Phytopathology® • 2022 • 112:1998-2011 • https://doi.org/10.1094/PHYTO-09-21-0376-R

Population Biology

e-Xtra*

Worldwide Population Structure of the Coffee Rust Fungus *Hemileia vastatrix* Is Strongly Shaped by Local Adaptation and Breeding History

Ana Sofia B. Rodrigues,[1] Diogo Nuno Silva,[1,2] Vitor Várzea,[2] Octávio S. Paulo,[1] and Dora Batista[1,2,†]

[1] Centre for Ecology, Evolution and Environmental Changes (cE3c), Computational Biology and Population Genomics Group (CoBiG2), Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal
[2] Centro de Investigação das Ferrugens do Cafeeiro (CIFC)/Linking Landscape, Environment, Agriculture and Food (LEAF), Instituto Superior de Agronomia, Universidade de Lisboa, 1349-017 Lisbon, Portugal
Accepted for publication 22 March 2022.

ABSTRACT

The devastating disease coffee leaf rust, caused by *Hemileia vastatrix*, has been a major constraint to worldwide coffee production. Recently, *H. vastatrix* populations were shown to be structured into three divergent genetic lineages with marked host specialization (C1, C2, and C3). However, there is yet no overall understanding of the population dynamics and adaptation of the most widespread and epidemiological relevant *H. vastatrix* group (C3). We used restriction site-associated DNA sequencing to generate 13,804 single nucleotide polymorphisms (SNPs) across a worldwide collection of 99 *H. vastatrix* isolates. Phylogenetic analyses uncovered a well-supported structuring within C3, with three main subgroups (SGs; SGI, SGII, and SGIII), which seem to reflect the historical distribution, breeding, and exchange of coffee cultivars. SGI shows a ladder-like diversification pattern and occurs across all four continents sampled, SGII is mainly restricted to Africa, and SGIII is observed only in Timor, revealing a higher genetic differentiation. Outlier and association tests globally identified 112 SNPs under putative positive selection, which impacted population structure. In particular, 29 overlapping SNPs per se seemed to have an extremely strong effect on *H. vastatrix* population divergence. We also found exclusive and fixed alleles associated with the SGs supporting local adaptation. Functional annotation revealed that transposable elements may play a role in host adaptation. Our study provides a higher-resolution perspective on the evolutionary history of *H. vastatrix* on cultivated coffee, showing its strong ability to adapt and the strength of the selective force imposed by coffee hosts, which should be taken into account when designing strategies for pathogen dissemination control and selective breeding.

Keywords: coffee leaf rust, disease control and pest management, fungal pathogens, genomics, host adaptation, host–parasite interactions, plant pathogen, population biology, population genomics, SNP association

Recurrent epidemics of fungal plant diseases constitute a growing worldwide problem and a major threat to the global economy and food security (Fones et al. 2020; Plissonneau et al. 2017). Many pathogens respond rapidly to selection pressure, leading to the inevitable breakdown of host resistance, which has been observed repeatedly in many crops and for many decades. Identifying the factors controlling the appearance and spread of these diseases and understanding the evolutionary dynamics of pathogen populations may thus be crucial for the design of more effective and durable control strategies. In recent years, population genomics has offered new opportunities to study the demography and population evolutionary history of important crop pathogens with unprecedented detail, illuminating dispersion routes and adaptive patterns (Vries et al. 2020).

Coffee leaf rust (CLR), caused by the obligate biotrophic rust fungus *Hemileia vastatrix* Berk. & Broome (Basidiomycota, Pucciniales), is a historically devastating disease that has been affecting global coffee production worldwide for more than one and a half centuries (McCook and Vandermeer 2015; Talhinhas et al. 2017). The disease causes premature defoliation, and, although normally nonlethal, it can lead to yield losses as high as 40% with serious social and economic consequences (Talhinhas et al. 2017). Within the >100 species in the genus *Coffea*, only *C. arabica* and *C. canephora* are considered economically relevant at a global scale, currently comprising 60 and 40% of world coffee production, respectively (International Coffee Organization 2020; McCook and Vandermeer 2015). *C. arabica*, which is a recent tetraploid hybrid between the diploid species *C. canephora* and *C. eugenioides* (Cenci et al. 2012; Clarindo and Carvalho 2009; Lashermes et al. 2000), holds a particular relevance in the market for high-quality coffee, but it is also the most susceptible species to *H. vastatrix* attacks. Nonetheless, *H. vastatrix* can infect all known cultivated *Coffea* species, but at different levels of severity (McCook 2006). Historical records from 1861 reveal the presence of CLR on wild *Coffea* species in Western Kenya (near Lake Victoria), implying a long history of interaction between *H. vastatrix* and coffee species prior to the beginning of coffee domestication, and suggesting eastern Africa as the region where the species first originated (Gichuru et al. 2012). Until the middle of the 19th century, even though the coffee plant was disseminated across the global tropics, the coffee rust fungus was contained in eastern Africa (McCook 2006). However, in 1869, it appeared for the first time as an epidemic in Sri Lanka (formerly Ceylon) and southern India, and, from there, it spread through the coffee production zones of the Indian Ocean Basin and the Pacific (McCook 2006). After this outbreak, the fungus reached the coffee farms of West Africa and finally crossed the Atlantic Ocean, where it expanded throughout the coffee production areas of the American continent (McCook 2006). Nowadays, *H. vastatrix* is distributed in almost every coffee production region around the world, and its current distribution and dissemination are intimately linked to the historical evolution of the global coffee

†Corresponding author: D. Batista; dorabatista@isa.ulisboa.pt

Current address of D. N. Silva: Lifebit Biotech Lda. Office 4, 219 Kensington High St., London, U.K. W8 6BD.

Funding: This work was cofunded by PORLisboa, Portugal2020, and European Union through FEDER funds (LISBOA-01-0145-FEDER-029189) and by the Foundation for Science and Technology (FCT) through Portuguese funds (PTDC/ASP-PLA/29189/2017).

*The *e-Xtra* logo stands for "electronic extra" and indicates supplementary tables and supplementary figures are published online.

The author(s) declare no conflict of interest.

industry. Only recently did the population evolutionary history of this pathogen begin to be uncovered by genome-wide single nucleotide polymorphism (SNP) data. For the first time, *H. vastatrix* populations were found to be structured into three divergent genetic lineages with marked host specialization (C1 and C2, infecting diploid coffee species; and C3, infecting tetraploid coffee species, e.g., *C. arabica* and derivative interspecific hybrids), revealing footprints of introgression (Silva et al. 2018). This contrasts sharply with previous genetic studies that mainly focused on local populations and globally pointed to unstructured populations regarding host and geographical origin (Cabral et al. 2016; Gouveia et al. 2005; Maia et al. 2013; Nunes et al. 2009; Rozo et al. 2012; Santana et al. 2018).

Rust fungi biology and epidemiology are known to be strongly shaped by complex coevolutionary histories with their host plants (Figueroa et al. 2020). Natural selection imposed by the host can be particularly more intense in managed agroecosystems than in natural ones, causing rapid evolution and dispersal of crop pathogens (Möller and Stukenbrock 2017). That is because the emergence of host specialization can be accelerated by the genetic homogeneity of plant hosts cultivated in high-density fields at large regional scales, which imposes strong directional selection while allowing the development of large pathogen populations (Corredor-Moreno and Saunders 2020; McDonald and Stukenbrock 2016; Stukenbrock and McDonald 2008). Even though similar agroecosystems can span over large spatial scales, local pathogen populations tend to be exposed to significant differences in the deployment of resistance genes, management systems, fungicide exposure, landscape, and annual fluctuations in temperature and precipitation (Pereira et al. 2020). Ultimately, our central question is how the interplay among local evolutionary forces the overwhelming evolutionary potential of pathogens and plant breeding strategies impact the demographic histories of plant pathogen populations.

In this work, the dynamics, history, and adaptive patterns of *H. vastatrix* populations were investigated, with a focus on the most widespread and epidemiologically relevant *H. vastatrix* genetic group (C3 lineage sensu Silva et al. 2018). For this, we expanded our previous investigation (Silva et al. 2018) to a larger worldwide-scale sampling of *H. vastatrix* isolates collected over time from tetraploid coffee hosts, including a broad range of pathotypes, and applied a restriction site-associated DNA (RAD) sequencing approach to obtain a set of genome-wide SNP markers. In particular, our objectives were to (i) evaluate the *H. vastatrix* genetic variation on a global scale and investigate population lineage subdivision, (ii) detect signatures of natural selection, and (iii) identify putative genomic regions involved in *H. vastatrix* adaptation to the coffee host.

MATERIALS AND METHODS

Sampling. A set of 99 *H. vastatrix* isolates from 23 geographical locations across four continents were retrieved from the historic spore collection maintained at Centro de Investigação das Ferrugens do Cafeeiro (CIFC), Instituto Superior de Agronomia, Universidade de Lisboa, Lisbon, Portugal (Supplementary Table S1). This is a unique collection of liquid nitrogen-preserved urediniospore samples representing all regions worldwide where CLR occurs, established in the early 1950s by the Portuguese plant pathologist Branquinho d'Oliveira (Talhinhas et al. 2017). The isolates used in this study were selected to represent maximum race (pathotype) diversity within each geographical region over time (from 1953 to 2013; Supplementary Table S1). Of these, 22 isolates were previously processed and sequenced by Silva et al. (2018). Total sampling includes isolates collected from different diploid and tetraploid coffee hosts, comprising 33 unique pathotypes determined previously on routinely performed tests at CIFC, based on inoculation assays on a set of coffee differentials bearing different resistance gene combinations under standard testing conditions (d'Oliveira 1954; Talhinhas et al. 2017). According to the virulence profiles on the differential plants, isolates are classified into pathotypes (i.e., races) comprising virulence genes as inferred by Flor's theory. These range from v1 to v9 in isolates derived from *C. arabica* and tetraploid interspecific hybrids, whereas those of the races that attack diploid coffee species are not known (v?). Pathotypes that are beyond the capacity of differential plants are designated as v* (d'Oliveira 1954; Talhinhas et al. 2017).

DNA extraction and RAD sequencing. A cetyltrimethylammonium bromide-based protocol modified from Kolmer et al. (1995) was used for DNA extraction. Genomic DNA concentration and quality were assessed by visual inspection on an agarose gel and with an ND-1000 Nanodrop spectrophotometer. Three micrograms of high-quality genomic DNA per individual was sent to Floragenex (Eugene, OR, U.S.A.) for RAD library preparation and sequencing. Libraries with sample-specific barcode sequences were produced from DNA digested with *Pst*I enzyme, and 100-bp single-end sequencing was performed using an Illumina HiSeq 2000 machine. The sequence data were deposited in the European Nucleotide Archive under accession code PRJEB47473 (https://www.ebi.ac.uk/ena/).

RAD sequencing quality filtering, assembly, and SNP calling. Sequence reads produced in this study were processed together with those obtained by Silva et al. (2018) for the isolates selected. Raw reads were trimmed, demultiplexed, and de novo assembled using IpyRAD 0.5.15 (Eaton and Overcast 2020). Assembly parameters described by Silva et al. (2018) were used. Reads with uncalled bases or distance to barcodes >1 were removed. All bases with a Phred quality score <20 were converted to Ns, and reads containing more than five Ns were discarded. Similarity threshold of 0.97 was used to cluster reads together and individuals into a locus. Only loci with a minimum coverage of four individuals were retained in the final dataset. To limit the risk of including paralogs in the analysis, loci sharing >50% heterozygous sites were not considered, and the maximum number of heterozygous sites in a consensus sequence (locus) allowed was eight. After clustering sequences, a data matrix for each locus and individual was generated. Replicates of seven samples (Supplementary Table S1) were sequenced and assembled separately and posteriorly used on the filtering process to eliminate possible artificial SNPs. Further filtering was performed using VCFtools v0.1.16 (Danecek et al. 2011) and customized script compare_pairs.py (https://github.com/CoBiG2/RAD_Tools/blob/master/compare_pairs.py) with the following criteria: each SNP had to be represented in ≥50% of the individuals, SNPs with a mismatch in at least one replicate were excluded, and the replicate with the largest number of SNPs after final filtering was retained. Furthermore, the minimum allele frequency of each SNP was 2%. Finally, to minimize the effects of linkage disequilibrium, downstream analyses were performed using only one SNP per locus by discarding all but the SNP closest to the center of the sequence in each locus. This final dataset composed of 99 *H. vastatrix* isolates and 11,118 SNP loci (All_99indv dataset; Table 1) was obtained using the python script vcf_parser.py (https://github.com/CoBiG2/RAD_Tools/blob/master/vcf_parser.py). Handling and exploration of alignment data matrices were performed using TriFusion v0.4.12 software (https://github.com/OdiogoSilva/TriFusion).

Phylogenetic analyses. Phylogenetic relationships among *H. vastatrix* samples were inferred using a single concatenated alignment that included loci with SNPs present in >50% of *H. vastatrix* isolates and a minor allele frequency >2% (Phylo_All_99indv dataset; Table 1). TriFusion was used to concatenate and convert the alignment matrices to the appropriate formats. Maximum-likelihood analyses using RAxML v8.2.12 (Stamatakis 2014) were conducted for the concatenated data matrices with the GTRCAT model of sequence evolution and with bootstrap support estimated from 1,000 replicates. RAxML runs were performed in the Cipres Science Gateway clusters (Miller et al. 2015).

These analyses were also performed using a matrix of SNP loci considered to be neutral (Phylo_neutral_loci dataset) and another

matrix consisting of all different and putatively adaptive SNP loci (Phylo_adaptive_loci dataset) obtained by the outlier/association analyses (see below; Table 1). For the sake of simplicity, putatively adaptive SNPs will be referred to hereon as adaptive SNPs. A matrix with only adaptive SNP loci that were common to outlier detection and association analyses (Phylo_common_loci dataset) was also used for assessing phylogenetic relationships among *H. vastatrix* isolates (Table 1). For these three datasets, *H. vastatrix* isolates with >70% missing data were excluded from the analyses.

Assessment of *H. vastatrix* population structure. Three distinct methods for clustering individuals were used to investigate the general pattern of population structure: (i) Structure (Pritchard et al. 2000); (ii) ALStructure (Cabreros and Storey 2019), which is a R package that estimates population structure by merging admixture models; and (iii) principal components (PC) analysis (PCA). Structure v.2.3.4 (Falush et al. 2003; Pritchard et al. 2000) was run using the admixture model. The *K* values tested ranged from 1 to 6. Ten replicates of each *K* were run, applying 50,000 steps of burn-in and 1,000,000 Monte Carlo Markov chain steps after burn-in. Structure_threader v1.2.2 (Pina-Martins et al. 2017) was used to parallelize the runs and find the *K* that best explains the data (Earl and vonHoldt 2012; Evanno et al. 2005). Structure_threader was also used to run ALStructure. These analyses were performed for the All_99indv dataset. PCA was performed with the script snp_pca_static.R (https://github.com/CoBiG2/RAD_Tools/blob/master/snp_pca_static.R), which implements the R package SNPRELATE v1.80 (Zheng et al. 2012). These analyses were performed using the original matrix of SNP loci from 99 *H. vastatrix* isolates (All_99indv dataset; Table 1) and seven more different subdatasets using only *H. vastatrix* isolates of C3 lineage sensu Silva et al. (2018) that infect tetraploid coffee species: (i) OnlyC3indv (includes only variable SNPs within C3 group), (ii) OnlyC3indv_outliers (only SNPs detected as outliers), (iii) Single-SNP_Assoc, (iv) Multi-SNP_Assoc (includes SNPs detected by the single- and multiassociation analyses, respectively), and finally (v) OnlyC3indv_adaptive_loci (includes all adaptive SNPs detected by all approaches), (vi) OnlyC3indv_neutral_loci (excludes all adaptive SNPs detected by all approaches), and (vii) OnlyC3indv_common_loci (includes only adaptive SNPs that were commonly detected in all approaches; Table 1). Subdatasets were obtained from the original matrix (All_99indv dataset) using custom shell scripts.

The degree of population differentiation was also assessed by calculating the pairwise fixation index (F_{ST}) values for each *H. vastatrix* group pair using VCFtools. The number of segregating sites, gene diversity, and fixation indexes (inbreeding coefficient F_{IT} and fixation index F_{IS}) were obtained using VCFtools and Genepop v4.2.2 (Rousset 2008).

Outlier detection of putative SNP loci under selection. To identify highly differentiated loci among populations potentially under selection an outlier approach based on F_{ST} distribution was applied. Two programs were used to perform the outlier detection: SelEstim v1.1.4 (Vitalis et al. 2014) and BayeScan v. 2.1 (Foll and Gaggiotti 2008). The methods implemented in these programs show the lowest rate of false positives (Narum and Hess 2011; Vitalis et al. 2014). SelEstim model assumes that all loci are under selection and estimates the intensity of selection at each locus. The posterior distributions of the locus-specific coefficients of selection are then compared with a distribution derived from the genome-wide effect of selection using Kullback-Leibler divergence that is calibrated with simulations from posterior predictive distribution based on observed data (Vitalis et al. 2014). BayeScan tests two alternative models for each locus, with or without selection, by implementing a Bayesian approach to estimate the posterior probability. Posterior odds are then obtained, and false discovery rate is calculated to control for multiple testing. For SelEstim analysis, 100 pilot runs of length 1,000 were followed by a main run of length 4×10^6, with a burn-in of 40,000 and a thinning interval of 20. A detection threshold of 0.0001 that corresponds to the 99.99% quantile of the Kullback-Leibler divergence distribution was selected. For BayeScan analysis, 20 pilot runs of length 5,000 followed by a main run of 500,000 iterations were performed. A burn-in of 50,000, a thinning interval of 10, and a detection threshold of 0.05 were used. Outlier detection analyses were performed using the OnlyC3indv dataset and clustered according to the phylogenetic subgroups (SGs) identified by the phylogenetic analyses. To reduce the chance of false positives, which can be common in this type of analyses (Gautier 2015; Vitalis et al. 2014), only SNPs that were detected as outliers by both programs were considered to be "real" outliers. The frequency of the alleles in each SG was then obtained for this set of loci using PLINK v1.07 (Purcell et al. 2007).

TABLE 1. Description of the datasets used in this study

Dataset	Number of loci	Number of SNPs[a]	Number of isolates	Analysis	Matrix description
All_99indv	11,118	11,118	99	PCA; outlier; single and multiassociation	Only central and variable SNPs present in >50% of *Hemileia vastatrix* isolates and MAF >2%
OnlyC3indv	2,422	2,422	91	PCA; outlier; single and multiassociation	Only central and variable SNPs within C3 group
OnlyC3indv_outliers	75	75	91	PCA; outlier; single and multiassociation	Only central and variable SNPs within C3 group detected by outlier analyses
Single-SNP_Assoc	46	46	91	PCA; outlier; single and multiassociation	Only central and variable SNPs within C3 group detected by single-association analysis
Multi-SNP_Assoc	67	67	91	PCA; outlier; single and multiassociation	Only central and variable SNPs within C3 group detected by multiassociation analysis
OnlyC3indv_neutral_loci	2,310	2,310	91	PCA; outlier; single and multiassociation	Only central and variable SNPs within C3 group considered neutral (i.e., excluding all 112 adaptive SNPs)
OnlyC3indv_adaptive_loci	112	112	91	PCA; outlier; single and multiassociation	Only central and variable SNPs within C3 group detected by at least one analysis (outlier, single or multiassociation)
OnlyC3indv_common_loci	29	29	91	PCA; outlier; single and multiassociation	Only central and variable SNPs within C3 group detected in common by all analyses (outlier, single and multiassociation)
Phylo_All_99indv	11,118	13,804	99	RaxML	Loci with all SNPs present in >50% of *H. vastatrix* isolates and MAF >2%
Phylo_neutral_loci	2,310	2,325	93	RaxML	Loci with SNPs considered neutral (i.e., excluding all 112 adaptive SNPs)
Phylo_adaptive_loci	112	127	93	RaxML	Loci with SNPs considered to be adaptive and detected by at least one analysis (outlier, single or multiassociation)
Phylo_common_loci	29	30	93	RaxML	Loci with SNPs considered to be adaptive and detected in common by all analyses (outlier, single and multiassociation)

[a] MAF, minor allele frequency; PCA, principal components analysis; SNP, single nucleotide polymorphism.

Genome-wide association analyses. For the OnlyC3indv dataset consisting of *H. vastatrix* isolates only from C3 lineage sensu Silva et al. (2018), single and multi-SNP correlations between SNPs and SGs were tested using a Bayesian variable selection regression model proposed by Guan and Stephens (2011) and carried out in piMASS v0.9. Three pairwise comparisons were tested: SGI versus SGII, SGI versus SGIII, and SGII versus SGIII. This method uses the phenotype (i.e., phylogenetic SGs) as the response variable and genetic variants (i.e., SNPs) as covariates to identify SNPs that may be associated with a particular phenotype (Guan and Stephens 2011). The posterior distribution of γ, or the posterior inclusion probability (PIP), identify SNPs that are statistically associated with phenotypic differences. In this study, SNPs with a PIP >99% empirical quantile ($PIP_{0.99}$ SNPs) were considered as highly associated with C3 lineages. PIP and the estimates of the phenotypic effect (β) were reported for all $PIP_{0.99}$ SNPs. A positive β in the combination phenotype1/phenotype2 (e.g., SGI/SGII) means that the frequency of the alternative allele is higher in phenotype2 (SGII in the example), and a negative β means that alternative allele is higher in phenotype1 (SGI in the example). The $|\beta|$ was considered to evaluate the phenotypic effect size of each $PIP_{0.99}$ SNP. Additional parameters of the model were estimated from the data: the number of SNPs in the regression model and the average phenotypic effect of a SNP that is included in the model (σSNP). For all pairwise analyses, the joint posterior probability distribution of model parameters was obtained using the Monte Carlo Markov chain method iterated for 4,000,000 generations, sampling at every 400 iterations and with the first 100,000 samples discarded as burn-in. A single-SNP approach was also performed with piMASS (Guan and Stephens 2011). For single-SNP pairwise analyses, SNPs >95% empirical quantile for Bayes factor (BF; $BF_{0.95}$ SNPs) were assumed to be correlated to the C3 lineages. From those, SNPs >99% empirical quantile for BF ($BF_{0.99}$ SNPs) were considered to have the strongest associations. Imputation of the missing genotypes was performed in BIMBAM v1.0 (Servin and Stephens 2007). fcgene v1.0.7 (Roshyara and Scholz 2014) was used to convert data to the appropriate format.

Functional annotation of putative adaptive SNP loci. The consensus sequence of the restriction site-associated DNA sequencing loci, considered as adaptive loci by outlier detection and association analyses, was generated using the python script loci_consensus.py (https://github.com/CoBiG2/RAD_Tools/blob/master/loci_consensus.py). Homology to noncoding and coding regions was investigated by locally querying consensus sequences against the NCBI nt database (RefSeq release 91, last modified 5 November 2018; and GenBank release 229, last modified 15 December 2018) using BLASTn v2.2.28+ (Altschul et al. 1997). An e-value threshold of 1e-30 was used. A protein BLAST against nonredundant sequences available in NCBI database using BLASTx 2.10.0+ (Altschul et al. 1997) was also performed directly on the NCBI webpage (on 27 October 2020) and using the MMseq2 software (Steinegger and Söding 2017). An e-value threshold of 1e-2 was used. Posteriorly, the python script xml2_best_hits.py (https://github.com/Duartb/RAD_Tools/blob/master/xml2_best_hits.py) was used to obtain a list with the identifiers corresponding to the best protein hit of each sequence for functional annotation of the loci. The UniProt database (UniProt Consortium) was then used to retrieve the Gene Ontology (GO) terms assigned to the biological process, molecular protein, and cellular component categories. The annotation was further improved by searching the putatively adaptive RAD loci of *H. vastatrix* against the pathogen–host interaction reference database (PHI-base) v.4.10 (Urban et al. 2019) with BLASTx using a threshold of 1e-1. Finally, the consensus sequences were also queried against the draft genome of *H. vastatrix* (PRJNA419278) (Porto et al. 2019) and against expressed sequence tag (EST) data of *H. vastatrix* (Talhinhas et al. 2014), with an e-value threshold of 1e-15 as the cutoff for restricting the alignments to the most significant ones.

RESULTS

Phylogenetic patterns of *H. vastatrix*. We identified a total of 62,856 SNPs across 99 *H. vastatrix* isolates. After the filtering steps, a final dataset matrix of 11,118 variable loci and 13,804 SNPs (Phylo_All_99indv) was obtained. Phylogenetic analysis of the 99 *H. vastatrix* isolates confirmed the three divergent and well-supported *H. vastatrix* clades (C1 to C3; bootstrap = 100) with marked host specialization (C1 and C2 infecting diploid coffee species and C3 infecting tetraploid coffee species, i.e., *C. arabica* and derivative interspecific hybrids) as previously reported by Silva et al. (2018) (Fig. 1). Interestingly, our reconstruction additionally revealed a clear and well-supported (bootstrap >95.0) structured pattern of three main SGs (SGI, SGII, and SGIII) within C3. SGI forms an independent clade with no apparent substructuring, comprising the higher number of isolates (n = 67), mainly of Asian origin, although it includes isolates from all geographical locations sampled, namely from Central and South America and Africa (Supplementary Fig. S1). In contrast, SGII (bootstrap = 100) and SGIII (bootstrap = 100) are sister lineages in a separated clade (bootstrap >95.0) and show a more specific and somewhat structured composition. Apart from a well-supported group clustering Brazilian isolates and one isolate from the Philippines, SGII comprises uniquely *H. vastatrix* isolates of African origin, whereas SGIII includes only three *H. vastatrix* isolates, all from Timor (Supplementary Fig. S1).

Population differentiation and genetic diversity. PCA revealed the same general pattern of three main clusters (C1, C2, and C3) as identified in the phylogenetic analyses, with the first and second PCs explaining 54.82 and 19.67% of the variance, respectively (Supplementary Fig. S4A), but no differentiation within C3 was observed. Structure (Supplementary Fig. S2) and ALStructure (Supplementary Fig. S3) corroborated the previous analyses, indicating that three is most likely the number of *H. vastatrix* groups. Pairwise F_{ST} estimates support the genetic differentiation between each of the three main *H. vastatrix* groups (C1, C2, and C3), with F_{ST} values ranging from 0.6537 (C1 and C2) to 0.8935 (C2 and C3; Table 2). When only isolates infecting tetraploid coffee species (i.e., *C. arabica* and derivative interspecific hybrids; C3) are analyzed (OnlyC3indv dataset with 2,422 SNPs), the PCA revealed population differentiation within this group (PC1, 8.99%; PC2, 5.08%; Fig. 2A). Four clusters were observed with all isolates of SGI clustering together as well as all isolates of SGIII, respectively. On the contrary, isolates of SGII showed some degree of variation, clustering into two different groups. However, low F_{ST} values were obtained for this dataset ($0.0770 < F_{ST} < 0.1187$), which would suggest high gene flow (Table 2).

Focusing on the C3 group (91 *H. vastatrix* isolates), genetic diversity indices were further estimated for the respective SGs (Supplementary Table S2). From a total of 2,422 SNPs (OnlyC3indv dataset), 1,988 segregated within SGI, 1,666 within SGII, and 475 within SGIII. The numbers of singletons in each group after excluding missing allele SNP loci were 359 (18.06%), 333 (19.99%), and 455 (95.79%) in SGI, SGII, and SGIII, respectively. The genetic diversity varied slightly among the three SGs, even considering the difference in the number of samples constituting each of them (Supplementary Table S2). We found the highest gene diversity in SGII (0.3256) and the lowest in SGI (0.2928). The inbreeding coefficient F_{IT} for each isolate considering the total C3 group was negative for most of the isolates, which indicates a slight to moderate excess of heterozygotic SNPs compared with theoretical expectation (Supplementary Fig. S5). Fixation index F_{IS} was also negative for the three C3 SGs, with SGIII showing the highest negative value and SGII the lowest (Supplementary Table S2). A significant excess of heterozygotes compared with that expected under Hardy-Weinberg equilibrium was also found for all three SGs ($P < 0.0001$).

Detection of SNP loci putatively under selection. To investigate the presence of signatures of selection, we focused on the

larger and more epidemiologically relevant *H. vastatrix* group C3, applying two different approaches. An outlier detection analysis was performed grouping C3 isolates according to the structure observed in the phylogenetic tree. Using SelEstim, we identified 92 SNP loci as outliers, while 131 outlier SNPs were detected by BayeScan (Fig 3 and Supplementary Table S3). The estimated α-coefficient for outlier SNPs detected in the BayeScan analysis was positive, indicating diversifying selection. A total of 75 SNP loci (3.10% of the analyzed loci) were identified by both methods and therefore considered to be potentially under the effect of positive selection (Supplementary Table S3).

In addition, single- and multi-SNP associations with C3 lineages were also tested. In single-SNP association analyses, a total of 208 $BF_{0.95}$ SNPs (>95% quantile BF) were found to be involved in the differentiation of the SGs, corresponding to 8.59% of the analyzed markers (Fig. 4A). For each pairwise comparison, the numbers of $BF_{0.95}$ SNPs were 122 for SG I versus SGII and SGI versus SGIII and 121 for SG II versus SGIII comparisons. When a stricter threshold (99% quantile BF) was applied, 46 $BF_{0.99}$ SNPs (1.90%) showed the strongest associations (Fig. 4A and Supplementary Table S4). Estimates of the phenotypic effects |β| associated with $BF_{0.99}$ SNPs for each comparison were low to moderate (SGI versus SGII, |β| = 0.3070; SGI versus SGIII, |β| = 0.0975; SGII versus SGIII |β| = 0.0156). These estimates were higher than the average |β| obtained for SGI versus SGII (|β|= 0.0538) and SGI versus SGIII (|β| = 0.0163) comparisons and considering the 2,422 overall SNPs, except for the SGII versus SGIII comparison (|β|= 0.0210). Twenty-seven of the SNP loci were shared among pairwise comparisons (Supplementary Table S4). For multi-SNP association analyses, estimates of the mean number of SNPs underlying SGs differentiation ranged from 42 to 65 (Supplementary Table S5). Sixty-seven different $PIP_{0.99}$ SNPs were revealed to be associated with C3 lineage differentiation, and eight of these SNP loci were shared by pairwise comparisons (Fig. 4B and Supplementary Table S6). The average effect of associated SNPs was high and similar among analyses but slightly higher in comparisons involving SGI (SGI versus SGII, σSNP = 1.765; SGI versus SGIII, σSNP = 1.133; SGII versus SGIII, σSNP = 1.112; Supplementary Table S5). The PIPs for the analyzed SNPs were quite similar among pairwise analyses (SGI versus SGII, PIP = 0.0177; SGI versus SGIII, PIP = 0.0269; SGII versus SGIII, PIP = 0.0131; Supplementary Table S5). In total, 112 different SNP loci (Supplementary Table S7) under putative selection were retrieved by all methods. From these, 29 SNP loci were common to all the analyses, 46 SNP loci were detected only by outlier analyses, 1 SNP locus was detected exclusively by single-SNP association analyses, and 36 SNP loci were found only by multi-SNP association analyses.

Effect of putatively adaptive SNP loci in the population and phylogenetic structure of the C3 lineage. We then assessed the influence of local adaptation in the C3 lineage structuring pattern. PCA using the 75 putative SNP loci (OnlyC3indv_outliers dataset) showed a high degree of population differentiation (PC1, 46.57%; PC2, 30.87%) with a pattern of three clusters corresponding to the three groups identified by the phylogenetic analysis (I, II, and III; Supplementary Fig. S4B). This same pattern of differentiation among C3 groups was also recovered when using the 46 associated $BF_{0.99}$ SNPs (Supplementary Fig. S4C) and the 67 $PIP_{0.99}$ SNPs (Supplementary Fig. S4D) separately, although a higher percentage of the variance is explained by the 46 associated $BF_{0.99}$ SNPs dataset (PC1, 50.22%; PC2, 32.92%, Supplementary Fig. S4C). Nevertheless, the clustering pattern is maintained when combining all 112 putatively adaptive SNPs (OnlyC3indv_adaptive_loci dataset; Fig. 2C) or only the 29 intersected SNPs (OnlyC3indv_common_loci dataset; Fig. 2D). In fact, all adaptive SNP datasets can recover the previous phylogenetic groups, but it is noteworthy that the set of 29 common SNPs per se is enough to reconstruct the global population structure. In contrast, the PCA of the neutral dataset (OnlyC3indv_neutral_loci dataset) revealed a different and less-structured pattern, with the PCs explaining a very low percentage of the variance (PC1, 7.10%; PC2, 5.12%; Fig. 2B). In this scenario, a cluster comprising isolates of SGI and SGIII is obtained and isolates of SGII are distributed in two main clusters. This pattern resembles the one obtained with all 2,422 SNPs (OnlyC3_indv dataset; Fig. 2A).

Phylogenetic reconstruction using the Phylo_neutral_loci dataset (without all adaptive SNP loci) corroborated the existence of the three main *H. vastatrix* groups (C1 to C3) (bootstrap = 100) but did not provide the previous observed clear substructure within C3 (Supplementary Fig. S6). However, the pattern obtained using only (i) the 112 different adaptive loci (Phylo_adaptive_loci dataset; Supplementary Fig. S7) or (ii) the 29 common SNP loci (Phylo_common_loci dataset; Supplementary Fig. S8) fully reconstruct the three divergent and well-supported SGs (SGI, SGII, and SGIII) observed for the total dataset (Phylo_All_99indv; Fig. 1).

Pairwise F_{ST} estimates further supported the strong contribution of putative adaptive SNPs for *H. vastatrix* population genetic differentiation. A higher and near-complete genetic differentiation among C3 SGs is particularly observed when considering the 46 associated $BF_{0.99}$ SNPs ($0.9966 < F_{ST} < 0.9983$) and the 29 intersected common SNPs ($0.9965 < F_{ST} < 0.9985$) compared with F_{ST} values obtained for the OnlyC3_indv ($0.0770 < F_{ST} < 0.1187$) and for the "neutral" SNP loci (OnlyC3indv_no_putative_loci dataset; $0.0016 < F_{ST} < 0.0523$; Table 2). For the remaining datasets, F_{ST} values corroborated the high genetic differentiation among SGs, but with lower magnitude and larger differences between the higher and lower differentiated groups (75 outlier loci ranging from 0.97 [SGI versus SGIII] to 0.93 [SGII versus SGIII]; 112 putative adaptive loci ranging from 0.83 [SGI versus SGIII] to 0.78 [SGII versus SGIII]; Table 2), particularly for $PIP_{0.99}$ SNPs (ranging from 0.74 [SGII versus SGIII] to 0.66 [SGI versus SGII]; Table 2).

To better understand SG differentiation, the allele frequency of the total 112 SNP loci under putative selection was analyzed (Supplementary Table S7). The reference allele was assumed as the allele that is most represented in the total dataset of 99 *H. vastatrix* isolates and the alternative allele was the less represented. In SGI, the reference allele was fixed for 73 SNP loci and was exclusive of the group for 18 SNP loci. The alternative allele was found to be exclusive of this group for 10 loci but at very low frequencies for the majority. On the contrary, for SGs II and III, the alternative allele was fixed in 42 and 46 loci whereas the reference allele was fixed in 37 and 39 loci, respectively. Exclusive alternative alleles were found in SGII and SGIII, respectively, but none of the groups had the reference allele restricted. SGI had SNPs with both the alternative and reference alleles exclusive of the group. In total, 28 exclusive alleles were found in SGI, 24 in SGII, and 25 in SGIII, and these alleles can distinguish the SGs completely (Fig. 5). Focusing on the 29 common SNPs, it was possible to observe that, in 28 SNPs, the reference allele was fixed in SGI. SGII and SGIII had either the reference or the alternative allele fixed (Table 3). Moreover, it is in these set of loci that the higher proportion of exclusive alleles was found (Table 3).

Functional annotation of SNP loci putatively under selection. The 112 different consensus sequences of loci identified as possibly being under selection and associated with C3 SGs were queried against the NCBI nucleotide and protein databases. Only one significant nucleotide hit was obtained (locus_246084), matching a retrotransposon Gypsy-like sequence of *H. vastatrix*, (e-value threshold of 1e-30; Supplementary Table S8A). In addition, 22 loci (e-value < 1e-2), corresponding to 19.64% of the total putative adaptive loci, matched to hypothetical/uncharacterized proteins in the NCBI protein database (Supplementary Table S8A). Functional annotation using the UniProt database assigned GO terms to 16.96% of these loci. Within the biological process category, these loci are all described as genes involved in "DNA integration." For the molecular function, the most represented categories are "aspartic-type endopeptidase activity" and "RNA binding." A

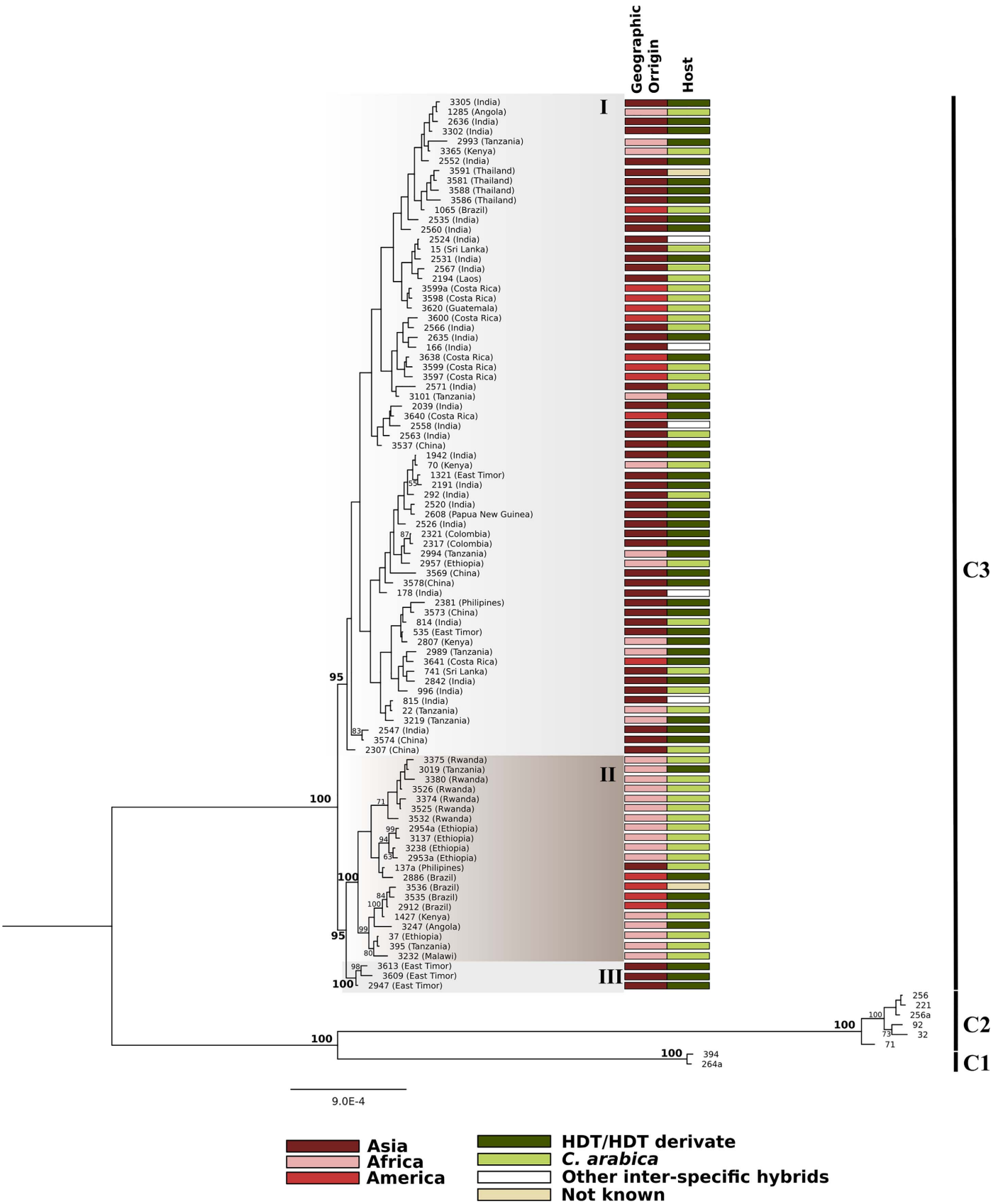

Fig. 1. Phylogenetic relationships among 99 isolates of *Hemileia vastatrix* using 13,804 single nucleotide polymorphisms under a maximum likelihood model. Bootstrap values >70 are provided above the branches. For each isolate, color-coded information about its geographical origin and coffee host is provided. HDT, Híbrido de Timor (Timor hybrid).

Fisher's exact test (false discovery rate <0.05) between All_99indv and OnlyC3indv_adaptive_loci datasets revealed no significant differences in the number of loci annotated into several GO terms. Within these functional classes, there are several genes encoding integrases, reverse transcriptases, RNA-directed DNA polymerase, and RNase H (Supplementary Table S8C), which might be related to the structural composition of retrotransposons, corresponding to 95.45% of the annotated loci.

The OnlyC3indv_adaptive_loci sequences (112 SNP loci) were also mapped against the draft genome and EST data of *H. vastatrix*. From these, 99 SNP loci (88.40%) successfully aligned with the *H. vastatrix* genome (e-value threshold of 1e-15; Supplementary Table S9). No significant differences between All_99indv and OnlyC3indv_adaptive_loci datasets were found regarding the number of loci alignments with the genome (Fisher's exact test: false discovery rate <0.05). Moreover, 11 loci, corresponding to 9.821% of the total putatively adaptive loci, aligned to *H. vastatrix* transcript sequences (Supplementary Table S10).

The potential virulence role of the loci found under positive selection was searched on the pathogen–host interaction database (PHI-base) by BLASTing the RAD loci sequences of *H. vastatrix* containing the 112 adaptive SNPs. A total of 15 *H. vastatrix* RAD loci had homology in the PHI-base to genes reported as showing a relevant role in fungal pathogenicity and virulence when a mutant phenotype was produced in other host–pathogen interactions (Supplementary Table S11). The majority of these genes belonged to the category of "reduced virulence" and "unaffected pathogenicity" in the PHI-base, but two genes were referenced as associated with "loss of pathogenicity" and one to "increased virulence (hypervirulence)" (Supplementary Table S11).

TABLE 2. Pairwise F_{ST} estimates of *Hemileia vastatrix* lineages and among C3 lineage subgroups for each dataset analyzed[a]

	Lineage/subgroup			
	C1	C2	C3	Dataset
C1	–			All_99indv
C2	0.6537	–		
C3	0.8764	0.8935	–	
	SG I	SG II	SG III	
SG I	–			OnlyC3indv
SG II	0.1187	–		
SG III	0.0867	0.0770	–	
	SG I	SG II	SG III	
SG I	–			OnlyC3indv_outliers
SG II	0.9598	–		
SG III	0.9698	0.9326	–	
	SG I	SG II	SG III	
SG I	–			Single-SNP_Assoc
SG II	0.9983	–		
SG III	0.9966	0.9969	–	
	SG I	SG II	SG III	
SG I	–			Multi-SNP_Assoc
SG II	0.6573	–		
SG III	0.6971	0.7422	–	
	SG I	SG II	SG III	
SG I	–			OnlyC3indv_adaptive_loci
SG II	0.8107	–		
SG III	0.8317	0.7816	–	
	SG I	SG II	SG III	
SG I	–			OnlyC3indv_neutral_loci
SG II	0.0523	–		
SG III	0.0016	0.0104	–	
	SG I	SG II	SG III	
SG I	–			OnlyC3indv_common_loci
SG II	0.9978	–		
SG III	0.9965	0.9985	–	

[a] SG, subgroup. Datasets as described in Table 1.

DISCUSSION

Worldwide population genetic divergence of *H. vastatrix* on cultivated coffee. To investigate the global genetic divergence of *H. vastatrix* populations, we generated and genotyped RAD tag–derived SNPs for the most extensive sampling of *H. vastatrix* isolates to date regarding geographical distribution and race diversity, based on the historic CIFC's rust collection from cultivated coffees, with focus on those infecting the most economically valuable coffee species (*C. arabica* and interspecific hybrids).

In this work, phylogenetic and clustering analyses of 99 *H. vastatrix* isolates confirmed the existence of the three well-diverged evolutionary lineages, highly correlated with coffee hosts (C1, C2, C3), found by Silva et al. (2018). The clear phylogenetic segregation between isolates infecting diploid coffee species (C1 and C2 lineages from *C. canephora*, *C. excelsa*, *C. liberica*, and *C. racemosa*) and isolates infecting tetraploid coffee species (C3 lineage from *C. arabica* and interspecific hybrids) described by Silva et al. (2018) is thus corroborated by our data. However, our scaled-up SNP assessment further allowed for the first time the identification of three well-supported and differentiated SGs (I, II, and III) within the *H. vastatrix* C3 lineage. Signals of a shallow structuring within C3 were detected by Silva et al. (2018), but the increased number of variable loci analyzed in our dataset (11,118 versus 6,783) was essential to uncovering a clear pattern of differentiation. The three diverged C3 SGs have some biogeographical alignment, but they appear primarily to reflect the historical distribution and exchange of coffee cultivars and the extent of coffee breeding strategies and pedigree. Despite its African origin on the Ethiopian plateau, *C. arabica* cultivation for commercial purposes may have begun on the Arabian Peninsula (Yemen). The early diffusion of coffee crops was slow, and, after several centuries, it spread from there to Asia (India, Sri Lanka, and Indonesia), later reaching other parts of Africa, and finally being introduced in the Americas, in Brazil (Ferreira et al. 2019). During this period, an intense trade of cultivars among the coffee production areas occurred, with a vast and uncontrolled circulation and exchange worldwide. At the same time, the CLR pathogen may have followed these commercial routes, having colonized all continents (McCook 2006). The close relationship found among geographically distant *H. vastatrix* isolates could be explained by high gene flow, resulting in part of spore dispersal facilitated by human transport, as a consequence of the exchange of infected coffee materials among coffee production regions worldwide, namely from Asia to Africa and America and from Africa to Asia and America (McCook 2006). Furthermore, this may have been combined with the high vagility of *H. vastatrix* spores, which makes it possible for this species to travel long distances mediated by wind currents (McCook and Vandermeer 2015; Talhinhas et al. 2017). In fact, one hypothesis for the introduction of rust in Brazil in 1970 is that wind currents could have carried rust spores across the Atlantic from Ivory Coast or Angola (McCook 2006). Although it seems more feasible that the rust could have arrived in a shipment from Africa, the close connection between coffee rust from Brazil and Africa is portrayed in our data by a well-supported group within SGII, clustering the majority of Brazilian isolates with other two isolates from Kenya and Angola, respectively. Nevertheless, the pattern of isolate clustering in our SGs, the existence of an SG restricted to Timor (SGIII), and their different levels of differentiation suggest additional factors shaping structure.

In our analyses, SGI appears as the largest dispersed *H. vastatrix* genetic group, showing low differentiation and exhibiting a

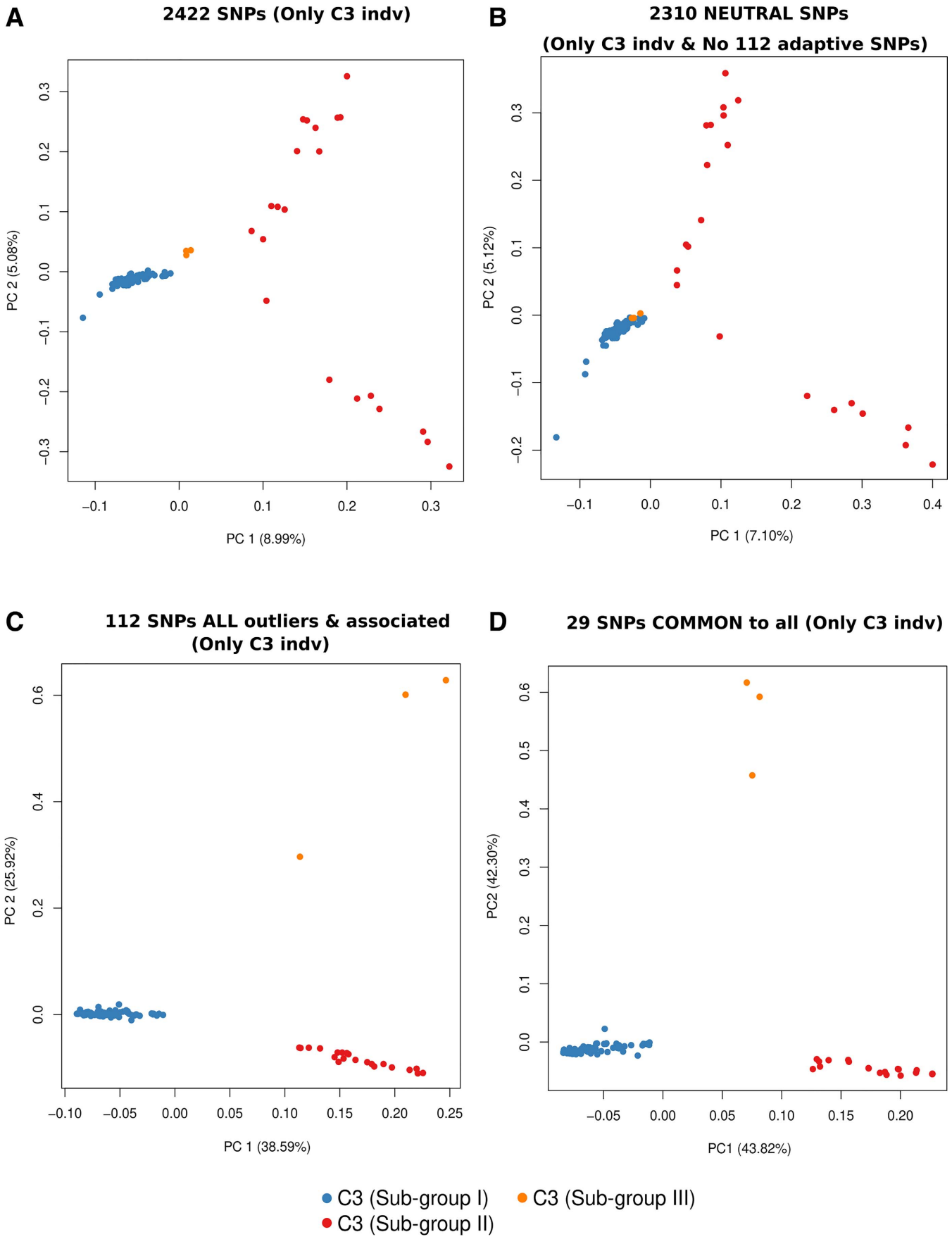

Fig. 2. Principal component analysis of genomic diversity within *Hemileia vastatrix* considering the different datasets assembled. **A,** OnlyC3indv, including only variable single nucleotide polymorphisms (SNPs) within the 91 isolates from the C3 group. **B,** OnlyC3indv_neutral_loci, excluding all putatively adaptive SNPs detected by all approaches. **C,** OnlyC3indv_adaptive_loci, including all putatively adaptive SNPs detected by all approaches. **D,** OnlyC3indv_common_loci, including only putatively adaptive SNPs that were commonly detected in all approaches. The percentage of variation explained by each principal component is provided in their respective label. Isolates are colored according to their assignment to the three C3 subgroups as provided in the legend.

ladder-like diversification pattern, which suggests rapid evolution and population expansion. Intensive selection, severe bottleneck events, and cultivation of few selected phenotypes narrowed down dramatically the genetic variability of domesticated Arabica coffee (Anthony et al. 2002). Such homogenization of cultivated varieties that are densely and globally distributed provides a highly conducive environment favoring pathogen adaptation and dispersion (Möller and Stukenbrock 2017). South Asia was the first region where coffee rust was introduced from eastern Africa, and breeding for resistance began as early as 1911, in India, leading to the release of the resistant cultivar Kent's, which was made available worldwide (Talhinhas et al. 2017). An explosive production of rust-resistant germplasm followed in many other countries, including highly effective sources of rust resistance based on the natural hybrid *C. arabica* × *C. canephora* found in Timor (Híbrido de Timor) broadly supplied by CIFC (Talhinhas et al. 2017). The pattern observed in SGI seems to link with the array of many segregated coffee materials of common origin, bringing in close phylogenetic connection *H. vastatrix* isolates from distant geographical locations. For this to happen, it would imply that infected coffee materials were exchanged in the past or that events of convergent evolution occurred. As undeniable as the occurrence of gene flow may be, there is no evidence of local *H. vastatrix* lineage admixture. Moreover, as a result of intensive breeding, in many countries, particularly India, new or experimental resistant coffee genotypes are continually deployed in field in massive amounts and in close range, exposing the pathogen to different assortments of resistance genes at the same time (Várzea and Marques 2005), which could have also contributed to the observed diversification pattern. In contrast, in Africa, not only are more conservative breeding procedures generally applied, but they also more regularly resort to wild coffee germplasm (McCook 2006). As rust occurs frequently among plants in forests across the native range of *C. arabica* in Ethiopia (Samnegård et al. 2014), it is reasonable to assume that wild rust populations could be a source of genetic variability by the inadvertent use of infected wild germplasm or by naturally spreading into coffee farms. These circumstances could have contributed to the existence of SGII as a specific genetic lineage in this region, revealing a higher degree of genetic differentiation with several small but well-supported groups of isolates. Similarly, SGIII lineage occurring only in east Timor, and represented by three single *H. vastatrix* isolates, also exhibits a high genetic diversity, which, together with the fact that these isolates present the highest level of virulence encountered so far, suggests that this group may have originated from adaptation to new host genotypes. This region is the cradle of the historic spontaneous hybrid Híbrido de Timor, and since its discovery in 1927, the derived populations have freely crossed and dispersed throughout the field (Talhinhas et al. 2017). In these unmanaged and wild coffee farms, it is not surprising that new genotypes with complex resistance gene assortments could have arisen, giving the opportunity for the development of highly differentiated and virulent rust races.

Evidence of local adaption mediating *H. vastatrix* C3 group differentiation. To better understand the complex structuring pattern of *H. vastatrix* populations, we searched for signatures of selection across the genome. Neutral and adaptive driving forces are usually difficult to disentangle, especially within complicated demographic histories (Shen et al. 2019). Conceptually different approaches were thus used to increase the chance of differentiating footprints of local adaptation from those of neutral evolutionary processes. Outlier and association analyses identified a total of 112 SNP loci (4.62% of the total loci) potentially under the effect of natural selection and associated with C3 *H. vastatrix* SGs. Within these candidates for local adaptation, 29 overlapping SNP loci, accounting for 1.20% of the total analyzed loci, were consistently significant in all analyses, which offer the greatest confidence on the signal detected. Interestingly, this small set of SNP loci showed to have a very strong contribution in shaping the divergence pattern within the C3 *H. vastatrix* lineage, as supported by our phylogenetic, PCA, and F_{ST} analyses. When the putatively adaptive loci (112 SNPs) are removed from the analyses, the overall structure within C3 disappears, with the exception of a shallow cluster containing the isolates of SGII, mainly of African origin. These results suggest that, in specific environments, neutral forces might be acting in parallel with adaptive traits, which could be expected under a closer interaction between coffee agricultural and wild environments as it occurs in Africa. Furthermore, pairwise F_{ST} estimates are very close to zero when the 112 adaptive loci are excluded, indicating that those "neutral" regions of the genome might not be indeed under the effect of selection and thus not probably contributing in general to host adaptation. On the contrary, our analysis using only the 112 total putative adaptive SNPs or only the 29 common SNPs recovered the clear division of the isolates into the three SGs correlated with the cultivated coffee host distribution and breeding. Our data reinforce previous assumptions that *H. vastatrix*' evolutionary dynamics are strongly shaped by the coffee host, acting as a major selective pressure (Silva et al. 2018; Várzea and Marques 2005). Here, diversifying selection could be operating in favor of different *H. vastatrix* genetic groups as a consequence of local adaptation. Moreover, the strength of selection in the regions of the *H. vastatrix* genome involved in host adaptation seems to be strong enough to overcome the apparent panmixia/general unstructured pattern of the remaining regions of *H. vastatrix* genome, and shape by itself the overall structure.

On the contrary, curiously, SGIII appears as a sister clade of SGII in phylogenetic analyses for Phylo_All_99indv, Phylo_adaptive_loci, and Phylo_common_loci datasets, but the sister status is not

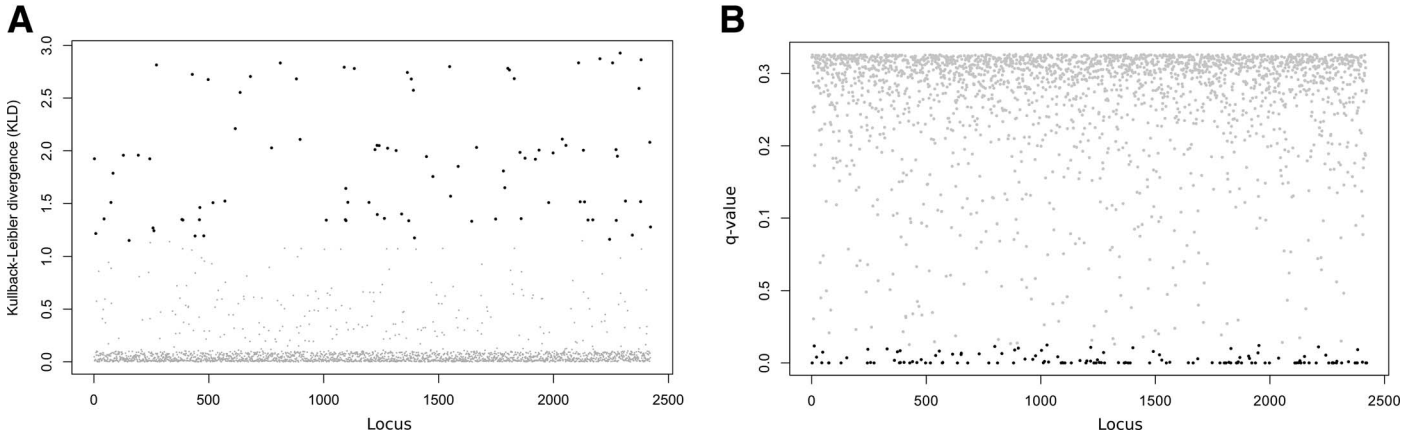

Fig. 3. Single nucleotide polymorphism (SNP) detection under F_{ST} outlier analyses. **A,** Scatter plot with Kullback-Leibler divergence (KLD) for each SNP in SelEstim analysis. Light gray dots indicate SNPs with KLD <99.99% empirical quantile. Black dots indicate SNPs with a KLD >99.99% empirical quantile. **B,** Scatter plot with q-value for each SNP in BayeScan analysis. Black dots indicate SNPs with a q-value <0.05. Dark gray dots indicate SNPs with a q-value >0.05.

maintained when the adaptive loci are removed (Supplementary Fig. S6). The analysis of allelic frequencies of the putatively adaptive loci also revealed that, for some SNPs, the same alleles tend to be fixed in these two SGs. This points to convergent evolution of African and Timor SGs attributable to similar selective pressures acting on these particular regions of the genome, putatively involved in *H. vastatrix* host adaptation. Convergent evolution in different protein families responsible, for example, for pathogenicity, was reported in fungal pathogens as result of host adaptation (Jwa and Hwang 2017; Shang et al. 2016). Meanwhile, multiple directional selections tend to fix specific alleles and can also play important roles in the process of colonization, leading to loss of genetic diversity and increase of population differentiation (Shen et al. 2019). Here, we observed that, for the majority of the SNPs in our candidate loci for local adaptation, C3 SGs tend to have one allele or the other fixed, while, considering these genomic regions, low levels of genetic diversity are found among isolates along with high population genetic differentiation. Nevertheless, as reported in previous studies (Wang et al. 2016), this apparent loss of genetic diversity did not seem to compromise the pathogen's successful adaptation. Probably, the genes associated with these loci are critically important for the species to survive in novel environments/coffee hosts. Moreover, some alleles were found to be exclusive of the SGs, revealing the potential of these SNP loci as candidate genetic markers for distinguishing *H. vastatrix* C3 lineages.

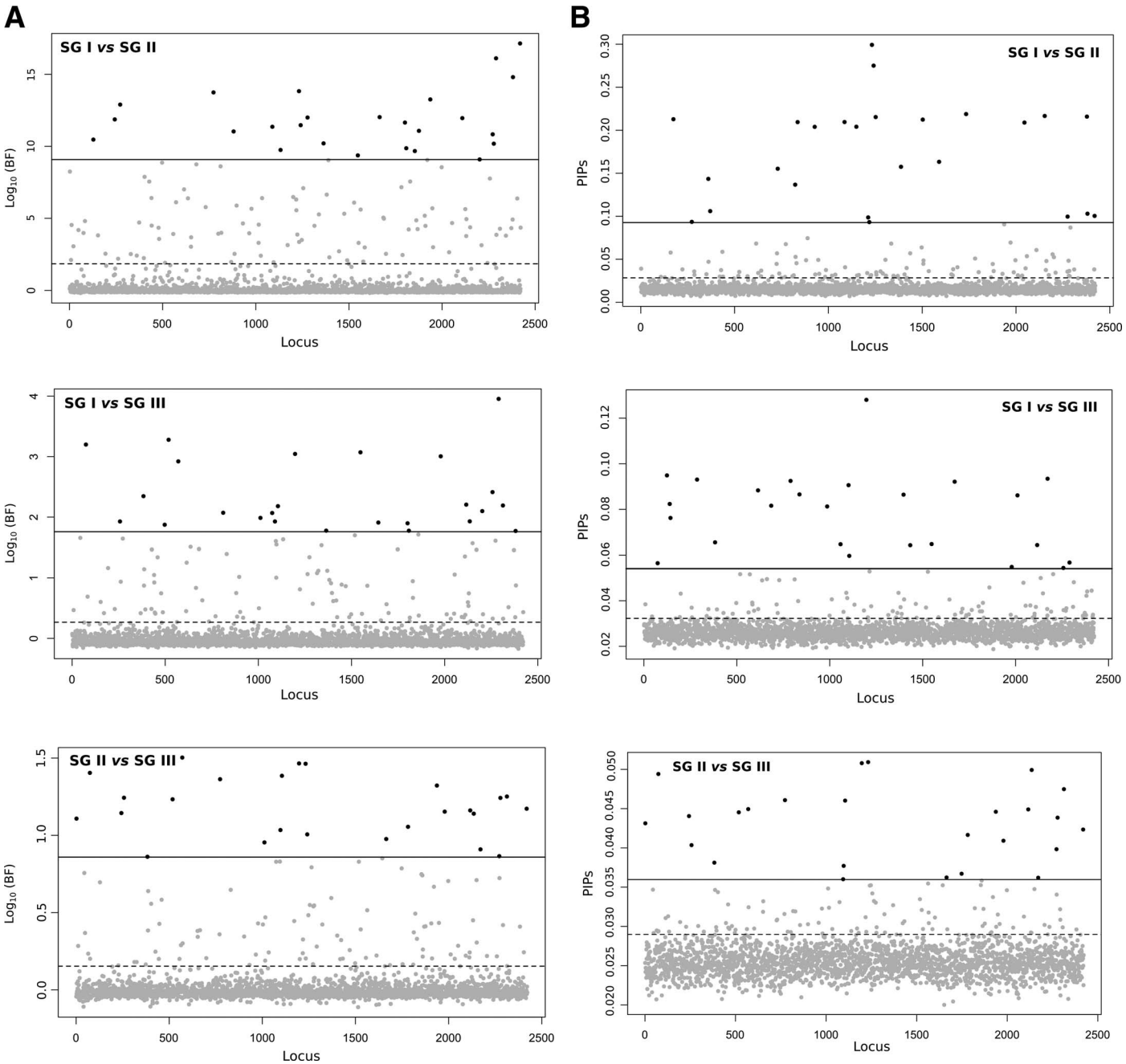

Fig. 4. Single nucleotide polymorphism (SNP) detection under association tests. **A,** Bayes factor (BF) for each SNP in each pairwise comparison in single-SNP association analysis. The horizontal gray dash lines correspond to the BF 95% empirical quantile threshold and the dark straight lines to the 99% empirical quantile. Light gray dots indicate SNPs with a BF <99% empirical quantile. Black dots indicate SNPs with a BF >99% empirical quantile. **B,** Posterior inclusion probabilities (PIPs) for each SNP in each pairwise comparison in multi-SNP association analyses. The horizontal gray dash lines correspond to the PIP 95% empirical quantile threshold and the dark straight lines to the 99% empirical quantile. Light gray dots indicate SNPs with a BF <99% empirical quantile. Black dots indicate SNPs with a BF >99% empirical quantile. SG, subgroup.

Candidate genes putatively involved in *H. vastatrix* adaptation. The identification of the genes associated with the putative adaptive loci was critically impaired by the lack of resources in public databases and the limited annotation of the available *H. vastatrix* draft genome. This is particularly noteworthy because rust species share fewer orthologs with other Basidiomycetes than any other Basidiomycota species between each other, besides exhibiting lineage-specific genes (Silva et al. 2015). Therefore, the content in species-specific genes and/or the low degree of similarity of *H. vastatrix* sequences to other available Pucciniales (rust fungi) sequences could explain the low number of significant protein matches found. Alternatively, our results may suggest that most

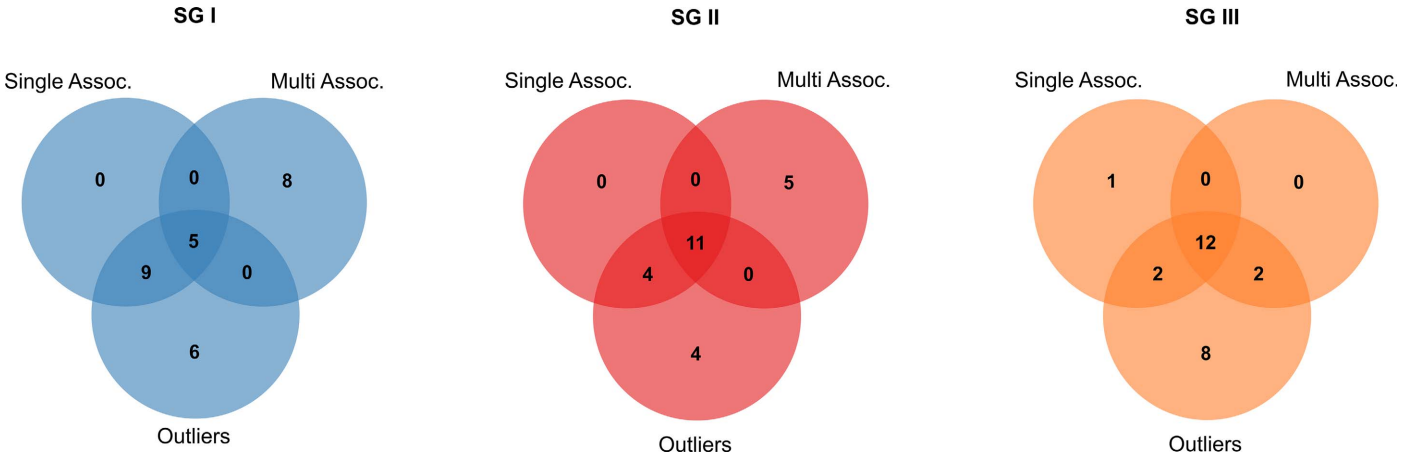

Fig. 5. Venn diagrams showing the distribution of exclusive alleles per analysis (outliers: SelEstim and BayeScan; single association: single-SNP association; and multiassociation: multi-SNP association) and within each C3 subgroup.

TABLE 3. List of 29 putatively adaptive SNP loci common to all analyses (outlier: SelEstim and BayeScan; association: single-SNP and multi-SNP)

Locus ID:SNP position	ALT[a]	REF	ALT frequency			Exclusive alleles	Fixed alleles	Protein description	PHI-base hit
			SGI	SGII	SGIII				
Locus_99:47	T	C	0	1[b]	0	ALT in SGII	ALT in SGII	No hits	–
Locus_6951:2	T	G	0	0	1[b]	ALT in SGIII	ALT in SGIII	No hits	–
Locus_21084:27	T	C	0	1[b]	0	ALT in SGII	ALT in SGII	No hits	–
Locus_22012:43	T	G	0	0	0.833[b]	ALT in SGIII	No	No hits	–
Locus_24931:32	C	T	0[c]	1	1	REF in SGI	REF in SGI	No hits	GCD7
Locus_41884:35	T	C	0	0	1[b]	ALT in SGIII	ALT in SGIII	No hits	yap2
Locus_59476:21	T	G	0	0	1[b]	ALT in SGIII	ALT in SGIII	No hits	–
Locus_65569:15	T	A	0	0	1[b]	ALT in SGIII	ALT in SGIII	Reverse transcription	–
Locus_83926:70	T	G	0	1[b]	0	ALT in SGII	ALT in SGII	Integrase catalytic domain-containing protein	–
Locus_138926:4	G	C	0	0	1[b]	ALT in SGIII	ALT in SGIII	No hits	–
Locus_141164:81	A	G	0	0	1[b]	ALT in SGIII	ALT in SGIII	No hits	–
Locus_150179:55	T	A	0	0	1[b]	ALT in SGIII	ALT in SGIII	No hits	–
Locus_155272:53	A	G	0	1[b]	0	ALT in SGII	ALT in SGII	No hits	–
Locus_156199:52	G	A	0	1[b]	0	ALT in SGII	ALT in SGII	No hits	–
Locus_190529:34	C	T	0[c]	1	1	REF in SGI	REF in SGI	No hits	–
Locus_198109:57	G	A	0	1[b]	0	ALT in SGII	ALT in SGII	No hits	–
Locus_206973:86	G	C	0	1[b]	0	ALT in SGII	ALT in SGII	No hits	–
Locus_217253:32	T	C	0	1[b]	0	ALT in SGII	ALT in SGII	No hits	–
Locus_219469:22	A	C	0	0	1[b]	ALT in SGIII	ALT in SGIII	No hits	–
Locus_237535:74	A	G	0.0238	0	1	No	No	No hits	–
Locus_240750:11	C	T	0	0	1[b]	ALT in SGIII	ALT in SGIII	No hits	–
Locus_242889:79	T	G	0	0	1[b]	ALT in SGIII	ALT in SGIII	Integrase catalytic domain-containing protein	glbC
Locus_254336:14	A	T	0[c]	1	1	REF in SGI	REF in SGI	RNA-directed DNA polymerase	–
Locus_255705:70	A	C	0	1[b]	0	ALT in SGII	ALT in SGII	No hits	–
Locus_255856:47	A	T	0	1[b]	0	ALT in SGII	ALT in SGII	No hits	–
Locus_256299:26	G	T	0[c]	1	1	REF in SGI	REF in SGI	No hits	–
Locus_258284:25	C	G	0	0	1[b]	ALT in SGIII	ALT in SGIII	No hits	–
Locus_271618:61	A	C	0[c]	1	1	REF in SGI	REF in SGI	No hits	–
Locus_274898:39	A	G	0	1[b]	0	ALT in SGII	ALT in SGII	RT RnaseH 2 domain-containing protein	–
Total exclusive alleles	–	–	5	11	12	–	–	–	–
ALT fixed	–	–	0	16	17	–	–	–	–
REF fixed	–	–	28	13	11	–	–	–	–
Total fixed alleles	–	–	28	29	28	–	–	–	–

[a] ALT, alternative allele; REF, reference allele; SG, subgroup; SNP, single nucleotide polymorphism.
[b] ALT exclusive.
[c] REF exclusive.

SNPs are not located in coding regions, but rather in genomic regions with a regulatory role. Even though approximately 19.64% of the 112 loci had hits with protein sequences, indicating that those loci are in coding regions. Moreover, 9.821% of them aligned with *H. vastatrix* EST data from Talhinhas et al. (2014), supporting their location in genes expressed during the early stages of infection. Functional annotation of the small proportion of loci containing SNPs putatively involved in *H. vastatrix* host adaptation (17%) allowed the global identification of candidate genes involved in processes of "DNA integration" that encode proteins with predicted "aspartic-type endopeptidase activity" and "RNA binding" or related functions. A similar proportion of loci annotated into several GO term categories was found between OnlyC3indv_adaptive_loci and All_99indv datasets, suggesting that there is no GO enrichment/depletion of putative adaptive loci in relation to the total number of loci. Nevertheless, the homogeneous prevalence of these functional categories in our data suggest they may play a relevant role in driving *H. vastatrix* host adaptation. It is remarkable that most of the annotated loci are orthologs to genes encoding predicted proteins that include typical functional domains or are components of the transposable elements structure, namely of the retrotransposon structure (Integrase catalytic domain-containing proteins, reverse transcriptases, RNA-directed DNA polymerase, and RT RnaseH 2 domain-containing proteins). In fungal genomes, the percentage of transposable elements is variable, with values ranging, for example, between 3% in the yeast *Saccharomyces cerevisiae* and 90% in the soybean pathogen *Phakopsora pachyrhizi* (see Lorrain et al. 2021 and references therein). In *H. vastatrix*, the first-draft genome assembly showed that repetitive elements represent nearly 75% of the total genome (Cristancho et al. 2014). Our results suggest a possible highly relevant role of transposable elements in *H. vastatrix* lineage differentiation and consequent local host adaptation/specialization. Several studies have shown the importance of transposable element activity in the virulence of several fungal pathogens (Lorrain et al. 2021), leading to gene and transcription modifications when inserted into coding and regulatory regions, as well as to the modulation of genome architecture and plasticity (Lorrain et al. 2021). For instance, genes linked to host specialization were found to be located in transposable element–rich genomic regions in the blast fungus *Magnaporthe oryzae* (Yoshida et al. 2016) and in the *Ceratocystis* genus (Fourie et al. 2020).

Identifying specific genes or alleles underlying adaptive phenotypes, namely host specialization, pathogenicity, virulence, and fungicide resistance (Grünwald et al. 2016), is important to the development of control strategies. Therefore, we investigated the potential virulence role of the candidate loci for host adaptation in *H. vastatrix*. Fifteen genes (13.39%) showed a relevant role in fungal pathogenicity and virulence when a mutant phenotype was produced in other host–pathogen interactions. For instance, locus12620 was found to be an ortholog to the *MgRho3* gene that encodes a protein required for the pathogenicity of *M. oryzae* in rice (*Oryza sativa*) (Zheng et al. 2007). Locus68839 and locus253132 were shown to be orthologs to genes *MoDeam* and *JmjC* domain-containing protein 5 (GH10 family), respectively, also important in the *M. oryzae–O. sativa* interaction. *MoDeam* gene encodes a GlcN-6-phosphate deaminase that is part of the N-acetylglucosamine (GlcNAc) catabolic pathway, and important for successful host colonization (Kumar et al. 2016). JmjC domain-containing protein 5 encodes a xylanase that plays a role in vertical penetration and horizontal expansion of *M. oryzae* in infected plants (Nguyen et al. 2011). Locus208173 was found to be an ortholog to the *Colletotrichum graminicola SID1* gene. This gene encodes a l-ornithine-N^5-monooxygenase protein known to be involved in siderophore biosynthesis in *C. graminicola*, and hence in modulation of the *Zea mays* (maize) immune system (Albarouki et al. 2014). These candidate genes may also be driving the successful infection of *H. vastatrix*, but how they play an important role in host adaptation remains to further study by functional analyses.

Conclusion. Using a population genomics approach, our study provides new insights on the evolutionary history of *H. vastatrix* by revealing, for the first time, a significant population subdivision within the *H. vastatrix* group infecting *C. arabica* and interspecific hybrids with three divergent genetic lineages (I, II, and III). It seems clear that the *H. vastatrix* lineage structuring is in direct connection with the host, reflecting a possible combination of factors mainly associated with the historical distribution and exchange of coffee materials, breeding strategies, and local management that possibly contributed to its divergence. Evidence of footprints of natural selection showed that local adaptation has likely played a crucial role in shaping the contemporary population structure of *H. vastatrix*. Population subdivision is explained by a small set of SNP loci putatively under the effect of positive selection, thus following an adaptive pattern with coffee hosts acting as a major selective pressure. Moreover, host adaptation might be linked to transposable element activity, although the underlying processes remain to be elucidated. Our study showed the high adaptive potential of *H. vastatrix* and how the close interactions with its host can model the population genetic structure, reinforcing the importance this may have from an epidemiological point of view. Taking this into account, additional measures may have to be considered for the efficient and sustainable management of CLR disease, including the design of selective breeding strategies.

ACKNOWLEDGMENTS

We thank Ana Paula Pereira and the technical staff from CIFC/Instituto Superior de Agronomia, namely Célia Lopes, Idalina Gomes, and Miguel Ribeiro, for the support provided on isolate multiplication and pathotype testing, as well as for the maintenance of coffee plants and preservation of CIFC rust collection.

LITERATURE CITED

Albarouki, E., Schafferer, L., Ye, F., Wirén, N., Haas, H., and Deising, H. B. 2014. Biotrophy-specific downregulation of siderophore biosynthesis in *Colletotrichum graminicola* is required for modulation of immune responses of maize. Mol. Microbiol. 92:338-355.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Anthony, F., Combes, M., Astorga, C., Bertrand, B., Graziosi, G., and Lashermes, P. 2002. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theor. Appl. Genet. 104:894-900.

Cabral, P. G. C., Maciel-Zambolim, E., Oliveira, S. A. S., Caixeta, E. T., and Zambolim, L. 2016. Genetic diversity and structure of *Hemileia vastatrix* populations on *Coffea* spp. Plant Pathol. 65:196-204.

Cabreros, I., and Storey, J. D. 2019. A likelihood-free estimator of population structure bridging admixture models and principal components analysis. Genetics 212:1009-1029.

Cenci, A., Combes, M.-C., and Lashermes, P. 2012. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. Plant Mol. Biol. 78:135-145.

Clarindo, W. R., and Carvalho, C. R. 2009. Comparison of the *Coffea canephora* and *C. arabica* karyotype based on chromosomal DNA content. Plant Cell Rep. 28:73-81.

Corredor-Moreno, P., and Saunders, D. G. O. 2020. Expecting the unexpected: Factors influencing the emergence of fungal and oomycete plant pathogens. New Phytol. 225:118-125.

Cristancho, M. A., Botero-Rozo, D. O., Giraldo, W., Tabima, J., Riaño-Pachón, D. M., Escobar, C., Rozo, Y., Rivera, L. F., Durán, A., Restrepo, S., Eilam, T., Anikster, Y., and Gaitán, A. L. 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. 5:1-11.

d'Oliveira, B. 1954. As ferrugens do cafeeiro. Rev. Café Port. 1:5-13.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Earl, D. A., and vonHoldt, B. M. 2012. Structure harvester: A website and program for visualizing structure output and implementing the Evanno method. Conserv. Genet. Resour. 4:359-361.

Eaton, D. A. R., and Overcast, I. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36:2592-2594.
Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. Mol. Ecol. 14:2611-2620.
Falush, D., Stephens, M., and Pritchard, J. K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567-1587.
Ferreira, T., Shuler, J., Guimarães, R., and Farah, A. 2019. Chapter 1. Introduction to coffee plant and genetics. Pages 1-25 in: Coffee: Production, Quality and Chemistry. A. Farah, ed. Royal Society of Chemistry, Cambridge, U.K.
Figueroa, M., Dodds, P. N., and Henningsen, E. C. 2020. Evolution of virulence in rust fungi — multiple solutions to one problem. Curr. Opin. Plant Biol. 56:20-27.
Foll, M., and Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977-993.
Fones, H. N., Bebber, D. P., Chaloner, T. M., Kay, W. T., Steinberg, G., and Gurr, S. J. 2020. Threats to global food security from emerging fungal and oomycete crop pathogens. Nat. Food 1:332-342.
Fourie, A., de Jonge, R., van der Nest, M. A., Duong, T. A., Wingfield, M. J., Wingfield, B. D., and Barnes, I. 2020. Genome comparisons suggest an association between *Ceratocystis* host adaptations and effector clusters in unique transposable element families. Fungal Genet. Biol. 143:103433.
Gautier, M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics 201:1555-1579.
Gichuru, E. K., Ithiru, J. M., Silva, M. C., Pereira, A. P., and Varzea, V. M. P. 2012. Additional physiological races of coffee leaf rust (*Hemileia vastatrix*) identified in Kenya. Trop. Plant Pathol. 37:424-427.
Gouveia, M. M. C., Ribeiro, A., Várzea, V. M. P., and Rodrigues, C. J. 2005. Genetic diversity in *Hemileia vastatrix* based on RAPD markers. Mycologia 97:396-404.
Grünwald, N. J., McDonald, B. A., and Milgroom, M. G. 2016. Population genomics of fungal and oomycete pathogens. Annu. Rev. Phytopathol. 54: 323-346.
Guan, Y., and Stephens, M. 2011. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. Ann. Appl. Stat. 5:1780-1815.
International Coffee Organization. 2020. World coffee production. http://www.ico.org/prices/po-production.pdf
Jwa, N.-S., and Hwang, B. K. 2017. Convergent evolution of pathogen effectors toward reactive oxygen species signaling networks in plants. Front. Plant Sci. 8:1687.
Kolmer, J. A., Liu, J. Q., and Sies, M. 1995. Virulence and molecular polymorphism in *Puccinia recondita* f. sp. *tritici* in Canada. Phytopathology 85:276-285.
Kumar, A., Ghosh, S., Bhatt, D. N., Narula, A., and Datta, A. 2016. *Magnaporthe oryzae* aminosugar metabolism is essential for successful host colonization. Environ. Microbiol. 18:1063-1077.
Lashermes, P., Combes, M. C., Topart, P., Graziosi, G., Bertrand, B., and Anthony, F. 2000. Molecular breeding in coffee (*Coffea arabica* L.). Pages 101-112 in: Coffee Biotechnology and Quality. T. Sera, C. R. Soccol, A. Pandey, and S. Roussos, eds. Springer Netherlands, Dordrecht, The Netherlands.
Lorrain, C., Oggenfuss, U., Croll, D., Duplessis, S., and Stukenbrock, E. 2021. Transposable elements in fungi: Coevolution with the host genome shapes, genome architecture, plasticity and adaptation. Pages 142-155 in: Encyclopedia of Mycology. O. Zaragoza, A. Casdevall, eds. Elsevier, Amsterdam.
Maia, T. A., Maciel-Zambolim, E., Caixeta, E. T., Mizubuti, E. S. G., and Zambolim, L. 2013. The population structure of *Hemileia vastatrix* in Brazil inferred from AFLP. Australas. Plant Pathol. 42:533-542.
McCook, S. 2006. Global rust belt: *Hemileia vastatrix* and the ecological integration of world coffee production since 1850. J. Glob. Hist. 1:177-195.
McCook, S., and Vandermeer, J. 2015. The big rust and the red queen: Long-term perspectives on coffee rust research. Phytopathology 105:1164-1173.
McDonald, B. A., and Stukenbrock, E. H. 2016. Rapid emergence of pathogens in agro-ecosystems: Global threats to agricultural sustainability and food security. Philos. Trans. R. Soc. B Biol. Sci. 371:20160026.
Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., Passarotti, M., Kaufman, S., and O'Leary, M. A. 2015. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. Evol. Bioinform. Online 11:43-48.
Möller, M., and Stukenbrock, E. H. 2017. Evolution and genome architecture in fungal plant pathogens. Nat. Rev. Microbiol. 15:756-771.
Narum, S. R., and Hess, J. E. 2011. Comparison of FST outlier tests for SNP loci under selection. Mol. Ecol. Resour. 11:184-194.
Nguyen, Q. B., Itoh, K., Van Vu, B., Tosa, Y., and Nakayashiki, H. 2011. Simultaneous silencing of endo-β-1,4 xylanase genes reveals their roles in the virulence of *Magnaporthe oryzae.* Mol. Microbiol. 81:1008-1019.
Nunes, C. C., Maffia, L. A., Mizubuti, E. S. G., Brommonschenkel, S. H., and Silva, J. C. 2009. Genetic diversity of populations of *Hemileia vastatrix* from organic and conventional coffee plantations in Brazil. Australas. Plant Pathol. 38:445.
Pereira, D., Croll, D., Brunner, P. C., and McDonald, B. A. 2020. Natural selection drives population divergence for local adaptation in a wheat pathogen. Fungal Genet. Biol. 141:103398.
Pina-Martins, F., Silva, D. N., Fino, J., and Paulo, O. S. 2017. Structure_threader: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. Mol. Ecol. Resour. 17:e268-e274.
Plissonneau, C., Benevenuto, J., Mohd-Assaad, N., Fouché, S., Hartmann, F. E., and Croll, D. 2017. Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. Front. Plant Sci. 8:1-15.
Porto, B. N., Caixeta, E. T., Mathioni, S. M., Vidigal, P. M. P., Zambolim, L., Zambolim, E. M., Donofrio, N., Polson, S. W., Maia, T. A., Chen, C., Adetunji, M., Kingham, B., Dalio, R. J. D., and Resende, M. L. V. 2019. Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression dynamics of candidate effectors dependent on host compatibility. PLoS One 14:e0215598.
Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., and Bender, D. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.
Roshyara, N. R., and Scholz, M. 2014. fcGENE: A versatile tool for processing and transforming SNP datasets. PLoS One 9:e97589.
Rousset, F. 2008. genepop'007: A complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8:103-106.
Rozo, Y., Escobar, C., Gaitán, Á., and Cristancho, M. 2012. Aggressiveness and genetic diversity of *Hemileia vastatrix* during an epidemic in Colombia. J. Phytopathol. 160:732-740.
Samnegård, U., Hambäck, P. A., Nemomissa, S., and Hylander, K. 2014. Local and regional variation in local frequency of multiple coffee pests across a mosaic landscape in *Coffea arabica* 's native range. Biotropica 46:276-284.
Santana, M. F., Zambolim, E. M., Caixeta, E. T., and Zambolim, L. 2018. Population genetic structure of the coffee pathogen *Hemileia vastatrix* in Minas Gerais, Brazil. Trop. Plant Pathol. 43:473-476.
Servin, B., and Stephens, M. 2007. Imputation-based analysis of association studies: Candidate regions and quantitative traits. PLoS Genet. 3:e114.
Shang, Y., Xiao, G., Zheng, P., Cen, K., Zhan, S., and Wang, C. 2016. Divergent and convergent evolution of fungal pathogenicity. Genome Biol. Evol. 8:1374-1387.
Shen, Y., Wang, L., Fu, J., Xu, X., Yue, G. H., and Li, J. 2019. Population structure, demographic history and local adaptation of the grass carp. BMC Genomics 20:467.
Silva, D. N., Duplessis, S., Talhinhas, P., Azinheira, H., Paulo, O. S., and Batista, D. 2015. Genomic patterns of positive selection at the origin of rust fungi. PLoS One 10:e0143959.
Silva, D. N., Várzea, V., Paulo, O. S., and Batista, D. 2018. Population genomic footprints of host adaptation, introgression and recombination in coffee leaf rust. Mol. Plant Pathol. 19:1742-1753.
Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.
Steinegger, M., and Söding, J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. 35:1026-1028.
Stukenbrock, E. H., and McDonald, B. A. 2008. The origins of plant pathogens in agro-ecosystems. Annu. Rev. Phytopathol. 46:75-100.
Talhinhas, P., Azinheira, H. G., Vieira, B., Loureiro, A., Tavares, S., Batista, D., Morin, E., Petitot, A. S., Paulo, O. S., Poulain, J., Da Silva, C., Duplessis, S., Silva Mdo, C., and Fernandez, D. 2014. 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. 5:230-232.
Talhinhas, P., Batista, D., Diniz, I., Vieira, A., Silva, D. N., Loureiro, A., Tavares, S., Pereira, A. P., Azinheira, H. G., Guerra-Guimarães, L., Várzea, V., and Silva, M. D. C. 2017. The coffee leaf rust pathogen *Hemileia vastatrix*: One and a half centuries around the tropics. Mol. Plant Pathol. 18:1039-1051.
Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S. Y., De Silva, N., Martinez, M. C., Pedro, H., Yates, A. D., Hassani-Pak, K., Hammond-Kosack, K. E. 2019. PHI-base: The pathogen–host interactions database. Nucleic Acids Res. 48:D613-D620.
Várzea, V. M. P., and Marques, D. V. 2005. Population variability of *Hemileia vastatrix* versus coffee durable resistance. Pages 53-74 in: Durable

Resistance to Coffee Leaf Rust. L. Zambolim, ed. Universidade Federal de Viçosa, Viçosa, Brazil.
Vitalis, R., Gautier, M., Dawson, K. J., and Beaumont, M. A. 2014. Detecting and measuring selection from gene frequency data. Genetics 196:799-817.
Vries, S., Stukenbrock, E. H., and Rose, L. E. 2020. Rapid evolution in plant–microbe interactions – an evolutionary genomics perspective. New Phytol. 226:1256-1262.
Wang, L., Wan, Z. Y., Lim, H. S., and Yue, G. H. 2016. Genetic variability, local selection and demographic history: Genomic evidence of evolving towards allopatric speciation in Asian seabass. Mol. Ecol. 25:3605-3621.
Yoshida, K., Saunders, D. G. O., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., Inoue, Y., Chuma, I., Tosa, Y., Cano, L. M., Kamoun, S., and Terauchi, R. 2016. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics 17:370.
Zheng, W., Chen, J., Liu, W., Zheng, S., Zhou, J., Lu, G., et al. 2007. A Rho3 homolog is essential for appressorium development and pathogenicity of *Magnaporthe grisea*. Eukaryot. Cell 6:2240-2250.
Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326-3328.