a protein is found with the malarial parasite *Plasmodium falciparum*, which uses a combination of alternative intron processing and exon splicing to produce variants of adenyl cyclase (Muhia et al. 2003). The variants are functional, and unique in that they differ by the number of transmembrane domains present at the amino-terminal portion of the protein. Most of the alternative processing occurs in the 5′ untranslated region and is specific to the sexual gametocyte stage of the parasite.

Expression of *P. sojae* EGL5 transcripts differed between plant infections and liquid culture. There was considerably less expression of PsEGL-V. 5 and 7, relative to PsEGL-V.1, 2, 3, and 8, during the soybean infections; however, it will take further studies to determine the biological relevance of this differential expression. It has not yet been determined if proteins are produced from each of the alternately spliced transcripts, nor is it known what the exact substrate specificity for the proteins produced from fully processed transcripts is. Analysis of the nature, minimum size, and linkage composition of the substrates hydrolyzed will allow us to determine if the enzymatic activities are redundant or unique.

We have demonstrated that *Phytophthora* undergoes alternate processing of transcripts. Further analysis of intron-containing transcripts can reveal their prevalence and provide an interesting contribution to oomycete biology, especially considering that the different *Phytophthora* species can process the same introns differently.

Acknowledgments The authors are grateful for the anonymous reviewer suggestion to investigate non-classical protein secretion pathways.

References

Baba Y, Shimonaka A, Koga J, Kubota H, Kono T (2005) Alternative splicing produces two endoglucanases with one or two carbohy-

drate-binding modules in *Mucor circinelloides*. J Bacteriol 187:3045–3051
Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795
Bendtsen JD, Kiemer L, Fausboll A, Brunak S (2005) Non-classical protein secretion in bacteria. BMC Microbiol 5:58
Birch PR, Sims PF, Broda P (1995) Substrate-dependent differential splicing of introns in the regions encoding the cellulose binding domains of two exocellobiohydrolase I-like genes in *Phanerochaete chrysosporium*. Appl Environ Microbiol 61:3741–3744
Boel E, Hansen MT, Hjort I, Hoegh I, Fiil NP (1984) Two different types of intervening sequences in the glucoamylase gene from *Aspergillus niger*. Embo J 3:1581–1585
Caten CE, Jinks JL (1968) Spontaneous variability of single isolates of *Phytophthora Infestans* .I. cultural variation. Can J Bot 46:329–348
Costanzo S, Ospina-Giraldo MD, Deahl KL, Baker CJ, Jones RW (2006) Gene duplication event in family 12 glycosyl hydrolase from *Phytophthora* spp. Fungal Genet Biol 43:707–714
Davies GJ, Gloster TM, Henrissat B (2005) Recent structural insights into the expanding world of carbohydrate-active enzymes. Curr Opin Struct Biol 15:637–645
Davis CA, Grate L, Spingola M, Ares M Jr (2000) Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. Nucleic Acids Res 28:1700–1706
Ebbole DJ, Jin Y, Thon M, Pan H, Bhattarai E, Thomas T, Dean R (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
Fedorov A, Roy S, Fedorova L, Gilbert W (2003) Mystery of intron gain. Genome Res 13:2236–2241
Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
Gilkes NR, Claeyssens M, Aebersold R, Henrissat B, Meinke A, Morrison HD, Kilburn DG, Warren RA, Miller RC Jr (1991a) Structural and functional relationships in two families of beta-1,4-glycanases. Eur J Biochem 202:367–377
Gilkes NR, Henrissat B, Kilburn DG, Miller RC Jr, Warren RA (1991b) Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. Microbiol Rev 55:303–315
Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp 41:95–98
Henrissat B (1991a) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280:309–316
Henrissat B (1991b) Sequence homology between a beta-galactosidase and some beta-glucosidases. Protein Seq Data Anal 4:61–62
Henrissat B, Davies GJ (2000) Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. Plant Physiol 124:1515–1519
Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163
Larrondo LF, Gonzalez B, Cullen D, Vicuna R (2004) Characterization of a multicopper oxidase gene cluster in *Phanerochaete chrysosporium* and evidence of altered splicing of the mco transcripts. Microbiology 150:2775–2783
Macarena S, Fernando LL, Monica V, Rafael V, Bernardo G (2005) Incomplete processing of peroxidase transcripts in the lignin degrading fungus *Phanerochaete chrysosporium*. FEMS Microbiol Lett 242:37–44
Money NP, Davis CM, Ravishankar JP (2004) Biomechanical evidence for convergent evolution of the invasive growth process among fungi and oomycete water molds. Fungal Genet Biol 41:872–876
Mourier T, Jeffares DC (2003) Eukaryotic intron loss. Science 300:1393
Moy P, Qutob D, Chapman BP, Atkinson I, Gijzen M (2004) Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. Mol Plant Microbe Interact 17:1051–1062
Muhia DK, Swales CA, Eckstein-Ludwig U, Saran S, Polley SD, Kelly JM, Schaap P, Krishna S, Baker DA (2003) Multiple splice variants encode a novel adenylyl cyclase of possible plastid origin expressed in the sexual stage of the malaria parasite *Plasmodium falciparum*. J Biol Chem 278:22014–22022
Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1–6
Raible F, Tessmar-Raible K, Osoegawa K, Wincker P, Jubin C, Balavoine G, Ferrier D, Benes V, de Jong P, Weissenbach J, Bork P, Arendt D (2005) Vertebrate-type intron-rich genes in the marine annelid *Platynereis dumerilii*. Science 310:1325–1326
Roy SW, Gilbert W (2005) The pattern of intron loss. Proc Natl Acad Sci USA 102:713–718
Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386
Sakon J, Adney WS, Himmel ME, Thomas SR, Karplus PA (1996) Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. Biochemistry 35:10648–10660
Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, Chapman J, Damasceno CM, Dorrance AE, Dou D, Dickerman AW, Dubchak IL, Garbelotto M, Gijzen M, Gordon SG, Govers F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Krampis K, Lamour KH, Lee MK, McDonald WH, Medina M, Meijer HJ, Nordberg EK, Maclean DJ, Ospina-Giraldo MD, Morris PF, Phuntumart V, Putnam NH, Rash S, Rose JK, Sakihama Y, Salamov AA, Savidor A, Scheuring CF, Smith BM, Sobral BW, Terry A, Torto-Alalibo TA, Win J, Xu Z, Zhang H, Grigoriev IV, Rokhsar DS, Boore JL (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261–1266
Van der Lee T, De Witte I, Drenth A, Alfonso C, Govers F (1997) AFLP linkage map of the oomycete *Phytophthora infestans*. Fungal Genet Biol 21:278–291
Wang H, Jones RW (1995a) A unique endoglucanase-encoding gene cloned from the phytopathogenic fungus *Macrophomina phaseolina*. Appl Environ Microbiol 61:2004–2006
Wang H, Jones RW (1995b) Cloning, characterization and functional expression of an endoglucanase-encoding gene from the phytopathogenic fungus *Macrophomina phaseolina*. Gene 158:125–128
Wang H, Jones RW (1997) Site-directed mutagenesis of a fungal beta-1,4-endoglucanase increases the minimum size required for the substrate. Appl Microbiol Biotechnol 48:225–231
Wang H, Jones RW (1999) Properties of the *Macrophomina phaseolina* endoglucanase (EGL 1) gene product in bacterial and yeast expression systems. Appl Biochem Biotechnol 81:153–160