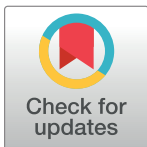


RESEARCH ARTICLE

Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression dynamics of candidate effectors dependent on host compatibility

Brenda Neves Porto¹, Eveline Teixeira Caixeta^{2*}, Sandra Marisa Mathioni³, Pedro Marcus Pereira Vidigal⁴, Laércio Zambolim⁵, Eunize Maciel Zambolim⁵, Nicole Donofrio⁶, Shawn W. Polson⁷, Thiago Andrade Maia⁵, Chuming Chen⁷, Modupe Adetunji⁷, Brewster Kingham⁸, Ronaldo José Durigan Dalio⁹, Mário Lúcio Vilela de Resende^{3*}



OPEN ACCESS

Citation: Porto BN, Caixeta ET, Mathioni SM, Vidigal PMP, Zambolim L, Zambolim EM, et al. (2019) Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression dynamics of candidate effectors dependent on host compatibility. PLoS ONE 14(4): e0215598. <https://doi.org/10.1371/journal.pone.0215598>

Editor: Richard A. Wilson, University of Nebraska-Lincoln, UNITED STATES

Received: November 23, 2018

Accepted: April 4, 2019

Published: April 18, 2019

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The Genome project is deposited at NCBI, BioProject ID: PRJNA419278 and BioSample ID: SAMN08048888. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession PHNK00000000. The version described in this paper is version PHNK01000000. The gff file containing the sequences of contigs and their respective annotations can be accessed in Figshare under DOI [10.6084/m9.figshare.7940411](https://doi.org/10.6084/m9.figshare.7940411) or direct URL:<https://doi.org/10.6084/m9.figshare.7940411>.

1 Programa de Pós-graduação em Biotecnologia Vegetal, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil, **2** Empresa Brasileira de Pesquisa Agropecuária (Embrapa-Café), Brasília, Distrito Federal, Brazil, **3** Departamento de Fitopatologia, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil, **4** Núcleo de Análises de Biomoléculas, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil, **5** Departamento de Fitopatologia, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil, **6** Department of Plant and Soil Sciences, University of Delaware, Newark, Delaware, United States of America, **7** Center for Bioinformatics and Computational Biology, Delaware Biotechnology Institute, Newark, Delaware, United States of America, **8** Sequencing and Genotyping Center, Delaware Biotechnology Institute, University of Delaware, Newark, Delaware, United States of America, **9** Instituto Agrônômico de Campinas, Centro de Citricultura “Sylvio Moreira”, Cordeirópolis, São Paulo, Brazil

* eveline.caixeta@embrapa.br (ETC); mlucio@dfp.ufla.br (MLVR)

Abstract

Coffee leaf rust caused by the fungus *Hemileia vastatrix* is one of the most important leaf diseases of coffee plantations worldwide. Current knowledge of the *H. vastatrix* genome is limited and only a small fraction of the total fungal secretome has been identified. In order to obtain a more comprehensive understanding of its secretome, we aimed to sequence and assemble the entire *H. vastatrix* genome using two next-generation sequencing platforms and a hybrid assembly strategy. This resulted in a 547 Mb genome of *H. vastatrix* race XXXIII (Hv33), with 13,364 predicted genes that encode 13,034 putative proteins with transcriptomic support. Based on this proteome, 615 proteins contain putative secretion peptides, and lack transmembrane domains. From this putative secretome, 111 proteins were identified as candidate effectors (EHv33) unique to *H. vastatrix*, and a subset consisting of 17 EHv33 genes was selected for a temporal gene expression analysis during infection. Five genes were significantly induced early during an incompatible interaction, indicating their potential role as pre-haustorial effectors possibly recognized by the resistant coffee genotype. Another nine genes were significantly induced after haustorium formation in the compatible interaction. Overall, we suggest that this fungus is able to selectively mount its survival strategy with effectors that depend on the host genotype involved in the infection process.

Funding: This work has been funded by the Brazilian Coffee Research and Development Consortium (Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – CBP&D/Café), by the Foundation for Research Support of the State of Minas Gerais (FAPEMIG), by the National Council of Scientific and Technological Development (CNPq), by the National Institutes of Science and Technology of Coffee (INCT/Café), by EMBRAPA Café, UFLA, UFV, and UDEL. BNP is granted by CAPES. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Coffea arabica (Arabica coffee) and *C. canephora* (Conilon coffee), are the two most cultivated coffee species in Brazil, which is the leading producer of coffee with approximately 43 million bags harvested in 2015 [1]. Due to the high quality of its beverage, *C. arabica* is the species most cultivated worldwide. However, a major drawback is that it is highly susceptible to *Hemileia vastatrix* Berkeley and Broome (1869), the coffee leaf rust pathogen. The fungus is widespread in coffee growing regions causing significant yield losses and its chemical control can lead to increased production costs. The best control strategy for the disease has been the use of resistant cultivars [2] [3], which have been developed through breeding programs in various countries. Another factor aggravating this disease is that the fungus has been able to overcome the resistance of recently released cultivars and has more than 50 physiological races described globally [3] [4]. In Brazil, previous studies differentiated 15 races named as I, II, III, VII, X, XIII, XV, XVI, XVII, XXII, XXIII, XXIV, XXV or XXXI, XXXIII, and XXXVII [3]. Race XXXIII was recently identified in Brazil [5] and overcame the resistance of cultivars Catimor, Tupi Amarelo and Catucaí Amarelo, which are derived from either Icatu or Híbrido de Timor, the major rust resistance parental donors [6] [7] [8] [3]. In this way, the presence of the rust virulent race XXXIII in the field may impose a serious threat to any coffee orchard worldwide.

The infection process of *H. vastatrix* on coffee leaves starts with the germination of urediniospores on the abaxial leaf surface, formation of appressoria over a stoma and subsequent penetration, spread of penetration hypha into the substomatal chamber, and further inter and intracellular colonization. At the tip of the penetration hypha, two thick lateral branches resembling an anchor are formed, a unique characteristic of *H. vastatrix*. Haustorial mother cells (HMC) are formed at each lateral branch of the anchor, giving rise to haustoria [9].

Fungal structures can develop at different times depending on the host genotype. In susceptible coffee cultivars, hyphae are formed from the majority of the infection sites, thus resulting in the formation of a large number of haustoria in the palisade and spongy parenchyma as well as in the upper epidermal cells. In these compatible interactions, where disease is the outcome, the infection process culminates with the colonization of mesophyll cells and a urediniosporic sporulation after 20 days post penetration [10] [4]. In resistant coffee cultivars, fungal growth is ceased usually after the formation of the first haustorium [11], and this is denoted as post-haustorial resistance. Other coffee genotypes may present pre-haustorial resistance as observed in the field and similar to non-host resistance, which prevents formation of haustoria [12] [13] [9].

Similar to other pathogenic fungi, *H. vastatrix* establishes its parasitic colonization by secreting effector proteins that modify the structure and function of host cells. Some of the effectors may also manipulate the plant immunity to increase pathogen fitness as observed in other plant-pathogen interactions [14] [15] [16]. R proteins encoded by resistance genes can recognize some of these effector proteins and thus trigger plant defense responses [17] [18]. The identification and characterization of these effectors, as well as their correlated R genes, are the first steps for understanding the molecular mechanisms underlying the gene-for-gene interaction and for a more focused breeding program for resistance to coffee leaf rust. Thus, it is crucial that genome and transcriptome data are available for a more thorough investigation and analysis of fungal effector proteins.

Over the past 10 years, advances in Next Generation Sequencing (NGS) technologies and decreasing sequencing costs allowed an increase in the number of sequenced genomes, especially of plant pathogenic fungi causing important diseases in major agricultural crops. To date, genomes of five important rust fungi have been published: *Melampsora larici-populina* (101.1 Mb, [19]), *Puccinia graminis* f. sp. *tritici* (88.6 Mb, [19]; [20]), *Puccinia striiformis* f. sp. *tritici* (64.8 Mb, [21]), *Melampsora lini* (189 Mb, [22]), and a partial draft genome of *H.*

vastatrix (333 Mb, [23]). Genome-wide analyses of these rusts revealed a large catalog of secreted proteins, some of which are candidate effectors.

In general, and when compared to non-biotrophic fungi, rust fungal species have large genome sizes, including *H. vastatrix* among the largest genomes of rust fungi. The large size of the fungal nuclear genome was measured by flow cytometry, as ranging from 733.5 Mb [24] to 796.8 Mb [25]. However, the lack of a well-assembled genome and the unknown full repertoire of effector proteins has significantly impeded the investigation of *H. vastatrix* biology and biotrophic interaction with its host as well as slowing the progress of a focused breeding program for resistant cultivars.

Therefore, the main goal of this study was to deep sequence the genome of *H. vastatrix* urediniospores of race XXXIII using a combined sequencing strategy with long (PacBio) and short (Illumina) reads and an integrated genome assembly method for achieving a high-quality reference genome. We used *in silico* methods to predict secreted proteins, and a subset of their corresponding genes were subject to time course-based, gene expression studies. Together, our data provide the most comprehensively sequenced and assembled coffee rust genome to date and an initial description and characterization of the putative secretome.

Material and methods

Fungal material

The *H. vastatrix* monopustular isolate Hv-02, characterized as race XXXIII, was used in this work. It was massively multiplied on coffee seedlings, *Coffea arabica* cv. Caturra (CIFC 19/1), as described previously [26] to result in prolific urediniospore production. Conditions (detailed below) were created to ensure the existence of sufficient concentration of fungal suspension for successive inoculations on seedlings. The fungal suspension was obtained by adding 2 mg of urediniospores in 2.0 ml sterile microfuge tubes containing 1000 μ l of sterile distilled water. The suspension was vortexed for 20 minutes and centrifuged at 2,500 \times g for 10 minutes. These procedures were performed three times to ensure the removal of impurities. Spores were uniformly spread on polyethylene Petri dishes containing a thin layer of agar and water. These Petri dishes were stored at 22°C for 16 hours in the dark to allow spore germination. After germination, urediniospores were scraped from Petri dishes, frozen in liquid nitrogen, and stored at -80°C until further processing. Liquid nitrogen was used to grind the samples for DNA/RNA extraction.

DNA and RNA extractions

Genomic DNA was extracted using the Mobio Power Microbial Maxi DNA Isolation kit (MOBIO Laboratories Inc., Carlsbad, CA, USA). After extraction, DNA concentration was measured using NanoDrop Spectrophotometer ND-2000 (Thermo Fisher Scientific Inc., Waltham, MA, USA), and its quality was determined by electrophoresis in agarose gel.

For RNA extraction, viable urediniospores were inoculated on two coffee cultivars: on the resistant coffee seedlings, Híbrido de Timor (CIFC 832/1), and on the susceptible coffee seedlings, *C. arabica* cv. Caturra (CIFC 19/1). Leaves were collected at 12, 24, 48 and 72 hours after inoculation and immediately frozen by immersion into liquid nitrogen. Samples were stored at -80°C until further processing. Total RNA was extracted using RNeasy Plant Mini kit (Qiagen, Hilden, Germany), according to manufacturer's instructions.

cDNA synthesis

The cDNA synthesis was performed using Impron II reverse transcriptase kit (Promega, Madison, WI, USA) with 1 μ g of total RNA added according to manufacturer's instructions.

The efficiency of the cDNA synthesis process was assessed by PCR using the following reference genes: β -tubulin and Cytochrome c oxidase subunit III. The integrity of PCR amplification products was verified by electrophoresis in 1.2% agarose gel.

Sequencing and assembly of the genome

Genomic DNA sequencing libraries of the race XXXIII (Hv33) were sequenced using two platforms: Pacific Biosciences PacBio RS II (Pacific Biosciences, Menlo Park, CA, USA), and Illumina HiSeq 2500 (Illumina, San Diego, CA, USA). All sequencing was performed at the University of Delaware Sequencing and Genotyping Center (Delaware Biotechnology Institute, Newark, DE, USA). Approximately 10 μ g of DNA from germinated urediniospores (protocol described above) were sheared to 5 kb fragments using the Covaris instrument (Covaris S2 Adaptive Focused Acoustic Disruptor with CryoPrep), and were used to build two single-end libraries, using PacBio Library Preparation kit, according to manufacturer's instructions, and sequenced in the Pacific Biosciences RSII Single-Molecule Sequencer. The first library was sequenced using 15 Single Molecule Real Time (SMRT) cells and combinations of two chemistries, namely, XL-C2 (11 SMRT) and P4-C2 (4 SMRT). The second library was sequenced based on 17 SMRT cells by using only the P5-C3 chemistry. For Illumina sequencing, approximately 10 μ g of DNA from germinated urediniospores (protocol described above and the same material used for the PacBio sequencing) was used to construct four libraries and sequenced with paired-end 151 bp reads.

FastQC (The Babraham Institute, Babraham, UK) was used to assess the quality of reads generated with the Illumina platform. CLC Genomics Server version 6.0.4 (CLC bio, Aarhus, Denmark) was used for quality trimming ($q < 0.001$ with no ambiguous nucleotides), adapter removal, and filtering of any reads shorter than 35bp. Overlapping paired end reads were merged previous to assembly using CLC. For the long reads generated with the PacBio platform, the quality of reads were assessed and filtered using default settings of PacBio SMRT Portal version 2.2.0.p2 (SMRT Analysis).

A hybrid genome assembly approach was used to obtain a consensus genome build: (i) assembly of the Illumina reads; (ii) assembly of the PacBio reads; and (iii) hybrid assembly approaches incorporating data from both platforms. The best of these approaches as determined by criteria including N50, longest contig/scaffolds, and read mapping accuracy criteria are summarized in Fig 1, although many additional approaches were also applied. Among the best performing approaches, high quality Illumina reads were separately *de novo* assembled using the CLC Genomics Workbench version 7.5 (CLC bio, Aarhus, Denmark) and the SOAP De-Novo version v2r215 [27] using kmer parameter sweeps, with the 121 bp kmer assembly from SOAP Denovo2 determined to be the best. PacBio long reads were also independently assembled using the PacBio SMRT Portal HGAP algorithm (versions 1 and 2; [28]), with version 2 performing best. The contigs from the HGAP assembly with the best assembly parameters were used as "Guide Contigs" for a new *de novo* assembly using CLC with the Illumina raw data.

The resulting set of scaffolds from both the SOAP *de novo* assembly and the HGAP/CLC Hybrid assembly were further joined and gaps were filled using PacBio long read data in PBJelly (v14.1.14; [29] in conjunction with the BLASR mapping tool ([30]; options: -minMatch 8 -minPctIdentity 70 -bestn 10 -nCandidates 20 -maxScore -500). Illumina raw data was mapped to the joined and filled scaffolds using CLC, and non-ambiguously mapping reads were used to correct nucleotide and InDel errors, typically in the PacBio-filled regions. Read mapping data was further utilized in CLC (Illumina reads) and PacBio SMRT Portal (PacBio reads) to identify potential mapping breakpoints and variations in genome coverage that

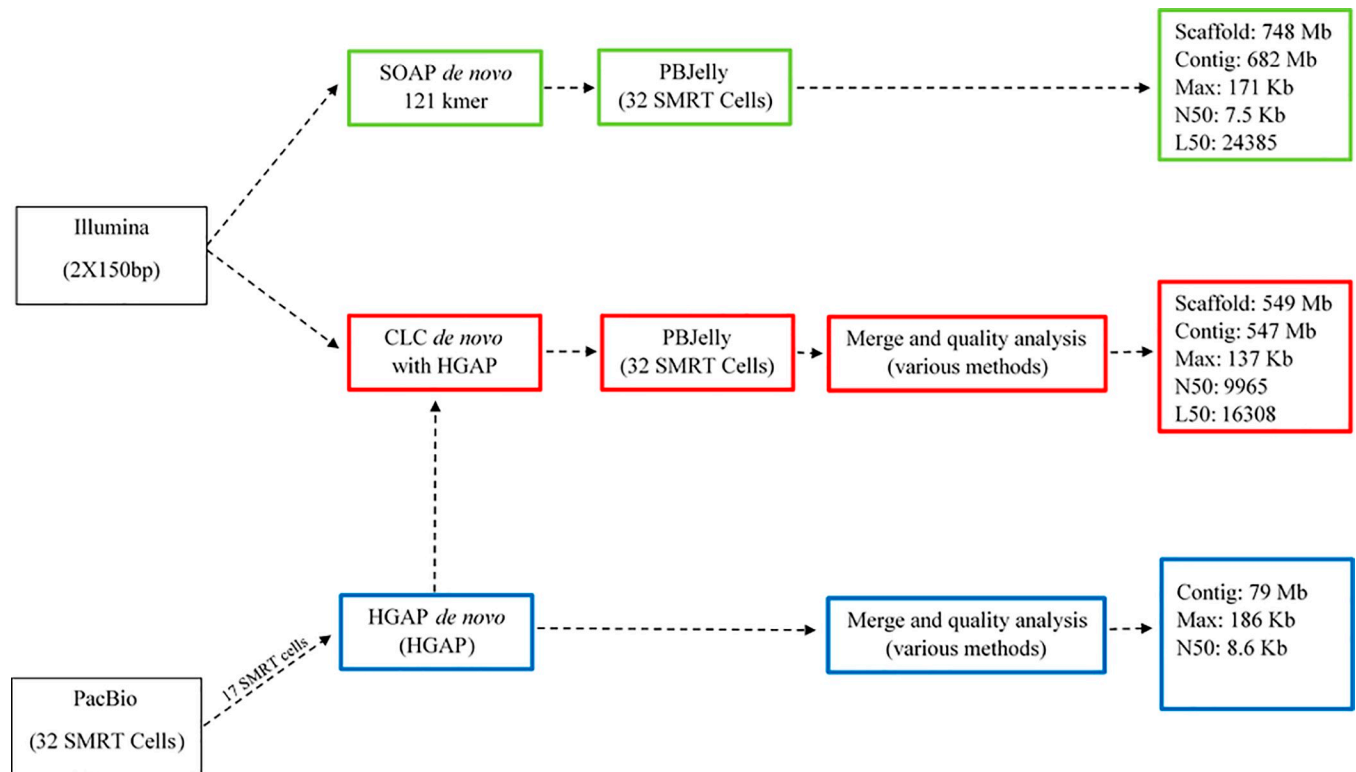


Fig 1. Schematic diagram describing different strategies used in the assembly the *H. vastatrix* genome (race XXXIII). Blue rectangles refer to the first assembly strategy, which used 17 SMRTcells with P5-C3 chemistry obtained on the PacBio RS2 platform in the HGAP assembler. Green rectangles refer to the second assembly strategy, which included a *de novo* assembly using reads obtained on the Illumina platform with SOAP *de novo* 2 assembler and subsequently improved with 32 SMRT cells of PacBio reads using PBJelly software. Red rectangles refer to the third strategy, which performed a *de novo* assembly using reads obtained from both next generation sequencing platforms with additional improvement with PBJelly software.

<https://doi.org/10.1371/journal.pone.0215598.g001>

might be indicative of mis-assembly, and manual genome finishing was performed utilizing available results. Illumina data from this study, and the 454 sequence data (NCBI SRA: SRR833265, SRR833266 –quality trimmed as previously described but to $q > 0.05$) used to generate the previously existing genome assembly (HvCat) (NCB: GCA_003057935.1) [23], were mapped to the Hv33 assemblies. Besides, by CLC Genomics Workbench, the HvCat assembly were mapped using both the default read mapping settings (80% identity over 50% read length) and a higher stringency setting (90% identity over 90% read length). Final SOAP and CLC/HGAP assemblies were evaluated using these data, as well as annotation results from each. The CLC/HGAP assembly performed significantly better, particularly in respect to percentage of Illumina reads which could be correctly aligned, as such this assembly is the one reported herein.

Annotation of genome and gene prediction

The structural annotation of genome and gene prediction were performed using the MAKER pipeline designed for emerging model organism genomes [31]. Our pipeline included masking of repeats using RepeatMasker (version 4.0.7; [32]) and the RepBase Puccinales library (build 1.24.19; [33]). This was followed by three iterative passes of gene calling using the Maker pipeline (version 2.31.10; [31]). First pass included all proteins from UniProt database under the order Puccinales (171,373 total proteins). The second pass combined transcripts assembled using CLC reference-guided Transcript Discovery (version 11.0.1), Trinity reference-guided

assembly [34], and Trinity *de novo* assembly of RNA-Seq data from three libraries of the same race of *H. vastatrix* [35]. Similarly, assembled transcripts from another *H. vastatrix* race HvCat downloaded from NCBI SRA (SRR1124793; [23]; and 352,146 NCBI downloaded ESTs from *H. vastatrix* CIFC isolate 178a (ERR106426; [36]) were used. All such sequence data from *H. vastatrix* races other than XXXIII was treated as being data from “related organisms” by Maker. Gene models from prior Maker iterations were used to train SNAP (GSL Biotech, Chicago, IL, USA) and Augustus (version 3.22; [37]) models, which were used along with a custom-trained GeneMark ES (version 2.5p; [38]) for *ab initio* gene calling in concert with the protein/transcript gene mapping data in Maker iterations 2 and 3. BUSCO (version 3.0.2; [39]) and CEGMA (version 2.5; [40]) were used to assess completeness of assembly/annotation with comparison to other genome assemblies from the order Puccinales–HvCat: *H. vastatrix* isolate HvCat (GCA_003057935.1), Mp: *Melampsora larici-populina* 98AG31 (GCF_000204055.1); Ml: *Melampsora lini* CH5 (JGI Genome portal: <https://genome.jgi.doe.gov/Melli1/Melli1.info.html>); Pg: *Puccinia graminis* f. sp. *tritici* CRL 75-36-700-3 (GCF_000149925.1); Ps: *Puccinia striiformis* f. sp. *tritici* (GCA_001936605.2) [19, 22]. BUSCO was run in both genome and protein modes and using the basidiomycota_odb9 conserved genes library.

Annotation of proteome

Closest known homologs for the predicted protein coding genes with transcriptome support from the *H. vastatrix* genome (Hv33) were determined using BLAST version 2.6.0 [41] to search the GenBank non-redundant database (GenBank nr) and UniProt Knowledgebase (UniProtKb) using BLASTp (e -value $< 10^{-5}$). The proteins were also aligned to the EuKaryotic Orthologous Groups (KOG) using Reverse Position-Specific BLAST (RPS-BLAST). The proteins were further annotated with gene ontology (GO) terms using Blast2GO [42].

Secretome analysis

The secretome of *H. vastatrix* was predicted using a pipeline combining different bioinformatics approaches. The proteins containing signal peptides were selected by SignalP version 4.0 [43] (cutoff D-Score = 0.45). The proteins flagged as related to secretory pathway signal peptide (SP) were selected by TargetP version 1.0 [44]. In addition, Wolf Psort [45] was also used to select proteins based on subcellular localization prediction. Then, the proteins without transmembrane domains were selected by TMHMM version 2.0 [46]. Following this, residues of cysteine and motifs characteristic of effectors, such as [YWF]xC, CxxC and CxxxC, were searched for in each secreted protein identified.

We predicted the secretome of four rust species (*P. graminis* f. sp. *tritici*, *P. striiformis* f. sp. *tritici*, *M. larici-populina* and *M. lini*) using the same pipeline that we used for predicting the secretome of *H. vastatrix*, and thus aiming an unbiased comparison. Then, the predicted secretomes were clustered based on sequence similarities using OrthoVenn [47] (e -value: 10^{-5} ; inflation value: 1.5).

Sequences of proteins of *H. vastatrix*, which did not have a hit, were compared with secreted proteins described in [36, 48] using the search tools BLASTp (e -value = 10^{-5}) and BLASTx (e -value = 10^{-10}), as described above. Those sequences were compared with RNA-Seq data described in [35], which were obtained by sequencing libraries from the *H. vastatrix*-coffee leaves interaction. These libraries were obtained using *C. arabica* cv. Caturra (CIFC 19/1) and Híbrido de Timor (CIFC 832/1) inoculated using fresh spores of *H. vastatrix* race XXXIII, seeking to establish compatible and incompatible interactions, respectively. In this study, only libraries obtained from incompatible interaction in two inoculation periods, i.e., 12 and 24 hours after inoculation were used.

Primer design

Primer Express 2.0 software (Applied Biosystem, Carlsbad, CA, USA) was used to design specific primers for genes that encode *H. vastatrix* candidate effectors (S1 Table).

The beta-tubulin (*β-tubulin*), cytochrome C oxidase subunit III (*CytIII*), 40S ribosomal protein, and glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) genes, previously used for expression analysis in *H. vastatrix* [49], were tested by RT-PCR as endogenous reference genes. The RT-PCR was carried out as an initial measure of amplified products using cDNA synthesized from RNA of pathogen samples, cDNA synthesized from RNA of coffee leaf samples only, and negative control consisting of nuclease-free water. The primers that showed amplification only for the cDNA from the pathogen were validated and used as endogenous reference genes.

Each 20 μL RT-PCR reaction volume included the following at final concentrations of: 1×buffer, 1.0 mM MgCl₂, 60 μM dNTP, 0.1 μM of each forward and reverse primers, 1 U Taq polymerase and 50 ng of cDNA sample. The PCR amplified products were separated by electrophoresis in 1.2% agarose gel.

Real-Time qPCR

Real-Time quantitative Polymerase Chain Reaction (RT-qPCR) was carried out in the ABI PRISM 7500 Real Time PCR Systems (Applied Biosystems, Carlsbad, CA, USA). The SYBR Green system (Thermo Fisher Inc., Waltham, MA, USA) was used to detect PCR amplified products. Each 10 μL RT-qPCR reaction volume contained 50 ng cDNA, 0.4 μM of each forward and reverse primers, 5 μL (50% v/v) of Power SYBR Green PCR Master Mix solution, and 2.2 μL of nuclease-free water. The thermal cycling conditions consisted of initial denaturation step at 95°C for 10 minutes, followed by 40 cycles at 94°C for 15 second and 60°C for 1 minute. At the end of each reaction, a dissociation curve analysis was performed by heating the amplicon from 60 to 95°C in order to confirm the specificity of the amplification. Standard curves using five points of cDNA serial dilution were used to optimize the concentration of cDNA and primer efficiencies. For each dilution point, cycle threshold value (Ct) efficiency was estimated for each primer, including target genes and selected reference genes.

Three technical replicates for each of three existing biological replications for each sample (resistant plants and susceptible plants) were performed at the following four time-points: 12, 24, 48, and 72 hours after inoculation (hai). The data were analyzed with REST software proposed by [50].

Results

Sequencing and assembly of *Hemileia vastatrix* genome

Fresh urediniospores of *H. vastatrix* race XXXIII incubated on polystyrene plates for 16 hours at 22°C in the dark (as described in Methods) resulted in approximately 80% germination. The sequencing of long read libraries rendered approximately 8.39 Gbp of data from 32 SMRT cells on the PacBio RS II platform. Following quality screening, 5.73 Gbp comprised of 920,326 long-reads were obtained. The N50 of PacBio reads was estimated at 8.6 kb with an expected genome coverage >20X. The Illumina HiSeq platform produced 264,721,502 short-reads. These reads totaled >100X coverage for the expected genome size of *H. vastatrix* (~733.5 Mb according to [24]).

Multiple strategies were tested for the hybrid genome assembly for the dual-platform sequencing, the most successful strategies are outlined in Fig 1. Our approach aimed to maximize advantages and minimize disadvantages of each strategy, completing gaps of scaffolds

and ensuring the best assembly of *H. vastatrix* genome. The CLC HGAP strategy was employed to take advantage of the PacBio long read data to produce long contigs (Fig 1, blue boxes). The 17 longest read SMRT cells (P5-C3 chemistry) were used for this approach, as computational limitations prevented all 32 from being used. This approach produced a significant number of long contigs, however they were error prone (particularly short insert/deletion) and did not cover the complete genome.

SOAP De novo2 assembler was used for an Illumina-only de novo assembly (Fig 1, green boxes). PBJelly on 32 SMRT cells of long read sequencing was used to fill scaffolds and join contigs. This assembly had the longest metrics in terms of total contig/scaffold size, with scaffold size closely reflecting the expected genome size, although a significant amount of total data was in very short contigs. This assembly also did not validate well as it produced fewer annotated gene calls than other assemblies, a higher number of missing (20.2%) and fragmented (19.8%) conserved single-copy genes (BUSCO). Besides, reference mapping of the Illumina data back to the contigs resulted in higher rates of broken read pairs (61.9%) and a low rate of reference coverage (33% of the reference did not align reads) relative to other strategies.

The best results, were achieved by a hybrid *de novo* genome assembly approach (Fig 1, red boxes). Contigs that were produced using the HGAP assembly of PacBio long-read data (Fig 1, blue boxes) were leveraged as guide reads for a subsequent assembly of Illumina short-reads using the CLC Genomic Workbench. These resulting scaffolds were further filled and joined using PBJelly software, and a final error polishing step mapping Illumina reads back to the improved assembly. This strategy resulted in a final assembled genome of 549 Mbp in 118,162 scaffolds (< 0.5% gaps) (Table 1). Reference mapping of the approximately 260 million Illumina reads (130M read pairs) back to this genome assembly result in a 95.3% mapping rate. Interestingly, mapping of the same reads to only the contigs >2000 bp resulted in a very comparable mapping rate of 94.4%.

Genome annotation and gene prediction

Gene model predictions and gene product annotations were performed on the Hv33 genome assembly. Repetitive elements were identified in 118,162 scaffolds using the Repeat Masker software [51]. This analysis revealed that 82% of scaffolds contained repetitive elements, with

Table 1. Genomic features for the *Hemileia vastatrix* assembled genome (race XXXIII).

Genome Feature	Number
Total length	549,560,213 bp*
Number of scaffolds	118,162
Longest scaffold	137,364 bp
Shortest scaffold	143 bp
N50	9,965 bp
L50	16,308
Count (% length) scaffolds > 1Kb	92,276 (96.0%)
Count (% length) scaffolds > 2Kb	58,535 (87.5%)
Count (% length) scaffolds > 10Kb	16,236 (49.9%)
Count (% length) scaffolds > 50Kb	86 (1.0%)
A+T	66%
G+C	33.6%
N	0.44%

*bp—base pairs

<https://doi.org/10.1371/journal.pone.0215598.t001>

transposable elements accounting for 43.6% of the total. Gypsy was found to be the most represented repetitive element in this genome (Table 2).

Annotation of the genome with Maker using custom-trained Augustus and SNAP gene models identified 33,483 predicted genes in the *H. vastatrix* genome with 33,153 putative protein coding genes supported by transcriptomic and/or genomic evidence from existing genome annotations within the order Pucciniales. Of these predicted genes, 13,364 protein coding genes could be supported by the currently available *H. vastatrix* transcriptomic data available with a Maker AED score less than 1. The CEGMA analysis found that 96.4% of expected core genes were found in the genome, while BUSCO identified 91.7% (13.1% fragmented) of expected single copy genes (from 1335 conserved Basidiomycota proteins) in the protein annotations (S1 Fig). Of the expected BUSCO single-copy genes, 3.7% were duplicated which is in line with or lower than four existing reference genome assemblies from other species in the order Pucciniales. Additionally, reference mapping of our Illumina sequence reads back to the Hv33 genome result in 95.3% mapping rate, indicating that a significant fraction of the genomic sequence space is included in the Hv33 assembly.

Functional annotation of proteome

Putative protein coding genes with transcriptomic support were compared against fungal protein sequences in the GenBank nr and UniProt databases. Results showed that 1,568 sequences were likely unique to *H. vastatrix* as these sequences did not have sequence similarity to any other sequence in the queried databases. Of sequences with a blast hit ($e\text{-value} < 10^{-5}$), 72.2% were most similar to *Puccinia graminis*, 27.0% were more similar to *Melampsora larici-populina*, and 0.8% were most similar to sequences from *H. vastatrix* that had already been deposited in the databases.

A functional clustering analysis was performed by using the Blast2GO program for 11,466 protein sequences found to be significantly similar ($e\text{-value} < 10^{-5}$) to those obtained from GenBank and Uniprot databases. From these sequences, 7,139 were further computationally characterized. A Gene Ontology hierarchical analysis revealed 12,501 annotation clusters in categories of molecular function, biological process, and cellular component (some proteins were annotated in two or more categories). The largest number of annotated genes (6,331) had

Table 2. Structural annotation analysis of *Hemileia vastatrix* genome (race XXXIII).

Type of repetitive element	Number of repetitive elements
Retrotransposons	
<i>Gypsy</i>	41,269
<i>Copy</i>	7,054
<i>LINEs</i>	426
<i>DIRs</i>	81
Transposons	
<i>EnSpm</i>	2,971
<i>Harbinger</i>	2,954
<i>Mariner</i>	837
<i>Helitron</i>	181
<i>MuDR</i>	16
<i>Hat</i>	10
Others	
Simple repeats	39,808
Low complexity	32,461

<https://doi.org/10.1371/journal.pone.0215598.t002>

a molecular function, with ‘ion binding’ and ‘hydrolase activity’ being the most represented GO categories with 889 and 538 proteins, respectively (S2A Fig). A total of 3,887 annotated genes had biological processes, followed by 2,283 annotated genes with a cellular component assigned. For the biological process category, ‘cellular metabolic process’, ‘primary metabolic process’ and ‘organic substance metabolic process’ were more represented with 1,243, 1,086 and 1,086 protein sequences, respectively (S2B Fig). For the cellular component category, the following GO-terms were most represented: ‘cell part’, ‘membrane-bounded organelle’, ‘protein complex’ and ‘non-membrane-bounded organelle’, comprising 1,326, 669, 511, and 464 proteins, respectively (S2C Fig). The analysis performed for the hierarchical levels numbers 6 and 8, led us to confirm that as these proteins are essentially associated with the nucleus, cytosol, ribosome, mitochondria, Golgi apparatus, and endoplasmic reticulum (S2D Fig).

Prediction and functional annotation of secretome

To identify potentially secreted proteins, we applied a sequential bioinformatics pipeline that employed SignalP, TargetP, and TMHMM network-based tools to analyze the predicted proteins with transcriptome support. Results revealed 615 putative secreted proteins. But those numbers drop to 452 after redundancy analysis. This subset of proteins was significantly similar (e -value $< 10^{-5}$) mainly to proteins from *P. graminis* f. sp. *tritici* (385) and *M. larici-populina* (109). In addition, 39 protein sequences showed similarity with previously deposited sequences from *H. vastatrix*, and 72 returned no results, indicating their uniqueness to the *H. vastatrix* race XXXIII genome.

The Blast2GO program was used for functional clustering of 543 proteins found to be similar to sequences of proteins obtained from the NCBI and UNIPROT databases. From these proteins, 161 were clustered in categories of molecular function. The largest number of proteins was associated with transferase activity, ion binding and hydrolase activity (Fig 2). By Wolf Psort analyses (Table 3), most of the secreted proteins had subcellular location assigned as mitochondria (241) and extracellular space (186).

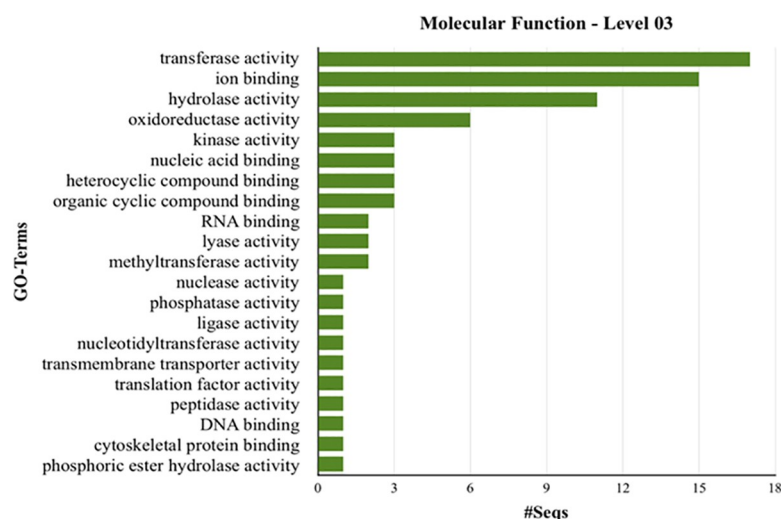


Fig 2. Functional clustering analysis of secreted proteins identified in *H. vastatrix* genome, race XXXIII. Secreted proteins were found to be significantly similar (e -value $< 10^{-5}$) to protein sequences obtained from NCBI and UNIPROT databases by using the Blast2GO bioinformatics platform. The y-axis consists of GO-terms described in the molecular function category for the hierarchical level #3 and the x-axis consists of protein sequences found for each GO-term in this category.

<https://doi.org/10.1371/journal.pone.0215598.g002>

Table 3. Prediction of subcellular localization of *Hemileia vastatrix* (race XXXIII) secretome using Wolf Psort.

Subcellular localization prediction	Number of proteins	Percentage (%)
Cytosol	26	4.23
Endoplasmic reticulum	3	0.49
Extracellular space	186	30.24
Golgi	1	0.16
Mitochondria	241	39.19
Nucleus	88	14.31
Peroxisome	37	6.02
Plasma membrane	26	4.23
No hit	7	1.14

<https://doi.org/10.1371/journal.pone.0215598.t003>

The KOG category analysis showed that approximately 27% of 615 secreted proteins have specific signatures and the most represented being replication, recombination and repair, transcription, and amino acid transport and metabolism (Fig 3).

We further analyzed whether the putative secreted proteins of *H. vastatrix* race XXXIII contained other effector-like sequences for the presence of cysteine residues, which are indicative signs of secretion in other plant pathogens. Cysteine residues were found in approximately 93% of the putative secreted proteins, and the majority contained either two, three, or four residues in their sequences. In addition, the amino acid motifs commonly found in fungal effectors, [YWF]xC, CxxC and CxxxC, were found in 365 of the secreted protein, either alone or grouped.

Transcriptional analysis of potential effector proteins

Seventeen genes that encode potential effector proteins identified in this study were selected for gene expression analysis using real-time quantitative RT-PCR (Figs 4–6). Some of these selected genes did not match with any other sequence in the queried databases (see methods for details) and other genes were identified as secreted protein of *H. vastatrix*, thus suggesting they are likely unique to *H. vastatrix*. These 17 potential effector proteins have predicted secretion signal and cysteine residues in their amino acid sequences and were also selected by

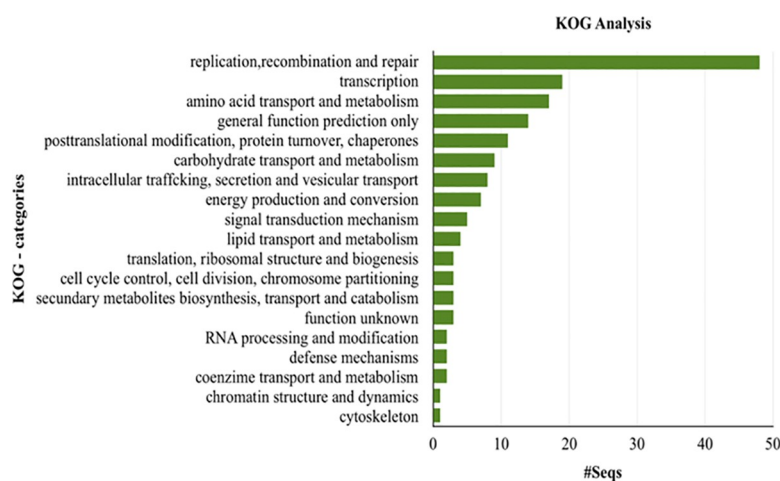


Fig 3. Functional clustering analysis of secreted proteins found in the genome of *H. vastatrix* race XXXIII, which specific signatures in KOG categories. The y-axis consists of KOG categories, and the x-axis consists of protein sequences found for each KOG category.

<https://doi.org/10.1371/journal.pone.0215598.g003>

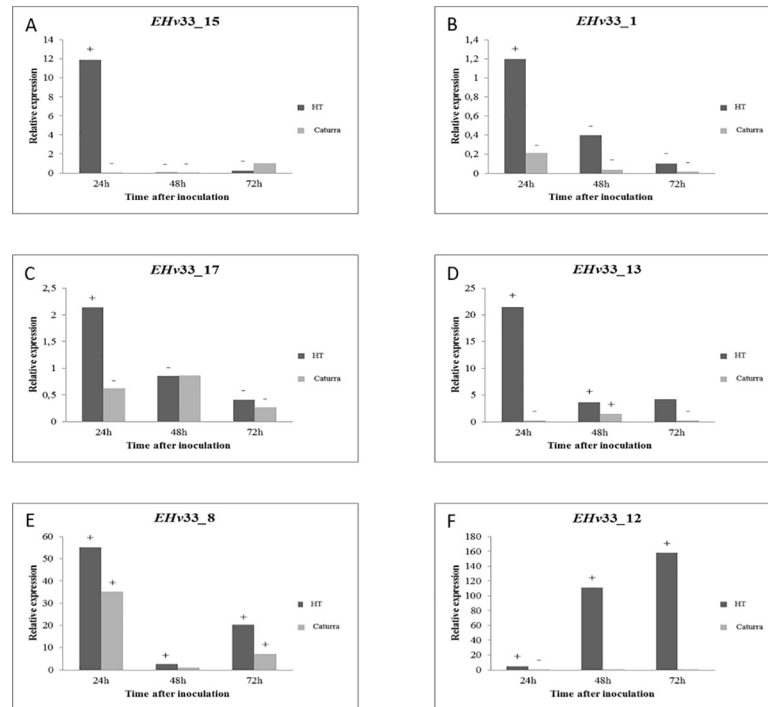


Fig 4. Quantitative RT-PCR-based analysis of gene expression performed for six EHV33 genes that encode putative candidate secreted effector proteins of *H. vastatrix*, race XXXIII. The relative expression pattern of target genes was estimated in plant samples of the hybrid of Timor and Caturra. Data were recorded at 12, 24, 48 and 72 hours after inoculation. The period of 12 hours after inoculation was used as reference sample. The expression level of target genes was normalized by using two endogenous genes of *H. vastatrix*, namely, β -tubulin and CytIII. (A) EHV33_15: the highest level of gene expression was recorded at 24 hours after inoculation (hai). (B) EHV33_1: the highest level of gene expression was recorded at 24 hai and, then, decreased over time. (C) EHV33_17: the highest level of gene expression was recorded at 24 hai and, then, decreased over time. (D) EHV33_13: the highest level of gene expression was recorded at 24 hai and, at the end, the expression estimate kept constant at 72 hai. (E) EHV33_8: the highest level of gene expression was recorded at 24 hai and, then, decreased at 48 hai and increased again at 72 hai. (F) EHV33_12: the increase of gene expression was recorded at 48 hai but, the highest level of gene expression was only recorded at 72 hai.

<https://doi.org/10.1371/journal.pone.0215598.g004>

BLAST hits to the database of *H. vastatrix* secretome from previously published studies [36, 48] and from a database of RNA-Seq libraries of resistant coffee infected with *H. vastatrix* (S2 Table).

Six of these genes were significantly expressed in the incompatible interaction (Fig 4) and nine were expressed in the compatible interaction (Fig 5 and Fig 6). Two genes, EHV33_3 and EHV33_14, did not show significant differences between compatible or incompatible interactions (S3 Fig).

A peak in expression was observed at 24 hours after inoculation (hai) in the incompatible interaction (no disease) for the following putative effector genes: EHV33_15, EHV33_1, EHV33_17, EHV33_13 and EHV33_8 (Fig 4A–4E). The EHV33_12 gene showed an expression increase at 48 hai and, the highest expression level of this gene was at approximately 72 hai (Fig 4F). The remaining EHV33 genes showed higher expression levels in the compatible interaction (Figs 5 and 7), where disease is the outcome. Genes EHV33_6, EHV33_5, EHV33_9 and EHV33_11 (Fig 5A–5D) showed a peak in expression at 24 hai; and EHV33_7, EHV33_2, EHV33_10 and EHV33_4 (Fig 6A–6D) at 48 hai, followed by a decrease in expression at 72 hai. The EHV33_16 gene showed significant expression level at 24 hai, followed by a gradual increase over time. The highest transcript accumulation was recorded at 72 hai (Fig 6E). A

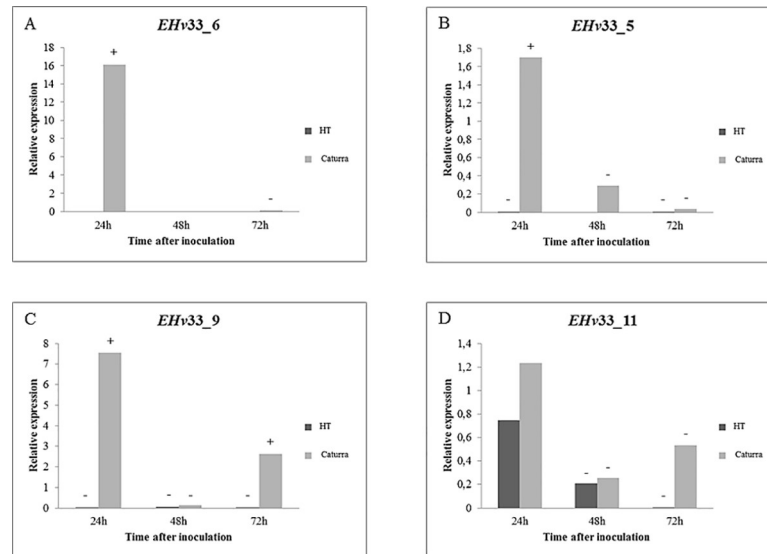


Fig 5. Quantitative RT-PCR-based analysis of gene expression performed for four EHv33 genes that encode putative candidate secreted effector proteins of *H. vastatrix*, race XXXIII. The relative expression pattern of target genes was estimated in plant samples of the hybrid of Timor and Caturra. Data were recorded at 12, 24, 48 and 72 hours after inoculation. The period of 12 hours after inoculation was used as reference sample. The expression level of target genes was normalized by using two endogenous genes of *H. vastatrix*, namely, β -tubulin and CytIII. (A) EHv33_6: the highest level of gene expression was recorded at 24 hours after inoculation (hai). (B) EHv33_5: the highest level of gene expression was recorded at 24 hai and, then, decreased over time. (C) EHv33_9: the highest level of gene expression was recorded at 24 hai and, then, decreased at 48 hai, and increased again at 72 hai. (D) EHv33_11: the highest level of gene expression was recorded at 24 hai and, then, decreased at 48 hai, and increased again at 72 hai.

<https://doi.org/10.1371/journal.pone.0215598.g005>

summary of these data is showed as a heat map (Fig 7) in susceptible (Caturra, Fig 7A) and resistant plants (Híbrido de Timor, Fig 7B). The heat map was generated by a log transformation of the RT-qPCR data presented as $\Delta\Delta CT$.

Discussion

H. vastatrix race XXXIII genome sequence summary

We utilized PacBio RS II and Illumina HiSeq to sequence the genome of the coffee rust fungus, *H. vastatrix* race XXXIII. This hybrid assembly approach of long and short reads yielded a genome approximately 576 Mb in size. We have noted that the population of contigs shorter than 2 kb appear to be dominated by fragments of regions found elsewhere in the assembly, but with enough sequence heterogeneity to exclude them from the longer contigs and to justify retaining them in this assembly. While such repetitive sequence data may represent allelic diversity, an alternative hypothesis is that this genome has undergone fairly large-scale duplication events. Such events could not only contribute to the difficulties that have arisen in assembling this genome, but could also explain how > 95% of sequence reads can recruit to a genome assembly that is 200-300Mbp shorter than the experimentally-predicted genome size of 733.5 Mb [24] to 796.8 Mb [25].

The combined use of a single monopustular isolate and long read sequencing (PacBio) in our approach was fundamental for obtaining this assembled genome with a higher level of contiguity ($N_{50} = \sim 10.0$ kb; Table 1), than the previous assembly of HvCat genome ($N_{50} = \sim 1.6$ kb; total genome 330 Mbp [23]). While the Hv33 genome assembly remains fragmented, it exhibited increased genome coverage, contiguity, and performs better for reference mapping of high-throughput sequencing data than the previous assembly. Reference mapping of the

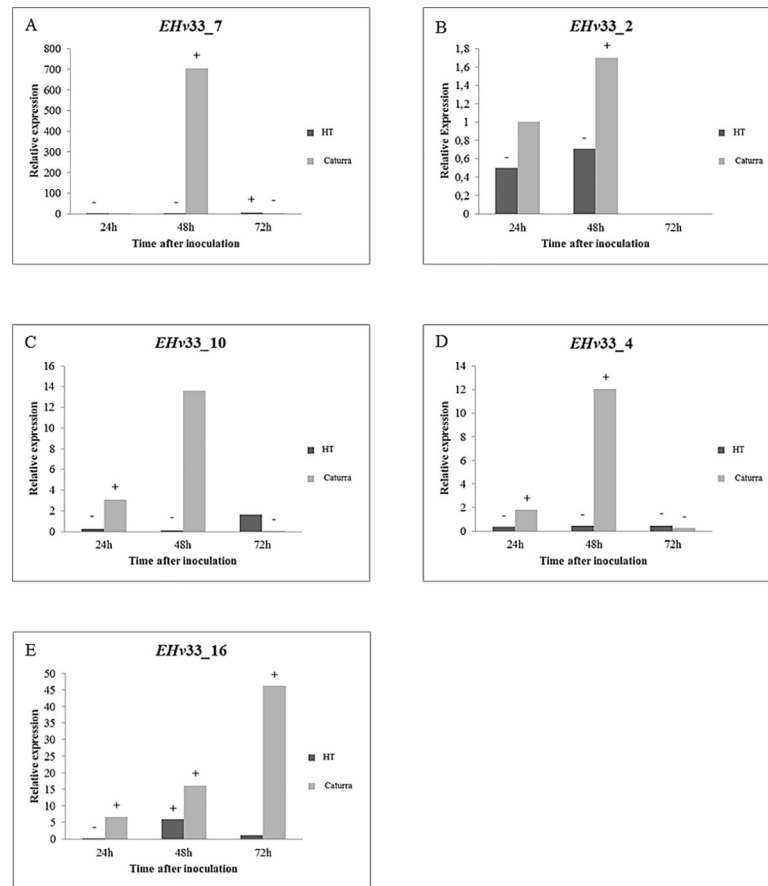


Fig 6. Quantitative RT-PCR-based analysis of gene expression performed for five EHv33 genes that encode putative candidate secreted effector proteins of *H. vastatrix*, race XXXIII. The relative expression pattern of target genes was estimated in plant samples of the hybrid of Timor and Caturra. Data were recorded at 12, 24, 48 and 72 hours after inoculation. The period of 12 hours after inoculation was used as reference sample. The expression level of target genes was normalized by using two endogenous genes of *H. vastatrix*, namely, β -tubulin and CytIII. (A) EHv33_7: the highest level of gene expression was recorded at 48 hours after inoculation (hai). (B) EHv33_2: the highest level of gene expression was recorded at 48 hai and, then, decreased over time. (C) EHv33_10: the highest level of gene expression was recorded at 48 hai and, then, decreased over time. (D) EHv33_4: the highest level of gene expression was recorded at 48 hai and, then, decreased over time. (E) EHv33_16: the highest level of gene expression was recorded at 72 hai.

<https://doi.org/10.1371/journal.pone.0215598.g006>

underlying sequence data from the HvCat assembly to Hv33 resulted in 80.8% of reads mapping, while a reciprocal analysis mapping our raw data to the HvCat genome resulted in a mapping rate of 60.4%. Other indicators including the >95% rate of raw data mapping to the Hv33 genome and the BUSCO results indicating that >90% of conserved single copy proteins are present, with a rate of duplication in line with other rust genomes (S1 Fig), also point to an assembly that covers the majority of the genomic sequence space.

H. vastatrix genome has been estimated to be one of the largest fungal genomes, only smaller than those predicted based on flow cytometry for *Gymnosporangium confusum* (893.2 Mb) and *Puccinia chrysanthemi* (806.5 Mb) [25], however these genomes have not yet been sequenced. Other economically important rusts, such as *P. graminis* f.sp. *tritici* (88.6 Mb), *P. striiformis* f.sp. *tritici* (64.8 Mb), *M. larici-populina* (101,1Mb), and *M. lini* (189 Mb), have relatively smaller genomes [19] [21] [22]. Parameters such as lifestyle, intraspecific genetic variability, selection pressure, and genetic drift, are considered determining factors for fungal genome size [52] [53].

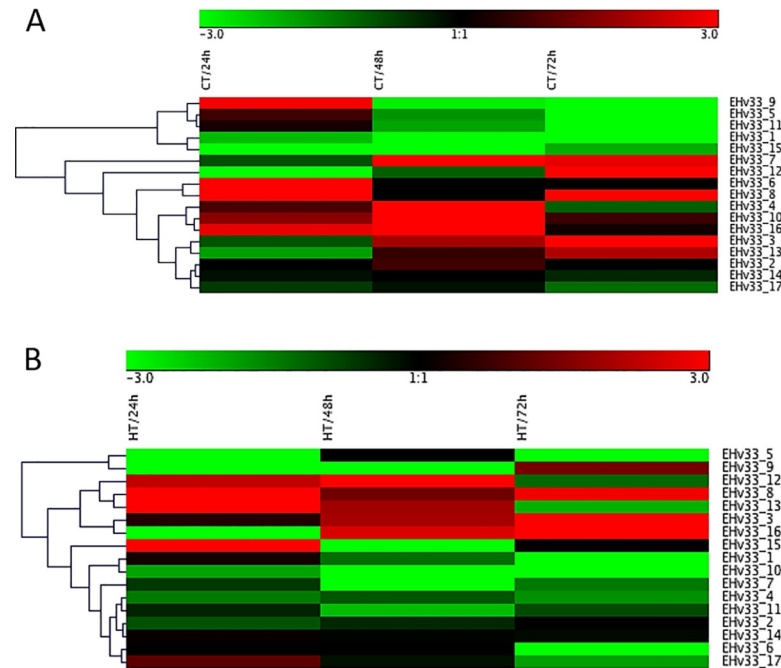


Fig 7. Heat map of 17 genes expression profile obtained from quantitative real time PCR. (A): genes expression profile (17) in resistant plant (Híbrido de Timor) during 24, 48 and 72 hai. (B): genes expression profile (17) in susceptible plant (Caturra) during 24, 48 and 72 hai. Red indicates a relatively high level of up-regulation, whereas green indicates a relatively low level of up-regulation.

<https://doi.org/10.1371/journal.pone.0215598.g007>

Having large genomes does not mean having more genes, as genome expansion may occur not only by acquisition of new genes, but also by increasing repetitive regions or by transposable element proliferation during the evolutionary process [52]. Our study identified about 82% of scaffolds contained repetitive elements, from which 43.6% consisted of transposable elements (Table 2). Studies involving different species responsible for leaf rusts, as *P. striiformis* f.sp. *tritici*, *P. graminis* f.sp. *tritici*, *M. larici-populina*, *M. lini* and *H. vastatrix*, showed that 17.8%, 42.5%, 40.5%, 31.67% and 48% of repetitive elements, respectively, consist of transposable elements [19] [21] [23] [22]. Differences in terms of transposable elements content may occur due to a very large genome size, which can contain large proportion of repetitive regions; and methods for hybrid genome assembly used, which may overestimate the real number of repetitive regions. The transposable elements are important sources of variability that occurs by means of mutations during the evolutionary process. The short lifespan of plant pathogenic organisms, however, makes them reproduce more often than their hosts, consequently evolving faster. Because of this, it is suggested that *H. vastatrix* has a great capacity of adaptation and high genetic variability during its co-evolution with the host over time.

The high similarity (~72%) between predicted proteins in *H. vastatrix* genome and protein sequences of *P. graminis* f.sp. *tritici* may not be solely because they belong to the same order, but also because they may have a similar gene content necessary for pathogenicity and may employ a similar pathogenic process. An advantageous point is that the genome sequence of *P. graminis* f.sp. *tritici* is publicly available [19] and can be used for further detailed comparative analysis of both pathogens.

The discovery of species-specific proteins and effectors is not trivial; while searches typically utilize criteria like being small and cysteine-rich to identify candidates, not all will function as effectors, and not all effectors will possess these hallmarks [54]. In the *H. vastatrix* genome,

12% of predicted proteins have no significant similarity (no matches) to protein sequences of fungi obtained from the NCBI and UNIPROT databases, and thus were considered unique to *H. vastatrix*.

Candidate secreted effector protein (CSEPs) of *H. vastatrix*

We analyzed predicted proteins of *H. vastatrix* genome, and identified 615 candidate secreted effector proteins (CSEPs). After redundancy analysis, this number dropped to 452 CSEPs (S3 Table). Previous studies predicted various numbers of CSEPs for *H. vastatrix*, as being 382 [36], 516 [48], from 659 to 775 [23] and 146 [55]. This variation in number of CSEPs is due to the type of data, software, and approaches used in each analysis. For instance, the range of predicted CSEPs, from 659 to 775, observed by [23], is indeed dependent on the prediction program used (659 CSEPs predicted using PProwler and 775 CSEPs using SignalP) and did not consider the analysis by TargetP (for subcellular localization of proteins) and TMHMM (to discriminate soluble and membrane proteins), which would decrease the predicted number. The smaller number of CSEPs, 382 [36], 516 [48], and 146 [55] was based on the analysis of expressed sequence tags (ESTs), thus, limiting identification of the full set of CSEPs. Using BLASTp, none of our 615 CSEPs were found in the study by [55]. However, we had 68 and 84 CSEPs found in studies by the [36] and [48].

Interestingly, when comparing *H. vastatrix* CSEPs with others rust fungi, 62.7% (385) of the *H. vastatrix* CSEPs are similar to *P. graminis* f. sp. *tritici* proteins, whereas only 17.7% (109) are similar to *M. larici-populina* proteins.

The classical secretory pathway [56] delivers the majority of effector proteins of the plant pathogenic fungi described thus far. Sequences of effector proteins frequently have no obvious similarity to each other or to other proteins in current databases [57], i.e., they are found to be species-specific effectors. Only 39 sequences from the predicted secretome had similarity to the public sequences of *H. vastatrix*. This low similarity can be explained by the small number of proteins of *H. vastatrix* available in the GenBank to date. In addition, 72 secreted proteins did not match to any protein sequence obtained from the databases. These proteins potentially represent novel genes and were considered unique to *H. vastatrix*, as they have not yet been described. The 111 (39+72) CSEPs of *H. vastatrix* were annotated using the Pfam software and no function has been assigned to them. Thus, the majority of EHv33 sequences reported in this study have no conserved domain described by the Pfam domain.

About 93% of 615 secreted proteins of *H. vastatrix* showed cysteine residues, the majority consisting of two, three or four residues in the sequence. According to Stergiopolous and de Wit [58], fungal effector proteins are usually rich in cysteine residues. These residues confer stability to proteins in extracellular space by means of formation of intra-molecular disulfide bridges that are characterized by their importance for structure and function. Changes in amino acid sequence may occur without changing the topology of proteins, except if it occurs in cysteine residues, making it ideal for the recognition and specificity [59]. Thus, the greater number of cysteine residues in an effector protein, the faster is its capacity for changes under any environmental selection pressure, thus favoring the evolution of new effectors and even new fungal physiological races. We identified large (>150 aa) proteins with up to 16 cysteine residues. While these proteins might not be considered typical effectors due to their size, the number of cysteine residues could make them candidate effectors with a high capacity for evolution [60].

Bioinformatics analysis can be used to identify motifs responsible for effector trafficking into host cells by predicting the presence and location of signal peptides [17]. Motifs of effectors such as [YWF]xC, CxxC and CxxxC have been found in conserved families of pathogen

effectors belonging to the Order Pucciniales [19]. In this study, these latter domains were found in 365 secreted proteins. While this indicates that these proteins may function as effectors, functional characterization assays are needed to decipher the molecular role of these proteins in pathogenesis and/or on pathogen fitness.

Temporal expression of *H. vastatrix* race XXXIII CSEPs in diverging host compatibility

Global expression of the predicted candidate effectors may indicate the effectors that are crucial for the disease establishment and progression, especially, the candidate effectors highly expressed during early infection [52]. We analyzed relative expression of 17 *H. vastatrix* race XXXIII CSEPs during a time course interaction with a susceptible and a resistant genotype, and our results show differing results depending on the cultivar used. These 17 candidates were chosen based upon possessing a predicted signal sequence and cysteine residues, and their uniqueness to *H. vastatrix*. More CSEPs were highly expressed in the compatible interaction than the incompatible interaction, which is expected as the effectors play a role in manipulating host defense and contribute to disease development. Six effectors showed expression during the incompatible interaction with Híbrido de Timor, the majority showing expression peaks at 24 hours after inoculation followed by a decrease. It is interesting to speculate that these effectors might play a role in defense and recognition by cognate R genes in the plant, however further experimentation is required to fully understand their functions.

A previous study from our group [35] showed that the interaction between a resistant genotype and race XXXIII starts very early, even before fungal penetration. Microscopic observations in the pathosystem Híbrido de Timor and *H. vastatrix* race XXXII revealed that cytological responses induced by the fungus can be observed in stomatal cells by 17 hours after infection (hai) and corresponded to hypersensitive-like cell response. Therefore, the fungus ceased its growth in the early stages of infection process, frequently in the penetration hyphae stage, indicating a pre-haustorial resistance. This resistance response was observed in 18% of infection sites at 17hai, reaching 65% and 93% at 24hai and 96hai, respectively [35]. Many effectors secreted by germinating spores are involved in signal transduction and establishment of the infection processes, in an attempt to suppress PAMP-triggered immunity (PTI), and then are recognized by plant receptors in resistant plants, triggering defense responses (effector-triggered immunity or ETI) [36] [61] [48]. This type of plant resistance, known as pre-haustorial resistance, was detected in coffee-rust interaction previously [62] [14]. From our predicted secretome, we found six candidate effectors significantly expressed only in the incompatible interaction and of these, five, EHv33_15, EHv33_1, EHv33_17, EHv33_13 and EHv33_8, were more abundant at 24 hours. This result suggests that they are pre-haustorial candidate effectors that might be attempting to suppress PTI, but additional experiments are required to confirm this further.

Diniz et al [63] found similar result to our prior cytological observations and suggested that this rapid plant defense response, which prevents haustorium formation, may be the basis for durable resistance to races of this fungal species. Therefore, those five candidate effectors identified can play important function in coffee resistance to the rust pathogen. On the other hand, if the pathogen overcomes the pre-haustorial resistance, the resistant plants will likely delay fungal growth in early stages of infection. This strategy is known as post-haustorial resistance [64] [10] and is usually shown after the formation of primary haustorium, which occurs about 48 hours after inoculation of *H. vastatrix* on coffee leaves. In our study, the post-haustorial expression pattern was detected for EHv33_12 effector candidate only, may be an ineffective attempt to suppress ETI.

H. vastatrix differentiates several specialized infection structures in the infection process. In susceptible genotypes, after the primary haustorium formation, the fungal growth continues, resulting in the formation of many intercellular hyphae and haustoria in mesophyll cells. Those changes in fungal development occur from 48 to 72 hours after inoculation [62] [48]. In this study, some effector candidates showed high expression levels in susceptible genotypes from 48 to 72 hours after inoculation. Thus, we can infer that the following genes are effector candidates possibly translocated into host cells via haustorium: EHv33_7, EHv33_2, EHv33_10, EHv33_4 and EHv33_16.

Overall, all these identified genes, pre- and post-haustorial candidate effectors, can be exploited to assist breeding programs, as biotechnological tools (effectoromics). Effectors are emerging as tools to accelerate and improve the identification and functional characterization of resistance genes that encode plant proteins able to recognize these effectors [18]. Thus, they can be used as markers to identify resistant genotypes in coffee breeding programs.

Ongoing challenges and future directions on *H. vastatrix* research

The genome sequence of the fungus *H. vastatrix* presented here is intended to help advance the knowledge of the pathogen and how it interacts with the plant successfully. With the more accurate number of predicted proteins and effectors, analyses for a more detailed understanding of their functional role and the set of main players in the interaction can now be performed. All the new findings can be useful in the development of molecular markers to distinguish races and, thus, provide a helpful tool to outperform difficulties found in the use of differential coffee cultivars.

Supporting information

S1 Fig. Assessment of the completeness of assembly/annotation with comparison to other genome assemblies from the order Puccinales by using the BUSCO bioinformatics platform. HvCat refers to *H. vastatrix* isolate (GCA_003057935.1); Mp: *Melampsora larici-populina* 98AG31 (GCF_000204055.1); Ml: *Melampsora lini* CH5 (JGI Genome portal: <https://genome.jgi.doe.gov/Melli1/Melli1.info.html>); Pg: *Puccinia graminis* f. sp. *tritici* CRL 75-36-700-3 (GCF_000149925.1); and Ps: *Puccinia striiformis* f. sp. *tritici* (GCA_001936605.2). (TIF)

S2 Fig. Functional clustering analysis of proteins identified in *H. vastatrix* genome, race XXXIII, which were found to be significantly similar (e-value < 10⁻⁵) to protein sequences obtained from NCBI and UNIPROT databases by using the Blast2GO bioinformatics platform. A) The y-axis consists of GO-terms described in the molecular function category for the hierarchical level #3. The x-axis consists of protein sequences found for each GO-term in this category. B) The y-axis consists of GO-terms described in the biological process category for the hierarchical level #3. The x-axis consists of protein sequences found for each GO-term in this category. C) The y-axis consists of GO-terms described in the cellular component category for the hierarchical level #3. The x-axis consists of protein sequences found for each GO-term in this category. D) The y-axis consists of GO-terms described in the cellular component category for the hierarchical levels #6 and #8. The x-axis consists of protein sequences found for each GO-term in this category. (TIF)

S3 Fig. Quantitative RT-PCR-based analysis of gene expression performed for two EHv33 genes that encode putative candidate secreted effector proteins of *H. vastatrix*, race XXXIII. The relative expression pattern of target genes was estimated in plant samples of the

hybrid of Timor and Caturra. Data were recorded at 12, 24, 48 and 72 hours after inoculation. The period of 12 hours after inoculation was used as reference sample. The expression level of target genes was normalized by using two endogenous genes of *H. vastatrix*, namely, β -tubulin and CytIII. **A)** EHv33_3: the gene expression increased over time, and the highest level was recorded at 72 hours after inoculation either in compatible or incompatible interaction. **B)** EHv33_14: there is no significant expression difference over time either in compatible or incompatible interaction.

(TIF)

S1 Table. Primers used for RT-PCR and qRT-PCR, with sequences designed using the Primer Express 2.0 software. a correspond to β -tubulin; b: cytochrome c oxidase subunit III; c: glyceraldehyde-3-phosphate dehydrogenase; EHv33: *Hemileia vastatrix* candidate effectors genes race XXXIII.

(XLSX)

S2 Table. Selection of 17 genes that encode potential effector proteins of *Hemileia vastatrix* race XXXIII identified in this study for gene expression analysis by RT-qPCR. *Data Base = NCBI/UNIPROT, RNA-Seq library by Lopes 2015 and Secreted proteins by Fernandez et al 2012 and Talhinhos et al 2014; ** No hits found = no hits found with any data base; ***Hits found = Hits found with NCBI/Uniprot data base with secreted protein *Hemileia vastatrix*.

(XLSX)

S3 Table. Putative Secreted Proteins (Singlets) *Hemileia vastatrix* race XXXIII.

(XLSX)

Acknowledgments

The authors would like to acknowledge the research funding and scholarships provided by INCT do Café, EMBRAPA Café, CNPq, CAPES, UFLA, FAPEMIG, UFV and UDEL.

Author Contributions

Conceptualization: Eveline Teixeira Caixeta, Laércio Zambolim, Eunize Maciel Zambolim, Mário Lúcio Vilela de Resende.

Data curation: Pedro Marcus Pereira Vidigal.

Formal analysis: Brenda Neves Porto, Eveline Teixeira Caixeta, Pedro Marcus Pereira Vidigal, Nicole Donofrio, Shawn W. Polson.

Funding acquisition: Laércio Zambolim, Mário Lúcio Vilela de Resende.

Investigation: Eveline Teixeira Caixeta.

Methodology: Pedro Marcus Pereira Vidigal, Nicole Donofrio, Thiago Andrade Maia.

Project administration: Eveline Teixeira Caixeta, Eunize Maciel Zambolim, Mário Lúcio Vilela de Resende.

Resources: Shawn W. Polson, Chuming Chen, Modupe Adetunji, Brewster Kingham.

Software: Chuming Chen, Modupe Adetunji.

Supervision: Eveline Teixeira Caixeta, Nicole Donofrio.

Validation: Sandra Marisa Mathioni, Ronaldo José Durigan Dalio.

Writing – original draft: Brenda Neves Porto, Eveline Teixeira Caixeta, Ronaldo José Durigan Dalio.

References

1. CONAB—Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Café, Primeiro Levantamento. 2016. Available from: http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_01_20_17_01_56_boletim_cafe_janeiro_2016.pdf
2. Gichuru EK, Ithiru JM, Silva MC, Pereira AP, Varzea VMP. Additional physiological races of coffee leaf rust (*Hemileia vastatrix*) identified in Kenya. *Trop Plant Pathol*. 2012; 37: 424–427. <https://doi.org/10.1590/S1982-56762012000600008>
3. Zambolim L. Current status and management of coffee leaf rust in Brazil. *Trop Plant Pathol*. 2016; 41: 1–8. <https://doi.org/10.1007/s40858-016-0065-9>
4. Talhinhas P, Batista D, Diniz I, Vieira A, Silva DN, Loureiro A, et al. The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. *Mol Plant Pathol*. 2017; 18: 1039–1051. <https://doi.org/10.1111/mpp.12512> PMID: 27885775
5. Capucho AS, Zambolim EM, Freitas RL, Haddad F, Caixeta ET, Zambolim L. Identification of race XXXIII of *Hemileia vastatrix* on *Coffea arabica* Catimor derivatives in Brazil. *Australas Plant Dis Notes*. 2012; 7: 189–191. <https://doi.org/10.1007/s13314-012-0081-7>
6. Avelino J, Cristancho M, Georgiou S, Imbach P, Aguilar L, Bornemann G, et al. The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food Secur*. 2015; 7: 303–321. <https://doi.org/10.1007/s12571-015-0446-9>
7. Van der Vossen H, Bertrand B, Charrier A. Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. *Euphytica*. 2015; 204: 243–256. <https://doi.org/10.1007/s10681-015-1398-z>
8. McCook S, Vandermeer J. The Big Rust and the Red Queen: Long-Term Perspectives on Coffee Rust Research. *Phytopathology*. 2015; 105: 1164–1173. <https://doi.org/10.1094/PHYTO-04-15-0085-RVW> PMID: 26371395
9. Ramiro DA, Escoute J, Petitot A-S, Nicole M, Maluf MP, Fernandez D. Biphasic haustorial differentiation of coffee rust (*Hemileia vastatrix* race II) associated with defence responses in resistant and susceptible coffee cultivars. *Plant Pathol*. 2009; 58: 944–955. <https://doi.org/10.1111/j.1365-3059.2009.02122.x>
10. Silva M do C, Várzea V, Guerra-Guimarães L, Azinheira HG, Fernandez D, Petitot A-S, et al. Coffee resistance to the main diseases: leaf rust and coffee berry disease. *Brazilian J Plant Physiol*. 2006; 18: 119–147. <https://doi.org/10.1590/S1677-04202006000100010>
11. Silva MC, Nicole M, Rijo L, Geiger JP, Rodrigues CJ Jr. Cytochemical Aspects of the Plant–Rust Fungus Interface during the Compatible Interaction *Coffea arabica* (cv. Caturra)–*Hemileia vastatrix* (race III). *Int J Plant Sci*. 1999; 160: 79–91. <https://doi.org/10.1086/314113>
12. Heath MC. A comparative study of non-host interactions with rust fungi. *Physiol Plant Pathol*. 1977; 10: 73–88. [https://doi.org/10.1016/0048-4059\(77\)90009-1](https://doi.org/10.1016/0048-4059(77)90009-1)
13. Heath M. Signalling between Pathogenic Rust Fungi and Resistant or Susceptible Host Plants. *Ann Bot*. 1997; 80: 713–720. <https://doi.org/10.1006/anbo.1997.0507>
14. Silva MC, Guerra-Guimarães L, Loureiro A, Nicole MR. Involvement of peroxidases in the coffee resistance to orange rust (*Hemileia vastatrix*). *Physiol Mol Plant Pathol*. 2008; 72: 29–38. <https://doi.org/10.1016/j.pmpp.2008.04.004>
15. Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S. Emerging Concepts in Effector Biology of Plant-Associated Organisms. *Mol Plant-Microbe Interact*. 2009; 22: 115–122. <https://doi.org/10.1094/MPMI-22-2-0115> PMID: 19132864
16. Cantu D, Segovia V, MacLean D, Bayles R, Chen X, et al., 2013 Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* 14: 270. <https://doi.org/10.1186/1471-2164-14-270> PMID: 23607900
17. Jones JDG, Dangl JL. The plant immune system. *Nature*. 2006; 444: 323–329. <https://doi.org/10.1038/nature05286> PMID: 17108957
18. Vleeshouwers VGAA Oliver RP. Effectors as Tools in Disease Resistance Breeding Against Biotrophic, Hemibiotrophic, and Necrotrophic Plant Pathogens. *Mol Plant-Microbe Interact*. 2014; 27: 196–206. <https://doi.org/10.1094/MPMI-10-13-0313-IA> PMID: 24405032
19. Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci*. 2011; 108: 9166–9171. <https://doi.org/10.1073/pnas.1019315108> PMID: 21536894

20. Upadhyaya NM, Garnica DP, Karaoglu H, Sperschneider J, Nemri A, Xu B, et al. Comparative genomics of Australian isolates of the wheat stem rust pathogen *Puccinia graminis f. sp. tritici* reveals extensive polymorphism in candidate effector genes. *Front Plant Sci. Frontiers*; 2015; 5: 759. <https://doi.org/10.3389/fpls.2014.00759> PMID: 25620970
21. Cantu D, Govindarajulu M, Kozik A, Wang M, Chen X, Kojima KK, et al. Next generation sequencing provides rapid access to the genome of *Puccinia striiformis f. sp. tritici*, the causal agent of wheat stripe rust. *PLoS One*. 2011; 6: 4–11. <https://doi.org/10.1371/journal.pone.0024230> PMID: 21909385
22. Nemri A, Saunders DGO, Anderson C, Upadhyaya NM, Win J, Lawrence GJ, et al. The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. *Front Plant Sci.* 2014; 5: 98. <https://doi.org/10.3389/fpls.2014.00098> PMID: 24715894
23. Cristancho MA, Botero-Rozo DO, Giraldo W, Tabima J, Riano-Pachon DM, Escobar C, et al. Annotation of a hybrid partial genome of the coffee rust (*Hemileia vastatrix*) contributes to the gene repertoire catalog of the Pucciniales. *Front Plant Sci.* 2014; 5: 1–11. <https://doi.org/10.3389/fpls.2014.00594> PMID: 25400655
24. Carvalho GMA, Carvalho CR, Barreto RW, Evans HC. Coffee rust genome measured using flow cytometry: Does size matter? *Plant Pathol.* 2014; 63: 1022–1026. <https://doi.org/10.1111/ppa.12175>
25. Tavares S, Ramos AP, Pires AS, Azinheira HG, Caldeirinha P, Link T, et al. Genome size analyses of Pucciniales reveal the largest fungal genomes. *Front Plant Sci. Frontiers*; 2014; 5: 422. <https://doi.org/10.3389/fpls.2014.00422> PMID: 25206357
26. Zambolim L, Chaves GM. Efeito de baixas temperaturas e do binomio temperatura-umidade relativa sobre a viabilidade dos uredosporos de *Hemileia vastatrix* Berk. et Br. e *Uromyces phaseoli typica* Arth. *Experientiae.* 1974; 17(7): 151–184. <https://doi.org/10.1111/j.1467.8527.2007.00388.4.x>
27. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience.* 2012; 1(1):18. <https://doi.org/10.1186/2047-217X-1-18> Erratum in: *Gigascience.* 2015;4:30. PMID: 23587118; PubMed Central PMCID: PMC3626529.
28. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods.* 2013; 10(6):563–9. <https://doi.org/10.1038/nmeth.2474> Epub 2013 May 5. PMID: 23644548.
29. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One.* 2012; 7(11):e47768. <https://doi.org/10.1371/journal.pone.0047768> Epub 2012 Nov 21. PMID: 23185243; PubMed Central PMCID: PMC3504050.
30. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using Basic Local Alignment with Successive Refinement (BLASR): Theory and Application, *BMC Bioinformatics* 2012, 13:238. <https://doi.org/10.1186/1471-2105-13-238> PMID: 22988817
31. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 2008; 18: 188–96. <https://doi.org/10.1101/gr.6743907> PMID: 18025269
32. Smit AFA, Hubley R, Green P. *RepeatMasker Open-4.0.* 2013–2015. Available from: <http://www.repeatmasker.org>
33. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA,* 2015; 6:11 <https://doi.org/10.1186/s13100-015-0041-9> PMID: 26045719
34. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011; 29(7):644–52. <https://doi.org/10.1038/nbt.1883> PMID: 21572440; PubMed Central PMCID: PMC3571712.
35. Freitas-Lopes RL. Análise citológica e perfil de expressão gênica de *Hemileia vastatrix* (raça XXXIII) na interação com o cafeeiro. Ph.D. thesis, Universidade Federal de Viçosa. 2015. Available from: <http://www.locus.ufv.br/handle/123456789/6947?show=full>
36. Fernandez D, Tisserant E, Talhinhos P, Azinheira H, Vieira A, Petitot AS, et al. 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. *Mol Plant Pathol.* 2012; 13: 17–37. <https://doi.org/10.1111/j.1364-3703.2011.00723.x> PMID: 21726390
37. Stanke M, Steinkamp R., Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research.* 2004; 32: W309–W312 <https://doi.org/10.1093/nar/gkh379> PMID: 15215400
38. Ter-Hovhannisyan V, Lomsadze A, Chernoff Y, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research,* 2008; 18 (12):1979–90 <https://doi.org/10.1101/gr.081612.108> PMID: 18757608

39. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31(19):3210–2. <https://doi.org/10.1093/bioinformatics/btv351> Epub 2015 Jun 9. PMID: 26059717.
40. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. 2007; 23(9):1061–7. Epub 2007 Mar 1. <https://doi.org/10.1093/bioinformatics/btm071> PMID: 17332020.
41. Altschul SF, Gish W, Pennsylvania T, Park U, Miller W, Myers EW, et al. Basic Local Alignment Search Tool 2 Department of Computer Science. *J Mol Biol*. 1990; 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
42. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005; 21: 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610> PMID: 16081474
43. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011; 8: 785–786. <https://doi.org/10.1038/nmeth.1701> PMID: 21959131
44. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J Mol Biol*. 2000; 300: 1005–1016. <https://doi.org/10.1006/jmbi.2000.3903> PMID: 10891285
45. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*. 2007; 35: W585–W587. <https://doi.org/10.1093/nar/gkm259> PMID: 17517783
46. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol*. 2001; 305: 567–580. <https://doi.org/10.1006/jmbi.2000.4315> PMID: 11152613
47. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 2015; 43: W78–W84. <https://doi.org/10.1093/nar/gkv487> PMID: 25964301
48. Talhinhos P, Azinheira HG, Vieira B, Loureiro A, Tavares S, Batista D, et al. 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*. 2014; 5: 88. <https://doi.org/10.3389/fpls.2014.00088> PMID: 24672531
49. Vieira A, Talhinhos P, Loureiro A, Duplessis S, Fernandez D, Silva M do C, et al. Validation of RT-qPCR reference genes for in planta expression studies in *Hemileia vastatrix*, the causal agent of coffee leaf rust. *Fungal Biol*. 2011; 115: 891–901. <https://doi.org/10.1016/j.funbio.2011.07.002> PMID: 21872186
50. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*. 2001; 29: 9
51. Bedell JA, Korf I, Gish W. MaskerAid: a performance enhancement to RepeatMasker. *Bioinformatics*. 2000; 16: 1040–1041. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11159316> PMID: 11159316
52. Spanu PD. The Genomics of Obligate (and Nonobligate) Biotrophs. *Annu Rev Phytopathol*. 2012; 50: 91–109. <https://doi.org/10.1146/annurev-phyto-081211-173024> PMID: 22559067
53. Stukenbrock EH, Croll D. The evolving fungal genome. *Fungal Biol Rev*. 2014; 28: 1–12. <https://doi.org/10.1016/j.fbr.2014.02.001>
54. Selin C, Kievit TR, Belmonte MF, Fernando WG. Elucidating the role of effectors in plant-fungal interactions: progress and challenges. *Frontiers in microbiology*. 2016; 7:600. <https://doi.org/10.3389/fmicb.2016.00600>
55. Maia T, Badel JL, Marin-Ramirez G, Rocha C de M, Fernandes MB, da Silva JCF, et al. The *Hemileia vastatrix* effector HvEC-016 suppresses bacterial blight symptoms in coffee genotypes with the *S_{H1}* rust resistance gene. *New Phytol*. 2017; 213: 1315–1329. <https://doi.org/10.1111/nph.14334> PMID: 27918080
56. De Wit PJGM, Mehrabi R, Van Den Burg HA, Stergiopoulos I. Fungal effector proteins: Past, present and future: Review. *Mol Plant Pathol*. 2009; 10: 735–747. <https://doi.org/10.1111/j.1364-3703.2009.00591.x> PMID: 19849781
57. Ellis JG, Rafiqi M, Gan P, Chakrabarti A, Dodds PN. Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. *Curr Opin Plant Biol*. 2009; 12: 399–405. <https://doi.org/10.1016/j.pbi.2009.05.004> PMID: 19540152
58. Stergiopoulos I, de Wit PJGM. Fungal Effector Proteins. *Annu Rev Phytopathol*. 2009; 47: 233–263. <https://doi.org/10.1146/annurev.phyto.112408.132637> PMID: 19400631
59. Povolotskaya IS, Kondrashov FA. Sequence space and the ongoing expansion of the protein universe. *Nature*. 2010; 465: 922–926. <https://doi.org/10.1038/nature09105> PMID: 20485343

60. Göhre V, Robatzek S. Breaking the Barriers: Microbial Effector Molecules Subvert Plant Immunity. *Annu Rev Phytopathol.* 2008; 46: 189–215. <https://doi.org/10.1146/annurev.phyto.46.120407.110050> PMID: 18422429
61. Staats CC, Kmetzsch L, Schrank A, Vainstein MH. Fungal zinc metabolism and its connections to virulence. *Front Cell Infect Microbiol.* 2013; 3: 65. <https://doi.org/10.3389/fcimb.2013.00065> PMID: 24133658
62. Ganesh D, Petitot AS, Silva MC, Alary R, Lecouls AC, Fernandez D. 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.* 2006; 170: 1045–1051. <https://doi.org/10.1016/j.plantsci.2005.12.009>
63. Diniz I, Talhinhas P, Azinheira HG, Várzea V, Medeira C, Maia I, et al. 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.* 2012; 133: 141–157. <https://doi.org/10.1007/s10658-011-9925-9>
64. Silva MC, Nicole M, Guerra-Guimarães L, Rodrigues CJ. Hypersensitive cell death and post-haustorial defence responses arrest the orange rust (*Hemileia vastatrix*) growth in resistant coffee leaves. *Physiol Mol Plant Pathol.* Academic Press; 2002; 60: 169–183.