Available online at www.sciencedirect.com
ScienceDirect
Fungal Genetics and Biology 43 (2006) 707–714
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

Gene duplication event in family 12 glycosyl hydrolase from *Phytophthora* spp.

Stefano Costanzo [a,c], M.D. Ospina-Giraldo [b], K.L. Deahl [c], C.J. Baker [d], Richard W. Jones [c,*]

[a] *Department of Natural Resource Sciences and Landscape Architecture, University of Maryland, College Park, MD 20742, USA*
[b] *Department of Biology, Lafayette College, Easton, PA 18042, USA*
[c] *Vegetable Laboratory, USDA-ARS, 10300 Baltimore Ave., Beltsville, MD 20705, USA*
[d] *Molecular Plant Pathology Laboratory, USDA-ARS, 10300 Baltimore Ave., Beltsville, MD 20705, USA*

Received 24 December 2005; accepted 27 April 2006
Available online 19 June 2006

Abstract

A total of 18 paralogs of xyloglucan-specific endoglucanases (EGLs) from the glycosyl hydrolase family 12 were identified and characterized in *Phytophthora sojae* and *Phytophthora ramorum*. These genes encode predicted extracellular enzymes, with sizes ranging from 189 to 435 amino acid residues, that would be capable of hydrolyzing the xyloglucan component of the host cell wall. In two cases, four and six functional copies of these genes were found in tight succession within a region of 5 and 18 kb, respectively. The overall gene copy number and relative organization appeared well conserved between *P. sojae* and *P. ramorum*, with apparent synteny in this region of their respective genomes. Phylogenetic analyses of *Phytophthora* endoglucanases of family 12 and other known members of EGL 12, revealed a close relatedness with a fairly conserved gene sub-family containing, among others, sequences from the fungi *Emericella desertorum* and *Aspergillus aculeatus*. This is the first report of family 12 EGLs present in plant pathogenic eukaryotes.
Published by Elsevier Inc.

Keywords: Xyloglucan hydrolase; Endoglucanase; *Phytophthora infestans*; *Phytophthora sojae*; *Phytophthora ramorum*

1. Introduction

Plant pathogens deploy an array of host cell wall modifying enzymes during pathogenesis (Gotesson et al., 2002). Host plants attempt to counter many of these enzymatic activities with proteinaceous inhibitors. The most thoroughly studied inhibitors have been the polygalacturonase inhibitor proteins (Cook et al., 1999; Desiderio et al., 1997; Leckie et al., 1999). For this family of proteins a specific pathogen endopolygalacturonase counterpart has been clearly identified (Have et al., 1998). Recently a new class of inhibitor has been identified, the xyloglucan-specific endoglucanase inhibitor proteins (XEGIP) (Naqvi et al., 2005; Qin et al., 2003). This inhibitor was first identified in tomato, and we have cloned and characterized XEGIP in potato. Our EST surveys indicate XEGIP is widespread among various plant genera. The counterpart to XEGIP is found only in family 12 glycoside hydrolases, however, there are no published reports on the occurrence of family 12 endoglucanases in plant pathogens. Several glycoside hydrolase family 12 endoglucanases have been identified and characterized from various bacterial and fungal saprophytes, and can be found in the CAZy Carbohydrate-Active Enzymes database (http://afmb.cnrs-mrs.fr/CAZY/). In an effort to define what role XEGIP could play in host defense from pathogens in the genus *Phytophthora*, we initiated a study to determine if family 12 endoglucanase-encoding genes occur in species of *Phytophthora*. We scanned and translated public genomic and EST databases using three conserved amino acid motifs found in family 12 endoglucanases (Goedegebuur et al., 2002). In this study, we report of a total of 18 family 12

* Corresponding author. Fax: +1 301 504 5555.
E-mail address: jonesr@ba.ars.usda.gov (R.W. Jones).

1087-1845/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.fgb.2006.04.006

endoglucanase genes identified in *Phytophthora sojae* and *Phytophthora ramorum* genomes, four of which carry an additional domain resembling a bacterial cell wall surface anchor domain. Preliminary results also show the presence of at least three copies of family 12 EGL in *Phytophthora infestans*.

2. Materials and methods

2.1. Culturing of Phytophthora spp.

Three *Phytophthora* species were used in this study: *P. infestans*, *P. sojae* and *P. ramorum*. A *P. infestans* isolate was obtained from the collection maintained at the Vegetable Laboratory, Beltsville Agricultural Research Center. The *P. sojae* isolates race 1, 4, 7 and 25 were kindly provided by Dr. S. Li at the National Soybean Pathogen Collection Center, Department of Crop Sciences, University of Illinois. All isolates were routinely cultured on rye agar medium supplemented with 2% sucrose (Caten and Jinks, 1968). 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.

2.2. 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 approximately two weeks at 22 °C in the dark before harvesting. 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.

2.3. RT-PCR

For gene expression analysis by RT-PCR, 5 μg of DNase-treated (TURBO DNA-free, Ambion, Austin, TX) total RNA was reverse-transcribed to cDNA (SuperScript III One-Step RT-PCR Kit, Invitrogen) with gene-specific primers (Table 2). A control reaction (minus RTase) was included to exclude spurious amplifications due to potential presence of contaminating DNA. An additional internal control for in-planta RT-PCR experiments was obtained by designing intron-spanning primers, based on the gene model of the cellulose synthase catalytic subunit located 40 kb upstream of the major EGL 12 gene cluster in the *P. sojae* genome. Forward and reverse primer sequences for the internal control are 5′-CTCGGGTTCCTCTACTAC-3′ and 5′-TCGTTGTCCAGAATGATG-3′, respectively. The expected amplicon size is 702 bp including the 94 bp intron and 608 bp without it. Annealing temperature for this primer set is 58 °C. Primers for PCR amplification of the complete open reading frame (ORF) of each EGL paralog were designed using the Primer3 web site (Rozen and Skaletsky, 2000), based on the *P. sojae* and *P. ramorum* genomic sequence data from the Joint Genome Institute. Amplification products were analyzed by 1% 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 *E. coli* colonies were screened by PCR and positive clones were sequenced.

2.4. Sequence scans and analysis

Preliminary tBlastP analyses of ESTs in GenBank were performed using the conserved amino acid sequences for family 12 consensus domains, referred to as Box 1 (NNLWG), Box 2 (YELMIW) and Box 3 (GTEPF) (Fig. 2). Overlapping ESTs were identified resulting in a family 12 endoglucanase gene of expected size. This gene was subsequently used to identify *P. sojae* and *P. ramorum* sequences by BlastN analysis (*E* value $1\,e^{-5}$), using the version 1.0 of their complete genome data produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/). Phylogenetic tree construction and molecular evolutionary analyses were conducted using *MEGA* version 3.0 (Kumar et al., 2004). A phylogenetic tree was derived based on the Neighbor-Joining algorithm, including protein sequences of 49 representatives of the glycoside hydrolase family 12 among those present in the CAZy Carbohydrate-Active Enzymes database (Coutinho and Henrissat, 1999). Bootstrap test with 1000 replications were conducted to examine the reliability of the interior branches (Felsenstein, 1985). Selective pressure on individual sites of codon alignments was determined by the molecular evolution analysis program Datamonkey (http://www.datamonkey.org). The random effects likelihood (REL) analysis (Pond and Frost, 2005) was implemented to test for purifying or diversifying selection. Amino acid sites under selective pressure were identified based on a Bayes factor of 95. Alignments were formatted using *BOXSHADE* version 3.21, available online 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/). Sequences of all the endoglucanase genes from *P. sojae* and *P. ramorum* were deposited in GenBank Accession Nos. AY645943, DQ206888–DQ206903 and DQ286231.

2.5. Infection assay

Soybean seeds (Glycine Max (L.) Merr.) of cv 'Williams' were germinated in moistened paper towels. Inoculation of

two-week-old seedlings was carried out by producing a small wound (1.5 cm) along the hypocotyls with a razor blade, and applying a plug of agar containing *P. sojae* mycelium directly on the wound. Infected hypocotyl sections of approximately 1 cm in length were removed above the lesion site 72 h after the inoculation. Hypocotyl sections from mock-inoculated plants were collected at the same time to be used as control samples. Excised tissue was flash frozen in liquid nitrogen and stored at −80 °C for subsequent total RNA extraction.

3. Results

3.1. Genomic organization and sequence analysis

Sequence analysis of *P. sojae* and *P. ramorum* genomes revealed the presence of family 12 endoglucanases in multiple copy number. Based on the annotated protein sequences, a considerably high level of synteny was observed between species, in the surrounding regions of the three scaffold sections harboring EGL 12. The cluster located on *P. sojae* scaffold 84 appears to have gained two additional copies, relative to *P. ramorum*, possibly due to genetic rearrangements (Fig. 1). In *P. sojae* a total of 10 predicted functional endoglucanase (EGL) copies (sEGL12-A–sEGL12-L) were identified, while *P. ramorum* had eight (rEGL12-A–rEGL12-H) copies. In *P. sojae*, six copies (sEGL12-A–F) were located in a large cluster spanning a region of approximately 18 kb on scaffold 84. Three additional, but incomplete, remnant copies of EGL were identified approximately 40 kb downstream from sEGL12-A, between sEGL12-A and sEGL12-B, and between sEGL12-C and sEGL12-D. The seventh and eighth full length copies (sEGL12-G and L) were identified on scaffold 270, in opposite reading frames and 2.4 kb from each other. The predicted gene model for sEGL12-L contained a 44 bp intron and a coding length of 609 bp. Our genomic and cDNA sequence data for sEGL12-L do not support the presence of the predicted intronic region, and revealed its coding length to be 570 bp, with a stop codon in the spurious intron. Sequence alignment of sEGL12-G and sEGL12-L with ClustalW revealed a region of 429 bp with 100% similarity shared by both copies. The final two copies, in *P. sojae*, were located on scaffold 66 separated from each other by 1.6 kb. These two copies (sEGL12-H and I) are approximately 500 bp longer than the rest of EGL copies found in *P. sojae*. Further analysis revealed that sEGL12-H and sEGL12-I have an additional serine and threonine-rich region following the glycosyl hydrolase domain. In *P. ramorum*, a similar arrangement presented a cluster of four copies located on a 5 kb region downstream from a predicted cellulose synthase catalytic subunit gene. Notably, a predicted cellulose synthase gene was also found on the *P. sojae* genome, but located 40 kb from the EGL gene cluster. Further analysis of this 40 kb region revealed numerous copies of predicted retroelement-related polyproteins including reverse transcriptases and integrases. A second cluster in *P. ramorum*, located on scaffold 73, had two copies with ORFs in opposite orientation and separated by 1.8 kb. The last two copies (rEGL12-G and H), present on scaffold 44, had a serine and threonine-rich region following the glycosyl hydrolase domain, similar to those (sEGL12-H and I) found in *P. sojae*.

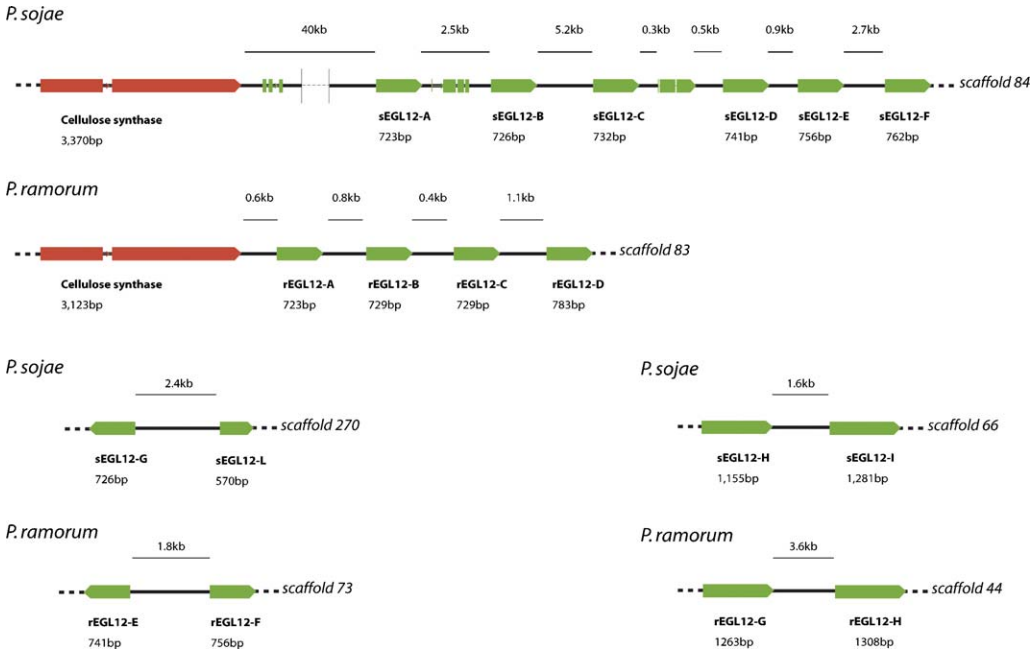


Fig. 1. Map of *Phytophthora sojae* and *Phytophthora ramorum* endoglucanase gene clusters. Conserved scaffold regions between the two *Phytophthora* spp. are shown side by side in three groups. Relative distances of adjacent EGL gene copies (green) are indicated above the solid line representing a scaffold section. The ORF lengths are reported below the individual EGL copy name.

3.2. Analysis of deduced amino acid sequence

The predicted ORF for all the identified EGL gene copies was verified using BioEdit software (Hall, 1999). Based on those predictions, each individual EGL copy was translated into amino acid sequences. These genes encode proteins varying in length from 189 to 435 amino acid residues. Protein sequence analysis from all EGL gene copies identified in both genomes revealed a high percentage of identical amino acid residues (Fig. 2). A signal peptide was predicted, based on a combination of prediction models implemented by the SignalP 3.0 server (Bendtsen et al., 2004; Nielsen et al., 1997), in all 18 EGL amino acid sequences with a cleavage site located between the 18th and 22nd amino acid residue (Table 1).

3.3. Functional and phylogenetic analyses of Phytophthora spp. EGLs

To detect expression of all EGL genes identified in *P. sojae* and *P. ramorum*, we conducted experiments using RT-PCR with gene-specific primers *in vitro*, and for *P. sojae*, also *in planta* (Table 2). Expression of all 18 copies of EGL genes was observed on total RNA extracted from two-week-old mycelium (Figs. 3 and 4). Also, in infected soybean tissue, gene expression could be clearly detected (72 h after inoculation) for all identified EGL copies but sEGL12-H (Fig. 3). The cDNAs encompassing each EGL region were cloned and sequenced to confirm identity. Phylogenetic analyses conducted with ClustalW (alignment) and the Neighbor-Joining method (for phylogenetic inference) indicated that *P. ramorum* and *P. sojae* endoglucanase genes could be classified in two major clusters (Fig. 5). One group constituted by sEGL12-(A–F) and rEGL12-(A–D) is separated from all the remaining copies by EGL sequences from various Ascomycete fungi. Copies sEGL12-D and E do not have a corresponding match in *P. ramorum*, but appear to be more closely related to sEGL12-C. A sequence alignment was generated from the conserved EGL12 domains (Box 1/Box 3) including 17 copies from *P. sojae* and *P. ramorum*. The *P. sojae* sEGL12-L sequence was excluded since it does not include a Box 3 region. Signal peptides were also omitted from the analysis. The aligned sequence data were analyzed by the REL method (Pond and Frost, 2005) with the Datamonkey program for the detection of selective pressure and 112 sites were found under negative or purifying selection and no positively selected sites could be detected (Fig. 6).

4. Discussion

Cellulases are members of a large family of glycosyl hydrolases enzymes involved in selective modification and degradation of cellulose, a main component of plant cell walls. These cellulolytic enzymes act as endo-β-glucanases, cellobiohydrolases and β-glucosidases and are generally classified in families based on their known enzymatic activities and sequence similarity. Family 12 endoglucanases are the only ones known to be inhibited by a xyloglucan-specific endoglucanase inhibitor protein (Ham et al., 1997, 2003, 2002). Thus far, only saprophytic microbes had been reported as a source of family 12 xyloglucan-specific

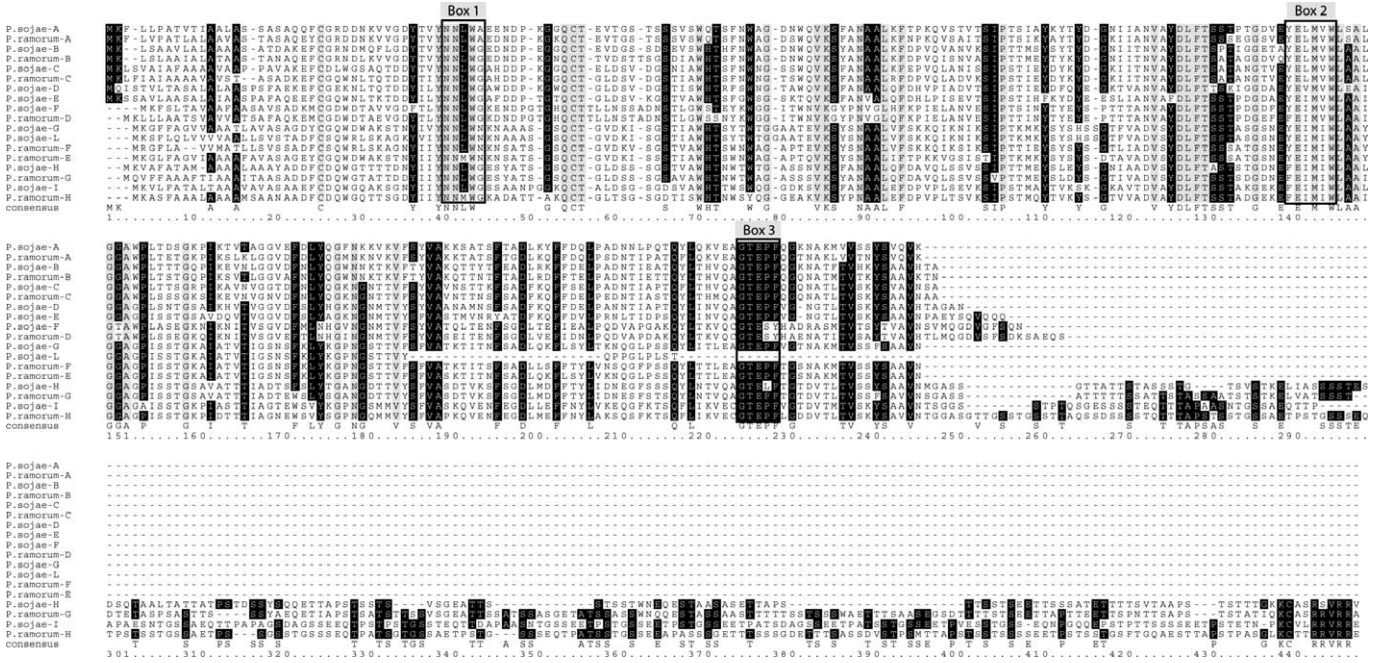


Fig. 2. Multiple sequence alignments of 18 *Phytophthora sojae* and *Phytophthora ramorum* family 12 glycosyl hydrolase members. Consensus line indicates amino acidic residues with identity threshold of 75% and above. A gray background indicates identical residues in all sequences. The predicted signal peptide has a cleavage site located in each copy between the 18th and 21st amino acid residue.

Table 1
Predicted characteristics of 18 EGL12 protein from *Phytophthora sojae* and *Phytophthora ramorum*

Protein	ORF size (bp)	aa[a]	SP cleavage site[b]	Serine/threonine-rich region[c]
sEGL12-A	723	240	20–21	–
sEGL12-B	726	241	19–20	–
sEGL12-C	732	243	21–22	–
sEGL12-D	741	246	21–22	–
sEGL12-E	756	251	21–22	–
sEGL12-F	762	253	18–19	–
sEGL12-G	726	241	19–20	–
sEGL12-H	1155	384	19–20	+
sEGL12-I	1281	426	19–20	+
sEGL12-L	570	189	20–21	–
rEGL12-A	723	240	20–21	–
rEGL12-B	729	242	19–20	–
rEGL12-C	729	242	19–20	–
rEGL12-D	783	260	19–20	–
rEGL12-E	720	239	19–20	–
rEGL12-F	714	237	18–19	–
rEGL12-G	1263	420	20–21	+
rEGL12-H	1308	435	19–20	+

[a] Number of amino acids.
[b] Predicted signal peptide cleavage site.
[c] Putative cell wall surface anchor domain.

endoglucanases. We demonstrate the presence of multiple, transcriptionally active copies of family 12 endoglucanase in the phytopathogen genus *Phytophthora*. Preliminary results from Southern analysis, and EST mining, also indicated the presence of at least three copies of EGL 12 in *P. infestans*.

Sequences and RT-PCR transcripts for these genes confirmed that they are in the proper reading frame and are all expressed during *in vitro* growth. In our analysis of the 10 *P. sojae* genes, differential expression between *in vitro* and *in planta* tests was observed only for sEGL12-H, where expression was very low *in planta*. A recent microarray study (Moy et al., 2004; Qutob et al., 2000), on patterns of gene expression in soybean and *P. sojae* during an infection time course, identified a putative endoglucanase EST from *P. sojae* (GenBank Accession No. BE583376) as being highly induced during the first two days after inoculation. Its sequence almost perfectly matches the first 500 bp of sEGL12-B and it was annotated as having similarities to an endoglucanase from *Emericella desertorum*. During the time course, the microarray spot corresponding to this EST had a 3-fold change in fluorescence emission within the first 6 h and reached a 16-fold increase 24 h after inoculation.

One distinctive feature of family 12 glycoside hydrolases is the absence of a cellulose-binding domain (CBD), which allows anchoring to crystalline cellulose substrates. The lack of a CBD is a paradigm of pathogen-encoded endoglucanases that may allow for diffusion of the endoglucanase through the host cell wall (Wang and Jones, 1995a; Wang and Jones, 1995b; Wang and Jones, 1997). An interesting feature of four members of the *Phytophthora* EGL 12 gene family is the serine and threonine-rich N terminal region which shares some homology to bacterial cell wall anchoring domains (Tettelin et al., 2001). Like all members of the Division Oomycota, *Phytophthora* species have a cell wall primarily comprised of β-glucans. The presence of endoglucanases within their genome could potentially be associated to hyphal extension mechanisms as well as the need to loosen the host cell wall during the infection process (McLeod et al., 2003). It is possible that EGL12-H and potentially EGL12-I, which are expressed during hyphal growth, have yet undefined targets within the cell wall of *Phytophthora*. It should be noted that the cellulose synthase gene is located in close proximity to the EGL12 gene clusters, suggesting a coordinated gene regulation. Further analysis of the *Phytophthora* cell wall may reveal specific sites for some members of the EGL12 family.

The target for EGL12 in plant cell walls is the xyloglucan framework. Xyloglucans (XG) play a key structural

Table 2
Gene-specific primers of 18 EGL12 genes from *Phytophthora sojae* and *Phytophthora ramorum* used in RT-PCR

Gene	Forward primer sequence	Reverse primer sequence	Amplicon size (bp)	Temperature (°C)[a]
sEGL12-A	CAACAAGGTGGTCGGAGACT	CTTCTGCAGGTACTGCGTCT	562	59
sEGL12-B	AAGACCCCGACACCAAGAC	CGTGAAGGTGGCGTTCTT	724	61
sEGL12-C	GCAGACGGACGACTACACAG	AGAACTTGGTGGTTGAGTTCAC	482	59
sEGL12-D	TCTCCACCGTTCTCACTGC	GCGGAGAAGCTGTTCATAGC	574	59
sEGL12-E	CTGGTCGGGCTCCAAGAC	GACGCTGGCTCTAAGCTATCC	595	61
sEGL12-F	GTGGAACGTCAAGGGCTATC	CTCCTTGCATCACCGAGTTC	509	58
sEGL12-G	GCCAAGAGCACTAACTACATC	GCGACGAACGAGAAGACG	467	58
sEGL12-H	TACTCATTTGTGGCTTCC	GACGCGACGAACACTGC	621	58
sEGL12-I	CTTCAAGACCAGCCAGTTC	CTCGCGACGAACGCGACGGA	661	61
sEGL12-L	ATGAAGAGCTTTCTCCAACTCG	ATACGGTCATCTTGGCGTTCG	627	58
rEGL12-A	TCCATCCCTACGAGCATCA	ACTTGACCTGCACCGAGTAG	427	58
rEGL12-B	CTGAAGGTTGTCGGGGACT	GGTCAGGTACTGCGTCGTC	561	59
rEGL12-C	CAAGGAGTTCTGCGGTCAGT	AAGCTGTTCGTCGTGTTGAC	510	58
rEGL12-D	GAAGGAAATGTGCGGTGACT	GATGGTGGCGTTCTCAGC	640	58
rEGL12-E	ATGAAGGGCCTCTTCGCTG	GTTGACAGCAGCAGAGTATGACG	717	59
rEGL12-F	AGTCTTCTAGTCGCGCCTTC	TTGGGCAAATGTAGTCGTTG	828	58
rEGL12-G	ATGCAGGTGTTCTTCGCTGC	AGCACGGCGGACGCGACGGCTA	1260	61
rEGL12-H	ATGAAGGCATCGTTCGCTG	CTCGCGACGGACGCGGCGAGTAG	1305	60

[a] Annealing temperatures used for PCR and RT-PCR.

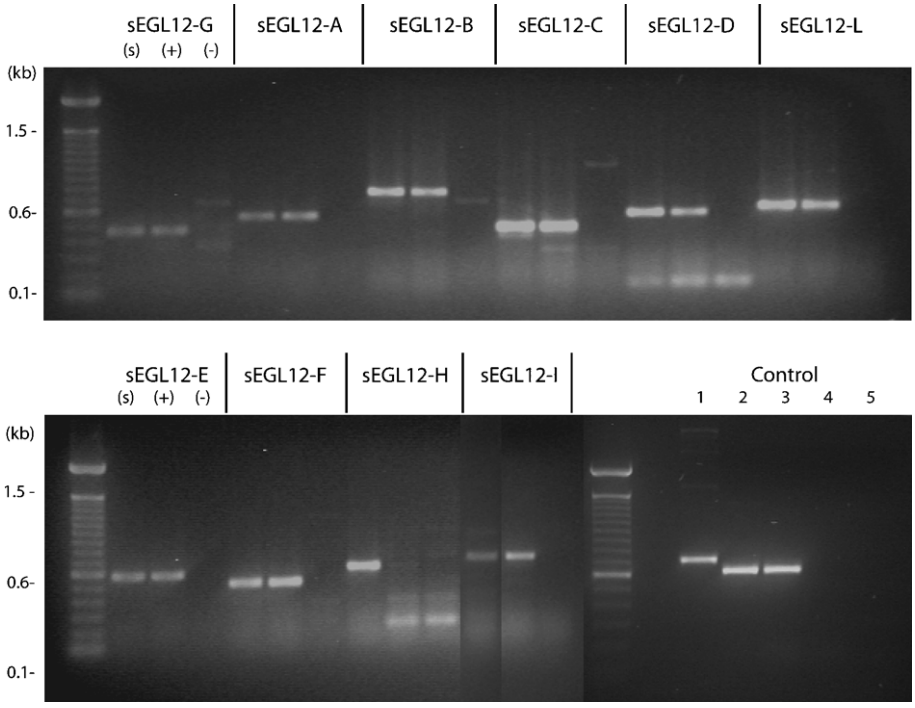


Fig. 3. sEGL12 (A–I) gene expression analysis. RT-PCR using individual gene-specific primers on total RNA from three different sources: (s) *P. sojae* mycelium; (+) infected soybean tissue; (−) mock-inoculated soybean tissue. Controls utilized primers spanning an intron region of the cellulose synthase catalytic subunit: lane 1 (− RTase/ + Taq/*P. sojae* genomic DNA); lane 2 (+ RTase/ + Taq/*P. sojae* total RNA); lane 3 (+ RTase/ + Taq/infected soybean tissue); lane 4 (+ RTase/ + Taq/mock-inoculated soybean tissue); lane 5 (− RTase/ + Taq/*P. sojae* total RNA). Molecular markers are given to the left of the figure.

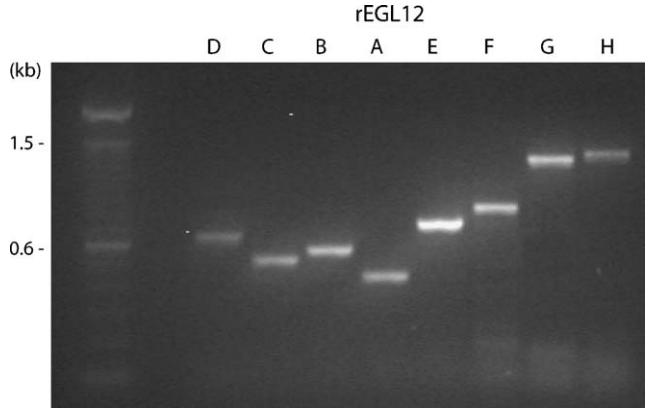
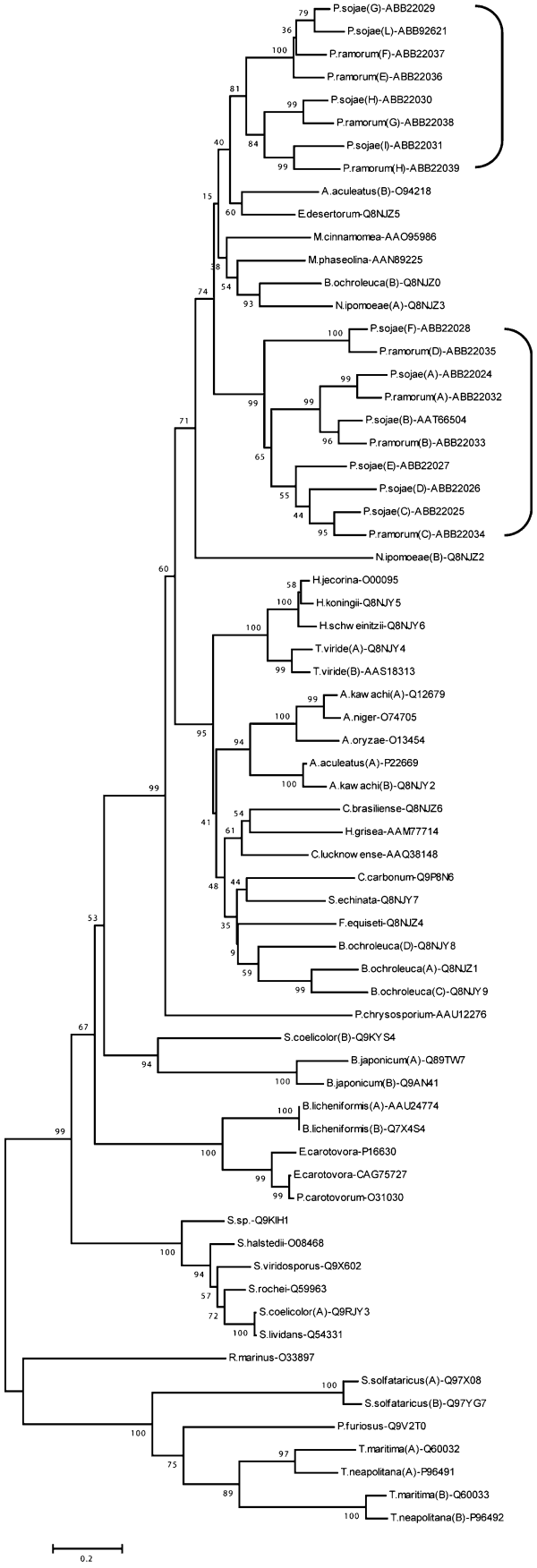


Fig. 4. rEGL12 (A–H) gene expression analysis. RT-PCR using individual gene-specific primers on total RNA from *Phytophthora ramorum* mycelium. Molecular weight marker sizes are given to the left of the figure.

role in the cell wall by tethering cellulose microfibrils (Carpita and Gibeaut, 1993; Hayashi, 1989; Yuan and Wu, 2001). Hydrolysis of this xyloglucan framework could allow sufficient loosening of the host cell wall to allow for physical penetration with limited apparent dissolution of the cell wall. Interestingly, in a study on glycoside hydrolases, family 12 members were the only ones with extensin-like activity, which induce wall extension by a mechanism resembling some of the changes found after auxin-induced growth (Cosgrove, 2000; Yuan and Wu, 2001). Endoglucanases of family 12 are capable of modifying both type-I and type-II walls, but considerable extension has been demonstrated only for type-I walls.

The presence of xyloglucan-specific endoglucanase inhibitor protein (XEGIP) suggests a role in defense from pathogen attack. We have functionally expressed sEGL12-B in both a PVX vector and in *E. coli*, to demonstrate this concept in a forthcoming study (unpublished). Given the large number of EGL 12 copies, it will be interesting to determine which EGLs interact with XEGIP and which ones may elude this host defense. The preponderance of EGL12-encoding genes in *Phytophthora* suggests a positive selection pressure for maintenance of such a large number of transcriptionally active copies. Evolutionary analyses on the 17 EGL12 copies from *P. sojae* and *P. ramorum* revealed that they are likely to be under purifying selection as determined by the random effects likelihood method. The presence of

Fig. 5. Phylogenetic analysis of 67 members of the Glycoside Hydrolase family 12, obtained by the Neighbor-Joining method. Protein ID follows the abbreviated organism name. Full names in alphabetical order are: *Aspergillus aculeatus*; *Aspergillus kawachii*; *Aspergillus niger*; *Aspergillus oryzae*; *Bacillus licheniformis*; *Bionectria ochroleuca*; *Bradyrhizobium japonicum*; *Chaetomium brasiliense*; *Chrysosporium lucknowense*; *Cochliobolus carbonum*; *Emericella desertorum*; *Erwinia carotovora*; *Fusarium equiseti*; *Humicola grisea*; *Hypocrea jecorina*; *Hypocrea koningii*; *Hypocrea schweinitzii*; *Macrophomina phaseolina*; *Malbranchea cinnamomea*; *Nectria ipomoeae*; *Pectobacterium carotovorum*; *Phanerochaete chrysosporium*; *Phytophthora ramorum*; *Phytophthora sojae*; *Pyrococcus furiosus*; *Rhodothermus marinus*; *Stachybotrys echinata*; *Streptomyces coelicolor*; *Streptomyces halstedii*; *Streptomyces lividans*; *Streptomyces rochei*; *Streptomyces* sp. *11AG8*; *Streptomyces viridosporus*; *Sulfolobus solfataricus*; *Thermotoga maritime*; *Thermotoga neapolitana*; *Trichoderma viride*. Two brackets indicate the *Phytophthora sojae* and *P. ramorum* sequences. Scale bar represent weighted sequence divergence. Bootstrap values are shown at the nodes.



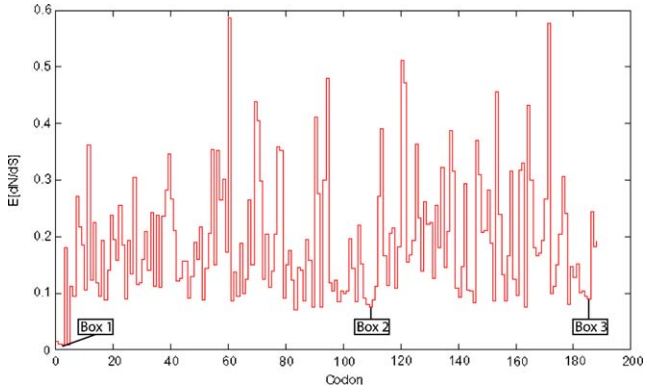


Fig. 6. Plot of posterior expected values for the nonsynonymous/synonymous rate ratio (dN/dS), calculated among all codon sites between Box 1 and Box 3 consensus domains of *Phytophthora* EGL 12.

numerous copies in close proximity along with remnant sequences indicates multiple gene duplication events, occurring at different times. The selection pressures cannot be simply due to host range as *P. sojae*, with a relatively narrow host range, has more copies than the wider host range *P. ramorum*. Gene duplication of endoglucanases appears to be a relatively uncommon event. In the literature, we have encountered few studies reporting examples of any duplication of cell wall degrading enzyme-encoding genes. When duplication occurs, only one to two extra copies were found clustered together. In *Phytophthora* we have observed an unexpected high level of duplication of endoglucanases. A similar pattern of gene clustering of cell wall degrading enzyme-encoding genes was reported only for a polygalacturonase gene family from *Phytophthora cinnamoni* (Gotesson et al., 2002), although, in that case, the distances between copies were much greater. It is likely that genes encoding numerous classes of plant cell wall modifying enzymes will be found clustered in the *Phytophthora* genome.

References

Bendtsen, J.D., Nielson, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783–795.

Carpita, N.C., Gibeaut, D.M., 1993. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J. 3, 1–30.

Caten, C.E., Jinks, J.L., 1968. Spontaneous variability of single isolates of *Phytophthora infestans*. I. Cultural Variation. Can. J. Bot. 46, 329–348.

Cook, B.J., Clay, R.P., Bergmann, C.W., Albersheim, P., Darvill, A.G., 1999. Fungal polygalacturonases exhibit different substrate degradation patterns and differ in their susceptibilities to polygalacturonase-inhibiting proteins. Mol. Plant Microbe Interact. 12, 703–711.

Cosgrove, D.J., 2000. Loosening of plant cell walls by expansins. Nature 407, 321–326.

Coutinho, P.M., Henrissat, B., 1999. Carbohydrate-Active Enzymes: An Integrated Database Approach. The Royal Society of Chemistry, Cambridge.

Desiderio, A., Aracri, B., Leckie, F., Mattei, B., Salvi, G., Tigelaar, H., Van Roekel, J.S., Baulcombe, D.C., Melchers, L.S., De Lorenzo, G., Cervone, F., 1997. Polygalacturonase-inhibiting proteins (PGIPs) with different specificities are expressed in *Phaseolus vulgaris*. Mol. Plant Microbe Interact. 10, 852–860.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Goedegebuur, F., Fowler, T., Phillips, J., Van Der Kley, P., Van Solingen, P., Dankmeyer, L., Power, S.D., 2002. Cloning and relational analysis of 15 novel fungal endoglucanases from family 12 glycosyl hydrolase. Curr. Genet. 41, 89–98.

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

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41, 95–98.

Ham, K.S., Wu, S.C., Darvill, A.G., Albersheim, P., 1997. Fungal pathogens secrete an inhibitor protein that distinguishes isoforms of plant pathogenssis-related endo-β-1,3-glucanases. Plant J. 11, 169–179.

Hayashi, T., 1989. Xyloglucans in the primary-cell wall. Annu. Rev. Plant Physiol. Plant Mol. Biol. 40, 139–168.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5, 150–163.

Leckie, F., Mattei, B., Capodicasa, C., Hemmings, A., Nuss, L., Aracri, B., DeLorenzo, G., Cervone, F., 1999. The specificity of polygalacturonase-inhibiting protein (PGIP): a single amino acid substitution in the solvent-exposed beta-strand/beta-turn region of the leucine-rich repeats (LRRs) confers a new recognition capability. EMBO J. 18, 2352–2363.

McLeod, A., Smart, C.D., Fry, W.E., 2003. Characterization of 1,3-beta-glucanase and 1,3;1,4-beta-glucanase genes from *Phytophthora infestans*. Fungal Genet. Biol. 38, 250–263.

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

Naqvi, S.M., Harper, A., Carter, C., Ren, G., Guirgis, A., York, W.S., Thornburg, R.W., 2005. Nectarin IV, a potent endoglucanase inhibitor secreted into the nectar of ornamental tobacco plants. Isolation, cloning, and characterization. Plant Physiol. 139, 1389–1400.

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.

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

Qin, Q., Bergmann, C.W., Rose, J.K., Saladie, M., Kolli, V.S., Albersheim, P., Darvill, A.G., York, W.S., 2003. Characterization of a tomato protein that inhibits a xyloglucan-specific endoglucanase. Plant J. 34, 327–338.

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

Rose, J.K., Ham, K.S., Darvill, A.G., Albersheim, P., 2002. Molecular cloning and characterization of glucanase inhibitor proteins: coevolution of a counterdefense mechanism by plant pathogens. Plant Cell 14, 1329–1345.

Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132, 365–386.

ten Have, A., Mulder, W., Visser, J., van Kan, J.A., 1998. The endopolygalacturonase gene Bcg1 is required for full virulence of *Botrytis cinerea*. Mol. Plant Microbe. Interact. 11, 1009–1016.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K., Fraser, C.M., 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 293, 498–506.

Wang, H., Jones, R.W., 1995a. A unique endoglucanase-encoding gene cloned from the phytopathogenic fungus *Macrophomina phaseolina*. Appl. Environ. Microbiol. 61, 2004–2006.

Wang, H., Jones, R.W., 1995b. Cloning, characterization and functional expression of an endoglucanase-encoding gene from the phytopathogenic fungus *Macrophomina phaseolina*. Gene 158, 125–128.

Wang, H., Jones, R.W., 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.

Yuan, S., Wu, Y., 2001. A fungal endoglucanase with plant cell wall extension activity. Plant Physiol. 127, 324–333.