

RESEARCH ARTICLE

Open Access

# 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

Lyndel W Meinhardt<sup>1\*</sup>, Gustavo Gilson Lacerda Costa<sup>2</sup>, Daniela PT Thomazella<sup>3</sup>, Paulo José PL Teixeira<sup>3</sup>, Marcelo Falsarella Carazzolle<sup>2</sup>, Stephan C Schuster<sup>4</sup>, John E Carlson<sup>5</sup>, Mark J Gultinan<sup>6</sup>, Piotr Mieczkowski<sup>7</sup>, Andrew Farmer<sup>8</sup>, Thiruvarangan Ramaraj<sup>8</sup>, Jayne Crozier<sup>9</sup>, Robert E Davis<sup>10</sup>, Jonathan Shao<sup>10</sup>, Rachel L Melnick<sup>1</sup>, Gonçalo AG Pereira<sup>3</sup> and Bryan A Bailey<sup>1</sup>

## Abstract

**Background:** The basidiomycete *Moniliophthora roreri* is the causal agent of Frosty pod rot (FPR) disease of cacao (*Theobroma cacao*), the source of chocolate, and FPR is one of the most destructive diseases of this important perennial crop in the Americas. This hemibiotroph infects only cacao pods and has an extended biotrophic phase lasting up to sixty days, culminating in plant necrosis and sporulation of the fungus without the formation of a basidiocarp.

**Results:** We sequenced and assembled 52.3 Mb into 3,298 contigs that represent the *M. roreri* genome. Of the 17,920 predicted open reading frames (ORFs), 13,760 were validated by RNA-Seq. Using read count data from RNA sequencing of cacao pods at 30 and 60 days post infection, differential gene expression was estimated for the biotrophic and necrotrophic phases of this plant-pathogen interaction. The sequencing data were used to develop a genome based secretome for the infected pods. Of the 1,535 genes encoding putative secreted proteins, 1,355 were expressed in the biotrophic and necrotrophic phases. Analysis of the data revealed secretome gene expression that correlated with infection and intercellular growth in the biotrophic phase and invasive growth and plant cellular death in the necrotrophic phase.

**Conclusions:** Genome sequencing and RNA-Seq was used to determine and validate the *Moniliophthora roreri* genome and secretome. High sequence identity between *Moniliophthora roreri* genes and *Moniliophthora perniciosa* genes supports the taxonomic relationship with *Moniliophthora perniciosa* and the relatedness of this fungus to other basidiomycetes. Analysis of RNA-Seq data from infected plant tissues revealed differentially expressed genes in the biotrophic and necrotrophic phases. The secreted protein genes that were upregulated in the biotrophic phase are primarily associated with breakdown of the intercellular matrix and modification of the fungal mycelia, possibly to mask the fungus from plant defenses. Based on the transcriptome data, the upregulated secreted proteins in the necrotrophic phase are hypothesized to be actively attacking the plant cell walls and plant cellular components resulting in necrosis. These genes are being used to develop a new understanding of how this disease interaction progresses and to identify potential targets to reduce the impact of this devastating disease.

\* Correspondence: lyndel.meinhardt@ars.usda.gov

<sup>1</sup>Sustainable Perennial Crops Lab, USDA/ARS, Bldg 001 Rm 223 Beltsville  
Agricultural Research Center-West, Beltsville, MD 20705, USA  
Full list of author information is available at the end of the article

## Background

Fungal plant pathogens can be classified as biotrophic, necrotrophic or hemibiotrophic pathogens based on how they interact with their host. Biotrophic pathogens cause only minor responses from the plant, particularly at initial stages of the disease. These biotrophic pathogens appear to evade plant defenses with stealthy methods [1]. Fungal biotrophs are often obligate pathogens, typically having narrow host ranges, possessing haustoria and secreting limited amounts of lytic enzymes [2]. On the other hand, infection by necrotrophic pathogens causes rapid cell death in hosts and elicit major molecular responses from the plant. Necrotrophs appear to utilize brute force and overwhelm the plant defenses. Necrotrophs are typically non-obligate pathogens, have wide host ranges and secrete copious amounts of lytic enzymes and toxins [2]. Hemibiotrophs initiate infection with a period of biotrophy, followed by a necrotrophic phase, and they possess properties of both groups. However, most of our understanding of how hemibiotrophs interact with their hosts is derived from these two extremes.

Both biotrophic and necrotrophic fungi share common elements but these may have different purposes when causing disease. During the host-interaction, the pathogens synthesize and secrete various peptides/proteins that block host responses (biotrophs) or kill the host cells (necrotrophs). Among biotrophs, the rust fungus of flax, *Melampsora lini*, excretes cysteine-rich avirulence elicitor proteins from the haustoria [3] and the bean rust fungus *Uromyces fabae* shows highly coordinated stage specific regulation of its secreted proteins [4]. The necrotrophic fungus *Fusarium graminearum* has 109 secreted cell wall degrading genes in its genome [5], while *Alternaria* species secrete non-host and host specific toxins that disrupt photosynthesis and kill plant cells [6]. Therefore, a detailed understanding of specific peptides/proteins secreted during the host-pathogen interaction is vital to elucidate the biotrophic and necrotrophic mechanisms.

*Moniliophthora roreri* (Cif.) H.C. Evans, Stalpers, Samson & Benny [7] causes Frosty Pod Rot (FPR), a devastating pod disease of *Theobroma cacao* L. (cacao), the source of cocoa powder and cocoa butter. Phylogenetically, *M. roreri* is related to another fungal pathogen, *Moniliophthora perniciosa* (Stahel) Aime and Phillips-Mora [8], which causes Witches' Broom Disease (WBD), a disease that infects all cacao meristematic tissues including flowers, shoots, and pods [8]. Together, these pathogens cause two of the most economically important diseases of *Theobroma cacao* in the Western Hemisphere [9,10]. While both of these fungal species are pathogenic on the plant genera *Theobroma* and *Herrania*, *M. roreri* is not known to have any other hosts, whereas *M. perniciosa* has distinct biotypes that infect different host species [11]. Historically, these *Moniliophthora* pathogens

have spread independently to cacao producing areas across the Western Hemisphere and they have typically resulted in production losses of 75% or higher in nearly all the cacao growing regions in the Americas [10,12].

*Moniliophthora roreri* and *M. perniciosa* are both hemibiotrophic pathogens, but have distinctive lifestyles and pathogenicity strategies. Unlike most other hemibiotrophic fungi, both *Moniliophthora* species have protracted biotrophic stages or phases that last three to six weeks. They also have distinctly different mycelial morphologies present at the beginning and the end of the disease process [10,13]. In both *Moniliophthora* diseases, the infected plant tissues are asymptomatic for 14 to 21 days. After that period of time, these tissues typically begin to show some form of altered growth or swelling that continues for the remainder of the biotrophic phase, culminating with the necrosis of the host tissues, which marks the beginning of the necrotrophic phase. In other hemibiotrophic fungi such as *Cladosporium fulvum*, *Mycosphaerella graminicola*, *Pyrenopeziza brassicae* and *Septoria tritici*, the biotrophic phase is asymptomatic with undifferentiated mycelia and lasts no more than two weeks [14].

As with hemibiotrophic *Colletotrichum* species, both *M. roreri* and *M. perniciosa* undergo distinct metabolic changes as they transition from the biotrophic phase to the necrotrophic phase of growth [15-19]. These metabolic changes suggest significant shifts in gene expression and in enzymatic processes.

In this study we present the genome of *M. roreri* and compare it to the genome of the closely related cacao pathogen, *M. perniciosa*. Transcriptional analysis of the genome will focus on genes with putative signal peptides from the *M. roreri* genome and compare them with expressed genes, identified through RNA sequencing of infected pods, to discover potential secreted proteins expressed during the disease process. Genes expressed in the early and late stages of FPR on pods were used to identify secreted genes that are potentially important during the biotrophic or necrotrophic phase of the disease cycle. The results presented here are the first analyses of genes encoding secreted proteins derived from the *M. roreri* genome. These new insights into the *M. roreri* genome, transcriptome and differential gene regulation provide a better understanding of the plant-pathogen interactions that occur during FPR.

## Results

### Genome structure

*Moniliophthora roreri* has a genome size of 52.3 Mbp, which is 7.7 Mbp larger than the 44.6 Mbp found in the *M. perniciosa* genome (Table 1). Despite the size difference, *M. roreri* had only 912 additional coding sequences (CDS) than *M. perniciosa*. 17,920 coding sequences were detected in *M. roreri*. The average GC% for the *M. roreri*

**Table 1 Genome Comparisons of *M. roseri* and *M. perniciosa***

	Genome size (bp)	Average (G + C)%	Total number contigs	N50 (bp)	Max contig size (bp)	Min contig size (bp)	Median contig size	Total tRNA	Avg gene Density	Total CDS	Avg (G + C) CDS%	Maximum CDS size (bp)	Median CDS size (bp)
<i>M. roseri</i>	52,334,075	46.88%	3,298	48,134	571,142	367	5,455	670*	0.42185#	17,920	49.55%	15,081	1,023
<i>M. perniciosa</i>	44,661,472	47.81%	3,087	48,096	476,325	500	3,300	550	0.51058#	17,008	49.83%	15,048	1,095

\*using tRNAscan default parameters yielded 331 contigs that have tRNA for a total of 670 tRNA's.  
 #CDS bases/total genome bases.

genome (46.88%) is lower than for the *M. perniciosa* genome (47.81%); however, the average G + C for the *M. roseri* CDS is nearly identical to the value for *M. perniciosa*. Gene density for *M. roseri* was also found to be lower than for *M. perniciosa*.

An analysis of the genomes of these pathogens has revealed that retro-transposons (LTRs) contribute to about 6.6% (3.46 Mbp) of the *M. roseri* genome compared to about 0.74% (0.332 Mbp) in the *M. perniciosa* genome. In *M. perniciosa*, transposable elements were reported to be active elements that may contribute to genetic variability of this pathogen [20]. Each genome was scanned for repetitive sequences and low complexity DNA using the Repeat Modeler and RepeatMasker programs. *M. roseri* had 7,060,129 bases associated with repeats and repetitive elements while *M. perniciosa* had 1,737,865 bases associated with repeats and repetitive elements (Additional file 1).

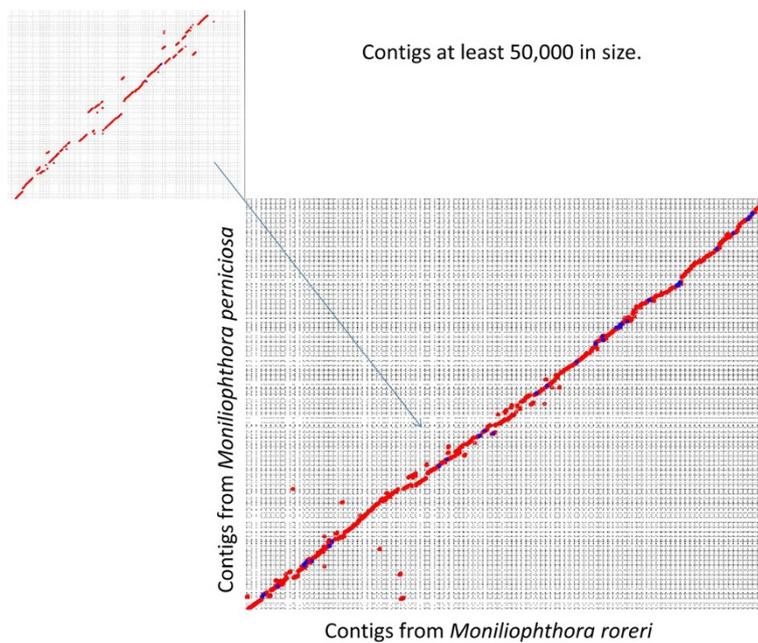
### Synteny

Using a bi-directional blast with an E-value of E-04 as the cut-off value for two proteins being considered a match, a whole genome comparison of the predicted proteins between the *M. roseri* and *M. perniciosa* genomes was performed. This analysis revealed that 16,713 protein encoding genes (93%) from *M. roseri* have sequence similarity to *M. perniciosa* protein encoding genes and 15,674 protein encoding genes (92%) from *M. perniciosa* have sequence similarity to *M. roseri* protein encoding genes. At lower E-values, E-25 and E 0.0, the *M. roseri* genome open reading frames (ORFs) are 87% and 40% similar to the *M. perniciosa* genome ORFs, respectively. In *M. roseri* there are 8,851 predicted genes or 49% that have no predicted function and are listed as hypothetical genes. Synteny between *M. roseri* and *M. perniciosa* was analyzed with the program MUMmer using the NUCmer setting that compares the contigs from both genomes at the nucleotide level. In this analysis, *M. roseri* contigs were queried using *M. perniciosa* contigs as the reference sequence. Contigs with high sequence identity resulted in a straight line graph, while translocations and inversions result in either a parallel shift in the line (up or down) or a line at a negative angle, respectively. Though both genomes are closely related at the nucleotide level, there are regions where inversions and translocations have

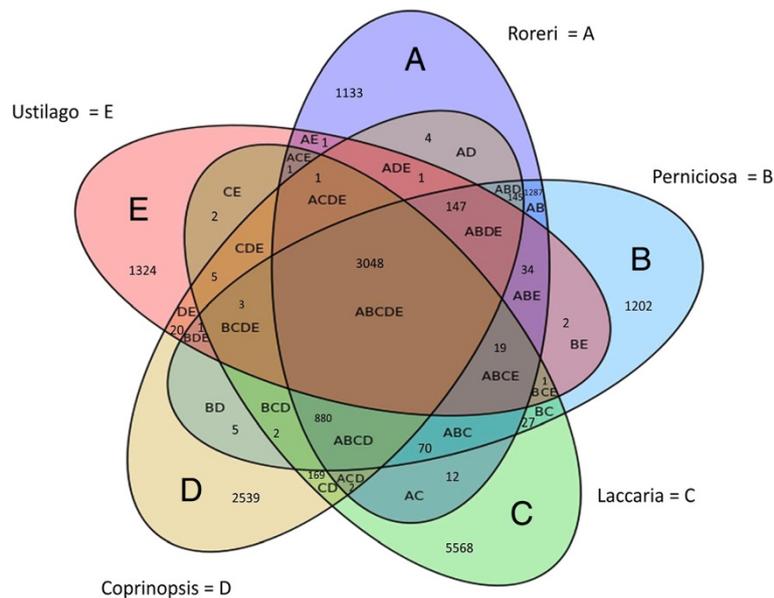
potentially occurred. This analysis confirms that these two genomes are very similar and most probably homologous, while it also shows that some regions in both genomes are unique. Due to the poor resolution of the MUMmer plot when all genome contigs were analyzed, we compared long intact regions of the genomes. Two hundred and fifty contigs from *M. roseri* and 233 contigs from *M. perniciosa* were compared that were at least 50 kb in length. Contigs with no corresponding sequence identity were eliminated, so that only contigs of high sequence identity between the two genomes were compared (Figure 1). The results yielded a final comparison of 222 contigs from *M. roseri* and 207 contigs from *M. perniciosa* and showed highly similar sequences with slight shifts off the diagonal line, which indicate translocations in the contigs (Figure 1 insert that shows an expanded region of the main figure).

### Comparisons to other basidiomycete fungi

Bidirectional blast analysis was conducted with *M. roseri*, *M. perniciosa*, *Laccaria bicolor*, *Coprinopsis cinerea* (*Coprinus cinereus*), and *Ustilago maydis*. These other genomes were selected because all are basidiomycetes but exhibit differing environmental adaptations representing a mushroom producing ectomycorrhizae a mushroom producing saprophyte and a pathogenic smut fungus, respectively. Figure 2 presents the Venn-diagram of the analysis, and the tax plot is shown in Table 2. Each of the five genomes is represented by an oval, and the numbers indicate how many genes are included within that intersect, at an E-value of E-04. Using this analysis, genes can be identified that are exclusive to a particular genome or to a particular group, which enables identification of putative similarities such as pathogenicity genes, and genus specific genes. The number of predicted genes particular to a single genome is: 1,133 for (A) *M. roseri*, 1,202 for (B) *M. perniciosa*, 5,568 for (C) *L. bicolor*, 2,539 for (D) *C. cinerea* and 1,324 for (E) *U. maydis*. Only seven of the 1,133 *M. roseri* genes were detected by RNA-Seq analysis (described below) and only 16 of the 1,133 *M. roseri* genes have putative functions (Table 3). Based on the annotation, several of these genes appear to have a role in genome integrity and function (a RNA polymerase Rpb1 C-terminal repeat domain-containing protein; a putative RecQ helicase; a P-loop containing nucleoside triphosphate hydrolase protein) while another set appears to be linked to retro



**Figure 1 MUMmer alignment dot plot of *M. roleri* and *M. pernicioso* contigs using only contigs 50 kb or larger.** The line graph represents the MUMmer results for the nucleotide comparison of *M. roleri* contigs to *M. pernicioso* contigs with a size cutoff of 50 kb. Only contigs larger than 50 kb were compared and contigs with no corresponding homology were eliminated from this particular analysis. Red circles represent positive strand alignments while blue circles represent negative strand alignments. The insert represent an enlargement of the MUMmer graph at that location. The 50 kb contigs utilized in this analysis for *M. roleri* account for 23,142,326 bp or 44% of the total 52,334,081 bp genome, where as for *M. pernicioso* 21,535,677 bp were, used which represents 48% of the 44,661,472 bp genome.



**Figure 2 Bi-directional Venn diagram.** Bi-directional blast results are present in a Venn diagram. The code used for this diagram is: **A** oval = *M. roleri* genes; the **B** oval = *M. pernicioso* genes; **C** oval = *L. bicolor* genes; **D** oval = *C. cinerea* genes and **E** oval = *U. maydis* genes. Intersects are labeled with the corresponding letters and a number which represents the number of specific genes in that particular intersect. An E-value E-04 was used as the homology cut off.

**Table 2 Tax plot results of the bi-directional blast analysis**

Bidirectional blast results			Venn intersection #	
Number of genes with homology at an E-value E0-4				
A-1133			1133	
AB-2982	BA-2508		1287	
AC-34	CA-86		12	
AD-13	DA-26		4	
AE-2	EA-17		1	
ABC-866	BAC-526	CAB-479	70	
ABD-540	BAD-486	DAB-395	145	
ABE-136	BAE-125	EAB-100	34	
ACD-25	CAD-101	DAC-57	2	
ACE-0	CAE-2	EAC-5	1	
ADE-0	DAE-2	EAD-8	1	
ABCD-3650	BACD-3819	CABD-3739	DABC-2957	880
ABCE-192	BACE-144	CABE-89	EABC-123	19
ABDE-348	BADE-355	DABE-265	EABD-254	147
ACDE-0	CADE-8	DACE-6	EACD-8	1
ABCDE-7998	BACDE-7711	CABDE-6535	DABCE-6126	3048
B-1202			1202	
BC-61	CB-249		27	
BD-33	DB-71		5	
BE-2	EB-22		2	
BCD-38	CBD-366	DBC-182	2	
BCE-0	CBE-5	EBC-9	1	
BDE-1	DBE-2	EBD-7	1	
BCDE-5	CBDE-18	DBCE-9	EBCD-22	3
C-5568			5568	
CD-904	DC-659		169	
CE-22	EC-40		2	
CDE-44	DCE-31	ECD-37	5	
D-2539			2539	
DE-29	ED-53		20	
E-1324			1324	

elements (a Gag protein, retrovirus-related Pol polyprotein from transposon TNT 1–94 and a reverse transcriptase protein). The remaining genes have similarity to a mix of toxins (volvatxin A2 precursor, a dipterans toxic crystal protein); heme binding (a Hemopexin domain-containing protein); cell adhesion (a calcium-binding tyrosine phosphorylation-regulated protein); pectin degradation (a probable pectin lyase precursor); possible oxidoreductases (two FAD NAD-binding domain-containing proteins); a putative CNVH-domain-containing protein and two capsule polysaccharide biosynthesis proteins. The ABCDE intersection contains 3,048 genes (data not shown), which are common to all of these basidiomycetes, while there are 34 genes associated with ABE and 70 with ABC. The *Moniliophthora* intersect (AB) has 1,287 genes, of which

**Table 3 *M. roreri* intersect**

A intersect	
Gene Id	Gene annotation
evm.model.sctg_0272_0002.2	Calcium-binding tyrosine phosphorylation-regulated
evm.model.sctg_0180_0002.8	Capsule polysaccharide biosynthesis
evm.model.sctg_0180_0002.6	Capsule polysaccharide biosynthesis
evm.model.sctg_0121_0002.16	CNVH-domain-containing protein
evm.model.sctg_0631_0001.1	Dipterans toxic crystal protein
evm.model.sctg_0013_0001.105	FAD NAD-binding domain-containing protein
evm.model.sctg_0013_0001.104	FAD NAD-binding domain-containing protein
evm.model.sctg_1050_0001.1	Gag protein
evm.model.sctg_0048_0006.5	Hemopexin domain-containing protein
evm.model.sctg_0344_0002.10	P-loop containing nucleoside triphosphate hydrolase protein
evm.model.sctg_0066_0001.15	Probable pectin lyase precursor
evm.model.sctg_0665_0003.1	Putative RecQ helicase
evm.model.sctg_0517_0002.1	Retrovirus-related pol polyprotein from transposon tnt 1-94
evm.model.sctg_1099_0001.3	Reverse transcriptase
evm.model.sctg_0028_0009.5	RNA polymerase Rpb1 C-terminal repeat domain-containing protein
evm.model.sctg_1043_0001.1	Volvatxin A2 precursor

only 106 or 8% have a putative function (Table 4). Among the AB genes there are five antibiotic biosynthesis monooxygenases; five feruloyl esterases involved in hemicellulose degradation; four cellulose binding proteins; four cell wall glycosyl hydrolases, involved in cell wall sugar degradation; four chitin binding proteins; four het domain proteins associated with fungal vegetative incompatibility and mycelial cell death; three pectate lyases involved in pectin degradation; three family 12 carbohydrate esterases, a family that includes pectin acetyl esterases, rhamnogalacturonan acetyl esterases and acetyl xylan esterases; three biotrophy-associated secreted proteins that are putative effectors and 2 amidohydrolases, enzymes that act on amide bonds. Most of these AB intersect genes were constitutively expressed, but RNA-Seq data also showed differential gene expression for 14 genes, which are indicated in Table 4 (Additional file 1). In the necrotrophic phase the following genes were upregulated: an antibiotic biosynthesis monooxygenase, two biotrophy-associated secreted proteins, a cellulose binding protein, a FAD binding domain

**Table 4 Moniliophthora hemibiotrophic intersect**

**AB intersect moniliophthora**

Gene Id	Gene annotation
evm.model.sctg_0025_0003.30	3-carboxymuconate cyclase
evm.model.sctg_0188_0002.6	Acetamidase formamidase
evm.model.sctg_0009_0002.120	Alpha beta hydrolase fold protein
evm.model.sctg_0055_0002.11	Amidohydrolase
evm.model.sctg_0260_0001.15	Amidohydrolase
evm.model.sctg_0053_0001.15	Antibiotic biosynthesis monooxygenase
evm.model.sctg_0072_0001.2	Antibiotic biosynthesis monooxygenase
evm.model.sctg_0210_0002.9 <sup>N</sup>	Antibiotic biosynthesis monooxygenase <sup>N</sup>
evm.model.sctg_0108_0002.35	Antibiotic biosynthesis monooxygenase domain protein
evm.model.sctg_0150_0003.5	Arabinofuranosidase
evm.model.sctg_0211_0001.12	Aryl sulfotransferase
evm.model.sctg_0005_0005.16 <sup>N</sup>	Biotrophy-associated secreted protein 2 <sup>N</sup>
evm.model.sctg_0066_0001.46 <sup>N</sup>	Biotrophy-associated secreted protein 2 <sup>N</sup>
evm.model.sctg_0104_0004.11	Biotrophy-associated secreted protein 2
evm.model.sctg_0099_0003.12	C6 zinc finger domain-containing protein
evm.model.sctg_0024_0001.46	Carbohydrate esterase family 12 protein
evm.model.sctg_0025_0003.31	Carbohydrate esterase family 12 protein
evm.model.sctg_0086_0003.12	Carbohydrate esterase family 12 protein
evm.model.sctg_0066_0001.40	Carboxyphosphoenolpyruvate phosphonmutase-like protein
evm.model.sctg_0063_0003.8	Cell wall glycosyl hydrolase
evm.model.sctg_0155_0002.9	Cell wall glycosyl hydrolase
evm.model.sctg_0155_0004.12	Cell wall glycosyl hydrolase
evm.model.sctg_0155_0004.13 <sup>B</sup>	Cell wall glycosyl hydrolase <sup>B</sup>
evm.model.sctg_0245_0001.21	Cellulose-binding family ii
evm.model.sctg_0176_0001.14 <sup>N</sup>	Cellulose-binding protein <sup>N</sup>
evm.model.sctg_0243_0004.9	Cellulose-binding protein
evm.model.sctg_0065_0001.23	Chitin binding
evm.model.sctg_0068_0003.6	Chitin binding
evm.model.sctg_0086_0003.11	Chitin binding
evm.model.sctg_0086_0003.19	Chitin binding
evm.model.sctg_0149_0001.3	Cysteine-rich secreted protein
evm.model.sctg_0011_0002.81	duf1446 domain containing protein
evm.model.sctg_0073_0003.23	duf1446 domain containing protein
evm.model.sctg_0091_0003.16	duf567 domain protein
evm.model.sctg_0008_0001.9	duf718 domain protein
evm.model.sctg_0119_0001.2	duf985 domain protein
evm.model.sctg_0121_0002.8	endo-β-xylanase precursor
evm.model.sctg_0132_0002.1	Epoxidase subunit a
evm.model.sctg_0072_0002.6	Erylsin b
evm.model.sctg_0055_0003.16	Ethyl tert-butyl ether degradation

**Table 4 Moniliophthora hemibiotrophic intersect**

**(Continued)**

evm.model.sctg_0042_0004.3	Excitatory amino acid transporter 2
evm.model.sctg_0050_0004.25	Exonuclease
evm.model.sctg_0022_0002.37	Extracellular chitosanase
evm.model.sctg_0054_0001.12 <sup>N</sup>	FAD binding domain protein <sup>N</sup>
evm.model.sctg_0133_0001.4	Feruloyl
evm.model.sctg_0015_0003.16	Fructose-bisphosphate aldolase
evm.model.sctg_0101_0001.1	Fusicoccadiene synthase
evm.model.sctg_0206_0002.8	General substrate transporter-like protein
evm.model.sctg_0214_0001.8	Glucokinase regulator family
evm.model.sctg_0011_0002.91 <sup>B</sup>	Glycoside hydrolase family 16 protein <sup>B</sup>
evm.model.sctg_0011_0002.92	Glycoside hydrolase family 16 protein
evm.model.sctg_0091_0003.6	Glycoside hydrolase family 29 protein
evm.model.sctg_0120_0002.15	Glycoside hydrolase family 29 protein
evm.model.sctg_0195_0003.1	Glycoside hydrolase family 78 protein
evm.model.sctg_0063_0003.36	Glycosyl family
evm.model.sctg_0002_0008.34	Glycosyl hydrolase family 32
evm.model.sctg_0120_0001.8	Het domain containing protein
evm.model.sctg_0018_0003.8	Het domain protein
evm.model.sctg_0068_0003.29	Het domain protein
evm.model.sctg_0175_0002.5	Het domain protein
evm.model.sctg_0014_0004.20	Hydantoinase
evm.model.sctg_0008_0001.24	Hydroxycinnamoyl shikimate quinate hydroxycinnamoyltransferase
evm.model.sctg_0091_0003.1 <sup>N</sup>	Integral membrane protein <sup>N</sup>
evm.model.sctg_0092_0002.20	Integral membrane protein
evm.model.sctg_0092_0002.22	Integral membrane protein
evm.model.sctg_0097_0001.11	Integral membrane protein
evm.model.sctg_0002_0010.51	Isochorismatase hydrolase
evm.model.sctg_0231_0001.20	Isoflavone reductase family protein
evm.model.sctg_0015_0001.8	Killer kp4 toxin
evm.model.sctg_0088_0001.8 <sup>N</sup>	Lactam utilization protein <sup>N</sup>
evm.model.sctg_0023_0001.50	Lea domain protein
evm.model.sctg_0178_0001.5 <sup>N</sup>	Lea domain-containing protein <sup>N</sup>
evm.model.sctg_0005_0005.46	Lipoprotein
evm.model.sctg_0049_0006.14	L-lysine-2,3-aminomutase
evm.model.sctg_0002_0010.48	Major facilitator superfamily transporter
evm.model.sctg_0240_0001.1	Major royal jelly protein
evm.model.sctg_0023_0002.33	mannitol-1-phosphate 5-dehydrogenase
evm.model.sctg_0158_0002.2	Metal dependent phosphohydrolase
evm.model.sctg_0081_0002.17 <sup>B</sup>	NAD-binding phosphogluconate dehydrogenase-like protein <sup>B</sup>
evm.model.sctg_0110_0001.5	NAD-binding phosphogluconate dehydrogenase-like protein
evm.model.sctg_0062_0002.32	nmra-like protein
evm.model.sctg_0026_0001.19	Pectate lyase
evm.model.sctg_0214_0001.16	Pectate lyase

**Table 4 Moniliophthora hemibiotrophic intersect (Continued)**

evm.model.sctg_0049_0006.8	Phosphoglycerate mutase family
evm.model.sctg_0049_0006.9	Phosphoglycerate mutase family
evm.model.sctg_0194_0001.15	Plasmid p 4b orf-3 family protein
evm.model.sctg_0094_0005.7	P-loop containing nucleoside triphosphate hydrolase protein
evm.model.sctg_0014_0004.25	Poxa3b laccase small subunit
evm.model.sctg_0162_0004.8	Proline racemase
evm.model.sctg_0018_0002.9	Proline-specific peptidase
evm.model.sctg_0089_0002.36	Protein tprxl
evm.model.sctg_0122_0004.1	Proteophosphoglycan ppg4
evm.model.sctg_0159_0004.4	Purine nucleoside
evm.model.sctg_0022_0002.64	Putative zinc metallopeptidase protein
evm.model.sctg_0100_0003.5	Related to tol protein
evm.model.sctg_0020_0001.61	Serine threonine sps1
evm.model.sctg_0217_0001.20	Surface cell-adhesion protein
evm.model.sctg_0105_0006.2 <sup>N</sup>	Thaumatococcus-like protein <sup>N</sup>
evm.model.sctg_0100_0003.7	Tol-like protein
evm.model.sctg_0098_0001.12	Transcription regulator
evm.model.sctg_0093_0001.28 <sup>N</sup>	Transferase family protein <sup>N</sup>
evm.model.sctg_0016_0005.17	Transposon en spm sub-class
evm.model.sctg_0188_0001.3	Twin-arginine translocation pathway signal
evm.model.sctg_0041_0004.5 <sup>N</sup>	Urea hydro-lyase cyanamide <sup>N</sup>
evm.model.sctg_0079_0008.3	Variable surface lipoprotein d1
evm.model.sctg_0122_0001.3	ww rsp5 wwp

<sup>N</sup> indicates expression during the necrotrophic phase and <sup>B</sup> indicates expression during the biotrophic phase

protein, an integral membrane protein, a lactam utilization protein, a LEA domain-containing protein, a thaumatococcus-like protein, a transferase family protein and a urea hydro-lyase cyanamide protein. In the biotrophic phase a cell wall glycosyl hydrolase, a family 16 glycoside hydrolase and a NAD binding phosphogluconate dehydrogenase-like protein were upregulated (Additional file 1, AB intersect sheet). The pathogenic intersect (ABE) has 30 genes with known functions and these are shown in Table 5. The biotrophic intersect (ABC) has 17 genes with putative functions (Table 6). RNA-Seq data revealed that most of ABC and ABE intersect genes are constitutively expressed (Additional file 1) at low to mid levels (adjusted means at 2 to 50 reads = Low level; 50 to 125 reads = mid level and >125 = high level of gene expression), with the exception of two genes from the ABC intersect, (a transcriptional family alpha/beta fold family protein and one hypothetical protein) that were differentially expressed in the necrotrophic phase (Additional file 1, ABE and ABC intersect sheets).

**Table 5 Pathogenic intersect**

**ABE intersect pathogenic fungi**

Gene Id	Gene annotation
evm.model.sctg_0099_0004.5	3-dehydroquinate dehydratase
evm.model.sctg_0005_0001.15	4-carboxymuconolactone decarboxylase
evm.model.sctg_0024_0002.8	4-hydroxyphenylpyruvate dioxygenase
evm.model.sctg_0220_0007.7	Acetyltransferase
evm.model.sctg_0108_0002.26	Arylsulfatase
evm.model.sctg_0002_0010.39	Cupin domain protein
evm.model.sctg_0004_0002.40	Cytochrome c oxidase assembly protein
evm.model.sctg_0029_0001.27	Extracellular invertase
evm.model.sctg_0040_0001.46	Flavin-containing amine oxidase dehydrogenase
evm.model.sctg_0006_0005.98	Formate nitrite transporter
evm.model.sctg_0084_0003.26	GAF domain nucleotide-binding protein
evm.model.sctg_0289_0002.8	Glycoside hydrolase family 45 protein
evm.model.sctg_0089_0002.33	gnat family acetyltransferase
evm.model.sctg_0185_0001.1	gpi-anchored protein
evm.model.sctg_0040_0003.2	Halooxid dehalogenase type ii
evm.model.sctg_0346_0003.4	Tubulin
evm.model.sctg_0011_0002.41	Iron permease ftr1
evm.model.sctg_0309_0002.8	Minor histocompatibility antigen h13
evm.model.sctg_0002_0009.45	Nitrilase 4
evm.model.sctg_0018_0008.47	Nucleoside-diphosphate-sugar
evm.model.sctg_0018_0007.18	Phytanoyl-CoA dioxygenase family protein
evm.model.sctg_0042_0011.14	Related to transposase
evm.model.sctg_0046_0002.12	rpel repeat protein
evm.model.sctg_0073_0003.11	Secreted hydrolase
evm.model.sctg_0354_0001.6	sirg protein
evm.model.sctg_0020_0001.23	spherulin 4-like cell surface protein
evm.model.sctg_0162_0003.2	Taurine catabolism dioxygenase
evm.model.sctg_0105_0006.1	Thaumatococcus-like protein 1-like
evm.model.sctg_0006_0005.105	Transcriptional regulator
evm.model.sctg_0138_0001.17	upf0132 domain protein

**Identification of secreted proteins**

Because of the hemibiotrophic nature of *M. roreri* and the role that secreted proteins are expected to play in that interaction, we analyzed the genome for genes that express secreted proteins. The first screening identified all of the putative proteins with signal peptides. In *M. roreri* 1,752 proteins (9.7%) were found to have signal peptides, which is slightly lower than *M. perniciosus* which has 1,810 proteins with signal peptides. Next the proteins with signal peptides were analyzed for transmembrane domains, to remove proteins that are expected to be incorporated into the membrane of the fungus. Ninety-two of the 1,752 proteins have two or more transmembrane

**Table 6 Biotrophic intersect**

ABC intersect biotrophic fungi	
Gene Id	Gene annotation
evm.model.sctg_0063_0001.9	alpha- -mannosylglycoprotein 6-beta-n-acetylglucosaminyltransferase a
evm.model.sctg_0124_0001.10	Aminoglycoside phosphotransferase
evm.model.sctg_0007_0001.162	Dolichol phosphate-mannose biosynthesis regulatory protein
evm.model.sctg_0112_0002.7	dolichyl-phosphate mannosyltransferase polypeptide 3
evm.model.sctg_0040_0002.2	gas1-like protein
evm.model.sctg_0016_0002.4	Glycoside hydrolase family 61 protein
evm.model.sctg_0063_0002.2	Glycoside hydrolase family 95 protein
evm.model.sctg_0001_0002.138	Hydantoinase
evm.model.sctg_0919_0001.1	Hypothetical proline-rich protein
evm.model.sctg_0166_0004.8	Nuclear fusion protein kar5
evm.model.sctg_0041_0003.8	Proteophosphoglycan 5
evm.model.sctg_0231_0001.2	Response regulator receiver domain-containing protein
evm.model.sctg_0229_0001.10	Ribosomal protein l36 containing protein
evm.model.sctg_0243_0004.11	Transcription factor iib
evm.model.sctg_0005_0002.50	Transcriptional family alpha beta fold family protein
evm.model.sctg_0096_0003.37	Transmembrane protein 167a
evm.model.sctg_0003_0004.13	Vacuolar h + -atpase assembly protein

domains. One hundred and twenty-five glycosylphosphatidylinositol (GPI) anchor proteins were also excluded from the list of secreted proteins (Table 7). While the GPI-anchored proteins are normally removed from secreted proteins, the posttranslational modification of glypiation is the only method of attachment to the membrane for these proteins, and because the bonds are subject to phospholipase cleavage, these proteins could still be released into the extracellular space via an enzymatic release process. After removal of proteins with transmembrane and GPI moieties the total number of *M. royeri* secreted proteins was 1,535, which is less than the 1,596 secreted proteins predicted for *M. perniciosa*.

#### Expression of the secretome

In this study we wanted to verify the expression of the predicted genes, including the secretome identified in the genome analysis. We utilized a replicate set of infected

pod tissues from a previous study to obtain transcriptomes representing the biotrophic and necrotrophic phases of the disease [21]. While our present focus is on the secreted proteins, the complete fungal transcriptome, for FPR, in different resistant plant backgrounds will be addressed in a separate treatise. The planting material used represents a segregating F1 population of progeny from highly susceptible clones (PA16 × SIC433 and PA16 × EEG29). Briefly, the pods used in this study were derived from flowers hand-pollinated 60 days prior to infection with a spore suspension of  $1 \times 10^8$  *M. royeri* spores ml<sup>-1</sup>. The infected pods were frozen in liquid nitrogen and lyophilized at 30 days post infection (DPI) and at 60 DPI. Based on typical progression of FPR symptoms [21], three malformed green pods were selected for RNA-Seq and RT-qPCR from 30 DPI pods (biotrophic phase) and three necrotic sporulating pods were selected for RNA-Seq and RT-qPCR from 60 DPI pods (necrotrophic phase).

The 30 DPI pod samples yielded 67,818,927; 64,099,928 and 64,033,574 reads and the number of fungal reads per sample were 149,675 (0.22%), 122,289 (0.19%) and 312,684 (0.49%), respectively. For the 60 DPI pod samples 73,037,363; 66,395,887 and 69,434,845 reads were sequenced, with fungal reads accounting for 603,498 (0.83%), 882,821 (1.33%) and 503,459 (0.73%), respectively. When all of the libraries were compared 13,759 genes (77% of the predicted CDS) from the *M. royeri* genome were found to be expressed in these infected pods. The secreted proteins were then separated from the other expressed genes. Among the 1,535 secreted protein genes identified 1,355 or 88%, were expressed at some level under the tested conditions. Using a P-value of 0.05, we found that 222 of the secreted protein genes were differentially expressed between the 30 DPI and 60 DPI pods. All of the expressed secreted protein genes are listed in Additional file 2. The data revealed phase-dependent differential gene expression, where 164 of the genes were up-regulated in the 60 DPI pods while 58 were upregulated in the 30 DPI pods. Since the hypothetical genes can provided no information regarding their putative function they were eliminated from further consideration with regards to their role in the disease process.

Genes expressed during the early stages of infection and disease development reflect gene products that function in the biotrophic phase of the disease development, which can be initially asymptomatic. Late-stage gene expression

**Table 7 Results of the genome based secretome analysis**

	total # of predicted proteins	Protein without signal peptides	Proteins with signal peptides	Proteins with more than 2 transmembrane domains	Proteins with GPI domains	Total number of secreted proteins Minus TMD and GPI domains	% proteins with signal peptides	Unique signal peptides
<i>M. royeri</i>	17920	16167	1752	92	125	1535	9.7	39
<i>M. perniciosa</i>	17016	15206	1810	92	122	1596	10.6	58

yields gene products that function in the necrotrophic phase, which is associated with pod chlorosis and necrosis, and possibly with sporulation of the fungus on pod tissue. To validate the differential expression of genes encoding secreted proteins as suggested by the RNA-Seq analysis, RT-qPCR was conducted, with a selected number of genes comparing expression in a set of seven malformed green pods at 30 DPI and seven necrotic pods at 60 DPI. At 30 DPI, 34 genes encoding secreted proteins with putative functions were expressed, with one being expressed only at this stage (Table 8). At 60 DPI, 105 genes encoding predicted secreted proteins with putative functions were expressed (Table 9). Forty-one of the 105 60 DPI genes were detected only in the necrotrophic phase. RT-qPCR confirmed the up-regulation of the 30 DPI and 60 DPI genes (Tables 8 and 9). Although the GPI-anchored proteins were removed from the secretome we did find a number of them highly expressed in both the biotrophic and necrotrophic phases. Table 10 provides a list of the most highly expressed genes encoding GPI-anchored proteins in both phases since controlled release of these proteins from the fungal membrane is possible.

Several gene families of secreted proteins were prominently upregulated in the biotrophic phase. In this phase the 34 genes with known functions were primarily associated with nutrient acquisition and with plant and fungal cell wall modification. The largest single group identified was glycoside hydrolases (GH). Eleven glycoside hydrolases genes were upregulated: two genes from GH family 5; five genes from GH family 18; and one gene each from the GH families 16, 28, 43, 92. The GH family 18 genes are homologous to chitinase genes (EC 3.2.1.14) [22] and GH family 5 genes are endo-enzymes that are capable of hydrolyzing both  $\beta$ -mannans and  $\beta$ -glucans [23] and are homologous to beta-1,3-glucanase, which includes endoglucanase, beta-mannanase, exo-1,3-glucanase, endo-1,6-glucanase, xylanase, and endoglycoceramidase. The GH 16 s are homologous to endobeta 1,3-glucanases (EC 3.2.1.39) and xyloglucan:xyloglucosyltransferase (EC 2.4.1.207) [24], GH 28 s are homologous to polygalacturonases and includes pectin degrading enzymes like polygalacturonase (EC 3.2.1.15), exopolygalacturonase (EPG; EC 3.2.1.67), exopoly- $\alpha$ -galacturonosidase (EC3.2.1.82) [25], GH 43 s are homologous to beta-xylosidases (EC 3.2.1.37) [26], while the GH 92 s are alpha-1,2-mannosidases [27] (Additional file 2). In addition to the GHs, other enzymes were identified such as galactose-1-epimerase, a carbonic anhydrase, a family 8 carbohydrate esterase (pectin methyltransferase), and a tryptophan dimethylallyl transferase. The plant cell wall-modifying enzymes identified in the biotrophic phase include an endo-polygalacturonase, a galactan 1,3-beta-galactosidase, a cell wall glycosyl hydrolase, a catalase, an acetylxylan esterase and a xylanase A. Additionally, three cerato-platanin genes, two hydrophobins, three PR-1-like

proteins, and an immunomodulatory protein were all induced in the biotrophic phase that could be involved in the fungal host interaction (Table 8).

Based on RNA-Seq and verified by RT-qPCR analysis, the majority of the genes within the 105 genes expressed during the necrotrophic phase are associated with fungal growth/pathogenicity and nutrient acquisition. Several genes families were identified with multiple expressed genes such as hydrophobins (10 genes), FAD binding domain containing proteins (6 genes), aryl-alcohol oxidases (4 genes), laccases (4 genes), extracellular triacylglycerol lipases (3 genes) and cerato-platanins (3 genes). Several of these highly upregulated genes are associated with lignin breakdown. The aryl-alcohol oxidases (AAOs) are FAD containing extracellular enzymes in the glucose-methanol choline oxidase (GMC) family [28]. AAOs are involved in lignin degradation where the enzyme interacts with p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol polymers that make up the lignin, and  $O_2$  yielding aromatic aldehydes and hydrogen peroxide [28]. Laccases are polyphenol oxidases that reduce phenolic compounds in lignin [29]. Glyoxal oxidases are extracellular enzymes that form hydrogen peroxide during the breakdown of lignin [30]. Eight GH families were identified among the 12 GH genes (two GH family 35; two GH family 61; two GH family 5 and one gene each for families 3, 9, 10, 16 and 18). The GH 35 s proteins are homologous to beta-galactosidases and GH 61 s proteins are homologous to endoglucanases. The GH 3 s proteins are homologous to beta-glucosidases; GH 9 s proteins are endoglucanases or cellulases; and GH 10s proteins are xylanases [31]. Numerous other enzymes and pathogenicity-related genes were identified such as carboxypeptidases, carbohydrate esterases, tripeptidyl peptidase, a necrosis inducing protein, a fungal peroxidase, cytochrome p450s, and expansins. The highest relative expressions were found for a dioxygenase, a Fbox and FNIP repeat-containing protein, a chitin synthase, a proline-rich antigen and one of the expansins (Table 9).

Several gene families were found to be differently regulated and had specific members upregulated in the biotrophic phase and others upregulated in the necrotrophic phase. Of the 41 hydrophobin genes found in the genome, 12 were differentially expressed; 10 were upregulated in the necrotrophic phase; two in the biotrophic phase. Among the eight cerato-platanin genes found in the genome six were differentially upregulated; three in each phase. The expression of genes encoding fungal-derived pathogenesis-related proteins was detected in samples from the biotrophic phase, with *MrPR-1n*, *MrPR-1g* and *MrPR-1i2* being upregulated. Among the other nine genes encoding PR-1-like proteins found in the *M. royeri* genome, one *MrPR-1d* was upregulated in the necrotrophic phase and five were constitutively expressed under the tested conditions (Additional file 2).

**Table 8 Biotrophic gene Expression of putative Secreted Proteins**

Transcript Id	Gene annotation	RNAseq				Primers	QPCR			
		30DPI	60DPI	Pvalue	Fold change		Induced 30-60 DPI		Repressed 30-60 DPI	
		Mean*	Mean*				Mean	St. E.	Mean	St. E.
e.m.s_0369_0002.2	Acetylxyylan esterase	638.4	91.6	5.8E-03	7.0					
e.m.s_0461_0001.1	Carbohydrate esterase family 8 protein	512.1	67.0	8.8E-04	7.6	MrCE8a	20.868	8.9736		
e.m.s_0215_0001.21	Carbonic anhydrase	16.3	0.4	2.2E-02						
e.m.s_0250_0003.8	Catalase cat1	294.9	59.3	6.4E-03	5.0	MrCAT1	114.906	48.407		
e.m.s_0155_0004.13	Cell wall glycosyl hydrolase	330.4	26.3	7.4E-05	12.6					
e.m.s_0192_0001.1	Cerato-platanin	3014.0	520.7	5.3E-03	5.8					
e.m.s_0192_0003.14	Cerato-platanin	4931.3	968.5	7.2E-03	5.1					
e.m.s_0192_0003.17	Cerato-platanin	7602.6	912.5	7.1E-04	8.3					
e.m.s_0258_0001.1	Defense-related protein scp domain-containing protein MrPR-1 g	16778.8	1200.4	5.6E-04	14.0	MrSCP	434.018	253.661		
e.m.s_0251_0004.1	Defense-related protein scp domain-containing protein MrPR-1i2	32.6	2.4	7.4E-02						
e.m.s_0516_0001.1	Defense-related protein scp domain-containing protein Mr-PR-1n	499.0	13.7	2.6E-06	38.3					
e.m.s_0041_0005.1	Endo-polygalacturonase	2621.4	32.4	1.4E-06	81.9	MrendoPG	911.453	592.910		
e.m.s_0407_0004.2	Galactan 1,3-beta-galactosidase	37.6	52.2	2.4E-02						
e.m.s_0027_0002.19	Galactose 1-epimerase	154.2	47.4	3.9E-02	3.3					
e.m.s_0011_0002.91	Glycoside hydrolase family 16 protein	1276.8	232.0	5.0E-03	5.5					
e.m.s_0030_0003.33	Glycoside hydrolase family 18 protein	6125.5	948.8	3.4E-03	6.5	MrGH18g	38.900	16.126		
e.m.s_0044_0001.33	Glycoside hydrolase family 18 protein	527.3	56.3	7.7E-04	9.4	MrGH18d	253.832	183.817		
e.m.s_0044_0001.6	Glycoside hydrolase family 18 protein	12287.0	3336.6	2.1E-02	3.7					
e.m.s_0114_0002.6	Glycoside hydrolase family 18 protein	12684.6	687.6	5.1E-06	18.5	MrGH18c	151.387	51.810		
e.m.s_0416_0001.5	Glycoside hydrolase family 18 protein	4897.1	266.4	3.1E-06	18.4					
e.m.s_0055_0003.9	Glycoside hydrolase family 28 protein	129.2	20.9	1.3E-02	6.1					
e.m.s_0067_0006.17	Glycoside hydrolase family 43	234.2	54.4	2.2E-02	4.3					
e.m.s_0043_0001.2	Glycoside hydrolase family 5 protein	3971.1	188.7	1.2E-06	21.1					
e.m.s_0531_0001.4	Glycoside hydrolase family 5 protein	3387.9	716.4	1.2E-02	4.7	MrGH5	69.7456	41.581		
e.m.s_0351_0003.10	Glycoside hydrolase family 92 protein	2935.9	894.7	4.6E-02	3.3					
e.m.s_0346_0003.2	Hydrophobin 3	431.5	65.6	1.8E-03	6.6	MrSc3P3	345.0789	196.349		
e.m.s_0346_0003.1	Hydrophobin sc3-like	6461.8	471.4	3.8E-05	13.7	MrSc3P2	356.391	177.576		
e.m.s_0409_0005.4	Hypothetical fad dependent oxidoreductase	566.2	119.8	1.4E-02	4.7					
e.m.s_0561_0002.1	Immunomodulatory protein	7180.5	746.5	3.2E-04	9.6					
e.m.s_0095_0005.5	Major facilitator superfamily multidrug-resistance	7.0	0.0	1.3E-02						
e.m.s_0053_0004.6	Membrane autotransporter barrel domain protein	755.7	203.5	2.5E-02	3.7	MrBDP	5.184	1.528		
e.m.s_0094_0001.8	O-methyltransferase	1131.7	488.6	8.2E-04						
e.m.s_0064_0001.5	Tryptophan dimethylallyltransferase	190.8	0.9	3.2E-04	211.1	MrTRY-DMA	235.864	134.155		
e.m.s_0039_0002.43	WSC domain protein	356.0	102.1	4.1E-02	3.5					
e.m.s_0478_0002.2	Xylanase A	1135.8	9.0	9.3E-04	126.0	MrXYLa	171.086	77.159		

Mean\* is the normalized mean based on the size factor of the libraries.

**Table 9 Necrotrophic genes expression of putative Secreted proteins**

Transcript Id	Gene annotation	RNAseq				Primers	QPCR			
		30DPI	60DPI	Pvalue	Fold change		Induced 30-60 DPI		Repressed 30-60 DPI	
		Mean*	Mean*				Mean	St. E.	Mean	St. E.
e.m.s_0228_0004.4	Alpha beta hydrolase	3.2	34.5	1.4E-02	10.8					
e.m.s_0245_0001.13	Alpha1,3-gluconase/mutanase	0.7	171.8	8.3E-03	245.4					
e.m.s_0185_0001.10	Aryl-alcohol oxidase	0.0	21.4	5.7E-04	21.4					
e.m.s_0185_0001.9	Aryl-alcohol oxidase	0.0	7.8	4.2E-02	7.8					
e.m.s_0274_0001.9	Aryl-alcohol oxidase	0.0	9.6	2.1E-02	9.6					
e.m.s_0491_0001.5	Aryl-alcohol oxidase	3.9	61.7	3.0E-02	15.9					
e.m.s_0100_0003.8	Arylsulfatase	2.6	42.4	2.9E-02	16.4					
e.m.s_0251_0003.2	Aspartic peptidase a1	11.5	46.7	3.6E-02	4					
e.m.s_0056_0003.12	Aspartic-type endopeptidase	0.7	25.6	2.5E-03	36.6					
e.m.s_0463_0001.8	Beta-lactamase class penicillin binding protein	6.6	324.6	1.5E-03	492.4					
e.m.s_0066_0001.46	Biotrophy-associated secreted protein 2 MrBAS1	7.6	496.7	1.0E-03	66.1	MrSP2a		26.509	15.802	
e.m.s_0005_0005.16	Biotrophy-associated secreted protein 2 MrBAS2	2.0	613.6	3.2E-04	322.6					
e.m.s_0227_0001.20	Calcineurin-like phosphoesterase	1.8	19.0	9.5E-03	10.6					
e.m.s_0066_0001.21	Carbohydrate esterase family 4 protein	0.7	83.8	1.8E-02	119.7					
e.m.s_0155_0002.4	Carbohydrate esterase family 4 protein	0.0	50.5	2.1E-06	50.5					
e.m.s_0008_0001.25	Carbonic anhydrase	0.7	121.3	7.0E-08	172.8					
e.m.s_0190_0004.1	Carboxypeptidase cpds	3.8	460.0	4.4E-07	121.1					
e.m.s_0190_0004.2	Carboxypeptidase cpds	3.1	28.5	5.1E-03	9.2					
e.m.s_0176_0001.14	Cellulose-binding protein	0.0	22.2	4.7E-04	22.2					
e.m.s_0058_0002.19	Cerato platanin	0.0	16.2	3.2E-03	16.2					
e.m.s_0058_0002.21	Cerato-platanin	2.1	36.2	2.7E-02	18.1					
e.m.s_0192_0001.2	Cerato-platanin	0.0	20.7	7.5E-04	20.7					
e.m.s_0192_0001.6	Cerato-platanin	0.0	494.9	2.2E-04	494.9					
e.m.s_0297_0002.3	Chitin deacetylase family 4	0.0	40.4	2.8E-02	40.5	MrCHIDACTb		5663.095	3460.583	
e.m.s_0245_0001.24	Chitin synthase	0.0	2158.9	2.8E-04	2159	MrCHS		2515.177	1862.483	
e.m.s_0008_0004.19	Copper amine oxidase	4.3	77.0	6.6E-03	17.9					
e.m.s_0003_0001.85	Cutinase	0.0	9.5	2.3E-02	9.5					
e.m.s_0017_0003.61	Cytochrome p450	0.0	32.2	9.1E-03	32.2					
e.m.s_0052_0004.10	Cytochrome p450	0.7	14.9	3.2E-02	21.1					
e.m.s_0470_0001.6	Cytochrome p450 monooxygenase	5.0	91.2	1.1E-04	18.3	MrCp450a		30.050	11.8905	
e.m.s_0059_0001.14	Deuterolysin m35 metalloprotease	2.8	137.0	5.0E-02	48.9					
e.m.s_0099_0003.13	Dioxygenase family protein	8.3	5904.3	3.6E-02	715.6					

**Table 9 Necrotrophic genes expression of putative Secreted proteins (Continued)**

e.m.s_0061_0002.41	Exo-beta 1,3-glucanase	2.6	16.7	4.4E-02	6.4			
e.m.s_0026_0005.53	Expansin family protein	0.0	1165.0	9.8E-04	1165	MrEXP-A	326.810	159.020
e.m.s_0086_0003.2	Expansin family protein	1.8	37.3	2.6E-04	20.7			
e.m.s_0280_0002.5	Extracellular dioxygenase	0.0	15.1	3.4E-03	15.1			
e.m.s_0107_0002.16	Extracellular triacylglycerol lipase precursor	0.0	9.8	2.0E-02	9.8			
e.m.s_0107_0002.18	Extracellular triacylglycerol lipase precursor	0.0	9.8	2.3E-02	9.8			
e.m.s_0352_0003.3	Extracellular triacylglycerol lipase precursor	0.0	9.7	2.1E-02	9.7			
e.m.s_0054_0001.12	FAD binding domain protein	31.7	306.9	2.3E-04	9.7			
e.m.s_0135_0001.27	FAD binding domain-containing protein	0.0	43.9	6.4E-06	43.9			
e.m.s_0277_0002.12	FAD binding domain-containing protein	6.5	91.2	1.1E-04	14.1			
e.m.s_0277_0002.13	FAD binding domain-containing protein	2.6	124.8	1.1E-04	48			
e.m.s_0375_0001.9	FAD binding domain-containing protein	2.0	85.8	9.5E-06	43.1			
e.m.s_0459_0001.6	FAD binding domain-containing protein	0.0	36.5	3.2E-04	36.5			
e.m.s_0469_0001.4	FAD binding domain-containing protein	0.0	8.4	3.7E-02	8.4			
e.m.s_0002_0010.20	F-box and fnip repeat-containing protein	114.0	3336.6	2.1E-02	29.3			
e.m.s_0087_0002.1	Fruit-body specific gene a	0.7	20.9	7.7E-03	29.8			
e.m.s_0053_0004.14	Fungal peroxidase	3.6	22.3	1.5E-02	6.3			
e.m.s_0111_0006.5	Glucosylglycosaminidase	0.0	27.8	8.3E-03	27.8			
e.m.s_0211_0001.14	Glucose oxidase	0.0	66.0	3.6E-07	66			
e.m.s_0085_0002.37	Glycoside hydrolase family 16 protein	0.0	30.0	7.7E-05	29			
e.m.s_0021_0002.67	Glycoside hydrolase family 18 protein	19.6	87.3	5.0E-02	4.5			
e.m.s_0261_0004.2	Glycoside hydrolase family 3 protein	25.4	127.4	1.7E-02	5			
e.m.s_0008_0004.3	Glycoside hydrolase family 30 protein	11.7	67.7	3.7E-02	6			
e.m.s_0232_0001.2	Glycoside hydrolase family 35 protein	33.9	128.7	4.3E-02	3.8			
e.m.s_0438_0002.3	Glycoside hydrolase family 35 protein	0.0	19.7	8.7E-04	19.7			
e.m.s_0007_0001.127	Glycoside hydrolase family 5 protein	27.1	163.4	5.6E-03	6			
e.m.s_0265_0002.8	Glycoside hydrolase family 5 protein	10.8	111.9	5.4E-05	10.4			
e.m.s_0241_0001.11	Glycoside hydrolase family 61 protein	0.0	26.6	9.6E-03	26.6			
e.m.s_0241_0001.12	Glycoside hydrolase family 61 protein	0.0	43.1	5.9E-03	43.1			
e.m.s_0155_0002.8	Glycoside hydrolase family 61 protein	13.5	115.5	1.3E-02	8.6			
e.m.s_0004_0003.50	Glycoside hydrolase family 9 protein	10.4	63.2	8.9E-03	6.3	MrGH9	4.311	1.827
e.m.s_0200_0003.3	Glycosyl hydrolase family 10	0.7	22.6	3.6E-02	32.2			
e.m.s_0333_0001.1	Glyoxal oxidase	4.9	47.9	5.1E-03	9.8			
e.m.s_0135_0001.28	gmc oxidoreductase	14.3	52.2	2.4E-02	3.6			

**Table 9 Necrotrophic genes expression of putative Secreted proteins (Continued)**

e.m.s_0074_0003.10	Hemerythrin hhe cation binding domain-containing protein	49.9	383.1	2.1E-02	7.7			
e.m.s_0039_0002.15	Hemolysin	0.7	16.1	2.3E-02	22.9			
e.m.s_0062_0001.18	Heterokaryon incompatibility protein het-c	28.2	116.3	1.5E-02	4.1			
e.m.s_0018_0008.40	Hydrophobin	13.2	86.6	3.6E-02	6.6			
e.m.s_0018_0008.54	Hydrophobin	0.7	61.9	1.2E-05	87.1			
e.m.s_0018_0008.56	Hydrophobin	0.0	256.2	4.3E-12	256	MrLM18	1573.702	1268.272
e.m.s_0058_0002.53	Hydrophobin	0.7	25.1	2.8E-03	35.7			
e.m.s_0058_0002.55	Hydrophobin	9.0	304.7	5.8E-07	34.2			
e.m.s_0149_0001.14	Hydrophobin	0.0	11.2	1.2E-02	11.2			
e.m.s_0149_0001.16	Hydrophobin	0.0	425.5	2.3E-06	425.6	MrLM19	257.737	133.646
e.m.s_0170_0002.12	Hydrophobin	0.0	9.3	2.4E-02	9.3			
e.m.s_0298_0001.2	Hydrophobin	0.0	11.4	1.2E-02	11.4			
e.m.s_0149_0001.15	Hydrophobin 2	0.0	205.6	1.6E-04	205.6			
e.m.s_0199_0001.3	Laccase	0.0	9.6	2.6E-02	9.6			
e.m.s_0210_0002.8	Laccase	0.0	119.6	1.0E-06	119.6			
e.m.s_0246_0002.7	Laccase	1.3	41.0	1.0E-02	31.8			
e.m.s_0279_0002.3	Laccase	0.0	21.1	3.4E-02	21.2			
e.m.s_0082_0001.16	Mannoprotein	88.8	298.1	4.7E-02	3.4			
e.m.s_0030_0003.64	Metalloproteinase	18.6	306.9	8.1E-06	16.5			
e.m.s_0406_0002.4	Necrosis inducing-like protein npp1 type	9.1	658.2	1.5E-04	73.1	MrNPP1	310.949	181.890
e.m.s_0051_0003.3	nhl repeat-containing protein	0.0	28.8	1.1E-04	28.8			
e.m.s_0120_0002.7	Oxalate decarboxylase	9.3	57.9	4.8E-02	6.2			
e.m.s_0398_0002.11	Oxidoreductase fad binding	1.8	111.0	5.4E-04	61.6			
e.m.s_0639_0001.6	Para-nitrobenzyl esterase	0.0	45.8	3.2E-04	45.8			
e.m.s_0082_0001.36	Peptidase m28	3.2	57.6	3.9E-02	18			
e.m.s_0007_0001.138	Peptide-n4-(n-acetyl-beta-glucosaminy)asparagine amidase a	0.0	12.6	3.5E-02	12.5			
e.m.s_0180_0002.16	Phosphatidylserine decarboxylase	4.3	451.8	5.1E-04	105.1			
e.m.s_0432_0001.5	PR-1 protein MrPR-1d	5.8	45.1	1.4E-03	7.8			
e.m.s_0568_0001.7	Proline-rich antigen	7.7	1372.5	7.7E-03	177.3			
e.m.s_0018_0003.1	Proline-specific peptidase	29.2	94.0	5.0E-02	3.2			
e.m.s_0726_0001.2	Pyrolysin	8.1	50.9	3.5E-02	6.3			
e.m.s_0125_0001.43	Riboflavin aldehyde-forming enzyme	61.2	229.9	3.8E-02	3.8			
e.m.s_0040_0001.30	Ribonuclease t1	0.0	11.4	1.1E-02	11.4			
e.m.s_0321_0001.8	Serine-rich protein	16.4	289.5	5.2E-04	18.1	MrSRP	14.072	5.613

**Table 9 Necrotrophic genes expression of putative Secreted proteins (Continued)**

e.m.s_0005_0005.47	Serine-type endopeptidase	224.0	811.1	4.1E-02	3.6	MrS-endo	3.482	0.866
e.m.s_0021_0002.36	Tripeptidyl peptidase A	0.0	20.8	6.5E-04	20.8			
e.m.s_0166_0004.10	Tripeptidyl peptidase A	1.8	39.5	1.8E-04	21.9			
e.m.s_0048_0003.5	Virus P4 KP4 toxin	0.0	118.9	4.5E-02	119	MrKP4	134.442	52.042
e.m.s_0014_0002.14	wsc domain-containing protein	0.0	42.6	8.7E-06	42.6			
e.m.s_0010_0002.57	Zinc metalloprotease	10.7	92.8	2.6E-03	8.7			

Mean\* is the normalized mean based on the size factor of the libraries.

**Table 10 GPI anchor proteins**

Gene Id	Gene name	30DPI	60 DPI	Pvalue
		Mean*	Mean*	
evm.model.sctg_0115_0006.6	Related to tgf beta induced protein ig-h3 precursor	1303.93	283.09	1.03E-02
evm.model.sctg_0044_0002.3	Chitin deacetylase 9	330.81	98.49	4.34E-02
evm.model.sctg_0035_0002.62	gmc oxidoreductase	15.03	10883.07	3.48E-03
evm.model.sctg_0474_0002.6	gmc oxidoreductase	1.99	6162.70	1.85E-04
evm.model.sctg_0154_0003.4	Nucleus protein	21.52	377.01	1.01E-05
evm.model.sctg_0023_0002.46	Aspartic peptidase a1	5.89	793.02	5.90E-05
evm.model.sctg_0233_0003.16	gmc oxidoreductase	12.78	268.23	9.28E-03
evm.model.sctg_0090_0005.3	Aspartic peptidase a1	4.60	105.14	4.91E-05
evm.model.sctg_0003_0001.29	Carbohydrate esterase family 4 protein	81.65	240.08	4.51E-02
evm.model.sctg_0470_0001.4	Carbohydrate esterase family 4 protein	23.70	94.23	1.66E-02
evm.model.sctg_0007_0001.101	Extracellular serine-rich	264.05	1547.88	2.37E-03
evm.model.sctg_0016_0004.31	Glycoside hydrolase family 16 protein	0.70	17.08	1.66E-02
evm.model.sctg_0061_0002.53	Glycosyl hydrolase 53 domain-containing protein	2.80	87.42	2.18E-02
evm.model.sctg_0069_0001.19	Glyoxal oxidase	15.69	92.15	3.58E-02
evm.model.sctg_0233_0003.16	gmc oxidoreductase	12.78	268.23	9.28E-03
evm.model.sctg_0149_0001.6	Macrofage activating glycoprotein	5.38	47.41	4.89E-02
evm.model.sctg_0149_0001.5	Macrofage activating glycoprotein	1.99	59.62	1.07E-04
evm.model.sctg_0154_0003.4	Nucleus protein	21.52	377.01	1.01E-05
evm.model.sctg_0020_0001.30	Serine-threonine rich	92.38	614.18	4.02E-03

Mean\* is the normalized mean based on the size factor of the libraries.

## Discussion

### Genome comparison

We have produced a high-quality draft genome of *M. roreri* using different next-generation sequencing technologies. After sequence assembly using the whole genome shotgun (WGS) strategy, a total of 3,298 contigs with a N50 value of 48.1 kb was assembled with a total genome size calculated to be 52.3 Mbp. Although larger, the *M. roreri* genome is structurally and organizationally similar to *M. pernicioso*, a sister taxa with a draft genome consisting of 3,087 contigs (N50 = 48 kb) and a genome size of 44.7 Mbp. The majority of the size difference between the genomes is associated with repetitive DNA. The *M. roreri* genome size is larger than the genome of *M. pernicioso*, but the number of genes located in each genome is similar.

One factor for the size difference between *M. roreri* and *M. pernicioso* is highly correlated to the difference in the number of Long Terminal Repeat–Transposable Elements (LTR-TEs) between the genomes. About 3.46 Mbp or approximately 7% of the *M. roreri* genome is attributed to LTR-TEs located in the genome, while only 0.332 Mbp or about 1% of the *M. pernicioso* genome is attributed to LTR-TEs. Transposable Elements (TEs) contribute no more than 10 to 15% in most fungal taxa [32]. For example, the *Saccharomyces cerevisiae* genome contains 3.1% TEs [33], the *Magnaporthe oryzae* genome ranges from 8.2–14% [34],

while 24% of the *Laccaria bicolor* genome is transposon derived [35]. Within the *M. pernicioso* genome these repetitive elements are active and appear to contribute to the genetic variability of the species [20]. Another factor contributing to the size difference between the genomes is associated with repetitive elements and low complexity DNA. *M. roreri* genome has four times as many repetitive elements (7,060,129 bp) as found in the *M. pernicioso* genome (1,737,865 bp). The classification and biological roles of these repetitive sequences and transposable elements in *M. roreri* are beyond the scope of this discussion but will be explored in depth in a separate treatise.

The high degree of similarity between *M. roreri* and *M. pernicioso* genomes supports their taxonomic relatedness, while distinctions have been identified that could contribute to their distinctive functional and morphologic differences. Bidirectional blast analysis of *M. roreri* with four other basidiomycetes revealed a set of 1,133 genes specific to *M. roreri*. The majority of these genes are hypothetical with only 16 genes having known homologies and most of these relate to TEs. Expression analysis of the A gene set (genes specific to *M. roreri*) revealed that only seven of these genes are expressed and all at low levels. Among the genes encoding hypothetical proteins that are specific to *M. roreri*, 129 were constitutively expressed and only three were differentially expressed in both the biotrophic and necrotrophic phases (Additional file 1, A set

sheet). Low to mid-level constitutive expression in genes from the ABE and ABC intersects (Tables 5 and 6) suggest that these genes are used in general functions common to the sampled basidiomycetes such as metabolism. The differential expression of the 14 genes shared by *M. roreri* and *M. perniciosa* (AB intersect) in the biotrophic and necrotrophic phases, as expression of these genes suggests similarities in the hemibiotrophic life cycle of these species.

### Secretome targeting the plant cell

RNA-Seq, was used to validate the genome CDS through transcriptome expression analysis of *M. roreri* infected cacao pods at 30 DPI and 60 DPI. These two sampling points compare genes expressed in the early stage or biotrophic phase and the late stage or necrotrophic phase of the FPR disease cycle. *In silico* analysis of genes with signal peptides, which have a high likelihood of being transported outside of the fungus, yielded 1,355 genes expressed at some level in the 30 and 60 DPI samples, as detected by RNA-Seq. Only 222 of these genes were differentially expressed between the biotrophic and necrotrophic phases, using a P-value of 0.05. The similarities in gene expression between the two phases are greater than the dissimilarities. For example, genes encoding plant cell wall-degrading cellulases were not upregulated in the biotrophic phase, but, their expression was detected. Although the functions of many of these genes are unknown their expression raises the potential that they serve critical roles in the *M. roreri* life cycle.

Many of the known secreted protein encoding genes that were upregulated in the biotrophic phase encode for putative glycoside hydrolases (GH) that act on oligosaccharides making up the plant cell wall. Of the 34 differentially expressed secreted protein genes associated with the biotrophic phase, 11 were glycoside hydrolases from six families of hydrolases, which enables *M. roreri* to interact with the different biochemical components of plant cell walls. This diverse array of polysaccharide degrading enzymes could function together with other hemicelluloses-degrading enzymes, such as those that hydrolyze xylan. All of these polysaccharide degrading enzymes appear to work in concert to loosen the connections between the plant cells, releasing nutrients that can be utilized by the fungus. Pectin is another major component of the intercellular matrix of the cacao pod husk [36]. *M. perniciosa*, the causal agent of WBD on cacao, has been observed to express genes for pectin lyase and pectin methylesterases during infection of parthenocarpic pods [18]. In FPR, pectin is also under attack during the biotrophic phase of *M. roreri*. Genes encoding pectin degrading enzymes are upregulated in this phase.

The biotrophic phase of *M. roreri* is a complex stage requiring controlled cell wall degradation and modification, allowing the invasion of the pod while the fungus

remains substantially undetected by plant defenses. All of the *M. roreri* genes encoding plant cell wall enzymes that are induced in the biotrophic phase are able to attack the matrix of the intercellular space. This interaction likely facilitates penetration and spread of the fungus between cells without eliciting major responses from the plant.

In contrast, the expression pattern of genes encoding secreted proteins during the necrotrophic phase of *M. roreri* suggest a more aggressive breakdown of plant structures and the upregulation of aryl-alcohol oxidases, laccases, and a glyoxal oxidase gene during the necrotrophic phase suggests that *M. roreri*, targets lignin for breakdown during the necrotrophic phase. This evidence is further supported by the fact that three of the four aryl alcohol oxidases and three of the four laccases genes are exclusively expressed during the necrotrophic phase (Table 9) [28-30]. There are other upregulated enzymes that may also have a role in these ligninolytic reactions such as a fungal peroxidase, a copper amine oxidase and two dioxygenases. Fungal peroxidases were induced in *M. perniciosa* under nutrient limited conditions [18], while intradiol dioxygenase was shown to be induced in *M. roreri* during necrotrophy in a prior study [21].

Similar to the biotrophic phase, eight glycoside hydrolase (GH) families were also induced in the necrotrophic phase. Although the specific details concerning the exact function of these GH families of genes are lacking, it is interesting to note the differential expression patterns between the biotrophic and necrotrophic phases. GH families shared with the biotrophic phase include families 5, 16 and 18. GH families 28, 43 and 92 are unique to the biotrophic phase, whereas families 3, 9, 10, 35, and 61 are unique to the necrotrophic phase. The GH families unique to the necrotrophic phase include enzymes involved in the degradation of plant cell walls into compounds that the fungus can consume. Complex stage-specific regulation of genes targeting plant cell walls in association with the biotroph/necrotroph shift was also observed in hemibiotrophic *Colletotrichum* species [37]. The differential expression of genes with seemingly related functions provides the opportunity to fine tune aspects of growth and development to the changing conditions and requirements of the two phases of FPR.

Other cell surface-modifying enzymes upregulated in the necrotrophic phase are triacylglycerol lipase precursors and a cutinase suggesting lipid breakdown and modification of the protective surfaces of the plant cells. A necrosis inducing-like protein (NLP) that could be responsible for cell death and nutrient leakage was also found to be up-regulated [38]. NLPs have been identified in many pathogenic associations such as WBD and other biotrophic interactions [39-41]. Two genes from another class of virulence genes, biotrophy-associated secreted proteins BAS (MrBAS1 and MrBAS2) are highly

expressed in the *M. royeri* necrotrophic phase. These effector proteins have been shown to accumulate in the biotrophic interface region for the ascomycete *Magnaporthe grisea* [42]; however, their roles in the necrotrophic phase of this interaction are yet to be determined. At the same time in the necrotrophic phase, multiple extracellular peptidases are being upregulated possibly in response to the proteins and peptides that are now available from the necrotic plant cells [43]. Specific activities of *M. royeri* associated with the necrotrophic phase appear to target cell wall and membrane breakdown in associations with enzymes that are linked with the switch of this fungus from biotrophic to necrotrophic phase. The differentially expressed secretome of the necrotroph suggests conditions for rapid growth of the fungus, where plant lignocellulosic cell walls are being degraded and multiple plant components are being released and are available to be used by the fungus as nutrients. The need for stealth has been removed and nutrient acquisition is predominantly to support the rapid growth associated with the necrotrophic phase [21].

#### Secretome targeting modification of the fungal cell wall

There is no evidence that haustoria or other cell wall-penetrating structures are formed during the biotrophic phase of *M. royeri*. The slow-growing *M. royeri* mycelia of the biotrophic phase are large pleomorphic cells that fill the spaces between the plant cells [21]. This abnormal mycelial growth habit does have some similarity to haustoria-like structures, although they are outside the cell wall and, possibly provide the fungus with more surface area for cell-to-cell interactions and exchange of metabolites. The necrotrophic mycelia are profuse and grow rapidly in necrotic tissues [21]. These mycelia are thin and elongated. The sporophores and resulting spores represent a second fungal morphology associated with the necrotrophic phase. Genes encoding proteins that are functional in and on the fungal cell wall were differentially expressed between the phases of FPR. Often unique members of the same gene family were differentially expressed.

#### Chitin synthesis and modification

There were five differentially expressed family 18 glycoside hydrolases associated with the biotrophic phase and one associated with the necrotrophic phase. These chitinolytic enzymes could be modifying the chitin structure and the surface of the mycelia. These modifications could protect the mycelia from plant defenses such as plant chitinases [44], thus allowing the fungus to grow intercellularly. Several chitinases are highly expressed in cacao pod tissues throughout development, and can be hyper-induced in response to *M. royeri* infection [21]. Another possibility is that the chitinases attack chitin oligosaccharides released from the fungus, after being attacked by plant chitinases. Additional extracellular chitinases could help eliminate

these chitin oligosaccharides and prevent the elicitation of plant defense responses [45].

A chitin deacetylase and a chitin synthase were highly upregulated in the necrotrophic phase. The induction of chitin synthase is consistent with the rapid growth of the fungus initiated during the necrotrophic phase. The chitin deacetylase and related enzymes like exo-beta-glucanases and family 4 carbohydrate esterases can modify chitin, either loosening the chitin polymer bonds or converting the chitin to chitosan. Like the rust fungus *Uromyces viciae-faba* [46], *M. royeri* may use chitin deacetylase and other enzymes to modify its own cell wall reducing the effectiveness of the plant enzymes. A resistant wall structure may be particularly important in the harsh environment created as plant tissues die. Although cacao defense genes were induced in the later stages of the biotrophic phase of FPR, their induction levels continued to increase once the necrotrophic phase has been initiated [21].

#### PR-1 related proteins

Pathogenesis-related 1 proteins (PR-1) are a class of plant genes induced by pathogens during the infection stage and they are partially responsible for systemic acquired resistance responses in those plants. It is of particular interest that plant-like defense proteins would be found in a pathogen and within two species of the same fungal genus. Of the 12 PR-1 related genes identified in the *M. royeri* genome, 10 have homology to PR-1 genes recently described in *Moniliophthora perniciosa* [47]. Three of these PR-1-like proteins are upregulated during the biotrophic phase, five are constitutively expressed and only *MrPr-1d* was upregulated in the necrotrophic phase. All of the MrPR-1-like proteins have the SCP/TAPS-like conserved domains (Figure 3). The *MrPR-1g* and *MrPR-1a* match the expression of their homologs in the biotrophic phase of *M. perniciosa* [47]. Also of interest is the C-terminal extension of the *MrPR-1g* protein, which is found in both *M. perniciosa* and *M. royeri* and could have a specific protein-protein mediated function as purposed by Teixeira et al., [47]. The *MrPR-1b* homolog was constitutively expressed in *M. royeri*, which differs from the case of *M. perniciosa* where it was upregulated in the necrotrophic phase. While the exact function of these PR-1 proteins in fungi is unknown, these fungal-derived PR-1 like proteins may be masking the fungal cell wall from plant defense mechanisms, or they may bind to them preventing their activity. Riviere et al., [48] demonstrated that *PR-1a* proteins in tobacco plants regulate the extracellular enzyme activity of  $\beta$ -(1-3)-glucanase activity and plants with PR-1 silenced genes had higher enzyme levels along with a decrease in the deposition of  $\beta$ -(1-3) glucans and callose. Callose deposition in the cell-wall sheath around the plasmodesmata openings is a dynamic regulator mechanism that appears to

0021_0002.73	MrPR-1a	-----RSQFLGAHNIERMSHGADPFQWSDDLASKADSWASKQQLKSGSEGLSD-----SRYGENIVAATGD	94
0001_0002.127	MrPR-1b	GGSSSGDIEQYLEAHNSVRAQHGAQPLTWSDEAASKAQWANNCKFHSGGSLG-----SFGENLAAGTSDS	265
0432_0001.5	MrPR-1d	-ALFEDWKQVVRQHNEYRARYGAPNLWSWDAIYPTDARYAGCKFOHNSGGG-----KYGENLAAGTGN	127
0433_0001.11	MrPR-1e	-----DVDOQLNAHNEVRVQHNAQPLTWNQVVAATQCNWANRQCTMCHSGG-----QYGENLAWGTG-S	84
0169_0002.7	MrPR-1f	-----SNSNLNAHNAERAOHGASALTWSKDVSAACQDWANQCNLDTSS--SA-----HYGENLAAGGAQ	85
0258_0001.1	MrPR-1g	-----DWLKLHNKERREHNAADKLSWDDLAQRAQQVYAEQCRLEV-----LDVAPLGNLGYGY	87
0301_0001.6	MrPR-1h	-----SGSDQEVYLNHNELDRAQHGASQLTWNKTELEHAACEWADRIFEFHSQGLG-----ELGENLSAGVGKD	143
0222_0002.9	MrPR-1i1	-----DDFVVEWLTTHNYERTFTGAGPLSWSGDLAYAAQWANNQ--TENPPP-----ADNRNYGENIG----R	93
0251_0004.1	MrPR-1i2	-----DEFIQEWNKLHNDEBRSYNAAPLSWNGDLAYAAQWAEQ--TENPPP-----ADSRNYGENIG--MYTKQ	97
0432_0001.6	MrPR-1j	-----TDEWKNTVLDQHNVNRAHYGAKPLTWSDELTPGTEWAWCNGQFEBHSAR-G-----RYGENLAAGTLVS	129
0516_0001.1	MrPR-1n	-----EAPLRGHNAERAKHGADKLDKLPDLTWSDEAASKAQWANNCKFHSGGSLG-----SFGENLAAGTSDS	84
1165_0001.2	MrPR-1o	-----TDEWKNTVLEQYNNVRAHYGAMPLTWSDELTPGTEWAWCNGQFEBHLVKFSCFRQQDFCRFLFRGISDAQGRYGENLAAGTVPVL	92
MpPR-1a.pro		-----KSQFLGVHNIERMSHGADPFQWSDDLASKADSWASKQQLKSGDGLS-----GIQYGENIVAATGD	94
MpPR-1b.pro		GVTNSNGDIEQYLAHNSVRAQHGAQPLTWSDEAASKAQWANNCKFHSGGSLG-----SFGENLAAGTSDS	305
MpPR-1c.pro		-----TDEWKNVLEQHNVNRAHYGAMPLTWSDELTPGTEWAWCNGQFEBHSAR-LA-----PYGCLAAGDPAS	116
MpPR-1d.pro		-----DDFVVEWLTTHNYERTFTGAGPLSWSGDLAYAAQWANNQ--TENPPP-----ADNRNYGENIG----R	136
MpPR-1e.pro		-----DIDQNLDAHNEVRVQHNAQPLTWNQVVAATQCNWANRQCTMCHSGG-----QYGENLAAGGGS	84
MpPR-1f.pro		-----SDSNLTAHNERDQHGAGKLEWSSDLASAQDWANQCNQFQTS--DS-----KYGENLARGSGQ	83
MpPR-1g.pro		-----TWLELHNKERGQYGAPELTDWSDGLAQAQWANNCKFHSGGSLG-----SFGENLAAGTSDS	81
MpPR-1h.pro		GPDGPSNNDKKIYLMHNKARAETHGAPPLTWSDELTPGTEWAWCNGQFEBHSAR-LA-----PYGCLAAGDPAS	164
MpPR-1i.pro		-----DSFKNKLELHNNEBRTTRQGLDLEWBDGLAWKACQVATQCNVDNPNQ-----WGENLAFNIGR	93
MpPR-1j.pro		-----TDEWKNVLEQHNVNRAHYGAMPLTWSDELTPGTEWAWCNGQFEBHSAR-HG-----RYGENLAAGTVPV	127
MpPR-1k.pro		-----SSSRLYVHNSKRQAKVPALEWNSHLMKCAQSVANTCGASTS--IN-----AVGNDR---TIY	92
0021_0002.73	MrPR-1a	FISVNAAVGTFLSDKQYNPENPV---YNHWTQAVVRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	169
0001_0002.127	MrPR-1b	YTIARAARKGWDVSDYDNNPK---ASHFTQVVKRATQVCGALVL-C--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	339
0432_0001.5	MrPR-1d	YGFSSGLKSWDEASKYDYNKPGFSTATGHFTQVVKRSKQVACAIANORGGTIFQQP--SKYIVCRYPBPGNFAGRFAE	214
0433_0001.11	MrPR-1e	FPIASAQKLVADVSEYDPNNPQYS---HFTQVVKRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	151
0169_0002.7	MrPR-1f	FTPEASAVNLWLDLSDKDYDNNPK---SSWTOVVKRSTNLECGAQA-KCSQKNGNDQDDEVTFYVYCYNFBPGNIVVQALD	161
0258_0001.1	MrPR-1g	-KPEVFEFLNAMEANYPNDVEE---ATPMMQIVYKFTQSVGGSSAS-CPGY-GENKW---TLNVGLYDVA--ISEDYAN	156
0301_0001.6	MrPR-1h	FDAASAQKLVADVSEYDPNNPQYS---HFTQVVKRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	217
0222_0002.9	MrPR-1i1	YTKCAFADAMKATKDFYP--DDL---SRWQIVVKRSTNLECGAQA-KCSQKNGNDQDDEVTFYVYCYNFBPGNIVVQALD	161
0251_0004.1	MrPR-1i2	Q-AFLA---WKQTKSSYP-GDA---SRWQIVVKRSTNLECGAQA-KCSQKNGNDQDDEVTFYVYCYNFBPGNIVVQALD	161
0432_0001.6	MrPR-1j	YGIENAMASWNNVSGSYDYGHPREFSTSTSHFTQVVKRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	208
0516_0001.1	MrPR-1n	LTAVEVLEFFRSFKKEYDPKPNQ---GTPTQVVKRSTNLECGAQA-KCSQKNGNDQDDEVTFYVYCYNFBPGNIVVQALD	155
1165_0001.2	MrPR-1o	YGFHHALENWMSVSSDRIVSEVPQELISRF	122
MpPR-1a.pro		FISVNAAVGTFLSDKQYNSANPA---YNHWTQAVVRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	169
MpPR-1b.pro		YTIARAARKGWDVSDYDNNPK---ASHFTQVVKRATQVCGALVL-C--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	379
MpPR-1c.pro		YGIEDAANLWLDLSDKDYDYPHGGFSTGHFTQVVKRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	193
MpPR-1d.pro		YGFSSGLKSWDEASKYDYNKPGFSTATGHFTQVVKRSKQVACAIANORGGTIFQQP--SKYIVCRYPBPGNFAGRFAE	214
MpPR-1e.pro		FPIASAQKLVADVSEYDPNNPQYS---HFTQVVKRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	151
MpPR-1f.pro		HAPEDAANLWLDLSDKDYDYPHGGFSTGHFTQVVKRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	159
MpPR-1g.pro		ESPDKVFKLWAEETGYNYP-HVSE---ATPMMQIVYKFTQSVGGSSAS-CPGY-GENKW---TLNVGLYDVA--ISEDYAN	146
MpPR-1h.pro		FDAASAQKLVADVSEYDPNNPQYS---HFTQVVKRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	241
MpPR-1i.pro		YTKCAFADAMKATKDFYP--DDL---SRWQIVVKRSTNLECGAQA-KCSQKNGNDQDDEVTFYVYCYNFBPGNIVVQALD	160
MpPR-1j.pro		YGIENAMASWNNVSGSYDYGHPREFSTSTSHFTQVVKRSSTQHQGAI-SEC--DGIFDPEYGTAKLVICLYSPAGNIVVQALD	209
MpPR-1k.pro		KVLPQTIGLQGFNAMIKAWIPD---DPYTLAMAADVTCVGCASA-NCIQAG-----IVLAGSYN	148
0021_0002.73	MrPR-1a	NVQP	173
001_0002.127	MrPR-1b	NVQA	343
0432_0001.5	MrPR-1d	NVGRPRRLQADLD	227
0433_0001.11	MrPR-1e	NVEV	155
0169_0002.7	MrPR-1f	NVQP	165
0258_0001.1	MrPR-1g	NIGPNGGQGGKKKPSK---EEDKIKTKPKPKKE-----SQSQEKEKKPEEHKEEQSSG-EDT SRLDRAASVIQGFK	224
0301_0001.6	MrPR-1h	NIQV	221
0222_0002.9	MrPR-1i1	NV	163
0251_0004.1	MrPR-1i2	NVEAQVRVPLAFRSYG	178
0432_0001.6	MrPR-1j	NVGIQV	214
0516_0001.1	MrPR-1n	NVQP	159
1165_0001.2	MrPR-1o	NVQP	122
MpPR-1a.pro		NVQP	173
MpPR-1b.pro		NVQA	383
MpPR-1c.pro		NVGKPRPQYQ	204
MpPR-1d.pro		NVGRPRRLQADLD	227
MpPR-1e.pro		NVEV	155
MpPR-1f.pro		NVQP	163
MpPR-1g.pro		VEGGGKKKSPSKNIAIPPLAKGGKKEKEEKEGPKHKEKEGPKHKEKEGPKHKEKEKAGGGEENDNLKHAEDVISRLK	230
MpPR-1h.pro		NVNL	245
MpPR-1i.pro		NAGDN	165
MpPR-1j.pro		NVSMQ	212
MpPR-1k.pro		NVQP	148

**Figure 3 PR-1 Alignment of the conserved SCP/TAPS protein domains.** Similarities are shown for the PR-1 like-proteins from *Moniliophthora roreri* and *Moniliophthora perniciosa*. Most of the sequences similarities are within the SCP/TAPS domain region. Conserved amino acids are highlighted with blue representing 100% identity and green representing at least a 60% identity. The putative active sites in these proteins are highlighted in red and the conserved cysteines are highlighted in yellow.

alter the size of the plasmodesmata orifice, which in turn can modify cell to cell transport by restricting the size of transport molecules [49]. Previous work has shown that *M. royeri* infected cacao pods at 30 DPI have a 1.6 to 2 fold reduction in glucose, phenylalanine and asparagine levels [21], which suggests the fungus has access to these plant metabolites, despite being limited to the intercellular spaces. During the necrotrophic phase (60 DPI) most pod metabolites are severely depleted. The *M. royeri* mycelia, which grow between plant cells, may have access to the plasmodesmata passages and the nutrients that pass through them. The PR-1 like proteins appear to be important to virulence, and it has been suggested that they support systemic spread of the fungus, possibly by limiting susceptibility to host defenses, or they may act as effectors suppressing host defenses [50]. The phase specific expression of the different PR-1 encoding genes is interesting, particularly because of the difference in the host interactions of the biotroph and necrotroph. The PR-1 proteins expressed by *M. royeri* in the different phases could be acting to alter the plant cell walls or as competitive inhibitors of plant PR-1 proteins, thus minimizing the plant defense response, or they could have antimicrobial activity, which would function to prevent subsequent infection and development of competing microorganisms.

#### **Hydrophobins**

Of the 41 hydrophobins found in the genome, three members of the hydrophobin gene family are up-regulated in the biotrophic phase and 10 members are up regulated in the necrotrophic phase. The hydrophobins interact at the fungal cell surface creating a hydrophobic surface layer and are important in many morphogenetic processes such as sporulation, fruiting body formation and infection structures [51]. Hydrophobins are grouped into three classes Ia, Ib, and II, with Ia and II occurring in ascomycetes and Ib in basidiomycetes [52]. These proteins have the unique ability to self assemble once secreted and they function at the fungal wall-air interface, limiting desiccation and providing protection against both chemical and enzymatic attack [53]. It is easy to postulate the need for different hydrophobins between the two phases of FPR, considering the need to limit detection in the biotrophic phase and the need to protect the surface from enzymatic and chemical attack in the necrotrophic phase. In animal systems at least, hydrophobins are known to prevent host recognition of pathogen-associated molecular patterns, thus preventing the activation of host immune response [54]. Hydrophobins are also of interest because of their potential uses in industrial processes [52].

#### **Cerato-platanins and related proteins**

Three cerato-platanin (CP) genes are up-regulated in the biotroph and four are upregulated in the necrotroph.

Cerato-platanin domain-containing proteins such as EPL (eliciting plant response-like), Snodprot, swollenins and expansins have a double  $\psi\beta$ -barrel fold and a carbohydrate binding site that have been shown to induce plant defense response [55,56]. Cerato-platanins can cause nutrient leakage and cell death in cacao [57,58]. Ectopic expression of the *Magnaporthe grisea* CP gene, *MgSM1*, upregulates the expression of plant defense genes such as PR-1, PR-5 and PDF1.2 and induces local hypersensitivity reactions (HR) [59]. This class of fungal proteins is called elicitors or effector proteins. A fungal effector protein found in *Colletotrichum truncatum* (CtCP1) is expressed during the infection stage [60]. Unlike the *MgSM1* protein, CtCP1 does not cause HR or elicit a plant defense response, [60] demonstrating that the activity of cerato-platanins can be both host and protein specific. An EPL gene from *Trichoderma* species produces a protein that binds to chitin and self assembles at the air/water interface like a hydrophobin [61]. When MSP1, a snodprot1 homolog protein, was knocked out in *Magnaporthe oryzae*, the fungus was unable to grow within the plant after normal infection suggesting this protein has a key function during the intercellular growth of this fungus [62]. Other CP-family proteins such as snodprot1 homologs are expressed in ecotomycorrhizal fungi *Pisolithus microcarpus* and *Laccaria bicolor* [63], suggesting a role for this family of genes during the plant fungal interaction.

Two expansins were upregulated in the necrotrophic phase of this study. Expansins were first discovered in plants, due to their role in developmental processes involving cell wall modifications, including cell enlargement, by "loosening" plant cell walls [64]. A functional expansin family protein was reported in the basidiomycete *Bjerkandera adusta* [65], and expansin-like proteins have been reported in other fungi, including plant pathogens. Brotman *et al.* [66] reported swollenin, an expansin family protein, in *Trichoderma* that is involved in colonization of plant roots and apparently elicits plant-defense responses. The expansin-like and cerato-platanin proteins may be involved in invasion and pathogenesis of plant tissues by phytopathogens that employ penetration of host/susceptible tissues. Since members of the cerato-platanin family are highly expressed in both phases of FPR, it further supports scenarios in which the fungus uses unique members of complex gene families to facilitate growth and spreading under the conditions unique to each phase.

Another indicator of the biotrophic/necrotrophic switch is the upregulation of several genes in the necrotrophic phase that are associated with fungal reproduction such as the heterokaryon incompatibility protein, a fruiting body-specific gene and a riboflavin aldehyde-forming enzyme. The reproductive stage of the fungus occurs after the necrotrophic phase and is a nutrient intensive process

[13]. Cytochrome P450s and hydrophobins have been previously found in necrotic cacao pods infected with FPR [21] and have been shown to be expressed during fungal reproduction, specifically in the fruiting body development [53,67]. However, unlike most basidiomycetes, *M. royeri* has lost the ability to form a basidiocarp or mushroom and sporulates directly from the mycelial mat that it forms on the diseased pod [13].

## Conclusion

The *M. royeri* genome and expression analyses provide insight into the molecular mechanisms of the biotrophic and necrotrophic phases of this important hemibiotrophic pathogen. RNA-Seq analysis revealed differential gene expression during the biotrophic and necrotrophic phases of *M. royeri* during the development of FPR disease of cacao. Focusing on the secreted proteins found in this genome, we observed specific secreted proteins putatively involved in plant cell wall degradation in the biotrophic phase, which seem to degrade and utilize the intercellular spaces. At the same time, they appear to be masking or modifying the fungal cell surface to avoid plant defenses, allowing the fungus to colonize the pod. In contrast, other secreted proteins upregulated in the necrotrophic phase suggest that the fungus continues to protect its mycelia from plant defenses, while actively releasing enzymes and toxins that have been reported to attack the plant cell wall components and utilizing plant nutrients that are released during plant cell death. These findings support the existing pathogenesis model for FPR while giving greater details regarding the identity of putative secreted proteins involved in the different stages, and information regarding the differential use of unique members of complex gene families. Due to the high number of genes that encode hypothetical proteins in this genome and that are expressed during the disease process more research is needed to fully understand the various phases of this hemibiotrophic fungus.

## Methods

### Biological material, libraries construction and sequencing

The *M. royeri* clone (MCA2977) was isolated from the state of Los Rios in Ecuador. It was grown and maintained on a yeast extract ( $5 \text{ gL}^{-1}$ ) and glycerol ( $20 \text{ mL L}^{-1}$ ) medium. Cultures were incubated at  $27^\circ\text{C}$  at 200 rpm for 7 d. Total DNA was extracted using a cetyl trimethylammonium bromide (CTAB) method [68].

### Genome assembly

*Moniliophthora royeri* DNA sequences were obtained by a hybrid sequencing approach using GS FLX and Illumina sequencing, and the whole-genome shotgun strategy. The GS FLX sequencer (454 Life Sciences/Roche) was used to produce 3,440,399 single-end reads with a

mean length of 230 bp and 417,891 reads from paired end runs with 2Kbp insert size. An Illumina Genome Analyzer was used to produce 30,826,974 read pairs ( $2 \times 76$  bp, insert size of 365 bp). The Illumina data was filtered to remove reads from mitochondrial genome and plasmids. Illumina data was error-corrected using Quake v0.3 with default parameters. In total, 24,405,509 read pairs passed Quake error correction. The final hybrid assembly of both 454 data and Illumina data was performed by using the Newbler v.2.6 assembler (default parameters with overlap parameter  $\geq 40$  and identity to  $\geq 95\%$ , <http://www.454.com/products/analysis-software/>) [69,70].

## Combined gene models

### Ab-initio gene models prediction

Genefinding prediction programs, Genemark-ES v2.3 [71] and Augustus v2.3.1 [72,73] programs optimized with the Exonerate v2.2.0 protein2genome model [74] were used to perform the *ab-initio* gene prediction. Genefinding prediction program Genemark.hmm [75] was executed in self-training mode and ORFs larger than 200 bp were retained. The Genemark-ES predicted proteins were aligned against the GenBank NR, using BLASTP with a e-value cutoff of  $1e-10$ . Finally, Evidence Modeller (EVM) [76] was used to combine the predictions from Genemark-ES, Augustus and Exonerate spliced protein alignments and generate a consensus gene prediction. Select gene models were manually inspected, especially incomplete genes 5' and 3' ends of the scaffold.

### Automatic annotation

The automatic annotation program AutoFACT [77] was used for functional annotation of gene models. The main contribution of AutoFACT is the capacity to conduct the annotation based on sequence similarity searches in several databases. For this, we used the BLASTp (e-value cutoff of  $1e-5$ ) to align the gene models against various proteins databases: non-redundant (NR) database at NCBI, SwissProt—databases containing only manually curated proteins [78], uniref90 and uniref100—databases containing clustered sets of proteins from UniProt, Pfam—database of proteins families [79] and KEGG—database of metabolic pathways [80]. Open reading frames were also annotated using Blast2GO (<http://www.blast2go.com/b2ghome>) [81].

### Transposable elements

The *M. royeri* and *M. perniciososa* genomes were scanned for transposable elements. Each genome was fragmented *in silico* into 1000-bp fragments. These fragments were compared to a database of full length LTR retrotransposons developed by the program LTR struc [82] of LTRs associated with *M. royeri* and *M. perniciososa* using the BLAST program. Fragments with significant E-values ( $E-04$ ) and

better were sequestered. The rest of the fragments with no hits to LTR retrotransposons were scanned against the repbase database [83], which is a collection of sequences representing repetitive DNA from different eukaryotic species using the Blast program.

The *M. royeri* and *M. perniciosa* genomes were also scanned for repetitive DNA using the program Repeat Masker with full Repbase. However, this scan did not find all the LTR and repetitive elements. Thus, repeat Modeler was also used to find *denovo* repeats and used as the database for input into Repeat Masker. This analysis is shown (Additional file 1 Repeat modeler sheet) *Repeat-Masker Open-3.0*. <<http://www.repeatmasker.org>>.

## Transcriptome

### *Transcriptome analysis in infected pods*

Briefly, the pods used in this study were derived from flowers hand pollinated 60 days prior to infection with a spore suspension of  $1 \times 10^8$  spores  $\text{ml}^{-1}$  [21]. The plant material used was a segregating F1 population of progeny from highly susceptible clones (PA16  $\times$  SIC433 and PA16  $\times$  EEG29). Seven infected pods were frozen in liquid nitrogen at 30 days post infection (DPI) and another seven at 60 DPI. Based on the typical progression of FPR symptoms [21], malformed green pods were selected for RNA-Seq and RT-qPCR from 30 DPI pods which represents the biotrophic phase. Necrotic sporulating pods were selected for RNA-Seq and RT-qPCR from 60 DPI pods, which represents the necrotrophic phase.

### RNA extraction

Each infected pod was broken up and coarsely ground under liquid nitrogen and approximately  $1 \text{ cm}^3$  was used for processing. The pod material was ground finely and transferred to a disposable 50 mL centrifuge tube containing 15 mL of 65°C extraction buffer [84]. Additional extraction methods were conducted as in Melnick *et al.* [18]. The cDNA was synthesized using the Invitrogen (Carlsbad, CA) Superscript VILO kit, following the manufacturer's directions.

### RNA sequencing

For the genome-wide analyses of expression patterns, cDNA was generated using a routine RNA library preparation TruSeq protocol developed by Illumina Technologies (San Diego, CA). Using the kit, mRNA was first isolated from total RNA by performing polyA selection step, followed by construction of single end sequencing libraries with an insert size of 160 bp. Single-end sequencing was performed on six samples (three 30 DPI and three 60 DPI) using the Illumina HiSeq platform. Samples were multiplexed with unique six-mer barcodes generating 404,820,524 filtered (for Illumina adapters/primers, and PhiX contamination)  $1 \times 50$  bp paired end reads.

## Expression analysis

Reads from 30 DPI and 60 DPI libraries were mapped to the nucleotide sequences of predicted protein coding genes of the *M. royeri* genome using the short read aligner Bowtie-0.12.7 which utilizes a Burrows-Wheeler index [85]. Count data were obtained for each coding sequence. Estimation and statistical analysis of expression levels using count data of each gene with 3 replicates for each library was performed using the DESeq package [86] and R x64 2.15.2 program. (<http://www.r-project.org/>).

After RNA-Seq, twenty five genes were chosen for analysis by RT-qPCR across the full set of infected pods. For RT-qPCR analysis, seven replicate malformed green pods 30 DPI and seven replicate necrotic sporulating pods (60 DPI) were used. RT-qPCR analysis was conducted following Bailey *et al.* [21], using Brilliant III<sup>®</sup> SYBR<sup>®</sup> Green Q-PCR Master Mix (Agilent, Santa Clara, CA). Primer sources, sequences for the *M. royeri* genes are in Additional file 3. *M. royeri* ESTs were chosen based upon results of RNA-Seq analysis. RT-qPCR was conducted to determine the changes in expression of *M. royeri* genes between the biotrophic and necrotrophic phases of FPR. The ddCt method was used to calculate the fold-change between the 30 and 60 day DPI samples [87,88].

## Statistical analysis of data

A two-way ANOVA of the RNA-Seq data was conducted using the general linear model (PROC GLM) followed by Tukey post-hoc testing ( $\alpha = 0.05$ ) using SAS 9.3 (SAS Institute Inc., Raleigh, NC, USA) to analyze pod data. For RT-qPCR, relative transcript levels were determined following Pfaffl [89].

## Secreted proteins

17,920 protein coding sequences were scanned for signal peptides using signalP [90,91]. The resulting proteins containing signal peptides were scanned for transmembrane proteins using the TMHMM program (Prediction of transmembrane helices in proteins) [92]. Proteins with no more than two transmembrane domains were sequestered for further analysis. The resulting proteins were scanned for GPI-anchored proteins [93,94] using FragAnchor, which is based on the tandem use of a Neural Network predictor and a Hidden Markov Model predictor (<http://navet.ics.hawaii.edu/~fraganchor/NNHMM/NNHMM.html>) [95]. Proteins that were highly probably and probable of having a GPI-anchor were discarded. Thus, using bioinformatics, proteins with signal peptides, with two or less transmembrane domains and without GPI-anchors were considered for possible secretion.

## Venn diagram

Using a bi-directional blast method the *M. royeri*, *M. perniciosa*, *Laccaria bicolor*, *Coprinus cinereus*, and *Ustilago*

*maydis* genomes were compared and used to generate a 5-way Venn diagram. In addition, five tax plots for all five fungal species were created using each genome as a reference, respectively. Protein sequences that share an expected E-value of E-04 were considered matches.

### Dot plots

Dot plots between the large contigs (50,000 bp or larger) of *M. royeri* and *M. perniciosa* were created using the whole genome aligner MUMmer3.22. The NUCmer and MUMmerplot programs from the MUMmer suite were used [96,97].

### Availability of supporting data

This *M. royeri* Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AFSO00000000. The version described in this paper is version AFSO01000000. The sequencing data used in this study can be found under NCBI BioProjects. BioProject PRJNA213737 is linked to the genome data and PRJNA229176 was established for the transcriptome data.

### Additional files

**Additional file 1:** This file contains the Repeat Modeler file and a complete list of the Venn diagram intersects; A, AB, ABC and ABE discussed in the text. There is a sheet for each set and intersect and RNA-Seq data are provided for each gene in that intersect.

**Additional file 2:** This file contains the complete list of the 1535 secreted genes and the 125 GPI-anchor containing proteins identified in the *M. royeri* genome. The file contains RNA-Seq data for each gene and RT-qPCR data for a number of selected genes. The corresponding *M. perniciosa* homolog is also listed with the homology E-value.

**Additional file 3:** This file contains the complete set of RT-qPCR primers used in this study. The file provides the primer name, nucleotide sequence, *M. royeri* gene Id, PCR fragment size and the nucleotide sequence for each fragment.

### Abbreviations

*M. royeri*: *Monilophthora royeri*; *M. perniciosa*: *Monilophthora perniciosa*; FPR: Frosty Pod Rot; CDS: Coding sequences; OFRs: Open reading frames; RNA-seq: RNA sequencing; WBD: Witches' broom disease; Mbp: Million base pairs; NLP: Necrosis inducing-like protein; NP-50: Statistical measure of average length of contigs in a Draft sequence.

### Competing interests

The authors declare that they have no competing interests.

### Author's contributions

LWM conceived the study, participated in its design and coordination, ARS lead and drafted the manuscript. GGLC participated in the sequence alignment and bioinformatic analysis of the genome. DPTT sequence analysis and helped to draft the manuscript. PJLTL sequence analysis and helped to draft the manuscript. MFC participated in the sequence alignment and bioinformatic analysis of the genome. SCS conducted 454 sequencing and participated in the sequence alignment. JEC conducted 454 sequencing and participated in the sequence alignment. MJG participated in its design and coordination of the study, PSU lead. PM conducted Illumina sequencing and participated in the sequence alignment. AF conducted RNA sequencing and bioinformatics analysis. TR conducted RNA sequencing and bioinformatics analysis. JC conducted field studies, collected samples and participated in the statistical analysis of data. RED sequence analysis and helped to draft the

manuscript. JS participated in the sequence alignment, bioinformatic analysis of the RNA-seq data and helped to draft the manuscript. RLM carried out molecular genetic studies, participated in the statistical analysis of data and helped to draft the manuscript. GAGP participated in its design and coordination of the study, Brazilian lead. BAB conceived the study, participated in its design and coordination, carried out molecular genetic studies and helped to draft the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

This work was funded by USDA ARS. We would like to acknowledge the contributions of the Center for Computational Engineering and Sciences at UNICAMP SP Brazil (FAPESP/CEPID project #2013/08293-7). References to a company and/or product by the USDA are only for the purposes of information and do not imply approval or recommendation of the product to the exclusion of others that may also be suitable. The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

### Author details

<sup>1</sup>Sustainable Perennial Crops Lab, USDA/ARS, Bldg 001 Rm 223 Beltsville Agricultural Research Center-West, Beltsville, MD 20705, USA. <sup>2</sup>Centro Nacional de Processamento de Alto Desempenho em São Paulo, Universidade Estadual de Campinas, CP 6141, Campinas 13083-970, SP, Brazil. <sup>3</sup>Laboratório de Genômica e Expressão, Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), CP 6109, Campinas 13083-970, SP, Brazil. <sup>4</sup>Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA 16802, USA. <sup>5</sup>Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA 16802, USA. <sup>6</sup>Department of Horticulture, Pennsylvania State University, University Park, PA 16802, USA. <sup>7</sup>Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Mary Ellen Jones, Room 921, 27599-3280 Chapel Hill, NC, USA. <sup>8</sup>National Center of Genomic Research, 2935 Rodeo Park Drive East Santa Fe, NM 87505 Santa Fe, USA. <sup>9</sup>CABI Bioscience UK Centre, Egham, UK. <sup>10</sup>Molecular Plant Pathology Lab, USDA/ARS, Bldg 004 Rm 119 Beltsville Agricultural Research Center West, Beltsville, MD 20705, USA.

Received: 29 September 2013 Accepted: 14 February 2014

Published: 27 February 2014

### References

1. Latunde-Dada AO: Colletotrichum: tales of forcible entry, stealth, transient confinement and breakout. *Mol Plant Pathol* 2001, **2**(4):187-198.
2. Oliver RP, Ipcho SVS: Arabidopsis pathology breathes new life into the necrotrophs-vs.-biotrophs classification of fungal pathogens. *Mol Plant Pathol* 2004, **5**(4):347-352.
3. Catanzariti AM, Dodds PN, Lawrence GJ, Ayliffe MA, Ellis JG: Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* 2006, **18**(1):243-256.
4. Link TI, Voegelé RT: Secreted proteins of *Uromyces fabae*: similarities and stage specificity. *Mol Plant Pathol* 2008, **9**(1):59-66.
5. Brown NA, Antoniw J, Hammond-Kosack KE: The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. *PLoS One* 2012, **7**(4):e33731.
6. Thomma BP: *Alternaria* spp.: from general saprophyte to specific parasite. *Mol Plant Pathol* 2003, **4**(4):225-236.
7. Evans HC, Stalpers JA, Samson RA, Benny GL: Taxonomy of *Monilia-Roreri*, an important pathogen of theobroma-cacao in South-America. *Can J Bot* 1978, **56**(20):2528-2532.

8. 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):1012–1022.
9. Phillips-Mora W, Wilkinson MJ: **Frosty pod of cacao: a disease with a limited geographic range but unlimited potential for damage.** *Phytopathology* 2007, **97**(12):1644–1647.
10. Meinhardt LW, Rincones J, Bailey BA, Aime MC, Griffith GW, Zhang DP, Pereira GAG: **Monilophthora perniciosa, the causal agent of witches' broom disease of cacao: what's new from this old foe?** *Mol Plant Pathol* 2008, **9**(5):577–588.
11. Ferreira LFR, Duarte KMR, Gomes LH, Carvalho RS, Leal GA, Aguiar MM, Armas RD, Tavares FCA: **Genetic diversity of polyporic isolates of Monilophthora perniciosa (Tricholomataceae).** *Genet Mol Res* 2012, **11**(3):2559–2568.
12. Phillips-Mora W, Wilkinson MJ: **Frosty pod: a disease of limited geographic distribution but unlimited potential for damage.** *Phytopathology* 2006, **96**(6):S138–S138.
13. Evans HC: *Pod Rot of Cacao caused by Monilophthora (Monilia) roreri*. 24th edition. London: Commonwealth Agricultural Bureau; 1981.
14. Joosten M, de Wit P: **THE TOMATO-CLADOSPORIUM FULVUM INTERACTION: a versatile experimental system to study plant-pathogen interactions.** *Annu Rev Phytopathol* 1999, **37**:335–367.
15. Perfect SE, Green JR: **Infection structures of biotrophic and hemibiotrophic fungal plant pathogens.** *Mol Plant Pathol* 2001, **2**(2):101–108.
16. Scarpari LM, Meinhardt LW, Mazzafera P, Pomella AWW, Schiavinato MA, Cascardo JCM, Pereira GAG: **Biochemical changes during the development of witches' broom: the most important disease of cocoa in Brazil caused by Crinipellis perniciosa.** *J Exp Bot* 2005, **56**(413):865–877.
17. Melnick RL, Marelli J, Bailey BA: **The molecular interaction of Theobroma cacao and Monilophthora perniciosa, causal agent of witches' broom, during infection of young pods.** *Phytopathology* 2011, **101**(6):S274–S274.
18. Melnick RL, Marelli JP, Sicher RC, Strem MD, Bailey BA: **The interaction of Theobroma cacao and Monilophthora perniciosa, the causal agent of witches' broom disease, during parthenocarpy.** *Tree Genet Genomes* 2012, **8**(6):1261–1279.
19. Thomazella DP, Teixeira PJ, Oliveira HC, Saviani EE, Rincones J, Toni IM, Reis O, Garcia O, Meinhardt LW, Salgado I, Pereira, GA: **The hemibiotrophic cacao pathogen Monilophthora perniciosa depends on a mitochondrial alternative oxidase for biotrophic development.** *New Phytol* 2012, **194**(4):1025–1034.
20. Mondego JM, Carazzolli MF, Costa GG, Formighieri EF, Parizzi LP, Rincones J, Cotomacci C, Carraro DM, Cunha AF, Carrer H, Vidal, RO, Estrela RC, Garcia O, Thomazella DPT, de Oliveira BV, Pires ABL, Rio MCS, Araujo MRR, de Moraes MH, Castro LAB, Gramacho KP, Goncalves MS, Neto JPM, Neto AG, Barbosa LV, Guitinan MJ, Bailey BA, Meinhardt LW, Cascardo JCM, Pereira GAG: **A genome survey of Monilophthora perniciosa gives new insights into Witches' Broom disease of cacao.** *Bmc Genomics* 2008, **9**:548.
21. Bailey BA, Crozier J, Sicher RC, Strem MD, Melnick R, Carazzolli MF, Costa GGL, Pereira GAG, Zhang DP, Maximova S, Guitinan M, Meinhardt L: **Dynamic changes in pod and fungal physiology associated with the shift from biotrophy to necrotrophy during the infection of Theobroma cacao by Monilophthora roreri.** *Physiol Mol Plant P* 2013, **81**:84–96.
22. Henrissat B: **A classification of glycosyl hydrolases based on amino acid sequence similarities.** *Biochem J* 1991, **280**(Pt 2):309–316.
23. Dias FM, Vincent F, Pell G, Prates JA, Centeno MS, Tailford LE, Ferreira LM, Fontes CM, Davies GJ, Gilbert HJ: **Insights into the molecular determinants of substrate specificity in glycoside hydrolase family 5 revealed by the crystal structure and kinetics of Cellvibrio mixtus mannosidase 5A.** *J Biol Chem* 2004, **279**(24):25517–25526.
24. Fibransah G, Masuda S, Koizumi N, Nakamura S, Kumasaka T: **The 1.3 Å crystal structure of a novel endo-beta-1,3-galactosidase of glycoside hydrolase family 16 from alkaliphilic Nocardia sp. strain F96.** *Proteins* 2007, **69**(3):683–690.
25. Markovic O, Janecek S: **Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution.** *Protein Eng* 2001, **14**(9):615–631.
26. Vandermarliere E, Bourgeois TM, Winn MD, van Campenhout S, Volckaert G, Delcour JA, Strelkov SV, Rabijns A, Courtin CM: **Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family.** *Biochem J* 2009, **418**(1):39–47.
27. Tiels P, Baranova E, Piens K, De Visscher C, Pynaert G, Nerinckx W, Stout J, Fudalej F, Hulpiau P, Tannler S, Geysens S, Van Hecke A, Valevska A, Vervecken W, Remaut H, Callewaert N: **A bacterial glycosidase enables mannose-6-phosphate modification and improved cellular uptake of yeast-produced recombinant human lysosomal enzymes.** *Nat Biotechnol* 2012, **30**(12):1225–1231.
28. Ferreira P, Hernandez-Ortega A, Herguedas B, Martinez AT, Medina M: **Aryl-alcohol oxidase involved in lignin degradation: a mechanistic study based on steady and pre-steady state kinetics and primary and solvent isotope effects with two alcohol substrates.** *J Biol Chem* 2009, **284**(37):24840–24847.
29. Mayer AM, Staples RC: **Laccase: new functions for an old enzyme.** *Phytochemistry* 2002, **60**(6):551–565.
30. Kersten PJ: **Glyoxal oxidase of Phanerochaete chrysosporium: its characterization and activation by lignin peroxidase.** *Proc Natl Acad Sci U S A* 1990, **87**(8):2936–2940.
31. Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP, Davies G: **Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases.** *Proc Natl Acad Sci U S A* 1995, **92**(15):7090–7094.
32. Wostemeyer J, Kreibich A: **Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution.** *Curr Genet* 2002, **41**(4):189–198.
33. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG: **Life with 6000 genes.** *Science* 1996, **274**(5287):546–563. 547.
34. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, Read ND, Lee YH, Carbone I, Brown D, Oh YY, Donofrio N, Jeong JS, Soanes DM, Djonovic S, Kolomiets E, Rehmeier C, Li W, Harding M, Kim S, Lebrun MH, Bohnert H, Coughlan S, Butler J, Calvo S, Ma LJ, Nicol R, et al: **The genome sequence of the rice blast fungus Magnaporthe oryzae.** *Nature* 2005, **434**(7036):980–986.
35. Labbe J, Murat C, Morin E, Tuskan GA, Le Tacon F, Martin F: **Characterization of transposable elements in the ectomycorrhizal fungus Laccaria bicolor.** *Plos One* 2012, **7**(8):e40197.
36. Adomako D: **Cocoa pod husk pectin.** *Phytochemistry* 1972, **11**(3):1145.
37. Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, Takano Y, Kubo Y, Shirasu K: **Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of Colletotrichum fungi.** *New Phytol* 2013, **197**(4):1236–1249.
38. Garcia O, Macedo JAN, Tiburcio R, Zapparoli G, Rincones J, Bittencourt LMC, Ceita GO, Micheli F, Gesteira A, Mariano AC, Schiavinato MA, Medrano FJ, Meinhardt LW, Pereira GA, Cascardo JC: **Characterization of necrosis and ethylene-inducing proteins (NEP) in the basidiomycete Monilophthora perniciosa, the causal agent of witches' broom in Theobroma cacao.** *Mycol Res* 2007, **111**:443–455.
39. Pemberton CL, Salmond GP: **The Nep1-like proteins—a growing family of microbial elicitors of plant necrosis.** *Mol Plant Pathol* 2004, **5**(4):353–359.
40. Zapparoli G, Barsottini MR, de Oliveira JF, Dyszy F, Teixeira PJ, Barau JG, Garcia O, Costa-Filho AJ, Ambrosio AL, Pereira GA, Dias SM: **The crystal structure of necrosis-and ethylene-inducing protein 2 from the causal agent of cacao's Witches' Broom disease reveals key elements for its activity.** *Biochemistry-US* 2011, **50**(45):9901–9910.
41. Cabral A, Oome S, Sander N, Kufner I, Nurnberger T, Van den Ackerveken G: **Nontoxic Nep1-like proteins of the downy mildew pathogen Hyaloperonospora arabidopsidis: repression of necrosis-inducing activity by a surface-exposed region.** *Mol Plant Microbe Interact* 2012, **25**(5):697–708.
42. Mosquera G, Giraldo MC, Khang CH, Coughlan S, Valent B: **Interaction transcriptome analysis identifies magnaporthe oryzae BAS1-4 as biotrophy-associated secreted proteins in rice blast disease.** *Plant Cell* 2009, **21**(4):1273–1290.
43. Paper JM, Scott-Craig JS, Adhikari ND, Cuomo CA, Walton JD: **Comparative proteomics of extracellular proteins in vitro and in planta from the pathogenic fungus Fusarium graminearum.** *Proteomics* 2007, **7**(17):3171–3183.
44. van den Burg HA, Harrison SJ, Joosten MH, Vervoort J, de Wit PJ: **Cladosporium fulvum Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection.** *Mol Plant Microbe Interact* 2006, **19**(12):1420–1430.
45. Roby D, Gadelle A, Toppan A: **Chitin oligosaccharides as elicitors of chitinase activity in melon plants.** *Biochem Biophys Res Commun* 1987, **143**(3):885–892.

46. Deising H, Siegrist J: **Chitin deacetylase activity of the rust *Uromyces-viciae-fabae* is controlled by fungal morphogenesis.** *Fems Microbiol Lett* 1995, **127**(3):207–211.
47. Teixeira PJPL, Thomazella DPT, Vidal RO, Do Prado PFV, Reis O, Baroni RM, Franco SF, Mieczkowski P, Pereira GAG, Mondego JMC: **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).
48. Riviere MP, Marais A, Ponchet M, Willats W, Galiana E: **Silencing of acidic pathogenesis-related PR-1 genes increases extracellular beta-(1->3)-glucanase activity at the onset of tobacco defence reactions.** *J Exp Bot* 2008, **59**(6):1225–1239.
49. Levy A, Guenoune-Gelbart D, Epel BL: **beta-1,3-Glucanases: plasmodesmal gate keepers for intercellular communication.** *Plant Signal Behav* 2007, **2**(5):404–407.
50. Prados-Rosales RC, Roldan-Rodríguez R, Serena C, Lopez-Berges MS, Guarro J, Martínez-del-Pozo A, Di Pietro A: **A PR-1-like protein of *Fusarium oxysporum* functions in virulence on mammalian hosts.** *J Biol Chem* 2012, **287**(26):21970–21979.
51. Kershaw MJ, Talbot NJ: **Hydrophobins and repellents: proteins with fundamental roles in fungal morphogenesis.** *Fungal Genet Biol* 1998, **23**(1):18–33.
52. Zelena K, Takenberg M, Lunkenbein S, Woche SK, Nimtz M, Berger RG: **Pfah2: a novel hydrophobin from the ascomycete *Paecilomyces farinosus*.** *Biotechnol Appl Biochem* 2013, **60**(2):147–154.
53. Wosten HA: **Hydrophobins: multipurpose proteins.** *Annu Rev Microbiol* 2001, **55**:625–646.
54. Bayry J, Aimanian V, Guijarro JJ, Sunde M, Latge JP: **Hydrophobins—unique fungal proteins.** *PLoS Pathog* 2012, **8**(5):e1002700.
55. De Oliveira AL, Gallo M, Pazzagli L, Benedetti CE, Cappugi G, Scala A, Pantera B, Spisni A, Pertinhez TA, Cicero DO: **The structure of the elicitor cerato-platanin (CP), the first member of the CP fungal protein family, reveals a double psi beta-barrel fold and carbohydrate binding.** *J Biol Chem* 2011, **286**(20):17560–17568.
56. Baccelli I, Comparini C, Bettini PP, Martellini F, Ruocco M, Pazzagli L, Bernardi R, Scala A: **The expression of the cerato-platanin gene is related to hyphal growth and chlamydospores formation in *Ceratocystis platani*.** *Fems Microbiol Lett* 2012, **327**(2):155–163.
57. Zaparoli G, Cabrera OG, Medrano FJ, Tiburcio R, Lacerda G, Pereira GG: **Identification of a second family of genes in *Moniliophthora perniciosa*, the causal agent of witches' broom disease in cacao, encoding necrosis-inducing proteins similar to cerato-platanins.** *Mycol Res* 2009, **113**:61–72.
58. Lombardi L, Faoro F, Luti S, Baccelli I, Martellini F, Bernardi R, Picciarelli P, Scala A, Pazzagli L: **Differential timing of defense-related responses induced by cerato-platanin and cerato-populin, two non-catalytic fungal elicitors.** *Physiol Plant* 2013, **149**:408–421.
59. Yang Y, Zhang H, Li G, Li W, Wang X, Song F: **Ectopic expression of MgSM1, a Cerato-platanin family protein from *Magnaporthe grisea*, confers broad-spectrum disease resistance in *Arabidopsis*.** *Plant Biotechnol J* 2009, **7**(8):763–777.
60. Bhadauria V, Banniza S, Vandenberg A, Selvaraj G, Wei Y: **EST mining identifies proteins putatively secreted by the anthracnose pathogen *Colletotrichum truncatum*.** *Bmc Genomics* 2011, **12**:327.
61. Frischmann A, Neudl S, Gaderer R, Bonazza K, Zach S, Gruber S, Spadiut O, Friedbacher G, Grothe H, Seidl-Seiboth V: **Self-assembly at air/water interfaces and carbohydrate binding properties of the small secreted protein EPL1 from the fungus *Trichoderma atroviride*.** *J Biol Chem* 2013, **288**(6):4278–4287.
62. Jeong JS, Mitchell TK, Dean RA: **The magnaporthe grisea snodprot1 homolog, MSP1, is required for virulence.** *Fems Microbiol Lett* 2007, **273**(2):157–165.
63. Peter M, Courty PE, Kohler A, Delaruelle C, Martin D, Tagu D, Frey-Klett P, Duplessis S, Chalot M, Podila G, Martin F: **Analysis of expressed sequence tags from the ectomycorrhizal basidiomycetes *Laccaria bicolor* and *Pisolithus microcarpus*.** *New Phytol* 2003, **159**(1):117–129.
64. Cosgrove DJ: **Loosening of plant cell walls by expansins.** *Nature* 2000, **407**(6802):321–326.
65. Quiroz-Castaneda RE, Martínez-Anaya C, Cuervo-Soto LI, Segovia L, Folch-Mallol JL: **Loosenin, a novel protein with cellulose-disrupting activity from *Bjerkandera adusta*.** *Microb Cell Fact* 2011, **10**:8.
66. Brotman Y, Briff E, Viterbo A, Chet I: **Role of swollenin, an expansin-like protein from *Trichoderma*, in plant root colonization.** *Plant Physiol* 2008, **147**(2):779–789.
67. Yamada M, Sakuraba S, Shibata K, Taguchi G, Inatomi S, Okazaki M, Shimosaka M: **Isolation and analysis of genes specifically expressed during fruiting body development in the basidiomycete *Flammulina velutipes* by fluorescence differential display.** *Fems Microbiol Lett* 2006, **254**(1):165–172.
68. Rincónes J, Scarpari LM, Carazzolle MF, Mondego JMC, Formighieri EF, Barau JG, Costa GGL, Carraro DM, Brentani HP, Vilas-Boas LA, de Oliveira BV, Sabha M, Dias R, Cascardo JM, Azevedo RA, Meinhardt LW, Pereira GA: **Differential gene expression between the biotrophic-like and saprotrophic mycelia of the witches' broom pathogen *Moniliophthora perniciosa*.** *Mol Plant Microbe In* 2008, **21**(7):891–908.
69. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**(5):821–829.
70. Sommer DD, Delcher AL, Salzberg SL, Pop M: **Minimus: a fast, lightweight genome assembler.** *BMC Bioinforma* 2007, **8**:64.
71. Ter-Hovhannisyantsyan V, Lomsadze A, Chernoff YO, Borodovsky M: **Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training.** *Genome Res* 2008, **18**(12):1979–1990.
72. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B: **AUGUSTUS: ab initio prediction of alternative transcripts.** *Nucleic Acids Res* 2006, **34**(Web Server issue):W435–W439.
73. Stanke M, Tzvetkova A, Morgenstern B: **AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome.** *Genome Biol* 2006, **7**(Suppl 1):S11 11–18.
74. Slater GS, Birney E: **Automated generation of heuristics for biological sequence comparison.** *BMC Bioinforma* 2005, **6**:31.
75. Borodovsky M, Lomsadze A, Ivanov N, Mills R: **Eukaryotic gene prediction using GeneMark.hmm.** *Curr Protoc Bioinformatics* 2003, **Chapter 4**:Unit4.6.
76. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR: **Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments.** *Genome Biol* 2008, **9**(1):R7.
77. Koski LB, Gray MW, Lang BF, Burger G: **AutoFACT: an automatic functional annotation and classification tool.** *BMC Bioinforma* 2005, **6**:151.
78. Szek BE, Huang H, McGarvey P, Mazumder R, Wu CH: **UniRef: comprehensive and non-redundant UniProt reference clusters.** *Bioinformatics* 2007, **23**(10):1282–1288.
79. Bateman A, Birney E, Cerruti L, Durbin R, Ewiler L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL: **The Pfam protein families database.** *Nucleic Acids Res* 2002, **30**(1):276–280.
80. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**(1):27–30.
81. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674–3676.
82. McCarthy EM, McDonald JF: **LTR\_STRUC: a novel search and identification program for LTR retrotransposons.** *Bioinformatics* 2003, **19**(3):362–367.
83. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: **Repbase update, a database of eukaryotic repetitive elements.** *Cytogenet Genome Res* 2005, **110**(1-4):462–467.
84. Bailey B, Bae H, Strem M, Samuels G, Evans H, Thomas S, Holmes K: **Molecular responses resulting from the endophytic association between *Trichoderma* species and cocoa (*Theobroma cacao*).** *Phytopathology* 2005, **95**(6):S5–S5.
85. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**(3):R25.
86. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**(10):R106.
87. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-Delta Delta C(T)</sup> method.** *Methods* 2001, **25**(4):402–408.
88. Schmittgen TD, Livak KJ: **Analyzing real-time PCR data by the comparative C(T) method.** *Nat Protoc* 2008, **3**(6):1101–1108.
89. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic Acids Res* 2001, **29**(9):e45.
90. Petersen TN, Brunak S, Von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods* 2011, **8**(10):785–786.
91. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**(3):567–580.

92. Sonnhammer EL, Von Heijne G, Krogh A: **A hidden Markov model for predicting transmembrane helices in protein sequences.** *Proc Int Conf Intell Syst Mol Biol* 1998, **6**:175–182.
93. Frieman MB, Cormack BP: **Multiple sequence signals determine the distribution of glycosylphosphatidylinositol proteins between the plasma membrane and cell wall in *Saccharomyces cerevisiae*.** *Microbiology* 2004, **150**(Pt 10):3105–3114.
94. Frieman MB, Cormack BP: **The omega-site sequence of glycosylphosphatidylinositol-anchored proteins in *Saccharomyces cerevisiae* can determine distribution between the membrane and the cell wall.** *Mol Microbiol* 2003, **50**(3):883–896.
95. Poisson G, Chauve C, Chen X, Bergeron A: **FragAnchor: a large-scale predictor of glycosylphosphatidylinositol anchors in eukaryote protein sequences by qualitative scoring.** *Genomics Proteomics Bioinformatics* 2007, **5**(2):121–130.
96. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes.** *Genome Biol* 2004, **5**(2):R12.
97. Delcher AL, Phillippy A, Carlton J, Salzberg SL: **Fast algorithms for large-scale genome alignment and comparison.** *Nucleic Acids Res* 2002, **30**(11):2478–2483.

doi:10.1186/1471-2164-15-164

**Cite this article as:** Meinhardt 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**:164.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

