CrossMark

ORIGINAL ARTICLE

Evolutionary Analysis of Pectin Lyases of the Genus *Colletotrichum*

Alicia Lara-Márquez[1] · Ken Oyama[1] · María G. Zavala-Páramo[2] ·
Maria G. Villa-Rivera[2] · Ulises Conejo-Saucedo[2] · Horacio Cano-Camacho[2]

Abstract Pectin lyases (PNLs) are important enzymes that are involved in plant cell wall degradation during the infection process. *Colletotrichum* is a diverse genus of fungi, which allows the study of the evolution of PNLs and their possible role in pathogen–host interactions and lifestyle adaptations. The phylogenetic reconstruction of PNLs from *Colletotrichum* and analysis of selection pressures showed the formation of protein lineages by groups of species with different selection pressures and specific patterns. The analysis of positive selection at individual sites using different methods allowed for the identification of three codons with evidence of positive selection in the oligosaccharide-binding region and two codons on the antiparallel sheet, which may influence the interaction with the substrate. Seven codons on the surface of the protein, mainly in the peripheral helices of the PNLs, could have an important function in evasion of plant defenses, as has been proposed in other enzymes. According to our results, it is possible that events of genetic duplication occurred in ancestral lines, followed by episodes of genetic diversification and gene loss, probably influenced by differences in the composition of the host cell wall. Additionally, different patterns of evolution in *Colletotrichum*

appear to be molded by a strong purifying selection and positive selection episodes that forged the observed evolutionary patterns, possibly influenced by host interaction or substrate specificity. This work represents a starting point for the study of sites that may be important for evasion of plant defenses and biotechnological purposes.

Keywords Pectin lyase · *Colletotrichum* sp. · Positive selection · Molecular evolution

✉ Horacio Cano-Camacho
hcano1gz1@mac.com

[1] Escuela Nacional de Estudios Superiores (ENES) Unidad Morelia, Universidad Nacional Autónoma de México (UNAM), Morelia, Michoacán, México

[2] Centro Multidisciplinario de Estudios en Biotecnología, FMVZ, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México

Introduction

The genus *Colletotrichum* ranks eighth in molecular plant pathology for its scientific and economic importance (Dean et al. 2012). The genus comprises more than 600 species capable of attacking more than 3200 species of monocots and dicot plants (O'Connell et al. 2012), causing large economic losses in important crops such as maize, sugar cane, sorghum, and beans. The production of fruits is mainly affected by infection by *Colletotrichum* sp., which are also considered post-harvest pathogens because the infection can be activated once the fruit has been stored (Prusky 1996; Lenné 2001; Dean et al. 2012). The genus is very diverse, and species have different host ranges and levels of virulence, ranging from destructive pathogens to endophyte fungi (Perfect et al. 1999; Gan et al. 2013). The genus *Colletotrichum* has been a main model of biochemical, physiological, and genetic studies, and intra-species pathogenic variation (race/cultivar specificity) was reported for the first time in *Colletotrichum lindemuthianum* (Barrus 1911).

According to comparative analyses of sequenced fungi genomes, there is large variation in the size and architecture of the genomes and differences in the size and distribution of some gene families encoding carbohydrate-active enzymes

(CAZymes). This variation appears to be phylum-specific in some cases and has also been attributed to the lifestyle of the fungi and host types, particularly in the case of pectin-degrading enzymes (Powell et al. 2008; Soanes et al. 2008; Demuth and Hahn 2009; Zhao et al. 2013; Lo Presti et al. 2015).

The genomes of *C. higginsianum* (O'Connell et al. 2012), *C. graminicola* (O'Connell et al. 2012), *C. orbiculare* 104-T (Gan et al. 2013), *C. fructicola* (formerly *C. gloeosporioides* Nara-gc5) (Gan et al. 2013), *C. sublineola* (Baroncelli et al. 2014a), and *C. acutatum* (*C. fioriniae*) (Baroncelli et al. 2014b) have been sequenced. The analyses revealed that *Colletotrichum* species have a complex repertoire of plant cell wall-degrading enzymes (PCWDEs). For example, expansion of CAZymes families in the sequenced genomes of *Colletotrichum* was detected, compared to other Ascomycetes and among species of the genus (O'Connell et al. 2012; Gan et al. 2013; Crouch et al. 2014). The expansion in pectinase content, which appears to be influenced by the pectin content of the host cell wall, is particularly notable (Gan et al. 2013). The genomes of *C. orbiculare, C. gloeosporioides*, and *C. higginsianum* have an expanded number of pectin-degrading enzymes compared to the number of pectinases of the monocots pathogen *C. graminicola*. This is consistent with the adaptation to the cell wall of dicots with pectin content (30%) higher than that of monocots (10%) and, according to transcriptomic analyses, the differences are also reflected in the expression patterns of pectinases (O'Connell et al. 2012; Gan et al. 2013).

Pectinolytic enzymes are considered important in the pathogenesis and as potential virulence factors due to the importance of pectin in the cell wall and because they are typically secreted in the early stages of pathogenesis (Cooper 1983; Annis and Goodwin 1997). According to different studies involving the process of penetration and acquisition of nutrients, a correlation between the levels of pectinolytic enzymes and maceration of host tissue has been observed (Walton 1994; Isshiki et al. 2001; Kim et al. 2001). The results of analyses performed to determine the potential of pectinases as virulence factors have been variable; however, their degradation products can induce the host defense system. Additionally, the presence of the inhibitory protein families of polygalacturonases (PG) (D'Ovidio et al. 2004; Federici et al. 2006), pectin methyl esterases (PME) (Balestrieri et al. 1990; Raiola et al. 2011), and pectin lyases (PNLs) (Bock et al. 1975; Bugbee 1993) in various plants represents the evidence of a coevolutionary arms race between the pectinolytic enzymes produced by the pathogen and the inhibitory proteins produced by the host (Dornez et al. 2010).

In some sequenced fungi genomes, positive selection signals have been found in virulence genes, elicitors, and some effectors (Aguileta et al. 2009, 2012; MacColl 2011). An analysis of 35 isolates of *Botrytis* showed evidence of positive selection in the *Bcpg1* and *Bcpg2* genes encoding endopolygalacturonases, suggesting that it could be advantageous to escaping host recognition (Cettul et al. 2008). In the emergent pathogen, *Zymoseptoria tritici* and ancestral species of various wild grasses, significant differences were reported in the life-cycle-dependent expression of pectinase genes in different life stages, as well as evidence of purifying selection (Brunner et al. 2013). This appears to be a reflection of the adaptive significance of pectinases to host preferences and possibly lifestyle.

Most research on the evolution of pectinases has been conducted on a large scale by analyzing CAZymes families in the sequenced fungal genomes, allowing for an overview of the role of pectinases in adaptation to lifestyle and host type. We were interested in analyzing the evolution of pectin lyases in *Colletotrichum* species, a very diverse genus that allows for investigation of the possible role of PNLs in host type adaptations and the lifestyle of the genus. The PNLs are the only enzymes capable of breaking the internal glycosidic bonds of highly methylated pectin such as fruit pectin without the action of other enzymes (Alaña et al. 1989). Additionally, these enzymes have been considered virulence factors and are members of family 1 of polysaccharide lyases (PL1), one of the most variables in the analyzed genomes.

To obtain insight into the evolutionary process of the PNLs of *Colletotrichum*, we performed a phylogenetic reconstruction of the genes encoding pectin lyases from *Colletotrichum* species reported in the databases, and we analyzed the selection pressures of the different clades identified in the PNL phylogeny to determine if they are under different evolutionary routes. We also analyzed the presence of positive and purifying selection using different methods to determine if there is evidence of selection at individual sites.

In addition, we identified the amino acids corresponding to the codons with selection evidence in the alignment of PNLs and in the three-dimensional model of PNL 2 from *C. lindemuthianum* that was used as a representative to analyze the possible implications of the identified amino acids.

Methodology

Sequence Data

The search of sequences encoding pectin lyases of 14 species of the genus *Colletotrichum* was performed using the Basic Local Alignment Search Tool (BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al. 1990) of the National Biotechnology Information Center (NCBI) website (https://www.ncbi.nlm.nih.gov/). We used characterized sequences from pectin lyases of *Colletotrichum* and other fungi as queries to perform searches on the nucleotide and genomic databases (BLASTN and BLAST Genomes)

available at the NCBI. The coding sequences of the genes in the genomes of *Colletotrichum* reported in the databases were obtained using the AUGUSTUS web server for the prediction of genes (Stanke and Morgenstern 2005) and FGENESH (Solovyev et al. 2006). The presence of the conserved functional domain Pec_lyase_C (Pfam PF00544) in the deduced amino acid sequences was analyzed using the Protein Families database (Pfam) (Bateman et al. 2004) through CLC Main Workbench 7.8 (http://www.clcbio.com) software (Online Resource Table 2). The coding sequence of pectin lyase 2 from *C. lindemuthianum* 1472 (GenBank JN034039) was used as a representative sequence to illustrate and analyze the results.

Multiple Sequence Alignments

A total of 36 nucleotide sequences of putative pectin lyases from different *Colletotrichum* species were aligned using their amino acid sequences as a guide with MUSCLE implemented in MEGA v6.06 (Tamura et al. 2013). A pairwise comparison analysis was performed to obtain the percent identity calculated as the ratio of identical alignment positions to overlapping alignment positions between the two sequences, and the Jukes–Cantor distance between two sequences using complete nucleotide and amino acid alignments was calculated using CLC Main Workbench 7.8 (http://www.clcbio.com) software (Online Resource Fig. 2). The analysis of secondary structures, domains, and signal peptide prediction from the sequences in this work was performed using CLC Main Workbench 7.8 (http://www.clcbio.com) software.

Phylogenetic Analyses

To analyze the molecular evolutionary history of the pectin lyases of the genus *Colletotrichum*, the alignment of the 36 PNL sequences was modified by removing non-aligned nucleotide portions and excluding ambiguously aligned sites. A pectin lyase sequence from the basidiomycete fungus *Rhizoctonia solani* (GenBank: AZST01000001) was added to the alignment as an external group. jModelTest 2.1 software (Posada 2008) was employed to determine the optimal model of nucleotide substitution for the coding sequence dataset. The Bayesian information criterion (BIC) and Akaike information criterion (AIC) were used to select the optimal model.

The phylogenetic tree was constructed with the Bayesian approach using MrBayes v3.2 (Ronquist et al. 2012) and maximum likelihood (ML). The Bayesian analysis was run for 1,000,000 generations with a sampling frequency of 100, and a burn-in of 25% of the sampled trees, and a 50% majority rule consensus tree of the remaining trees was generated. Maximum likelihood analyses were performed

using RaxmlGUI v1.31, (Silvestro and Michalak 2011) the graphical front-end for RAxML (Stamatakis 2014) with the GTR + gamma model, executing 500 non-parametric bootstraps. The phylogenetic tree generated using the Bayesian approach was used for the analysis of selection. FigTree v1.4.3 (http://tree.bio.ed.ac.uk) was used for viewing and editing the trees.

Branch Analysis of Selective Pressure

The selective pressures in the different groups comprising the tree were analyzed using ML method in the CODEML program of PAMLX v1.3.1 (Xu and Yang 2013). We used the one-ratio model, which assumes that all branches have only one ratio to identify the selective pressures in the PNL genes, and several two-ratio models, which allow different ω values between the clusters to check if the data fit significantly better than in the one-ratio model. A likelihood ratio test (LRT) was performed to accept or reject the hypotheses, calculated using $2 \times (\ln L1 - \ln L0)$. The degrees of freedom were determined by the difference in the number of free parameters between the null and alternative models and the test statistic was approximated to a χ^2 distribution to determine statistical significance (Bielawski and Yang 2003).

Site-Specific Analyses of Positive Selection

To detect positive selection at specific sites of PNLs, we used site-specific models in the CODEML program of PAMLX v1.3.1 software (Xu and Yang 2013). The LRT was used to compare the pairs of models: M1 (neutral), and M2 (selection), M0 (one-ratio), and M3 (discrete), M7 (β distribution), and M8 (beta and omega variations). The M0 model estimates the overall data. The M1 model estimates the codon site proportion p_0 with $\omega_0 < 1$ and the proportion p_1 ($p_1 = 1 - p_0$) with $\omega_1 = 1$. Model M3 allows ω to vary across sites within n discrete categories, $n \geq 3$. The M2 model allows an additional class of positively selected sites with proportion p_2 ($p_2 = 1 - p_1 - p_0$), with ω_2 estimated from the data. The M7 model specifies that ω follows a beta distribution and ω is allowed to vary between 0 and 1. In the M8 model, a proportion of sites p_0 has a ω in the beta distribution and the p_1 proportion of sites is assumed to be positively selected (Yang et al. 2000). The likelihood ratio test statistic was calculated as $2 \times (\ln L1 - \ln L0)$; the degrees of freedom were determined by the difference in the number of free parameters between the null and alternative models; and the test statistic was approximated to a χ^2 distribution to determine statistical significance and accept or reject the hypothesis of the null model. Sites with Bayes Empirical Bayes (BEB) posterior probabilities > 95% were considered positive (Yang et al. 2005).

To analyze positive selection at specific sites of the PNL genes, we used the single-likelihood ancestor counting (SLAC) method that maps changes in the phylogeny to estimate the selection on a site-by-site basis and calculate the number of non-synonymous and synonymous substitutions at each site using ML reconstructions of ancestral sequences (Pond et al. 2005c, a). The fixed effects likelihood (FEL) method that estimates the ratio of non-synonymous and synonymous substitutions does not assume an a priori distribution of rates (Pond et al. 2005b). Both methods were performed using the HyPhy package implemented on the Datamonkey web server (http://www.datamonkey.org) (Pond et al. 2005c; Delport et al. 2010). The best-fitting nucleotide substitution model was searched for using the automatic model selection tool on the server. We identified codons with P values of <0.2 in the SLAC and FEL results for informational purposes and accepted those with P values <0.1.

We used the mixed-effects evolution model (MEME) to identify evidence of either episode or pervasive positive selection, which in protein-coding genes commonly occurs according to Murrell et al. (2012). MEME analysis was performed with the HyPhy package available on Datamonkey web server (Pond et al. 2005c; Delport et al. 2010). The best nucleotide substitution model was selected using the automatic model tool of the server, and we identified and accepted codons with P values <0.1 in the MEME results.

Branch-Site Analyses of Positive Selection

To analyze the selective pressures that occur in the internal branches of the tree, we implemented the internal branches FEL (IFEL) (Pond et al. 2006a) and the Bayesian unresolved test for episodic diversification (BUSTED) (Murrell et al. 2015) methods on the Datamonkey web server (http://www.datamonkey.org) (Pond et al. 2005c; Delport et al. 2010). The IFEL method is based on FEL method and can discriminate substitutions that occur in the internal branches from those occurring at the tips of the tree. We identified codons with P values <0.2 and accepted codons with P values <0.1. The BUSTED method considers three ω categories ($\omega_1 \leq \omega_2 \leq 1 \leq \omega_3$) shared by all branches and sites so it can detect positive selection in all phylogenetic branches and at specific sites of genes. The alternative model allows for $\omega_3 > 1$ on the foreground branch, whereas the null model considers $\omega_3 = 1$ on this branch. The LRT was compared with a χ^2 distribution ($df = 2$) and positively selected sites were identified at a significance level of $P < 0.05$. Codons identified with evidence of negative or purifying selection using the applied methods were compiled (Online Resource Table 4) and used for structural analysis.

Test for Recombination

To analyze the possible effects of recombination on the PNL phylogeny and on the positive selection results, we analyzed the existence of putative recombination breakpoints using the GARD algorithm (Pond et al. 2006b), available at http://www.datamonkey.org/. Potential breakpoints were detected using the AIC. For topological incongruity analysis, GARD applies the Shimodaira and Hasegawa test (KH test) (Pond et al. 2006c). The sequences of the non-recombinant fragments according to the GARD analysis were analyzed separately to identify positively selected sites with the above-mentioned methods and the trees were analyzed using Phylo.io version 1.0.0 (http://phylo.io/#) (Robinson et al. 2016).

Structural Analyses

A full-length atomic model of pectin lyase 2 of *C. lindemuthianum* was constructed by iterative template fragment-assembly simulations using the hierarchical method for protein structure prediction implemented in I-TASSER (Roy et al. 2010) using the mature protein sequence and the crystallized structures of pectin lyase B (PDB 1QCX) and pectin lyase A (PDB 1IDK) from *Aspergillus niger* as templates. The models generated using this method were evaluated considering the quality values implemented by I-TASSER (c-score $= 1.52$, TM-score $= 0.93 \pm 0.06$, and estimated RMSD $= 3.5 \pm 2.4$ Å). From the best-quality model according to I-TASSER, a high-resolution refined protein structure was generated using ModRefiner (Xu and Zhang 2011) software. After the model was refined, an RMSD of 0.422 and a TM-score of 0.9963 to the initial model and an energy force field value of $-154,324$ kJ/mol were obtained. The quality of the model was assessed by plating the dihedrals φ and Ψ onto Ramachandran plots using SPDBV v4.1 (Arnold et al. 2005), obtaining 4 non-glycine residues outside the allowed regions corresponding to 1%. The Visual Molecular Dynamics program (VMD v 1.9.3) (Humphrey et al. 1996) was used for the management and visualization of the generated structural model and for localization of sites with evidence of selection and other important sites in the analysis.

Results

We identified 36 genes encoding pectin lyases from 14 species of *Colletotrichum* reported in databases (Table 1; Online Resource Table 1). The coding sequences are highly variable at the nucleotide and amino acid levels. The Jukes–Cantor distance between nucleotide sequences was from 0.00–0.01 (even between sequences from different species) to 0.55 and the identity to amino acid level was from 100 to 99.73

Table 1 List of species used in this study

Organisms	Strain	Accession no.	Source/Host	Other host	Lifestyle	Refs.
C. lindemuthianum	1472	JN034039	Phaseolus vulgaris		Plant pathogen	Martin et al. (2017)
C. lindemuthianum	83.501	MASP02000055 MASP02000590	Phaseolus vulgaris		Plant pathogen	Martin et al. (2017)
C. gloeosporioides	sp. malvae	AF156984	Malva pusilla		Plant pathogen	
C. gloeosporioides	Nara GC5	XM_007279216 KB020251 KB020859 KB021063	Strawberry	Coffee, stylo, almond, avocado, apple, citrus, Cashew, cacao	Plant pathogen, endophyte	Wang et al. (2008; Cannon et al. (2012)
C. gloeosporioides	CG 14	AMYD01000361	Avocado cv. Fuente	Olive, yam, Mango, passion-fruit, tamarillo, persimmon, jasmine, tomato	Plant pathogen	Cannon et al. (2012)
C. orbiculare	MAFF 240422	KB725956 KB725912 KB725976 KB726116	Cucumis sativus	Melon/cucumber, Malvaceae, Phaseolus	Plant pathogen	Cannon et al. (2012)
C. simmondsii	CBS 122122	JFBX01000473 JFBX01000279	Carica papaya	Safflower	Plant pathogen	(Helder et al. (2016)
C. fioriniae	PJ7	XM_007601225 XM_007591291	Fragaria × ananassa		Plant pathogen, entomopathogen	Marcelino et al. (2008)
C. salicis	CBS 607.94	JFFI01002365 JFFI01002131	Salix sp.	Acer, populus	Plant pathogen	Helder et al. (2016)
C. tofieldiae	0861	LFIV01000012 LFIV01000159	Arabidopsis thaliana	Tofieldia calyculata, Lupinus polyphyllus	Endophyte	Damm et al. (2009; Hiruma et al. (2016)
C. incanum	MAFF 238712	KV842124 KV841940 KV841764	Raphanus sativus, Japanese daikon radish	Arabidopsis, lily, maize	Plant pathogen	Gan et al. (2016)
C. incanum	MAFF 238704	LFIW01002722	Raphanus sativus	Soybean	Plant pathogen	Helder et al. (2016)
C. orchidophilum	IMI 309357	MJBS01000013 MJBS01000071	Phalaenopsis sp.		Plant pathogen	
C. higginsianum	IMI 349063	XM_018305831 XM_018301484 XM_018299863	Brassica rapa subsp. Chinensis		Plant pathogen	
C. graminicola	M1.001	XM_008097032		Corn, Poaceae, Agrostis, Centipedegrass, sorghum	Plant pathogen	Cannon et al. (2012)
C. chlorophyti	NTL11	MPGH01000035 MPGH01000111	Solanum lycopersicum		Plant pathogen	
C. sublineola	TX430BB	JMSE01000824	Sorghum bicolor	wild rice	Plant pathogen	Baroncelli et al. (2014a)
C. nymphaeae	SA-01	JEMN01001530 JEMN01001095	Fragaria x ananassa	Apple bitter rot	Plant pathogen	Helder et al. (2016)

to 33.63% (Online Resource Fig. 2). However, all deduced amino acid sequences contain the conserved Pec-lyase_C domain (Pfam Accession PF00544), which characterizes these proteins. The protein alignment used as a guide for the alignment of nucleotide sequences shows the single-domain, its position, as well as the amino acids reported as catalytic in PNLs (Fig. 1).

The phylogenetic analyses of the PNL protein-coding genes showed similar overall topologies using both of the reconstruction methods: Bayesian and ML (Fig. 2). The sequences integrated two main clades designated C1 and C2, in which C1 showed two subclades or lineages of conserved PNLs (S1C and S1D) from *C. orbiculare* MAFF 240422, *C. gloeosporioides* Nara gc5, and *C. incanum* MAFF238712. On the other hand, in C2, a sequence of *C. higginsianum* IMI 349063 is basal and a large subclade included the rest of the sequences in several lineages. The phylogenetic analysis of *Colletotrichum* species to the taxonomic level resulted in nine clades (Cannon et al. 2012). Although the objective of this study was not taxonomic, we marked the correspondence of the clades in agreement with Cannon et al. (2012) to assist phylogenetic analyses of pectin lyases (Fig. 2). According to these analyses, two sequences of *C. orbiculare* MAFF 240422, three of *C. gloeosporioides* Nara gc5 (both pathogens of dicots), and one of *C. incanum* MAFF238712, a pathogen of dicots and monocots (Gan et al. 2016) (taxonomic clades orbiculare/gloeosporioides/spaethianum, respectively) were exclusive in C1. In C2, a sequence of *C.*

higginsianum IMI 349063 is basal, and subclades S2A and S2B integrated two large lineages of PNLs with the same pattern of distribution to each lineage in two new subclades for the same species (see S2A1 vs. S2B1 and S2A2 vs. S2B2) (taxonomic clades orbiculare/gloeosporioides, and acutatum/spaethianum/destructivum). The PNLs included in S2A1 and S2B1 corresponded to pathogens of dicot plants. The PNLs in S2A2 and S2B2 integrated two lineages each, one with sequences belonging to pathogens of dicots and graminicolous monocots and a second with sequences from pathogens of dicots, monocots, or both. Notably, S2A2 contains a lineage of PNLs belonging to *C. graminicola* M1 001 and *C. sublineola* TX430BB (taxonomic clade graminicola), two species that exclusively infect monocot plants. S2A2 also included the sequences of *C. tofieldiae* 0881, a dicot pathogen, and *C. incanum* MAFF238712, a pathogen of dicots and monocots. PNLs from *C. orchidophilum* IMI 309357, a monocot pathogen, were integrated into S2A2 and S2B2. Notably, the sequences of *C. chlorophyti* NTL11, a dicot pathogen, were included as basal in S2A and S2A2.

Next, we used codon-based models to test for heterogeneous selection pressures imposed on PNL genes of *Colletotrichum* using the phylogenetic tree. First, we calculated ω for all ramifications under a one-ratio model giving a value of 0.232 ($P < 0.001$, Table 2), indicating as expected, that the analyzed genes are under purification selection since it is common in enzymes to keep their functions experiment constrained selective pressures (Brunner et al. 2013). The
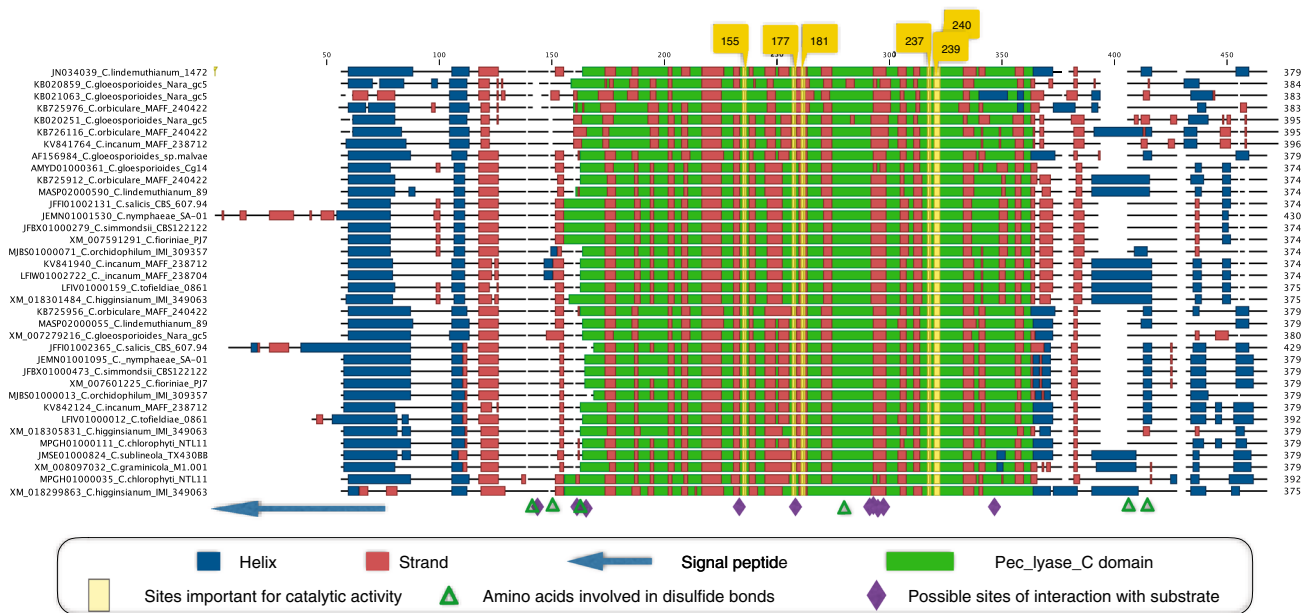
Fig. 1 Schematic representation of the alignment of the 36 PNL sequences of *Colletotrichum* sp. The scheme shows the prediction of secondary structure elements, signal peptide prediction, Pec-lyase_C domain location, the position of catalytic amino acids, and other important sites reported in PNLs. The code of structural elements and important sites is shown in the lower box. The CLC Main Workbench 7.8 software was used to make the schematic representation of the secondary structure and domain of PNLs
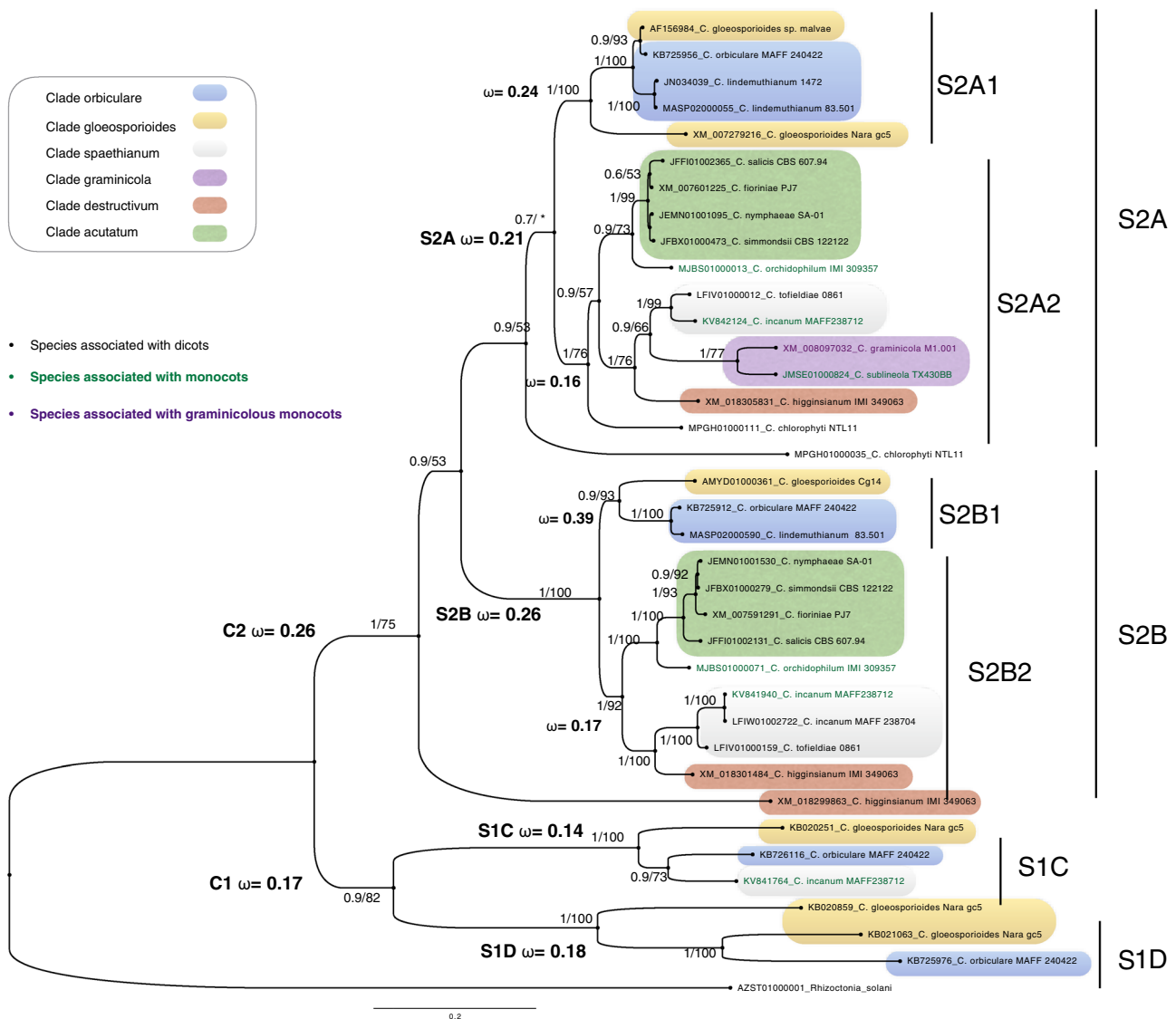
Fig. 2 Phylogenetic relationships of pectin lyase genes of *Colletotrichum* sp. The phylogenetic analysis shows the 50% majority rule consensus Bayesian tree. The Bayesian posterior probabilities values are indicated. Numbers below the diagonal indicate bootstrap values inferred from the ML analyses. The asterisks represent branches that were not supported by 50% or more of the bootstraps. The scale bar represents the number of substitutions per site. The clades defined and used for the analysis in the tree are shown in bold with the ω obtained in the branch model analysis using PAMLX v1.3.1. The phylogenetic tree was edited using FigTree v1.4.3 software. The colors of the boxes correspond to the clade to which they belong, and the letter color corresponds to the type of host. The numbers indicate the use of more than one sequence in a subclade

free-ratio model shows a better fit to our data than the one-ratio model ($P < 0.001$). We then applied two-ratio models to determine the existence of different selection pressures among clusters formed in the phylogenetic reconstruction (Fig. 2). According to the results, the main clades C1 and C2 and the subclades have experienced significantly different selective pressures (Table 2). Although the phylogenetic reconstruction of the pectin lyases from *Colletotrichum* shows evidence of purifying selection (ω 0.232), the analysis also shows significant evidence that may represent relaxation in the purifying selection or the action of positive selection, which would explain the formation of the different subclades observed. The basal clade of the phylogenetic tree (C1 ω = 0.17 $P < 0.001$) shows the lowest ω, so the genes that comprise it are clearly under purifying selection.

Multiple ML methods were implemented to test the selective pressures imposed on individual sites (Table 3). A potential positively selected site was detected using the M2 and M3 models and four were detected using the M8 model. However, only the comparison between the M3 versus M0 model was statistically significant, $2\Delta\ln L = 784,55$ ($P < 0.001$), and the 324 site that was positively selected

Table 2 Likelihood ratio test of branch models on PNLs genes

Model	Np[a]	Ln likelihood	Omega values (ω)		Model compare	2Δln$L$[b] (p value)
A: omega = 1	73	− 16666.365				
B: one-ratio	74	− 16043.873	0.232		B vs. A	1244.982 ($P < 0.001$)
C: free-ratio	145	− 15756.565	variable ω by branch		C vs. B	574.616 ($P < 0.001$)
D: C1 and C2[b]	76	− 16027.751688	0.266	0.170	D vs. B	32.24443 ($P < 0.001$)
E: S2A and S2B[b]	76	− 16040.135579	0.211	0.284	E vs. B	7.476648 ($P < 0.05$)
F: S2A1 and S2A2[b]	76	− 16031.361713	0.243	0.160	F vs. B	25.02438 ($P < 0.001$)
G: S2B1 and S2B2[b]	76	− 16035.829	0.397	0.173	G vs. B	16.089806 ($P < 0.001$)
H: S1C and S1D[b]	76	− 16032.273902	0.144	0.182	H vs. B	23.200002 ($P < 0.001$)

[a]Number of parameters

[b]Clades corresponding to the main branches of the phylogenetic tree (Fig. 2)

Table 3 Site model tests on PNLs genes of *Colletotrichum* sp.

Model	Np[a]	Ln likelihood	Parameter estimates	Model compare	Positive selection sites[b]	2Δln$L$ (P value)
M0 (one ratio)	74	− 16043.87	$\omega = 0.23262$		None	
M1 (nearly neutral)	75	− 15845.40	$P_0 = 0.655$ $P_1 = 0.344$ $\omega_0 = 0.162$ $\omega_1 = 1.000$		Not allowed	
M2 (positive selection)	77	− 15845.40	$P_0 = 0.655$ $P_1 = 0.344$ $P_2 = 0.000$ $\omega_0 = 0.162$ $\omega_1 = 1.000$ $\omega_2 = 27.922$	M2 vs. M1	324 (0.837)	ns
M3 (discrete)	82	− 15651.59	$P_0 = 0.184$ $P_1 = 0.355$ $P_2 = 0.407$ $P_3 = 0.049$ $P_4 = 0.004$ $\omega_0 = 0.008$ $\omega_1 = 0.133$ $\omega_2 = 0.412$ $\omega_3 = 0.843$ $\omega_4 = 1.681$	M3 vs. M0	324 (0.798)	784.55 ($P < 0.001$)
M7 (β)	75	− 15657.65	$P = 0.717$ $q = 1.978$			
M8 (β and ω)	77	− 15656.28	$P_0 = 0.995$ $p = 0.734$ $q = 2.079$ $p_1 = 0.004$ $\omega = 1.759$	M8 vs. M7	285 (0.637), 308 (0.566), 324 (0.965*), 351 (0.662)	2.73 ($P = 0.25$)

The position of the positively selected sites referred to the coding sequence of the pectin lyase of *C. lindemuthianum* (Genbank: JN034039)

ns not significant

[a]Number of parameters

[b]Positively selected sites were indicated by the BEB approach. Asterisks indicate posterior probability $P > 95\%$

using model M3 did not meet the statistical significance criterion of later probability $P > 0.95$ for the BEB method. Although the comparisons of nested models M1 and M2 (P value = 0) and M7 and M8 (P value = 0.25) were not statistically significant, the comparison of the M3 and M0 models was a test of heterogeneity of ω values between sites, providing evidence of the participation of positive selection in the diversification of PNLs. We also applied several models using the maximum likelihood approach in the HyPhy program on the Datamonkey server. The evaluation of positive selection using SLAC and FEL models showed four and two sites, respectively, under selection at a P value of 0.2 (Table 4), but no sites under positive selection were found at a P value of 0.1, which may indicate weak positive selection signals.

Because these methods tend to be conservative, an analysis for signatures of episodic positive selection was performed using MEME (Murrell et al. 2012). MEME is a mixed-effects evolution model capable of identifying episodic and pervasive positive selection at the individual site level, in which episodes of positive selection are confined to a small subset of branches in a phylogenetic tree. We identified 15 sites with evidence of positive selection at a P value < 0.1 using this method (Table 4), four of which had already been detected using site-specific methods. We also explored the variation of evolutionary rates in specific sites using branch-site models using IFEL and BUSTED methods implemented on the Datamonkey server. We identified three sites with evidence of positive selection using IFEL at a P value < 0.2 and one site with evidence of positive selection

Table 4 Phylogenetic test of positive selection in PNLs genes of *Colletotrichum* sp.

MODEL	PAML[b]	SLAC[c]	FEL[c]	IFEL[c]	MEME[d]	BUSTED[e]	TOTAL[a]
Positive selected sites	285, 308, 324*, <u>351</u>	<u>24</u>, 311, <u>338</u>, <u>351</u>	<u>66</u>, <u>338</u>	<u>39</u>, <u>66</u>*, 325, <u>338</u>	<u>24</u>, <u>39</u>, <u>66</u>, <u>74</u>, <u>100</u>, 128, <u>254</u>, 270, 275, <u>300</u>, 323, <u>338</u>, 345, <u>348</u>, <u>351</u>	<u>24</u>, 34, <u>39</u>, <u>74</u>, <u>100</u>, 112, 181, <u>254</u>, 285, <u>300</u>, 303, <u>338</u>, 339, <u>348</u>, <u>351</u>, 367, 370	24, 39, 66, 74 100, 254, 300, 338, 348, 351
Positive selected sites GARD-inferred trees		24, <u>338</u>*, <u>351</u>	<u>66</u>, <u>338</u>	<u>39</u>, <u>66</u>*, 149, 164, <u>338</u>*	<u>24</u>, <u>39</u>, <u>66</u>, <u>74</u>, <u>100</u>, 128, 254, 270, 275, <u>300</u>, 323, <u>338</u>, <u>345</u>, <u>348</u>, <u>351</u>, <u>370</u>	34, 73, <u>74</u>, <u>100</u>, 103, 120, 127, 181, 239, 247, 285, 291, <u>300</u>, 303, 324, <u>338</u>, 339, <u>345</u>, <u>348</u>, <u>351</u>, <u>370</u>	66, 74, 100, 300, 338, 345, 348, 351, 370

Codons identified by more that one method are underlined

[a]Sites with evidence of selection detected by at least two of the methods with the criteria of statistical significance

[b]Potential sites identified as positively selected in PAML under the M8 model are shown despite not passing the LRT statistical test for comparative purposes. Positively selected sites are indicated by BEB approach. Asterisks indicate posterior probability $P > 95\%$

[c]Codons with P values < 0.2 for SLAC, FEL, and IFEL methods. Asterisks indicate P values < 0.1

[d]Codons with P values < 0.1 for MEME

[e]Codons with P value < 0.05 for BUSTED

at a P value < 0.1. Using the BUSTED method, 17 sites with evidence of episodic positive selection at a P value < 0.05 were identified, nine of which had already been identified by other methods. To improve the accuracy of the results, sites with evidence of selection should be detected using at least two of the methods with the criteria of statistical significance (P value < 0.1 for SLAC, FEL, MEME, and IFEL and P value < 0.05 for BUSTED). With these parameters, we detected 10 sites with evidence of positive selection in the PNL genes analyzed.

Recombination can affect several analyses, including phylogenetic reconstruction and analysis of positive selection (Anisimova et al. 2003). Therefore, we assessed gene recombination using the genetic algorithm recombination detection (GARD) (Pond et al. 2006c). The GARD analysis showed evidence of a breakpoint at position 307 of the alignment used for the selection analysis, with significant topological incongruence using AIC of 20.7581 and a P value of 0.05 for the topological incongruity analysis using the KH test. Phylogenetic reconstructions generated by the GARD from partitions corresponding to non-recombinant alignment fragments allow for the analysis of the existence of topological inconsistencies among phylogenies (Fig. 3) and the identification of sequences or clades involved in the possible recombination event. We found topological inconsistencies in the basal clades of trees (C1) formed by sequences of *C. gloeosporioides* Nara gc5, *C. incanum* MAFF238712, and *C. orbiculare* MAFF 240422. We also observed two possible recombinants with topological incongruities in a sequence of *C. higginsianum* IMI 349063 (GenBank

XM_018299863) and a sequence of *C. chlorophyti* NTL 11 (GenBank MPGH01000035). Low values were also observed in the metric comparisons between trees in clade C2, in which a sequence of *C. chlorophyti* is notable (GenBank MPGH01000111). To analyze the effect of recombination on selection, the above analyses were repeated using the non-recombinant partitions. We identified one site with evidence of positive selection in the analysis using the SLAC method and two positively selected sites using the IFEL method, both with a P value < 0.1, which shows an improvement in the reliability of the analyses. Considering only the sites with evidence of positive selection detected using at least two methods with the appropriate statistical significance criteria, nine sites with evidence of positive selection were obtained, seven of which were identified in the initial analysis (Table 4; Online Resource Table 3).

Next, to analyze the possible role of the sites identified to have evidence of positive selection in the PNL sequences of *Colletotrichum* used in this work, we identified amino acids with a catalytic role and amino acids that are important in the interaction with the substrate, which have been reported in pectin lyases and crystallized pectate lyases. The sites identified in the analysis of multiple alignments of protein sequences (Online Resource Fig. 1) and in pectin lyase 2 of *C. lindemuthianum* (GenBank JN034039) were used to illustrate the results (Fig. 4a).

Three invariant residues have been reported in the active region of the pectin lyase (PelC) of *Erwinia chrysanthemi* (Herron et al. 2003), which have the same orientation and conformation in pectin lyase B (PLB 1QCX) of *A. niger*
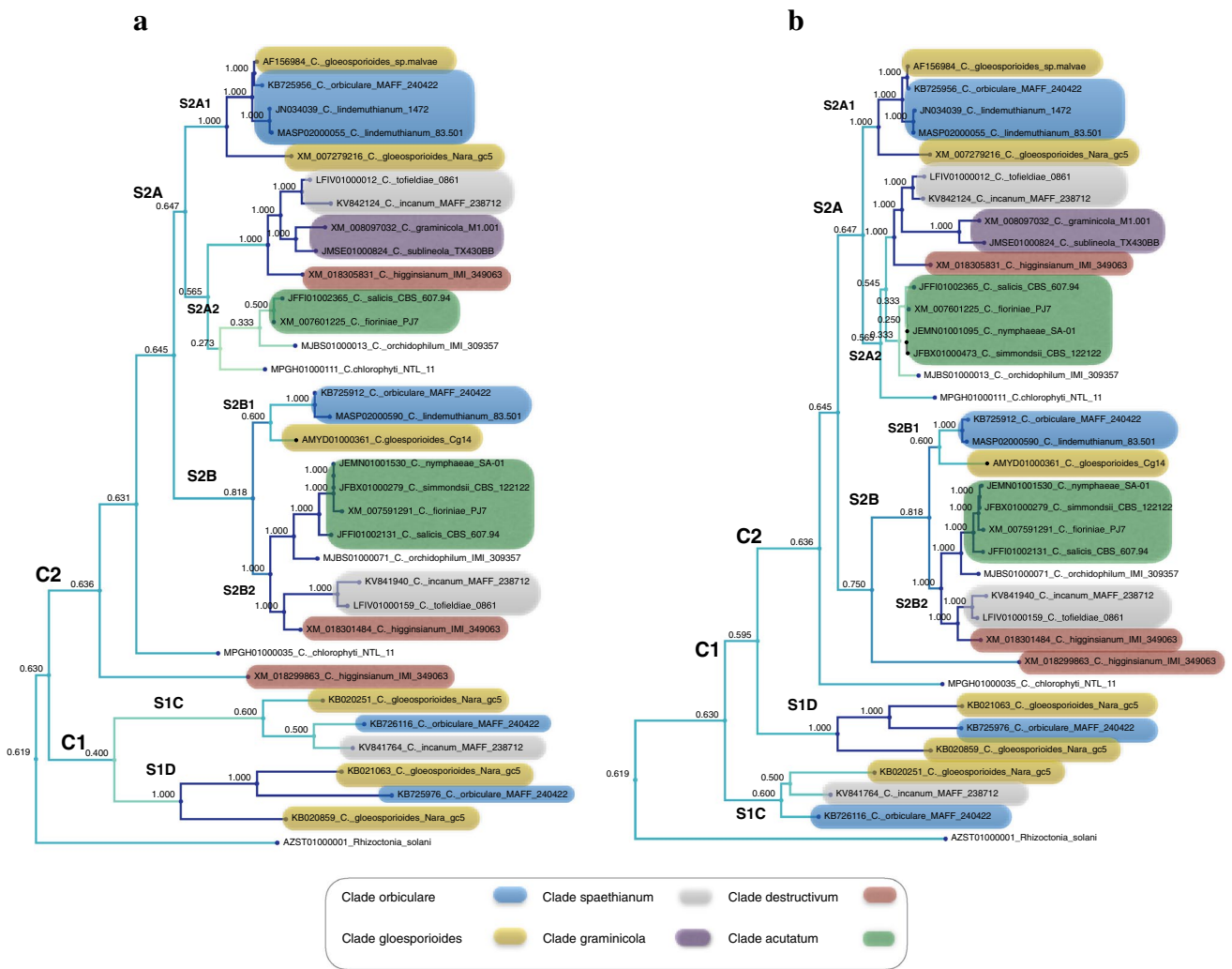
Fig. 3 Comparisons of phylogenetic trees generated from the non-recombinant fragments using GARD. **a** GARD-inferred tree fragment 1-306. **b** GARD-inferred tree fragment 310–996. The colors of the boxes correspond to the clade to which they belong. Missing sequences in the tree that correspond to the non-recombinant fragment 1–306 were deleted during the GARD analysis on the Datamon- key server. The numbers correspond to metric comparisons, and a score of 1 indicates that the subtree structure of the node is identical to the subtree structure of its best corresponding node according to the analysis using phylo.io version 1.0.0 (http://phylo.io/#) (Robinson et al. 2016)

(Vitali et al. 1998). We found these residues in the structural model of pectin lyase 2 from *C. lindemuthianum* (Fig. 4b) and conserved in the multiple alignments of the analyzed pectin lyases (Asp 174, Arg 256, and Pro 258 according to the PNL sequence of *C. lindemuthianum*). It is believed that the role of Arg 256 is to provide a positive charge to neutralize the negatively charged intermediate generated by the initial proton abstraction during the β-elimination pathway (Vitali et al. 1998), whereas the role of Ca^{2+} in pectate lyases is provided by amino acids Arg 196 and Val 200 in pectin lyases. The amino acids Lys 259 and Asp 241 have been considered important due to their possible stabilizing role and are found in the Pec_lyase_C domain (Vitali et al. 1998; Herron et al. 2000; Sánchez-Torres et al. 2003). Only Ser 254, identified with positive selection evidence, was found in the Pec_lyase_C domain very close to Arg 256, for which an important role in pectin lyase activity has been confirmed (Vitali et al. 1998).

According to an analysis of the PLB of *A. niger*, the putative oligosaccharide-binding region is formed by a large aromatic cluster (Mayans et al. 1997; Vitali et al. 1998) and favors binding to the uncharged substrate (highly esterified pectin). A similar cluster was found in the structure of pectin lyase 2 of *C. lindemuthianum* (Fig. 4a, b) and in the remaining sequences used in this work, with some changes (that maintain the aromatic nature of the cluster). It would be interesting to analyze the possible effect of the variation found between amino acids that have been reported as part
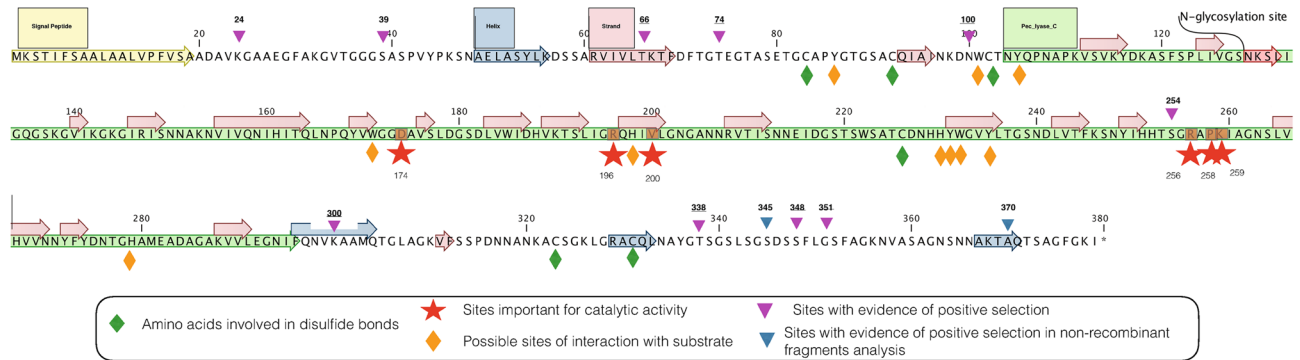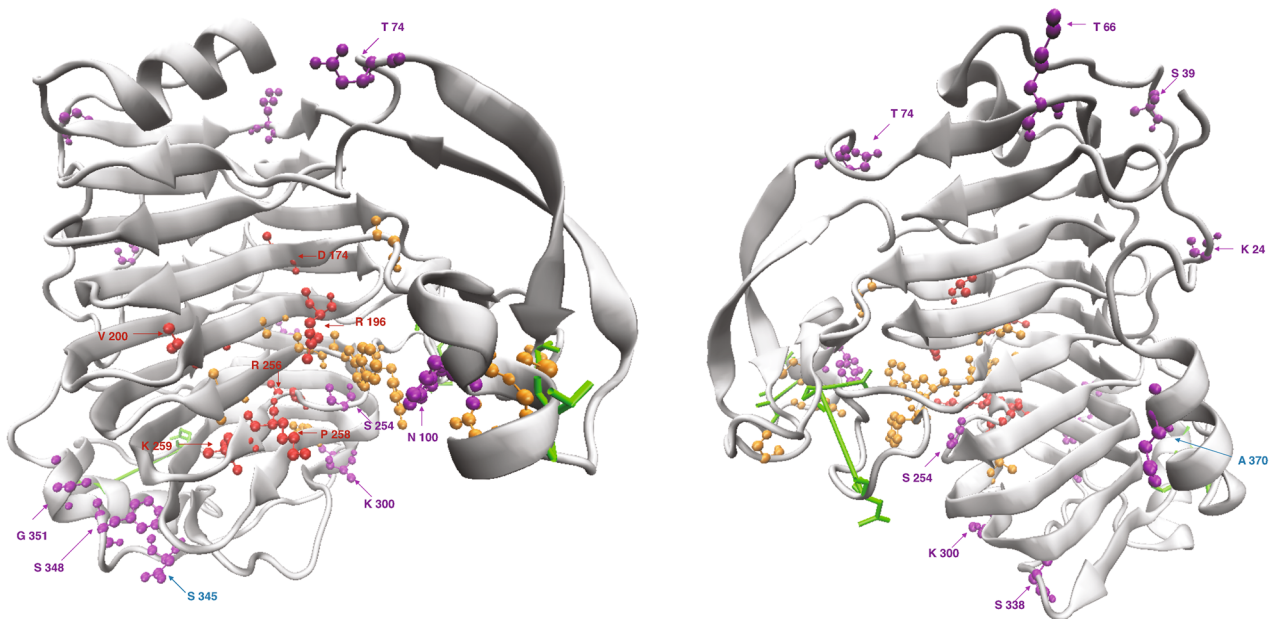
Fig. 4 Important sites in the analysis of PNLs and sites with evidence of selection with PNL 2 of *Colletotrichum lindemuthianum* used as a representative. **a** Amino acid sequence of PNL 2 of *C. lindemuthianum* (GenBank JN034039) showing the predicted signal peptide, predicted N-glycosylation site, secondary structure elements represented in blue helices and pink strands, and the Pec-lyase-C domain shown in green. The code of the important sites in the sequence is shown in the lower box. Sites with evidence of positive selection in the initial analysis and in the analysis of non-recombinant fragments are underlined. **b** Structural model of PNL 2 of *C. lindemuthianum* generated using I-TASSER (Roy et al. 2010) and refined with ModRefiner (Xu and Zhang 2011), showing the location of important sites in the sequence. The colors and numbering correspond to those in the PNL 2 sequence in panel a. Three-dimensional structure images were generated using VMD v 1.9.3 software (Humphrey et al. 1996)

of the oligosaccharide-binding region among the PNLs of *Colletotrichum*. However, in this work, we found no evidence of positive selection in any of these amino acids but found evidence of positive selection in Asn 100, Lys 300, and Ser 254; as noted above, Ser 254 is very close to the catalytic amino acid Arg 256 in the structural model of pectin lyase 2 of *C. lindemuthianum*. These amino acids are found in the same oligosaccharide-binding region and, therefore, may be related to interaction with the substrate.

The sites identified to have evidence of positive selection Gly 351, Ser 348, Ser 345, Ala 370, Lys 24, Ser 39, Thr 66 were in the N- and C-terminal tails, which correspond to the peripheral helices in the three-dimensional structure, and sites Thr 74 and Asn100 were found on the antiparallel sheet. A high degree of structural conservation between pectin and pectate lyases in regions distant from the active site has been reported and the N- and C-terminal tails contribute to maintaining the β-helix architecture (Vitali et al. 1998). Our negative selection results support these reports because sites identified to have evidence of negative selection appear to be related to the preservation of the structure of the central body of the protein, the peripheral helices, and

the antiparallel sheet (*P* values < 0.1 for SLAC, FEL, and IFEL methods) (Fig. 5, Online Resource Table 4). We also found three conserved disulfide bonds between Cys 83 and Cys 102, Cys 92 and Cys 226, Cys 323 and Cys 331, which correspond to those found in PLB of *A. niger* (Vitali et al. 1998), that stabilize the structure. Interestingly, in our work, the most sites with evidence of positive selection were found in these regions.

Discussion

The phylogenetic analysis of *Colletotrichum* by Crouch et al. (2014) mentions the taxonomic clades orbiculare and gloeosporioides as part of the basal lineages and, according to our analyses, the enzymes of C1 that also include a sequence of *C. incanum* MAFF238712 (taxonomic clade spaethianum) could represent enzymes related to the ancestral lineages of the genus. Apparently, purifying selection has maintained the function of these enzymes, and the rest of the sequences encoding PNLs in clade C2 show evidence of relaxation in purifying selection. C2 revealed two large lineages of PNLs (Fig. 2), products of an ancestral genetic duplication. Notably, there was a topology in which these subclades included similar new genetic duplications of PNLs in each species, which were grouped as representatives of taxonomic clades orbiculare/gloeosporioides (S2A1 and S2B1), and acutatum/spaethianum/destructivum (S2A2 and S2B2). In agreement with Nei and Kumar (2000) the topology of the tree suggest that pectin lyases come from events of genetic duplication from a common ancestor and multiple events of duplication that occurred before and after the speciation.

According to the phylogenetic analyses conducted in *Colletotrichum*, pathogenic species of graminicolous monocots form a monophyletic group in the graminicola clade

and are the product of a trait derived from a relatively recent origin of ancestral lines of non-graminicolous *Colletotrichum* (O'Connell et al. 2012; Crouch et al. 2014). In this study, we only found a sequence of *C. graminicola* and *C. sublineola* (species associated with monocots) in S2A2. The absence of a similar subclade in S2B2 can be explained by a reduction in the number of pectinases, which has been reported in *C. graminicola* and other monocot pathogenic fungi in response to a lower pectin concentration in the host cell walls (Soanes et al. 2008; O'Connell et al. 2012; Zhao et al. 2013; Lo Presti et al. 2015). The analysis of *C. incanum* sequences is particularly interesting because it has been reported to be closely related to the graminicola clade, whose members are strictly restricted to monocot hosts and to the destructive clade, whose members are mostly associated with dicot infections (Gan et al. 2016). Our results confirm these relationships because in subclade S2A2, a *C. incanum* sequence is related to the sequences of the graminicola and destructivum clades, and two of the sequences of *C. incanum* are related to the destructivum clade in subclade S2B2. However, in subclades S2A2 and S2B2, the *C. incanum* sequences are grouped with *C. tofieldiae*, a species of the spaethianum clade. Moreover, a *C. incanum* sequence was also found in C1, where it is apparently under a purifying selection regime with sequences of the orbiculare and gloeosporioides clades in a well-supported relationship.

Adaptation to a specific substrate can reduce the diversity of non-synonymous codon sites relative to synonymous sites due to the functional optimization of the enzymatic activity, which would leave evidence of purifying selection on these enzymes (Brunner et al. 2013). In a selection analysis conducted in paralogs of the endopolygalacturonases of *Botrytis cinerea*, no variability was found in one of them, which was attributed to the efficiency of the enzyme in the late phases of infection (Cettul et al. 2008). According to

Fig. 5 Structural model of PNL 2 of *C. lindemuthianum* showing sites with evidence of positive and purifying selection. Sites with evidence of purifying selection are shown in blue, and sites with positive selection are shown in purple
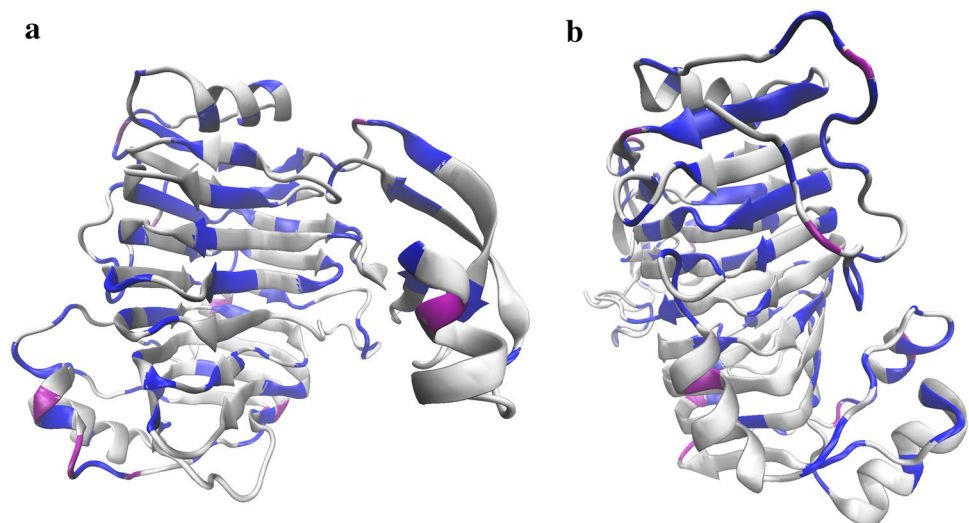
our analysis, it is likely that, during the evolutionary history of PNLs, episodes of gene duplication occurred and were followed by diversification processes. It is believed that most of the families of PCWDEs originated from genetic duplication (Wapinski et al. 2007). An evolutionary study of PL1 (PNL family) conducted in 103 fungi representative of Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota suggested that the last common ancestor of the fungi possessed numerous paralogs of genes of the PL1 family, some of which were lost during the evolution of fungal taxa. However, events of genetic duplication also have occurred in some cases, particularly in plant pathogenic fungi (Zhao et al. 2013). It appears that pectinase evolution has a history of expansions and contractions related to nutritional strategies, which are also apparent in the history of PNLs within *Colletotrichum*.

It has been suggested that expansions and contractions of gene families may play a role that is as important in adaptation as the changes induced by nucleotide substitution (Lynch and Conery 2000). Duplicate genes may become pseudogenes by the accumulation of deleterious mutations, but it is believed that they can be maintained if there are advantages in increasing the dose or divergence in the regulatory control (Wapinski et al. 2007; Cornell et al. 2007). They may undergo a relaxed purifying selection period that allows for evolutionary divergence (neofunctionalization) (Lynch and Conery 2000). After genetic duplication, the natural selection can favor the fixation of mutations in one or both copies (positive selection), allowing for diversification. The purifying selection then acts to maintain the new function (Lynch and Conery 2000; Bielawski and Yang 2003).

In the selection analysis performed on *Colletotrichum* PNL-encoding genes, we identified 10 sites with evidence of positive selection using different evolutionary models. The methods that allow for analysis of the signatures of episodic positive selection were more informative, probably due to the proximity of the sequences analyzed or because of the manner of evolution in this type of gene. Recombination is an important process for most fungi, and it can occur in meiosis, by mitotic recombination in some fungi or through the parasexual cycle (Lamb 2003). The recombination can have an adverse effect on the power and accuracy of molecular evolution analysis tools such as phylogenetic reconstructions (Anisimova et al. 2003) and positive selection detection (Shriner et al. 2003). Thus, it is recommendable to analyze if there is evidence of recombination in the sequences of interest and topological incongruities in the phylogenetic reconstructions, and a selection analysis should be conducted using non-recombinant fragments. According to the analysis of the phylogenetic reconstructions of the non-recombinant fragments generated by GARD, there are small topological differences, mainly in C1 and S2A2 (Fig. 3). From the positive selection analysis of the non-recombinant fragments,

evidence of positive selection was found for one less codon (9 codons) than in the selection analysis performed with the complete alignment (10 codons). However, two new codons with evidence of positive selection were detected and more codons with positive selection were found using the SLAC and IFEL methods with P value < 0.1. Consideration of the possible recombination event resulted in the detection of new codons with positive selection evidence and an increase in confidence values.

Only one site with positive selection evidence was identified in the Pec_lyase_C domain very close to Arg 256, for which an important role in pectin lyase activity has been confirmed (Vitali et al. 1998). Additionally, we found two sites on a three-dimensional model generated from PNL2 of *C. lindemuthianum* that are also found in the putative oligosaccharide-binding region, which is characterized by the formation of a large aromatic cluster, seven sites in the N- and C-terminal tails, and two in the antiparallel sheet that may be involved in the interaction with the substrate. Selection analyses conducted in other PCWDEs have reported a low proportion of sites with evidence of positive selection (Cettul et al. 2008; Brunner et al. 2009), which is expected because the evolutionary shift from one enzymatic function to another involves crossing an energetic barrier in a fitness landscape (Romero and Arnold 2009). Whereas some residues are critical for maintaining protein stability, others maintain their catalytic activity and, because most mutations are destabilizing, mutations that lead to the modification of enzyme activity may decrease protein stability, so compensatory mutations are needed to restore global stability (DePristo et al. 2005; Tokuriki et al. 2008). Thus, it is expected in the case of PNLs (and other PCWDEs), where the same catalytic function is maintained, there will be a limited number of sites where mutations that confer advantageous characteristics can be fixed. Thus, it is logical to discover weak and limited positive selection signals and strong negative selection signals that maintain the structural and functional integrity of the PNLs.

The selection analysis of a family of endopolygalacturonases from *B. cinerea* revealed diversifying selection in the majority of the analyzed genes, which can contribute to escape from recognition by the host (Cettul et al. 2008). Similar results were reported in an analysis performed on 189 strains of *Mycosphaerella graminicola*, in which sites with evidence of positive selection in β-xylosidase genes were located on the surface of the protein, suggesting a relation to the interaction with the host defense system (Brunner et al. 2009).

Several inhibitory proteins of PCWDEs have been identified; the most studied of which are xylanase inhibitory proteins (Fierens et al. 2007; Dornez et al. 2010) and polygalacturonase inhibitory proteins (Albersheim and Anderson 1971; Nuss et al. 1996; De Lorenzo et al. 2001). These

proteins are apparently part of polymorphic families of plant defense systems, are highly specific inhibitors, have a specificity determined by a few amino acids, and have generally been found to inhibit enzymes in a competitive manner by blocking the active site (Leckie et al. 1999; Gebruers et al. 2001; Federici et al. 2001, 2006; D'Ovidio et al. 2004; Fierens et al. 2007; Pollet et al. 2009; Dornez et al. 2010). PNL inhibitory proteins have been identified in extracts of onion, French bean, sweet paprika, white cabbage, cucumber, and sugar beet and apparently have inhibitory specificity (Bock et al. 1975; Bugbee 1993), but the sequences of these proteins have unfortunately not been reported. However, the reports of PNL inhibitory proteins represent evidence of a possible coevolutionary arms race and suggest the possible role of the seven sites with evidence of positive selection located in the N- and C-terminal tails of the structural model. Notably, these regions appear to have strong negative selection signals, which may indicate their possible importance at a structural level and in the interaction with the host defense system.

The rest of sites with evidence of positive selection identified in this work are a potential source of analysis because they are in the oligosaccharide-binding region and, therefore, may be related to interaction with the substrate. However, to study their role in pathogenesis, different types of analysis are required to determine whether a relationship exists between amino acids with evidence of positive selection and the preference of substrate, catalytic differences, stability, and interaction with the host defense system.

Conclusions

The history of the adaptive evolution of PNLs in *Colletotrichum* sp. is complex and can be influenced by multiple factors. On one hand, we observed a lower number of PNL-encoding genes in *Colletotrichum* species associated with monocot plants than with dicot plants. This is consistent with the reports of pectinases in *Colletotrichum* genomes and other fungal genomes where the loss of pectinolytic genes has been associated with pectin content in the host cell wall. The ability of *C. incanum* MAFF238712 to infect both dicotyledonous and monocotyledonous plants has been reported, and this is clearly reflected in our analysis where their PNL-encoding genes are related to sequences of species associated with dicotyledonous plants together with *C. incanum*. This seems to be another indication of the influence of pectin content in the cell wall on the pectinases evolution and specifically on the PNLs evolution in the genus *Colletotrichum*.

On the other hand, it is probable that several events of genetic duplication occurred from ancestral lines where the new copies have been diversified during episodes of

relaxation of the great purifying selection force that maintains the enzymatic function. These episodes of diversification have forged different evolutionary trajectories that appear to form patterns among *Colletotrichum* species, even though they share the same type of host. With current information it is difficult to define the forces that have influenced the diversification and evolution of observed PNL lineages; however, taking into account the background and possible role of amino acids with evidence of positive selection identified, it is likely to be related to host interaction and substrate specificity.

Investigation of the evolution of PNLs and of pectinolytic enzymes is fundamental to studying their role in pathogenesis and biotechnological importance because it generates information that can be used to study the role of key sites for interaction with the host in more detail and focus the analyses on sites that are potentially important due to their catalytic efficiency or substrate specificity.

References

Aguileta G, Refrégier G, Yockteng R et al (2009) Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. Infect Genet Evol 9:656–670. doi:10.1016/j.meegid.2009.03.010

Aguileta G, Lengelle J, Chiapello H et al (2012) Genes under positive selection in a model plant pathogenic fungus, *Botrytis*. Infect Genet Evol 12:987–996. doi:10.1016/j.meegid.2012.02.012

Alaña A, Gabilondo A, Hernando F (1989) Pectin lyase production by a *Penicillium italicum* strain. Appl Environ Microbiol 55:1612–1616

Albersheim P, Anderson AJ (1971) Proteins from plant cell walls inhibit polygalacturonases secreted by plant pathogens. Proc Natl Acad Sci 68:1815–1819

Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. doi:10.1016/S0022-2836(05)80360-2

Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236

Annis SL, Goodwin PH (1997) Recent advances in the molecular genetics of plant cell wall-degrading enzymes produced by plant pathogenic fungi. Eur J Plant Pathol 103:1–14. doi:10.1023/A:1008656013255

Arnold K, Bordoil L, Kopp J, Schwede T (2005) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201. doi:10.1093/bioinformatics/bti770

Balestrieri C, Castaldo D, Giovane A et al (1990) A glycoprotein inhibitor of pectin methylesterase in kiwi fruit (*Actinidia chinensis*). Eur J Biochem 193:183–187

Baroncelli R, Sanz-Martín JM, Rech GE et al (2014a) Draft genome sequence of *Colletotrichum sublineola*, a destructive pathogen of cultivated Sorghum. Genome Announc 2:e00540–e00514. doi:10.1128/genomeA.00540-14

Baroncelli R, Sreenivasaprasad S, Sukno SA et al (2014b) Draft genome sequence of *Colletotrichum acutatum sensu lato* (*Colletotrichum fioriniae*). Genome Announc 2:e00112–e00114. doi:10.1128/genomeA.00112-14

Barrus MF (1911) Variation of varieties of beans in their susceptibility to anthracnose. Phytopathology 1:190–195

Bateman A, Coin L, Durbin R et al (2004) The Pfam protein families database. Nucleic Acids Res 32:D138–D141. doi:10.1093/nar/gkh121

Bielawski JP, Yang Z (2003) Maximum likelihood methods for detecting adaptive evolution after gene duplication. J Struct Funct Genom 3:201–212

Bock H, Dongowski G, Gobel H, Krause M (1975) Nachweis der hemmung mikrobieller pektin und pektat lyase durch pftanzeneigene inhibitoren. Nahrung 19:411–416

Brunner PC, Keller N, McDonald BA (2009) Wheat domestication accelerated evolution and triggered positive selection in the beta-xylosidase enzyme of *Mycosphaerella graminicola*. PLoS ONE 4:e7884. doi:10.1371/journal.pone.0007884

Brunner PC, Torriani SFF, Croll D et al (2013) Coevolution and life cycle specialization of plant cell wall degrading enzymes in a hemibiotrophic pathogen. Mol Biol Evol 30:1337–1347. doi:10.1093/molbev/mst041

Bugbee WM (1993) A pectin lyase inhibitor protein from cell walls of sugar beet. Phytopathology 83:63–68

Cannon PF, Damm U, Johnston PR, Weir BS (2012) *Colletotrichum* - current status and future directions. Stud Mycol 73:181–213. doi:10.3114/sim0014

Cettul E, Rekab D, Locci R, Firrao G (2008) Evolutionary analysis of endopolygalacturonase-encoding genes of *Botrytis cinerea*. Mol Plant Pathol 9:675–685. doi:10.1111/j.1364-3703.2008.00492.x

Cooper JA (1983) The mechanisms and significance of enzymic degradation of host cell walls by parasites. In: Callow JA (ed) Biochemical plant pathology. Wiley, Chichester, pp 101–137

Cornell MJ, Alam I, Soanes DM et al (2007) Comparative genome analysis across a kingdom of eukaryotic organisms: specialization and diversification in the fungi. Genome Res 17:1809–1822. doi:10.1101/gr.6531807

Crouch J, O'Connell R, Gan P et al (2014) The genomics of *Colletotrichum*. In: Genomics of plant-associated fungi: monocot pathogens. Springer, Berlin, pp 69–102

D'Ovidio R, Raiola A, Capodicasa C et al (2004) Characterization of the complex locus of bean encoding polygalacturonase-inhibiting proteins reveals subfunctionalization for defense against fungi and insects. Plant Physiol 135:2424–2435. doi:10.1104/pp.104.044644

Damm U, Woudenberg JHC, Cannon PF, Crous PW (2009) *Colletotrichum* species with curved conidia from herbaceous hosts. Fungal Divers 39:45–87

De Lorenzo G, D'Ovidio R, Cervone F (2001) The role of polygalacturonase-inhibiting proteins (PGIPs) in defense against pathogenic fungi. Annu Rev Phytopathol 39:313–335. doi:10.1146/annurev.phyto.39.1.313

Dean R, van Kan JAL, Pretorius ZA et al (2012) The top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol 13:414–430. doi:10.1111/j.1364-3703.2011.00783.x

Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457. doi:10.1093/bioinformatics/btq429

Demuth JP, Hahn MW (2009) The life and death of gene families. Bioessays 31:29–39. doi:10.1002/bies.080085

DePristo MA, Weinreich DM, Hartl DL (2005) Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet 6:678–687. doi:10.1038/nrg1672

Dornez E, Croes E, Gebruers K et al (2010) Accumulated evidence substantiates a role for three classes of wheat xylanase inhibitors in plant defense. Crit Rev Plant Sci 29:244–264. doi:10.1080/07352689.2010.487780

Federici L, Caprari C, Mattei B et al (2001) Structural requirements of endopolygalacturonase for the interaction with PGIP (polygalacturonase-inhibiting protein). Proc Natl Acad Sci 98:13425–13430. doi:10.1073/pnas.231473698

Federici L, Di Matteo A, Fernandez-Recio J et al (2006) Polygalacturonase inhibiting proteins: players in plant innate immunity? Trends Plant Sci 11:65–70. doi:10.1016/j.tplants.2005.12.005

Fierens E, Rombouts S, Gebruers K et al (2007) TLXI, a novel type of xylanase inhibitor from wheat (*Triticum aestivum*) belonging to the thaumatin family. Biochem J 403:583–591. doi:10.1042/BJ20061291

Gan P, Ikeda K, Irieda H et al (2013) Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol 197:1236–1249. doi:10.1111/nph.12085

Gan P, Kumakura N, Narusaka M et al (2016) Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. Genome Biol Evol 8:1467–1481. doi:10.1093/gbe/evw089

Gebruers K, Debyser W, Goesaert H et al (2001) *Triticum aestivum* L. endoxylanase inhibitor (TAXI) consists of two inhibitors, TAXI I and TAXI II, with different specificities. Biochem J 353:239–244

Helder P, Uma M, Martin U et al (2016) PhytoPath: an integrative resource for plant pathogen genomics. Nucleic Acids Res 44:D688–D693. doi:10.1093/nar/gkv1052

Herron SR, Benen J, Scavetta RD et al (2000) Structure and function of pectic enzymes: virulence factors of plant pathogens. Proc Natl Acad Sci 97:8762–8769

Herron SR, Scavetta RD, Garrett M et al (2003) Characterization and implications of Ca^{2+} binding to pectate lyase C. J Biol Chem 278:12271–12277. doi:10.1074/jbc.M209306200

Hiruma K, Gerlach N, Sacristán S et al (2016) Root endophyte *Colletotrichum tofieldiae* confers plant fitness benefits that are phosphate status dependent. Cell 165:464–474. doi:10.1016/j.cell.2016.02.028

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33-8-27–33-8-28

Isshiki A, Akimitsu K, Yamamoto M, Yamamoto H (2001) Endopolygalacturonase is essential for citrus black rot caused by *Alternaria citri* but not brown spot caused by *Alternaria alternata*. Mol Plant Microbe Interact 14:749–757. doi:10.1094/MPMI.2001.14.6.749

Kim H, Ahn J-H, Görlach JM et al (2001) Mutational analysis of β-glucanase genes from the plant-pathogenic fungus *Cochliobolus carbonum*. Mol Plant Microbe Interact 14:1436–1443. doi:10.1094/MPMI.2001.14.12.1436

Lamb B (2003) Meiotic recombination in fungi: mechanisms and controls of crossing-over and gene conversion. In: Arora DK, Khachatourians GG (eds) Fungal genomics. Elsevier, Amsterdam, pp 15–41

Leckie F, Mattei B, Capodicasa C et al (1999) The specificity of polygalacturonase-inhibiting protein (PGIP): a single amino acid substitution in the solvent-exposed β-strand/β-turn region of the leucine-rich repeats (LRRs) confers a new recognition capability. EMBO J 18:2352–2363. doi:10.1093/emboj/18.9.2352

Lenné JM (2001) Some major plant diseases. In: Plant pathologist's pocketbook, 3rd edn. CABI, Wallingford, pp 4–18

Lo Presti L, Lanver D, Schweizer G et al (2015) Fungal effectors and plant susceptibility. Annu Rev Plant Biol 66:513–545. doi:10.1146/annurev-arplant-043014-114623

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

MacColl ADC (2011) The ecological causes of evolution. Trends Ecol Evol 26:514–522. doi:10.1016/j.tree.2011.06.009

Marcelino J, Giordano R, Gouli S et al (2008) *Colletotrichum acutatum* var. *fioriniae* (teleomorph: *Glomerella acutata var. fioriniae var. nov.*) infection of a scale insect. Mycologia 100:353–374

Martin U, Alayne C, Rutherford K et al (2017) PHI-base: a new interface and further additions for the multi-species pathogen–host interactions database. Nucl Acids Res 45:D604–D610. doi:10.1093/nar/gkw1089

Mayans O, Scott M, Connerton I et al (1997) Two crystal structures of pectin lyase a from *Aspergillus* reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases. Structure 5:677–689

Murrell B, Wertheim JO, Moola S et al (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8:e1002764. doi:10.1371/journal.pgen.1002764

Murrell B, Weaver S, Smith MD et al (2015) Gene-wide identification of episodic selection. Mol Biol Evol 32:1365–1371. doi:10.1093/molbev/msv035

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York

Nuss L, Mahé A, Clark AJ et al (1996) Differential accumulation of PGIP (polygalacturonase-inhibiting protein) mRNA in two near-isogenic lines of *Phaseolus vulgaris* L. upon infection with *Colletotrichum lindemuthianum*. Physiol Mol Plant Pathol 48:83–89. doi:10.1006/pmpp.1996.0008

O'Connell RJ, Thon MR, Hacquard S et al (2012) Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat Genet 44:1060–1065. doi:10.1038/ng.2372

Perfect SE, Hughes HB, O'Connell RJ, Green JR (1999) *Colletotrichum*: a model genus for studies on pathology and fungal-plant interactions. Fungal Genet Biol 27:186–198. doi:10.1006/fgbi.1999.1143

Pollet A, Sansen S, Raedschelders G et al (2009) Identification of structural determinants for inhibition strength and specificity of wheat xylanase inhibitors TAXI-IA and TAXI-IIA. FEBS J 276:3916–3927. doi:10.1111/j.1742-4658.2009.07105.x

Pond SK, Pond SK, Muse SV (2005a) Site-to-site variation of synonymous substitution rates. Mol Biol Evol 22:2375–2385. doi:10.1093/molbev/msi232

Pond SLK, Pond SLK, Frost SDW, Frost SDW (2005b) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208–1222. doi:10.1093/molbev/msi105

Pond SLK, Pond SLK, Frost SDW, Frost SDW (2005c) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531–2533. doi:10.1093/bioinformatics/bti320

Pond SLK, Frost SDW, Grossman Z et al (2006a) Adaptation to different human populations by HIV-1 revealed by codon-based analyses. PLoS Comp Biol 2:e62. doi:10.1371/journal.pcbi.0020062

Pond SLK, Posada D, Gravenor MB et al (2006b) GARD: a genetic algorithm for recombination detection. Bioinformatics 22:3096–3098. doi:10.1093/bioinformatics/btl474

Pond SLK, Posada D, Gravenor MB et al (2006c) Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891–1901. doi:10.1093/molbev/msl051

Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256. doi:10.1093/molbev/msn083

Powell AJ, Conant GC, Brown DE et al (2008) Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens. BMC Genom 9:147. doi:10.1186/1471-2164-9-147

Prusky D (1996) Pathogen quiescence in postharvest diseases. Annu Rev Phytopathol 34:413–434. doi:10.1146/annurev.phyto.34.1.413

Raiola A, Lionetti V, Elmaghraby I et al (2011) Pectin methylesterase is induced in *Arabidopsis* upon infection and is necessary for a successful colonization by necrotrophic pathogens. Mol Plant Microbe Interact 24:432–440. doi:10.1094/MPMI-07-10-0157

Robinson O, Dylus D, Dessimoz C (2016) Phylo.io: interactive viewing and comparison of large phylogenetic trees on the web. Mol Biol Evol 33:2163–2166. doi:10.1093/molbev/msw080

Romero PA, Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol 10:866–876. doi:10.1038/nrm2805

Ronquist F, Teslenko M, van der Mark P et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. doi:10.1093/sysbio/sys029

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738. doi:10.1038/nprot.2010.5

Sánchez-Torres P, Visser J, Benen JAE (2003) Identification of amino acid residues critical for catalysis and stability in *Aspergillus niger* family 1 pectin lyase A. Biochem J 370:331–337. doi:10.1042/BJ20021071

Shriner D, Nickle DC, Jensen MA, Mullins JI (2003) Potential impact of recombination on sitewise approaches for detecting positive natural selection. Genet Res 81:115–121

Silvestro D, Michalak I (2011) raxmlGUI: a graphical front-end for RAxML. Org Diver Evol 12:335–337. doi:10.1007/s13127-011-0056-0

Soanes DM, Alam I, Cornell M et al (2008) Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS ONE 3:e2300. doi:10.1371/journal.pone.0002300

Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. doi:10.1186/gb-2006-7-s1-s10

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. doi:10.1093/bioinformatics/btu033

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucl Acids Res 33:W465–W467. doi:10.1093/nar/gki458

Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. doi:10.1093/molbev/mst197

Tokuriki N, Stricher F, Serrano L, Tawfik DS (2008) How protein stability and new functions trade off. PLoS Comp Biol 4:e1000002. doi:10.1371/journal.pcbi.1000002

Vitali J, Schick B, Kester HC et al (1998) The tree-dimensional structure of *Aspergillus niger* pectin lyase B at 1.7-A resolution. Plant Physiol 116:69–80

Walton J (1994) Deconstructing the cell wall. Plant Physiol 104:1113–1118

Wang Y-T, Lo H-S, Wang P-H (2008) Endophytic fungi from taxus *mairei* in Taiwan: first report of *Colletotrichum gloeosporioides* as an endophyte of Taxus *mairei*. Bot Stud 39:39–43

Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. Nature 449:54–61. doi:10.1038/nature06107

Xu B, Yang Z (2013) PAMLX: a graphical user interface for PAML. Mol Biol Evol 30:2723–2724. doi:10.1093/molbev/mst179

Xu D, Zhang Y (2011) Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys J 101:2525–2534. doi:10.1016/j.bpj.2011.10.024

Yang Z, Nielsen R, Goldman N, Pedersen A-MK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449. doi:10.1016/S0378-1119(99)00294-2

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118. doi:10.1093/molbev/msi097

Zhao Z, Liu H, Wang C, Xu J-R (2013) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genom 14:274. doi:10.1186/1471-2164-14-274