Title

Adaptive evolution of Moniliophthora PR-1 proteins towards its pathogenic lifestyle.

Journal

BMC Ecology and Evolution, 21(1)

Authors

Vasconcelos, Adrielle
José, Juliana
Tokimatu, Paulo
et al.

Publication Date

2021-05-14

DOI

10.1186/s12862-021-01818-5

Peer reviewed

BMC Ecology and Evolution

RESEARCH

Open Access

Adaptive evolution of *Moniliophthora* PR-1 proteins towards its pathogenic lifestyle

Adrielle A. Vasconcelos[1], Juliana José[1], Paulo M. Tokimatu[1], Antonio P. Camargo[1], Paulo J. P. L. Teixeira[2], Daniela P. T. Thomazella[1], Paula F. V. do Prado[1], Gabriel L. Fiorin[1], Juliana L. Costa[3], Antonio Figueira[3], Marcelo F. Carazzolle[1], Gonçalo A. G. Pereira[1*] and Renata M. Baroni[1]

Abstract

Background: Plant pathogenesis related-1 (PR-1) proteins belong to the CAP superfamily and have been characterized as markers of induced defense against pathogens. *Moniliophthora perniciosa* and *Moniliophthora roreri* are hemibiotrophic fungi that respectively cause the witches' broom disease and frosty pod rot in *Theobroma cacao*. Interestingly, a large number of plant *PR-1*-like genes are present in the genomes of both species and many are up-regulated during the biotrophic interaction. In this study, we investigated the evolution of PR-1 proteins from 22 genomes of *Moniliophthora* isolates and 16 other Agaricales species, performing genomic investigation, phylogenetic reconstruction, positive selection search and gene expression analysis.

Results: Phylogenetic analysis revealed conserved PR-1 genes (PR-1a, b, d, j), shared by many Agaricales saprotrophic species, that have diversified in new PR-1 genes putatively related to pathogenicity in *Moniliophthora* (PR-1f, g, h, i), as well as in recent specialization cases within *M. perniciosa* biotypes (PR-1c, k, l) and *M. roreri* (PR-1n). PR-1 families in *Moniliophthora* with higher evolutionary rates exhibit induced expression in the biotrophic interaction and positive selection clues, supporting the hypothesis that these proteins accumulated adaptive changes in response to host–pathogen arms race. Furthermore, although previous work showed that MpPR-1 can detoxify plant antifungal compounds in yeast, we found that in the presence of eugenol *M. perniciosa* differentially expresses only *MpPR-1e, k, d,* of which two are not linked to pathogenicity, suggesting that detoxification might not be the main function of most *MpPR-1*.

Conclusions: Based on analyses of genomic and expression data, we provided evidence that the evolution of PR-1 in *Moniliophthora* was adaptive and potentially related to the emergence of the parasitic lifestyle in this genus. Additionally, we also discuss how fungal PR-1 proteins could have adapted from basal conserved functions to possible roles in fungal pathogenesis.

Keywords: Witches' broom disease, Phytopathogen, Gene evolution, Fungi, Phylogenetics, Positive selection, Adaptation

Introduction

Pathogenesis related-1 (PR-1) proteins are part of the CAP (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1) superfamily, also known as SCP/TAPS proteins (sperm-coating protein/Tpx-1/Ag5/PR-1/Sc7), and are present throughout the eukaryotic kingdom [1, 2]. In plants, PR-1 proteins are regarded as markers of induced defense responses against pathogens [3]. These

*Correspondence: goncalo@unicamp.br
[1] Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil
Full list of author information is available at the end of the article

(See figure on next page.)

Fig. 1 Characterization of PR-1 gene families in *M. perniciosa and M. roreri* genomes. **a** Heatmap of the number of gene copies per family of *PR-1-like* candidates per *Moniliophthora* isolate. Identification of genomes are in columns and PR-1 family names are in rows. **b** Amplification by PCR of *MpPR-1c* and *MpPR-1d* genes in the genomic DNA of eight *M. perniciosa* isolates. 1 Kb Ld = 1 Kb Plus DNA Ladder (Invitrogen), Neg = PCR negative control (no DNA). Expected fragment sizes were 687 bp for *MpPR-1c* and 902 bp for *MpPR-1d*. **c** Synteny analysis of a 10 Kb portion of the genome where the *MpPR-1j, c, d* genes are found in the three biotypes of *M. perniciosa* and *M. roreri*. The genomes analyzed were C-BA3, S-MG2, R-CO2, L-EC1 and L-EC2. Only identity above 75% to the C-biotype reference is shown

proteins have also been ascribed roles in different biological processes in mammals, insects, nematodes and fungi, including reproduction, cellular defense, virulence and evasion of the host immune system [4–12]. In *Saccharomyces cerevisiae*, Pry proteins (*Pathogen related in yeast*) bind and export sterols and fatty acids to the extracellular medium, an activity that has also been demonstrated for other proteins of the CAP superfamily through functional complementation assays [13–17].

The basidiomycete fungi *Moniliophthora perniciosa* and *Moniliophthora roreri* are hemibiotrophic phytopathogens that cause, respectively, the Witches' Broom disease (WBD) and Frosty Pod Rot of cacao (*Theobroma cacao*). Currently, three biotypes are recognized for *M. perniciosa* based on the hosts that each one is able to infect. The C-biotype infects species of *Theobroma* and *Herrania* (Malvaceae); the S-biotype infects plants of the genus *Solanum* (e.g., tomato) and *Capsicum* (pepper); and the L-biotype is associated with species of lianas (Bignoniaceae), without promoting visible disease symptoms [18–20].

With the genome and transcriptome sequencing of the C-biotype, 11 *PR-1*-like genes, named *MpPR-1a* to *k*, were identified in *M. perniciosa* [21]. Interestingly, many of these genes are upregulated during the biotrophic interaction of *M. perniciosa* and *T. cacao*, which constitutes a strong indication of the importance of these proteins in the disease process [21, 22]. In this context, efforts have been made to elucidate the role of these molecules during the interaction of *M. perniciosa* with cacao, such as the determination of the tridimensional structure of MpPR-1i [23] and the functional complementation of *MpPR-1* genes in yeast *Pry* mutants [24]. These studies revealed that seven MpPR-1 proteins display sterol or fatty acid binding and export activity, suggesting that they could function as detoxifying agents against plant lipidic toxins [24].

Despite these advances, studies with a deeper evolutionary perspective have not yet been performed for MpPR-1 proteins. Evolutionary analysis can be an important tool for the inference of gene function and the identification of mechanisms of evolution of specific traits, such as pathogenicity. Genes that are evolving under negative selection pressures are likely to play a crucial role in basal metabolism [25]. On the other hand, genes that are evolving under positive selection may have changed to adjust their function to a relatively new environmental pressure [26]. Thus, it can be hypothesized that *Moniliophthora* PR-1 might have accumulated adaptive substitutions in response to selective pressures related to a pathogenic lifestyle, and the analysis of these substitutions may reveal protein targets and specific codons that are potentially important for the pathogenicity in *Moniliophthora*.

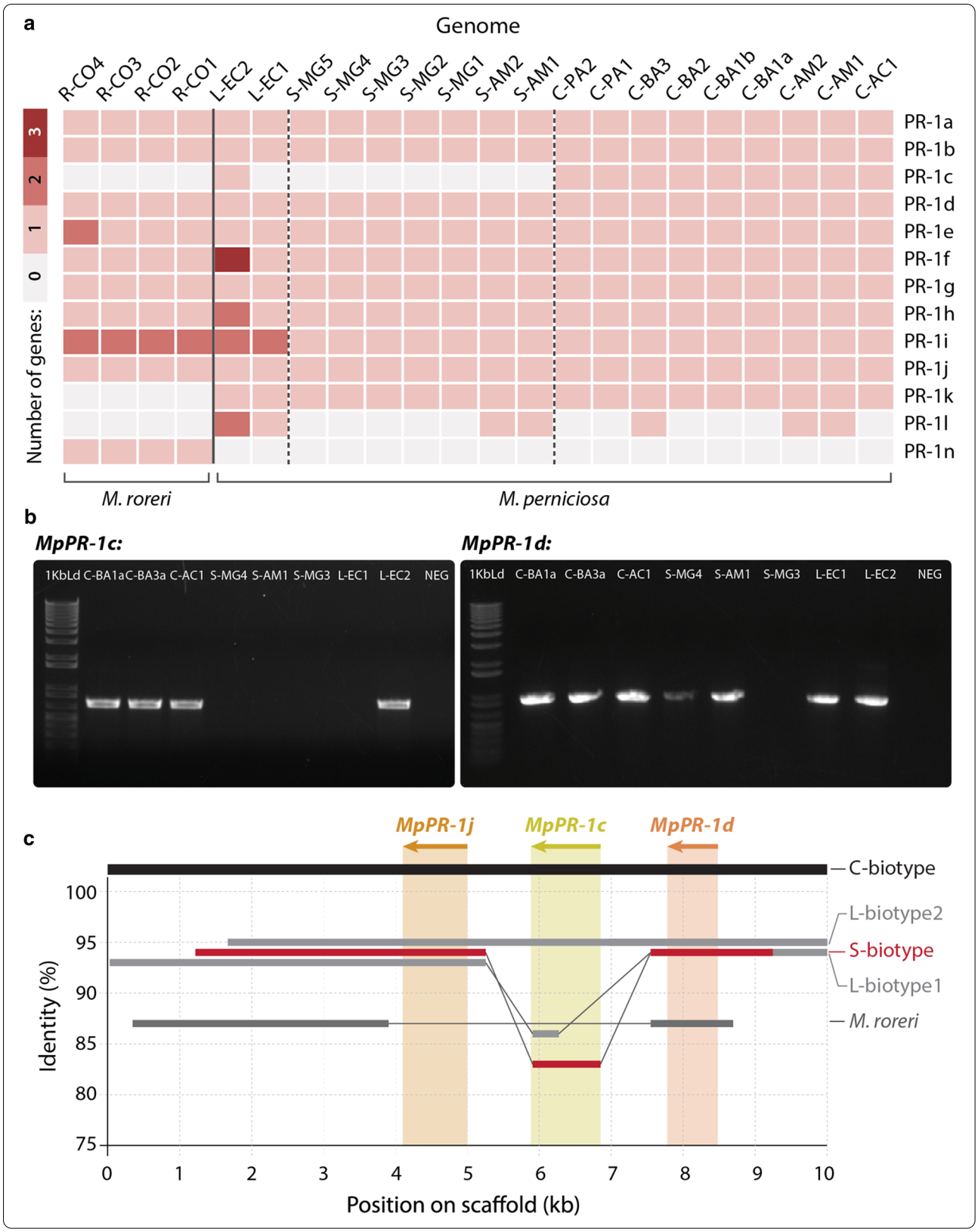
In this study, we performed a two-level evolutionary analysis of *Moniliophthora PR-1* genes: (i) their macroevolution in the order Agaricales, which consists mainly of saprotrophic fungi, being the *Moniliophthora* species one of the few exceptions; (ii) and their microevolution within *M. roreri* and *M. perniciosa* and its biotypes that differ in host-specificity. By characterizing PR-1 proteins encoded by 22 *Moniliophthora* genomes, reconstructing their phylogenetic history, searching for evidence of positive selection and analyzing expression data, we identified an increased diversification in these proteins in *Moniliophthora* that is potentially adaptive and related to its pathogenic lifestyle.

Results

Characterization of PR-1 gene families in *Moniliophthora*

Previous work had already reported the identification of 11 *PR-1*-like genes in the genome of *M. perniciosa* isolate CP02 (C-biotype), which were named *MpPR-1a* to *MpPR-1k* [21]. Likewise, 12 *PR-1*-like genes were identified in the genome of *M. roreri* (MCA2977) [27]. With the sequencing and assembly of 18 additional genomes of *M. perniciosa* isolates and other four genomes of *M. roreri* isolates, it was possible to characterize the *PR-1* gene families in the different biotypes of *M. perniciosa* and in its sister species *M. roreri* in order to identify differences at the species and biotype levels. Figure 1a shows the number of genes identified as *PR-1* per isolate.

The examination of orthogroups containing *PR-1-like* hits revealed that the PR-1i orthogroup has the highest number of duplications with two copies in *M. roreri* and in L-biotype. Moreover, a new PR-1-like orthogroup with seven candidates that are more similar to *MpPR-1i* (67% identity) was found. This newly identified gene was named "*MpPR-1l*" and was not found in *M. roreri*. It has the same number and structure of introns and


a
Genome
R-CO4
R-CO3
R-CO2
R-CO1
L-EC2
L-EC1
S-MG5
S-MG4
S-MG3
S-MG2
S-MG1
S-AM2
S-AM1
C-PA2
C-PA1
C-BA3
C-BA2
C-BA1b
C-BA1a
C-AM2
C-AM1
C-AC1
Number of genes:
3
2
1
0
PR-1a
PR-1b
PR-1c
PR-1d
PR-1e
PR-1f
PR-1g
PR-1h
PR-1i
PR-1j
PR-1k
PR-1l
PR-1n
M. roreri
M. perniciosa
b
MpPR-1c:
1KbLd C-BA1a C-BA3a C-AC1 S-MG4 S-AM1 S-MG3 L-EC1 L-EC2 NEG
MpPR-1d:
1KbLd C-BA1a C-BA3a C-AC1 S-MG4 S-AM1 S-MG3 L-EC1 L-EC2 NEG
c
MpPR-1j
MpPR-1c
MpPR-1d
C-biotype
L-biotype2
S-biotype
L-biotype1
M. roreri
Identity (%)
100
95
90
85
80
75
0
1
2
3
4
5
6
7
8
9
10
Position on scaffold (kb)

exons as the *MpPR-1i* gene and they are closely located in the same scaffold, which is evidence of a duplication event within *M. perniciosa*. Interestingly, the sequences corresponding to *MpPR-1l* in the five S-biotype isolates from Minas Gerais were found in another orthogroup, in which *MpPR-1l* was fused to the adjacent gene in the genome (a putative endo-polygalacturonase gene containing the IPR011050 domain: Pectin lyase fold) with no start codon found between the two domains. Furthermore, we found that the *MpPR-1i* gene and, consequently, its predicted protein is truncated in almost all S-biotype isolates from MG (except for S-MG2) (Fig. 4).

Examining these gene families to look for other putatively species-specific *PR-1* in *Moniliophthora*, we observed that the *MpPR-1k* and *MpPR-1c* genes are not found in the *M. roreri* genomes analyzed in this work, while *MrPR-1n* constitutes an exclusive family in this species. The *MrPR-1o* gene previously identified by Meinhardt et al. [27] was not predicted in any genome as a gene in this work. The protein sequence of *MrPR-1o* has higher identity with *MrPR-1j* (70%), *MpPR-1j* (66%) and *MpPR-1c* (56%), but it is shorter than all 3 protein sequences and does not have a signal peptide like other PR-1 proteins, which suggested that *MrPR-1o* is a pseudogenized paralog of *MrPR-1j*.

The absence of *MpPR-1c* in all S-biotype genomes suggested that this gene could be biotype-specific within *M. perniciosa*, however, it was predicted in the L-biotype genome L-EC2. Therefore, we sought to confirm the presence or absence of this gene in different *M. perniciosa* by PCR amplification and synteny analysis. The absence of *MpPR-1c* in the S-biotype isolates and in the L-biotype L-EC1 was confirmed, as well as its presence in L-EC2 (Fig. 1b). We also amplified the *MpPR-1d* gene, which was predicted in all genomes, in almost all tested isolates, except for S-MG3 because of a mismatch in the annealing regions of both primers. Even though our PCR results indicated that *MpPR-1c* is not present in the S-biotype, synteny analysis of the genome region where Mp*PR-1j-c-d* are found in tandem [22] revealed that, in fact, *MpPR-1c* is partially present in these S-biotype genomes (Fig. 1c), suggesting again that the duplication event of PR-1j occured in the ancestral of *Moniliophthora* but this paralog was also pseudogenized in the evolution of the S-biotype.

PR-1 genes evolution along the Agaricales order

To study the macroevolution of PR-1 proteins, we identified orthologous sequences of genes encoding PR-1-like proteins in 16 genomes of species from the Agaricales order, including three selected *M. perniciosa* genomes (one of each biotype) and 1 *M. roreri* genome for comparisons. The phylogenetic reconstruction of Agaricales PR-1 proteins revealed a basal separation of two major clades, hereafter called clade1 and clade2 (Fig. 2a).

The first PR-1 clade includes most Agaricales species outside *Moniliophthora* showing *PR-1* genes that diverged early in the phylogeny, before the appearance of PR-1a-l-n orthologues. From this first clade including the early diverged PR-1 proteins, subsequently diverged PR-1d and PR1-j. The separation between PR-1d and j proposed for *Moniliophthora* only occurs in *Volvariella volvacea*, while for all other species, paralogous of PR-1d diverged early. In *Moniliophthora*, PR-1j and PR-1d have more recent common ancestors, and the only paralogous of these *Moniliophthora* PR-1s originates from a possible duplication of *MpPR-1j* in the C-biotype of *M. perniciosa*, which was previously named *MpPR-1c*, therefore exclusive to this species and biotype.

The second clade includes all other PR-1 families and other Agaricales PR-1s that do not have a common ancestor with a single PR-1 from *Moniliophthora*. Clade2 is divided into two subclasses in its base, one of them composed of the PR-1b clade, which is distributed among 14 species. In the second subgroup, PR-1a shows a common ancestor in a total of 16 species, being the most common PR-1 here. The great diversification of a PR-1a-like ancestor in *Moniliophthora* resulted in the formation of at least 4 new and exclusive *PR-1* genes (*k, g, i, l*) with high evolutionary rates reflected on the branch lengths (Fig. 2b). PR-1n showed a putative ortholog in *Armillaria mellea*, the only other plant pathogen in the Agaricales dataset, however, this connection has low branch support.

Recent diversification of PR-1 genes in *Moniliophthora*

Through the investigation of the phylogenetic history of PR-1 proteins within 22 *Moniliophthora* isolates (Fig. 3), we found that the previous classification of *MpPR-1a* to *k* and *MrPR-1n* represents monophyletic clades in the tree, except for *MpPR-1c* which is a recent paralogous of *MpPR-1j*. The evolution of *Moniliophthora* PR-1 also reflected the basal divergence between two large clades as observed in the Agaricales PR-1 tree (Fig. 2). The only incongruence between the two phylogenetic trees is the relative position of PR-1k, which appeared after the divergence of PR-1a in Agaricales and before it in *Moniliophthora*. This incongruence may be due to the extreme differentiation of PR-1k, with longer branch lengths in *Moniliophthora*, being its position on the Agaricales tree more reliable.

Among the PR-1 families, the proteins with the greatest number of changes in the tree are PR-1g, i, and k, which are exclusive PR-1s in the genus *Moniliophthora*, as pointed out by the previous phylogenetic analysis. In addition, PR-1h also showed a greater branch length

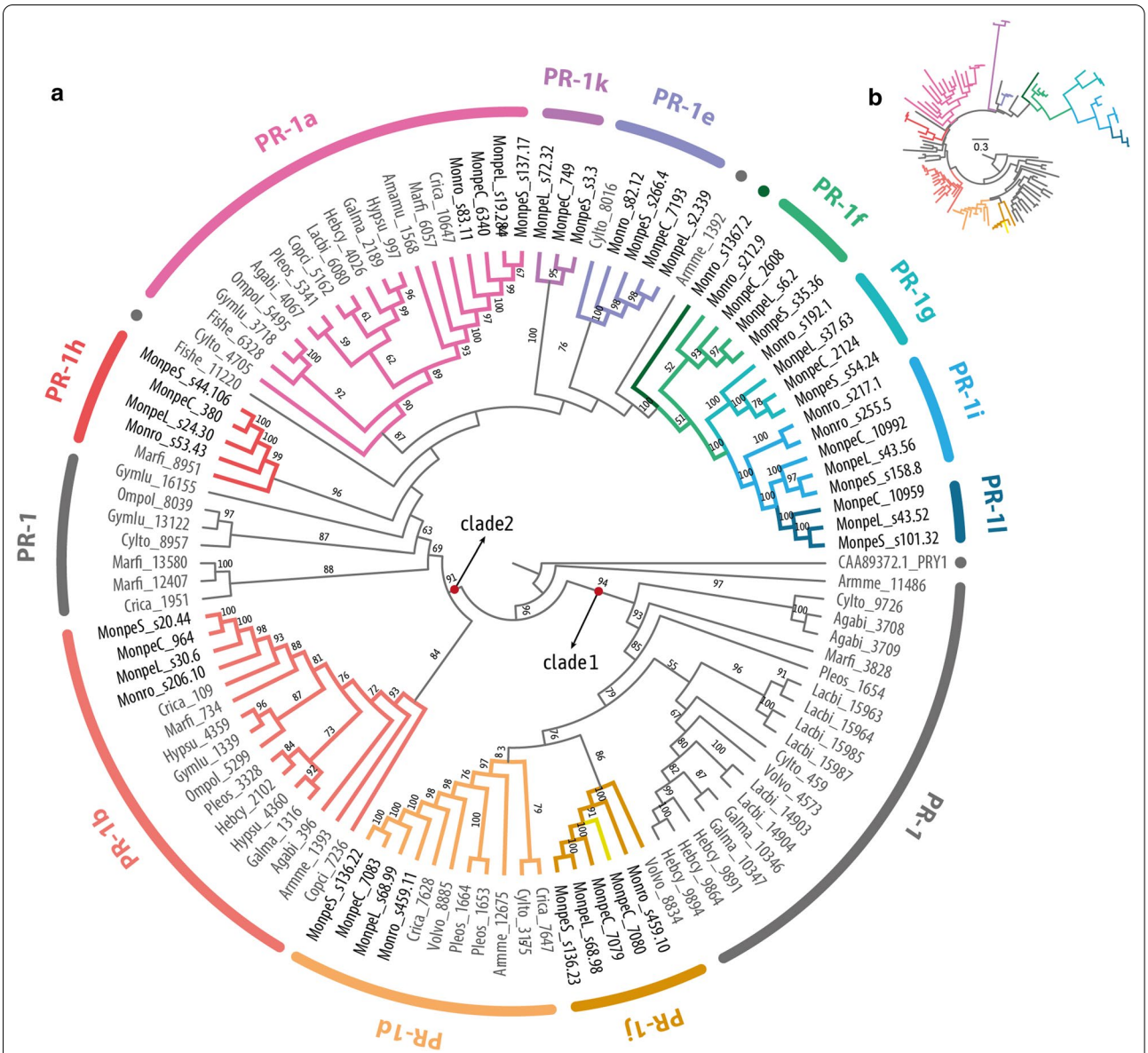


Fig. 2 Phylogenetic cladogram of PR-1 proteins in Agaricales (Basidiomycota). **a** Phylogenetic relationships were inferred by maximum likelihood and branch support was obtained using 1000 bootstraps. Only branch support values greater than 70 are shown. PR-1c is indicated as a yellow branch inside the PR-1j clade, and PR-1n is indicated with a dark green dot and branch. Proteins with ancestral divergence to more than one family were named with the letters of the derived families. Full species names are in Additional file 1. The branch lengths were dimensioned for easy visualization. **b** The same phylogeny shown in "**a**" is presented with the respective branch lengths augmented for the *Moniliophthora* specific PR-1 genes

than the others, being a family of PR-1s only shared between *Moniliophthora* and *Marasmius* in the Agaricales PR-1 tree (Fig. 2). PR-1c also presented a large number of changes in relation to its ancestor PR-1j. This greater number of changes in these MpPR-1s, and their exclusive presence in comparison to the other Agaricales, indicate a recent potential adaptive process of diversification of these proteins in *Moniliophthora*.

Positive selection shaping PR-1 families in *Moniliophthora*

Based on the observations of high diversification of PR-1 families within *Moniliophthora*, we hypothesized that positive selection could be shaping these proteins either in the C-biotype or in the S-biotype. To test this

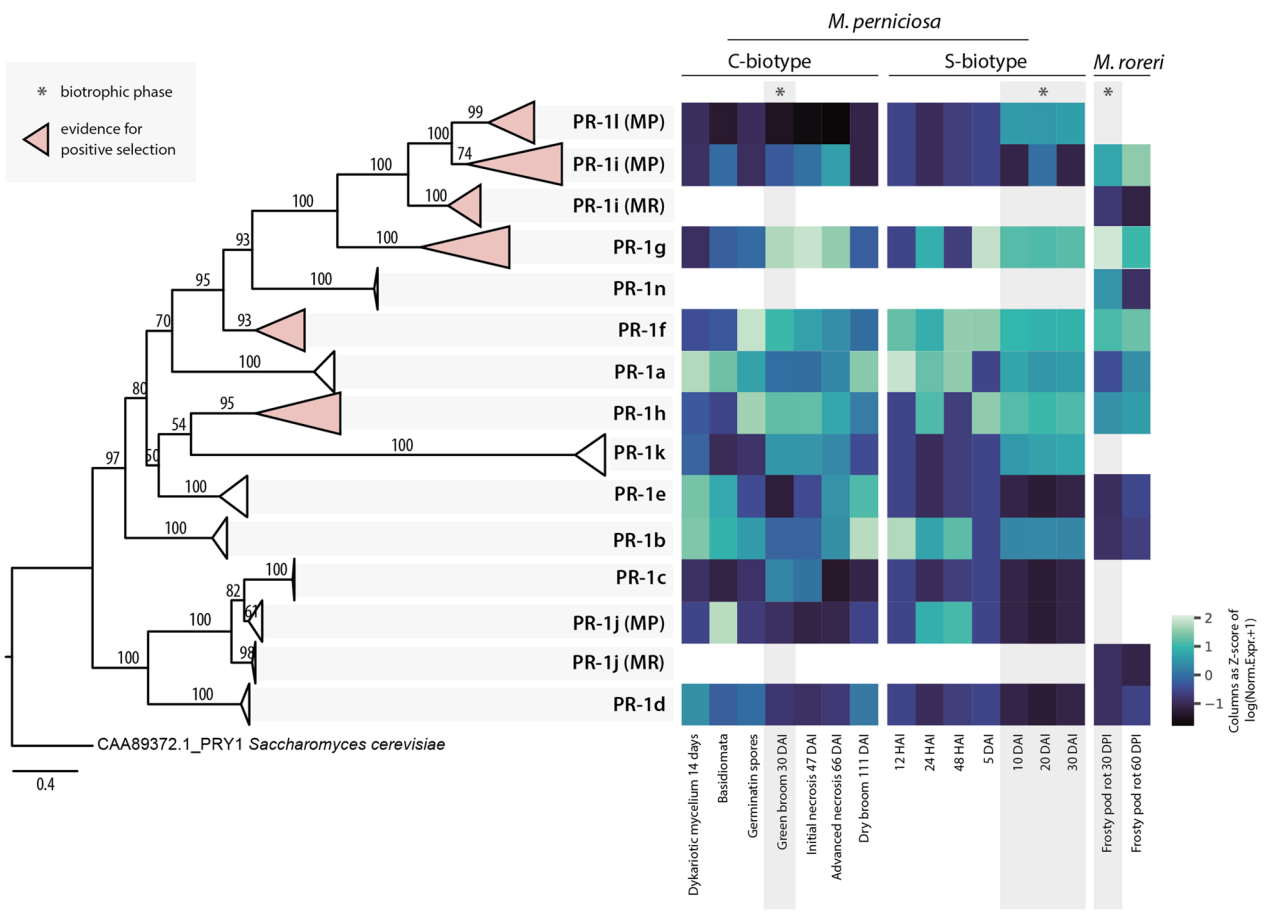


Fig. 3 Phylogenetic reconstruction of PR-1 proteins in *Moniliophthora* and heatmap of expression Z-scores of PR-1. Phylogenetic relationships of PR-1 proteins from 18 *M. perniciosa* and 4 *M. roreri* isolates were inferred by maximum likelihood and branch support was obtained using 1000 bootstraps. The PRY1 protein of *Saccharomyces cerevisiae* was used as an outgroup. Clades filled with pink color indicate PR-1 families with evidence of positive selection detected with the site model test of CodeML. A version of this tree with non-collapsed branches can be found in Additional file 4. For each PR-1 family, the Z-score of log transformed expression levels of *MpPR-1* and *MrPR-1* from transcriptomic data was calculated for conditions (columns) and plotted as a heatmap. The heatmap includes *MpPR-1* data from seven conditions of the C-biotype of *M. perniciosa* from the Witches' Broom Transcriptomic Atlas, seven different time points of S-biotype infection in MicroTom tomato plants, and two conditions of *M. roreri* infection in cacao pods (frosty pod rot). Conditions highlighted with a grey background and asterisk indicate the biotrophic stage of the plant-pathogen interaction

hypothesis, we tested the branch-sites evolutionary model for each PR-1 family. None of these tests brought evidence of positive selection in any PR-1 family for the C-biotype branches. For the S-biotype branches, a signal of positive selection was detected for PR-1g on one site of the protein sequence.

Considering that the existence of *M. perniciosa* biotypes are very recent in the evolutionary timescale and that C-biotype itself has almost no genetic variation among its sequences, which makes it very difficult to apply separate dN/dS tests, we tested both C- and S-biotypes together. We tested the hypothesis that there was a single selective pressure shaping PR-1 families throughout the *M. perniciosa* and *M. roreri* evolution regardless of the biotype, using the site model test. In these tests, sites with positive selection signs were detected in five families (PR-1f, g, h, i, l) (Table 1). The PR-1n family was not included in these tests because all sequences were identical.

PR-1g stands out for having the highest omega and for being one of the most expressed genes during the green broom stage [22]. Three of the codons under positive selection are part of the 'keke' domain, which is possibly involved in the interaction with divalent ions or proteins [21]. Among the sites detected under positive selection for PR-1i, one is found in the caveolin binding

Table 1 Omega (dN/dS) values and protein sites (amino acid: alignment position) detected with significant probability of positive selection for each PR-1 family in *Moniliophthora*

Family	Omega	Sites under positive selection (p > 0.95)
PR-1a	4.12	None
PR-1b	2.31	None
PR-1c	1	None
PR-1d	2.07	None
PR-1e	1	None
PR-1f	2.74	K: 49, S: 157
PR-1g	7.05	P: 211, P: 234, A: 242, S: 260, S: 271
PR-1h	4.33	S: 78, Y: 107, P: 141, S: 155, E: 187, D: 206, L: 209, M: 224, R: 259, Q: 267
PR-1i	4.38	T: 40, Q: 54, D: 56, R: 118, K: 132, A: 141, L: 150
PR-1l	6.39	Q: 65, Q: 76, –: 164, N: 176, R: 192, K: 193, E: 194, F: 196
PR-1j	2.76	None
PR-1k	2.50	None

selection, both revealed processes of diversification in the PR-1 phylogenies. It is possible that these families have also undergone selective pressures in their evolution, but the short time of evolution of *M. perniciosa* in relation to the genus has reduced the accuracy of the dN/dS tests in these exclusive families.

Adaptive evolution of PR-1 is reflected on expression data

It has already been shown that *MpPR-1* genes of the C-biotype have distinct expression profiles in several different conditions of the WBD Transcriptome Atlas, which were also confirmed by quantitative RT-PCR [21, 22]. *MpPR-1a, b, d, e* are ubiquitously expressed during the necrotrophic mycelial stage, while *MpPR-1j* is mainly expressed in primordia and basidiomata. Six *MpPR-1s* are highly and almost exclusively expressed during the biotrophic stage of WBD: *MpPR-1c, f, g, h, i, k* [21]. However, in contrast to *MpPR-1i*, the newly discovered *MpPR-1l* is not expressed in any of the conditions ana-

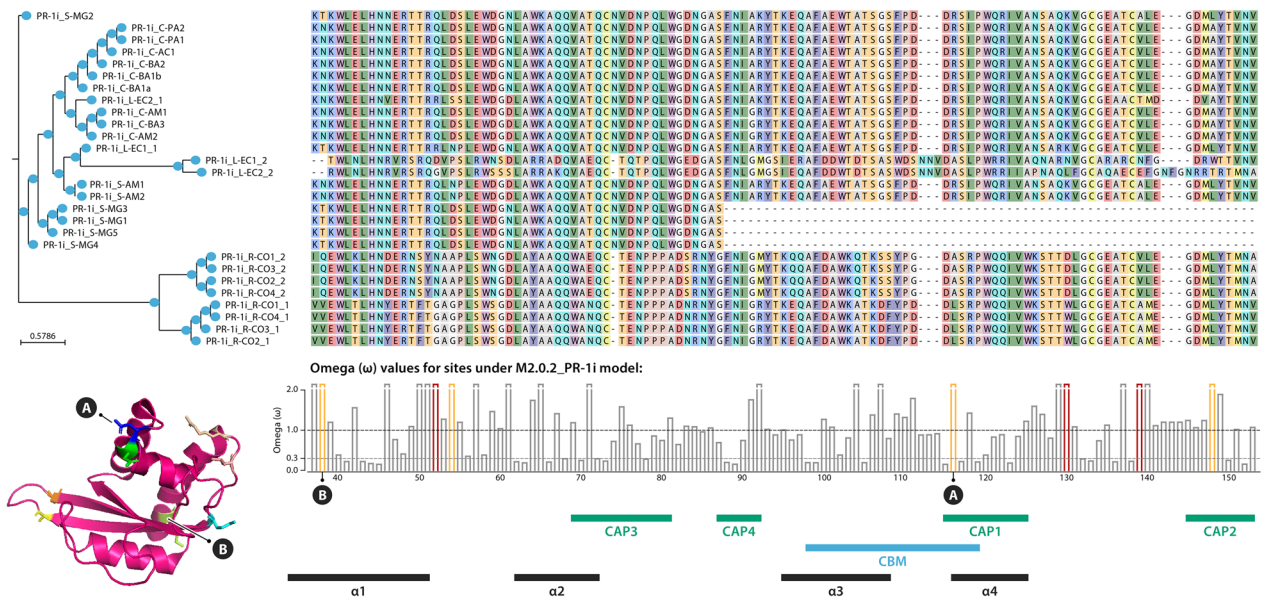


Fig. 4 Sequence alignment and phylogeny of PR-1i proteins in *Moniliophthora* isolates. Only a slice of the middle portion of the alignment is shown to highlight the sites with positive selection signs, indicated by red (p-value ≤ 0.01) or orange (p-value ≤ 0.05) bars in the bar chart of omega values below the alignment. Below the bar chart, annotations indicate the locations along the sequence of the CAP domains, caveolin-binding motif (CBM) and alpha-helices (α). On the 3D crystal structure of MpPR-1i protein (PBD:5V50) and on the bar chart, "A" indicates the site under positive selection detected in the CBM and "B" indicates the site under positive selection in alpha-helix 1. For figure of the whole sequence alignment and omega value bar chart of PR-1i and the other PR-1 families, see Additional file 5

motif (CBM), an important region for binding to sterols, and another site is in the α-helix 1, which together with α-helix 4 form the cavity for ligation to palmitate [23] (Fig. 4).

Although the two exclusive PR-1 families of *M. perniciosa*, PR-1c and PR-1k, do not present evidence of positive

lyzed, suggesting that this gene may not be functional in the C-biotype.

Expression data of *MpPR-1* genes from the S-biotype during a time course of the biotrophic interaction with MT tomato revealed that *MpPR-1f, g, h, l,* and *k* are highly expressed during 10–30 days after infection (d.a.i.) (Fig. 3).

These results are similar to the expression profile verified for the C-biotype during the biotrophic interaction with *T. cacao*, with the exceptions that *MpPR-1c* is absent in the S-biotype and, instead of *MpPR-1i*, *MpPR-1l* is expressed during tomato infection. S-biotype *MpPR-1*s are highly expressed starting at 10 d.a.i., which is usually when the first symptoms of stem swelling are visible in MT tomato [28]. *MpPR-1a* and *b* appear to have ubiquitous expression profiles since they show similar expression levels in almost all conditions. *MpPR-1j, d, e*, and *i* did not show significant expression in these libraries. Because *MpPR-1i* is truncated in the S-MG1 genome, quantification of expression was also done with S-MG2 as a reference, since it has a complete *MpPR-1i* gene. However, we still obtained the same expression profiles as S-MG1 for all *MpPR-1*. This could suggest that *MpPR-1l* is expressed in S-biotype even with a fusion to the adjacent gene. In the C-biotype, this adjacent gene is only expressed during the biotrophic interaction.

In *M. roreri*, it has been previously reported that *MrPR-1n*, *MrPR-1g* and *MrPR-1i2* were upregulated in samples from the biotrophic phase (30 days post infection of pods), *MrPR-1d* was upregulated in the necrotrophic phase (60 days post infection of pods) and five other *MrPR-1* were constitutively expressed under these conditions [27]. The heatmap in Fig. 3 shows that similar to *M. perniciosa's PR-1* expression profile, *MrPR-1g* is the most expressed *PR-1* gene during the biotrophic stage. Moreover, while *MrPR-1h* and *MrPR-1f* were not differentially expressed when comparing the biotrophic and necrotrophic stages, they also showed higher expression when compared to other *MrPR-1*s that belong to the conserved families.

Expression of recently diversified MpPR-1 was not induced by plant antifungal compounds

It has been previously demonstrated that CAP proteins of *M. perniciosa* bind to a variety of small hydrophobic ligands with different specificities. Thus, it has been suggested that the *MpPR-1* genes induced in the biotrophic interaction could function in the detoxification of hydrophobic molecules produced by the host as a defense strategy [24]. In this context, we investigated if *MpPR-1* genes, especially the ones induced in WBD (*c, f, h, i, k, g*), are differentially expressed by the presence of the plant antifungal compounds eugenol or α-tomatin, which are similar to sterol and fatty acids, respectively. However, when the necrotrophic mycelia of *M. perniciosa* was treated with eugenol, only *MpPR-1e, k, d* were up-regulated, while *MpPR-1f* was down-regulated in α-tomatin-treated samples (Fig. 5). In all samples, among all *MpPR-1* genes, *MpPR-1a* and *MpPR-1b* had the highest expression levels, while *MpPR-1c, i, g, j* have the lowest (TPM ≤ 2).

Discussion

The evolution of PR-1 and the emergence of pathogenicity among saprotrophs

The plant pathogen *M. perniciosa* has at least 11 genes encoding *PR-1*-like secreted proteins, which were previously identified and characterized in the genome of the C-biotype CP02 isolate [21]. Many of these genes were shown to be highly expressed during the biotrophic interaction of *M. perniciosa* and cacao, suggesting that MpPR-1 proteins have important roles during this stage of WBD. *M. perniciosa* has two other known biotypes (S and L) that differ in host specificity and virulence, the closest related species *M. roreri* that also is a *T. cacao* pathogen, and other nine *Moniliophthora* species: one described as a non-pathogenic grass endophyte [29], three of biotrophic/parasitic habit [30], and five species of unascertained lifestyle, found in dead or decaying vegetal substrates [31–33]. Because the majority of *Moniliophthora* related fungi in the Agaricales order are saprotrophs, the occurrence of parasitic *Moniliophthora* species raises the question about the emergence of biotrophic/parasitic lifestyle in this lineage of Marasmiaceae [30, 34]. The evolutionary scenario of host–pathogen arms race that emerges through the diversification of the *Moniliophthora* genus in the Agaricales order and of host-specific biotypes in *M. perniciosa* isolates, is especially suitable for the study of adaptive evolution in pathogenicity-related genes. Besides, the knowledge on putative adaptations gained through pathogen evolution are also specially interesting for further development of strategies against the pathogen. Based on our findings, Fig. 6 presents a model for the adaptive evolution of PR-1 proteins in *Moniliophthora*.

Through the characterization of the evolution of PR-1 proteins in Agaricales, we observe that at least one copy of PR-1 is present in all the sampled fungi, with most of the Agaricales species encoding between 1 and 7 PR-1 proteins. This is in contrast with *Moniliophthora* genomes, which encode among 10–12 proteins. *Moniliophthora* PR-1 proteins are derived independently from both ancient clades in the Agaricales gene tree. The longer branch-lengths in PR-1 families exclusive to *Moniliophthora* along with the evidence of evolution under positive selection identified in independently diverged clades suggest that the diversification of PR-1 in the genus was adaptive and related to its pathogenic lifestyle. The accentuated adaptive evolution of PR-1 in *Moniliophthora* is not only reflected in the genomic evolution of these genes, but also in their expression context. *PR-1c, f, g, h, i, k, l, n* are upregulated during the biotrophic interaction, while *PR-1*s that are also conserved in other Agaricales species are mainly expressed in mycelial stages of *M. perniciosa* (*MpPR-1a, b, d, e*). Most Agaricales species

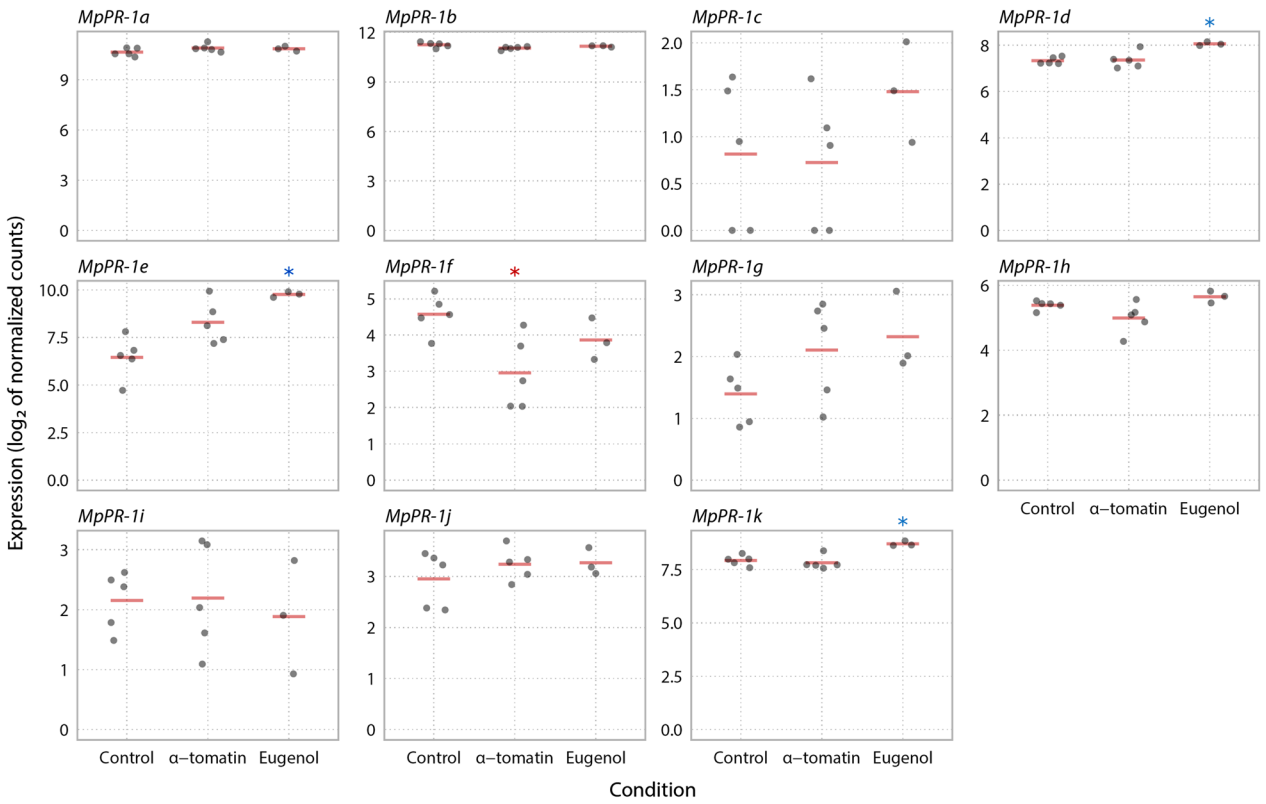


Fig. 5 Expression profile of *MpPR-1* genes in response to two plant antimicrobial compounds. The necrotrophic mycelium of *M. perniciosa* C-biotype (C-BA1a) was grown in liquid media in the presence of eugenol, α-tomatin or DMSO (mock condition) for 7 days. The expression values (log2 transformed) for each *MpPR-1* were obtained by RNA-Seq and subsequent quantification of read counts and between-sample normalization using size factors. Red bars indicate the mean of expression values within a group of replicates. Asterisks indicate that *MpPR-1e, k, d, f* are differentially expressed (s-value < 0.005) when compared to the mock condition, with blue asterisk indicating up-regulation and red indicating down-regulation

are not plant pathogens, which is also an evidence that PR-1 in *Moniliophthora* diverged from a few ancestral PR-1s that are related to basal metabolism in fungi and have been evolving under positive selective pressure possibly because of a benefit for the biotrophic/pathogenic lifestyle.

The emergence of SCP/TAPs proteins as pathogenicity factors has been reported in other organisms, such as the yeast *Candida albicans* and the filamentous fungus *Fusarium oxysporum* [10, 35]. Even though their specific function and mode of action may be different and remains to be characterized in plant pathogens, the recent evolution of these proteins towards their pathogenic role in the *Moniliophthora* genus could have contributed for the transition from a saprotrophic to parasitic lifestyle. Accelerated adaptive evolution evidenced by positive selection signs has also been observed in other virulence-associated genes of pathogenic fungi, such as the genes *PabaA*, *fos-1*, *pes1*, and *pksP* of *Aspergillus fumigatus*, which are involved in nutrient acquisition and oxidative stress response [36], and several gene families in *C. albicans*, including cell surface protein genes enriched in the most pathogenic *Candida* species [37].

Adaptive evolution of PR-1 within *Moniliophthora* species and its biotypes

The high diversification of PR-1 families observed within *Moniliophthora* was reinforced by our findings of positive selection in families that have also augmented expressions during infection: PR-1f, g, h, i, l. Among these five families, PR-1g and PR-1i are two of the most diversified families in *Moniliophthora* and have a more recent common ancestor with PR-1f than with the other PR-1s, placing this monophyletic clade of PR-1f, g, i as the key one to diversification and adaptive evolution of these proteins in the genus. Many sites that potentially evolved under selective pressure were also found in PR1-h, indicating that a parallel process of adaptive evolution occurred in this family.

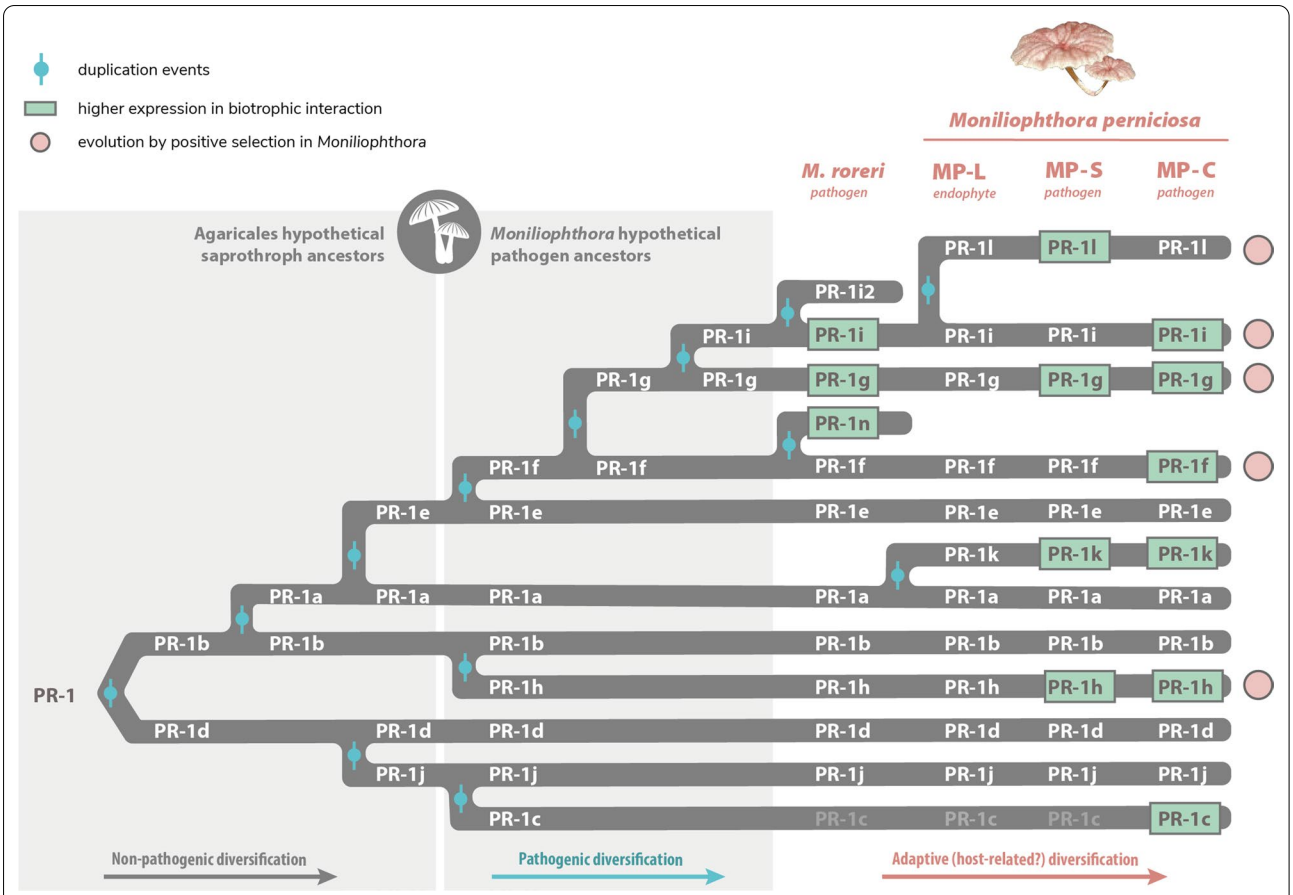


Fig. 6 Proposed model for the adaptive evolution of PR-1 proteins in *Moniliophthora* towards the pathogenic lifestyle. All *Moniliophthora* PR-1 proteins derived independently from two ancient clades (PR-1b-like and PR-1d-like) within the Agaricales order, as indicated in PR-1 phylogeny. The subsequent diversification of PR-1a and PR-1e from PR-1b, and PR-1j from PR-1d, occured in putative saprotroph lineages before the divergence of *Moniliophthora* genus, suggesting a diversification not related to pathogenicity. Within *Moniliophthora* hypothetical pathogenic ancestors, five other PR-1 proteins were derived (c from j, h from b, f-g-i from e) and most of these new lineages showed evidence of positive selection in *M. perniciosa* samples (indicated by pink circles). New PR-1 copies (n and i2 in *M. roreri*, l and k in *M. perniciosa*) diverged within *M.* species. Recently diversified PR-1 genes in *Moniliophthora,* not only show an elevated rate of evolution and positive selection evidence but are also predominantly expressed during the biotrophic interaction (indicated by green highlights). This supports the hypothesis that these proteins accumulated adaptive changes related to pathogen lifestyle that might also contribute to the host specialization observed in *Moniliophthora* species and biotypes

Within the diversification of PR-1 in *Moniliophthora*, four cases of putative species-specific evolution of PR-1 families were found: PR-1c, PR-1k and PR-1l in *M. perniciosa* and PR-1n in *M. roreri*. Although vestigial sequences indicate that PR-1c probably emerged as a paralog of PR-1j in the ancestral of both species, it was only kept in the evolution of C-biotype, in which a change in expression profile occurred, thus placing *MpPR-1c* as a case of PR-1 diversification within *M. perniciosa* biotypes and a potential candidate for host specificity. Another candidate for biotype-specific diversification is PR-1l, which diverged from a duplication of PR-1i. Even though PR-1l was found in all three *M. perniciosa* biotypes, it was expressed only in the S-biotype during the biotrophic interaction, instead of PR-1i, which is expressed in *M. roreri* and in the C-biotype. This suggests that the divergence of PR-1i can be host-specific, but further experiments are necessary to clarify if they are either a cause or consequence of *M. perniciosa* pathogenicity.

Even though almost all PR-1 families are present in the genomes of L-biotype isolates, this biotype has an endophytic lifestyle and does not cause visible disease symptoms in their hosts [18, 38]. There is no available expression data for the L-biotype, so it is unknown whether their PR-1 genes could have any role related to their lifestyle or these genes are not pseudogenized yet due to a small evolutionary time. Evans [18] reported that the L-biotype can induce weak symptoms in seedlings

of the Catongo variety of *T. cacao*. Therefore, it could be possible that host susceptibility is an important factor for the manifestation of WBD symptoms.

From basal metabolism to key roles in disease: how PR-1 proteins could have functionally adapted for pathogenicity?

It was previously shown that Pry proteins detoxify and protect yeast cells against eugenol [14] and that MpPR-1 proteins can bind to hydrophobic compounds secreted by plants, indicating that they could antagonize the host defense response [24]. When the necrotrophic mycelia of *M. perniciosa* was cultivated with eugenol, expression of *MpPR-1d* and *MpPR-1k* was up-regulated, which is in agreement with the ability of these two proteins to bind to plant and fungal sterol compounds [24]. *MpPR-1e* expression was also highly induced by eugenol, even though it was previously shown to bind only to fatty acids, but not sterols. Additionally, *MpPR-1f*, which is up-regulated during the biotrophic interaction like *MpPR-1k*, was down-regulated by α-tomatin. Given the above, it appears that *M. perniciosa* does not rely on MpPR-1 for cellular detoxification or this function is not transcriptionally regulated by plant hydrophobic compounds in the necrotrophic mycelia, or even, they could have different roles other than detoxification.

Considering that some PR-1 proteins are not associated with infection and are conserved in other saprotrophic fungi, here we hypothesize that the primary function of PR-1 in fungi can be related to the export of sterols from basal metabolism, such as ergosterol, the most abundant sterol in fungal cell membrane [39, 40]. PRY of *S. cerevisiae* transports acetylated ergosterol to the plasma membrane [13] and MpPR-1d, which belongs to a relatively conserved PR-1 family in Agaricales, can also efficiently bind to ergosterol [24]. Furthermore, ergosterol acts as a PAMP molecule (pathogen-associated molecular pattern) in plants [41], resulting in the activation of defense-related secondary metabolites and genes, including plant PR-1s [3, 42–44], which are likely to have a role in sequestering sterols from the membranes of microbes [16] and stress signaling [45, 46]. Additionally, PR-1 receptor-like kinases (PR-1-RLK) from *T. cacao* are also upregulated on WBD and could be binding to the same ligand of PR-1 [47]. Given that, it is possible that MpPR-1 could have evolved different adaptive roles through neofunctionalization. Besides export of hydrophobic compounds of basal metabolism, they could be acting in the protection of the cell membrane against the disruption caused by antifungal compounds, in the detoxification of hydrophobic compounds like phytoalexins secreted by the host, or it could even be possible that those *MpPR-1s* expressed during infection are sequestering the membrane sterols of the fungus itself in order to prevent detection by a possible ergosterol recognition complex from the host [48], thus compromising the elicitation of plant immunity in a similar fashion of MpChi, a chitinase-like effector that is highly expressed by *M. perniciosa* during the biotrophic stage of WDB [49].

It has been shown that the ability of MpPR-1 proteins to bind to sterols can be altered by a point mutation in the caveolin binding motif [24], highlighting the significance of understanding those sites under positive selection that are detected in important regions of the proteins, such as the candidate sites found in the CBM and alpha-helix 1 of PR-1i. These findings are central to learn how changes in the nucleotide or protein sequences could impact binding affinity and function. Even though this is speculative, as the specific role of PR-1 remains unknown, these results can guide further validation experiments and maybe demonstrate another case of adaptive evolution of fungal effectors.

Conclusions

Based on genomic and transcriptomic data, we presented evidence of adaptive evolution of PR-1 proteins in processes underlying the pathogenic lifestyle in *Moniliophthora*. These results reinforce the power of evolutionary analysis to reveal key proteins in the genomes of pathogenic fungi and contribute to the understanding of the evolution of pathogenesis. Our results indicate a set of PR-1 families that are putatively related to pathogenicity in the genus (PR-1f, g, h, i) and specialization within *M. perniciosa* biotypes (PR-1c, k and l) and *M. roreri* (PR-1n). The positive selection analysis also indicates protein sites that are putatively related to those adaptations. *PR-1* genes and sites with evidence of adaptations are strong candidates for further study and should be evaluated in order to understand how changes in these sites can affect structure, binding affinity and function of these proteins.

Methods

Identification of PR-1-like gene families

In this study, we used a dataset of families of genes predicted in 22 genomes of *Moniliophthora* (unpublished) and 16 genomes of other fungal species of the order Agaricales, which were obtained from the Joint Genome Institute (JGI) Mycocosm database [50]. The *Moniliophthora* genomes included are 7 isolates of the S-biotype (collected at the states of Amazonas and Minas Gerais, in Brazil), 9 isolates of the C-biotype (collected at the states of Amazonas, Pará, and Acre, in Brazil), 2 isolates of the L-biotype (from Colombia) and 4 samples of *M. roreri* (from Colombia). Additional file 1 contains the list of species and isolates, their genome identification and source (collection location or reference publication).

To identify candidate *PR-1* gene families, we performed a search for genes encoding the CAP/SCP/PR1-like domain (CDD: cd05381, Pfam PF00188) using the HMMER software [51]. The assignment of protein sequences to families of homologues (orthogroups) was done using Orthofinder (v. 1.1.2) [52]. In addition, we searched all gene families for families containing *PR-1* candidate genes with Blastp [53] using the known 11 MpPR-1 sequences [21] as baits, in order to search for possible candidates that were not previously identified and/or that have been wrongly assigned to other orthogroups due to incorrect gene prediction. To verify the presence of the SCP PR1-like/CAP domain (InterPro entry IPR014044) in the sequence, the InterProscan platform [54] was used. All *Moniliophthora* PR-1 candidate sequences identified in this study are deposited in GenBank under Accession numbers MW659198–MW659445.

Sequence alignment and phylogenetic reconstruction

For the inference of the phylogenetic history of the gene, the protein sequences of the PR-1 homologue families identified in the 22 *Moniliophthora* isolates were aligned with the PRY1 sequence of *S. cerevisiae* (GenBank ID CAA89372.1), which was used as outgroup. Multiple sequence alignments were performed with Mafft (v. 7.407) [55] using the iterative refinement method that incorporates local alignment information in pairs (L-INS-i), with 1000 iterations performed. Then, the alignments were used for phylogenetic reconstruction using the maximum likelihood method with IQ-Tree (v. 1.6.6) [56], which performs the selection of the best replacement model automatically, with 1000 bootstraps for branch support. Bootstraps were recalculated with BOOSTER (v. 0.1.2) for better support of branches in large phylogenies [57]. Likewise, the phylogenetic inference for PR-1 of the Agaricales group of species was performed with the alignment of the homologous proteins identified in the 16 species obtained from Mycocosm, 3 isolates of *M. perniciosa* (C-BA3, S-MG3, L-EC1, one representing each biotype), an isolate of *M. roreri* (R-CO1), and PRY1 of *S. cerevisiae* as the outgroup. To improve alignment quality, trimAl package [58] was used. For dN/dS analysis, considering each gene family independently, the phylogenetic reconstruction was performed using IQ-Tree (v. 1.6.6) [56] with the multiple local alignment of the protein sequences obtained with Mafft (v. 7.407) [55], and the codon-based alignment of the nucleotide sequences was performed with Macse (v. 2.01) [59].

Detection of positive selection signals

To search for genes and regions that are potentially under positive selection in each of the PR-1 families of the 22 isolates, the CodeML program of the PAML 4.7 package [60] was used with the ETEToolkit tool [61]. CodeML implements a modification of the model proposed by [62] to calculate the omega (rate of non-synonymous mutations (dN)/rate of synonymous mutations (dS)) of a coding gene from the multiple alignment sequences and phylogenetic relationships that have been previously inferred.

In order to detect positive selection signals in isolates or specific positions in the sequences, we performed tests with the "branch-site" model, which compares a null model (bsA1) in which the branch under consideration is evolving without restrictions ($dN/dS = 1$) against a model in which the same branch has sites evolving under positive selection (bsA) ($dN/dS > 1$) [63]. In these tests, those branches that were tested for significantly different evolving rates from the others (foreground branches ωfrg) are marked in the phylogenetic trees—in this case, the branches corresponding to the isolates of the C-biotype or S-biotype. To detect signs of positive selection at specific sites throughout the sequences, regardless of the isolate, we used the "sites" model (M2 and M1, NSsites 0 1 2) to test all branches of the phylogenetic trees.

In both tests, the models are executed several times with different initial omegas (0.2, 0.7, 1.2), and the models with the highest probability are selected for the hypothesis test, in which a comparison between the alternative model and the null model is made through a likelihood ratio test. If the alternative model is the most likely one (p-value < 0.05), then the possibility of positive selection ($\omega > 1$) can be accepted, and sites with evidence of selection (probability > 0.95) are reported by Bayes Empirical Bayes analysis (BEB) [63].

Gene amplification and synteny analysis of *PR-1c*

In order to confirm the presence or absence of *MpPR-1c* and *MpPR-1d* genes in the genomes of *M. perniciosa* isolates, these genes were amplified by polymerase chain reaction (PCR) from isolates C-AC1, C-BA1a, C-BA3, S-AM1, S-MG3, S-MG4, L-EC1 and L-EC2. The necrotrophic mycelia of these isolates was cultivated in 1.7% MYEA media (15 g/L agar; 5 g/L yeast extract, 17 g/L malt extract) at 28 °C for 14 days, then harvested and ground in liquid nitrogen for total DNA isolation with the phenol–chloroform method [64]. PCRs were performed with primers designed for *MpPR-1c* (F: 5′-GGATCCCGACTTGACAACTCCATCTCG-3′, R: 5′-GAGCTCTCACTCAAACTCCCCGTCATAAT-3′) and *MpPR-1d* (F: 5′-GGATCCCCCTCGCAATGGGTTTTC-3′, R: 5′-GTCGACTCAGTCAAGATCAGCCTGGAGA-3′) and amplifications cycles consisting of an initial stage of 94 °C for 3 min, 35 cycles of 95 °C for 30 s, 60 °C for 50 s and 72 °C for 1 min, and final extension at 72 °C for 10 min.

For synteny analysis, the positions of *PR-1j*, *PR-1c* and *PR-1d* genes were searched in the scaffolds of genomes C-BA3, S-MG2, R-CO2, L-EC1 and L-EC2 by blastn. The scaffolds were then excised 5000 bp upstream and 5000 bp downstream from the starting position of *PR-1j* in the scaffolds. The resulting 10,000 bp excised scaffolds were used for synteny analysis with Mummer (v. 4.0.0beta2) [65], using the C-BA3 sequence as the reference.

MpPR-1 expression data

MpPR-1 expression data in RPKM (Reads Per Kilobase per Million mapped reads) values from the C-biotype of *M. perniciosa* in seven biological conditions (dikaryotic mycelium 14 days, basidiomata, germinating spores, green broom, initial necrosis, advanced necrosis, dry broom) were downloaded from the Witches' Broom Disease Transcriptome Atlas (v. 1.1) (http://bioinfo08.ibi.unicamp.br/atlas/).

MpPR-1 expression data of *M. perniciosa* treated with plant antifungal compounds were obtained from RNA-seq data (unpublished). The C-BA1a isolate's necrotrophic mycelia was initially inoculated in 100 mL of liquid MYEA media and cultivated for 5 days under agitation of 150 rpm at 30 °C, then 5 mL of this initial cultivation were transferred to 50 mL of fresh MYEA liquid media containing eugenol (500 µM), α-tomatin (80 µM) or DMSO (250 µL, solvent control) and cultivated again under agitation of 150 rpm at 30 °C for 7 days. The total RNA was extracted using the Rneasy® Plant Mini Kit (Quiagen, USA) and quantified on a fluorimeter (Qubit, Invitrogen). cDNA libraries were prepared in five biological replicates for each treatment, plus biological control. The cDNA libraries were built from 1000 ng of total RNA using Illumina's TruSeq RNA Sample Prep kit, as recommended by the manufacturer. The libraries were prepared according to Illumina's standard procedure and sequenced on Illumina's HiSeq 2500 sequencer. The quality of raw sequences was assessed with FastQC (v.0.11.7) [66]. Read quantification was performed by mapping the generated reads against 16,084 gene models of the C-BA1a genome using Salmon (v.0.14.1) in mapping-based mode [67]. Read counts were normalized to Transcripts Per Million (TPM) values for plotting. Differential expression analysis was performed with the DESeq2 (v.1.22.2) package using Wald test and Log fold change shrinkage by the *apeglm* method (IfcThreshold = 0.1, s-value < 0.005) [68]. TPM values and DESeq2 results for *MpPR-1* genes in these experimental conditions are available at Additional file 2.

MpPR-1 expression data in TPM for the S-biotype was obtained from RNA-seq libraries of infected MicroTom tomato plants in seven different time points after inoculation (12 h, 24 h, 48 h, 5 days, 10 days, 20 days, 30 days) (unpublished). The quality of raw sequences was assessed with FastQC (v. 0.11.7) [66]. Next, Trimmomatic (v.0.36) [69] was used to remove adaptor-containing and low-quality sequences. Quality-filtered reads were then aligned against the S-MG1 or S-MG2 reference genome using HISAT2 (v.2.1.0) with default parameters [70]. Reads that mapped to coding sequences were counted with featureCounts (v.1.6.3) [71]. TPM values for *MpPR-1* genes in these experimental conditions are available at Additional file 3.

MrPR-1 expression data in TPM was obtained from RNA-Seq reads of *M. roreri* in the biotrophic (30 days after infection) and necrotrophic (60 days after infection) stages of frosty pod rot from [27]. Reads were mapped and quantified with Salmon (v.0.14.1) [67] using 17,910 gene models of *M. roreri* MCA 2997 (GCA_000488995) available at Ensembl Fungi.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12862-021-01818-5.

Additional file 1. List of fungal genomes and their source (collection site or reference publication. Excel file with table containing the species names and biotypes of the fungal genomes used for the identification of *PR-1*-like genes, the identification names we used for the genomes, and their source, which for *M. perniciosa* and *M. roreri* isolates corresponds to their collection site, and for the other Agaricales species corresponds to their reference publication.

Additional file 2. *MpPR-1* quantification data and differential expression results of *M. perniciosa* treated with eugenol or alpha-tomatin treatment. Excel file with three spreadsheets (three tabs). First spreadsheet contains a matrix of expression values in Transcripts Per Million (TPM) for *MpPR-1* genes of the necrotrophic mycelia of *M. perniciosa* (C-biotype) treated with eugenol (500 µM), α-tomatin (80 µM) or DMSO (250 µL) (solvent control) for 7 days. Quantification was performed from RNA-Seq reads using the C-BA1a genome as reference. Second spreadsheet contains the results table for *MpPR-1* genes in the differential expression analysis comparing the expression profiles between Eugenol vs Control treatments, while comparison between α-tomatin vs Control treatments is shown in the third spreadsheet.

Additional file 3. *MpPR-1* expression data in S-biotype infection in tomato. Excel file containing the matrix of expression values in Transcripts Per Million (TPM) for *MpPR-1* genes of *M. perniciosa* S-biotype infection in MicroTom tomato in various time points of infection (12 h, 24 h, 48 h, 5 days, 10 days, 20 days, 30 days). Quantification was performed from RNA-Seq reads using the S-MG1 (data in first spreadsheet) or S-MG2 genome (data in second spreadsheet) as reference.

Additional file 4. Phylogenetic reconstruction of PR-1 proteins in *Moniliophthora* isolates (version with non-collapsed branches). Figure of phylogenetic tree with non-collapsed branches of PR-1 proteins identified from genomes of 18 *M. perniciosa* and 4 *M. roreri* isolates, inferred by maximum likelihood and branch support obtained using 1000 bootstraps. The PRY1 protein of *Saccharomyces cerevisiae* was used as an outgroup.

Additional file 5. Protein sequence alignment and omega (dN/dS) values of each PR-1i family of *Moniliophthora* genomes. PDF file containing figures of the protein sequence alignment for each PR-1 family from *Moniliophthora* isolates and a bar chart of omega (dN/dS) values calculated

for each amino acid site along the alignment using the site model test of codeML. Sites with positive selection signs are indicated by red (p-value ≤ 0.01) or orange (p-value ≤ 0.05) bars in the bar chart of omega values below the alignment. Figures are organized by alphabetical order of the names of PR-1 families.

Acknowledgements
We are thankful to Dr. Jorge M. Mondego for helping with the conceptualization of the experiments of *M. perniciosa* treated with antifungal compounds and M.sc. Bárbara A. Pires and Dr. Mario O. Barsottini for helping with the execution of these experiments. We thank M.sc. Leandro C. do Nascimento for performing the assembly of *Moniliophthora* genomes.

Authors' contributions
JJ and RMB conceived and supervised this project. AAV performed identification of PR-1-like candidate genes, evolutionary analysis, and most expression analysis from RNA-seq data, executed PCR experiments and generated figures. PJPLT and DPTT conceived the project of genomics of *Moniliophthora* isolates. PJPLT, DPTT, PFVP and GLF executed genomic data acquisition of *Moniliophthora* isolates. JLC executed RNA-seq data acquisition of MT plants infected with S-biotype, and PJPLT analyzed this data. PMTF performed gene prediction, annotation, and assignment of orthogroups from genomes. APC helped with genomic and RNA-seq analysis. RMB conceived and executed RNA-seq data acquisition of *M. perniciosa* treated with antifungal compounds and AAV analyzed this data. AAV and JJ wrote the original draft. JJ and APC improved the design of figures. JJ, RMB, PJPLT, DPTT, GLF, APC, PFVP and GAGP edited the draft. MFC, AF and GAGP contributed with project supervision and funding acquisition. All authors read and approved the final manuscript.

Funding
This work was supported by the São Paulo Research Foundation (FAPESP) Grants to M.F.C (#2013/08293-7), G.A.G.P. and A. F. (#2016/10498-4), and FAPESP fellowships to A.A.V (#2017/13015-7), A.P.C. (#2018/04240–0), P.J.P.L.T. (#2009/51018–1), G.L.F (#2011/23315-1, #2013/09878-9, #2014/06181-0), P.F.V.P. (#2013/05979-5, #2014/00802-2), J.L.C. (#2013/04309-6) and R.M.B (#2017/13319-6). The funders were not involved in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials
The datasets supporting the conclusions of this article are included within the article and its additional files. All *Moniliophthora* PR-1 candidate sequences identified in this study are deposited in GenBank under Accession numbers MW659198–MW659445.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil. [2]Departamento de Ciências Biológicas, Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), Universidade de São Paulo, Piracicaba, SP, Brazil. [3]Centro de Energia Nuclear Na Agricultura, Universidade de São Paulo, Piracicaba, SP, Brazil.

Received: 19 March 2021 Accepted: 26 April 2021
Published online: 14 May 2021

References

1. Cantacessi C, Campbell BE, Visser A, Geldhof P, Nolan MJ, Nisbet AJ, et al. A portrait of the "SCP/TAPS" proteins of eukaryotes—developing a framework for fundamental research and biotechnological outcomes. Biotechnol Adv. 2009;27(4):376–88.
2. Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, cancer, and immune defense. Endocr Rev. 2008;29(7):865–97.
3. van Loon LC, Rep M, Pieterse CMJ. Significance of inducible defense-related proteins in infected plants. Annu Rev Phytopathol. 2006;44:135–62.
4. Asojo OA, Goud G, Dhar K, Loukas A, Zhan B, Deumic V, et al. X-ray structure of Na-ASP-2, a pathogenesis-related-1 protein from the nematode parasite, Necator americanus, and a vaccine antigen for human hookworm infection. J Mol Biol. 2005;346(3):801–14.
5. Chalmers IW, McArdle AJ, Coulson RM, Wagner MA, Schmid R, Hirai H, et al. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. BMC Genomics. 2008;9(1):89.
6. Ding X, Shields J, Allen R, Hussey RS. Molecular cloning and characterisation of a venom allergen AG5-like cDNA from *Meloidogyne incognita*. Int J Parasitol. 2000;30(1):77–81.
7. Gao B, Allen R, Maier T, Davis EL, Baum TJ, Hussey RS. Molecular characterisation and expression of two venom allergen-like protein genes in *Heterodera glycines*. Int J Parasitol. 2001;31(14):1617–25.
8. Hawdon JM, Narasimhan S, Hotez PJ. Ancylostoma secreted protein 2: cloning and characterization of a second member of a family of nematode secreted proteins from *Ancylostoma caninum*. Mol Biochem Parasitol. 1999;99(2):149–65.
9. Lozano-Torres JL, Wilbers RHP, Warmerdam S, Finkers-Tomczak A, Diaz-Granados A, van Schaik CC, et al. Apoplastic venom allergen-like proteins of cyst nematodes modulate the activation of basal plant innate immunity by cell surface receptors. PLoS Pathog. 2014;10(12):e1004569.
10. Prados-Rosales RC, Roldán-Rodríguez R, Serena C, López-Berges MS, Guarro J, Martínez-del-Pozo Á, et al. A PR-1-like protein of *Fusarium oxysporum* functions in virulence on mammalian hosts. J Biol Chem. 2012;287(26):21970–9.
11. Schneiter R, Di Pietro A. The CAP protein superfamily: function in sterol export and fungal virulence. Biomol Concepts. 2013;4(5):519–25.
12. Zhan B, Liu Y, Badamchian M, Williamson A, Feng J, Loukas A, et al. Molecular characterisation of the Ancylostoma-secreted protein family from the adult stage of *Ancylostoma caninum*. Int J Parasitol. 2003;33(9):897–907.
13. Choudhary V, Schneiter R. Pathogen-related yeast (PRY) proteins and members of the CAP superfamily are secreted sterol-binding proteins. Proc Natl Acad Sci USA. 2012;109(42):16882–7.
14. Darwiche R, Mène-Saffrané L, Gfeller D, Asojo OA, Schneiter R. The pathogen-related yeast protein Pry1, a member of the CAP protein superfamily, is a fatty acid-binding protein. J Biol Chem. 2017;292(20):8304–14.
15. Darwiche R, Schneiter R. Cholesterol-binding by the yeast CAP family member Pry1 requires the presence of an aliphatic side chain on cholesterol. J Steroids Horm Sci. 2016. https://doi.org/10.4172/2157-7536.1000172.
16. Gamir J, Darwiche R, Van't Hof P, Choudhary V, Stumpe M, Schneiter R, et al. The sterol-binding activity of PATHOGENESIS-RELATED PROTEIN 1 reveals the mode of action of an antimicrobial protein. Plant J Cell Mol Biol. 2017;89(3):502–9.
17. Kelleher A, Darwiche R, Rezende WC, Farias LP, Leite LCC, Schneiter R, et al. Schistosoma mansoni venom allergen-like protein 4 (SmVAL4) is a novel lipid-binding SCP/TAPS protein that lacks the prototypical CAP motifs. Acta Crystallogr D Biol Crystallogr. 2014;70(Pt 8):2186–96.
18. Evans HC. Witches' broom disease of cocoa Crinipellis perniciosa) in Ecuador. Ann Appl Biol. 1978;89(2):185–92.
19. Evans HC. Cacao diseases—the trilogy revisited. Phytopathology®. 2007;97(12):1640–3.
20. Purdy L, Schmidt R. STATUS OF CACAO WITCHES' BROOM: biology, epidemiology, and management. Annu Rev Phytopathol. 1996;34(1):573–94.
21. Teixeira PJPL, Thomazella DT, Vidal RO, do Prado PFV, Reis O, Baroni RM, et al. The fungal pathogen *Moniliophthora perniciosa* has genes similar to plant PR-1 that are highly expressed during its interaction with cacao. PLoS ONE. 2012;7(9):e45929.

22. Teixeira PJPL, Thomazella DPdT, Reis O, do Prado PFV, do Rio MCS, Fiorin GL, et al. High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. Plant Cell. 2014;26(11):4245–69.
23. Baroni RM, Luo Z, Darwiche R, Hudspeth EM, Schneiter R, Pereira GAG, et al. Crystal structure of MpPR-1i, a SCP/TAPS protein from *Moniliophthora perniciosa*, the fungus that causes Witches' Broom Disease of Cacao. Sci Rep. 2017;7(1):7818.
24. Darwiche R, El Atab O, Baroni RM, Teixeira PJPL, Mondego JMC, Pereira GAG, et al. Plant pathogenesis-related proteins of the cacao fungal pathogen *Moniliophthora perniciosa* differ in their lipid-binding specificities. J Biol Chem. 2017;292(50):20558–69.
25. Oleksyk TK, Smith MW, O'Brien SJ. Genome-wide scans for footprints of natural selection. Philos Trans R Soc B Biol Sci. 2010;365(1537):185–205.
26. Manel S, Perrier C, Pratlong M, Abi-Rached L, Paganini J, Pontarotti P, et al. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. Mol Ecol. 2016;25(1):170–84.
27. Meinhardt LW, Costa GGL, Thomazella DP, Teixeira PJP, Carazzolle MF, Schuster SC, et al. Genome and secretome analysis of the hemibiotrophic fungal pathogen, *Moniliophthora roreri*, which causes frosty pod rot disease of cacao: mechanisms of the biotrophic and necrotrophic phases. BMC Genomics. 2014;15(1):164.
28. Deganello J, Leal GA, Rossi ML, Peres LEP, Figueira A. Interaction of *Moniliophthora perniciosa* biotypes with Micro-Tom tomato: a model system to investigate the witches' broom disease of *Theobroma cacao*. Plant Pathol. 2014;63(6):1251–63.
29. Aime MC, Phillips-Mora W. The causal agents of witches' broom and frosty pod rot of cacao (chocolate, Theobroma cacao) form a new lineage of Marasmiaceae. Mycologia. 2005;97(5):1012–22.
30. Niveiro N, Ramírez NA, Michlig A, Lodge DJ, Aime MC. Studies of Neotropical tree pathogens in Moniliophthora: a new species, *M. mayarum*, and new combinations for *Crinipellis ticoi* and *C. brasiliensis*. MycoKeys. 2020;66:39–54.
31. Kerekes J, Desjardin D, Desjardin D. A monograph of the genera Crinipellis and Moniliophthora from Southeast Asia including a molecular phylogeny of the nrITS region. Fungal Divers. 2009;37:101–52.
32. Kropp BR, Albee-Scott S. *Moniliophthora aurantiaca* sp. nov., a Polynesian species occurring in littoral forests. Mycotaxon. 2012;120(1):493–503.
33. Takahashi H. Four new species of Crinipellis and Marasmius in eastern Honshu, Japan. Mycoscience. 2002;43(4):343–50.
34. Teixeira PJPL, Thomazella DPdT, Pereira GAG. Time for chocolate: current understanding and new perspectives on Cacao Witches' broom disease research. PLoS Pathog. 2015;11(10):e1005130.
35. Braun BR, Head WS, Wang MX, Johnson AD. Identification and characterization of TUP1-regulated genes in *Candida albicans*. Genetics. 2000;156(1):31–44.
36. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, et al. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLOS Genet. 2008;4(4):e1000046.
37. Butler G, Rasmussen MD, Lin MF, Santos MAS, Sakthikumar S, Munro CA, et al. Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature. 2009;459(7247):657–62.
38. Griffith GW, Hedger JN. Spatial distribution of mycelia of the liana (L-) biotype of the agaric *Crinipellis perniciosa* (Stahel) Singer in tropical forest. New Phytol. 1994;127(2):243–59.
39. Mohd As'wad AW, Sariah M, Paterson RRM, Zainal Abidin MA, Lima N. Ergosterol analyses of oil palm seedlings and plants infected with Ganoderma. Crop Prot. 2011;30(11):1438–42.
40. Zhao XR, Lin Q, Brookes PC. Does soil ergosterol concentration provide a reliable estimate of soil fungal biomass? Soil Biol Biochem. 2005;37(2):311–7.
41. Nürnberger T, Brunner F, Kemmerling B, Piater L. Innate immunity in plants and animals: striking similarities and obvious differences. Immunol Rev. 2004;198:249–66.
42. Kasparovsky T, Milat M-L, Humbert C, Blein J-P, Havel L, Mikes V. Elicitation of tobacco cells with ergosterol activates a signal pathway including mobilization of internal calcium. Plant Physiol Biochem. 2003;41(5):495–501.
43. Klemptner RL, Sherwood JS, Tugizimana F, Dubery IA, Piater LA. Ergosterol, an orphan fungal microbe-associated molecular pattern (MAMP). Mol Plant Pathol. 2014;15(7):747–61.
44. Lochman J, Mikes V. Ergosterol treatment leads to the expression of a specific set of defence-related genes in tobacco. Plant Mol Biol. 2006;62(1–2):43–51.
45. Chen Y-L, Lee C-Y, Cheng K-T, Chang W-H, Huang R-N, Nam HG, et al. Quantitative peptidomics study reveals that a wound-induced peptide from PR-1 regulates immune signaling in tomato. Plant Cell. 2014;26(10):4135–48.
46. Chien P-S, Nam HG, Chen Y-R. A salt-regulated peptide derived from the CAP superfamily protein negatively regulates salt-stress tolerance in Arabidopsis. J Exp Bot. 2015;66(17):5301–13.
47. Teixeira PJPL, Costa GGL, Fiorin GL, Pereira GAG, Mondego JMC. Novel receptor-like kinases in cacao contain PR-1 extracellular domains. Mol Plant Pathol. 2013;14(6):602–9.
48. Khoza TG, Dubery IA, Piater LA. Identification of candidate ergosterol-responsive proteins associated with the plasma membrane of *Arabidopsis thaliana*. Int J Mol Sci. 2019;20(6):1302.
49. Fiorin GL, Sanchéz-Vallet A, Thomazella DPdT, do Prado PFV, do Nascimento LC, Figueira AVdO, et al. Suppression of plant immunity by fungal chitinase-like effectors. Curr Biol CB. 2018;28(18):3023-3030.e5.
50. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 2014;42(Database issue):D699-704.
51. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7(10):e1002195.
52. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1):157.
53. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinform. 2009;10(1):421.
54. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. Nucleic Acids Res. 2009;37(Database issue):D211–5.
55. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
56. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
57. Lemoine F, Domelevo Entfellner J-B, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature. 2018;556(7702):452–6.
58. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics (Oxf, Engl). 2009;25(15):1972–3.
59. Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol Biol Evol. 2018;35(10):2582–4.
60. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
61. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol. 2016;33(6):1635–8.
62. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol. 1994;11(5):725–36.
63. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 2005;22(12):2472–9.
64. Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol:chloroform. CSH Protoc. 2006. https://doi.org/10.1101/pdb.prot4455.
65. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2):R12.
66. Babraham Bioinformatics—FastQC a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 6 Mar 2021.
67. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417–9.
68. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014. https://doi.org/10.1186/s13059-014-0550-8.

69. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
70. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.
71. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.

Publisher's Note