



**HAL**  
open science

## Proteome dataset of *Hemileia vastatrix* by LC–MS/MS label-free identification

Natália Faustino Cury, Daiane Gonzaga Ribeiro, Jonathan Dias de Lima, Pollyana da Nóbrega Mendes, Diana Fernandez, Wagner Fontes, Mariana S Castro, Marcelo V Sousa, Natália F Martins, Angela Mehta

### ► To cite this version:

Natália Faustino Cury, Daiane Gonzaga Ribeiro, Jonathan Dias de Lima, Pollyana da Nóbrega Mendes, Diana Fernandez, et al.. Proteome dataset of *Hemileia vastatrix* by LC–MS/MS label-free identification. *Data in Brief*, 2022, 43, pp.108433. 10.1016/j.dib.2022.108433 . hal-03948572

**HAL Id: hal-03948572**

**<https://hal.inrae.fr/hal-03948572>**

Submitted on 20 Jan 2023

**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.



Distributed under a Creative Commons Attribution 4.0 International License



## Data Article

# Proteome dataset of *Hemileia vastatrix* by LC–MS/MS label-free identification



Natália Faustino Cury<sup>a</sup>, Daiane Gonzaga Ribeiro<sup>a,b</sup>,  
Jonathan Dias de Lima<sup>c</sup>, Pollyana da Nóbrega Mendes<sup>a</sup>,  
Diana Fernandez<sup>a,d</sup>, Wagner Fontes<sup>e</sup>, Mariana S. Castro<sup>e</sup>,  
Marcelo V. Sousa<sup>e</sup>, Natália F. Martins<sup>f</sup>, Angela Mehta<sup>a,\*</sup>

<sup>a</sup> Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, PqEB, Av. W5 Norte (final), Brasília, DF 70770-917, Brazil

<sup>b</sup> Centro de Análises Proteômicas e Bioquímicas, Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Taguatinga, DF 71966-700, Brazil

<sup>c</sup> Laboratório de Inorgânica e Materiais (LIMA - IQ/UnB), Departamento de Ciências de Materiais, Universidade de Brasília, Planaltina, DF 73.345-010, Brazil

<sup>d</sup> IRD - Institut de Recherche pour le Développement, PHIM Plant Health Institute, Univ Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France

<sup>e</sup> Laboratório de Bioquímica e Química de Proteínas, Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF, 70919-900, Brazil

<sup>f</sup> Embrapa Agroindústria Tropical, Rua Dra. Sara Mesquita, no 2.270, Bairro Planalto do Pici, Fortaleza, CE 60511-110, Brazil

## ARTICLE INFO

## Article history:

Received 10 May 2022

Revised 23 June 2022

Accepted 27 June 2022

Available online 3 July 2022

Dataset link: [MassIVE MSV000087665](https://massive.ucsb.edu/external/vol1/study?id=MSV000087665)  
(Original data)

## Keywords:

*Hemileia vastatrix*

Coffee rust

Proteome

LC–MS/MS

## ABSTRACT

Here we describe the proteome of the fungus *Hemileia vastatrix* by label free mass spectrometry (LC–MS/MS). *H. vastatrix* is the causal agent of coffee rust disease, causing great economic losses in this crop. The objective of our work was to identify *H. vastatrix* proteins potentially involved in host colonization and infection, by exploring the shotgun proteomics approach. A total of 742 proteins were identified and are associated with several crucial molecular functions, biological processes, and cellular components. The proteins identified contribute to a better understanding of the metabolism of the fungus and may help identify target proteins for the development of specific drugs in order to control coffee rust disease. All data can be accessed

\* Corresponding author.

E-mail address: [angela.mehta@embrapa.br](mailto:angela.mehta@embrapa.br) (A. Mehta).

at the Centre for Computational Mass Spectrometry – Massive MSV000087665 -<https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=cc71ad75f767451abe72dd1ce0019387>

© 2022 The Author(s). Published by Elsevier Inc.  
This is an open access article under the CC BY license  
(<http://creativecommons.org/licenses/by/4.0/>)

## Specifications Table

Subject	Biological sciences
Specific subject area	Omics: Proteomics Fungus proteomics
Type of data	Raw data by nano UPLC-MS/MS
How the data were acquired	<ul style="list-style-type: none"> <li>LC-MS/MS was performed on an LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) coupled to an Ultimate 3000 RSLCnano UPLC system (Thermo Scientific Dionex).</li> <li>MS was operated in DDA mode, acquiring precursor ions at 120000 resolution and fragmenting the top 15 precursors by HCD.</li> <li>MS/MS data was analysed by de novo sequencing and sequence database searching.</li> </ul>
Data format	The set of spectra obtained was stored in .RAW.
Description of data collection	Urediniospores of <i>H. vastatrix</i> were collected from infected <i>Coffea arabica</i> L. plants and spores with a germination rate higher than 80% were used for germination.
Data source location	<ul style="list-style-type: none"> <li>Institution: Embrapa Recursos Genéticos e Biotecnologia</li> <li>City/Town/Region: Brasília</li> <li>Country: Brazil</li> </ul>
Data accessibility	Public <b>Repository name:</b> Centre for Computational Mass Spectrometry – MASSIVE <b>Data identification number:</b> Massive MSV000087665 <b>Direct URL to data:</b> <a href="https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=cc71ad75f767451abe72dd1ce0019387">https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=cc71ad75f767451abe72dd1ce0019387</a> <b>Instructions for accessing these data:</b> Web access: <a href="https://massive.ucsd.edu/">https://massive.ucsd.edu/</a> MassIVE ID = MSV000087665 PorteomeXchange ID = PXD026810 DOI = <a href="https://doi.org/10.25345/C5FK0D">10.25345/C5FK0D</a>

## Value of the Data

- The proteins identified in this study contribute to better understand the metabolism of *Hemileia vastatrix*.
- The data obtained can help researchers and agricultural industries to identify target proteins for the development of specific drugs in order to control coffee rust disease.
- The dataset of *H. vastatrix* proteins represents valuable information that contributes to the Pucciniales proteome repertoire.

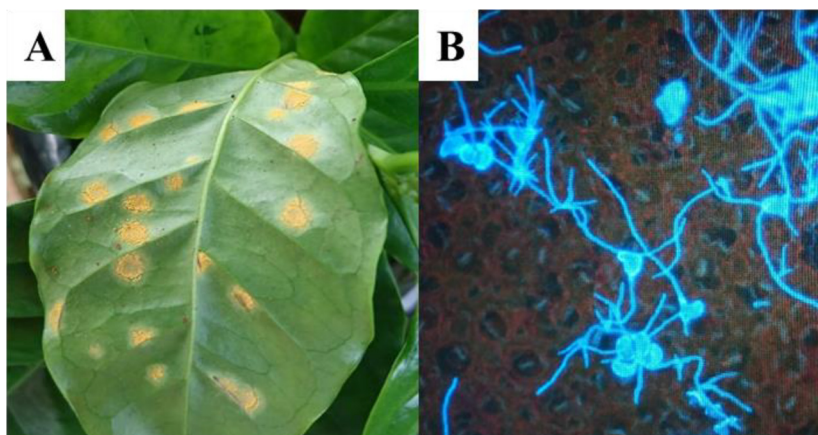
## 1. Data Description

The dataset described here was obtained from the proteome analysis of the fungus *Hemileia vastatrix* Berkeley and Broome (Basidiomycota, Pucciniales). A total of 742 proteins of *H. vastatrix* were identified using the PEAKS software and the proteins were deposited in the Massive repository under the ID MSV000087665. All the files presented in the Massive repository are

**Table 1**

Description of the folders deposited in the Massive / ProteomeXchange repositories.

Files	File description
Methods_and_protocols	Configuration used in chromatography and mass spectrometry analyses.
Raw_spec_files	Set of obtained spectra.
Sequence_database	Database sequences retrieved from the Uniprot repository in May 2021, filtered for the order Pucciniales, taxon ID 5258.
Peaklist_files	File in the MGF (Mascot generic file) format, containing the peaklist used for identification in the Peaks software, required for validation in the Massive workflow.
Result_files	File in the MZID (mzIdentML) format, containing the identification results exported by the Peaks software, based on the search of the peaklist in the database, both mentioned above, also used for validation in the Massive workflow.
Search_engine_files	All files generated by Peaks software, in which the complete set of spectra and protein identification were analyzed. The tables of identified peptides and proteins can be found in the "export" subfolder.
Supplementary_files	The iterative sequence of mappings (Uniprot, DB2DB on BioDBnet and BLASTKoala) and their results.



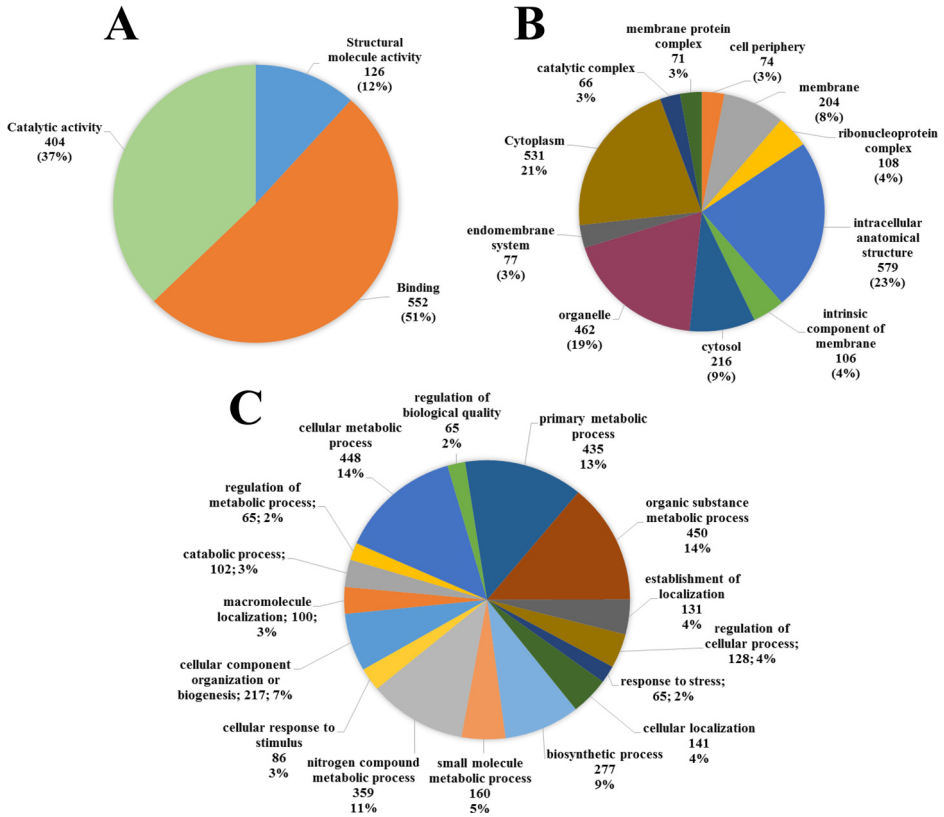
**Fig. 1.** (A) *Hemileia vastatrix* urediniospores germinated at coffee (*Coffea arabica*) leaf surface. (B) Rust (*H. vastatrix*) pustules on a coffee leaf.

described in Table 1. The proteins identified in *H. vastatrix* germinating urediniospores (Fig. 1) were classified according to their molecular functions, cellular components, and biological processes categories (Fig. 2).

## 2. Experimental Design, Materials and Methods

The urediniospores of *H. vastatrix* (race II, isolate Hv01) were collected from artificially infected leaves of *C. arabica* (var. Catuaí Amarelo) plants grown in a greenhouse (Fig. 1). Approximately 10 mg of urediniospores were spread in 10 mL distilled water and allowed to germinate in Petri dishes kept in the dark at 24 °C.

Germinated spores (> 80%) and non-germinated spores (altogether called germinating urediniospores (gU) sample) were harvested after 24 hours by centrifugation at 12000 rpm for 2 min. The gU from five Petri dishes were collected into one single tube to form the *H. vastatrix* gU sample used for protein extraction as described by Ribeiro et al. [1]. For tryptic digestion, the sample was solubilized with 60 µL of 50 mM ammonium bicarbonate (NH<sub>4</sub>HCO<sub>3</sub> pH 8.5), then



**Fig. 2.** Diagram showing the molecular functions (A), biological processes (B), and cellular sites (C) of identified proteins in germinating urediniospores of *Hemileia vastatrix*.

25 µL of RapiGest™ SF – Waters (0.2% v/v) was added. The sample was reduced with dithiothreitol (100 mM), alkylated with iodoacetamide (300 mM) and proteins were digested using 200 ng of trypsin at 37 °C for 19 h.

The peptides were desalted as described in Rappsilber et al. [2], with some modifications [1], solubilized with 0.1% formic acid and injected into an Ultimate 3000 RSLCnano UPLC (Thermo Scientific Dionex) for reversed-phase nano-chromatography. A total of 1 µg of solubilized peptides were injected into a trap column (2 cm × 100 µm), containing ReproSil-Pur C18-AQ 5 µm reverse phase resin (Dr. Maisch GmbH, Germany) in triplicate. The sample was eluted from the trap column to the analytical column (32 cm × 75 µm containing ReproSil-Pur C18-AQ 3 µm reverse phase resin –Dr. Maisch GmbH, Germany). A gradient of 2–35% acetonitrile, 0.1% formic acid was used and the sample was sprayed into the ionization source of the spectrometer.

The LTQ Orbitrap Elite mass spectrometer was operated in data-dependent acquisition (DDA) mode, generating MS1 spectra in the Orbitrap analyzer (with resolution of 120 000 FWHM at 400 m/z) between the masses 300–1650 m/z and dynamic exclusion of 10 ppm. The 15 most intense ions were chosen for each MS1 spectrum automatically with charges higher than two and directed toward higher energy collision-induced dissociation (HCD). The configuration for HCD was: 2.0 m/z isolation window with automatic gain control (AGC) of 1 × 10<sup>6</sup>, and maximum fill time of 100 ms, with normalized collision energy at 35% and threshold for the selection of 3000.

Alignment of spectra and quantification of peptides were performed using Progenesis® QI for proteomics v.1.0 software [3] and proteins were identified using Peaks® 7.0 software [4]. The

sequences were deduced from the fragmentation information and the search performed in the Uniprot (Universal Protein Resource) repository in May 2021, filtered to the order Pucciniales (Taxon ID: 5258). The search was performed based on de novo sequencing and PSM with the following parameters: tolerance for the mass of the precursor of 10 ppm, and of 0.05 Da for the fragments, tolerance of up to 2 missing cleavages, carbamidomethylation of cysteines as a fixed modification and methionine oxidation as a variable modification. Protein identifications were considered as being reliable at FDR < 1%, presenting at least two unique peptides. Finally, the proteins identified were functionally annotated using Blast2GO software [5].

## Ethics Statements

This research involved neither human participants nor animals.

## Declaration of Competing Interest

The authors declare that they have no known financial interests and personal relationships that could inappropriately influence the work reported in this paper.

## Data Availability

[MassIVE MSV000087665 \(Original data\)](#) (Centre for Computational Mass Spectrometry – MASSIVE).

## CRedit Author Statement

**Natália Faustino Cury:** Writing – original draft, Formal analysis; **Daiane Gonzaga Ribeiro:** Methodology, Writing – original draft; **Jonathan Dias de Lima:** Methodology; **Pollyana da Nóbrega Mendes:** Methodology, Writing – original draft; **Diana Fernandez:** Methodology, Investigation; **Wagner Fontes:** Formal analysis, Data curation, Software; **Mariana S. Castro:** Software; **Marcelo V. Sousa:** Resources; **Natália F. Martins:** Project administration, Resources; **Angela Mehta:** Investigation, Supervision, Writing – review & editing.

## Acknowledgements

This research was supported by Bayer (Grants for Targets) (Project code 20.21.00.042.00.00), Embrapa Recursos Genéticos e Biotecnologia, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Institut de Recherche pour le Développement (IRD) and CONCAFÉ (Project code 10.18.20.001.00.00). The authors thank Dr Marcelo O. Rodrigues from Instituto de Química – UNB for the inestimable help and Dr. Eduardo S.G. Mizubuti (Universidade de Viçosa) for providing the isolate used in this study.

## References

- [1] D.G. Ribeiro, R.F. de Almeida, W. Fontes, M. de Souza Castro, M.V. de Sousa, C.A.O. Ricart, R.N.V. da Cunha, R. Lopes, J.E. Scherwinski-Pereira, A. Mehta, Stress and cell cycle regulation during somatic embryogenesis plays a key role in oil palm callus development, *J. Proteom.* 192 (2019) 137–146, doi:[10.1016/j.jprot.2018.08.015](https://doi.org/10.1016/j.jprot.2018.08.015).
- [2] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, *Nat. Protoc.* 2 (8) (2007) 1896–1906, doi:[10.1038/nprot.2007.261](https://doi.org/10.1038/nprot.2007.261).
- [3] T. Välikangas, T. Suomi, L.L. Elo, A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation, *Briefings Bioinf.* 19 (6) (2018) 1344–1355, doi:[10.1093/bib/bbx054](https://doi.org/10.1093/bib/bbx054).

- [4] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, G. Lajoie, PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry, *Rapid Commun. Mass Spectrom.* 17 (20) (2003) 2337–2342, doi:[10.1002/rcm.1196](https://doi.org/10.1002/rcm.1196).
- [5] A. Conesa, S. Götz, Blast2GO: a comprehensive suite for functional analysis in plant genomics, *Int. J. Plant Genom.* 2008 (2008) 619832, doi:[10.1155/2008/619832](https://doi.org/10.1155/2008/619832).