



**HAL**  
open science

## Phylogenetic analysis of *Hemileia vastatrix* and related taxa using a genome-scale approach

Diogo N. Silva, Ana Vieira, Pedro Talhinhos, Helena Gil Azinheira, Maria Do Ceu Silva, Diana Fernandez, Octavio S Paulo, Sébastien Duplessis, Dora Batista

► **To cite this version:**

Diogo N. Silva, Ana Vieira, Pedro Talhinhos, Helena Gil Azinheira, Maria Do Ceu Silva, et al.. Phylogenetic analysis of *Hemileia vastatrix* and related taxa using a genome-scale approach. 24. International Conference on Coffee Science. ASIC 2012, Nov 2012, San Jose, Costa Rica. hal-01267895

**HAL Id: hal-01267895**

**<https://hal.science/hal-01267895>**

Submitted on 3 Jun 2020

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

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

# Phylogenetic analysis of *Hemileia vastatrix* and related taxa using a genome-scale approach

SILVA, DIOGO N.\*<sup>‡</sup>, VIEIRA, ANA\*, TALHINHAS, PEDRO\*, AZINHEIRA, HELENA G.\*, SILVA, MARIA DO CÉU\*, FERNANDEZ, DIANA<sup>§</sup>, DUPLESSIS, SÉBASTIEN<sup>‡</sup>, PAULO, OCTÁVIO S<sup>‡</sup>, BATISTA, DORA\*.

\* CIFC/IICT - Centro de Investigação das Ferrugens do Cafeeiro/Instituto de Investigação Científica Tropical, Quinta do Marquês, 2784-505 Oeiras, Portugal, <sup>‡</sup>Computational Biology and Population Genomics group, Centro de Biologia Ambiental, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal, <sup>§</sup>Institut de Recherche pour le Développement, UMR 186 IRD/CIRAD-UM2 Résistance des Plantes aux Bioagresseurs, 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France, <sup>‡</sup>Institut National de la Recherche Agronomique UMR 1136 INRA/Université de Lorraine Interactions Arbres/Micro-organismes, Centre INRA de Nancy, 54280 Champenoux, France.

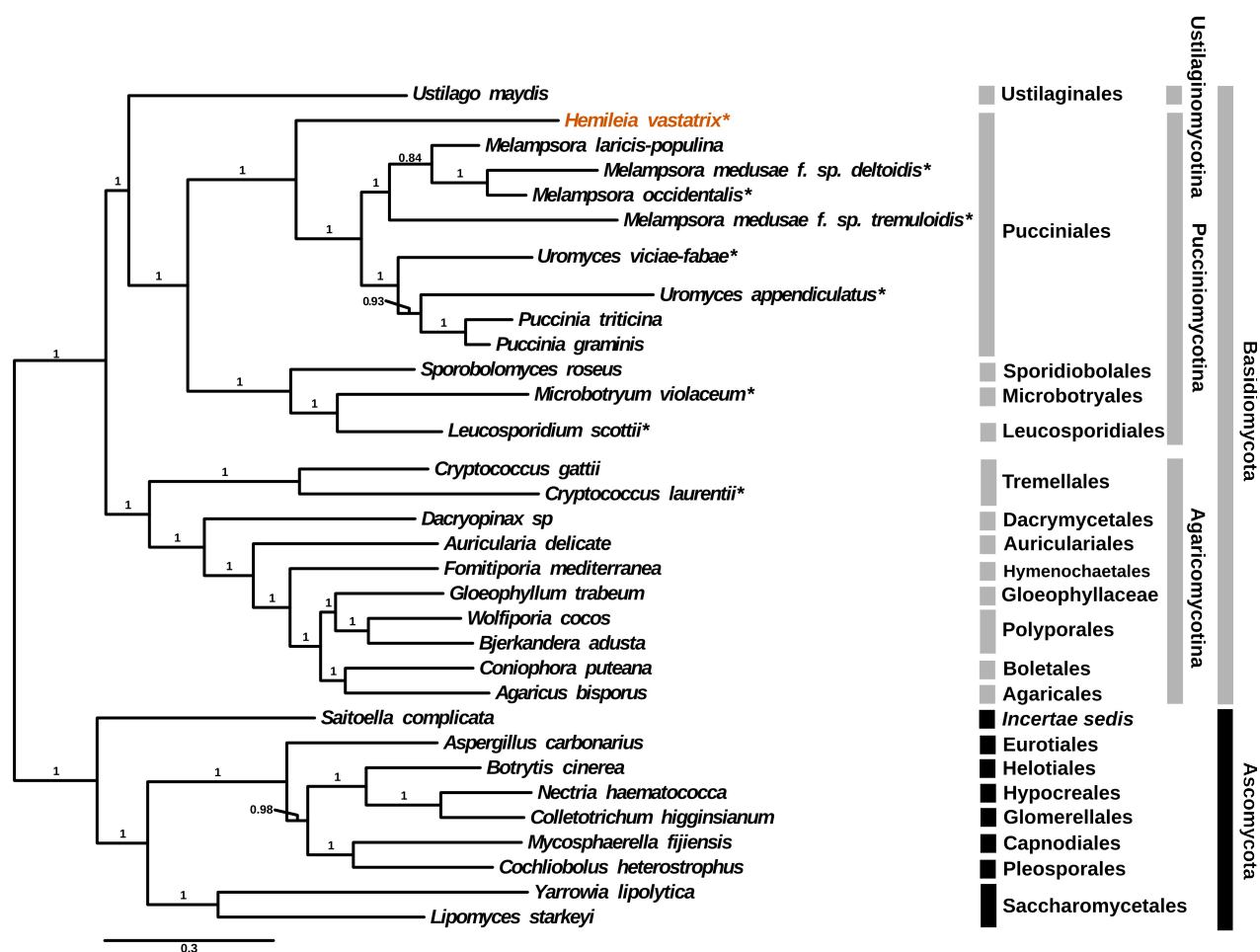
## SUMMARY

For more than a century, Coffee Leaf Rust caused by the biotrophic fungus *Hemileia vastatrix* (Berkeley & Broome), has increasingly stood out as one of the major factors hampering Arabica coffee production. Since its first historical outburst in the 19<sup>th</sup> century in Sri Lanka, this disease has rapidly spread worldwide and currently occurs in nearly all the regions of the world where coffee is grown. Despite its widespread distribution and negative economic impact, little is known about the pathogen's evolutionary origin and phylogenetic placement in the fungal tree of life. Attaining this knowledge, however, would provide fundamental insights on the evolutionary context in which *H. vastatrix* emerged, which in turn could have important implications to understand the evolution of pathogenicity in this fungus and in other rusts as well. With this in mind, we are undertaking a genome-scale approach on 32 fungal species to allow *H. vastatrix* and other rust fungi to be placed in the fungal tree of life with unprecedented details. The complete proteomes of three Pucciniomycotina species as well as of 11 other Basidiomycetes and nine Ascomycetes, approximately 250 000 publicly available EST sequences from several Pucciniomycotina species and transcriptome data recently obtained for *H. vastatrix* were used. A high quality matrix that includes orthologs, co-orthologs and recent paralogs was prepared with a sophisticated orthology detection strategy. On a first approach, we were able to identify at least 378 single-copy orthologs that were used for the phylogenetic analyses. Providing robust and resolved phylogenetic relationships will lay the ground for the identification of genes or gene families that are exclusive to *H. vastatrix* as well as genes of rapid evolution and/or under positive selection, which would be prime targets for functional studies aiming at disease control and prevention. This information will ultimately contribute significantly to advance our knowledge on *H. vastatrix*'s pathogenicity.

## INTRODUCTION

*Hemileia vastatrix* (Berkeley & Broome), the causal agent of Coffee Leaf Rust, is a biotrophic fungus considered to be one of the major burdens to Arabica coffee production world-wide. However, despite the growing relevance of its devastating impact on the sustainable production of coffee crops, very little has been investigated on the evolutionary origin and phylogenetic relationships of *H. vastatrix* with its related taxa. *Hemileia vastatrix*, which belongs to the Basidiomycota phylum of fungi, order Pucciniales (rust fungi), has only been included in a few phylogenetic studies of the rust fungi that employ a small number of loci [1-3]. Following major revisions that rejected the original hypothesis that the evolutionary primitive rust fungi were those occurring in ferns [4], these phylogenetic studies consistently place *H. vastatrix* as the one of the most primitive lineages of the Pucciniales.

In the last few years, the continuous generation of genomic data for several fungal species in the last



**Figure 1.** Maximum Likelihood tree estimate of the 378 gene data set for 32 fungal taxa using RAxML. Values above branches represent bootstrap values from 250 replicates. Taxa with an asterisk (\*) appended are represented by EST data, otherwise they are represented by a complete genome. On the right, the first set of grey bars contains the taxonomic order while the second contains the sub-phylum.

few years is creating unprecedented opportunities of going beyond and studying, for instance, the evolutionary drivers leading to the formation of pathogenic species [5,6]. Since genomes hold a vast and intricate record of their carriers' evolutionary history, it is now becoming possible to harness this informative potential and go beyond the phylogenetic reconstruction of species. By employing hundreds or thousands of loci in a comparative framework across multiple species, large regions of the genome can be screened for signatures of natural selection in specific historical periods of target species or group of species with greater detail and robustness [7]. These data can also be integrated with information from the fossil record to produce precise and accurate time-calibrated phylogenies, in which the phenotypic evolution of key pathogenic traits can be mapped and correlated with available environmental data from past geologic eras. However, due to the relatively recent availability of such genomic resources and scalable computational methods to analyze this amount of data, such integrative effort is still lacking.

In this work we report the first preliminary results of a phylogenomics project aiming at an integrative approach including a wider range of fungal species, here mainly centered in *H. vastatrix* and the rust fungi

## MATERIALS AND METHODS

Sequence data from 23 complete genomes, including the comprehensive analyses of rust fungi genomes [8] and 9 EST databases were collected from multiple public sources and from *H. vastatrix*'s transcriptome sequencing project [9,10]. Putative orthology relationships were first assessed using the deduced proteomes of the 23 complete genomes (171,987 proteins) using a Markov Clustering algorithm (MCL). Selected clusters were used as core orthologs to be complemented with EST data. Only those clusters that contained sequences of *H. vastatrix* were selected. The resulting clusters containing both genomic and EST protein data were then aligned

with MAFFT v6.903b. To better explore our data set, two approaches were followed for the phylogenomic analyses. First, a Maximum Likelihood (ML) tree estimation was performed with RAxML using the rapid bootstrap algorithm (“-f a -x”), the PROTCATWAG substitution model and 250 bootstrap replicates. Second, a phylogenomic network was constructed with SplitsTree 4 using the NeighbourNet algorithm in order to visualize potential conflicting signals in the data.

## RESULTS AND DISCUSSION

From the orthology assessment strategy employed in this work, 378 clusters (single-copy orthologs) were identified after the initial processing of the MCL output, completion with EST data and subsequent selection of clusters that would maximize the presence of *H. vastatrix* in the data matrix. This data matrix was remarkably complete for the complete genome taxa (average of 148,371 aligned sites) and patchy for the EST taxa (average of 9,161 aligned sites; 17,504 aligned sites for *H. vastatrix*). Even though the proportion of aligned data for EST taxa relatively low compared to complete genome taxa, it still represents a large absolute number of aligned sites for a group of taxa that is commonly analysed with a few loci.

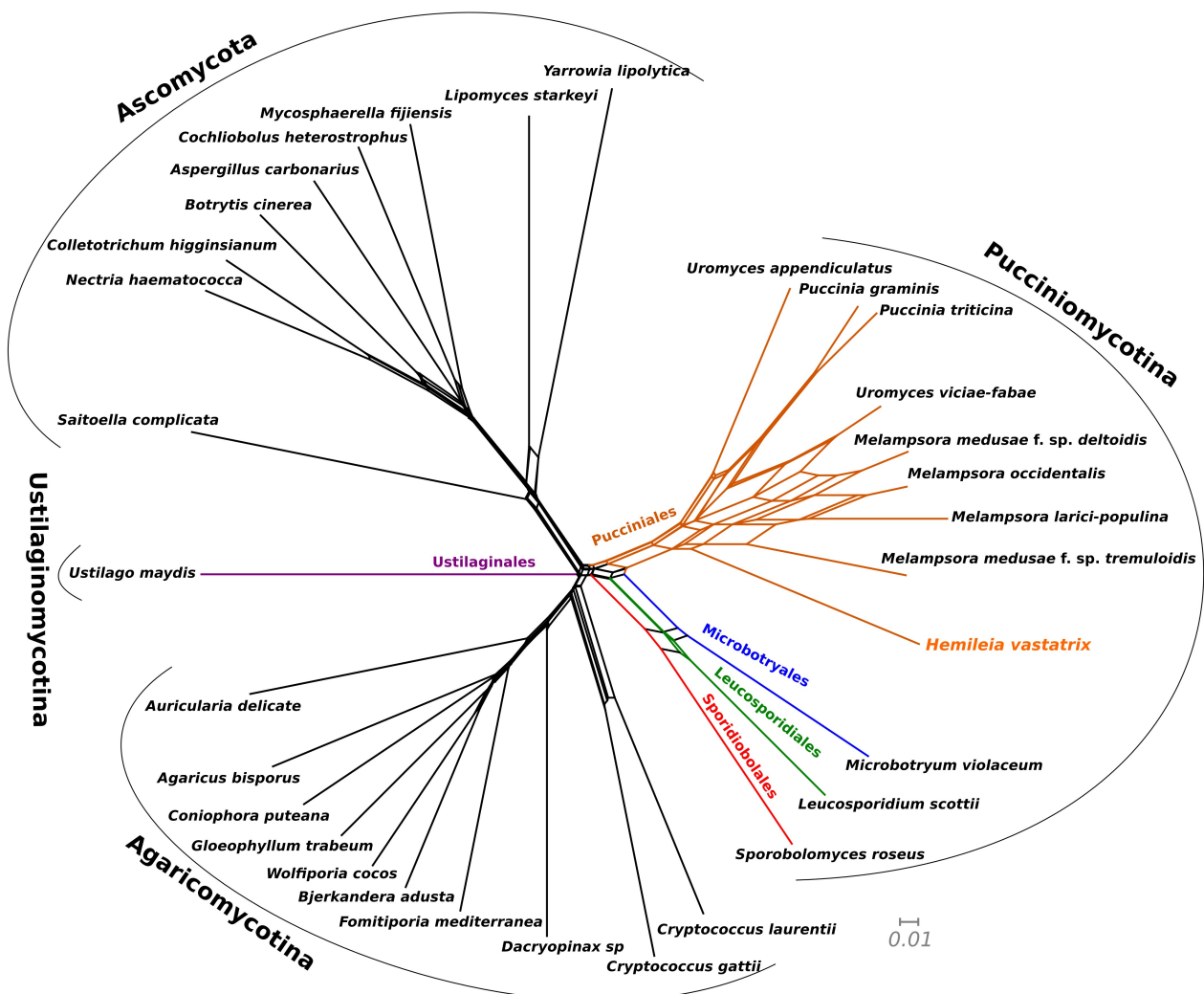


Figure 2. Neighbour-net based on the 378 gene data set for 32 fungal species. All supraclass and most supraordinal relationships are clearly defined by wide bosses, indicating limited conflict in the data. However, there is a substantial conflicting signal within the Pucciniales and there is uncertainty on the position of the Ustilaginomycotina relative to the Pucciniomycotina and Agaricomycotina

Our ML tree reconstruction resulted in an exceptionally resolved and well supported phylogeny, particularly within the Pucciniales, compared to other recent attempts [11] (Fig 1). These results support the traditional view that *H. vastatrix* occupies a basal position within the order [2] and also highlight the polyphyletic relationships of the *Uromyces* genus [12]. In addition, it is also noteworthy that the relationships recovered by the ML approach among the three sub-phyla

Ustilaginomycotina, Pucciniomycotina, Agaricomycotina, which were not resolved until very recently [11] already challenge the current view by placing the Ustilaginomycotina as a sister group of the Pucciniomycotina instead of the Agaricomycotina. Using the alternative approach of the NeighbourNet algorithm implemented in SplitsTree, the same general pattern is recovered with two additional nodes (Fig 2). First, there are considerably more possible splits within the Pucciniales, which could indicate a source of genuine conflicting phylogenetic signal in our data or simply reflect the higher proportion of missing data for this group with the inclusion of EST data. Second, the position of the Ustilaginomycotina is again ambiguous in relation to the other two sub-phyla. Solving these problems will most likely require that addition of more high quality data from further loci and taxa and the application of additional filtering and quality control steps on the data matrix [13,14]. Establishing a robust and well resolved phylogeny will then ensure the first and most important step in our project, by laying the basis from which further analyses can be undertaken such as divergence time estimation, detection of natural selection and mapping of pathogenic traits or evolution of gene families on the evolutionary history of the rust fungi and *H. vastatrix*.

## ACKNOWLEDGEMENTS

This work is being funded by Portuguese national funds through Fundação para a Ciência e a Tecnologia (projects PTDC/AGR-GPL/119943/2010 and PTDC/AGR-GPL/114949/2009), by grant UI88/5537/2011 (Universidade de Aveiro, Portugal) and by CEA/Genoscope-INRA-IRD Collaborative project (<http://www.genoscope.cns.fr/spip/Identification-of-virulence.html>) (France), whose funding is gratefully acknowledged.

## LITERATURE CITED

- [1] Wingfield B.D., Ericson L., Szaro T., Burdon JJ. 2004. Phylogenetic Patterns in the Uredinales. *Australasian Plant Pathology* 33: 327–335.
- [2] Aime M.C. 2006. Toward Resolving Family-level Relationships in Rust Fungi (Uredinales). *Mycoscience* 47: 112–122.
- [3] Grasso V., Sierotzki H., Garibaldi A., Gisi U. 2006. Relatedness Among Agronomically Important Rusts Based on Mitochondrial Cytochrome b Gene and Ribosomal ITS Sequences. *Journal of Phytopathology* 154: 110–118.
- [4] Savile D.B.O. 1979. Fungi as aids in higher plant classification. *Botanical Review* 45: 377–503.
- [5] Delsuc F., Brinkmann H., Philippe H. 2005. Phylogenomics and the Reconstruction of the Tree of Life. *Nature Reviews Genetics* 6: 361–375.
- [6] Raffaele S., Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*. 10:417-430.
- [7] Boussau B., Daubin V. 2010. Genomes as Documents of Evolutionary History. *Trends in Ecology & Evolution* 25: 224–32.
- [6] Dos Reis M., Inoue J., Hasegawa M., Asher R.J., Donoghue P.C.J., Yang Z. 2012. Phylogenomic Datasets Provide Both Precision and Accuracy in Estimating the Timescale of Placental Mammal Phylogeny. *Proceedings of The Royal Society Biological Sciences* 279: 3491-3500.
- [8] Duplessis S., Cuomo C.A., Lin Y.-C., et al. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America*. 108:9166-9171.
- [9] Talhinhos P., Azinheira H.G., Loureiro A., Batista D., Vieira B., Pina-Martins F., Tisserant E., Petitot A.S., Paulo O.S., Duplessis S., Silva M.C., Fernandez D. 2010. Overview of the functional virulent genome of the coffee leaf rust pathogen *Hemileia vastatrix*. *Proceedings of the 23rd International Conference on Coffee Science ASIC*. 414-422.
- [10] Fernandez D., Tisserant E., Talhinhos P., Azinheira H., Vieira A., Petitot A-S., Loureiro A., Poulain J., da Silva C., Silva M.C., Duplessis S. 2012. 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. *Molecular Plant Pathology* 13: 17-37
- [11] Ebersberger I., Simoes R.M., Kupczok A., Gube M., Kothe E., Voigt E., Haeseler A. 2012. A Consistent Phylogenetic Backbone for the Fungi. *Molecular Biology and Evolution* 29: 1319–1334.
- [12] Maier W., Wingfield B.D., Mennicken M., Wingfield M.J. 2007. Polyphyly and Two Emerging Lineages in the Rust Genera *Puccinia* and *Uromyces*. *Mycological Research* 111: 176–185.
- [13] Laurie J.D., Ali S., Linning R., et al. 2012. Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. *The Plant Cell*. 24:1733-1745.
- [14] Schirawski J., Mannhaupt G., Munch K., et al. 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. *Science*. 330:1546-1548.