DE GRUYTER Journal of Integrative Bioinformatics 2026; 20250013

Kimberly Rojas-Rojas, Paz Vieto-Cárdenas, Bernal Azofeifa-Bolaños, Carolina Sancho-Blanco and Stefany Solano-González*

In silico comparative genomics and proteomics to identify potential candidates to monitor coffee plant-fungi pathogenesis

https://doi.org/10.1515/jib-2025-0013
Received February 27, 2025; accepted April 3, 2026; published online May 15, 2026

Abstract: Fungal pathogens pose a major threat to coffee production, yet the molecular mechanisms underlying these infections remain poorly understood. Omics data from fungal pathogens affecting coffee plants offer valuable but largely unexplored insights into host-pathogen interactions. In this study, we applied a computational approach to analyze publicly available genomic and proteomic data from three coffee cultivars and three fungal pathogens, including *Hemileia vastatrix* (coffee leaf rust fungus) and *Colletotrichum higginsianum* (anthracnose fungus). We identified candidate genes involved in the plant's defense response and potential fungal effector proteins associated with pathogenesis, providing novel targets for disease monitoring and management. Fungal analysis revealed in *H. vastatrix* a total of 2,058 potential effectors and in *C. higginsianum* 4,475, these effectors were categorized as cytoplasmic or apoplastic. Orthofinder analysis highlighted four informative groups, clustered based on function, role in infection processes, and total count of positive parameters allowing the identification of kinases, deacetylases, chitin-binding, recognition, GTP-binding, Ras, and Rho family proteins. Our results provide the first set of computationally called proteomes for the analyzed species, contextualized to coffee-fungi pathogenesis and provide a set of genes and proteins to further validate experimentally.

Keywords: plant bioinformatics; effector proteins; R genes; secreted proteins; virulence; *H. vastatrix*

1 Introduction

Plants' molecular resistance to pathogens is intricately linked to the unique genetic information of each species or variety [1]. Understanding these molecular interactions requires the systematic identification, mapping, and functional annotation of genes and proteins involved in defense mechanisms. In recent years, significant efforts have been made to sequence the genomes of agronomically important plants and fungal pathogens. However, genomic resources remain limited for rust (*Hemileia vastatrix*, coffee leaf rust fungus) and anthracnose-causing (*Colletotrichum higginsianum*, anthracnose fungus) pathogens, particularly in several Latin American

---

***Corresponding author: Stefany Solano-González**, Laboratorio de Bioinformática Aplicada, Escuela de Ciencias Biológicas, Universidad Nacional, Heredia, Costa Rica, E-mail: stefany.solano.gonzalez@una.cr. https://orcid.org/0000-0002-1167-2174
Kimberly Rojas-Rojas and Paz Vieto-Cárdenas, Laboratorio de Bioinformática Aplicada, Escuela de Ciencias Biológica s, Universidad Nacional, Heredia, Costa Rica
Bernal Azofeifa-Bolaños, Laboratorio de Biotecnología de Plantas, Escuela de Ciencias Biológicas, Universidad Nacional, Heredia, Costa Rica
Carolina Sancho-Blanco, Laboratorio de Análisis Genómico, Escuela de Ciencias Biológicas, Universidad Nacional, Heredia, Costa Rica

coffee-producing countries such as Brazil, Colombia, Costa Rica, and Guatemala [2, 3]. This scarcity of comprehensive genomic data hinders a deeper understanding of plant-fungal interactions and constrains the development of effective molecular-based disease management strategies. To date *C. higginsianum,* has a publicly available genome.

Coffee's major fungal pathogens include *Mycena citricolor* [4–6], *Colletotrichum kahawae* [7], *C. higginsianum* [8, 9], and the well-known *H. vastatrix* [10, 11]. The plant immune system responds through its resistance genes, known as R genes, which play a crucial role in the defense mechanisms and have been used to identify and develop disease-resistant cultivars [12]. In coffee, this has been particularly implemented with the SH genes [11, 13]. Although their molecular and functional characteristics have not been fully explored. While certain R genes have been identified during infection with *C. kahawae* [14], there is limited information available regarding their response to *M. citricolor* attack [15]; nonetheless, additional genes related to plant defense mechanisms during *H. vastatrix* infection in coffee has been discovered [16].

During infection processes, the pathogenic fungi secrete specific proteins, termed secreted proteins [17], that interact with the cascade of specific molecules from the coffee plant as a response to the infection process [18, 19]. The success of the fungal invasion and its reproduction depends mainly on molecular mechanisms that counteract the recognition of defense molecules or deactivate plant responses; some of these mechanisms involve effector proteins. Genes that recognize these effector proteins have been identified in coffee, particularly for *H. vastatrix* [20]. In the case of *C. kahawae*, only one effector protein has been reported [21]. However, the unavailability of its genome creates a general deficiency in comparative genomic analysis for the species of interest in this study. Fortunately, its closely related species, *C. higginsianum,* has a publicly available genome, making it a suitable starting point for analysis.

Genomic data on coffee plant fungal pathogens in Latin America are limited; however, some studies have been conducted in Brazil, which remains the leading producer and researcher in coffee pathogenesis, facilitating the provision of in silico data for identifying molecular markers and single-nucleotide mutations [12, 22, 23]. This approach is crucial to monitoring and promptly identifying potential fungal infections, which aids in protecting cultivars and tailors their agronomic management to specific needs. Our study identified potential target elements related to or involved with the pathogenic coffee-fungi interaction through an *in-silico* analysis to provide specific genes or proteins for its future experimental validation, facilitating variety selection and genetic improvement studies.

2 Workflow

From the National Centre for Biotechnology Information (NCBI) RefSeq Database [24] we retrieved genome sequences from three coffee species and three of its fungal pathogens (Supplementary Tables S1 and S2, respectively) (Figure 1A). To determine potential genome improvements through re-assembly, we utilized MeDuSa v1 [25]. For coffee genomes, *Solanum lycopersicum* SL3.0 genome (RefSeq ID accession no. **GCF_000188115.4**) was used as a reference, given the phylogenetic proximity between *Rubiaceae* (coffee) and *Solanaceae* (tomato) families, which share a common ancestor [26]. For fungal genomes, we employed as references *Ustilago maydis* (RefSeq ID accession no. **GCF_000328475.2**) and *Puccinia triticina* (RefSeq accession no. **ID GCF_026914185.1**) genomes. The former is a model organism of plant-microorganism interactions, and its functional and structural annotation is very well characterized [27], and the latter is another rust belonging to the *Pucciniales* order and a well-studied microorganism as one of the main pathogens of wheat worldwide [28]. Following this, we evaluated the quality of the assemblies with the Quality Assessment Tool for Genome Assemblies, QUAST v.4.6.0 [29] (Figure 1.B). Once all genomes were retrieved or re-assembled, we called for the corresponding proteins in those whose proteomes were not available, corresponding to *M. citricolor*, *H. vastatrix*, *Coffea arabica* var. *caturra* (now onwards defined as *C. caturra*), *C. eugenioides,* and *Coffea canephora* using AUGUSTUS v3.4.0 [30]. For this protein-call process, we used *U. maydis* and *S. lycopersicum* as references for fungi and plants, respectively.

To search for effector proteins in fungal proteomes, we utilize Predector v1.2.6 [31]. and selected 10 different parameters related to pathogenicity, virulence, signal peptides, and effector proteins

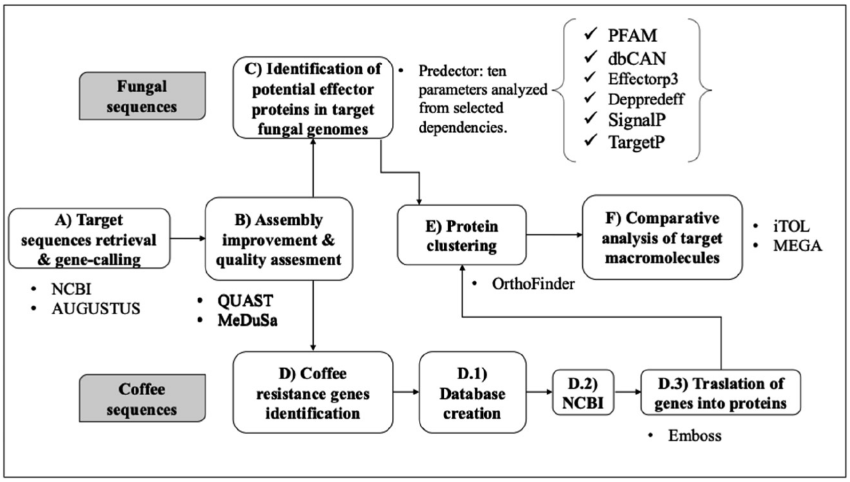


Figure 1: Schematic view for analysis of coffee and fungal genome sequences to identify molecules related to the pathogenic process. A) Retrieval of sequences and gene-calling using Augustus v3.4.0. 2. B) Assembly improvement and quality assessment using MeDuSa v1.6 and Quast v.4.6.0, respectively. C) Identification of potential effector proteins in objective fungal genomes using 10 different parameters from six tools within Predector v1.2.6 software. D). Identification of resistance genes in coffee genomes. D.1) establishment of a resistance gene database for coffee plants. D.2) Retrieval of genes validated in the literature from PubMed NCBI. D.3) conversion of genes into proteins using EMBOSS. E) Protein clustering using OrthoFinder v.2.5.4 output to identify potential orthologues to the coffee resistance database and relevant proteins to fungal infection processes. F) Comparative analysis of targeted sequences to determine informative regions with MEGA v.11 and iTOL [32] (https://itol.embl.de).

(Supplementary Table S3); all considering the pathogen-host-interaction (PHI-base) and the Carbohydrate-Active Enzyme Annotation (dbCAN) databases [32–34] (Figure 1C).

On the other hand, coffee genomic and proteomic data were employed to identify resistance genes/proteins. To achieve this, we created a database of 10 resistance genes for which literature reported experimental validation (Supplementary Table S5); and translated these genes into proteins using EMBOSS/6.6.0 for further comparison (http://emboss.sourceforge.net/) [35] (Figure 1.D.1–D.3).

For both approaches, coffee and fungal proteomes were clustered into informative groups of proteins with OrthoFinder v.2.5.4 [36], to identify potential orthologues to the coffee resistance database (Figure 1E) and proteins associated with fungal infection processes [37]. The orthologous groups (OG) were selected based on the presence of the most virulence pathogenicity parameters and then through a secondary validation process involving a manual review of their annotations in the NCBI and UniProt (The UniProt Consortium, 2023).

In the analysis of Coffee orthofinder outputs, only orthogroups (OGs) containing genes from the resistance genes database were selected. Subsequently, these selections underwent manual functional annotation using NCBI and Uniprot.

For both approaches, the trees selected for coffee and fungal final orthogroups were visualized with iToL (https://itol.embl.de), and the corresponding alignments were visualized in MEGA [38] (Figure 1F) to determine potential conserved and variable regions, or motifs across protein sequences. The schematic view of the process is detailed in Figure 1.

3 Application

3.1 Genome re-assembly and quality assessment

Initially, *C. canephora, C. caturra* and *C. eugenioides* genomes' consisted of 12, 2,833 and 3,530 contigs, respectively (Supplementary Table S1). After MeDuSa [25] and Quast [29] processing, the contig counts were 12, 2,829 and 1,163, respectively. Overall, only *C. eugenioides* exhibited a relevant reduction in gaps and contigs (data not shown), whereas the other two only showed marginal or no improvement, compared to the original values. A similar pattern was observed for *H. vastatrix*, *M. citricolor,* and *C. higginsianum* fungal genomes, which reported 116,750, 12,462, and 25 contigs, respectively.

3.2 Identification and in silico characterization of potential fungal effector proteins

Predector pipeline gave no prediction for *M. citricolor* proteome; therefore, it was not included in the analysis. On the other hand, a total of 18,491 called proteins were analyzed, 3,842 from *H. vastatrix* and 14,649 from *C. higginsianum*. Our Predector annotation within the PHI-base, resulted in 0.60 % of proteins annotated as lethal effectors and 2.26 % as virulence effectors for *H. vastatrix*. The corresponding values in *C. higginsianum* conveyed 1.04 % and 7.30 %, respectively.

When considering the total count of effector proteins and their most likely post-translational category (apoplastic or cytoplasmic) determined by EffectorP, which is a Predector dependency, functioning like a classifier of effector proteins, trained on apoplastic and cytoplasmic effectors, implementing machine learning algorithms [39]. These allowed us to identify in *H. vastatrix* a total of 2058 potential effectors, accounting for 53.6 % of its complete proteome. From these, 1984 was cytoplasmic and 74 apoplastic effectors. Regarding *C. higginsianum,* a total of 4,475 potential effectors were identified, accounting for 30.5 % of its complete proteome: assigning 3,917 as cytoplasmic and 558 as apoplastic effectors. Interestingly, 52 proteins for *H. vastatrix* and 189 for *C. higginsianum* were assigned to both cytoplasmic and apoplastic locations.

On the other hand, complete proteomes for the fungal organisms were processed using Orthofinder. We focused our analysis exclusively on OG with a positive assignment in either PHI-base or EffectorP. As a result, we examined 40 OG, classified based on 10 parameters, including function, role in infection processes, and total count of positive parameters (Supplementary Table S3). Following this second selection, OG00000054, OG00000065, OG00000068, and OG00000226, were chosen and their OrthoFinder phylogenetic trees were visualized (Figure 2). In OG0000054, most proteins were classified as kinases (Figure 2A, Supplementary Table S4). In OG0000065, most proteins were classified as deacetylases, chitin-binding proteins, and recognition proteins (Figure 2B, Supplementary Table S4). OG0000068 presented proteins related to the Ras family (Figure 2D, Supplementary Table S4), whereas OG0000226, primarily corresponded to GTP-binding proteins, Ras family proteins, and Rho proteins (Figure 2C, Supplementary Table S4). Ras and Rho proteins are a type of small GTPases, acting as molecular switches by the activation/inactivation of their GTP [40]. This activation interacts with specific effectors and triggers cell responses, such as those involved in pathogenicity, by the formation of virulence structures and the production of ROS species [41, 42].

Finally, alignments were conducted for the chosen OGs. Also, conserved motifs were absent in most of the proteins within these OGs, except for OG0000068, where the DTAG (Q/T) (E/D) motif was identified (Figure 3).

3.3 Identification of coffee proteins associated with resistance and pathogen susceptibility

A database of 10 validated genes (Supplementary Table S5) associated with plant responses to phytopathogens was created through a literature review. It was used in addition to the coffee proteomes to be processed with OrthoFinder. A total of 40,109 OGs were identified. However, only four of these (OG0004334, OG0005261, OG0006253 and OG0010612) were related to the resistance genes database, comprising 17 genes with reported functional annotations, while four lacked annotations. As each OG comprised an unannotated protein, phylogenetic relationships were determined to understand its potential function (Figure 4).

4 Discussion

4.1 Genome re-assembly and quality assessment

Incontiguous genomes dramatically reduce the efficacy of algorithms employed in genomic comparisons; therefore, reassembly improvements yielding contiguous genomes are significant in facilitating the identification of

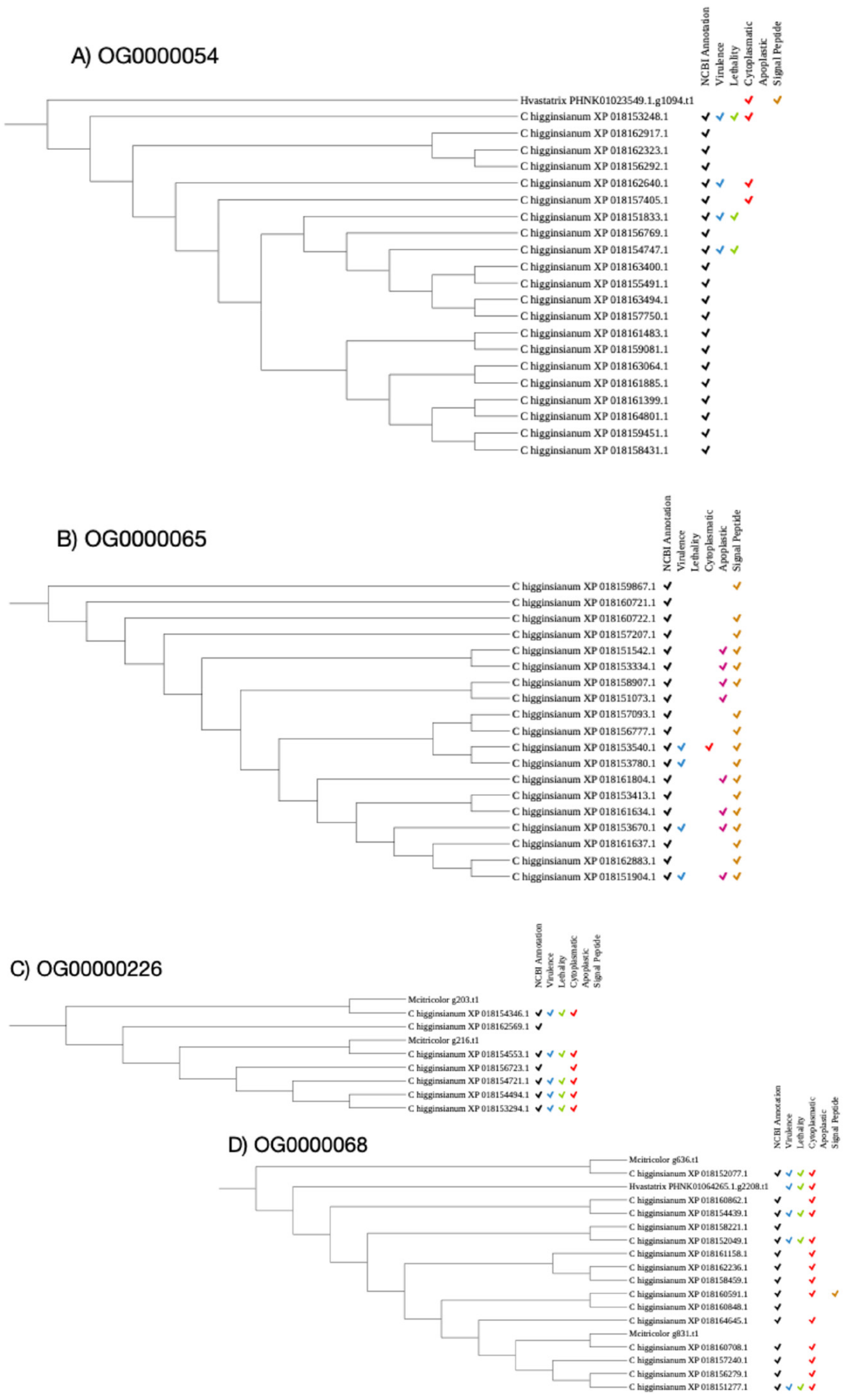
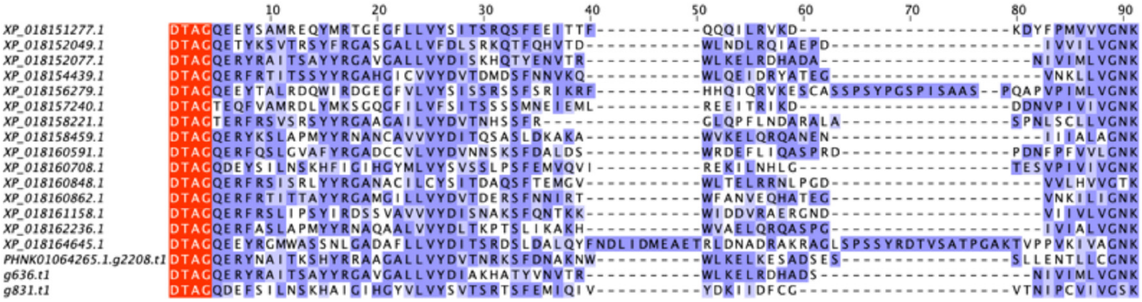


Figure 2: Panoramic display of OrthoFinder gene trees for fungal orthologue groups. A) OG54, B) OG65, C) OG226, and D) OG68 gene trees with a focus on protein members distinguished by their functional significance (virulence or lethality), subcellular localization (cytoplasmic or apoplastic), the presence or absence of a signal peptide, and pertinent annotations sourced from NCBI. Additional data regarding the reported function for each protein entry is available at the Supplementary Table S4.

pertinent genetic elements [43, 44]. On the other hand, the notable scarcity of well-assembled public genomes poses challenges for key genes *in silico* identification, which is usually used in molecular approaches aimed at molecular breeding programs and pathogen detection mechanisms. Our results highlight that coffee genomes exhibited superior original assembly compared with fungal genomes. This is because the original assembly of coffee genomes displayed the highest N50 and lowest L50 values, which were consistently upper 38 mpb, in contrast to fungal genomes, where the highest N50 value was approximately 10,000 bp (Supplementary Table S1).



```
                                10        20        30        40        50        60        70        80        90
XP_018151277.1          DTAGQEEYSAMREQYMRTGEGFLLVYSITSRQSFEEITTF---------QQQILRVKD---------------------KDYFPMVVVGNK
XP_018152049.1          DTAGQETYKSVTRSYFRGASGALLVFDLSRKQTFQHVTD----------WLNDLRQIAEPD---------------------IVVILVGNK
XP_018152077.1          DTAGQERYRAITSAYYRGAVGALLVYDISKHQTYENVTR----------WLKELRDHADA---------------------NIVIMLVGNK
XP_018154439.1          DTAGQERFRTITSSYYRGAHGICVVYDVTDMDSFNNVKQ----------WLQEIDRYATEG---------------------VNKLLVGNK
XP_018156279.1          DTAGQEEYTALRDQWIRDGEGFVLVYSISSRSSFSRIKRF---------HHQIQRVKESCASSPSYPGSPISAAS---PQAPVPIMLVGNK
XP_018157240.1          DTAGTEQFVAMRDLYMKSGQGFILVFSITSSSSMNEIEML---------REEITRIKD---------------------DDNVPIVIVGNK
XP_018158221.1          DTAGTERFRSVSRSYYRGAAGAILVYDVTNHSSFR--------------GLQPFLNDARALA-----------------SPNLSCLLVGNK
XP_018158459.1          DTAGQERYKSLAPMYYRNANCAVVVYDITQSASLDKAKA----------WVKELQRQANEN---------------------IIIALAGNK
XP_018160591.1          DTAGQERFQSLGVAFYRGADCCVLVYDVNNSKSFDALDS----------WRDEFLIQASPRD-----------------PDNFPFVVLGNK
XP_018160708.1          DTAGQDEYSILNSKHFIGIHGYMLVYSVSSLPSFEMVQVI---------REKILNHLG---------------------TESVPIVIVGNK
XP_018160848.1          DTAGQERFRSISRLYYRGANACILCYSITDAQSFTEMGV----------WLTELRRNLPGD---------------------VVLHVVGTK
XP_018160862.1          DTAGQERFRTITTAYYRGAMGILLVYDVTDERSFNNIRT----------WFANVEQHATEG---------------------VNKILIGNK
XP_018161158.1          DTAGQERFRSLIPSYIRDSSVAVVVYDISNAKSFQNTKK----------WIDDVRAERGND---------------------VIIVLVGNK
XP_018162236.1          DTAGQERFASLAPMYYRNAQAALVVYDLTKPTSLIKAKH----------WVAELQRQASPG---------------------IVIALVGNK
XP_018164645.1          DTAGQEEYRGMWASSNLGADAFLLVYDITSRDSLDALQYFNDLIDMEAETRLDNADRAKRAGLSPSSYRDTVSATPGAKTVPPVKIVAGNK
PHNK01064265.1.g2208.t1 DTAGQERYNAITKSHYRRAAGALLVYDVTNRKSFDNAKNW---------WLKELKESADSES-----------------SLLENTLLCGNK
g636.t1                 DTAGQERYRAITSAYYRGAVGALLVYDIAKHATYVNVTR----------WLKELRDHADS---------------------NIVIMLVGNK
g831.t1                 DTAGQDEFSILNSKHAIGIHGYVLVYSVTSRTSFEMIQIV---------YDKIIDFCG---------------------VTNIPCVIVGSK
```

Figure 3: Amino acid sequence section for orthogroup OG0000068 protein members. In red color the conserved DTAG-Q/T motif is observed throughout the sequence.

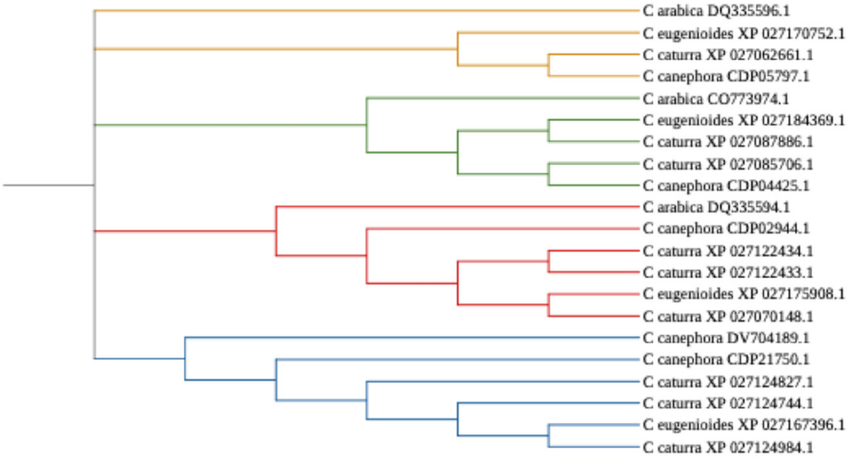


Figure 4: Maximum likelihood analysis of orthogroups resulting from the analysis of coffee proteomes and genes associated with plant responses to pathogens. The tree represents orthogroups clustered with key fungal pathogenic proteins, color-coded according to their functions, yellow represents non-specific disease resistance proteins, green is for proteins related to stress responses and involved in antagonistic functions of salicylic acid (SA) and jasmonic acid, red denotes proteins related to the CaPR1b gene, and blue signifies proteins related to the CaLOX gene. The name of each protein is shown beneath each organism.

4.2 *In silico* identification and characterization of potential fungal effector proteins

Our classifications evidenced higher lethal and virulence protein counts in *C. higginsianum* compared to *H. vastatrix*, aligning with the reported hemibiotrophic and obligate biotrophic nature of both fungi, respectively [45, 46].

In silico effector protein (EP) identification exhibits differences in amino acid sequences that do not resemble the model used by the computational gene prediction program (s), reducing homology [47, 48]. However, through genomic comparisons between fungal species, signal peptides (short amino acid sequences that guide protein localization) and other features used for EP identification [49] have revealed significant instances of uncommon homology or orthology relationships among proteins from distantly related pathogenic species, suggesting potential lateral gene transfer or significant virulence function [48]. This integrative approach demonstrated that *C. higginsianum* has an overall higher count for effector proteins (regardless of the classification) than *H. vastatrix*. This disparity could arise from the chromosome-level assembly of the former in contrast to the contig assembly of the latter, or possibly from some yet undiscovered biological interaction.

Experimental validation is a must after computational predictions; for example, in this study a total of 2058 potential effectors were annotated, including orthologues to HvEC_050, HvEC_051, HvEC_054, HvEC_072 (data not shown). Nonetheless only 516 secreted proteins in *H. vastatrix* [45, 50, 51] have been reported experimentally. These Hv genes have been described by Maia et al. [20] as promising effector candidates with potential essential roles in *H. vastatrix*-coffee interactions, influencing plant resistance mechanisms. These reported experimental findings suggest that there may be an interplay between the pathogen and the host during the early stages of infection, specifically during urediniospore germination. This implies the existence of pre-haustorial resistance, which is attributed to the abundance of cysteine residues in the amino acid sequences encoded by most HvECs genes. Cysteine-rich proteins are often implicated in recognition and specificity mechanisms, making them strong candidates for involvement in pre-haustorial resistance during early host-pathogen interactions. Recognition typically serves as a reliable indicator of pre-haustorial resistance [20]. The establishment and assessment of the Protein Interaction Network between *H. vastatrix* and coffee have, for the first time, revealed a set of interactions with significant potential involvement in the plant resistance process.

Selected orthogroups OG00000054, OG00000065, OG00000068, and OG00000226, were categorized based on their function, role in infection processes, and the total count of positive parameters (Supplementary Table S3). Of the 22 proteins within OG00000054, the majority belonged to *C. higginsianum* [21], classified as kinases, with a unique protein from *H. vastatrix* lacking functional annotation at the NCBI (Figure 2A). Interestingly, proteins XP_ 018151833.1 and XP_018154747.1 from *C. higginsianum* were annotated as virulent and lethal effectors but showed no cellular location. Proteins XP_ 018162,640.1 and XP_ 018157405.1 presented a cytoplasmic location; however, only the former was classified as a virulence effector. On the other hand, protein XP_018153248.1 was the only *C. higginsianum* protein effector assigned as both virulent and lethal with a cytoplasmic location. For *H. vastatrix* annotated protein (PHNK01023549.1.g1094.t1) we could only classify its cellular location as cytoplasmic, notably from the OG was the only one with a signal peptide. Proteins from this OG might be a good target to experimentally identify potential or significant differential expression during the infection process with compatible and incompatible interactions with coffee, understanding its effect as lethal or virulent.

The 19 proteins from OG0000065 corresponded solely to *C. higginsianum* and were mainly classified as deacetylases, chitin-binding proteins, and recognition proteins (Figure 2B).

The 19 proteins from OG0000065 were exclusive to *C. higginsianum* and were mainly classified as deacetylases, chitin-binding proteins, or recognition proteins (Figure 2B). In contrast to OG00000054, most OG0000065 proteins contained a signal peptide, except for three (XP_018160721.1, XP_018160722.1 and XP_018151073.1) within this same group. In addition, none were annotated as lethal effectors, and only four (XP_018153540.1, XP_018153780.1, XP_018153670.1, and XP_018151904.1) were classified as virulent effectors.

It is noteworthy that proteins from this OG are targeted to the apoplast, indicating that their mechanisms operate between the plant cell wall and the plasma membrane [52, 53] with XP_ 018153540.1 being the only exception as it is directed to the cytoplasm; this might indicate a potential dual function or a misannotation.

OG0000068 included members from the three fungal species studied and consisted primarily of proteins associated with the Ras family (Figure 2C) targeting the cytoplasm; these align with cytoplasmic effectors having a higher proportion of positively charged amino acids. In contrast, apoplastic effectors are enriched for cysteine residues [39]. Interestingly, the identified conserved DTAG (Q/T) (E) motif (Figure 3), associated with protein kinases [54], might have a relationship between OG0000068 proteins interaction with the coffee molecular machinery response. On the other hand, from the 18 proteins within the group, only XP_018152077.1, PHNK01064265.1, XP_018154,439.1, XP_0181158221.1, and XP_018151,277.1 were assigned as both lethal and virulent effectors; with *M. citricolor* protein lacking functional annotation from NCBI. This OG clustered with a Rab GTPase family 11 from *C. higginsianum*. This is interesting as Rab GTPase family is associated with vesicle transportation from endosomes to the plasma membrane [55] and markers of organelle identity [56]; however, from the 18 proteins, only XP_018160591.1 reports a signal peptide. Nonetheless, proteins within this group might be useful to monitor potential combinatorial behavior regulating membrane traffic [57].

OG0000226 comprises nine proteins, the smallest group from the analysis, predominantly composed of GTP-binding, Ras family, and Rho proteins (Figure 2C). From this group, the majority presented both virulent and lethal effector annotations, targeting the cytoplasm, except for four proteins (Figure 2D). As observed for other OGs, *M. citricolor* proteins lack functional annotation. However, g.203.t1 presented a guanylate cyclase domain and g.216.t1 was related to Sen54 protein, both related to splicing and signaling processes [58, 59]. Proteins from this OG might offer insights into novel infection roles and serve as potential candidates for early resistance detection, given their potential involvement in signaling pathways often activated in the early stages of infection.

4.3 *In silico* identification of coffee proteins associated with resistance and pathogen susceptibility

OG0004334, OG0005261, OG0006253, and OG0010612 were concatenated, and the corresponding tree was color-coded according to their functionality as follows: yellow comprised proteins related to non-specific disease resistance of race 1 (NDR1), which are involved in the activation of the family of resistance proteins with leucine-rich repeats (CC-NB-LRR). Green OG consists of WRKY proteins related to stress response and plays antagonistic roles with salicylic acid (SA) and jasmonic acid (JA)/ethylene (ET). Red OG comprises proteins related to the CaPR1b genes, which have been used as markers for salicylic acid-mediated disease resistance, and the blue OG is associated with the CaLOX gene, which encodes linoleate 13S-lipoxygenase and is correlated with pest resistance. Subsequently, WRKY motifs were identified in some proteins from the OG (Supplementary Figure S1). This is particularly associated with the kinase signaling network [60], which aligns with some of the proteins identified for the fungal pathogens.

5 Conclusions

The lack of high-quality genomes in public repositories has hindered the development of molecular markers for the early identification of fungal pathogens in coffee cultivation. Despite being limited to available and incomplete genomes, our work provides a list of key 21 resistance genes and 68 potential effector proteins involved in pathogenic fungi-coffee interaction by an *in silico* computational study that can support experimental validation to monitor early stages fungal infections through plant-pathogen interaction. This can contribute to the selection and genetic improvement of coffee varieties, particularly benefiting major producers like Costa Rica, in aligning production with global sustainability standards. Moreover, future efforts will greatly benefit from the integration of artificial intelligence to accelerate effector discovery and enhance prediction accuracy. Machine learning models trained on known fungal effectors, combined with experimental data and environmental parameters, could uncover novel, biologically relevant patterns in large protein datasets [61], potentially revealing mechanisms tailored to local conditions such as soil characteristics and temperature.

Acknowledgments: We want to thank the Centro Nacional de Computación Avanzada from CeNAT, CONARE, located in Costa Rica, for providing access to the supercomputing cluster Kabré.

Research ethics: Not applicable.

Informed consent: Not applicable.

Author contributions: Rojas-Rojas K and Vieto-Cárdenas P: investigation, methodology, data curation, formal analysis, drawing figures, preparing tables and writing original draft. Sancho-Blanco C and Azofeifa-Bolaños B: writing original draft and reviewing and editing. Solano-González S: conceptualization, methodology, investigation, validation, resources, visualization, writing – original draft, project administration, reviewing and editing, supervision and advisory role.

Use of Large Language Models, AI and Machine Learning Tools: None declared.

Conflict of interest: The authors state no conflict of interest.

Research funding: This work was supported by the Universidad Nacional at Grant SIA-0717-19 and the Fondo Especial para la Educación Superior (FEES).

Data availability: Data is available within the manuscript and/or supplementary material.

References

1. Ernesto G-P, Edmundo L-G. Genes de resistencia a enfermedades en plantas. Rev Mex Fitopatol 2004;22:414—22.
2. 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—21.
3. 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—51.
4. Lyu X, Shen C, Fu Y, Xie J, Jiang D, Li G, et al. A small secreted virulence-related protein is essential for the necrotrophic interactions of *Sclerotinia sclerotiorum* with its host plants. PLoS Pathog 2016;12:1—28.
5. Yu Y, Xiao J, Zhu W, Yang Y, Mei J, Bi C, et al. Ss-Rhs1, a secretory Rhs repeat-containing protein, is required for the virulence of *Sclerotinia sclerotiorum*. Mol Plant Pathol 2017;18:1052—61.
6. Zhu W, Wei W, Fu Y, Cheng J, Xie J, Li G, et al. A secretory protein of necrotrophic fungus Sclerotinia sclerotiorum that suppresses host resistance. PLoS One 2013;8:e53901.
7. Diniz I, Figueiredo A, Loureiro A, Batista D, Azinheira H, Varzea V, et al. A first insight into the involvement of phytohormones pathways in coffee resistance and susceptibility to *Colletotrichum kahawae*. PLoS One 2017;12:1—20.
8. Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, van Themaat EVL, van der Does HC, et al. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. PLoS Pathog 2012;8:e1002643.
9. Takahara H, Hacquard S, Kombrink A, Hughes HB, Halder V, Robin GP, et al. *Colletotrichum higginsianum* extracellular LysM proteins play dual roles in appressorial function and suppression of chitin-triggered plant immunity. New Phytol 2016;211:1323—37.
10. Alzate NO. Importancia del gen sh3 y su relación con la resistencia a la enfermedad roya del cafeto *Hemileia vastatrix* bajo la modalidad de compilación bibliográfica 2015;88. Available from: http://repository.unad.edu.co/bitstream/10596/3385/1/4585204.pdf.
11. Ramiro DA, Escoute J, Petitot AS, 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—55.
12. Alvarenga SM, Caixeta ET, Hufnagel B, Thiebaut F, Maciel-Zambolim E, Zambolim L, et al. *In Silico* identification of coffee genome expressed sequences potentially associated with resistance to diseases. Genet Mol Biol 2010;33:795—806.
13. Rodrigues CJ, Bettencourt AJ, Rijo L. Races of the pathogen and resistance to coffee rust. Annu Rev Phytopathol 1975;13:49—70.
14. Diniz I, Azinheira H, Figueiredo A, Gichuru E, Oliveira H, Guerra-Guimarães L, et al. Fungal penetration associated with recognition, signaling and defence-related genes and peroxidase activity during the resistance response of coffee to Colletotrichum kahawae. Physiol Mol Plant Pathol 2019;105:119—27.
15. Kotlobay AA, Sarkisyan KS, Mokrushina YA, Marcet-Houben M, Serebrovskaya EO, Markina NM, et al. Genetically encodable bioluminescent system from fungi. Proc Natl Acad Sci USA 2018;115:12728—32.
16. 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—57.
17. Toruño TY, Stergiopoulos I, Coaker G. Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. Annu Rev Phytopathol 2016;54:419—41.
18. Barka GD. Identification, molecular characterization and differential expression studies of genes activated during coffea arabica L. - Hemileia vastatrix Berk. & Broome interaction [Internet]. Universidade Federal de Viçosa 2017. Available from http://hdl.handle.net/123456789/9087.
19. Hou S, Liu Z, Shen H, Wu D. Damage-associated molecular pattern-triggered immunity in plants. Front Plant Sci 2019;10:1—16.
20. Maia T, Badel JL, Marin-Ramirez G, Rocha Cde M, Fernandes MB, da Silva JCF, et al. The *Hemileia vastatrix* effector HvEC-016 suppresses bacterial blight symptoms in coffee genotypes with the S*H*1 rust resistance gene. New Phytol 2017;213:1315—29.
21. Gichuru EK, Agwanda CO, Combes MC, Mutitu EW, Ngugi ECK, Bertrand B, et al. Identification of molecular markers linked to a gene conferring resistance to coffee berry disease (*Colletotrichum kahawae*) in *Coffea arabica*. Plant Pathol 2008;57:1117—24.
22. Mondego JMC, Vidal RO, Carazzolle MF, Tokuda EK, Parizzi LP, Costa GGL, et al. An EST-based analysis identifies new genes and reveals distinctive gene expression features of Coffea arabica and Coffea canephora. BMC Plant Biol 2011;11:1—22.
23. Vinecky F, da Silva FR, Andrade AC. *In Silico* analysis of cDNA libraries SH2 and SH3 for the identification of genes responsive to drought in coffee. Coffee Sci 2012;7:1—19.
24. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 2007;35:D61—5.
25. Medlock GL, Moutinho TJ, Papin JA. Medusa: software to build and analyze ensembles of genome-scale metabolic network reconstructions. PLoS Comput Biol 2020;16:1—11.
26. Lefebvre-Pautigny F, Wu F, Philippot M, Rigoreau M, Priyono ZM. High resolution synteny maps allowing direct comparisons between the coffee and tomato genomes. Tree Genet Genomes 2010;6:565—77.
27. Kämper J, Kahmann R, Bölker M, Ma LJ, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature 2006;444:97—101.

28. Ren X, Wang C, Ren Z, Wang J, Zhang P, Zhao S, et al. Genetics of resistance to leaf rust in wheat: an overview in a genome-wide level. Sustainability 2023;15:1−27.
29. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics 2013;29:1072−5.
30. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: *ab Initio* prediction of alternative transcripts. Nucleic Acids Res 2006;34:W435−9.
31. Jones DAB, Rozano L, Debler JW, Mancera RL, Moolhuijzen PM, Hane JK. An automated and combinative method for the predictive ranking of candidate effector proteins of fungal plant pathogens. Sci Rep 2021;11:1−13.
32. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 2021;49:W293−6.
33. Janowska-Sejda EI, Lysenko A, Urban M, Rawlings C, Tsoka S, Hammond-Kosack KE. PHI-Nets: a network resource for ascomycete fungal pathogens to annotate and identify putative virulence interacting proteins and siRNA targets. Front Microbiol 2019;10:1−18.
34. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, et al. PHI-base: the pathogen-host interactions database. Nucleic Acids Res 2019;48:D613−20.
35. Rice P, Longden L, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet 2000;16:276−7.
36. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 2015;16:1−14.
37. Hill EH, Solomon PS. Extracellular vesicles from the apoplastic fungal wheat pathogen Zymoseptoria tritici. Fungal Biol Biotechnol 2020;7:1−14.
38. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 2021;38:3022−7.
39. Sperschneider J, Dodds PN. EffectorP 3.0: prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. Mol Plant Microbe Interact 2022;35:146−56.
40. Dautt-Castro M, Rosendo-Vargas M, Casas-Flores S. The small GTPases in fungal signaling conservation and function. Cells 2021;10:1−29.
41. An B, Li B, Qin G, Tian S. Function of small GTPase Rho3 in regulating growth, conidiation and virulence of Botrytis cinerea. Fungal Genet Biol 2015;75:46−55.
42. Kayano Y, Tanaka A, Takemoto D. Two closely related Rho GTPases, Cdc42 and RacA, of the en-dophytic fungus *Epichloë festucae* have contrasting roles for ROS production and symbiotic infection synchronized with the host plant. PLoS Pathog 2018;14:1−30.
43. Jung H, Ventura T, Sook Chung J, Kim WJ, Nam BH, Kong HJ, et al. Twelve quick steps for genome assembly and annotation in the classroom. PLoS Comput Biol 2020;16:1−25.
44. Majidian S, Agustinho DP, Chin CS, Sedlazeck FJ, Mahmoud M. Genomic variant benchmark: if you cannot measure it, you cannot improve it. Genome Biol 2023;24:1−25.
45. Silva Mdo C, Guerra-Guimarães L, Diniz I, Loureiro A, Azinheira H, Pereira AP, et al. An overview of the mechanisms involved in Coffee-*Hemileia vastatrix* interactions: plant and pathogen perspectives. Agronomy 2022;12:326.
46. Yan Y, Yuan Q, Tang J, Huang J, Hsiang T, Wei Y, et al. *Colletotrichum higginsianum* as a model for understanding host−pathogen interactions: a review. Int J Mol Sci 2018;19:2142.
47. Sperschneider J, Dodds PN, Gardiner DM, Manners JM, Singh KB, Taylor JM. Advances and challenges in computational prediction of effectors from plant pathogenic fungi. PLoS Pathog 2015;11:1−7.
48. Nachtweide S, Stanke M. Multi-genome annotation with AUGUSTUS. In: Kollmar M, editor. Gene Prediction: Methods and Protocols, Methods in Molecular Biology. New York: Humana; 2019:139−60 pp.
49. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 2019;37:420−3.
50. Fernandez D, Tisserant E, Talhinhas 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.
51. Talhinhas 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;5:1−17.
52. Chen XR, Wang Y, Kale SD, Fang Y, Srivastava V. Editorial: apoplastic effectors — what roles do they play in plant-pathogen interactions? Front Microbiol 2023;14:1−2.
53. Guerra-Guimarães L, Tenente R, Pinheiro C, Chaves I, Silva Mdo C, Cardoso FMH, et al. Proteomic analysis of apoplastic fluid of *Coffea arabica* leaves highlights novel biomarkers for resistance against *Hemileia vastatrix*. Front Plant Sci 2015;6:1−16.
54. Ung PMU, Rahman R, Schlessinger A. Redefining the protein kinase conformational space with machine learning. Cell Chem Biol 2018;25:916−24.e2.
55. Junutula JR, Schonteich E, Wilson GM, Peden AA, Scheller RH, Prekeris R. Molecular characterization of Rab11 interactions with members of the family of Rab11-interacting proteins. J Biol Chem 2004;279:33430−7.
56. Thomas LL, Highland CM, Fromme JC. Arf1 orchestrates Rab GTPase conversion at the trans-Golgi network. Mol Biol Cell 2021;32:1104−20.
57. Pfeffer SR. Rab GTPases: specifying and deciphering organelle identity and function. Trends Cell Biol 2001;11:487−91.

58. Francis SH, Busch JL, Corbin JD. cGMP-dependent protein kinases and cGMP phosphodiesterases in nitric oxide and cGMP action. Pharmacol Rev 2010;62:525—63.
59. Trotta CR, Paushkin SV, Patel M, Li H, Peltz SW. Cleavage of pre-tRNAs by the splicing endonuclease requires a composite active site. Nature 2006;441:375—7.
60. Chen X, Li C, Wang H, Guo Z. WRKY transcription factors: evolution, binding, and action. Phytopathol Res 2019;1:1—15.
61. Mediouni M, Makarenkov V, Diallo AB. Towards an interpretable machine learning model for predicting antimicrobial resistance. J Glob Antimicrob Resist 2025:47—51. https://doi.org/10.1016/j.jgar.2025.08.011.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/jib-2025-0013).