

Diversity in the invasive cacao pathogen *Moniliophthora roreri* is shaped by agriculture

Jorge R. Díaz-Valderrama^{1,2}  | Ricardo Zambrano³ | Sergio Cedeño-Amador³ |
 Ulises Córdova-Bermejo^{4,5} | Gianmarco G. Casas⁶  | Nilda García-Zurita⁷ |
 Juan A. J. Sánchez-Arévalo⁷ | Enrique Arévalo-Gardini^{8,9} | Darío Dávila¹⁰ | Juan Ruiz¹⁰ |
 Xiomay Pinchi-Dávila¹¹ | Zarella R. Quispe-Chacón¹² | Julio A. Chia-Wong¹²  |
 Oscar P. Hurtado-Gonzales¹³ | Carlos A. Rodríguez-Callañaupa¹⁴ | Casto Maldonado-Fuentes¹⁵ |
 Eucebio Pérez-Callizaya¹⁶ | Santos Leiva-Espinoza² | Manuel Oliva-Cruz² |
 Dylan M. Guevara-Ruiz¹⁷ | Esperanza Torres-Rojas¹⁷ | Elizabeth Johnson¹⁸ |
 Wilbert Phillips-Mora¹⁹ | M. Catherine Aime¹

¹Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana, USA

²Instituto de Investigación para el Desarrollo Sustentable de Ceja de Selva, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Amazonas, Peru

³Industrial y Agrícola Cañas C.A., Guayaquil, Ecuador

⁴Centro de Investigación y Promoción del Campesinado, Piura, Peru

⁵Asociación de Pequeños Productores de Cacao de Piura, Piura, Peru

⁶Department of Forest Engineering, Federal University of Viçosa, Viçosa, Minas Gerais, Brazil

⁷Exportadora Romex S.A., Tarapoto, San Martín, Peru

⁸Instituto de Cultivos Tropicales, Tarapoto, San Martín, Peru

⁹Facultad de Agronomía, Universidad Nacional Autónoma de Alto Amazonas, Yurimaguas, Loreto, Peru

¹⁰Herbarium Amazonense, Centro de Investigaciones de Recursos Naturales, Universidad Nacional de la Amazonía Peruana, Iquitos, Peru

¹¹Department of Plant Pathology, University of Georgia, Athens, Georgia, USA

¹²Facultad de Agronomía, Universidad Nacional Agraria de la Selva, Tingo María, Huánuco, Peru

¹³US Department of Agriculture, Animal and Plant Health Inspection Service, Beltsville, Maryland, USA

¹⁴Servicio Nacional de Sanidad Agraria, Quillabamba, Cusco, Peru

¹⁵Facultad de Agronomía, Universidad Mayor de San Andrés, Sapecho, La Paz, Bolivia

¹⁶Cooperativa El Ceibo, Sapecho, La Paz, Bolivia

¹⁷Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá, Colombia

¹⁸Inter-American Institute for Cooperation on Agriculture, Hope Botanical Gardens, Kingston, Jamaica

¹⁹Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba, Costa Rica

Correspondence

M. Catherine Aime, Department of Botany and Plant Pathology, Purdue University, West Lafayette, IND, USA.
 Email: maime@purdue.edu

Funding information

Peru National Fund for Scientific, Technological Development and

Abstract

Frosty pod rot disease of cacao (FPR), caused by the fungus *Moniliophthora roreri*, has severely impacted the production of cocoa in Latin America since its discovery. Prior to the 1950s, FPR was known only from Colombia and Ecuador. However, beginning in the 1970s, its geographical range has dramatically expanded throughout most of the

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Plant Pathology* published by John Wiley & Sons Ltd on behalf of British Society for Plant Pathology.

Technological Innovation, Grant/Award Number: 030-2019-FONDECYT-BM-INC. INV; U.S. Department of Agriculture, Grant/Award Number: Hatch 1010662; Purdue University; Golden Key International Honour Society; Mycological Society of America

cacao-producing regions of the Americas. The origin of the pathogen remains unknown. In this study, we evaluated the genetic diversity of *M. roreri* from areas spanning, as much as possible, its current geographical range using simple-sequence repeat markers and a publicly available single-nucleotide polymorphism data set. Two hotspots of genetic diversity were found: coastal Ecuador and the inter-Andean Magdalena Valley of Colombia, neither of which correspond to the Amazonian origin of the host. However, both areas were early centres of intense cultivation of cacao. Our results indicate that *M. roreri* was introduced into both areas from its centre of origin, where intensive cacao cultivation probably led to the increase of inoculum and further dissemination of the disease. Current invasions can be traced to two genotypes responsible for all known instances of the pathogen in Central America, the Caribbean, Peru and Bolivia. We also report for the first time *M. roreri* in Maynas (Peruvian Amazon), which is probably the result of a recent introduction from Colombia.

KEYWORDS

airborne fungi, Basidiomycota, dissemination of pathogens, Marasmiaceae, Moniliasis, tropical phytopathogens

1 | INTRODUCTION

Agriculture is a major force for the movement of plant species and consequently of microorganisms associated with them, such as pathogens. The low genetic diversity background of cultivated species tends to facilitate the spread of pathogens, causing a decline in the productivity of agricultural settings in the invaded areas. Investigating the history of the invasion routes of pathogens from their centres of diversity can offer insights into invasive success and lead to strategies to mitigate their impact (Stukenbrock & McDonald, 2008). Here, we present the case of the tropical crop cacao (*Theobroma cacao*), the raw material of chocolate, and the fungus *Moniliophthora roreri*, one of its major pathogens in Latin America.

Cacao originated in the upper Amazonian region shared by Peru, Ecuador, Colombia and Brazil (Osorio-Guarín et al., 2017; Thomas et al., 2012). The earliest evidence for its use and domestication is in the Mayo-Chinchipe basin in the Upper Amazon of Ecuador and Peru (Olivera-Núñez, 2018; Zarrillo et al., 2018). In other words, the centre of origin of the crop and the earliest known domestication centre of cacao point to the same region in South America. By the time early colonialists arrived in South America, cacao was being cultivated outside its centre of origin by indigenous peoples in coastal Ecuador and the Maracaibo basin of Venezuela, a region that neighbours Colombia to the northwest (Díaz-Valderrama et al., 2020). During the colonial period, the cacao crop underwent expansive dissemination in response to the growing international demand for chocolate, with coastal Ecuador and Venezuela as the most important global suppliers of cacao (Díaz-Valderrama et al., 2020).

Cacao cultivation is limited to the use of selected clone cultivars that have a narrow genetic basis. Clonal propagation of cacao to regions outside its centre of origin has decreased the genetic diversity

of the crop, which has subsequently increased the vulnerability to diseases and makes the development of fully disease-resistant cultivars more challenging (Phillips-Mora et al., 2009). One of these, frosty pod rot (FPR), is among the most important cacao diseases in the Americas. It is caused by the basidiomycete *M. roreri* (Agaricales, Marasmiaceae), one of the most threatening plant pathogens in the world according to Thurston (1973). *M. roreri* is an unusual agaric that lacks the capacity to reproduce sexually, relying on large production of asexual spores (or conidia) for dissemination between cacao fields (Díaz-Valderrama & Aime, 2016a, 2016b). The pathogen infects only the fruits of *Theobroma* and related *Herrania* species. Symptoms include malformations, especially in young fruits, necrosis and eventual mummification of fruits; internally, seeds become rotten, watery and useless for chocolate manufacturing (Aime & Phillips-Mora, 2005). External symptoms take 45–90 days after infection to develop, meaning that infected pods that appear symptomless can be disseminated across large geographical distances by humans (Phillips-Mora et al., 2007). The first documentation and recorded outbreak of the disease occurred in Ecuador in 1895 and the late 1910s, respectively, but reports of unnamed cacao diseases with similar symptomology date to at least as early as 1817 in Norte de Santander, the Colombian side of the Maracaibo basin (Díaz-Valderrama et al., 2020).

Before the 1950s, the known geographical range of *M. roreri* remained constrained to Colombia, Ecuador and western Venezuela (Phillips-Mora & Wilkinson, 2007). Then, beginning in 1956 with the first report in Panama, *M. roreri* disseminated to all cacao-producing countries in Central America (Phillips-Mora, Cawich, et al., 2006; Phillips-Mora, Coutiño, et al., 2006; Phillips-Mora & Wilkinson, 2007), and by 2016, it had expanded eastward into Jamaica (Johnson et al., 2017). The pathogen has also been spreading south, reaching northern Peru

in 1988, southern Peru in 1998, Bolivia in 2012 and, more recently, the western state of Acre in Brazil (Hernández et al., 1990; Jiménez et al., 2021; Phillips-Mora et al., 2015; Ríos-Ruiz & Rodríguez, 1998). Spread of FPR has completely altered global cacao cultivation in the 20th century. Before the outbreak of the disease in Ecuador in the 1910s, the country was a major producer in the world; currently, the total cacao production from West Africa and Southeast Asia—areas still free of FPR—constitutes nearly 85% of the world's production (Díaz-Valderrama et al., 2020).

M. roreri was originally considered to be indigenous to Ecuador (Phillips-Mora et al., 2007), although recent studies have found high levels of genetic diversity for *M. roreri* in Colombia that challenge this view (Ali et al., 2015; Jaimes et al., 2016; Phillips-Mora et al., 2007). However, no study has included extensive sampling from either the upper Amazon, where the crop originated and was first domesticated (Zarrillo et al., 2018), or from coastal Ecuador, where the pathogen was first documented in 1895 (Díaz-Valderrama et al., 2020)—data necessary for a complete analysis of *M. roreri* origin and spread. Regardless, most studies have found that *M. roreri* disseminated throughout Central America after one or very few introductions of the same or nearly identical genotypes from South America (Díaz-Valderrama & Aime, 2016a; Phillips-Mora et al., 2007). A second genotype seems responsible for disease in Peru (Díaz-Valderrama & Aime, 2016a). The recent invasions of *M. roreri* into Jamaica in 2016 and Bolivia in 2012 (Johnson et al., 2017; Phillips-Mora et al., 2015) raise the question of whether these represent further incursions of the same invasive genotypes or represent the evolution of new invasive genotypes.

In this study, we (a) developed a statistically robust set of simple-sequence repeat (SSR) molecular markers for genotyping of *M. roreri* isolates; (b) assembled a comprehensive sampling of *M. roreri* from most of its known range, including new isolates from coastal Ecuador, Amazonian Peru, Bolivia and Jamaica; (c) used population genetics to determine the centre(s) of genetic diversity for this pathogen; and (d) inferred a centre of origin for the pathogen based on the generated SSR data set, sequences of the internal transcribed spacer rDNA regions 1 and 2 and the 5.8S subunit of the ribosomal repeat (ITS) of samples collected, and a publicly available single-nucleotide polymorphism (SNP) data set.

2 | MATERIALS AND METHODS

2.1 | Collection and isolation of *M. roreri* samples

A total of 228 samples of *M. roreri* were collected for this study (Table S1). Isolations took place on the same day of collection. We collected white stroma, if present, from the surface of infected pods. Then, we dissected the pods and collected internal necrotic tissue from pulp and beans with sterile forceps. We placed the white stroma and internal tissue in Eppendorf tubes containing 600 µl of Nuclei Lysis Solution from the Wizard Genomic Purification Kit (Promega) for DNA extraction without the need to perform pure

culture isolation (Figure 1). We also performed isolations in pure culture for a subset of samples. This included surface sterilization of pods in 2.5% sodium hypochlorite solution for 3 min and a rinse in sterile water for 2 min. Pods were dissected, and necrotic internal tissue was placed on potato dextrose agar (PDA) plates. Subculturing was performed until pure cultures were obtained. Samples used in this study came from 207 *M. roreri*-infected pods from 185 trees (Table S1). Two different samples may have come from the same pod, but sampled material always came from different locations within the pod depending on the type of material collected, that is, internal necrotic tissue, white stroma, or pure culture isolation (Table S1); if same-pod samples were from the same type of material, these were taken from different places, for example, seed versus pulp necrotic tissue.

The collections took place in Ecuador, Colombia, Peru, Bolivia, Costa Rica and Jamaica from 2015 to 2017 (Figure 2). We estimated the minimum samples to collect based on the number of cacao farmers in each country (Table S1). Sixteen isolates were from the inter-Andean upper and middle Magdalena valley, Colombia; 46 from Ecuador (Guayas basin, coastal region), 42 of which were collected in a single growing season from an 850-acre cacao farm; 14 from Maynas in the Peruvian Upper Amazon; and 126 across the main cacao-cultivating regions of Peru from north to south (34 samples from Piura, in the western side of the Andes, next to the Pacific coast; and 92 from cacao-producing regions in the eastern side, in the Upper Amazon). Additionally, two isolates from the first reports of *M. roreri* in Mexico and Belize (Phillips-Mora, Cawich, et al., 2006; Phillips-Mora, Coutiño, et al., 2006) were included in the analysis. Samples from Jamaica and Bolivia (three and nine samples, respectively) came from the same areas where the first report of *M. roreri* in these countries occurred (Johnson et al., 2017; Phillips-Mora et al., 2015). Most samples were isolated from *T. cacao*, but some were from other *Theobroma* species: eight from cultivated *T. grandiflorum* from Ecuador and one from wild *T. bicolor* from Maynas, Peru (Table S1). Samples from *T. cacao* came from the three main cacao groups *Trinitario*, *Criollo* and *Forastero*, from Peruvian native varieties and hybrid cultivars (Table S1). Five isolates, including one from Panama, analysed in a previous study (Díaz-Valderrama & Aime, 2016a), were also included in this analysis.

2.2 | Molecular characterization of *M. roreri* samples

Sixteen SSR markers were used to characterize *M. roreri* samples (Table S2). Eleven SSR loci selected had been previously used (Díaz-Valderrama & Aime, 2016a). Additional SSR loci were detected and primers flanking the SSR sequences were designed with the program QDD v. 3.1.2b (Megléczy et al., 2014) by screening the genome of *M. roreri* MCA2952 (Díaz-Valderrama & Aime, 2016a). The SSR marker selection criteria included number of motif repeats, the scaffold size where loci were located and position in the scaffold, to ensure loci were as evenly distributed throughout the *M. roreri*



FIGURE 1 Direct collection method for *Moniliophthora roreri*. (a) Identification of malformation on cacao pods, an early symptom of frosty pod rot disease (FPR). (b) Mummified cacao pods (symptom) with white stroma (sign) in an advanced stage of the disease. (c) Symptoms (brown spots) and signs (white mycelium) of FPR on *Theobroma bicolor*-infected pod. (d) Symptoms (mummification) and signs (white stroma) of FPR on *Theobroma grandiflorum*-infected pod. (e) Use of Eppendorf tube containing nuclei lysis solution (see Materials and Methods) to collect the white stroma on surface of infected pod. (f) Dissection of pod. (g) Collection of necrotic tissue to be placed in a Eppendorf tube with nuclei lysis solution. Scale bars = 5 cm. [Colour figure can be viewed at wileyonlinelibrary.com]

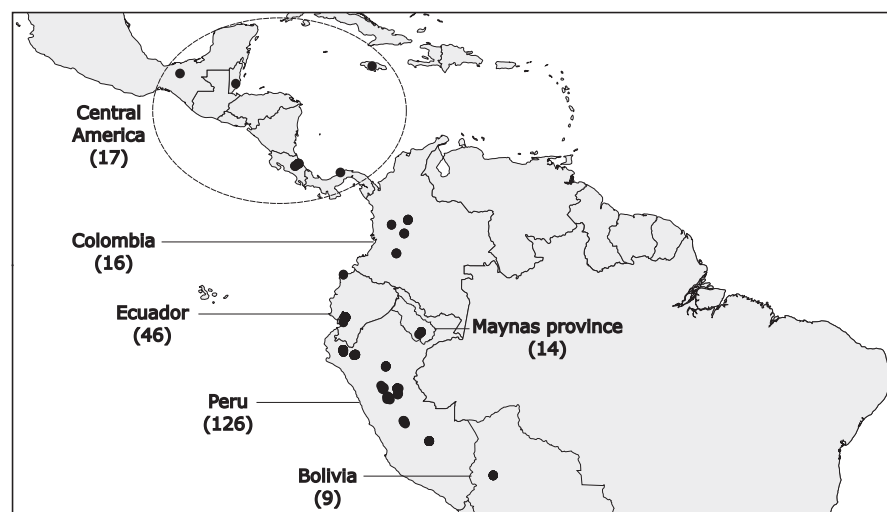


FIGURE 2 Geographical origin of samples collected for this study. The numbers of samples collected in each region are in parentheses.

genome as possible. The genotyping of samples was performed via capillary electrophoresis with a modified M13-tailed primer method as previously described (Díaz-Valderrama & Aime, 2016a). To ensure reproducibility, PCRs on all the markers for a subset of five samples in each region were performed three times. In some instances,

reactions generated null alleles. Null alleles in SSR genotyping are frequent and might have an effect on genetic diversity estimations. In these cases, PCRs were replicated three times with positive controls to verify the legitimacy of the null allele. To ensure specificity, all 16 SSR markers were tested against one isolate of *Moniliophthora*

perniciosa, another major pathogen of cacao and a sister species of *M. roleri* (Aime & Phillips-Mora, 2005), from Maynas. Thirteen markers failed to amplify; only Mr_SSR1, Mr_SSR_38 and Mr_SSR40 were able to cross-amplify *M. perniciosa*, but none of these produced shared alleles between *M. perniciosa* and *M. roleri*. The SSR alleles were scored with Geneious v. 10.0.5. Most of the downstream analyses, unless specified, were done in R v. 3.6.2 via RStudio v. 1.2.1335. Population genetics analyses were performed with R package poppr (Kamvar et al., 2014).

A data set generated from 88 single-nucleotide polymorphisms (SNPs) on 172 *M. roleri* samples from a previous study (Ali et al., 2015) was also analysed to compare with results from our SSR markers. This data set contains 86 samples from Costa Rica, 66 from Colombia, 11 from Ecuador, 6 from Bolivia, 2 from Peru and 1 from Venezuela. The same population genetics analysis pipeline that follows was applied to the SNP data set by replacing nucleotides with numbers (A = 1; C = 2; G = 3; T = 4; a deletion = 5).

2.2.1 | Clone correction and resolution of markers

The resolving power of our markers was evaluated with a genotype accumulation curve analysis, which shows the discriminating capacity of the markers between samples and inspects whether the number of genotypes found increases by adding an extra marker (Arnaud-Haond et al., 2007). These curves were generated by randomly sampling 1000 times from 1 up to $n-1$ loci (n = number of markers used; 16 for SSRs and 88 for SNPs). Then, the number of observed genotypes was counted, and results were box-plotted. Additionally, the sampling was examined with rarefaction curves built with VEGAN R package (Dixon, 2003).

Clone correction is the process of retaining just one copy of each multilocus genotype (MLG) in each location for analysis. However, it is possible that nearly identical MLGs, that is, MLGs that differ in one or very few alleles, may be part of the same multilocus lineage (MLL) and that those allelic differences are due to mutations during mitosis or even scoring errors, and not to sexual recombination events. Because *M. roleri* is a clonally reproducing fungus (Díaz-Valderrama & Aime, 2016a, 2016b), the identification of MLLs was based on the procedure established for population genetics analyses in clonal organisms (Arnaud-Haond et al., 2007). First, we examined the spectrum of genetic diversity by plotting the frequency distribution of the genetic distances among MLGs (Rozenfeld et al., 2007); for this Nei genetic distances were calculated (Kamvar et al., 2014). If at the low distance extreme of the spectrum of genetic diversity there is a small peak, that is, the spectrum of genetic diversity has a bimodal rather than a unimodal distribution (Arnaud-Haond et al., 2007), it is possible that MLGs having genetic distances at that threshold (the small distance peak) or lower could belong to the same MLL. The bimodality of the spectrum was statistically corroborated using the Hartigan's dip test in the R package diptest; a p value less than 0.05 indicates the dip value (D) of the spectrum statistically differs from zero, that is, the spectrum of genetic diversity is at least bimodal

(Hartigan, 1985). Once the low distance peak threshold was identified, we performed a Nei-distance UPGMA dendrogram with bootstrap support (1000 bootstraps). Then, we carefully evaluated whether every MLG in a cluster formed at a distance equal or lower to that threshold belonged to the same MLL. To accomplish this, we identified and temporarily removed the loci with allelic differences among MLGs in each cluster. The probability that the now identical MLGs arose from hypothetical independent sexual reproductive events (p_{sex}) was calculated without the different allele loci; a p_{sex} value less than 0.01 meant that nearly identical MLGs were probably derived from the same hypothetical sexual event; that is, they were part of the same MLL (Arnaud-Haond et al., 2007). The allele variants identified for samples in each MLL were then adjusted by replacing them with the most common allele in the MLL. This resulted in each MLL having a representative genotype that was used in further analyses.

2.2.2 | Genetic structure and spatial analysis

The genetic structure of *M. roleri* samples was analysed with the LEA package (Frichot & François, 2015). LEA implements a non-negative matrix factorization algorithm to estimate admixture coefficients like other population genetics Bayesian clustering programs (Pritchard et al., 2000). LEA does not rely on Hardy-Weinberg equilibrium or linkage disequilibrium assumptions; therefore, it is suitable for asexual organisms like *M. roleri*. Estimation of the optimal number of hypothetical ancestral populations (K) was performed using the cross-entropy criterion (Frichot & François, 2015). The cross-entropy values were estimated for K values between 1 and 10, with 50 independent runs for each K with an algorithm regularization parameter of 100. The run with the lowest cross-entropy value was selected for further analysis and visualization in STRUCTURE PLOT v. 2.0 (Ramasamy et al., 2014). Spatial analysis of the population admixture proportions of samples at the geographical scale was performed using mapplots package (<https://CRAN.R-project.org/package=mapplots>).

2.2.3 | Genetic diversity and potential of recombination

Several diversity indices were performed over the SSR and SNP data sets. Clonal richness (R) (Arnaud-Haond et al., 2007) was estimated and the genetic variation in isolates in each region was quantified in terms of gene and genotypic diversity. For the former, the Nei diversity index, while for the latter, the Shannon-Wiener (H), Simpson (λ) and Stoddart and Taylor (G) indices (Kamvar et al., 2014) were calculated. The distribution of the gene and genotypic diversity among regions was calculated with the evenness index (E_j), which provides a ratio between abundant over rare genotypes (Grünwald et al., 2003). The genetic differences among regions were evaluated with the pairwise genetic diversity statistics (G_{ST}) analyses computed with the mmod

package (Winter, 2012). Similarly, the diversity and evenness indices of individual SSR/SNP loci were calculated to determine the most informative ones. Additionally, the potential of sexual recombination in the samples based on the SSR data set was tested via the standardized index of association (I_A^S) with 10,000 random permutations and using the parametric bootstrap shuffling method implemented in poppr (Kamvar et al., 2014). All values were calculated with both the entire (without clone correction) and the clone-corrected data set.

2.2.4 | Association analyses

Association analyses were performed via a principal component analysis (PCA) with the entire data set using the ade4 package (Dray & Dufour, 2007), and a Nei-distance UPGMA tree with 1000 bootstraps with the clone-corrected data set. Additionally, a minimum spanning network analysis was performed with poppr (Kamvar et al., 2014), to examine the relationships and genetic distances among genotypes and to ultimately infer centres of diversity/origin. This was accomplished using a discrete dissimilarity matrix reflecting the number of differing alleles or SNPs. The location in the network of the genotypes from countries recently invaded and from countries where the disease was originally described, as well as from other species of cacao, was examined. The geographical areas of genotypes located in the centre of the network corresponds to the centre of origin of *M. royeri* as shown in other pathogens (Couch et al., 2005).

2.3 | ITS sequence analysis

In order to identify the most ancestral genotype, and because SSR markers were not specific to *M. pernicioso*, we used ITS sequences of our *M. royeri* isolates and publicly available sequences of *M. pernicioso* (Table S3) to conduct a minimum spanning network analysis as detailed above. The ITS sequences generated in this study were obtained via PCR and Sanger sequencing as previously described (Aime & Phillips-Mora, 2005). Sequences were edited with Sequencher v. 5.0 (Gene Codes Corp.). The alignment was performed with MUSCLE on MEGA X (<https://www.megasoftware.net/>). The ITS data set was transformed in a similar way to the SNP data set for the network analysis.

3 | RESULTS

3.1 | SSR data set

3.1.1 | Clone correction and identification of MLGs and MLLs

Fifty-seven MLGs were initially identified across the 228 *M. royeri* isolates in this study (Table S4). The spectrum of genetic diversity with Nei genetic distances of samples was examined with unique MLGs (clone-corrected data set) and found to have a bimodal distribution

($D = 0.076$; $p < 2.2 \times 10^{-6}$) with a small peak in the beginning of the spectrum, setting up the threshold value for MLL identification at 0.13 (Figure S1a). Therefore, MLGs that clustered together at a 0.13 Nei distance or lower were evaluated to see whether they belonged to the same MLL via p_{sex} . Five MLLs were identified ($p_{\text{sex}} < 0.01$; Figure S1b) and their differing alleles were adjusted (Table S4). The final SSR data set used in all downstream analyses contained a total of 50 MLGs including the representative genotype for each of the five identified MLLs (Figure S1b). Hereinafter we refer to MLGs and MLLs as genotypes to facilitate reading. One genotype from Central America was found in Ecuador (MLG_5) and the genotype found in Bolivia was repeatedly observed within the Peruvian samples (MLL_5). All 14 samples from Maynas province corresponded to a unique genotype not seen in any other region, MLG_17 (Table S5). Samples from non-cacao hosts *T. grandiflorum* and *T. bicolor* had the same genotypes as samples from *T. cacao*, except for samples JD_E13 and JD_E25, both from *T. grandiflorum*, with unique genotypes (Table S5). Thirty-six samples came from 16 pods, and we found that samples from three pods had different genotypes. In other words, we found two clones infecting a single pod in 3 out of 16 possible instances.

3.1.2 | SSR markers

The highest levels of polymorphism at the individual SSR-locus level were found in Ecuador and Colombia (Table S6). The average number of alleles (N_a) per locus found in these countries was 2.88 and 2.75, respectively, while in Peru and Central America N_a was 1.38 and 1.25, respectively; in Maynas province and Bolivia only one allele per locus was found (Table S6). Overall, the most polymorphic locus was Mr_SSR1 with nine alleles, followed by Mr_SSR27 with eight and Mr_SSR18 and Mr_SSR33 both with seven (Table S6). The loci with the highest Nei gene diversity and with the most evenly distributed alleles among samples were Mr_SSR1, Mr_SSR4 and Mr_SSR27 ($N_a > 5$; $N_ei > 0.74$, $E_s > 0.70$). In Colombia, locus Mr_SSR25 was the most informative as it had the highest gene diversity, and their alleles were among the most evenly distributed ($N_ei = 0.81$; $E_s = 0.93$). Also, the diversity of Mr_SSR23 and Mr_SSR25 loci in Colombia severely declined in all other regions (Table S6). In Ecuador, the most informative loci were Mr_SSR1, Mr_SSR4 and Mr_SSR33 based on Nei gene diversity and E_s indices. Regarding null alleles, 21 samples had at least one null allele: 19 from Ecuador, 1 from Colombia and 1 from Peru. Marker Mr_SSR4 generated the highest number of null alleles in the samples analysed (seven samples), followed by Mr_SSR33 and Mr_SSR28 with six and five, respectively (Table S4). On average, only 2.3 (1%) samples had null alleles at each locus.

The genotype accumulation curve showed that the 16 SSR markers can discriminate almost all genotypes encountered in the data set and that the addition of extra markers would not increase the number of encountered genotypes (Figure S2a), that is, the number of markers is sufficient to discriminate all available genotypes. The rarefaction curves showed that the incorporation of more samples

from Central America, Peru, Bolivia and Maynas into the analysis would probably not contribute with new *M. royeri* genotypes, that is, our sampling strategy was adequate for capturing most predicted variability in these regions (Figure S2b). Conversely, rarefaction curves show that the diversity in Colombia and Ecuador remained undersampled as the curves had not reached a plateau, especially in Ecuador (Figure S2b).

3.1.3 | Genetic structure, diversity, relationship of samples and spatial correlation

The genetic structure among *M. royeri* genotypes performed in LEA identified that samples derived from six hypothetical ancestral populations ($K = 6$; Figure S3a). The region with samples deriving from the highest number of ancestral populations was Ecuador with $K = 4$, followed by samples from Colombia with $K = 3$ (Figure 3). In all the other regions (Central America, Maynas province, mainland Peru and Bolivia), *M. royeri* samples derived from only one ancestral population (Figure 3). The ancestral population deriving the genotypes of

the Central American samples was also found in Ecuador, occupying 4.4% of the population admixture in this country; the ancestral population deriving the genotype of the Maynas samples was found in Colombia, occupying 44.3% of the population admixture; and the ancestral population deriving the genotypes in Peru and Bolivia was found in both Ecuador and Colombia, occupying 2.1% and 10.5% of the population admixture, respectively (Figure 3b).

The pairwise G_{ST} values computed between samples from each region showed a high degree of genetic difference ($G_{ST} \geq 0.2$) in region pairs other than Peru and Bolivia, which were almost genetically identical (Table 1). When looking at the regions with a longer history of *M. royeri*, that is, Colombia and Ecuador, Colombia was genetically more similar to regions recently invaded ($G_{ST} = 0.20$ – 0.44) than Ecuador ($G_{ST} = 0.44$ – 0.62 ; Table 1).

Clonal richness and evenness were more informative in the data set without clone correction (Table 2). In Colombia and Ecuador, more than half of the samples had unique genotypes (R values: 0.53 and 0.62, respectively; Table 2). Also, the genotypes found in Colombia and Ecuador were more evenly distributed compared to other regions (E_s values: 0.91 and 0.81, respectively; Table 2). Gene

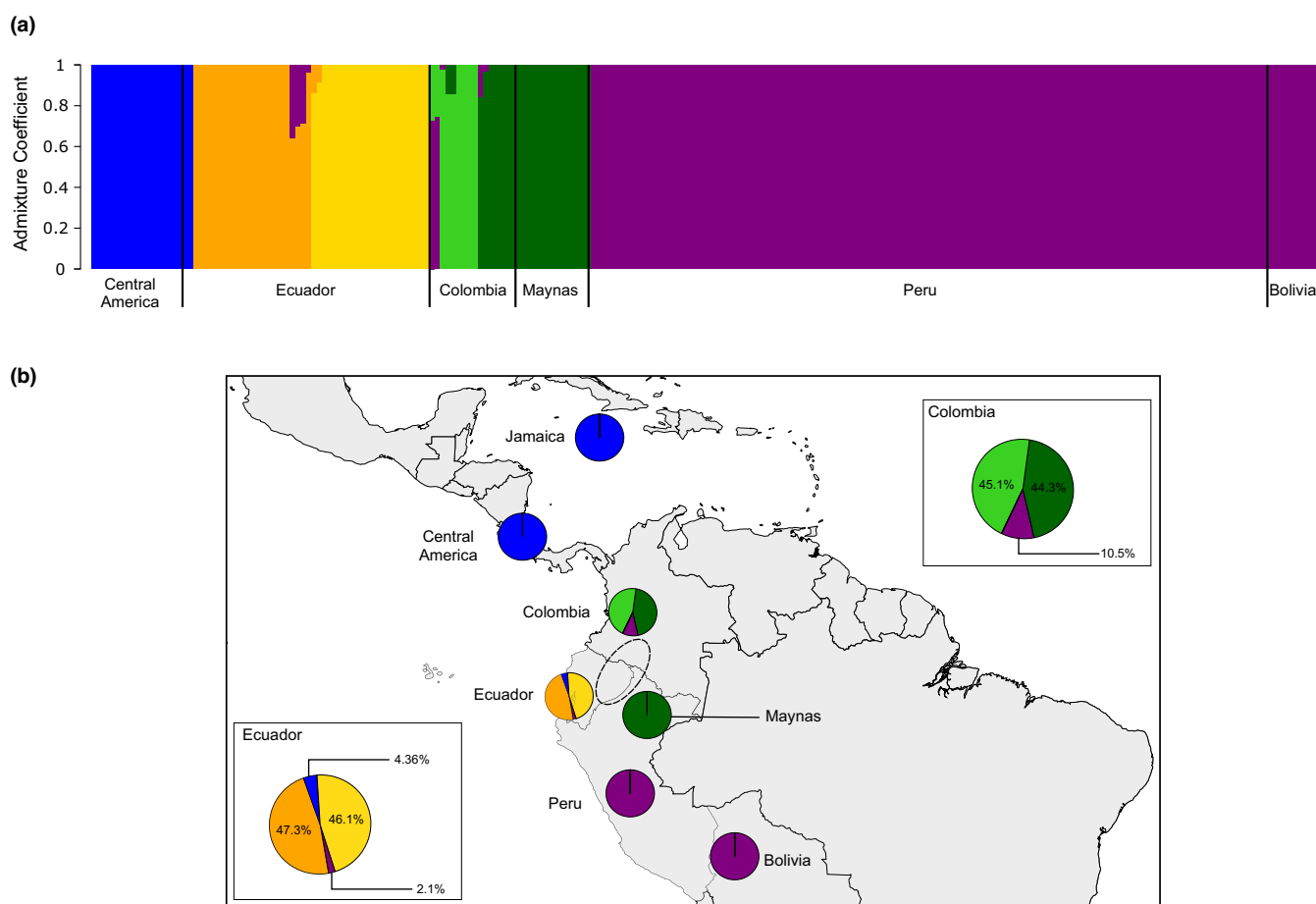


FIGURE 3 Population genetic structure of *Moniliophthora royeri* based on the simple-sequence repeat (SSR) data analysed in LEA. (a) Plot generated using $K = 6$ as the optimal number of hypothetical ancestral populations, which have their own colour in the plot. Each *M. royeri* sample is represented by a single vertical line coloured according to its coefficient for each ancestral population (Table S4). (b) Population admixture proportions of *M. royeri* in the regions sampled; the dashed line encompasses the Amazonian area shared by Ecuador, Peru and Colombia that remains unexplored in terms of *M. royeri* genetic diversity but is near the probable centre of origin for the pathogen. [Colour figure can be viewed at wileyonlinelibrary.com]

Data set	Region	Ecuador	Peru	Maynas	Bolivia	Central America
Non-clone corrected	Peru	0.56				
	Maynas	0.71	0.98			
	Bolivia	0.57	-0.01	1.00		
	Central America	0.52	0.89	0.93	0.91	
	Colombia	0.28	0.39	0.33	0.39	0.49
Clone-corrected	Peru	0.44				
	Maynas	0.62	0.80			
	Bolivia	0.45	-0.07	1.00		
	Central America	0.46	0.73	0.86	0.81	
	Colombia	0.27	0.32	0.20	0.27	0.44

TABLE 1 Genetic differentiation between regions sampled based on pairwise genetic diversity statistics (G_{ST}) comparisons with the simple-sequence repeat (SSR)-marker data set

TABLE 2 Genetic diversity of *Moniliophthora roreri* based on analyses of the simple-sequence repeat (SSR)-marker data set

Data set	Region	N	Diversity index								
			MLG	R	Nei	H	λ	G	E_s	I_A^S	p^a
Non-clone-corrected	Central America	17	5	0.25	0.07	1.28	0.65	2.86	0.72	0.127	NA
	Colombia	16	9	0.53	0.53	2.10	0.87	7.53	0.91	0.121	0.0001
	Maynas	14	1	0.00	0.00	0.00	0.00	1.00	NaN	NaN	NA
	Ecuador	46	29	0.62	0.31	3.19	0.95	19.96	0.81	0.114	0.0001
	Peru	126	7	0.05	0.01	0.44	0.17	1.20	0.36	0.021	NA
	Bolivia	9	1	0.00	0.00	0.00	0.00	1.00	NaN	NaN	NA
	All	228	50	0.22	0.37	2.30	0.70	3.28	0.25	0.379	0.0001
Clone-corrected	Central America	5	5	1.00	0.13	1.61	0.80	5.00	1.00	0.014	NA
	Colombia	9	9	1.00	0.59	2.20	0.89	9.00	1.00	0.041	0.023
	Maynas	1	1	0.00	NaN	0.00	0.00	1.00	NaN	NA	NA
	Ecuador	29	29	1.00	0.36	3.37	0.97	29.00	1.00	0.092	0.0001
	Peru	7	7	1.00	0.14	1.95	0.86	7.00	1.00	-0.097	NA
	Bolivia	1	1	0.00	NaN	0.00	0.00	1.00	NaN	NA	NA
	All	52	50	0.96	0.54	3.90	0.98	48.30	0.98	0.184	0.0001

N, number of samples; MLG, number of multilocus genotypes or multilocus lineages; R, clonal richness; Nei, Nei gene diversity; H, Shannon-Wiener index; λ , Simpson index; G, Stoddart and Taylor index; E_s , evenness index; I_A^S , standardized index of association; NA, not available; NaN, not a defined number.

^aProbability value of I_A^S .

and genotypic diversity were more informative in the data set with clone correction. Colombia and Ecuador were the most genetically diverse regions (Nei values: 0.59 and 0.36; H values: 2.20 and 3.37, respectively; Table 2). No evidence of sexual recombination was found in any region as the I_A^S was always significantly greater than zero ($p < 0.05$), or in the case of Peru, the uneven and low allelic (gene) diversity prevented the bootstrapping method to calculate a p value for I_A^S (Table 2).

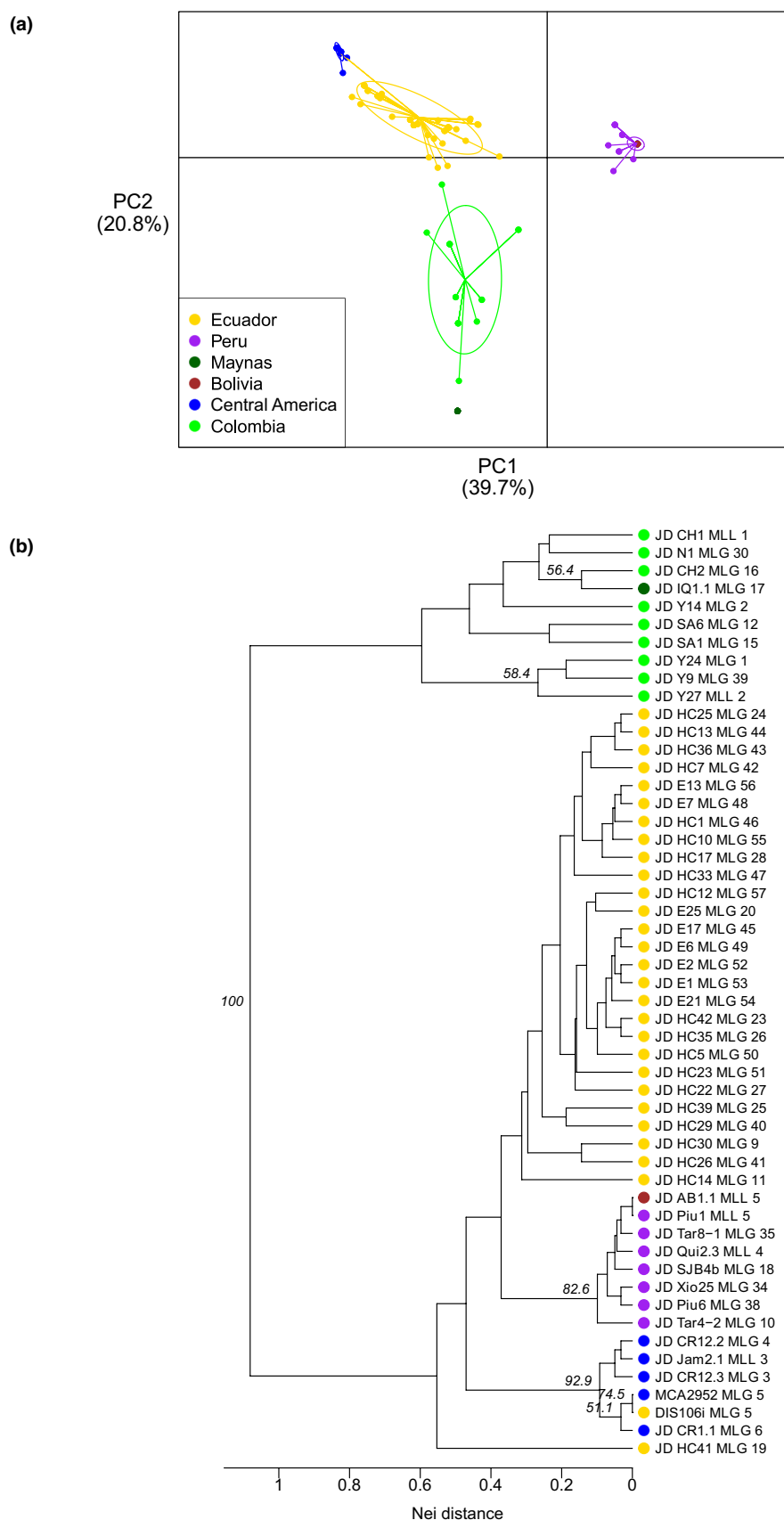
The PCA plot depicting two of the highest variance-explaining principal components (60.5%) showed some Ecuadorian genotypes had closer affinities to some Colombian ones (Figure 4a), while the bootstrapped dendrogram clearly separated Ecuador and Colombia samples (Figure 4b). However, both association analyses (PCA and bootstrapped dendrogram) consistently grouped together *M. roreri* samples from Central America and Ecuador; Colombia and Maynas

province; and Peru and Bolivia (Figure 4). Additionally, the network analysis showed the relationship and the geographic pattern of dissemination of *M. roreri* genotypes (Figure 5a). Ecuadorian genotypes are located in the centre of the network, from which the Central American, Peruvian and Bolivian genotypes were derived (Figure 5a). The network analysis also showed that the Maynas province genotype probably came from Colombia (Figure 5a). Genotypes from samples from other *Theobroma* species were interconnected to genotypes from *T. cacao* and did not form host-specific lineages (Figure 5a).

3.2 | ITS sequence analysis

Because we were unable to amplify any other species of *Moniliophthora* with our complete set of SSR markers, determining

FIGURE 4 Association analysis of *Moniliophthora roreri* samples based on the 16-simple-sequence repeat (SSR) locus data. (a) Principal component analysis (PCA) plot depicting the first two principal components accounting for 60.5% of the variation. Each dot represents a sample; inertia ellipses enclose 95% confidence clouds for each region/country; lines connect each sample to the centre of its ellipse. All samples from Bolivia share the same multilocus genotype (MLG) with samples from Peru, and two samples from Ecuador share the same MLG with samples from Central America, so the dots overlap. (b) UPGMA tree based on Nei genetic distance with 1000 bootstraps; bootstrap values greater than 50% are shown on branches; only one sample representing each MLG or multilocus lineage (MLL) is shown in the tree and is marked with a coloured dot based on the legend in (a). [Colour figure can be viewed at wileyonlinelibrary.com]



the most ancestral *M. roreri* genotype with these markers was not possible. Therefore, we analysed the ITS sequences of our *M. roreri* isolates and of other isolates of *M. perniciosa* (Figure 5b, Table S3).

The *M. roreri* clones that are more similar to the ancestor of modern *M. roreri* isolates (based on ITS sequence) were compared to the SSR network (Figure 5a). Isolates JD_E1, JD_HC28, JD_HC41 from Guayas

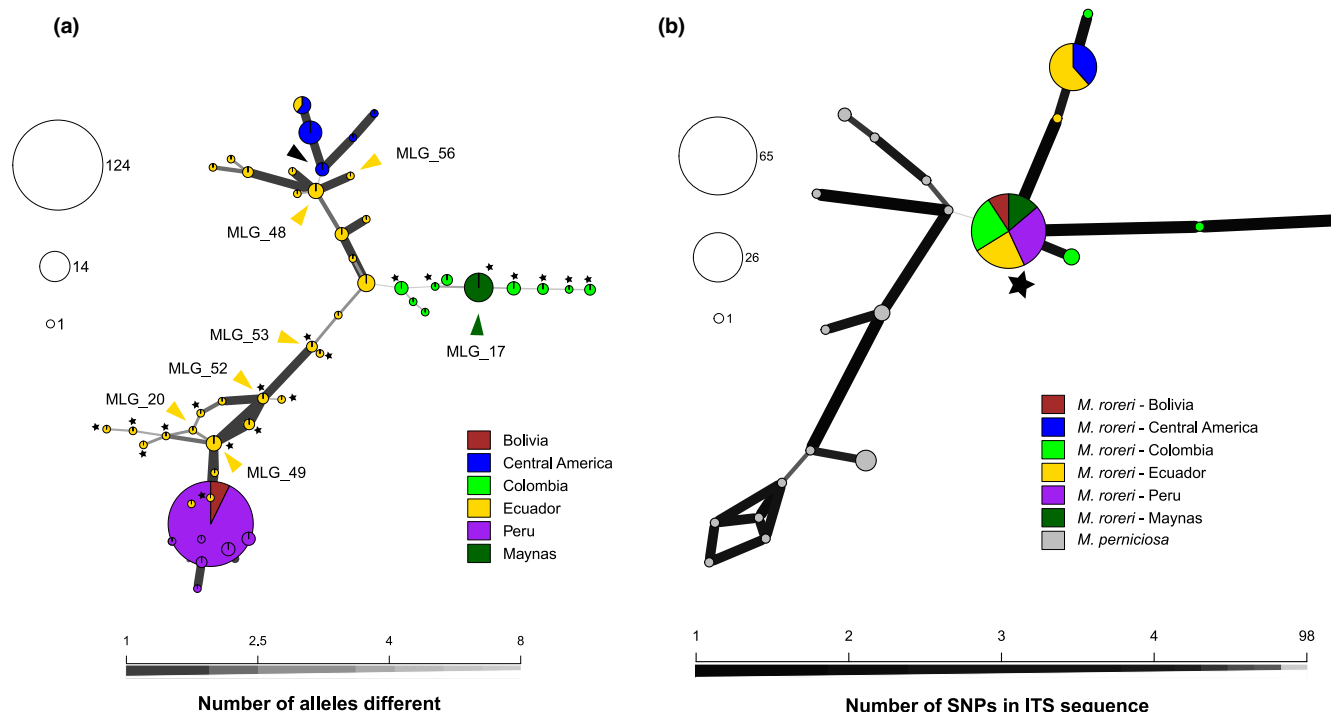


FIGURE 5 Minimum spanning network of *Moniliophthora roreri* isolates. Nodes (circles) represent multilocus genotypes/lineages (MLGs/MLLs); node size, the number of samples; colours, regions; and connecting line widths and shading, relatedness based on number of different alleles (simple-sequence repeat [SSR] data) or single-nucleotide polymorphisms (SNPs) (internal transcribed spacer [ITS] data); the widths are scaled proportionally to the inverse of the observed genetic distance, that is, the thicker the line, the more related the connected nodes are; line lengths are arbitrary. (a) Network built with the entire SSR data set of *M. roreri*. Coloured arrowheads indicate nodes containing at least one MLG from a host other than *Theobroma cacao* (Table S5); the black arrowhead indicates the MLG from Jamaica. (b) Network built with the ITS sequences including *M. pernicioso* (Table S3). The black star(s) indicate the isolate(s) at the base of the ITS network that share(s) the most similarity with the sister taxa. [Colour figure can be viewed at wileyonlinelibrary.com]

(in coastal region of Ecuador) and JD_Y27, JD_Y30 from Yacopí (in the inter-Andean middle Magdalena Valley of Colombia) shared the earliest ITS genotype and were present in two of the three main lineages (Figure 5), implicating an unsampled area between Yacopí and Guayas as the most likely centre of origin of *M. roreri*.

3.3 | SNP data set

To compensate for limited sampling from Colombia, we used an available SNP data set containing 66 isolates from throughout cacao-producing areas of Colombia (Ali et al., 2015). This data set was clone-corrected in the same manner as the SSR data set (Table S4 and Figure S1c,d). The SNP spectrum of genetic diversity had a bimodal distribution ($D = 0.017$; $p = 8.27 \times 10^{-4}$) and the threshold value for MLL identification was 0.0365, with 11 MLLs being identified ($p_{\text{sex}} < 0.01$; Figure S1c). The final SNP data set used in further analyses contained 33 MLGs/MLLs (Figure S1d). Only one MLL was found in Costa Rica, which was present in both Colombia and Ecuador. Also, only one MLL was found in Peru/Bolivia, which was only present in Ecuador but not in Colombia (Figure S1d). Just as in the clone correction with the SSR data set, hereinafter we will refer to MLGs and MLLs identified in the SNP data set as genotypes.

The genetic structure analysis on the SNP data set performed in LEA identified that samples derive from 10 hypothetical ancestral populations ($K = 10$; Figure S3b). The region with samples deriving from the highest number of ancestral populations was Colombia followed by samples from Ecuador with $K = 4$ (Figure S3). In the other regions covered by the SNP data set (Costa Rica, Peru and Bolivia), *M. roreri* samples derived from only one ancestral population (Figure S3c).

Within the non-clone-corrected SNP data set, the highest number of unique genotypes belonged to Colombia and Ecuador (R values: 0.47 and 0.30 and E_s values: 0.76 and 0.66, respectively; Figure S4a). Because there was only one genotype found in both Peru/Bolivia and Costa Rica, they have zero values for R (Figure S4a). Additionally, Colombia and Ecuador had similar values of Nei gene diversity (0.36 and 0.35, respectively), while Colombia had higher genotypic diversity than Ecuador (results with clone correction; Figure S4a). Finally, the haplotype network based on the SNP data set put Ecuadorian genotypes embedded within the network conformed by the Colombian genotypes in the centre (Figure S4b).

4 | DISCUSSION

The population genetics analyses of plant pathogens invading new geographical regions may be used to reveal their invasion routes and

TABLE 3 Comparison of the genetic diversity of *Moniliophthora roreri* from the literature

Study	Marker type	No. of markers	Region	N	Diversity index			
					g	R	H	H/ln (g)
Gutarra et al. (2013)	RAPD	14	Peru	21	3	0.10	0.39	0.35
Phillips-Mora et al. (2007) ^a	AFLP/ISSR and ITS	4 AFLP, 7 ISSR, 11 in total	Colombia	18	15	0.82	0.20	0.07
			Ecuador	36	11	0.29	0.15	0.06
			Venezuela	4	3	0.67	NA	NA
			Peru	2	1	0.00	0.00	NaN
			Central America	37	5	0.11	0.01	0.00
Maridueña-Zavala et al. (2016)	ITS-RFLP	4	Ecuador	90	50	0.55	0.21	0.05
Grisales Ortega and Kafuri (2007)	RAPD	49	Colombia	170	6	0.03	0.32	0.18
Moreira (2006)	SSR	7	Ecuador	25	6	0.21	0.12	0.07
			Peru	25	9	0.33	0.38	0.17
Jaimes et al. (2016)	SSR	23	Colombia	120	117	0.97	4.75	1.00
Ali et al. (2015) ^b	SNP	88	Colombia ^c	67	32	0.47	3.47	1.00
			Ecuador	11	4	0.30	1.49	1.07
This study	SSR	16	Colombia	16	10	0.60	2.30	1.00
			Ecuador	46	39	0.84	3.37	0.92
			Central America	17	5	0.25	1.61	1.00
			Peru/Bolivia	135	8	0.05	1.95	0.94

Note: N, number of samples used; g, number of genotypes and in the case of RAPD studies, number of groups detected; R, clonal richness, calculated in this study; H, Shannon–Wiener index; scaling, H/ln (g) as suggested to compare genetic diversity studies with varying number of samples (Grünwald et al., 2003, 2017); NA, not available in the referred study. NaN, not a defined number.

^aDiversity indices calculated by Phillips-Mora et al. (2007) did not include ITS.

^bDiversity indices for data by Ali et al. (2015) were calculated in the present study. H should be calculated based on clone-corrected data sets, but only Jaimes et al. (2016) and this study provide such a calculation.

^cOne sample out of the 67 in this study is from Venezuela.

allow inference of their centres of origin and diversity (Stukenbrock & McDonald, 2008). The invasive history of the cacao pathogen *M. roreri* started about 60 years ago; since then, the pathogen has increased its range throughout the Americas via clonal reproduction (Díaz-Valderrama & Aime, 2016a, 2016b). Currently, global cacao production is under threat from *M. roreri* (Díaz-Valderrama et al., 2020). However, its centre of origin and the factors that triggered the invasive stage of the fungus remain unknown. In this study we developed a robust set of SSR markers to explore these unresolved questions via population genetics analyses.

We performed extensive sampling in areas not previously sampled or incompletely sampled for *M. roreri*. These included Amazonian Peru (126 samples total)—the region with the highest Shannon diversity index for the host, cacao (Thomas et al., 2012)—Bolivia, Jamaica and southern Peru. We also included samples from two regions suspected to harbour high levels of genetic diversity for *M. roreri*—the coastal region of Ecuador, one of the two centres of pre-Colombian domestication of cacao (Díaz-Valderrama et al., 2020) and the Magdalena Valley of Colombia. Our analyses show that our sampling was sufficient for most regions except for Colombia and Ecuador, where additional genotypes are likely. Our findings in Ecuador are

even more striking given that 42 of the 46 Ecuadorian samples analysed came from a single farm in the coastal province of Guayas.

Several studies have investigated genetic diversity of *M. roreri* with varying interpretations (Table 3). Most of these have concluded that Ecuador or Colombia are the regions with the highest diversity in terms of clonal richness and Shannon–Wiener index (Table 3). Only one study found higher genetic diversity in Peru than in Ecuador (Moreira, 2006), but this is probably due to an artefact of the nature of the seven SSR markers employed, which were designed for analyses on *M. perniciosa* and as a result three of them were not polymorphic in *M. roreri* (Moreira, 2006). Detailed examination of the methods used in previous studies on genetic diversity of *M. roreri* revealed some issues (Table 3). Except for Jaimes et al. (2016), those studies did not provide results with clone-corrected data sets. Thus, in the current study, we performed genetic diversity analyses with and without clone correction of the generated SSR data set, as suggested for clonal organisms (Arnaud-Haond et al., 2007; Kamvar et al., 2014). One study based on SNP markers concluded that the Magdalena Valley of Colombia is the centre of genetic diversity of *M. roreri* (Ali et al., 2015). However, it did not provide any diversity index measurement other than the number of genotypes found across the

regions they sampled. Here, we reanalysed the same SNP data set and compared results with the 16-SSR data set. We found nearly the same gene diversity in Colombia and Ecuador in the SNP data set, and the genotypic diversity found with the SSR data set, as opposed to the SNP data set, was higher in Ecuador than in Colombia. These results are inconsistent with an exclusive origin of *M. royeri* in the Magdalena Valley of Colombia. Additionally, several other lines of evidence support this: the SSR genotypes found in Ecuador are different from the genotypes found in Colombia; there is a high genetic differentiation between Ecuador and Colombia revealed by a pairwise G_{ST} comparison; and the ancestral admixture proportions between these two regions highly differ. This scenario would not be possible if Colombia were the sole centre of diversity.

Our results on the genetic diversity indices and genetic structure indicate that *M. royeri* has diversified in two regions: coastal Ecuador and the inter-Andean upper and middle Magdalena Valley of Colombia, neither of which correspond to the centre of origin for the host (Thomas et al., 2012). However, agricultural systems can be major drivers for the emergence of plant pathogens in regions outside of the centre of origin of the host (Stukenbrock & McDonald, 2008). These two hotspots of diversification in *M. royeri* may reflect historic patterns of intense cultivation of this crop, rather than the origin of the pathogen. The coast of Ecuador and the Maracaibo basin (Venezuela and Colombia) have a long history of cultivation of cacao dating to pre-Colombian times. Coastal Ecuador continued to be a centre of cultivation during the colonial period (Díaz-Valderrama et al., 2020). The middle Magdalena Valley is south of Norte de Santander in Colombia, located in the Maracaibo basin, and the region where a disease corresponding to the symptoms of *M. royeri* was first observed about 200 years ago (Phillips-Mora et al., 2007).

The genetic diversity of *M. royeri* in the Amazonian region shared by Ecuador, Peru and Colombia has not been analysed. If the centre of origin of *M. royeri* coincided with the centre of origin of cacao, we would expect high levels of diversity within this Amazonian region and that association analyses would interconnect the independent diversification events that occurred in coastal Ecuador and inter-Andean Colombia. The network built with *M. royeri* ITS sequences supports this hypothesis as, irrespective of their SSR genotype, samples from Ecuador, Colombia, Maynas, Peru and Bolivia share the ITS genotype located in the origin of the *M. royeri* network (Figure 5b). From the point of origin, *M. royeri* appears to have migrated into at least two regions of intense cultivation where it underwent extensive diversification, coastal Ecuador and the Maracaibo basin. Subsequently, at least three separate introductions of *M. royeri* have occurred from these diversification centres in a severe founder effect fashion, in which genetic diversity drastically decreased. The first moved from Ecuador or Colombia (the Central American SNP data set-genotype was found in both Ecuador and Colombia) into Panama in 1958 and then spread northward, reaching Mexico and Belize in 2005 (Phillips-Mora, Cawich, et al., 2006; Phillips-Mora, Coutiño, et al., 2006) and then into the Caribbean, reaching Jamaica in 2016 (Johnson et al., 2017) (Figure 5a). A second invasive *M. royeri*

lineage is responsible for the relatively newer invasion that has moved from Ecuador or Colombia southward into northern Peru in 1988 (Hernández et al., 1990), to southern Peru in 1998 (Ríos-Ruiz & Rodríguez, 1998) and recently invaded Bolivia (Phillips-Mora et al., 2015). A third lineage in Maynas province appears to represent a more recent invasion.

An alternative hypothesis may suggest that *M. royeri* had already diversified in coastal Ecuador and inter-Andean Colombia before the introduction and expansion of cacao cultivation. After these events, new confrontations between the extant and introduced genotypes of the pathogen may have originated the high diversity observed in this study. However, the centre of diversity of *Theobroma* and *Herrania*, the only known hosts of *M. royeri*, in the upper Amazon of Peru, Ecuador and Colombia (González-Orozco et al., 2020; Thomas et al., 2012), and the lack of sexual recombination in *M. royeri*, conflicts with its diversification in non-Amazonian areas.

Only a single genotype was recovered in Maynas province, this having closest genetic affinities to the Colombian samples. Therefore, it is likely that *M. royeri* was introduced to this region from Colombia. The Maynas province is a geographically isolated area embedded in the Peruvian Amazon forest and the capital city, Iquitos, can only be reached via river or by air. The Loreto department, to which Maynas province belongs, is a minor cacao-producing region in Peru yet the one with the highest growth rate in terms of cacao cultivated area in recent years (MINAGRI, 2018). In 1995, there were only 236 ha of cacao but by 2016 this had increased to 7053 ha (Evans et al., 1998; MINAGRI, 2018). This agricultural expansion may explain the clonal dissemination of the Maynas genotype into the province. Although the date of this introduction could not be determined, it is unlikely that it predates the first official introduction in Peru in 1988 (Hernández et al., 1990), as the genotypes in mainland Peru have had enough time during its expansion to undergo some diversification (mutations), and the Maynas lineage does not appear to have done so as yet.

The genotypes found in *M. royeri* samples from *T. grandiflorum* were also found in commercial cultivars of *T. cacao*. The outer perimeter of this farm was planted with *T. grandiflorum* trees whose fruits would get infected by *M. royeri*. Similarly, the genotype from *T. bicolor* from Maynas province was the same as that found in *T. cacao* trees throughout the province. These genotypes from hosts other than *T. cacao* were not genetically related, meaning there is no signal of host specificity. Conversely, it speaks of its host range of the pathogen beyond *T. cacao* (Phillips-Mora & Wilkinson, 2007), and its capacity to persist in cacao plantations by means of infecting an alternative host, which serves as a reservoir of primary inoculum for the next cacao fructification and production season.

Although a centre of origin for *M. royeri* could not be definitively established, there is indirect evidence that *M. royeri* originated within the upper Amazonian area that today extends through parts of modern-day Colombia, Peru and Ecuador. From this region, it migrated, probably in association with its host plant, to at least two areas of intense cultivation—in coastal Ecuador and in the inter-Andean valleys of Colombia—where genetic diversity of the

pathogen was increased and intensified. From these genetic pools, at least two genotypes have arisen that were well adapted to become invasive—a lineage that spread northward throughout Central America, and a lineage that has more recently been spreading southward. The Maynas province of Peru may harbour a third invasive genotype that has most recently emerged and was introduced from Colombia.

ACKNOWLEDGEMENTS

Financial support was provided by the Peru National Fund for Scientific, Technological Development and Technological Innovation, contract number 030-2019-FONDECYT-BM-INC.INV. We also acknowledge travel research grants from the Mycological Society of America, Golden Key International Honour Society and the Department of Botany and Plant Pathology at Purdue University and USDA Hatch 1010662. We are indebted to all the farmers who kindly opened their doors to us for collection and camaraderie. We also want to thank agronomists and friends who helped us during collections. Some of these people are Wilmer Silva, Christian Saucedo, Wilder Rodríguez, Daniel Cruz, Rodrigo Baselly, Hugo Bautista, José Imán, Donald Flores, John Sarango and Cielo Ramos. We want to thank Luis Moncayo, Edison Carvajal and Danilo Vera for their support for our work in Ecuador. Thanks to Marcela Arteaga for sharing culture media and laboratory equipment in Jaén. Thanks also to Clara Ramírez, Manuel Alburqueque and José Luis Juárez for arranging logistics for the collection work in Piura. Thanks to Francisco Alejo for logistics in Quillabamba. Thanks to Gladys Chipana, Juan José Aparicio and Teresa Gilles for logistics in Bolivia. Thanks to John Freddy León from the Yacopi Technical Unit of Fedecacao for the sampling logistics in Colombia. Samples from Peru were collected under permits D000473-2021-MIDAGRI-SERFOR-DGGSPFFS and 005-2015-SERFOR-DGGSPFFS. Samples from Colombia were collected under permit 0255 (14 March 2014; Minister of Environment and Sustainable Development – contract no. 2 for project no. 385 – 2012 from the 'Corredor Tecnológico Agroindustrial' agreement, Cundinamarca). The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The ITS DNA sequences are available at GenBank at <https://www.ncbi.nlm.nih.gov/genbank/> with accession numbers MW898034–MW898105. All other data that support the findings of this study are available in the supplementary material of this article.

ORCID

Jorge R. Díaz-Valderrama  <https://orcid.org/0000-0002-1015-3751>

Gianmarco G. Casas  <https://orcid.org/0000-0001-5491-8771>

Julio A. Chia-Wong  <https://orcid.org/0000-0002-0947-9949>

REFERENCES

- Aime, M.C. & Phillips-Mora, W. (2005) The causal agents of witches' broom and frosty pod rot of cacao (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae. *Mycologia*, 97, 1012–1022.

- Ali, S.S., Shao, J., Strem, M.D., Phillips-Mora, W., Zhang, D., Meinhardt, L.W. et al. (2015) Combination of RNAseq and SNP nanofluidic array reveals the center of genetic diversity of cacao pathogen *Moniliophthora roreri* in the upper Magdalena Valley of Colombia and its clonality. *Frontiers in Microbiology*, 6, 850.
- Arnaud-Haond, S., Duarte, C.M., Alberto, F. & Serrão, E.A. (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology*, 16, 5115–5139.
- Couch, B.C., Fudal, I., Lebrun, M.-H., Tharreau, D., Valent, B., van Kim, P. et al. (2005) Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*, 170, 613–630.
- Díaz-Valderrama, J.R. & Aime, M.C. (2016a) The cacao pathogen *Moniliophthora roreri* (Marasmiaceae) possesses biallelic A and B mating loci but reproduces clonally. *Heredity*, 116, 491–501.
- Díaz-Valderrama, J.R. & Aime, M.C. (2016b) The cacao pathogen *Moniliophthora roreri* (Marasmiaceae) produces rhizolytic thallic conidia and their size is influenced by nuclear condition. *Mycoscience*, 57, 208–216.
- Díaz-Valderrama, J.R., Leiva-Espinoza, S.T. & Aime, M.C. (2020) The history of cacao and its diseases in the Americas. *Phytopathology*, 110, 1604–1619.
- Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14, 927–930.
- Dray, S. & Dufour, A.-B. (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22, 1–20.
- Evans, H.C., Krauss, U., Rios Rutz, R., Zeceovich Acosta, T. & Arévalo-Gardini, E. (1998) Cacao in Peru. *Cocoa Growers Bulletin*, 51, 7–22.
- Frichot, E. & François, O. (2015) LEA: an R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6, 925–929.
- González-Orozco, C.E., Galán, A.A.S., Ramos, P.E. & Yockteng, R. (2020) Exploring the diversity and distribution of crop wild relatives of cacao (*Theobroma cacao* L.) in Colombia. *Genetic Resources and Crop Evolution*, 67, 2071–2085.
- Grisales Ortega, S.P. & Kafuri, L.A. (2007) Análisis de variabilidad genética en *Moniliophthora roreri* con AP-PCR y RAPD en Antioquia, Colombia. *Revista Colombiana de Biotecnología*, 9, 15–32.
- Grünwald, N.J., Goodwin, S.B., Milgroom, M.G. & Fry, W.E. (2003) Analysis of genotypic diversity data for populations of microorganisms. *Phytopathology*, 93, 738–746.
- Grünwald, N.J., Everhart, S.E., Knaus, B.J. & Kamvar, Z.N. (2017) Best practices for population genetic analyses. *Phytopathology*, 107, 1000–1010.
- Gutarra, B.A., Silva, M.J., Márquez, K.J. & León, B. (2013) Análisis de la diversidad genética de 21 aislamientos del hongo *Moniliophthora roreri* basado en marcadores RAPD. *Apuntes de Ciencia y Sociedad*, 3, 100–110.
- Hartigan, P.M. (1985) Algorithm AS 217: computation of the dip statistic to test for unimodality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 34, 320–325.
- Hernández, T.T.A., Aranzazu, H.F., Arévalo-Gardini, E. & Rios, R.R. (1990) La moniliasis del cacao en el Perú. *Agrotrópica (Brasil)*, 2, 56–58.
- Jaimes, Y.Y., Gonzalez, C., Rojas, J., Cornejo, O.E., Mideros, M.F., Restrepo, S. et al. (2016) Geographic differentiation and population genetic structure of *Moniliophthora roreri* in the principal cocoa production areas in Colombia. *Plant Disease*, 100, 1548–1558.
- Jiménez, D.L., Alvarez, J.C. & Mosquera, S. (2021) Frosty pod rot: a major threat to cacao plantations on the move. *Tropical Plant Pathology*, 47, 187–200.
- Johnson, E.S., Rutherford, M.A., Edgington, S., Flood, J., Crozier, J., Cafá, G. et al. (2017) First report of *Moniliophthora roreri* causing frosty pod rot on *Theobroma cacao* in Jamaica. *New Disease Reports*, 36, 2.
- Kamvar, Z.N., Tabima, J.F. & Grünwald, N.J. (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281.

- Maridueña-Zavala, M.G., Villavicencio-Vásquez, M.E., Cevallos-Cevallos, J.M. & Peralta, E.L. (2016) Molecular and morphological characterization of *Moniliophthora roreri* isolates from cacao in Ecuador. *Canadian Journal of Plant Pathology*, 38, 460–469.
- Megléc, E., Pech, N., Gilles, A., Dubut, V., Hingamp, P., Trilles, A. et al. (2014) QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. *Molecular Ecology Resources*, 14, 1302–1313.
- MINAGRI. (2018) *Análisis de la cadena productiva del cacao con enfoque en los pequeños productores de limitado acceso al mercado*. Lima, Peru: Ministerio de Agricultura y Riego, Peru.
- Moreira R.F.C., 2006. *Estructura genética de populações de Crinipellis perniciososa e Moniliophthora roreri utilizando marcadores RAPD e SSR*. Thesis, São Paulo, Universidade Estadual Paulista.
- Olivera-Núñez Q, 2018. *Jaén, arqueología y turismo* (Yanapay Andina Consultores, Ed.). Jaén, Perú: Municipalidad Provincial de Jaén.
- Osorio-Guarín, J.A., Berdugo-Cely, J., Coronado, R.A., Zapata, Y.P., Quintero, C., Gallego-Sánchez, G. et al. (2017) Colombia: a source of cacao genetic diversity as revealed by the population structure analysis of germplasm bank of *Theobroma cacao* L. *Frontiers in Plant Science*, 8, 1994.
- Phillips-Mora, W., Aime, M.C. & Wilkinson, M.J. (2007) Biodiversity and biogeography of the cacao (*Theobroma cacao*) pathogen *Moniliophthora roreri* in tropical America. *Plant Pathology*, 56, 911–922.
- Phillips-Mora, W., Baqueros, F., Melnick, R.L. & Bailey, B.A. (2015) First report of frosty pod rot caused by *Moniliophthora roreri* on cacao in Bolivia. *New Disease Reports*, 31, 29.
- Phillips-Mora W, Castillo J, Arciniegas A, Mata A, Sánchez A, Leandro M et al., 2009. Overcoming the main limiting factors of cacao production in Central America through the use of improved clones developed at CATIE. In: *Proceedings of the 16th International Cacao Research Conference, COPAL; 16-21st November 2009*. Bali, Indonesia, pp. 93–99.
- Phillips-Mora, W., Cawich, J., Garnett, W. & Aime, M.C. (2006) First report of frosty pod rot (moniliasis disease) caused by *Moniliophthora roreri* on cacao in Belize. *Plant Pathology*, 55, 584.
- Phillips-Mora, W., Coutiño, A., Ortiz, C.F., López, A.P., Hernández, J. & Aime, M.C. (2006) First report of *Moniliophthora roreri* causing frosty pod rot (moniliasis disease) of cacao in Mexico. *Plant Pathology*, 55, 584.
- Phillips-Mora, W. & Wilkinson, M.J. (2007) Frosty pod of cacao: a disease with a limited geographic range but unlimited potential for damage. *Phytopathology*, 97, 1644–1647.
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Ramasamy, R.K., Ramasamy, S., Bindroo, B.B. & Naik, V.G. (2014) STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *Springerplus*, 3, 1–3.
- Ríos-Ruiz R & Rodríguez C, 1998. La moniliasis [*Moniliophthora roreri* (Cif. & Par.) Evans et al.] del cacao en la cuenca del alto Urubamba: primera ocurrencia e impacto económico esperado. In: *XV Congreso Peruano de Fitopatología*. Pucallpa, Peru, p. 14.
- Rozenfeld, A.F., Arnaud-Haond, S., Hernández-García, E., Eguíluz, V.M., Matías, M.A., Serrão, E. et al. (2007) Spectrum of genetic diversity and networks of clonal organisms. *Journal of the Royal Society Interface*, 4, 1093–1102.
- Stukenbrock, E.H. & McDonald, B.A. (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, 46, 75–100.
- Thomas, E., van Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G. & van Etten, J. (2012) Present spatial diversity patterns of *Theobroma cacao* L. in the Neotropics reflect genetic differentiation in Pleistocene refugia followed by human-influenced dispersal. *PLoS One*, 7, e47676.
- Thurston, U.D. (1973) Threatening plant diseases. *Annual Review of Phytopathology*, 11, 27–52.
- Winter, D.J. (2012) MMOD: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, 12, 1158–1160.
- Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I. et al. (2018) The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nature Ecology & Evolution*, 2, 1879–1888.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Díaz-Valderrama, J. R., Zambrano, R., Cedeño-Amador, S., Córdova-Bermejo, U., Casas, G. G., García-Zurita, N., et al. (2022). Diversity in the invasive cacao pathogen *Moniliophthora roreri* is shaped by agriculture. *Plant Pathology*, 71, 1721–1734. <https://doi.org/10.1111/ppa.13603>