

Genome sequence of *Hemileia vastatrix* Berk. and Br. (Race I), the causal agent of coffee leaf rust, isolate from Risaralda, Colombia

Carlos A. Ángel C.,¹ Gustavo A. Marín-Ramírez,¹ Carlos E. Maldonado²

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT Coffee leaf rust, caused by the fungus *Hemileia vastatrix* (Basidiomycota; Pucciniomycota), is a devastating disease spread worldwide. To improve the available genomes, we use PacBio HiFi sequencing enhanced by Dovetail Omni-C chromatin conformation capture to assemble a highly contiguous 747.98 Mb genome of an isolate collected from *Coffea arabica*.

KEYWORDS coffee leaf rust, *Hemileia vastatrix*, *Coffea arabica*, genomes, PacBio HiFi, Dovetail Omni-C, annotation, transposable elements

Coffee leaf rust (CLR), caused by *Hemileia vastatrix* Berk and Br., is the most important coffee disease, causing epidemics and yielding losses up to 80% in susceptible varieties (1, 2). Here, we present a high-quality highly contiguous whole-genome assembly of *H. vastatrix* Race I, improving the previous reported assemblies (3–5).

We isolated *H. vastatrix* from a *Coffea arabica* var. Caturra plantation in the rural area of Pereira (Risaralda, Colombia), at 4°44'46.25"N, 75°36'14.59" W. Twenty successive inoculation cycles of 40–50 days on 5-month-old plants of the Caturra variety were done to increase the urediniospore mass to obtain enough spores for DNA isolation and phenotyping. The isolate was inoculated on 26 coffee genotypes carrying different combinations of CLR resistance genes (*SH*). Compatible infection reaction was observed on carriers of the genes *SH2* and *SH5*, assigning the isolate to Race I (6) (Fig. 1).

The DNA isolation and sequencing were performed by the Arizona Genomics Institute. High molecular weight DNA was isolated from urediniospores using a modified CTAB protocol (7). A PacBio 30 Kb library was constructed from size-selected DNA (BluePippin system) using the SMRTbell express template prep kit 2.0, and two SMRTcells 8M v1 were sequenced on a Sequel II sequencer using the SeqII 1.0 chemistry. Dovetail Genomics constructed an Omni-C library using a proximity-ligation protocol on immobilized chromatin, followed by library construction using NEBNext Ultra enzymes and Illumina-compatible adapters, followed by sequencing on an Illumina HiSeqX platform. PacBio produced 3,439,300 circular consensus sequences (HiFi reads) with an N50 of 16,166 bp, while 139.2 million paired reads of 150 bp were produced from the Omni-C library, resulting in coverage greater than 70× and 58×, respectively, for an ~719 Mbp genome estimated by GCE v1.0.2 (8). For polishing, 50 million DNBseq 150-PE reads, covering 21× the genome, were produced by Complete Genomics from DNA isolated using the DNeasy Plant Kit (Qiagen).

A primary genome assembly was generated by Hifiasm (9) using HiFi reads. The assembly was scaffolded by HiRise pipeline (10) using the Omni-C sequences and polished by Polca (11) using DNBseq data. Contaminants were trimmed by FCS, FCS-GX (<https://github.com/ncbi/fcs/>), and NCBI Genome Workbench (12). The final polished assembly had a total length of 747,980,314 bp, a consensus quality of Q40.82, and a GC content of 33.78%. Assembly statistics were obtained by QUAST 5.2.0 (13) (Table 1), reporting a BUSCO genome completeness score of 97.93% using the fungi_odb9 as

Editor Jason E. Stajich, University of California, Riverside, California, USA

Address correspondence to Carlos A. Ángel C., carlosariangel@gmail.com, or Carlos E. Maldonado, carlosernesto.maldonado@cafedecolombia.com.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 20 June 2023

Accepted 8 September 2023

Published 4 October 2023

Copyright © 2023 Angel et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

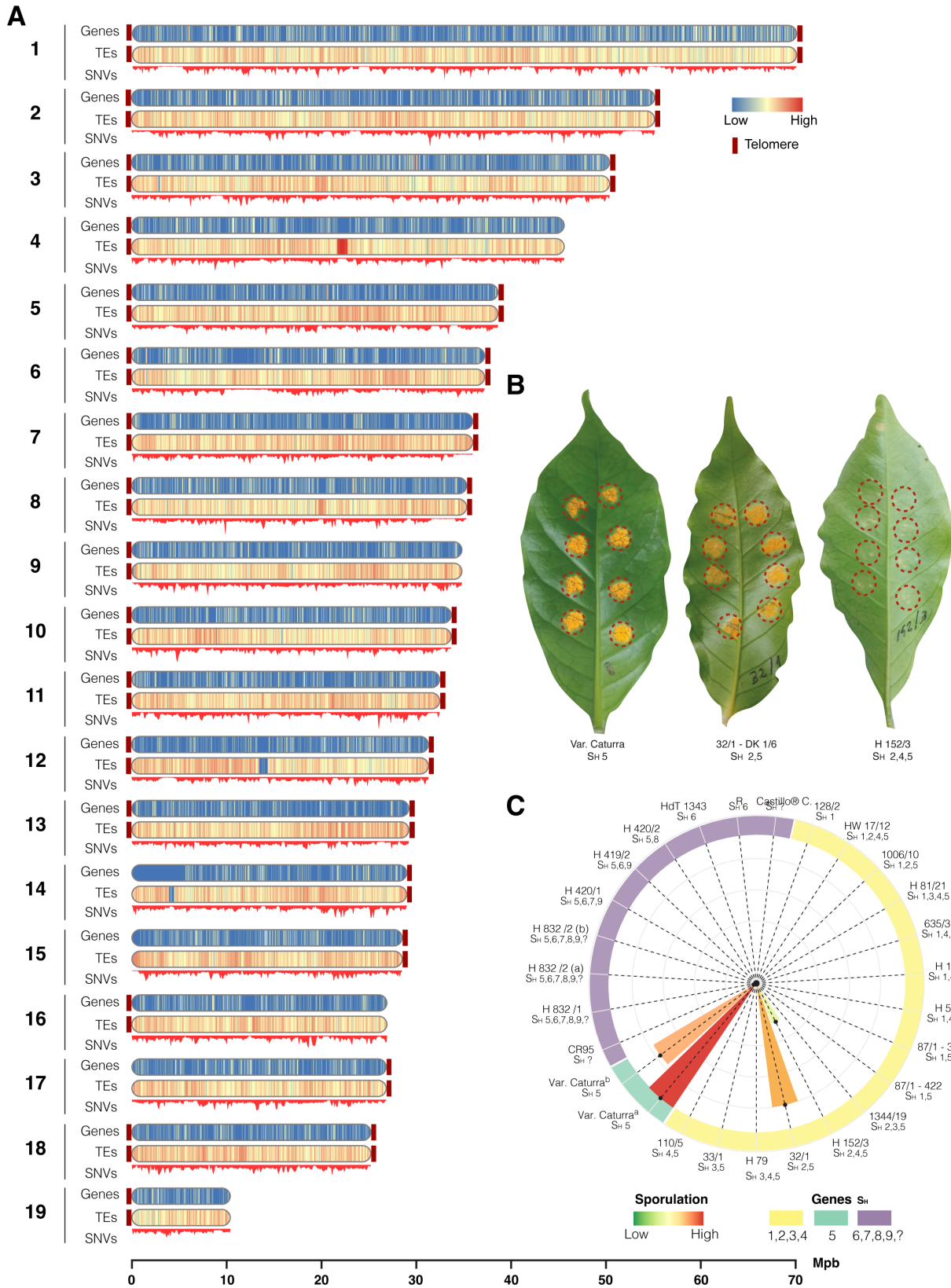


FIG 1 Phenotyping and genomic analysis of *Hemileia vastatrix* Race I. (A) Features of the 19 largest scaffolds including telomere positions, heat maps of genes and transposable elements in 10 Kbp windows, and distribution of single nucleotide variants (SNVs) from polishing. (B) Phenotypic characterization of the *H. vastatrix* isolate inoculated on hosts carrying different CLR resistance genes (*SH*). (C) Percentage of sporulated inoculation sites on 26 coffee differential clones; Var. Caturra^a and Var. Caturra^b are the inoculation positive controls incubated under controlled and uncontrolled environment, respectively.

Downloaded from https://journals.asm.org/journal/mra on 12 May 2026 by 200.3.167.2.

TABLE 1 Genomic features and statistics of *Hemileia vastatrix* Race I after assembly, scaffolding, and polishing

	Primary assembly HiFiasm v. 0.14	Dovetail HiRise assembly	Polished assembly
Total length (bp)	775,474,960	775,477,860	747,980,110
N50	19,794,549	35,487,552	35,997,362
L50	14	9	8
N90	531,823	531,823	25,694,141
L90	55	26	18
Largest scaffold	45,329,991	71,418,495	71,415,227
Number of contigs	985		
Number of scaffolds		956	794
Number of gaps	0	29	29
GC (%)	33.91	33.91	33.78

lineage data set. Transposable elements were *de novo* called using the EDTA pipeline (14), which masked 81.66% of the genome. Evidence-based annotation was performed by Augustus (15, 16), calling 12,870 protein genes on the masked sequence. Telomere repeat sequences screening was done by Tapestry v1.0.1 (17), the motif CCCTAA/TTAGGG. Telomeres were found at the end of the 19 longest scaffolds, 11 of those were assembled telomere to telomere (Fig. 1).

ACKNOWLEDGMENTS

Research funded by Colombian Coffee Growers Federation (FNC)/National Coffee Research Center (CENICAFE - Crossref Funder ID 100019597), project ID: PAT102006. The authors wish to thank Isabel C. Ramirez, Jorge D. Ocampo, and Carlos A. Zuluaga, from Plant Pathology – CENICAFE, for the *in vivo* maintenance of the Race I isolate and spores' production. To Dave Kudrna and Jayson Talag from the Arizona Genomics Institute (AGI) for wet lab work and primary assembly. To Dovetail Genomics (now Cantata Bio, LLC) for scaffolding services. To Biommit personnel and Andres F. López for their support on data analysis.

AUTHOR AFFILIATIONS

¹Plant Pathology, Colombian National Coffee Federation (FNC)/National Coffee Research Center (CENICAFE), Chinchiná, Colombia

²Plant Breeding, Colombian National Coffee Federation (FNC)/National Coffee Research Center (CENICAFE), Chinchiná, Colombia

AUTHOR ORCID*s*

Carlos A. Ángel C.  <http://orcid.org/0000-0003-4312-1390>

Gustavo A. Marín-Ramírez  <http://orcid.org/0000-0002-2145-9756>

Carlos E. Maldonado  <http://orcid.org/0000-0001-9433-6979>

FUNDING

Funder	Grant(s)	Author(s)
Centro Nacional de Investigaciones de Café (CENICAFÉ)	PAT102006	Carlos E. Maldonado Carlos A. Ángel C. Gustavo A. Marín-Ramírez

AUTHOR CONTRIBUTIONS

Carlos A. Ángel C., Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision | Gustavo A. Marín-Ramírez, Data curation, Formal

analysis, Visualization, Writing – original draft, Writing – review and editing | Carlos E. Maldonado, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The Hv_R1_Cat_CENPAT genome assembly is available at NCBI under the BioProject accession number [PRJNA912191](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA912191), BioSample [SAMN32232808](https://www.ncbi.nlm.nih.gov/biosample/SAMN32232808), GenBank assembly accession [GCA_030280995.1](https://www.ncbi.nlm.nih.gov/genbank/GCA_030280995.1) WGS PacBio and DNBseq sequences are available on NCBI's SRA database with the IDs [SRR22911636](https://www.ncbi.nlm.nih.gov/sra/SRR22911636) and [SRR22911637](https://www.ncbi.nlm.nih.gov/sra/SRR22911637). The evidence-based structural annotation for this assembly is available on the CENICAFE: institutional repository (<https://doi.org/10.38141/10799/dataset02>) as well as the transposable elements library derived from *de novo* calling (<https://doi.org/10.38141/10799/dataset01>).

REFERENCES

- Rivillas CA, Serna CA, Cristancho MA, Gaitán A. 2011. La roya del café en Colombia: impacto manejo Y costos del control. *Bol Téc Cenicafé* 36:1–51.
- Zambolim L. 2016. Current status and management of coffee leaf rust in Brazil. *Trop plant pathol* 41:1–8. <https://doi.org/10.1007/s40858-016-0065-9>
- Cristancho MA, Botero-Rozo DO, Giraldo W, Tabima J, Riaño-Pachón DM, Escobar C, Roza Y, Rivera LF, Durán A, Restrepo S, Eilam T, Anikster Y, Gaitán AL. 2014. Annotation of a hybrid partial genome of the coffee rust (*Hemileia vastatrix*) contributes to the gene repertoire catalog of the Pucciniales. *Front Plant Sci* 5:594. <https://doi.org/10.3389/fpls.2014.00594>
- Porto BN, Caixeta ET, Mathioni SM, Vidigal PMP, Zambolim L, Zambolim EM, Donofrio N, Polson SW, Maia TA, Chen C, Adetunji M, Kingham B, Dalio RJD, MLV. 2019. Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression dynamics of candidate effectors dependent on host compatibility. *PLoS One* 14:e0215598. <https://doi.org/10.1371/journal.pone.0215598>
- Tobias PA, Edwards RJ, Surana P, Mangelson H, Inácio V, do Céu Silva M, Várzea V, Park RF, Batista D. 2022. A Chromosome-level genome resource for studying virulence mechanisms and evolution of the coffee rust pathogen *Hemileia vastatrix*. *Genomics*. <https://doi.org/10.1101/2022.07.29.502101>
- Rodrigues CJ Jr, Bettencourt AJ, Rijo L. 1975. Races of the pathogen and resistance to coffee rust. *Annu Rev Phytopathol* 13:49–70. <https://doi.org/10.1146/annurev.py.13.090175.000405>
- Porebski S, Bailey LG, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Rep* 15:8–15. <https://doi.org/10.1007/BF02772108>
- Liu BH, Shi YJ, Yuan JY, Fan W. 2013. Estimation of genomic characteristics by analyzing K-MER frequency in *de novo* genome project. *Quant Biol* 35:62–67.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with Hifiasm. *Nat Methods* 18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>
- Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE. 2016. Chromosome-scale shotgun assembly using an *in vitro* method for long-range linkage. *Genome Res* 26:342–350. <https://doi.org/10.1101/gr.193474.115>
- Zimin AV, Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLOS Comput Biol* 16:e1007981. <https://doi.org/10.1371/journal.pcbi.1007981>
- Kuznetsov A, Bollin CJ. 2021. NCBI genome workbench: desktop software for comparative genomics, visualization, and genBank data submission. *Methods Mol Biol* 2231:261–295. https://doi.org/10.1007/978-1-0716-1036-7_16
- Mikheenko A, Prijibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>
- Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, Jiang N, Hirsch CN, Hufford MB. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* 20:275. <https://doi.org/10.1186/s13059-019-1905-y>
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res* 34:W435–W439. <https://doi.org/10.1093/nar/gkl200>
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntentically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24:637–644. <https://doi.org/10.1093/bioinformatics/btn013>
- Davey JW, Davis SJ, Mottram JC, Ashton PD. 2020. Tapestry: Validate and edit small Eukaryotic genome assemblies with long reads. *Bioinformatics*. <https://doi.org/10.1101/2020.04.24.059402>