



**HAL**  
open science

## Overview of the functional virulent genome of the coffee leaf rust pathogen *Hemileia vastatrix* with an emphasis on early stages of infection

Pedro Talhinhos, Helena G. Azinheira, Bruno Vieira, Andreia Loureiro, Silvia Tavares, Dora Batista, Emmanuelle Morin, Anne-Sophie Petitot, Octavio S. Paulo, Julie Poulain, et al.

### ► To cite this version:

Pedro Talhinhos, Helena G. Azinheira, Bruno Vieira, Andreia Loureiro, Silvia Tavares, et al.. Overview of the functional virulent genome of the coffee leaf rust pathogen *Hemileia vastatrix* with an emphasis on early stages of infection. *Frontiers in Plant Science*, 2014, 5, 10.3389/fpls.2014.00088 . hal-01268536

**HAL Id: hal-01268536**

**<https://hal.science/hal-01268536>**

Submitted on 28 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Overview of the functional virulent genome of the coffee leaf rust pathogen *Hemileia vastatrix* with an emphasis on early stages of infection

Pedro Talhinhos<sup>1</sup>, Helena G. Azinheira<sup>1\*</sup>, Bruno Vieira<sup>2</sup>, Andreia Loureiro<sup>1</sup>, Sílvia Tavares<sup>1</sup>, Dora Batista<sup>1</sup>, Emmanuelle Morin<sup>3,4</sup>, Anne-Sophie Petitot<sup>5</sup>, Octávio S. Paulo<sup>2</sup>, Julie Poulain<sup>6</sup>, Corinne Da Silva<sup>6</sup>, Sébastien Duplessis<sup>3,4</sup>, Maria do Céu Silva<sup>1</sup> and Diana Fernandez<sup>5</sup>

<sup>1</sup> Centro de Investigação das Ferrugens do Cafeeiro/BioTrop/Instituto de Investigação Científica Tropical, Oeiras, Portugal

<sup>2</sup> Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal

<sup>3</sup> Institut National de la Recherche Agronomique, Centre INRA Nancy Lorraine, UMR 1136 INRA/Université de Lorraine Interactions Arbres/Micro-organismes, Champenoux, France

<sup>4</sup> Université de Lorraine, UMR 1136 INRA/Université de Lorraine Interactions Arbres/Micro-organismes, Faculté des Sciences et Technologies, Vandoeuvre-lès-Nancy, France

<sup>5</sup> Institut de Recherche pour le Développement, UMR 186 IRD-Cirad-UM2 Résistance des Plantes aux Bioagresseurs, Montpellier, France

<sup>6</sup> Genoscope, Centre National de Séquençage, Commissariat à l'Énergie Atomique, Institut de Génétique, Evry, France

## Edited by:

David L. Joly, Université de Moncton, Canada

## Reviewed by:

Hossein Borhan, Agriculture and Agri-Food Canada, Canada  
Ralf Thomas Voegelé, Universität Hohenheim, Germany

## \*Correspondence:

Helena G. Azinheira, Centro de Investigação das Ferrugens do Cafeeiro, BioTrop, Instituto de Investigação Científica Tropical, Quinta do Marquês, 2784-505 Oeiras, Portugal  
e-mail: hgazinheira@gmail.com

*Hemileia vastatrix* is the causal agent of coffee leaf rust, the most important disease of coffee *Arabica*. In this work, a 454-pyrosequencing transcriptome analysis of *H. vastatrix* germinating urediniospores (gU) and appressoria (Ap) was performed and compared to previously published *in planta* haustoria-rich (H) data. A total of 9234 transcripts were identified and annotated. Ca. 50% of these transcripts showed no significant homology to international databases. Only 784 sequences were shared by the three conditions, and 75% were exclusive of either gU (2146), Ap (1479) or H (3270). Relative transcript abundance and RT-qPCR analyses for a selection of genes indicated a particularly active metabolism, translational activity and production of new structures in the appressoria and intense signaling, transport, secretory activity and cellular multiplication in the germinating urediniospores, suggesting the onset of a plant-fungus dialogue as early as at the germ tube stage. Gene expression related to the production of carbohydrate-active enzymes and accumulation of glycerol in germinating urediniospores and appressoria suggests that combined lytic and physical mechanisms are involved in appressoria-mediated penetration. Besides contributing to the characterization of molecular processes leading to appressoria-mediated infection by rust fungi, these results point toward the identification of new *H. vastatrix* candidate virulence factors, with 516 genes predicted to encode secreted proteins.

**Keywords:** appressorium, coffee leaf rust, germinating urediniospore, haustorium, pyrosequencing, transcriptome

## INTRODUCTION

Rust diseases have been a long standing threat for centuries and have reshaped cultivation of crops and breeding strategies. Coffee leaf rust caused by *Hemileia vastatrix* Berk & Broome is the major disease of *Arabica* cultivated coffees (*Coffea arabica* L.) (Silva et al., 2006). *H. vastatrix* is considered as one of the most primitive phylogenetic lineages of the Pucciniales (Aime, 2006; Silva et al., 2012) and has no alternate host known so far. Since the 19th century, when it caused suppression of the coffee cultivation in Sri Lanka, the disease gained a worldwide distribution, reaching nearly all regions of the world where coffee is grown with severe economical damages. Breeding and selection of coffee resistant genotypes to different fungal races from several parts of the world has been successful (Silva et al., 2006), but as a consequence of the high adaptive potential of the pathogen, the emergence of new rust pathotypes and the corresponding breakdown of resistance has been observed in many improved coffee varieties in several countries (Várzea and Marques, 2005; Diniz et al., 2012; Cressey,

2013). Thus, currently coffee leaf rust still stands as the major constraint to *Arabica* coffee production.

During infection of their hosts, rust fungi differentiate several specialized infection structures such as germ tubes, appressoria, stomatal vesicles, infection hyphae, haustoria, and spore-forming cells. Until recently, most of the biological knowledge gained at the molecular level on rust fungi was derived from EST sequencing, mainly from ungerminated and germinating urediniospores, rust-infected tissues, isolated haustoria and some spore types at other stages of their complex life cycle (for a review see Duplessis et al., 2012; Fernandez et al., 2013). However, some differentiation stages are not sufficiently covered yet and lack description and information, such as appressoria formation (Fernandez et al., 2013). For instance, two studies conducted in *Phakospora pachyrhizi* and *Puccinia triticina* reported that a high proportion of genes of unknown functions were expressed at the appressorial stage (Hu et al., 2007; Stone et al., 2012). Appressoria may be also differentiated by other pathogenic fungi, enabling

host cuticle penetration through physical and/or chemical mechanisms. Key features of these specialized structures include the production of an extracellular matrix for adhesion to the surface, the accumulation of molar concentrations of glycerol for generating turgor pressure and the differentiation of a penetration hypha (Deising et al., 2000). Appressoria differentiated from urediniospores typically form over host stomata, and a penetration hypha is subsequently formed at the base of the appressorium to invade the substomatic chamber. There are evidences that mechanical pressure (about 0.35 MPa) is exerted by the penetration hypha when penetrating the stoma (Terhune et al., 1993). This is considerably less than the pressure exerted by some fungi that penetrate directly through the cuticle, such as *Magnaporthe oryzae* or *Colletotrichum* spp. (Howard et al., 1991; Chen et al., 2004), but enough to distort stomatal guard cell lips (Terhune et al., 1993). Nevertheless, rust fungi must also possess machinery for lytic penetration of host cuticle and cell wall, since hyphae produced by germinating basidiospores are capable of direct penetration into host epidermal cells (Voegele et al., 2009).

Until very recently, no genomic resources were available for *H. vastatrix*. After several years of lagging behind other rust fungi on genomic research, Fernandez et al. (2012) reported on the 454-transcriptome sequencing of rust-infected coffee leaves. This study generated 22,774 contigs of which 30% were assigned to *H. vastatrix*. Analysis of these *in planta* expressed sequence tags (ESTs) revealed that the majority (60%) had no homology in public genomic databases, representing potential coffee rust-specific genes. Nevertheless, *H. vastatrix* candidate effectors likely related to host infection and orthologous to other rust fungi, were identified among 382 predicted secreted proteins (Fernandez et al., 2012). Still, there is no knowledge of transcripts expressed at early stages of infection that could provide a more integrative scenario on the molecular mechanisms governing this pathosystem.

Complementing the knowledge gained into the *in planta* transcriptome of coffee rust, here we report on the in-depth transcriptome analysis of *H. vastatrix* by 454-based RNA-Seq during urediniospore germination and appressorium formation, two early and key stages of infection. Comparison of these specific stages with infected leaves allows an integrative characterization of transcript expression profiles during the course of biotrophic growth and infection. In particular, the identification of genes related with appressorium formation leads to novel insights into a stage that has been poorly described at the molecular level.

## MATERIALS AND METHODS

### BIOLOGICAL MATERIAL, RNA ISOLATION AND cDNA SYNTHESIS

*Hemileia vastatrix* isolate CIFC 178a (race XIV: genotype  $v_2v_3v_4v_5$ ) was multiplied on its differential host plant (*C. arabica* accession CIFC H147/1, carrying the resistance factors  $S_{H2}$ ,  $S_{H3}$ ,  $S_{H4}$  and  $S_{H5}$ ). An *in vitro* method was used to produce germinating urediniospores and appressoria to ensure the generation of cDNA libraries with no contaminating plant sequences (Azinheira et al., 2001; Vieira et al., 2012). For the germinating urediniospores sample (gU), 19 mg of spores were spread in sterile distilled water in Petri dishes and incubated for 18 h at 24°C under darkness. For the appressoria sample (Ap), 15 mg of spores were spread over oil-collodion membranes (Vieira et al., 2012) in

Petri dishes, sprayed with water and incubated for 24 h at 24°C and 100% relative humidity, under darkness. For an accurate sample characterization, urediniospore germination and appressoria formation were quantified, showing that the germinating urediniospores sample (gU) comprised over 50% of germinating urediniospores. The appressoria sample (Ap) comprised over 60% of germinating urediniospores with appressoria. These are considered rather fair rates for *H. vastatrix* (Azinheira et al., 2001).

Samples gU and Ap were harvested, immediately frozen in liquid nitrogen and the RNA was isolated from each sample with the RNeasy Plant minikit (Qiagen, Hilden, Germany), including an in-solution DNase treatment following the manufacturer's instructions. RNA concentration and integrity were evaluated by spectrometry (Lambda EZ201, Perkin-Elmer, Waltham-MA, USA) and capillary electrophoresis (Bioanalyzer 2100, Agilent, Santa Clara-CA, USA) respectively. Following the SMARTer Pico PCR cDNA Synthesis Kit (Clontech, Saint-Germain-en-Laye, France) protocol, cDNA were synthesized from 1 µg total RNA using SMARTScribe Reverse Transcriptase (Clontech) and amplified using Advantage 2 Polymerase (Clontech). cDNA fragments, which ranged between 500 and 3000 bp, were purified with the NucleoSpin Extract II kit (Macherey-Nagel, Düren, Germany) and their quality and concentration were evaluated by electrophoresis.

### PYROSEQUENCING AND ASSEMBLY OF 454 READS

For each sample, 20 µg cDNA was used for 454-pyrosequencing run on half of a picotitre plate on a Genome Sequencer FLX System using long-read GS FLX Titanium chemistry (Roche; www.454.com) at the Genoscope (Centre National de Séquençage, Evry, France; www.genoscope.cns.fr) following standard procedures recommended by Roche.

Raw sequences obtained for gU and Ap samples were assembled into contigs using Newbler 2.5 (Roche) with default parameters. For comparative purposes, the MIRA 3.2 assembler (<http://sourceforge.net/apps/mediawiki/mira-assembler>) was also employed. The relative abundance (Ra) of transcripts was calculated as the ratio between the number of 454 reads per contig and the length of the assembled contig (Vega-Arreguín et al., 2009).

### BIOINFORMATIC ANALYSIS OF TRANSCRIPTS

As previously described (Fernandez et al., 2012), sequence homology searches were performed against several databases: the NCBI non-redundant (nr) nucleotide and protein databases (www.ncbi.nlm.nih.gov), the genome sequences of *Melampsora larici-populina* and *Puccinia* spp. (Cantu et al., 2011; Duplessis et al., 2011a; www.jgi.doe.gov and www.broadinstitute.org, respectively); the euKaryotic Orthologous Group (KOG) database (Tatusov, 2003); the Pathogen-Host Interaction (PHI-base v3.2) reference database (Winnenburg et al., 2007; www.phi-base.org); the Phytopathogenic Fungi and Oomycete EST Database (COGEME v1.6; Soanes and Talbot, 2006); and a Pucciniales EST database (168,199 ESTs retrieved from GenBank in November 2012—unchanged number as of December 2013). Besides these, 16,831 transcripts from the *M. larici-populina*

frozen gene catalog (<http://genome.jgi-psf.org/Mellp1/Mellp1.download ftp.html>) and 20,567 (*P. graminis* f. sp. *tritici*) and 11,638 (*P. triticina*) from the *Puccinia* spp. transcript catalogue ([http://www.broadinstitute.org/annotation/genome/puccinia\\_group/MultiDownloads.html](http://www.broadinstitute.org/annotation/genome/puccinia_group/MultiDownloads.html)) were also considered. Homology searches were performed using BLAST algorithms (Altschul et al., 1997) with a cut-off criterion ( $e$ -value  $< 10^{-5}$ ). For each search against a given database, only the best hit was considered. The assignment of 454-contig sequences into KOG functional categories was obtained using Reverse psi-BLAST (RPSBLAST; Altschul et al., 1997) against the KOG database.

Open reading frames (ORFs) were predicted with the translation tool getorf from the European Molecular Biology Open Software Suite (EMBOSS; <http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf>) using default parameters. ORFs below 18 amino acids were not considered. A secretome bioinformatics pipeline was employed to define a tentative set of secreted proteins encoded by *H. vastatrix* transcripts, using SignalP v4.0 (Petersen et al., 2011), TargetP v1.1 (Emanuelsson et al., 2000) and TMHMM v2.0 (Krogh et al., 2001).

The catalytic and carbohydrate-binding modules (or functional domains) of enzymes that degrade, modify, or create glycosidic bonds (carbohydrate-active enzymes—CAZymes) was investigated by blastp comparison of predicted polypeptides to the CAZymes database ([www.cazy.org](http://www.cazy.org); Cantarel et al., 2009) and to the CAZymes from *M. larici-populina* and *P. graminis* f. sp. *tritici* (Duplessis et al., 2011a). Similarly, proteins involved in membrane transport were investigated by blastp searches against predicted polypeptides in the Transporter Classification Database ([www.tcdb.org](http://www.tcdb.org); Saier et al., 2006, 2009).

Contigs from gU and Ap samples, as well as those predicted as fungal from a 21 day infected-coffee leaf sample (sample H; Fernandez et al., 2012), were compared using a best reciprocal BLAST hit approach with BioEdit 7.0.4.1 (Hall, 1999). Pairs of contigs with an  $e$ -value lower than  $10^{-30}$  were considered as representing the same transcript and assembled. Ra values were calculated for each transcript present in more than one library, and these values were compared across the libraries in order to evaluate variations in expression levels. For such,  $\tau$  values were calculated for each gene based on the normalized Ra values, in order to account for differences in library sizes. The expression specificity index ( $\tau$ ) is defined as  $\tau = \frac{\sum_{i=1}^n (1 - x_i)}{n - 1}$ , where  $n$  is the number of tissues and  $x_i$  is the expression profile component normalized by the maximal component value (Yanai et al., 2005). The genes with the most stable expression across the three libraries were selected (105 genes with  $\tau$  values below 0.25). Average Ra values were calculated among these genes for each library (0.02684 for gU, 0.03158 for Ap and 0.03039 for H) and Ra values for each contig in each library were normalized to the gU sample, following the strategy described by Ekblom et al. (2010) based on the guidelines provided by Mank et al. (2008).

### RT-qPCR

Germinating urediniospores (gU) and appressoria (Ap) samples for *H. vastatrix* isolate 178a were obtained as described above. *In planta* time course samples were collected at 18 h (mostly containing appressoria) and 1, 2, 3, 7, 14, and 21 days after

inoculation for the compatible interaction between isolate 178a and the *C. arabica* genotype H147/1, as previously described (Diniz et al., 2012; Vieira et al., 2012). Fungal germination, appressoria formation and the differentiation of infection structures *in planta* were monitored by light microscopy as previously described (Vieira et al., 2012). RNA extraction, cDNA synthesis and RT-qPCR experiments were performed as previously described (Vieira et al., 2012), using Hv00099, 40S ribosomal protein and glyceraldehyde-3-phosphate dehydrogenase as reference genes (Vieira et al., 2011) and ungerminated urediniospores as the control sample. A set of 43 genes was selected for RT-qPCR analysis based on RNA-Seq expression profiles and assigned functions. Primers (Supplementary Data 1) were designed as previously described (Vieira et al., 2012).

## RESULTS AND DISCUSSION

### 454-PYROSEQUENCING DATA FOR GERMINATING UREDINIOSPORES AND APPRESSORIA SAMPLES

Given that different 454-pyrosequencing data assemblers are available and are known to generate diverse results (Kumar and Blaxter, 2010), MIRA 3.2 and Newbler 2.5 were compared in this study. In general, MIRA produced shorter and more numerous contigs. The overall homology scores of contigs to a Pucciniales EST database was better for the Newbler assembly (data not shown), suggesting a better quality of the assemblage which led us to use Newbler assembly in this study.

For samples gU and Ap (Table 1), a total of 455,807 sequence reads (113,404,366 nucleotides) was generated and assembled into 9108 contigs (4267 for gU and 3626 for Ap), with ca. 24% sequences remaining as either too short/low quality sequences (7%) or singletons (17%). Among those, 1214 contigs (13%) were  $< 100$  bp and not further considered in the analysis. The remaining 7894 contigs (Supplementary Data 2) had a mean length of 656 bp (Table 1), with 16% contigs larger than 1 kb (3.7% larger than 2 kb). Mean number of reads per contig was 41.0, with 11% contigs over 50 reads. Mean relative abundance (Ra) was 0.1153, with 18% contigs (1424) representing transcripts with a medium to high rate of expression ( $Ra > 0.05$ ).

In the absence of genomic information for *H. vastatrix*, contigs were compared to sequences deposited in databases (summary in Table 2 and results by contig listed in Supplementary Data 2), and 54% contigs had homology ( $e$ -value  $< 10^{-5}$ ) to the NCBI nr nucleotide database using blastn (Supplementary Data 2, columns G–I).

A total of 13,951 sequences obtained from the 21-days *H. vastatrix* infected-coffee leaf samples (H library) and previously predicted as of plant origin (Fernandez et al., 2012) were compared to the gU+Ap sequences, from which 22 showed an homology  $e$ -value below  $10^{-60}$  (19 of which had  $e$ -value of 0.0; Supplementary Data 3). This analysis showed that only 0.1% of the sequences predicted as of plant origin (Fernandez et al., 2012) were wrongly assigned to this class. Similarly, among 2060 contigs previously classified as “not attributed/not resolved,” 28 had homology to gU+Ap sequences with an  $e$ -value below  $10^{-60}$  (21 of which had  $e$ -value of 0.0). These 50 contigs ( $e$ -value  $< 10^{-60}$ ) were incorporated in the present study, together with the 4415 fungal contigs initially identified in the H library,

summing a total of 4465 H contigs (Fernandez et al., 2012) to our dataset.

**COMPARISON TO THE *IN PLANTA* EXPRESSED FUNGAL SEQUENCES**

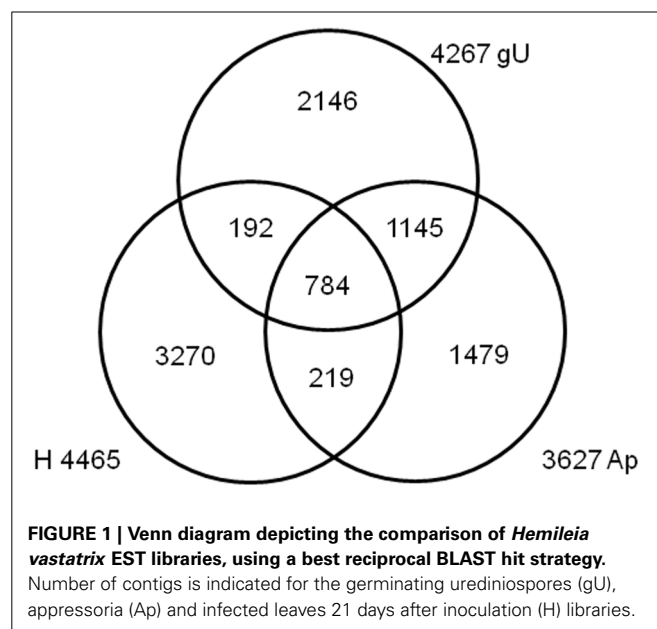
A best reciprocal BLAST strategy was used to compare the contigs from both gU and Ap libraries, as well as the fungal contigs from the H library (Fernandez et al., 2012). This enabled the identification and re-assembly of 784 sequences shared by the three libraries, 1145 shared only by gU and Ap, 219 by Ap and H and 192 by gU and H (Figure 1 and Supplementary Data 4, columns

B–D). The remaining 6894 sequences (75%) are exclusive of each library. Altogether, 9234 unique *H. vastatrix* sequences were identified, which represents >50% of the total number of genes predicted from the genomes of *M. larici-populina* (16,399 genes) and *P. graminis* f. sp. *tritici* (17,773 genes) (Duplessis et al., 2011a). In order to further ascertain a measure of the genome coverage obtained in this study, we compared each of the three libraries (gU, Ap, and H) separately, along with the total set of 9234 sequences, against the FUNYBASE database containing 246 families of single-copy orthologs obtained from 21 genomes (core

**Table 1 | Descriptive statistics for *Hemileia vastatrix* 454 pyrosequenced cDNA libraries of germinated urediniospores (gU) and appressoria (Ap).**

Library	gU	Ap
Number of bases	67773266	45631100
Number of sequences	269199	186608
Mean size of reads (bp)	251.8	244.5
Number of contigs	4267	3626
Mean size of contigs (bp)	676	632
Size of contigs (bp)*	188/546/1293/ 3754	192/530/1139/ 4860
Mean number of reads per contig	48.5	32.2
Reads per contig*	4/10/63/3326	4/9/52/2077
Mean relative abundance (Ra)	0.1456	0.0797
Relative abundance (Ra)*	0.0087/0.0190/ 0.0967/26.72	0.0085/0.0187/ 0.0777/14.41

\*Values correspond to 10/50/90/100 percentiles.



**Table 2 | *Hemileia vastatrix* transcript homology in databases and KOG functional categories classification.**

Library	gU	Ap	Total	% of all gU contigs	% of all Ap contigs	% of all contigs
NCBI nr_blastn	2119	2126	4245	49.66	58.62	53.78
Mlp_genome_tblastx	2356	2159	4515	55.21	59.53	57.20
Pgt_genome_tblastx	2294	2099	4393	53.76	57.87	55.65
Pt_genome_tblastx	2410	2171	4581	56.48	59.86	58.03
Pst_genome_tblastx	2408	2159	4567	56.43	59.53	57.85
EST_Pucciniales_tblastx	2507	2334	4841	58.75	64.35	61.33
SwissProt_blastx	1301	1362	2663	30.49	37.55	33.73
PHIbase_tblastx	482	444	926	11.30	12.24	11.73
COGEME_tblastx	2016	2042	4058	47.25	56.30	51.41
KOG	1691	1690	3381	39.63	46.59	42.83
Posttranslational modification, protein turnover, chaperones				12.5	13.6	
Translation, ribosomal structure and biogenesis				12.4	19.7	
Intracellular trafficking, secretion, and vesicular transport				8.2	5.7	
Energy production and conversion				7.9	9.1	
Signal transduction mechanisms				7.5	6.3	
Lipid transport and metabolism				5.7	6.4	

Summary of the number and percentage of hits in homology searches of the *H. vastatrix* germinated urediniospores (gU) and appressoria (Ap) contig libraries against the NCBI nr database (NCBI nr\_blastn), rust genomic and transcriptomic databases (Mlp\_genome\_tblastx, Pgt\_genome\_tblastx, Pt\_genome\_tblastx, Pst\_genome\_tblastx, EST\_Pucciniales\_tblastx), the SwissProt database (SwissProt\_blastx) and functional databases (PHIbase\_tblastx, COGEME\_tblastx, KOG).



fungal genes) (Marthey et al., 2008). Each individual library contained only about half of the 246 genes (39% in H and 55% both in Ap and gU; data not shown), but the gU+Ap+H library included 174 (71%) of those core genes (Supplementary Data 4, columns BN-BP). These results indicate that the 9234 unique *H. vastatrix* transcripts provide a significant coverage of the *H. vastatrix* functional genome.

According to their RNA-Seq expression values and assigned functions, the expression profiles of a set of 43 genes was analyzed by RT-qPCR along the time course of a compatible interaction (Table 3).

## GENE FUNCTION

Over 72% of the 9234 *H. vastatrix* transcripts had no specific KOG category assigned (No hits, “Function unknown” or “General function prediction only”; Table 4 and Supplementary Data 4, columns BK-BM). Within the remaining transcripts, the most represented KOG categories are “Translation, ribosomal structure and biogenesis” (14%) and “Post-translational modification, protein turnover, chaperones” (12%), while other nine categories represent 5–8% each.

A total of 4040 transcripts (44% of total) presented homologues against the NCBI non-redundant nucleotide database (Supplementary Data 4, columns V–X). A similar value (45%) was obtained by comparison (tblastx) with the Pucciniales EST database (Supplementary Data 4, columns AY-BA), with the most frequent organisms being *M. larici-populina*, *P. triticina*, *P. graminis* f. sp. *tritici* and *P. pachyrhizi* in similar proportions to those reported for the gU and Ap libraries (Supplementary Data 4, columns Y-AN). A total of 2992 transcripts (32%) have homology to all available rust genome sequences, suggesting that the corresponding genes are conserved among the Pucciniales. Only 16 transcripts showed homology to *P. graminis* f. sp. *tritici* or *P. triticina* mitochondrial sequences (Supplementary Data 4, columns AS-AX). A total of 141 and 148 transcripts did not show homology to *M. larici-populina* and *P. striiformis* gene models respectively, although showing significant homology to their genome sequences, which could indicate actual genes that were not predicted by automatic annotations in the corresponding genomes (Supplementary Data 4, columns AB-AE and AO-AR). Interestingly, 4707 transcripts (51% of total) showed no homology to the rust genes identified in genome sequences or EST databases (at a cut off  $e$ -value of  $10^{-10}$ ), suggesting they may correspond to highly divergent or specific *H. vastatrix* genes. In fact, among these, only 3.2% have a specific KOG category assigned, with an overrepresentation of categories involved in cellular structure, nucleic acid activity and signaling (“Cytoskeleton,” “RNA processing and modification,” “Transcription” and “Signal transduction mechanisms”).

A total of 3573 transcripts showed homology to annotated fungal genes listed in the COGEME database (Supplementary Data 4, columns BH-BJ), *Ustilago maydis* (21%), *Giberella* spp. (13%), and *M. oryzae* (13%) being the most represented species. Further, 588 transcripts showed homology to fungal pathogenicity-related genes listed in the PHI database (Supplementary Data 4, columns BE-BG; 94% of which also have homologues in the COGEME database), mostly from *M. oryzae* (19%), *Candida albicans* (18%),

and *Cryptococcus neoformans* (14%). About 70% of these 588 transcripts had specific KOG categories assigned, with categories such as “Cell cycle control, cell division, chromosome partitioning,” “Secondary metabolites biosynthesis, transport and catabolism” and “Signal transduction mechanisms” overrepresented as compared to KOG categories assigned to all genes (Table 4).

For each transcript, the size of the predicted polypeptide was compared to the size of the corresponding polypeptide in the *P. graminis* f. sp. *tritici* genome. Among the 3686 hits to the *P. graminis* f. sp. *tritici* genome, 24% were >90% the length of their orthologs (46% proteins were >50% size; Supplementary Data 4, columns BQ-BS).

As *H. vastatrix* 454 cDNA libraries were not normalized, the number of reads contained in each contig can be considered a relative expression level of each gene. For each contig, the number of reads was divided by the length of the contig, resulting in a Relative Abundance (Ra) index (Supplementary Data 4, columns G-O). Comparison among contigs from different libraries required a normalization step in order to account for differences in library sizes (Supplementary Data 4, columns P-S). Based on the comparative analysis of the expression levels identified in each library, nine different expression profiles were defined (Supplementary Data 4, column U). The most frequent profiles observed reflect the fact that 75% of contigs are exclusive of a single library, with 87% of the 9234 predicted transcripts presenting profiles 1, 2, or 3 (Table 5, row 4).

The analysis of relative abundance values according to the KOG category of each gene (Table 5) suggests a particularly active metabolism, translational activity and production of new structures in the Ap sample and both intense signaling and secretory activity and cellular multiplication in germinating urediniospores. In the H sample, over-represented KOG categories suggest intense signaling and nutrient uptake from the host to the fungus, as previously pointed out (Fernandez et al., 2012).

## SECRETED PROTEINS

A total of 467 putative secreted proteins were predicted with a secretion prediction pipeline composed of the SignalP, TargetP and TMHMM programmes (Supplementary Data 4, columns BT to CI for SignalP, CJ to CO for TargetP, CP to CS for TMHMM, CT for TMHMM vs. SignalP comparison and CU for final secretion prediction score). Besides these, other transcripts showing high homology ( $e$ -value <  $10^{-30}$ ) to the *M. larici-populina* or the *P. graminis* f. sp. *tritici* predicted secreted proteins (Duplessis et al., 2011a) were also selected (Supplementary Data 4, column CV) even if not detected by the pipeline. Since these sequences were shorter than their orthologs, the signal peptide may be lacking from sequence. From this list of 516 transcript encoding putative secreted proteins, 87 and 70% entries presented less than 300 amino acids, and 200 amino acids, respectively (Supplementary Data 4, column BR). Also, 82 of these translated gene sequences are highly enriched in cysteine residues (5–15% of all amino acids; Supplementary Data 4, column CW), the vast majority of which (78) is less than 200 aa, similar to what was reported for *M. larici-populina* small secreted proteins (Hacquard et al., 2012). Nearly 60% of these 82 sequences contain

**Table 3 | Heatmap of *Hemileia vastatrix* genes expression profiles in germinating urediniospores (gU), *in vitro*-obtained appressoria (Ap) and *in planta* samples (from 18h to 21d after inoculation) obtained by RT-qPCR by comparison with 454 pyrosequencing-derived relative abundance transcript levels.**

Gene	gU	Ap	18h	1d	2d	3d	7d	14d	21d	KOG function	Annotation (according to Supplementary Data 4)
Hv00125	1.16367	19.3825	2.11073	0.40276	0.26266	0.0869	0.21031	0.0255	0.00246	Signaling	Fungal conserved hypothetical protein similar to gas1
Hv00156	0.70207	19.7096	9.06186	9.12287	2.74469	1.95129	4.4071	0.3216	0.02838	Metabolism	Hexokinase
Hv00175	0.45556	0.29447	3.12743	4.31055	9411.5	2563.85	10548	175.507	26.569	Metabolism	Long-chain fatty acid CoA ligase
Hv00191	0.03	56.1	36.3	19.5	8.2	39.8	1.02	1.05	3.12	Metabolism	Acetyl-CoA C-acyltransferase
Hv00297	0.51073	2.35799	0.06277	0.01619	0.52742	0.08316	0.13036	0.02135	0.00623	Oxidative stress	Manganese superoxide dismutase
Hv00334	0.66625	1.31543	0.24244	0.13565	0.33702	0.06125	0.11578	0.03311	0.01143	Signaling	GTP-binding protein ypt1
Hv00373	2.04393	2.14363	1.9811	1.89263	1.972	1.81027	2.05546	1.59249	1.60879	Signaling	Mitogen-activated protein kinase (MAPK)
Hv00409	35.2036	216.796	1.14688	0.81432	2.465	0.72411	1.46818	0.00581	0.00044	Signaling	Fungal conserved hypothetical protein similar to Gas2
Hv00489	6.33369	48.0067	1.78482	1.64577	0.00442	12.5255	0.67139	0.01991	0.01893	Signaling	Fungal conserved hypothetical protein similar to Gas1
Hv00491	0.74366	19.0622	0.90118	0.66227	70.355	0.32046	1.68915	0.85444	0.01163	Metabolism	Malate dehydrogenase
Hv00616	1.11902	1.31567	0.99129	1.26598	0.986	0.92218	0.94056	0.89716	1.10502	Signaling	Mitogen-activated protein kinase, Hog-1
Hv00622	1.24335	12.1162	3.22935	6.22324	5.49718	0.76849	0.93931	0.2717	0.08024	Metabolism	Acyl-CoA oxidase
Hv00643	2.24832	24.0975	1.78299	0.86742	0.16458	0.17574	0.03718	0.01411	0.00131	Cell structure	Chitinase/extracellular matrix protein
Hv00704	1.02253	12.4865	0.01606	0.01829	0.03554	0.00998	0.016	0.21356	0.19162	Metabolism	Family 37 glycoside hydrolase (neutral trehalase)
Hv00717	0.4797	16.8843	1.80606	1.85386	0.02887	0.00811	0.52011	0.00046	8E-05	Metabolism	Fungal conserved hypothetical protein similar to malate synthase A
Hv00898	0.63376	3.81303	2.08204	2.04629	2.14348	0.25311	8.06796	0.06247	0.01499	Signaling	Fungal conserved hypothetical protein similar to catalytic subunit of cyclic AMP-dependent protein kinase
Hv01055	0.17642	67.4803	0.17342	0.19745	0.38368	0.10773	2.01277	0.27596	0.00106	Metabolism	Enoyl-CoA hydratase
Hv01133	1.16988	3.04568	0.9911	0.80173	0.34996	0.08809	0.19209	0.10159	0.02631	Metabolism	Citrate synthase (putative peroxysomal)
Hv01266	1.6161	9.49352	1.90843	0.73394	0.69469	0.079	0.11407	0.00913	0.00477	Metabolism	Isocitrate lyase
Hv01268	6.62721	6.03687	3.0546	1.30407	1.81178	0.39949	0.42311	0.02548	0.00541	Oxidative stress	Manganese superoxide dismutase
Hv01400	21.7861	32.1225	0.39251	0.44691	0.86842	0.24384	0.39086	0.01384	0.39722	Metabolism	Fungal conserved hypothetical protein similar to glycerol-3-phosphatase 1
Hv01431	0.58351	23.194	3.98369	0.01802	0.03502	1.14611	0.01576	0.60223	0.00794	Signaling	Catalytic subunit of cyclic AMP-dependent protein kinase
Hv01534	2.43435	54.6528	99.8917	39.3294	0.03511	4.09924	0.0158	7.29429	0.07934	Transport	Copper-transporting atpase 1
Hv01548	9.07846	12.0714	12.7266	5.93327	3.30629	2.60159	0.74004	1.02832	0.58709	Signaling	Adenylyl cyclase
Hv01594	0.63282	0.05241	0.3587	0.40842	2805.94	0.81017	766.871	18.6638	8.90275	Metabolism	Carnitine palmitoyltransferase II protein
Hv01628	3.76138	38.715	0.02326	0.02649	0.05147	26.5421	7.63213	0.00082	1.58465	Metabolism	Acyl-CoA dehydrogenase
Hv01629	1.71268	5.68132	0.47669	0.6158	0.21829	0.0037	0.06997	0.06322	0.01645	Metabolism	Short-chain specific acyl-coa dehydrogenase, mitochondrial precursor
Hv01805	0.60063	3.41873	0.69978	0.79676	1334.16	6916.29	171696	4467.17	303.319	Metabolism	Fungal conserved hypothetical protein similar to enoyl-CoA hydratase
Hv01932	0.61077	2.80827	3.4634	1.02717	1.43976	0.34943	1.16013	0.43472	0.16774	Signaling	Mitogen activated protein kinase kinase, map2k-ste7
Hv01988	0.58754	5.48309	68.5928	69.5714	2221.03	2952.97	133758	2386.12	256.303	Metabolism	Carnitine palmitoyltransferase II protein
Hv02022	0.3602	44.8358	10.3906	3.76459	0.04084	0.93193	2.73293	0.62952	0.04667	Signaling	G-protein beta subunit
Hv04402	0.69534	12.3966	0.01395	0.01589	0.03087	0.00867	0.01389	0.00049	0.08973	Metabolism	Family 20 Glycosyltransferase (Trehalose-phosphate synthase TPS, UDP-forming)
Hv06436	0.55549	31.6094	2.61774	1.81984	0.01325	0.56496	0.00596	0.23063	0.11324	Signaling	cAMP-dependent protein kinase type 3
Hv06448	1.1389	6.48226	3.70593	4.01523	0.04881	0.01371	0.02197	0.00078	0.0225	Metabolism	NAD-dependent glycerol-3-phosphate dehydrogenase
Hv06788	0.30129	7.70897	1.99418	1.18127	0.92582	0.0721	0.19353	0.17328	0.02286	Metabolism	Glycerol kinase

(Continued)

**Table 3 | Continued**

Gene	gU	Ap	18h	1d	2d	3d	7d	14d	21d	KOG function	Annotation (according to Supplementary Data 4)
Hv06883	3.62681	34.4684	2.06459	0.81996	1.73846	0.50857	1.0788	0.41016	0.71987	Signaling	Fungal conserved hypothetical protein similar to serine/threonine-protein kinase PRP4m
Hv07140	1	195.751	246.896	4713.12	1426.44	446.111	786.522	33.3738	0.07913	Signaling	Mitogen-activated protein kinase HOG1
Hv07174	1.79233	37.7162	2.69853	3.49281	0.02431	0.00683	0.01094	0.33983	0.00929	Signaling	G-protein beta subunit
Hv07400	1.24777	41.1467	11.1695	7.05488	4.73257	3.38405	13.2912	0.32332	0.02379	Metabolism	Carnitine palmitoyltransferase II protein
Hv08812	0.33774	114.518	4.75008	37.0662	113.31	39.9515	22.0792	0.16746	0.02904	Metabolism	Glycerol-3-phosphate dehydrogenase [NAD+]
Hv08827	1.03811	20.2036	1.9725	1.07745	1.87259	0.37207	1.24266	0.34945	0.54076	Signaling	Protein phosphatase 1 regulatory subunit 7
Hv08833	1.14907	1.23147	0.80187	0.41262	0.00392	0.0011	0.27852	0.40565	0.18488	Metabolism	Citrate synthase
Hv09049	4.05405	113.847	5.59448	2.96287	2.48302	0.10502	0.44865	0.35815	0.73447	Metabolism	Glycerol kinase

Values represent “fold change” related to the levels of expression in resting urediniospores (except for transcript 07140, where germinating urediniospores were used because of no amplification was obtained in resting urediniospores); Color scale: green to red denote lowest to highest expression levels across time points for each gene.

a [YFW]×C motif (Supplementary Data 4, column CX), while only 20% of the remaining 434 sequences (<5% cysteines) possess that motif. An overrepresentation of this particular motif was similarly observed for the small secreted proteins of the poplar rust fungus (Hacquard et al., 2012). Secreted proteins transcripts tend to present high relative expression values: while they represent 5.6% of all 9234 genes in this study, they represent 12–14% of genes with Ra > 1 (Table 5). Moreover, the sum of all Ra values for predicted secreted proteins is higher in gU than in Ap or H, although more genes were identified in H (Figure 2). In addition, 46 of these genes encoding predicted secreted proteins present homology to genes in the PHI database (Supplementary Data 5) whose mutants in various fungi exhibit either loss of pathogenicity or reduced virulence phenotype, 21 of which were up-regulated in the gU library, 13 in Ap and 11 in H according to Ra values.

Four transcripts (00303, 00357, 01043, and 04304) encoding predicted secreted proteins are orthologous of the rust transferred protein (RTP1) genes (Pretsch et al., 2013) from *U. fabae*, *M. occidentalis* or *M. medusae* f. sp. *deltoidis* (Supplementary Data 6). Orthologs of these four genes were also identified in *M. larici-populina* and *Puccinia* spp. genomes (Supplementary Data 7) and the overall similarity among genes is quite low (Supplementary Data 6, columns Z–AD). Three of these transcripts were previously identified in *H. vastatrix* transcripts (Fernandez et al., 2012; see Supplementary Data 6). Transcript 04304 was exclusively detected in the H library (Supplementary Data 6), corroborating the observations by Vieira et al. (2012) and the expression profile of RTP1 in *U. fabae* (Kemen et al., 2005). Transcript 01043 was detected in Ap and H, and transcripts 00303 and 00357 were identified in the three libraries. Different expression profiles could be observed, transcript 00303 being highly expressed in Ap, transcript 01043 more expressed in Ap and H and transcript 00357 showing similar Ra values across the three libraries. *H. vastatrix* RTP1 orthologs show distinct expression profiles for the different members of this single gene family. Such an observation confirms the very dynamic and specific transcriptional process at the gene family level that was reported for gene families encoding small secreted proteins of *M. larici-populina* during time course infection of poplar leaves (Duplessis et al., 2011b).

Some *H. vastatrix* transcripts are orthologous of haustorially expressed secreted proteins (HESP) identified in *Melampsora lini* (Dodds et al., 2006; Barrett et al., 2009) (Supplementary Data 7). Among these, HESP-178 is orthologous to the transcripts 01506 and 04456, detected respectively in gU and Ap, and in H libraries (Supplementary Data 8). HESP-379 is orthologous to transcript 00258, which was identified in the three libraries at decreasing levels of expression along the differentiation stages, confirming previous observations (Fernandez et al., 2012). HESP-767 is orthologous to transcript 09298 only identified in library Ap. No homology was detected to *Melampsora* spp. Avr genes (Dodds et al., 2004), presumably because of their poor conservation across the Pucciniales (Catanzariti et al., 2006; Duplessis et al., 2011a).

Several transcripts present homology to genes involved in the alleviation of oxidative stress caused by ROS. By instance, two transcripts (00297 and 01268) with homology to Mn-type



**Table 4 | Distribution of *Hemileia vastatrix* genes by KOG categories considering all genes and different groups of genes according to their homology or predicted function.**

Total	All genes		Gene with homology to the available rust genomes (Mlp, Pgt, Pst)		Genes with no homologies		Secreted proteins		PHI hits	
	9234	%	2992	%	4707	%	516	%	588	%
Amino acid transport and metabolism	155	5.30	119	5.68	1	0.534759	1	0.77	25	5.26
Carbohydrate transport and metabolism	153	5.23	114	5.44	5	2.673797	7	5.38	24	5.05
Cell cycle control, cell division, chromosome partitioning	77	2.63	61	2.91	6	3.208556	9	6.92	27	5.68
Cell wall/membrane/envelope biogenesis	58	1.98	35	1.67	1	0.534759	7	5.38	9	1.89
Chromatin structure and dynamics	73	2.50	54	2.58	8	4.278075	2	1.54	7	1.47
Coenzyme transport and metabolism	29	0.99	25	1.19	0	0	3	2.31	5	1.05
Cytoskeleton	141	4.82	63	3.01	33	17.64706	9	6.92	18	3.79
Defense mechanisms	19	0.65	8	0.38	0	0	0	0.00	4	0.84
Energy production and conversion	233	7.97	182	8.68	4	2.139037	6	4.62	40	8.42
Extracellular structures	20	0.68	7	0.33	6	3.208556	2	1.54	0	0.00
Inorganic ion transport and metabolism	101	3.45	68	3.24	2	1.069519	5	3.85	24	5.05
Intracellular trafficking, secretion, and vesicular transport	187	6.40	135	6.44	11	5.882353	7	5.38	22	4.63
Lipid transport and metabolism	175	5.98	136	6.49	3	1.604278	6	4.62	37	7.79
Nuclear structure	25	0.85	15	0.72	3	1.604278	5	3.85	5	1.05
Nucleotide transport and metabolism	31	1.06	22	1.05	1	0.534759	0	0.00	2	0.42
Posttranslational modification, protein turnover, chaperones	351	12.00	290	13.84	8	4.278075	22	16.92	65	13.68
Replication, recombination and repair	48	1.64	33	1.57	2	1.069519	3	2.31	5	1.05
RNA processing and modification	147	5.03	105	5.01	20	10.69519	2	1.54	12	2.53
Secondary metabolites biosynthesis, transport and catabolism	103	3.52	70	3.34	2	1.069519	5	3.85	45	9.47
Signal transduction mechanisms	238	8.14	148	7.06	32	17.1123	11	8.46	77	16.21
Transcription	162	5.54	91	4.34	18	9.625668	9	6.92	15	3.16
Translation, ribosomal structure and biogenesis	398	13.61	315	15.03	21	11.22995	9	6.92	7	1.47
General function prediction only	340	3.68	234	7.82	21	0.446144	12	2.33	78	13.27
Function unknown	168	1.82	92	3.07	9	0.191205	11	2.14	9	1.53
No hits	6159	66.70	796	26.60	4526	96.15466	382	74.03	91	15.48

For the non-specific categories ("No hits," "Function unknown" or "General function prediction only"), % refers to the total number of genes. For the remaining categories, % refers to the total number of genes with specific KOG categories assigned; highlighted cells correspond to information that is referred to in the article.

**Table 5 | Distribution (%) of *Hemileia vastatrix* genes by expression profile across the three differentiation stages, considering all genes and different groups of genes according to their predicted function and relative abundance.**

Expression profile	1 (gU= Ap<H)	2 (gU< Ap>H)	3 (gU> Ap=H)	4 (gU= Ap>H)	5 (gU> Ap<H)	6 (gU< Ap=H)	7 (gU= Ap=H)	8 (gU< Ap<H)	9 (gU> Ap>H)	Total
All genes	35.70	27.93	23.47	1.70	2.66	0.83	0.11	2.19	5.41	9234
Genes present in the three libraries	3.44	46.43	2.68	6.63	6.89	6.12	1.28	13.78	12.76	784
<b>KOG CATEGORIES</b>										
Amino acid transport and metabolism	29.29	35.00	15.00	3.57	5.00	2.86	0.00	5.71	3.57	155
Carbohydrate transport and metabolism	28.10	43.79	13.07	2.61	2.61	1.96	0.00	4.58	3.27	153
Cell cycle control, cell division, chromosome partitioning	12.70	31.75	28.57	3.17	6.35	3.17	0.00	4.76	9.52	77
Cell wall/membrane/envelope biogenesis	16.67	38.89	27.78	1.85	3.70	0.00	0.00	1.85	9.26	58
Chromatin structure and dynamics	25.37	35.82	20.90	1.49	1.49	0.00	1.49	5.97	7.46	73
Coenzyme transport and metabolism	11.54	53.85	23.08	0.00	3.85	0.00	0.00	7.69	0.00	29
Cytoskeleton	20.45	36.36	28.41	2.27	2.27	2.27	0.00	2.27	5.68	141
Defense mechanisms	18.75	56.25	18.75	0.00	0.00	6.25	0.00	0.00	0.00	19
Energy production and conversion	13.84	49.11	16.52	4.46	4.02	0.45	0.00	5.36	6.25	233
Extracellular structures	16.67	38.89	16.67	0.00	5.56	0.00	0.00	0.00	22.22	20
Inorganic ion transport and metabolism	20.00	33.00	30.00	3.00	1.00	2.00	1.00	0.00	10.00	101
Intracellular trafficking, secretion, and vesicular transport	18.06	28.39	26.45	5.16	5.16	0.65	0.65	2.58	12.90	187
Lipid transport and metabolism	15.38	51.92	18.59	2.56	2.56	1.92	0.00	1.92	5.13	175
Nuclear structure	29.41	17.65	35.29	0.00	5.88	0.00	0.00	5.88	5.88	25
Nucleotide transport and metabolism	12.90	45.16	25.81	6.45	0.00	0.00	0.00	3.23	6.45	31
Posttranslational modification, protein turnover, chaperones	16.56	44.70	13.91	2.32	5.63	3.64	0.66	4.64	7.95	351
Replication, recombination and repair	14.63	26.83	46.34	0.00	4.88	0.00	0.00	0.00	7.32	48
RNA processing and modification	27.59	28.28	24.83	4.14	4.83	3.45	0.00	3.45	3.45	147
Secondary metabolites biosynthesis, transport and catabolism	40.00	28.42	15.79	3.16	1.05	1.05	0.00	2.11	8.42	103
Signal transduction mechanisms	30.10	28.57	21.43	3.06	4.59	0.00	0.00	3.06	9.18	238
Transcription	26.95	30.50	20.57	0.71	2.84	2.13	0.00	2.84	13.48	162
Translation, ribosomal structure and biogenesis	6.79	63.97	8.88	1.57	4.18	2.35	0.78	7.57	3.92	398
All genes except No hits, Function unknown or General function prediction only	19.69	41.59	18.88	2.72	3.87	1.84	0.31	4.14	6.97	2924
General function prediction only	17.57	36.49	29.39	3.04	2.03	0.68	0.00	2.70	8.11	340
Function unknown	25.60	32.14	24.40	2.98	5.95	0.00	0.00	1.79	7.14	168
Ra > 0.05 (%)	37.52	24.81	14.07	2.12	2.72	0.45	0.00	4.54	13.77	660
Ra > 0.1 (%)	38.86	22.83	18.75	1.90	1.36	0.54	0.00	2.45	13.32	368
Ra > 1 (%)	28.57	23.21	33.93	3.57	0.00	0.00	0.00	1.79	8.93	56
<b>SECRETED PROTEINS</b>										
All (%)	34.50	27.13	17.25	1.55	3.49	0.97	0.39	6.20	8.53	516
Ra > 0.05 (%)	12.50	26.25	12.50	3.75	2.50	2.50	0.00	13.75	26.25	80
Ra > 0.1 (%)	13.95	23.81	21.43	2.38	0.00	2.38	0.00	7.14	28.57	43
Ra > 1 (%)	12.50	25.00	37.50	12.50	0.00	0.00	0.00	12.50	0.00	8

Highlighted cells correspond to information that is referred to in the article.

	Number of genes			Relative abundance		
	gU	Ap	H	gU	Ap	H
<b>Secreted proteins</b>	<b>249</b>	<b>242</b>	<b>301</b>	<b>78.22</b>	<b>29.53</b>	<b>25.56</b>
<200aa	79	61	129	31.8625	11.2911	14.0697
>5% Cysteines	27	29	48	8.1476	1.7498	3.7998
[YFW]xC motif containing	54	61	78	6.3661	11.7250	1.5273
<b>CAZymes</b>	<b>89</b>	<b>90</b>	<b>57</b>	<b>6.809</b>	<b>3.358</b>	<b>1.225</b>
Carbohydrate Esterase	9	12	11	0.4511	0.3405	0.2845
Carbohydrate-Binding Module	2	5	2	0.0578	0.1237	0.0663
Glycoside Hydrolase	43	40	32	2.8917	1.8615	0.6192
Glycosyltransferase	32	30	12	0.3682	0.4281	0.1109
<b>Transporters</b>	<b>138</b>	<b>127</b>	<b>98</b>	<b>6.590</b>	<b>2.994</b>	<b>2.539</b>
<b>Metabolic pathways</b>	<b>349</b>	<b>395</b>	<b>305</b>	<b>14.320</b>	<b>8.995</b>	<b>22.990</b>
<b>Pathogenicity genes</b>	<b>286</b>	<b>292</b>	<b>186</b>	<b>6.6649</b>	<b>5.4458</b>	<b>4.9834</b>
Signal transduction mechanisms	38	40	27	0.551	0.5401	0.6832
Posttranslational modification, protein turnover, chaperones	22	28	19	0.2415	0.4502	0.5434
Energy production and conversion	18	26	18	0.2759	0.4666	0.3882
Lipid transport and metabolism	19	24	13	0.2290	0.3573	0.1802
Secondary metabolites biosynthesis, transport and catabolism	7	16	10	0.0575	0.2008	1.0602
<b>KOG categories</b>						
Amino acid transport and metabolism	67	85	82	0.6158	1.0915	7.0027
Carbohydrate transport and metabolism	65	88	74	0.6730	1.3595	1.3705
Cell cycle control, cell division, chromosome partitioning	51	40	39	0.9816	0.6809	0.5840
Cell wall/membrane/envelope biogenesis	33	31	21	0.8085	0.8442	1.2180
Chromatin structure and dynamics	43	38	39	1.7490	7.0806	3.3650
Coenzyme transport and metabolism	17	18	11	0.1554	0.3872	0.1487
Cytoskeleton	78	67	63	2.8448	7.7092	5.9115
Defense mechanisms	8	12	4	0.0569	0.1171	0.1062
Energy production and conversion	130	152	98	1.9564	3.0311	1.8325
Extracellular structures	12	12	5	4.2132	2.1092	0.0562
Inorganic ion transport and metabolism	61	50	33	2.9121	1.6024	1.0367
Intracellular trafficking, secretion, and vesicular transport	139	97	75	1.8922	1.6718	1.2310
Lipid transport and metabolism	98	112	56	1.1721	2.0621	0.7688
Nuclear structure	17	10	13	0.3586	1.9497	0.1668
Nucleotide transport and metabolism	19	19	9	0.1962	0.3500	0.2885
Posttranslational modification, protein turnover, chaperones	204	228	159	3.0086	5.1990	3.0526
Replication, recombination and repair	35	18	18	0.3704	0.2931	0.2156
RNA processing and modification	84	64	74	1.0546	1.1414	0.2150
Secondary metabolites biosynthesis, transport and catabolism	41	46	48	2.5150	1.0394	9.7508
Signal transduction mechanisms	132	112	111	4.0205	8.0394	6.7880
Transcription	98	77	68	1.9524	1.7127	1.2772
Translation, ribosomal structure and biogenesis	212	325	151	4.4119	11.4528	4.9655
General function prediction only	201	171	131	3.1023	5.5329	10.8636
Function unknown	101	77	75	2.2658	1.4814	1.5117
No hits	2524	1846	3166	311.775	186.672	304.552
<b>All genes</b>	<b>4267</b>	<b>3626</b>	<b>4465</b>	<b>350.37</b>	<b>245.69</b>	<b>355.71</b>

**FIGURE 2 | Heatmaps of the number of genes and sum of their relative abundance values (=number of transcripts/transcript length) in the three libraries (germinating urediniospores, gU; appressoria, Ap;**

**infected leaves 21 days after inoculation, H) for transcripts according to the main categories under analysis. Color scale: green to red denote lowest to highest expression values for each gene.**

superoxide dismutase, orthologs of the *Cryptococcus gattii* pathogenicity-required manganese superoxide dismutase gene (*sod2*) (Narasipura et al., 2005) were identified. Transcript 00297 is ortholog of *U. fabae* gene *Uf058*, *P. graminis* f. sp. *tritici* gene PGTG\_04728 and *M. larici-populina* gene 107563. These *H. vastatrix* genes are up-regulated in germinating urediniospores and their proteins are predicted to be secreted, suggesting an early role in response to plant defense responses (Table 3).

### CAZymes

The comparison of *H. vastatrix* transcripts to the carbohydrate-active enzymes (CAZymes) database (www.cazy.org; Cantarel et al., 2009) and to the predicted *M. larici-populina* and *P. graminis* f. sp. *tritici* CAZymes (Duplessis et al., 2011a) enabled the identification of 148 putative CAZymes in the coffee rust fungus. This number represents ca. 45 % of the CAZymes in the poplar and the wheat stem rust fungi genomes (Supplementary Data 9 and Supplementary Data 4, column DB), similar to those arising from the comparison of the total number of transcripts predicted in this study to the number of genes in those two genomes. However, the number of *H. vastatrix* CAZymes transcripts varies according to the type and family of enzymes. For instance, 13 and 14 genes belonging to the Glycoside Hydrolase family 47 were identified in *M. larici-populina* and in *P. graminis* f. sp. *tritici* respectively, while only two transcripts were detected in *H. vastatrix*. Several other gene families are found in comparable numbers in the three fungal species, even so in the most abundant families (e.g., Carbohydrate Esterase family 4, Glycoside Hydrolase families 5 and 16, Glycosyltransferase family 2). On the contrary, some transcript families found in *H. vastatrix* are absent from the genomes of *M. larici-populina* (e.g., Glycosyltransferase families 25 and 43) or *P. graminis* f. sp. *tritici* (Glycoside Hydrolase families 51 and 92). Additionally, eight Glycoside Hydrolase family 7 genes were identified both in *M. larici-populina* and in *P. graminis* f. sp. *tritici*, but none in *H. vastatrix* transcripts. CAZyme transcripts were more frequently expressed in the gU library and less in H (Figure 2). Among CAZymes transcripts identified in this study, 31 presented homology to genes in the PHI database.

### TRANSPORTERS

A comparison to the Transporter Classification Database (www.tcdb.org; Saier et al., 2006, 2009) and to transporters from *M. larici-populina* and *P. graminis* f. sp. *tritici* (Duplessis et al., 2011a) enabled the identification of 215 transcripts encoding putative transporters. This represents ca. 60% of the number of transporters inferred from the *M. larici-populina* and *P. graminis* f. sp. *tritici* genome sequences (Duplessis et al., 2011a), again a similar proportion to that reported for other *H. vastatrix* transcript categories. However, deviations to this proportion occur in different transporter families. By instance, a family expansion is apparent in *H. vastatrix* for the F-ATPase family (H<sup>+</sup>- or Na<sup>+</sup>-translocating F-type, V-type and A-type ATPase) with 25 different transcripts predicted in *H. vastatrix*, against 20 in *M. larici-populina*, 22 in *P. graminis* f. sp. *tritici*, and 19–25 in a selection of basidiomycetes (Duplessis et al., 2011a). Similarly, variations in the Ra values can be related to the transporter type (Supplementary Data 10 and 11). In general, these results

suggest that the transport capacity is at least as high in gU or Ap as in H.

Among the 215 *H. vastatrix* transcripts encoding putative transporters, 60 show homology to the PHI database. Both Ra and gene number are higher in gU and lower in H (Figure 2). Forty are ATP-dependent transporters, including members of the ATP-binding cassette superfamily, (transcripts 01804, 07317, and 09267) and members of the P-type ATPase superfamily (transcripts 00176, 00302, 00402, 01534, and 07365), which are mostly expressed in the gU and Ap libraries. These results are corroborated by RT-qPCR for transcript 01534, with induction of expression both in *in vitro* and *in planta* appressorial samples (Table 3). The transcript 00302 is an ortholog of the *M. oryzae* P-type ATPase gene (*pde1*), required for the development of penetration hyphae and the proliferation of the fungus (Balhadère and Talbot, 2001) and was detected in the three *H. vastatrix* libraries at relatively constant expression levels. Orthologs of this gene were also identified in other rusts species (Broeker et al., 2006; Jakupović et al., 2006; Yin et al., 2009; Duplessis et al., 2011a), with elevated expression values in *P. pachyrhizi* appressoria (Stone et al., 2012). Among members of the Voltage-gated K<sup>+</sup> Channel  $\beta$  subunit family are two transcripts (00184 and 00427) identified in the three *H. vastatrix* libraries, and one (04218) only identified in the H library but at higher Ra values.

### METABOLIC PATHWAYS

The availability of nutrients for the fungus is very scarce at the early stages of the infection process and energy must be obtained from urediniospore contents. Carbohydrate metabolism by glycolysis/tricarboxylic acid cycle (TCA)/glyoxylate shuttle and lipids metabolism seems to be crucial to the success of the penetration process (Solomon et al., 2004). In the present study, orthologs of genes coding several key enzymes of glycolysis and TCA pathways were identified that presented higher Ra values in gU and Ap datasets (Supplementary Data 12, panels A and E). Polyols and trehalose are among the sugars mobilized during germination (D'Enfert et al., 1999; Voegelé and Mendgen, 2003). One of the major roles of trehalose seems to be the regulation of glycolysis. In the trehalose biosynthetic pathway, the intermediate trehalose 6-phosphate plays an important metabolic regulatory role by controlling glycolysis through hexokinase. In *H. vastatrix*, transcript 00156, orthologous of a hexokinase, is upregulated in Ap according to RT-qPCR results. Two transcripts (04402 and 04553) were identified orthologous of trehalose-6-phosphate synthase genes in *M. larici-populina* (gene 33497) and *P. graminis* f. sp. *tritici* (PGT\_06208), and RT-qPCR results showed the accumulation of transcript 04402 in appressoria. *H. vastatrix* transcript 00704, an ortholog of a neutral trehalase (*M. larici-populina* gene 116200), was detected in the three libraries, RT-qPCR showing a peak of expression in the appressoria and at 21 days after inoculation, suggesting a close control of trehalose/trehalose-6-P levels at these stages.

The glycolysis pathway leads to the production of pyruvate after conversion into acetyl-CoA. This pathway is fundamental for cell survival since it provides intermediate metabolites and other important small molecules, such as ATP and NADH. In the present dataset, all enzymes involved in this pathway were



detected (Supplementary Data 12, panel A). A close connection between glycolysis and other pathways such as pentose phosphates and  $\beta$ -oxidation suggests the existence of a tight control of carbohydrate mobilization and utilization. Dihydroxy acetone phosphate, produced by aldolase by the glycerol-3-phosphate shuttle, can lead to the formation of glycerol (Supplementary Data 12, panel C) (Cronwright et al., 2002). In *H. vastatrix*, transcript 08812, ortholog of a glycerol 3-phosphate dehydrogenase, was identified in the Ap library. RT-qPCR analysis (Table 3) further revealed its expression during pre- and post-penetration events, strongly decreasing at late colonization stages. Similarly, transcript 01400 (glycerol 3-phosphatase gene ortholog), was accumulated in gU and Ap samples (Table 3). In *Saccharomyces cerevisiae*, the role of glycerol has been described in the maintenance of the cytosolic redox state (Cronwright et al., 2002). Besides, in fungi such as *Magnaporthe* or *Colletotrichum*, the important turgor pressure built in appressoria is mediated by the accumulation of very large amounts of glycerol in the cell (de Jong et al., 1997; Soanes et al., 2012). In *H. vastatrix*, transcripts with homology to glycerol 3-phosphatase and NAD<sup>+</sup>-dependent glycerol 3-phosphate dehydrogenase (transcripts 01400 and 06448, respectively) showed higher expression during germination and appressoria formation according both to 454 pyrosequencing and RT-qPCR results. Increased levels of these enzymes were also described in *P. pachyrhizi* at the appressorial stage (Stone et al., 2012). The glycerol formed is metabolized by the action of a glycerol kinase (transcripts 06788 and 09049) the expression of which is also observed during appressorial formation according both to 454 pyrosequencing and RT-qPCR results, suggesting the importance of the maintenance of glycerol contents. While the sum of Ra values suggests higher expression of genes related to metabolism in the H library, a higher number of genes was identified in the Ap library (Figure 2).

Beyond the glycerol-3-phosphate shuttle, glycerol generation may also be achieved by the mobilization of storage lipids through degradation of triacylglycerol by triacylglycerol lipases (EC 3.1.1.3) (Thines et al., 2000). In fact, flexibility in lipid metabolism and ability to divert intermediates from glycolysis identified in *M. oryzae* was suggested to be important for rapid glycerol accumulation during appressorium development (Dean et al., 2005). In this study, the results suggested a high rate of lipid metabolism during germination and appressoria formation. Among the 16 putative lipases (transcripts 00223, 00443, 00530, 00606, 01163, 01746, 01917, 02201, 04308, 06521, 06529, 07167, 07216, 07621, and 09011), 12 were found in the gU library and nine in the Ap library, while only two transcripts were expressed in the H library (Supplementary Data 4). Lipid metabolism is important for ATP generation and as a source of intermediates to secondary metabolic pathways. Fatty acids are oxidized by  $\beta$ -oxidation, a pathway that has been referred crucial for appressorium formation, in addition to the glyoxylate cycle, to enable utilization of acetyl-CoA for central carbohydrate metabolism (Kretschmer et al., 2012; Soanes et al., 2012). The present study enabled the identification of orthologs of all genes involved in  $\beta$ -oxidation pathways in *M. larici-populina* and *P. graminis* f. sp. *tritici* (Supplementary Data 12, panel D). The comparison among the three *H. vastatrix* libraries revealed that fatty acid degradation

increased in Ap as indicated by the increased expression of transcripts coding for several  $\beta$ -oxidation enzymes such as long-chain fatty acid CoA ligase (transcript 00175), acyl-CoA dehydrogenase (transcript 01629), enoyl CoA hydratase (transcript 01055), 3-hydroxyacyl-CoA dehydrogenase (transcript 01628), and 3-ketoacyl-CoA thiolase (transcript 00191). A similar profile was detected for acyl-CoA oxidase (transcript 00622). RT-qPCR profiles for these transcripts further revealed a second peak of expression at 2 days for transcripts 00191 and 01628, and at 7 days for transcript 00175. Transcripts 01594, 01988 and 07400 are orthologs of *M. oryzae* carnitine acetyl-transferase gene (*crat1*), involved in transport of peroxisomal acetyl-CoA. *M. oryzae* deletion mutants for this gene show reduced appressoria melanisation, and are not able to elaborate penetration pegs or infection hyphae (Ramos-Pamplona and Naqvi, 2006). Interestingly, in *H. vastatrix* these transcripts were only identified in gU and Ap libraries, further suggesting their potential involvement in appressorium-mediated infection. RT-qPCR analyses illustrate different expression profiles for these three transcripts: while 07400 is induced during appressoria formation both *in vitro* and *in planta*, with a second peak of induction at 7 days, transcripts 01594 and 01988 are mostly over-expressed during hyphal colonization of host tissues, after 2 days for transcript 01594 and as early as appressoria differentiation for transcript 01988 (Table 3).

The glyoxylate cycle provides means for cells to assimilate two-carbon compounds into the TCA cycle and channel these via gluconeogenesis to the biosynthesis of glucose, (Supplementary Data 12, panel E). Generally, induction of the glyoxylate cycle indicates that a cell is using lipid metabolism as its predominant source for ATP generation, involving  $\beta$ -oxidation of fatty acids and the production of acetyl CoA. In *H. vastatrix*, results showed the presence of transcripts coding for all enzymes of the glyoxylate cycle. Ra values, as well as RT-qPCR analysis for transcripts 00491, 00717, 01133, 01266, and 08833, suggest an increasing level of expression during appressoria formation. The fact that glyoxylate cycle allows the connection between lipid and carbon metabolism may be particularly important for foliar pathogenic fungi that need to germinate and develop specific infection structures before having access to plant nutrients (Wang et al., 2003).

## SIGNALING

A total of 25 *H. vastatrix* transcripts presented homology to genes involved in signaling, whose mutants in various fungi exhibit either loss of pathogenicity or reduced virulence phenotype (recorded in the PHI database) (Supplementary Data 5). In the cAMP pathway, transcripts 01548 and 08827 are orthologs of pathogenicity-required adenylyl cyclase (*cdc35*) and adenylyl cyclase (*cac1*) genes, respectively from *Candida albicans* (Rocha et al., 2001) and *Colletotrichum lagenarium* (Yamauchi et al., 2004), necessary for filamentous growth. Matching the relevance of this gene for spore germination and differentiation of infection structures from appressoria, RT-qPCR analysis showed the accumulation of transcript 08827 at the appressorial stage and of transcript 01548 at early infection stages, from urediniospore germination until 3 days (Table 3). Also, transcripts 00898 and 01431 are orthologs of the *Colletotrichum trifolii* pathogenicity-required catalytic subunit of cyclic AMP-dependent protein kinase gene

(*pkac*), necessary for penetration and sporulation (Yang and Dickman, 1999). RT-qPCR profiles showed induction of their expression in appressorial samples both obtained *in vitro* and *in planta* (Table 3), compatible with the involvement of these genes in penetration. Another protein kinase involved in the cAMP pathway is the *M. oryzae/C. trifolii* pathogenicity-required *cpkA* gene, ortholog of *H. vastatrix* transcript 06436, required for appressorium formation and pathogenesis (Mitchell and Dean, 1995; Yang and Dickman, 1999). RT-qPCR results further suggested an activation of the expression of this transcript in the appressorial stage (Table 3).

Several MAP kinases and serine/threonine kinases were identified, and RT-qPCR results further corroborated induction of their expression in germinating urediniospores and/or in appressoria. By instance, orthologs of the *Ustilago maydis kpp6* and *ubc3* genes (Mayorga and Gold, 1999; Brachmann et al., 2003), the *Cryphonectria parasitica cpmk1* gene (Park et al., 2004), the *Claviceps purpurea cpmk1* gene (Mey et al., 2002) and the *Cryptococcus neoformans* var. *grubii hog1* gene (Bahn et al., 2005) were identified. *H. vastatrix* transcripts 06883 and 07140, orthologs respectively of the *C. purpurea* and of *C. parasitica cpmk1* genes, were both identified only in the gU library. RT-qPCR analysis showed distinct expression profiles, with transcript 06883 induced at pre-penetration stages only, and transcript 07140 expressed at all infection stages except late colonization and resting urediniospores (Table 3). On the contrary, both the *ubc3* type transcript 00373 and the *hog1* type transcript 00616 were identified in the three libraries showing stable expression profiles, as corroborated by RT-qPCR results (Table 3). In *M. oryzae*, the Pmk1 MAP-kinase pathway has a major role in controlling appressorium morphogenesis (Soanes et al., 2012). Also, two MAPK kinases (MAP2K) were identified, corresponding to transcripts 01932 and 01813, of the Ste7 and the Mkk1 types respectively (Hamel et al., 2012), both of them identified in the gU and Ap libraries. RT-qPCR results further showed that the expression of transcript 01932 is observed at early infection stages (Table 3). The *H. vastatrix* transcripts 00125 and 00489, and 00409, orthologs respectively of the of the *M. oryzae* MAP kinase-regulated *gas1* and *gas2* genes (Xue et al., 2002), were identified in the three libraries, with high expression levels ( $Ra > 1$ ) in the gU and Ap libraries. RT-qPCR results corroborate the induction of expression of these transcripts in germinating urediniospores and in appressoria, with a second peak of expression recorded for transcripts 00409 and 00489 respectively at 2 and 3 days (Table 3). Interestingly, transcript accumulation for *gas* orthologs in *P. pachyrhizi* and *Uromyces appendiculatus* purified haustoria were also reported (Link et al., 2013). The present results indicated that *gas* expression is not solely related to the rust haustorial infection structure, but also to earlier stages such as spore germination. A group of *H. vastatrix* transcripts show homology to G protein subunits genes from the PHI base. Heterotrimeric G-proteins transduce extracellular signals to various downstream effectors (e.g., MAP kinases) in eukaryotic cells. Transcript 06565 shows homology to the *Cryptococcus neoformans* virulence-related *gpa1* gene (Alspaugh et al., 1997) and to the *M. larici-populina* heterotrimeric G-protein  $\alpha$  subunit 3 gene (*gpa3*) (Duplessis et al., 2011a). In *H. vastatrix*, transcript 06565 was only detected in the

gU library, suggesting an involvement in pre-penetration events, in agreement with the profile of its *M. larici-populina* ortholog (gene 47478) (Duplessis et al., 2012). Orthologs of G-protein  $\beta$  subunit genes involved in appressorium formation, including genes *mgb1* from *M. oryzae* (Nishimura et al., 2003), *cgb1* from *Cochliobolus heterostrophus* (Ganem et al., 2004) and *Bpp1* from *U. maydis* (Müller et al., 2004), were identified in *H. vastatrix* (transcripts 00968 and 02022 for *mgb1* and 07174 for *cgb1/Bpp1*), all of them in the gU library, indicative of a possible role in appressoria formation in *H. vastatrix*. RT-qPCR expression profiling further showed induction of transcripts 02022 and 07174 during early pre-penetration events (Table 3).

#### OTHER GENES IDENTIFIED IN RUST TRANSCRIPTOMIC/GENOMIC STUDIES

Several orthologs of *U. fabae in planta*-induced genes (PIGs) were identified in *H. vastatrix* (Supplementary Data 13). The *U. fabae* PIGs genes showed induced expression in *Vicia faba* infected leaves as compared to germinating urediniospores (Jakupović et al., 2006). The majority of *U. fabae* transcripts with a predicted function have orthologs in *H. vastatrix* (Supplementary Data 13).

A comparison of *H. vastatrix* genes to *P. pachyrhizi* genes expressed in germinating urediniospores (Posada-Buitrago and Frederick, 2005) reveals that the two most expressed *P. pachyrhizi* genes, *Pp0104* and *Pp0417*, have no significant homologies in *H. vastatrix* (Supplementary Data 14). In the same way, the comparison of the 9234 *H. vastatrix* transcript to a collection of ESTs and proteins differentially expressed in *P. pachyrhizi* appressoria (Stone et al., 2012) shows a very limited number of genes in common between both studies (Supplementary Data 14). Interestingly however, the comparison of 4492 *P. pachyrhizi* haustorial ESTs (Link et al., 2013) to the 9234 *H. vastatrix* transcripts identified 1668 hits to 1132 unique *H. vastatrix* transcripts. Half of them corresponded to *H. vastatrix* transcripts not detected in the H library (Supplementary Data 14). A similar situation was observed when comparing the 7561 *U. appendiculatus* haustorial ESTs (Link et al., 2013) to the *H. vastatrix* transcripts (data not shown).

Among the 156 *M. larici-populina* annotated genes that are >10-fold up-regulated in infected leaves as compared to urediniospores (Duplessis et al., 2011a), only 22% have orthologs in the present *H. vastatrix* dataset, including 12 transporters (mostly sugar and ion transporters), 10 secreted proteins and six glycoside hydrolases (Supplementary Data 4, columns Y-AA). Among the 235 *P. graminis* f. sp. *tritici* annotated genes that are the >10-fold up-regulated in infected leaves as compared to urediniospores (Duplessis et al., 2011a), 49% have orthologs in the present *H. vastatrix* dataset, although half of these are predicted ribosomal genes. Unlike for *M. larici-populina*, none of these *P. graminis* f. sp. *tritici* genes include glycoside hydrolases or secreted proteins and only two transporters were identified.

An expanded number of multigene families have been reported in *M. larici-populina* and *P. graminis* f. sp. *tritici* as compared to other Basidiomycetes (Duplessis et al., 2011a). Among those expanded families, the number of *H. vastatrix* transcripts-based predicted genes is higher than those for *M. larici-populina* or *P. graminis* f. sp. *tritici* for the major facilitator superfamily,

helicase or chitinase, and under-represented for families such as serine/threonine protein kinase and sugar transporter (Table 6). While the current study does not cover all differentiation stages of the *H. vastatrix* life cycle and transcripts expressed at low level may not be represented, it is interesting to note that some gene families are over-represented in comparison to annotated genome sequences.

## CONCLUSIONS

In this study, 7894 contigs were obtained by 454 pyrosequencing of cDNA from *H. vastatrix* germinating urediniospores and appressoria. These transcripts, along with 4465 *in planta* expressed contigs (Fernandez et al., 2012), were assembled into 9234 annotated transcripts. This number represents an important fraction (>50%) of the genes predicted in rust sequenced genomes so far (Duplessis et al., 2012). In addition, this elevated gene number for *H. vastatrix* is corroborated by other database comparisons, such as the core fungal genes database (FUNYBASE), the carbohydrate-active enzymes (CAZy) database or the Transporter Classification Database (TCDB). Database comparisons further indicate that half of these transcripts (4707) present no significant homology to genomic or transcriptomic data from other rusts, potentially representing novel or very divergent *H. vastatrix* genes.

Annotation of *H. vastatrix* transcripts and comparison of their relative abundance in each of the three sampling stages suggest a particularly active metabolism, translational activity, production of new structures and signaling in appressoria and intense transport, secretory activity and cellular multiplication in germinating urediniospores (Figure 2). Transcripts encoding putative carbohydrate-active enzymes and different types of transporters are more expressed in germinating urediniospores and appressoria, and lesser at late infection stages. Among transcripts involved in metabolic pathways,

an active lipid metabolism was observed at pre-penetration stages compared to late infection stages, while amino acid and carbohydrate metabolism was more active in post-penetration samples. Moreover, the homology of *H. vastatrix* transcripts to genes known to be involved and/or required for pathogenicity in other fungal plant pathogens, namely in appressoria-mediated infection, enabled the identification of an array of putative pathogenicity factors, a large proportion of which are expressed as early as during germ-tube elongation. Also, while melanized cuticle-breaching appressoria have been thoroughly investigated over the last few decades, namely in *M. oryzae* and *Colletotrichum* spp. (Deising et al., 2000; Kleemann et al., 2012), the present study represents an important insight into genes expressed in non-melanized stomata-penetrating appressoria. To this end, induction of expression of genes related to the production of carbohydrate-active enzymes and to the accumulation of glycerol in germinating urediniospores and appressoria suggests that combined lytic and physical mechanisms are involved in appressoria-mediated penetration of coffee leaf stomata.

This early activation of signaling, transport and secretory pathways suggests a precocious plant-fungus dialogue, which is corroborated by the possible induction of an hypersensitive reaction in stomatal cells of some resistant coffee varieties as early as at the appressorial stage (Silva et al., 2002; Ganesh et al., 2006; Diniz et al., 2012), thus prompting further studies targeting the identification of virulence/avirulence factors (and their resistance/susceptibility counterparts) expressed at these early stages of the plant-fungus interaction.

## AUTHOR CONTRIBUTIONS

This study was conceived and directed by Pedro Talhinhas, Helena G. Azinheira, Sébastien Duplessis, Maria do Céu Silva, and Diana Fernandez. The laboratorial experiments were conducted by Pedro Talhinhas, Helena G. Azinheira, Andreia Loureiro, Sílvia Tavares, and Anne-Sophie Petitot. 454-pyrosequencing was conducted by Julie Poulain and Corinne Da Silva. Bioinformatic analyses were conducted by Bruno Vieira, Emmanuelle Morin, and Octávio S. Paulo. Biological interpretation of bioinformatics analyses were conducted by Pedro Talhinhas, Helena G. Azinheira, Andreia Loureiro, Sílvia Tavares, Sébastien Duplessis, and Diana Fernandez. Pedro Talhinhas, Helena G. Azinheira, Andreia Loureiro, Sílvia Tavares, Dora Batista, Octávio S. Paulo, Sébastien Duplessis, Maria do Céu Silva, and Diana Fernandez wrote the paper. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

This work was undertaken through a French-Portuguese collaborative project (Partenariat Hubert Curien PHC-Pessoa 22583XM) funded by the Ministère des Affaires Étrangères et Européennes of France. The work was funded by Portuguese national funds through Fundação para a Ciência e a Tecnologia (projects PTDC/AGR-AAM/71866/2006 and PTDC/AGR-GPL/114949/2009 and grants SFRH/BPD/47008/2008, SFRH/BPD/65965/2009 and SFRH/BPD/88994/2012), and by CEA/Genoscope-INRA-IRD Collaborative project (<http://>

**Table 6 | Comparison of the number of *Hemileia vastatrix* (Hv) genes to the number of members of gene families in *Melampsora larici-populina* (Mlp) and *Puccinia graminis* f. sp. *tritici* (Pgt) reported (Duplessis et al., 2011a) as expanded in relation to other Basidiomycetes.**

Gene family	Hv	Mlp	Pgt
Amino acid transporter	12	11	12
Carbohydrate deacetylase	11	11	10
Cell division/GTP binding protein	13	6	13
Chitinase	33	14	9
Helicase	19	14	13
Histone H3	9	11	8
Major facilitator superfamily	50	29	17
Serine/threonine protein kinase	36	87	73
Sugar transporter	7	19	15
Superoxide dismutase	6	7	20
Zinc finger protein	29	4	67

Colour scale: green to red denote lowest to highest number of genes for each gene family.



www.genoscope.cns.fr/spip/Identification-of-virulence.html) (France), whose funding is gratefully acknowledged.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpls.2014.00088/abstract>

## REFERENCES

- Aime, M. C. (2006). Toward resolving family-level relationships in rust fungi (Uredinales). *Mycoscience* 47, 112–122. doi: 10.1007/s10267-006-0281-0
- Alsbaugh, J. A., Perfect, J. R., and Heitman, J. (1997). *Cryptococcus neoformans* mating and virulence are regulated by the G-protein alpha subunit GPA1 and cAMP. *Genes Dev.* 11, 3206–3217. doi: 10.1101/gad.11.23.3206
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Azinheira, H. G., Silva, M. C., Guerra-Guimarães, L., Mendgen, K., Rodrigues, Jr., and C., Pinto Ricardo, C. (2001). “Development of infection structures of *Hemileia vastatrix* on artificial membranes,” in *11th Conference of the Mediterranean Phytopathological Union, 17–20 September 2001* (Évora: Andalus Academic Publishing), 353–355.
- Bahn, Y. S., Kojima, K., Cox, G. M., and Heitman, J. (2005). Specialization of the HOG pathway and its impact on differentiation and virulence of *Cryptococcus neoformans*. *Mol. Biol. Cell* 16, 2285–2300. doi: 10.1091/mbc.E04-11-0987
- Balhadère, P. V., and Talbot, N. J. (2001). PDE1 encodes a P-type ATPase involved in appressorium-mediated plant infection by the rice blast fungus *Magnaporthe grisea*. *Plant Cell* 13, 1987–2004. doi: 10.1105/TPC.010056
- Barrett, L. G., Thrall, P. H., Dodds, P. N., van der Merwe, M., Linde, C. C., Lawrence, G. J., et al. (2009). Diversity and evolution of effector loci in natural populations of the plant pathogen *Melampsora lini*. *Mol. Biol. Evol.* 26, 2499–2513. doi: 10.1093/molbev/msp166
- Brachmann, A., Schirawski, J., Müller, P., and Kahmann, R. (2003). An unusual MAP kinase is required for efficient penetration of the plant surface by *Ustilago maydis*. *EMBO J.* 22, 2199–2210. doi: 10.1093/emboj/cdg198
- Broeker, K., Bernard, F., and Moerschbacher, B. M. (2006). An EST library from *Puccinia graminis* f. sp. *tritici* reveals genes potentially involved in fungal differentiation. *FEMS Microbiol. Lett.* 256, 273–281. doi: 10.1111/j.1574-6968.2006.00127.x
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663
- Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K. K., et al. (2011). Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS ONE* 6:e24230. doi: 10.1371/journal.pone.0024230
- Catanzariti, A.-M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A., and Ellis, J. G. (2006). Haustorially-expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* 18, 243–256. doi: 10.1105/tpc.105.035980
- Chen, Z. J., Nunes, M. A., Silva, M. C., and Rodrigues, C. J. (2004). Appressorium turgor pressure of *Colletotrichum kahawae* might have a role in coffee cuticle penetration. *Mycologia* 96, 1199–1208. doi: 10.2307/3762135
- Cressey, D. (2013). Coffee rust regains foothold. *Nature* 493, 587. doi: 10.1038/493587a
- Cronwright, G. R., Rohwer, J. R., and Prior, B. A. (2002). Metabolic control analysis of glycerol synthesis in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 68, 4448–4456. doi: 10.1128/AEM.68.9.4448-4456.2002
- D’Enfert, C., Bonini, B. M., Zapella, P. D. A., Fontaine, T., da Silva, A. M., and Terenzi, H. F. (1999). Neutral trehalases catalyse intracellular trehalose breakdown in the filamentous fungi *Aspergillus nidulans* and *Neurospora crassa*. *Mol. Microbiol.* 32, 471–484. doi: 10.1046/j.1365-2958.1999.01327.x
- Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434, 980–986. doi: 10.1038/nature03449
- Deising, H. B., Werner, S., and Wernitz, M. (2000). The role of fungal appressoria in plant infection. *Microbes Infect.* 2, 1631–1641. doi: 10.1016/S1286-4579(00)01319-8
- de Jong, J. C., McCormack, B. J., Smirnov, N., and Talbot, N. J. (1997). Glycerol generates turgor in rice blast. *Nature* 389, 244–245. doi: 10.1038/38418
- Diniz, I., Talhinhas, P., Azinheira, H. G., Várzea, V., Medeira, C., Maia, I., et al. (2012). Cellular and molecular analyses of coffee resistance to *Hemileia vastatrix* and nonhost resistance to *Uromyces vignae* in the resistance-donor genotype HDT832/2. *Eur. J. Plant Pathol.* 133, 141–157. doi: 10.1007/s10658-011-9925-9
- Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Ayliffe, M. A., and Ellis, J. G. (2004). The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* 16, 755–768. doi: 10.1105/tpc.020040
- Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Teh, T., Wang, C. I. A., Ayliffe, M. A., et al. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8888–8893. doi: 10.1073/pnas.0602577103
- Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., et al. (2011a). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9166–9171. doi: 10.1073/pnas.1019315108
- Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F., et al. (2011b). *Melampsora larici-populina* transcript profiling during germination and time-course infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Mol. Plant-Microbe Interact.* 24, 808–818. doi: 10.1094/MPMI-01-11-0006
- Duplessis, S., Joly, D. L., and Dodds, P. N. (2012). “Rust effectors,” in *Effectors in Plant-Microbe Interactions*, eds F. Martin and S. Kamoun (New York, NY: Wiley), 155–193.
- Eklblom, R., Balakrishnan, C. N., Burke, T., and Slate, J. (2010). Digital gene expression analysis of the zebra finch genome. *BMC Genomics* 11:219. doi: 10.1186/1471-2164-11-219
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903
- Fernandez, D., Talhinhas, P., and Duplessis, S. (2013). “Rust fungi: achievements and future challenges on genomics and host-parasite interactions,” in *The Mycota XI, Agricultural Applications 2nd Edn.*, ed F. Kempken (Berlin: Springer-Verlag), 315–341.
- Fernandez, D., Tisserant, E., Talhinhas, P., Azinheira, H. G., Vieira, A., Loureiro, A., et al. (2012). 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta expressed pathogen secreted proteins and plant functions expressed in a late compatible plant-rust interaction. *Mol. Plant Pathol.* 13, 17–37. doi: 10.1111/j.1364-3703.2011.00723.x
- Ganem, S., Lu, S. W., Lee, B. N., Chou, D. Y., Hadar, R., Turgeon, B. G., et al. (2004). G-protein beta subunit of *Cochoilobolus heterostrophus* involved in virulence, asexual and sexual reproductive ability, and morphogenesis. *Eukaryot. Cell* 3, 1653–1663. doi: 10.1128/EC.3.6.1653-1663.2004
- Ganesh, D., Petitot, A., Silva, M. C., Alary, R., Lecouls, A. C., and Fernandez, D. (2006). Monitoring of the early molecular resistance responses of coffee (*Coffea arabica* L.) to the rust fungus (*Hemileia vastatrix*) using real-time quantitative RT-PCR. *Plant Sci.* 170, 1045–1051. doi: 10.1016/j.plantsci.2005.12.009
- Hacquard, S., Joly, D. L., Lin, Y.-C., Tisserant, E., Feau, N., Delaruelle, C., et al. (2012). A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol. Plant-Microbe Interact.* 25, 279–293. doi: 10.1094/MPMI-09-11-0238
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41, 95–98.
- Hamel, L.-P., Nicole, M.-C., Duplessis, S., and Ellis, B. E. (2012). Mitogen-activated protein kinase signaling in plant-interacting fungi: distinct messages from conserved messengers. *Plant Cell* 24, 1327–1351. doi: 10.1105/tpc.112.096156
- Howard, R. J., Ferrari, M. A., Roach, D. H., and Money, N. P. (1991). Penetration of hard substrates by a fungus employing enormous turgor pressures. *Proc. Natl. Acad. Sci. U.S.A.* 88, 11281–11284. doi: 10.1073/pnas.88.24.11281
- Hu, G. G., Linning, R., McCallum, B., Banks, T., Cloutier, S., Butterfield, Y., et al. (2007). Generation of a wheat leaf rust, *Puccinia triticina*, EST database from stage-specific cDNA libraries. *Mol. Plant Pathol.* 8, 451–467. doi: 10.1111/j.1364-3703.2007.00406.x



- Jakupović, M., Heintz, M., Reichmann, P., Mendgen, K., and Hahn, M. (2006). Microarray analysis of expressed sequence tags from haustoria of the rust fungus *Uromyces fabae*. *Fungal Genet. Biol.* 43, 8–19. doi: 10.1016/j.fgb.2005.09.001
- Kemen, E., Kemen, A. C., Rafiqi, M., Hempel, U., Mendgen, K., Hahn, M., et al. (2005). Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Mol. Plant Microbe Interact.* 18, 1130–1139. doi: 10.1094/MPMI-18-1130
- Kleemann, J., Rincon-Rivera, L. J., Takahara, H., Neumann, U., Ver Loren van Themaat, E., van der Does, H. C., et al. (2012). Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog.* 8:e1002643. doi: 10.1371/journal.ppat.1002643
- Kretschmer, M., Klose, J., and Kronstad, J. W. (2012). Defects in mitochondrial and peroxisomal  $\beta$ -oxidation influence virulence in the maize pathogen *Ustilago maydis*. *Eukaryot. Cell* 11, 1055–1066. doi: 10.1128/EC.00129-12
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Kumar, S., and Blaxter, M. L. (2010). Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics* 11:571. doi: 10.1186/1471-2164-11-571
- Link, T. I., Lang, P., Scheffler, B. E., Duke, M. V., Graham, M. A., Cooper, B., et al. (2013). The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. *Mol. Plant Pathol.* doi: 10.1111/mpp.12099. [Epub ahead of print].
- Mank, J. E., Hultin-Rosenberg, L., Zwahlen, M., and Ellegren, H. (2008). Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *Am. Nat.* 171, 35–43. doi: 10.1086/523954
- Marthey, S., Aguilera, G., Rodolphe, F., Gendraulat, A., Giraud, T., Fournier, E., et al. (2008). FUNYBASE: a FUNgal phylogenomic dataBASE. *BMC Bioinformatics* 9:456. doi: 10.1186/1471-2105-9-456
- Mayorga, M. E., and Gold, S. E. (1999). A MAP kinase encoded by the *ubc3* gene of *Ustilago maydis* is required for filamentous growth and full virulence. *Mol. Microbiol.* 34, 485–497. doi: 10.1046/j.1365-2958.1999.01610.x
- Mey, G., Oeser, B., Lebrun, M. H., and Tudzynski, P. (2002). The biotrophic, non-appressorium-forming grass pathogen *Claviceps purpurea* needs a Fus3/Pmk1 homologous mitogen-activated protein kinase for colonization of rye ovarian tissue. *Mol. Plant Microbe Interact.* 15, 303–312. doi: 10.1094/MPMI.2002.15.4.303
- Mitchell, T. K., and Dean, R. A. (1995). The cAMP-dependent protein kinase catalytic subunit is required for appressorium formation and pathogenesis by the rice blast pathogen *Magnaporthe grisea*. *Plant Cell* 7, 1869–1878.
- Müller, P., Leibbrandt, A., Teunissen, H., Cubasch, S., Aichinger, C., and Kahmann, R. (2004). The G $\beta$ -subunit-encoding gene *bpp1* controls cyclic-AMP signaling in *Ustilago maydis*. *Eukaryot. Cell* 3, 806–814. doi: 10.1128/EC.3.3.806-814.2004
- Narasipura, S. D., Chaturvedi, V., and Chaturvedi, S. (2005). Characterization of *Cryptococcus neoformans* variety *gattii* SOD2 reveals distinct roles of the two superoxide dismutases in fungal biology and virulence. *Mol. Microbiol.* 55, 1782–1800. doi: 10.1111/j.1365-2958.2005.04503.x
- Nishimura, M., Park, G., and Xu, J. R. (2003). The G-beta subunit MGB1 is involved in regulating multiple steps of infection-related morphogenesis in *Magnaporthe grisea*. *Mol. Microbiol.* 50, 231–243. doi: 10.1046/j.1365-2958.2003.03676.x
- Park, S. M., Choi, E. S., Kim, M. J., Cha, B. J., Yang, M. S., and Kim, D. H. (2004). Characterization of HOG1 homologue, CpMK1, from *Cryphonectria parasitica* and evidence for hypovirus-mediated perturbation of its phosphorylation in response to hypertonic stress. *Mol. Microbiol.* 51, 1267–1277. doi: 10.1111/j.1365-2958.2004.03919.x
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Posada-Buitrago, M. L., and Frederick, R. D. (2005). Expressed sequence tag analysis of the soybean rust pathogen *Phakopsora pachyrhizi*. *Fungal Genet. Biol.* 42, 949–962. doi: 10.1016/j.fgb.2005.06.004
- Pretsch, K., Kemen, A., Kemen, E., Geiger, M., Mendgen, K., and Voegelé, R. (2013). The rust transferred proteins—a new family of effector proteins exhibiting protease inhibitor function. *Mol. Plant Pathol.* 14, 96–107. doi: 10.1111/j.1364-3703.2012.00832.x
- Ramos-Pamplona, M., and Naqvi, N. I. (2006). Host invasion during rice-blast disease requires carnitine-dependent transport of peroxisomal acetyl-CoA. *Mol. Microbiol.* 61, 61–75. doi: 10.1111/j.1365-2958.2006.05194.x
- Rocha, C. R., Schröppel, K., Harcus, D., Marcil, A., Dignard, D., Taylor, B. N., et al. (2001). Signaling through adenylyl cyclase is essential for hyphal growth and virulence in the pathogenic fungus *Candida albicans*. *Mol. Biol. Cell* 12, 3631–3643. doi: 10.1091/mbc.12.11.3631
- Saier, M. H. Jr., Tran, C. V., and Barabote, R. D. (2006). TCDB: the transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Res.* 34, D181–D186. doi: 10.1093/nar/gkj001
- Saier, M. H. Jr., Yen, M. R., Noto, K., Tamang, D. G., and Elkan, C. (2009). The Transporter Classification Database: recent advances. *Nucleic Acids Res.* 37, D274–D278. doi: 10.1093/nar/gkn862
- Silva, D. N., Vieira, A., Talhinhas, P., Azinheira, H. G., Silva, M. C., Fernandez, D., et al. (2012). “Phylogenetic analysis of *Hemileia vastatrix* and related taxa using a genome-scale approach,” in *Proceedings of the 24th International Conference on Coffee Science, 11–16 November 2012, San José*, ed Association for Science and Information on Coffee (Paris), 1404–1408.
- Silva, M. C., Nicole, M., Guerra-Guimarães, L., and Rodrigues, C. J. Jr. (2002). Hypersensitive cell death and post-haustorial defence responses arrest the orange rust (*Hemileia vastatrix*) growth in resistant coffee leaves. *Physiol. Mol. Plant Pathol.* 60, 169–183. doi: 10.1006/pmpp.2002.0389
- Silva, M. C., Várzea, V., Guimarães, L. G., Azinheira, H. G., Fernandez, D., Petitot, A., et al. (2006). Coffee resistance to the main diseases: leaf rust and coffee berry disease. *Braz. J. Plant Physiol.* 18, 119–147. doi: 10.1590/S1677-04202006000100010
- Soanes, D. M., Chakrabarti, A., Paszkiewicz, K. H., Dawe, A. L., and Talbot, N. J. (2012). Genome-wide transcriptional profiling of appressorium development by the rice blast fungus *Magnaporthe oryzae*. *PLoS Pathog.* 8:e1002514. doi: 10.1371/journal.ppat.1002514
- Soanes, D. M., and Talbot, N. J. (2006). Comparative genomic analysis of phytopathogenic fungi using expressed sequence tag (EST) collections. *Mol. Plant Pathol.* 7, 61–70. doi: 10.1111/j.1364-3703.2005.00317.x
- Solomon, P. S., Lee, R. C., Wilson, T. J. G., and Oliver, R. P. (2004). Pathogenicity of *Stagonospora nodorum* requires malate synthase. *Mol. Microbiol.* 53, 1065–1073. doi: 10.1111/j.1365-2958.2004.04178.x
- Stone, C. L., McMahon, M. B., Fortis, L. L., Nuñez, A., Smythers, G. W., Luster, D. G., et al. (2012). Gene expression and proteomic analysis of the formation of *Phakopsora pachyrhizi* appressoria. *BMC Genomics* 13:269. doi: 10.1186/1471-2164-13-269
- Tatusov, R. L. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41
- Terhune, B. T., Bojko, R. J., and Hoch, H. C. (1993). Deformation of stomatal guard cell lips and microfabricated artificial topographies during appressorium formation. *Exp. Mycol.* 17, 70–78. doi: 10.1006/emyc.1993.1006
- Thines, E., Weber, R. W., and Talbot, N. J. (2000). MAP kinase and protein kinase A-dependent mobilization of triacylglycerol and glycogen during appressorium turgor generation by *Magnaporthe grisea*. *Plant Cell* 12, 1703–1718. doi: 10.1105/tpc.12.9.1703
- Várzea, V. M. P., and Marques, D. V. (2005). “Population variability of *Hemileia vastatrix* vs coffee durable resistance,” in *Durable Resistance to Coffee Leaf Rust*, eds L. Zambolim, E. Zambolim, and V. M. P. Várzea (Viçosa: Universidade Federal de Viçosa), 53–74.
- Vega-Arreguín, J. C., Ibarra-Laclette, E., Jiménez-Moraila, B., Martínez, O., Vielle-Calzada, J. P., Herrera-Estrella, L., et al. (2009). Deep sampling of the Palomero maize transcriptome by a high throughput strategy of pyrosequencing. *BMC Genomics* 10:299. doi: 10.1186/1471-2164-10-299
- Vieira, A., Talhinhas, P., Loureiro, A., Duplessis, S., Fernandez, D., Silva, M. C., et al. (2011). Validation of RT-qPCR reference genes for *in planta* expression studies in *Hemileia vastatrix*, the causal agent of coffee leaf rust. *Fungal Biol.* 115, 891–901. doi: 10.1016/j.funbio.2011.07.002
- Vieira, A., Talhinhas, P., Loureiro, A., Thürich, J., Duplessis, S., Fernandez, D., et al. (2012). Expression profiling of genes involved in the biotrophic colonisation of *Coffea arabica* leaves by *Hemileia vastatrix*. *Eur. J. Plant Pathol.* 133, 261–277. doi: 10.1007/s10658-011-9864-5
- Voegelé, R. T., Hahn, M., and Mendgen, K. (2009). “The uredinales: cytology, biochemistry, and molecular biology,” in *The Mycota, 5. Plant Relationships*. Vol. 2, ed H. Deising (Berlin: Springer), 69–98.
- Voegelé, R. T., and Mendgen, K. (2003). Rust haustoria: nutrient uptake and beyond. *New Phytol.* 159, 93–100. doi: 10.1046/j.1469-8137.2003.00761.x

- Wang, Z. Y., Thornton, C. R., Kershaw, M. J., Debaio, L., and Talbot, N. J. (2003). The glyoxylate cycle is required for temporal regulation of virulence by the plant pathogenic fungus *Magnaporthe grisea*. *Mol. Microbiol.* 47, 1601–1612. doi: 10.1046/j.1365-2958.2003.03412.x
- Winnenburg, R., Baldwin, T. K., Urban, M., Rawlings, C., Köhler, J., and Hammond-Kosack, K. E. (2007). PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res.* 34, D459–D464. doi: 10.1093/nar/gkm858
- Xue, C., Park, G., Choi, W., Zheng, L., Dean, R. A., and Xu, J. R. (2002). Two novel fungal virulence genes specifically expressed in appressoria of the rice blast fungus. *Plant Cell* 14, 2107–2119. doi: 10.1105/tpc.003426
- Yamauchi, J., Takayanagi, N., Komeda, K., Takano, Y., and Okuno, T. (2004). cAMP-pKA signaling regulates multiple steps of fungal infection cooperatively with Cmk1 MAP kinase in *Colletotrichum lagenarium*. *Mol. Plant Microbe Interact.* 17, 1355–1365. doi: 10.1094/MPML.2004.17.12.1355
- Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21, 650–659. doi: 10.1093/bioinformatics/bti042
- Yang, Z., and Dickman, M. B. (1999). *Colletotrichum trifolii* mutants disrupted in the catalytic subunit of cAMP-dependent protein kinase are nonpathogenic. *Mol. Plant Microbe Interact.* 12, 430–439. doi: 10.1094/MPML.1999.12.5.430
- Yin, C., Chen, X., Wang, X., Han, Q., Kang, Z., and Hulbert, S. H. (2009). Generation and analysis of expression sequence tags from haustoria of the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici*. *BMC Genomics* 10:626. doi: 10.1186/1471-2164-10-626
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 22 January 2014; paper pending published: 06 February 2014; accepted: 24 February 2014; published online: 14 March 2014.
- Citation: Talhinhas P, Azinheira HG, Vieira B, Loureiro A, Tavares S, Batista D, Morin E, Petitot A-S, Paulo OS, Poulain J, Da Silva C, Duplessis S, Silva MC and Fernandez D (2014) Overview of the functional virulent genome of the coffee leaf rust pathogen *Hemileia vastatrix* with an emphasis on early stages of infection. *Front. Plant Sci.* 5:88. doi: 10.3389/fpls.2014.00088
- This article was submitted to *Plant-Microbe Interaction*, a section of the journal *Frontiers in Plant Science*.
- Copyright © 2014 Talhinhas, Azinheira, Vieira, Loureiro, Tavares, Batista, Morin, Petitot, Paulo, Poulain, Da Silva, Duplessis, Silva and Fernandez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.