POPULATION GENETICS AND GENOMICS TO UNDERSTAND THE INVASIVE HISTORY OF THE CACAO PATHOGEN *MONILIOPHTHORA RORERI*

by

Jorge Ronny Díaz-Valderrama

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Botany & Plant Pathology

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. M. Catherine Aime, Chair
Department of Botany & Plant Pathology
Dr. Charles Woloshuk
Department of Botany & Plant Pathology
Dr. Stephen S. Goodwin
Department of Botany & Plant Pathology
Dr. Jin-Rong Xu
Department of Botany & Plant Pathology
Dr. Timothy Y. James
Department of Ecology and Evolutionary Biology, University of Michigan

Approved by:

Dr. Cristopher J. Staiger
Head of the Graduate Program

En el mundo del Microsoft Word existe un botón al que vamos a llamar "la P al revés" y luce así: ¶

Tengo que confesar que esta P al revés me fue muy útil para poner esta tesis en el caprichoso formato que la universidad quiere. De hecho, su uso es la primera recomendación que te dan. Y es la primera vez en mi vida que yo, a puertas de un doctorado, la uso. Sin embargo, cuando yo tenía ocho años, María Jesús, de dieciséis, ya la usaba. Lo sé porque una vez este símbolo desapareció extrañamente de la lista de botones de Microsoft Word 95. María Jesús armó un escándalo, con mucho enojo y desesperación adolescente, y logró que todos en casa la traten de ayudar a recuperar este botón para que pueda continuar con su trabajo. Recuerdo claramente que nadie tenía idea de lo que ella estaba hablando. Pero ahora comprendo todo. Este botón te ayuda a alcanzar una alta organización, consistencia e incluso perfección en la alineación e interlineado de tus documentos. Y María Jesús es, desde siempre, muy organizada y perfeccionista en todas sus actividades. Esta P al revés la define en ese aspecto. Desafortunadamente no existen botones de Word que definan lo increíble que es un ser humano. Cuando María Jesús hacía sus prácticas profesionales, destinó 700 de 900 soles, que era su primer sueldo, para apoyarme económicamente porque yo empezaba a estudiar en una universidad privada en Lima, es decir lejos de nuestra Cajamarca. Y lo hizo cada mes por todo un año, hasta que logré obtener una beca. Y lo hizo porque sentía orgullo que yo estudie en dicha prestigiosa universidad. Y lo hizo porque quiso que salga adelante. Es por eso hermana, que te dedico esta tesis, a sabiendas de que nada de lo que hice, haga o te pueda ofrecer será más grande que esos 700 soles.

ACKNOWLEDGMENTS

My deep gratitude goes to all the people who positively influenced me during the last years and helped me to carry out this work. You, my friends, are so many, and I hope I don't forget to acknowledge any of you. If I do, please forgive me. Thank you, Cathie Aime, for your mentoring, for fully supporting me since I joined your lab way back in 2011. Thank you, Drs. Charles Woloshuk, Stephen Goodwin, Jin-Rong Xu, Tim James and Michael Gribskov, for your constructive input and advice throughout these years.

Thank you, my friends from Peru, Ecuador and Bolivia, who made my collection trips possible. Thank you, Wilmer Silva, for being my right-hand man in most of these trips; thank you, Sixto Díaz, for facilitating transportation. Thank you, Sergio Cedeño, Wilder Rodríguez and Daniel Cruz, for making your farms in Ecuador available to collect, and Carlos Zambrano, thanks so much for collecting samples. Thank you, Clara Ramírez, for putting me in touch with researchers and farmers from Piura. Thank you, Ulises Córdova, Donald Flores, Abelardo Palacios, Hugo Bautista, Boasdil Vidal, Juan Luis Talledo, John Sarango, Francisco Pasos, José Santos Adrianzen, Manuel Alburqueque, José Luis Juárez and José Imán, for facilitating, driving, walking and collecting with me in cacao farms in this region. Thank you, Amelia Valderrama and Rigoberto Marín, for putting me in contact with researchers from Jaén, and for hosting us in your own house. Thank you, Gianmarco Goycochea, Cielo Ramos, Marcela Arteaga and Rodrigo Baselly, for helping me with the collections in Jaén. Thank you, Indira Marín, Nelly Silva and your family, for making our stay in Jaén wonderful. Thank you, Tania Hoyos, for putting me in contact with researchers from San Martín. Thank you Nilda García and Juan Alfonso Sánchez, for helping us to collect in San Martin. Thank you, Juan Ruiz and Darío Dávila, for sharing your knowledge and helping us to collect in Iquitos, Maynas. My lifelong friend Christian Saucedo, thanks for using your vacation days to help me collect during our entire stay in Iquitos. Thank you, Xiomy Pinchi, for collecting samples in Ucayali. Thank you, Alejo Zapata, for facilitating my trip to Quillabamba, Cusco; thank you, Carlos Rodríguez, for taking me to the precise places to collect in this region. Thank you, Oscar Hurtado, Julio Chia-Wong and Zarella Quispe, for facilitating samples from Huánuco, Ucayali and San Martin. Thank you,

Gladys Chipana, Aparicio Torres and Fernando Manzaneda, for putting me in contact with researchers and farmers from Sapecho, Bolivia. Thank you, Casto Maldonado and Eusebio Pérez for helping us collect samples. Thank you, Teresa Gilles, for hosting us so well in La Paz.

Thank you, José Córdova, Elmer Saldaña, Juan Castillo, Orlando Neyra, Salomón Martínez, Fernando Correa, Pedro Guerrero, Segundo Guerrero from Piura; Eduardo Mendoza, Orlando Silva and Hernán Sangama from San Martin; Visitación Castro and Alejandro Montoya from Quillabamba; and Néstor Escobar, Nemio Márquez and Hipólito Pérez from Bolivia, for kindly opening the doors of your farms so we can collect in them. Thank you, Esperanza Rojas Torres and Dylan Guevara, for facilitating samples from Colombia, and thank you Drs. Wilbert Phillips-Mora and Elizabeth Johnson for facilitating samples from Costa Rica and Jamaica. Thank you, Rajdeep Khangura, for all the hours spent when we started to write SYLOCAL. Thank you, Kevin Bello, for sharing your knowledge and for making my scripts much smoother. Thank you, Drs. Francis Martin and Christoffer Bugge Harder, and 1000 Fungal Genome consortium for access to unpublished genome data. These genome sequence data were produced by the US Department of Energy Joint Genome Institute in collaboration with the user community.

Thank you, my dear lab mates. Rachel Koch, you have been like my sister throughout these years and I learned a lot from you, thank you for the conversations and your research. Thank you, Pedro Pablo Parra, Paula Gómez, Steve Russell, Luis Ramírez and Teeratas Kijpornyongpan, for the constructive discussions, feedback and your friendship. Thank you, Silenze Esquivel, Marlon Rodríguez, Seth Helfers, Nicole Klinger, Brandon Garver and Victoria Sato for letting me be your instructor in the lab and for collecting some of my data or keeping my cultures alive. Thank you, Lucia Molina, for the many hours you spent helping me process samples in a time I really needed help. Thank you, Dillon Husbands, Fanni Lakatos, Blaise Jumbam, for all the laughs and the great moments in the lab. Thank you, John Klimek, Hector Urbina, Danny Haelewaters, Sam Brown, Lucy Liu, Kirk Rumple, Jarius Chittenden, Shannon Newerth, Cade Kane, Mary Claire Noble and Nikolai Lutas, for making working in the lab so awesome.

Thank you, John Cavaletto, for teaching me how to teach. Thank you, Mariam Valladares, Anaí Caparó, José Bonilla, Luis Maldonado, Nicolette Albright and Yue Li, for spending your time helping me in some parts of my research. Thank you, Drs. Peter Goldsbrough, Chris Staiger and Dieudonné Baributsa, for economic support at different stages of my doctoral program. Dr. Clifford Sadof, thank you for allowing me to have experience in extension talks. Dr. Linda Naimi, thank you for teaching me how to persevere in adverse moments. Thank you, Pam Mow, Lisa Gross, Tyson McFall, Stevie Keen and Stacie Miller, for all the favors and paperwork done for me. Thanks to the Mycological Society of America, the Golden Key International Honour Society and the Department of Botany and Plant Pathology for support in the form of research and travel grants.

Don't think I have forgotten of you, my other friends, the ones who made my life enjoyable during these years. Thank you, Eliana Torres, Randol Rodríguez, Ariana Torres, Vanessa Sarria, Amanda Cardoso, Miguel Álvarez, Davis Chacón, Celina Gómez, José Gómez, Minerva Dorantes, Sergio Cortez, Clara Chávez and your beautiful family, Nick Harby, Paul Schields, Emily Gunter, Victor Maqque, and my Peruvian and Zamorano families. Finally, thank you, Blanca Valderrama and Jorge Díaz Mondragón, and María Jesús Amaya Valderrama and Karen Díaz Valderrama, parents and sisters of mine, for your support and for being my motivation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Díaz-Valderrama, Jorge R. PhD
Institution: Purdue University
Degree Received: May 2019
Title: History of Cacao Cultivation and Studies on The Center of Origin and Genomics of One of its Major Pathogens, *Moniliophthora roreri*
Committee Chair: M. Catherine Aime

Cacao (*Theobroma cacao* L.) is an ancestrally cultivated crop that has been the source of one of the most beloved commodities, chocolate. Its worldwide demand has shaped the history of its cultivation. In Chapter 1, the center of origin of cacao, its center of domestication, the most outstanding movements of germplasm from the Pre-Columbian to the Republican era, the appearance and discovery of major diseases, among other important economic, agricultural and social aspects regarding cacao cultivation are reviewed. The following chapters focus on one of the major pathogens of cacao in the Americas, *Moniliophthora roreri* causing frosty pod rot disease. Chapter 2 presents evidence that the center of origin of *M. roreri* is not limited to the Magdalena Valley in Colombia, as other studies have suggested, but extends to Ecuador and the Peruvian Upper Amazon. Chapter 3 focuses on the *A* and *B* mating type loci diversity of *M. roreri* and reports a new *A* mating allele in Colombia and new mating types in Colombia, Ecuador and infers the presence of even more mating types in Ecuador and the Peruvian Upper Amazon. Additionally, Chapter 3 introduces rapid approaches to collect *M. roreri* and to diagnose mating types. Finally, Chapter 4 touches the genomic aspect of *M. roreri* and its close relatives within the Marasmiineae suborder. It presents the most complete genome of a *Moniliophthora* species generated so far and describes the evolution of predicted effectors and other proteins that might be involved in pathogenicity in this suborder. It also releases a custom program called SyLOCAL that evaluates synteny of a cluster of genes between two genomes.

CHAPTER 1. HISTORY OF CULTIVATION OF CACAO AND THE SPREAD OF ITS MAIN DISEASES

1.1 Introduction

Cacao, *Theobroma cacao* L., is a tropical tree originating in South America that has had a great impact on the history of humanity (Motamayor et al. 2002, McNeil 2006). For example, the fruit of the cacao tree played a role in pre-Columbian spirituality; the beans constituted the first forms of currency thus moving the economy of ancient civilizations; its consumption as a drink captured Spanish conquerors; and the invention of chocolate candy amazed the whole of Europe, ultimately leading to the expansion of its cultivation throughout the world tropics (Dand 1997, McNeil 2006). The spread of the crop to places out if its center of origin brought many inherent issues: the increased demand for labor, especially during the colonial times, and the emergence and dissemination of diseases and pests (Turner 1974, Matson et al. 1997, Stukenbrock and McDonald 2008). The problem got bigger because of the narrow genetic background of cultivars used all over the world (Bennett 2003). Currently, it is estimated that diseases alone account for more than 30% of global cacao losses (Hebbar 2007). Additionally, political decisions in many countries affected cacao cultivation, for better or worse. All these circumstances implied subjection of people, reckless explorations for new cultivars and even collapse of entire countries' economies. Despite of all of this, cacao represents hope of progress for the people who continue cultivating it.

The aim of this literature review is to provide a broad, chronological account of the most important events and human decisions that impacted on cacao cultivation, from Pre-Columbian (Figure 1) to colonial (Figure 2) and republican (Figure 3) times. The movements of cacao germplasm for breeding purposes are also described. Additionally, this review explores the causes for both the prosperous and adverse times in the history of cacao cultivation, with a special emphasis on cacao diseases. Some social issues are included in this review, but profound analyses must be completed to better understand their real impact.

1.2 Cacao germplasm

Cacao, *Theobroma cacao* L., was traditionally classified into two major groups: *Criollo* and *Forastero* (Cheesman 1944). *Criollo* cacao refers to the one highly cultivated by the native people from Mesoamerica in pre-Columbian times, introduced by humans from the Southwestern and Northeastern regions of Venezuela and Colombia, respectively (Motamayor et al. 2002). *Criollo* cacao has superior quality beans but the trees are weak and susceptible to diseases (Ciferri and Ciferri 1957, Cuatrecasas 1964). On the other hand, *Forastero* cacao refers to all forms of cacao other than *Criollo*; therefore the term was questioned for its inaccuracy on representing this variable cacao group (Cheesman 1944). The other most important term in the traditional classification of cacao is *Trinitario*. This applies to hybrid cultivars between *Criollo* and *Forastero* (Ciferri 1949). *Trinitario* cacao appeared and started to expand after the cacao production downfall in Trinidad in the 1720s due to a disease, referred to as the "blast" or "blight" (Dand 1997, Leiter and Harding 2004). *Trinitario* cacao combined the quality of the *Criollo* cacao and showed disease tolerance conferred by the *Forastero* genetic background (Cuatrecasas 1964). However, the term still comprised a very variable cacao group (Motamayor et al. 2003).

Currently, modern molecular technologies have allowed classification of cacao germplasm into eleven genetic clusters, named according to geographic origin or to the most representative cultivar of the cluster (Motamayor et al. 2008). These are: *Criollo, Marañon, Curaray*, *Iquitos*, *Nanay*, *Contamana, Amelonado* (which was proved to be the parental genotype for the old *Trinitario* group; Motamayor et al. 2003), *Purús*, *Guiana, Nacional Ecuatoriano* (Motamayor et al. 2008) and *Nacional Boliviano* (Zhang et al. 2012). Therefore, from this list one can easily observe that the *Forastero* cacao group comprised ten current genetic clusters and that Cheesman's concerns about this group were justifiable (Cheesman 1944). Furthermore, in Peru there are three recognized native cacao cultivars without molecular characterization: the "Chuncho" cacao, the "Blanco Piurano" and the "Cacao Amazonas Peru" (which has a protected designation of origin) that might constitute different genetic clusters because of their unique organoleptic properties (Soria V. 1970, García Carrión 2010, Thomas et al. 2012, Guimac Cedillo 2017, Rojas et al. 2017, Quiñones et al. 2018).

1.3 Center of origin of cacao

Field observations suggested cacao may have had a Mesoamerican origin (Mora Urpi 1958, Miranda 1962). Other authors suggested a Mesoamerican for the *Criollo* cacao and a South American for the *Forastero* (Cuatrecasas 1964, Laurent et al. 1994, Whitkus et al. 1998). The main driver for the Mesoamerican hypothesis was the presence of putative native and wild cacao in the Lacandon forest, in Chiapas, Southern Mexico (Miranda 1962, Cuatrecasas 1964). However, the most likely hypothesis is that cacao originated in South America (van Hall 1914, Cheesman 1944, Baker et al. 1954, Schultes 1984). Follow-up studies have aimed to unveil the specific region of origin. In one study, isozyme markers suggested the origin probably took place in Ecuador and Colombia due to higher allelic diversity of samples in comparison to the diversity found in Peru; however, they recognized this might be the result of a sampling bias (Warren 1994). The development of more powerful markers such as microsatellites has helped demonstrate that the cacao cultivated in Mesoamerica, i.e., *Criollo* cacao, came from a few individuals from Southwestern Venezuela in a genetic "founder effect" fashion (Motamayor et al. 2002), supporting the original hypothesis of the South American center of origin. This study also showed the putative wild cacao in the Lacandon forest is of the same genotype as the commonly cultivated *Criollo* cacao, bringing down the former Mesoamerican hypothesis (Motamayor et al. 2002). In Brazil the largest source of genetic variability was found in the Upper Amazonian region compared to the Lower Amazon (Sereno et al. 2006, Silva et al. 2010). Another study that includes most of the countries where cacao is cultivated, found that cacao samples with the highest allelic diversity come from the Upper Amazonian regions of Peru and Brazil (Motamayor et al. 2008). These results were corroborated in another study (Thomas et al. 2012, Osorio-Guarín et al. 2017), which concluded that the Peruvian and Ecuadorian Upper Amazon and the Amazonian region shared by Peru, Colombia and Brazil are the center of origin of cacao. This study also showed that this region harbors the highest *Theobroma* species richness compared to other regions in South and Central America (Thomas et al. 2012).

1.4 Center of domestication of cacao

The origins of domestication and usage of cacao have been the focus of hot debate in recent years. It is widely accepted that cacao was domesticated and first cultivated in Mesoamerica by the Mayans ca. 1500 B.P. (McNeil 2006). However, bioarcheological studies over the last decades have contributed to a revision of this date. First, the search for theobromine, cacao's main alkaloid, residues in Pre-Classic Mayan pottery pushed the earliest cacao usage to 2550-1700 B.P (Hurst et al. 2002, Powis et al. 2002). Then, similar analyses revealed the most ancient usage of cacao in Mesoamerica corresponds to the Olmec society between 3800-3000 B.P (Powis et al. 2008, 2011). The search for even earlier evidence of cacao usage and domestication did not stop. State-of-the-art biochemical, spectrometric and molecular analyses of pottery vessels from the Mayo-Chinchipe culture from the Ecuadorian Upper Amazon revealed not only that cacao was consumed as early as 5450–5300 B.P. but that the earliest known center of domestication was in the Upper Amazonian region of South America as opposed to Mesoamerica (Zarrillo 2012, Valdez 2013, Zarrillo et al. 2018). Therefore, most science now points to both the center of origin and the center of domestication of cacao as having occurred in the same region of South America (Thomas et al. 2012, Zarrillo et al. 2018).

1.5 Historical events in the cultivation of cacao and movements of germplasm

1.5.1 Pre-Columbian Era:

1.5.1.1 Mesoamerica

The introduction of cacao to Mesoamerica has been also a matter of debate. The ancient people from the coast of Ecuador (e.g., the highly skilled Valdivia culture which spanned from 6400-3450 B.P.; Zarrillo 2012) may have influenced the development of civilizations in Peru and Mesoamerica by maritime communication and exchange of products (Wolters 1999). There was a hypothesis that cacao was introduced to Mesoamerica by Valdivia traders via Pacific Ocean routes (Wolters 1999). However, this was rapidly rejected by molecular studies clearly showing *Criollo* cacao in Mesoamerica came from the Maracaibo Basin in Venezuela (Motamayor et al. 2002, 2008) via inland routes and supported by the

fact that colonialists did not find the crop in the Caribbean islands in the years of conquest (Bergmann 1969).

Once *Criollo* cacao reached and spread out in Mesoamerican pre-Columbian societies, it became the top agricultural product for centuries. Its cultivation was mainly restricted to specific districts in the humid and warm lowlands of the Gulf of Mexico (Tabasco, Mexico), Pacific Ocean (Soconusco, Chiapas, Mexico; Suchitepéquez, Guatemala; and Izalco, El Salvador) and the Caribbean (Sula Valley, Honduras) inland coastlines, Lacandon Jungle (Chiapas, Mexico), and Petén Basin (Guatemala), which all underwent intensive cacao production systems; additionally, there were semi-intensive and minor cacao-producing regions scattered from the lowlands of as north as Colima, Mexico to as far south as Quepos, Costa Rica (Bergmann 1969, Whitmore and Turner 1992, Caso Barrera and Aliphat Fernández 2006). Estimates account that cacao-producing areas under the Aztec realm sent annually at least 22 tons of cacao beans to the capital Tenochtitlan as tribute before Spanish conquest (Millon 1955).

1.5.1.2 South America

It is commonly accepted that in pre-Columbian South America cacao was valued in its wild state for the pulp, which was eaten directly or squeezed for juice and fermented beverages, while the seeds were discarded (Sauer 1993, Smith 1999, Clement et al. 2010). This is true for several indigenous groups like the *Machiguengas* in Amazonian Cusco, Peru (Rojas et al. 2017), or the *Tukuna* in the Colombian Amazon (Glenboski 1983), among others (Bletter and Daly 2006). Indigenous groups from the Brazil and French Guiana Amazon may have even started a "proto-domestication" process of cacao, selecting for pulp content, since some researchers claimed to have seen natives planting wild cacao seeds in their villages (Barrau 1979, Clement et al. 2010). These observations may have been the result of the decades of subjection indigenous people suffered by colonialists during the Amazonian cacao boom (Walker 2007). However, an incipient process of domestication in pre-Columbian times is supported by the high levels of homozygosity (the two highest after the homozygosity of the fully domesticated *Criollo* cacao from Mesoamerica) in the

Amelonado and the *Guiana* genetic groups found in these Amazonian regions (Thomas et al. 2012), and by their relative fruit uniformity (Clement et al. 2010, Thomas et al. 2012).

Furthermore, there are numerous accounts that cacao and its seeds were as important, in terms of consumption and spirituality, before arrival of Europeans in South America as they were in Mesoamerica (Wagner 1987, Ogata 2002, Méndez Ramírez 2015). Historical texts by friar Pedro Simon (1627), cosmographer Juan López de Velasco (1574) and bishop Lucas Fernández de Piedrahíta (1881) provide some examples. These documents describe abundance of cacao even higher than in Mesoamerica, extensive plantations and forests of cacao, and the elaboration of a drink called *chorote*, traditional of the natives from the Maracaibo Basin, Northwestern Venezuela, prepared from roasted and ground cacao beans in a different manner than the *chocolatl* drink from Mesoamerica. There is also a government report from 1602 of the discovery of a 100,000-tree plantation managed by natives near the Maracaibo Lake (Reyes and Capriles de Reyes 2000, Carmen 2005). Such abundance of cacao can only be explained with cultivation because in the wild cacao is only found scattered in mixed forests (Allen 1988, Ogata 2002). Also, this cultivation must have been done by the natives; colonialists would have been unable to bring enough African slaves or laborers from Mesoamerica to cultivate such amount of cacao since that region was experiencing labor shortage because of dramatic population reduction caused by lethal diseases (Lovell 1992, Ogata 2002).

The importance of cacao for ancient societies of Ecuador may be also underestimated. Juan López de Velasco in his accounts from the 16th century documented an abundance of cacao in the town of Santiago de las Montañas, in the eastern region of the Loja province (López de Velasco 1574). Additionally the Spanish, during their first exploration journey towards Peru in 1526–27, encountered “blooming plantations of cacao” in current Esmeraldas province, northern coast of Ecuador (Prescott 1847). It is very difficult to corroborate Prescott’s statement because his sources do not refer to cacao (Bergmann 1969). However, pre-Columbian domestication and cultivation of cacao in Ecuador was recently demonstrated in the Mayo-Chinchipe basin (Zarrillo et al. 2018), which make Prescott’s statement more likely. It is also likely that the cacao found by the Spanish in Esmeraldas

corresponded to the *Nacional Ecuatoriano* genetic group, which was introduced from the Upper Amazon to the coast in pre-Columbian times and likely underwent a persistent process of human selection (Lerceteau et al. 1997, Motamayor et al. 2008, Loor Solorzano et al. 2012). Moreover, ancient DNA from Mayo-Chinchipe pottery artifacts revealed the cacao consumed in the Upper Amazon of Ecuador 5450–5300 B.P. are genetically more similar to cultivated *Curaray* and *Purús* groups than to the *Nacional Ecuatoriano* (Zarrillo et al. 2018). Nowadays, *Curaray* and *Purús* groups can be found in the Upper Amazonian region of Ecuador and their geographical range extends to the Upper Amazon in Brazil and to the Northern Amazon of Colombia (Motamayor et al. 2008, Zarrillo et al. 2018). Additionally, these two cacao groups have close genetic affinities to the *Criollo* Mesoamerican group, suggesting they both played an important role in the domestication and further spread of the crop throughout the Amazon Basin and Venezuela in pre-Columbian times (Loor Solorzano et al. 2012, Zarrillo et al. 2018).

Additionally, it is not unreasonable to think that cacao was also important for ancient Peruvian societies. Juan López de Velasco also mentions abundance of cacao in some provinces in "Pirú" without specifying which ones (López de Velasco 1574). However, the most astonishing evidence for cacao relevance in Peru is a 3000-year-old ceramic vessel from the Tembladera people (North coast) depicting two Amazonian organisms: a monkey and what are highly likely cacao fruits (Bonavia 1994, Ogata et al. 2006). This reveals that interactions between coastal and Amazonian peoples in ancient Peru occurred since at least 3500-2900 B.P.; they presumably exchanged cacao among other agricultural goods, which explains the presence of cacao in Tembladera's pottery (Ogata et al. 2006). Another source of support for the relevance of cacao in ancient Peru is the "Chachapoyas" society. It flourished from 800 to 1470 A.D. in the cloud forests of modern Utcubamba province, department of Amazonas, in the junction of the North Andes and the Upper Amazon of Peru (Church and von Hagen 2008, Friedrich et al. 2010). The "Cacao Amazonas Peru" cultivar is native from this region and it is currently highly cultivated with a protected designation of origin (INDECOPI 2016, Guimac Cedillo 2017). It has been described as a "cultural step from the wild cocoas of the neighboring Amazon slope" (Soria V. 1970). During the 15th century, the Incas were trying to conquer the "Chachapoyas" mainly to get

access to many tropical forest resources such as cacao and other agricultural and natural products (Church and von Hagen 2008, Ruiz Estrada 2017). Therefore, it is likely the "Chachapoyas" exploited the "Cacao Amazonas Peru" native cultivar. Coincidentally, the "Chachapoyas" domain is in the Southeastern end of the Mayo-Chinchipe binational basin (Valdez 2013), where the first known center of domestication of cacao is located (Zarrillo et al. 2018). Unfortunately, the "Cacao Amazonas Peru" was not included in the genetic characterization of the more relevant studies of cacao germplasm (Motamayor et al. 2008, Thomas et al. 2012). It is possible the genetic make-up of the "Cacao Amazonas Peru" is close to the ones found in Mayo Chinchipe pottery, i.e., *Curaray* and *Purús* groups (Zarrillo et al. 2018), or to the *Nacional Ecuatoriano*, which was also domesticated in this region (Loor Solorzano et al. 2012).

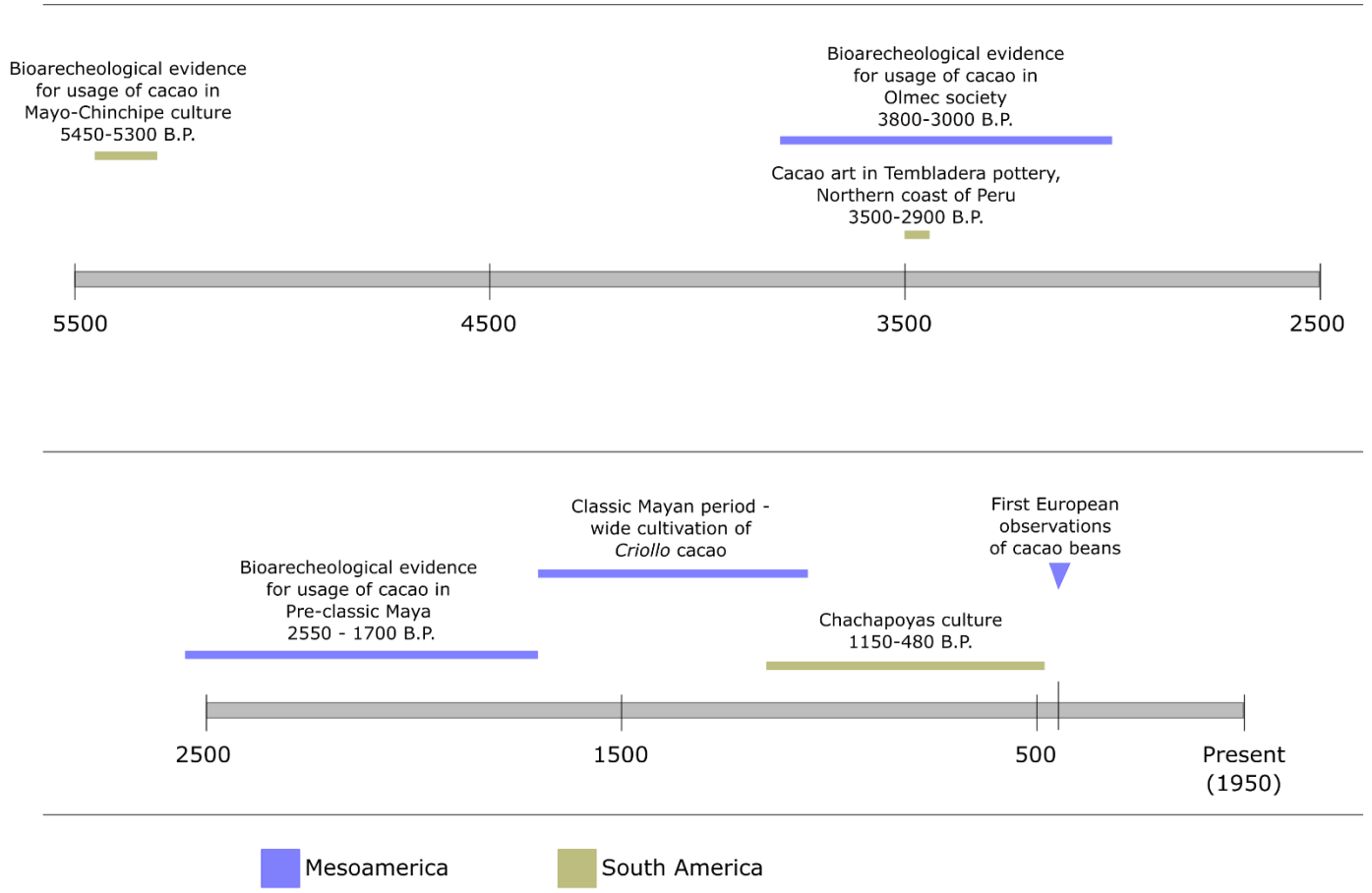


Figure 1.1 Main events regarding cacao usage by Pre-Columbian societies reviewed in the text. Bars represent a time span while the arrow points to a specific year.

1.5.2 Colonial era

Christopher Columbus, his brother Bartholomew, his son Ferdinand and the rest of his crew were the first known Europeans to come into contact with cacao beans (Ferdinand described them as “almonds”) during their fourth voyage to the Western Hemisphere in 1502 (Keen 1992). They observed how indigenous people from the North Coast of present-day Honduras transported, among other goods, high quantities of “almonds”. The Columbus men noticed how the Indians were very diligent with their cargo, but never realized the significance these “almonds” would have in the upcoming centuries (Bergmann 1969, Keen 1992).

1.5.2.1 Mesoamerica

The Spanish promptly became truly interested in cacao after the conquest of Mexico in 1519-1521, when they realized that the Aztec royals and nobles consumed a special beverage, called *chocolatl*, made mainly from the roasted and ground cacao beans (Alden 1976, Young 1994). Once the Viceroyalty of New Spain was established in 1535, the Spanish coerced cacao cultivation by taking over existing plantations, expanding cultivated areas and forcing indigenous people to work in cacao farms to pay imposed levies (Alden 1976). All of this was in response to the increasing local demand for the *chocolatl* drink by the new settlers, and even all classes of indigenous people, once reserved only for the Aztec nobility (Erneholm 1948). An example of the degree of cacao expansion during these times follows. Before the colonial period, in the beginning of the 16th century, exports as a form of tribute from Soconusco to Tenochtitlan were estimated at 5 tons per year (Millon 1955). By the end of the century, the cacao expansion allowed for the existence of 1.6 million trees in the region and exports between 3000 and 6000 loads per year (Gasco 1987), which is equivalent to 75-150 tons by Millon's (1955) calculations. By this time, Spain was the only consumer of cacao in Europe and its consumption would start to popularize in the whole continent by the beginning of the 17th century (Dand 1997).

An unexpected yet terrible population decline occurred during the first decades of the colonial period in New Spain. The population of native people was sharply decreasing

because of infectious diseases; for example in Guatemala, from 2,000,000 people in 1520 to 427,850 in 1550; and in El Salvador, from 400–500,000 in 1524 to 70,000 in 1570 (Lovell and Lutz 1992). In other words, there were many fewer people working on cacao plantations. Therefore, the local supply diminished and the need to import cacao beans increased significantly by the beginning of the 17th century (Erneholm 1948). The main supplier to the viceroyalty during this century was Guayaquil, from where cacao was cheaper, while their highly priced *Criollo* cacao was shipped overseas (Erneholm 1948). This would become the most representative pattern in the cacao import/export situation for the rest of the colonial period in Mesoamerica (Erneholm 1948). For example, by the end of the colonial period, there were approximately 400,000 trees in Soconusco and exports from that region were estimated at 500 loads (12.5 tons), a 80–90% reduction in the total exports compared to the first decades of the colonial period (Gasco 1987).

The Lacandon society was a Mayan-derived group that resisted Spanish subjection until the end of the 17th century. They developed advanced agricultural production systems, and cacao was one of their main products. However, in 1695 the Lacandon society could no longer resist the oppression and were finally conquered by the Spanish, who exiled them from their forest (Caso Barrera and Aliphat Fernández 2006). Thus, their cacao plantations were abandoned and the ancestral knowledge of cultivation in the Lacandon rainforest was forever lost (Caso Barrera and Aliphat Fernández 2006). “Wild” cacao trees in the Lacandon forest have the same *Criollo* genotype and are the remnants from its cultivation in this Mayan society (Motamayor et al. 2002, 2008).

1.5.2.2 South America

1.5.2.2.1 Venezuela

It is debatable whether cacao cultivation started before or after the colonial period (Pittier 1935). Some researchers believe that Venezuelan native people practiced “collection agriculture” in which they only harvested cacao fruits from the wild, and that “plantation agriculture” started with the arrival of the Spanish (Venturini 1983, Reyes and Capriles de Reyes 2000). However, there is strong evidence that cacao was cultivated by native people from Venezuela before the Spanish (Ogata 2002, Méndez Ramírez 2015). Regardless,

Venezuelan cacao, of *Criollo* genetic background (Motamayor et al. 2002), had a reputation of good quality and was highly prized in the European market, and there is no doubt that its cultivation was the main and most dynamic economic activity during the colonial period (Díaz Morales 2000a).

The Maracaibo basin, Northwestern Venezuela, was the first place in South America to export cacao to Europe by the 1560–70s (Erneholm 1948). However, this fertile region was practically forgotten by the Spanish who were more interested in searching for precious metals than working the land; then, the Dutch took advantage of this and thanks to the strategic location of the Curaçao island they dominated the cacao exports from the Maracaibo basin in the 17th century, mostly as contraband (Erneholm 1948, Malavé Mata 1974). On the other hand, East of the Maracaibo basin, specifically in La Guaira, in the old province of Caracas, important plantations were taking form with the work of African slaves and subjected native people (Ferry 1981, McCook 2002b, Delgado C. 2008). Suddenly, during the decades of 1630–40s, Venezuela would suffer its first cacao crisis due to a blight disease referred as the "alhorra" (Ferry 1981). The "alhorra" destroyed most cacao trees in La Guaira and severely affected in all the province (Ferry 1981, McCook 2002b). There were no signs of recovery until the 18th century (Ferry 1981). Maybe, the "alhorra" disease is another reason for Spain to have overlooked cacao cultivation in the Maracaibo basin during the 17th century. Unfortunately, no literature on the symptoms of this disease exist.

In 1706 the Dutch shipped cacao from Maracaibo worth half of the national production to the Netherlands, which warned the Spanish Crown (Israel 1989). The foundation of the viceroyalty of New Granada in 1717 by Spain and the creation of the *Guipuzcoana* company in 1728 were decisive to recover full sovereignty over the Maracaibo basin and the cacao international trade (Malavé Mata 1974, Ferry 1989). The *Guipuzcoana* was the only company allowed to export cacao and other agricultural products to Spain. Because of its rigorous policies, it would double the number of planted cacao trees in the Caracas province, with the Tuy region being the heart of these plantations (Ferry 1989). The crop then became more extensive and the cacao production was fairly constant throughout the

18th century (Malavé Mata 1974, Delgado C. 2008). Smallholder families came together by means of arranged marriages and developed large *haciendas* of cacao (Ferry 1981). For example, in the Coro jurisdiction only there are accounts of more than 1.7 million trees among 168 *haciendas* (de Olavarriaga 1722). From 1700 to 1797 cacao exports increased from 1,500 to 6,750 tons per year (Palma 1953). These prosperous times in Venezuela made *hacienda* families part of the elite and wealthy class of colonial Caracas (Ferry 1989). During the first half of the 18th century, Venezuela was the top cacao supplier in the world and its closest competitor was Guayaquil (Díaz Morales 2000a). After 1765, exports from Guayaquil overpassed Venezuela's (Clarence-Smith 2000).

1.5.2.2.2 Ecuador

The *Nacional Ecuatoriano* cacao was exclusively cultivated during the entire colonial period in Ecuador (Suárez Capello et al. 1993, Loor Solorzano et al. 2012). The Guayas basin was the chief *Nacional Ecuatoriano* cacao-producing region and Guayaquil, the second-biggest port in the viceroyalty of Peru, after Callao in Lima. Because of the inter-colonial trade restrictions ruling in the beginning of the colonial period, the cacao production and exportation system in Ecuador was not very important (Erneholm 1948). It just started to develop by the end of the 16th and beginning of the 17th centuries as the Mesoamerican native population and the local demand were decreasing and increasing, respectively (Erneholm 1948, Clayton 1975, Lovell and Lutz 1992). This made cacao a common and highly priced export to Central America during the early 17th century (Clayton 1975). Just as in Venezuela, cacao plantations were maintained by the subordination of Africans and natives under the *encomienda* system (Bryant 2006). Throughout the 17th century, cacao exports suffered mainly because of pirate attacks, the most important being the one in 1624 by the Dutch which destroyed Guayaquil and paralyzed maritime communications (Erneholm 1948, Clayton 1974, Lavaina Cuetos 2014). Ecuador did not export to Europe because Venezuela had a better geographical position to do so. Exports to Mesoamerica were also constantly under struggle because of the colonial protection policies. Thus contrabands and illegal exports were common during that time (Clayton 1975, Clarence-Smith 2000). Up to the first half of the 18th century,

Ecuador was the second largest cacao producer in the world after Venezuela (Díaz Morales 2000b).

The middle of the 18th century was a time of conflict in Europe because of disputes for possession and redistribution of colonies and territories, culminating in the Seven Years War from 1756 to 1763. The end of this war marked the beginning of a new era in colonial commerce, especially for cacao in Ecuador (Clarence-Smith 2000). In 1765 Spain enacted a decree on inter-colonial trade in which taxes and regulations were greatly softened, which minimized contraband and enabled the beginning of the first cacao boom era of the country, from the 1770s to the 1840s (Alden 1976, Cárdenas Vega 2017). According to some authors, cacao was the economic activity that supported independence movements (Suárez Capello et al. 1993), which finally happened in 1820.

1.5.2.2.3 Brazil

The main economic activity that allowed the Portuguese to settle in Brazil during the 16th century was the production of sugarcane (Prado Júnior 1967). However, the increasing competition in the sugar international trade added to the weakening of the Portuguese empire in India generated an economic crisis in the 17th century (Subrahmanyam 2012, Chambouleyron 2014). On the other hand, Portugal witnessed how cacao was the main engine of the economy in early-17th-century Caracas (Ferry 1989). Therefore, the Crown enacted several measurements to promote cacao cultivation in the current States of Pará and Maranhão (Chambouleyron 2014). The creation of demonstrative orchards in Belém, the largest settlement and port of Northern Brazil at the time, and the license to evangelizing groups to teach cacao cultivation were some of the measurements (Alden 1976). Although there might have been earlier ephemeral plantations, the first official cacao plantations in Brazil took place in 1674, when a Jesuit missionary collected seeds in Pará, current Amazonas State, and planted them in Maranhão, current Pará State (Alden 1976, Chambouleyron 2014). These newly planted trees served as a source for seeds throughout the region in the following years (Chambouleyron 2014). In 1681 cacao cultivation acquired taxation and custom duty benefits which encouraged settlers to cultivate more, thus expansion of cultivation started to appear (Alden 1976,

Chambouleyron 2014). Just as in other regions, expansion implied the need for more people to work, who in this case were almost exclusively Amazonian natives (Walker 2007). It was very common in the 17th century that Portuguese explorers went into the deep Amazon and, allowed by royal policies, ransomed inter-tribal captives to use them as a labor force (Kiemen 1948). Since the Portuguese had noted wild cacao was very abundant along the river banks, "ransom journeys" also served to carry over Amazonian resources like cacao seeds (Alden 1976). These explorations took place as far as current Yurimaguas, in the Peruvian Upper Amazon, according to the journals of Father Samuel Fritz in 1695 (Edmundson 1922).

Despite promotion of cacao cultivation, the crop did not become an important export because most of the beans came from wild trees, which would only offer one harvest per year and usually yielded lower quality beans compared to their cultivated counterparts (Alden 1976). Additionally, spoilage of improperly dried cacao beans during transport to Belém and shipping to Europe was another important limiting factor (Alden 1976, Miller and Nair 2006). By the 1730s cultivation of cacao was becoming well adopted and expeditions to collect wild cacao were less frequent (Alden 1976, Miller et al. 2006). Since then there were steady and increasing exports of cacao beans which averaged 600 tons/year during the prosperous and revolutionary Pombaline's regime, 1750–1777 (Hemming 2008). This period is referred to as the Amazonian cacao boom of the 18th century (Alden 1976) and, inadvertently, may mark the beginning of dispersal of one of the most devastating diseases of cacao, witches' broom disease (WBD). Up to this point, the cacao cultivated throughout the current states of Amazonas and Pará was from the *Amelonado* genetic group (Motamayor et al. 2008, Thomas et al. 2012).

During the 1780s, the naturalist and explorer Alexandre Rodrigues Ferreira set Amazonian expeditions to study the flora and fauna of the region. In one of them, along the Negro river close to the current municipality of Barcelos, Amazonas, he encountered growers complaining about a cacao disease referred to as the *lagartão* (lizard in Portuguese), which was spread out in neighboring provinces and able to kill cacao trees after two years of planting (Ferreira 1786). Researchers strongly believe this *lagartão* disease to be WBD

because infected branches adopt the shape of a lizard, suggesting Ferreira was the first to document WBD (Viera 1942, Silva 1987).

Cacao cultivation in Brazilian Amazon expanded by approximately two-fold increase during the last decades of the colonial era; exports account for 753 tons in 1775 to 1678 tons in 1818 (Erneholm 1948). Cacao cultivation did not have so much chance to expand because of the Amazonian rubber tree boom (Barham and Coomes 1994). Therefore, its expansion had to move southeast, to the state of Bahia. By 1746, cacao seeds from Pará (*Amelonado* genotype) were already introduced to Bahia, in the current municipality of Canavieiras, and gradually disseminated throughout the state, but no official exports from Bahia were registered in the colonial period (Erneholm 1948, Walker 2007).

1.5.2.3 The Caribbean

The history of cacao in the Caribbean during colonial times would have a tremendous impact on the cultivation of the crop throughout the world for the next centuries. The genetics for the most widely used cultivars of cacao, *Trinitario*, was born in this region. Here we will focus on Trinidad and Jamaica, the two largest producers in the region.

The earliest records of cacao in Trinidad date back to a Spanish introduction in 1525 (Dand 1997) but commercially its production started around 1678 with *Criollo* material brought from Venezuela (Knapp 1920, Shephard 1932, Bekele 2004, Leiter and Harding 2004). *Criollo* was the only cacao genetic group and the only exported product in Trinidad by the beginning of the 18th century (Cuatrecasas 1964, Motilal and Sreenivasan 2012). Similarly, the Spanish introduced cacao to Jamaica around 1638-1640 (Fagan 1984, Dand 1997) with seeds from Guatemala and likely Caracas (Momsen and Richardson 2009), i.e., seeds of *Criollo* genetic background. Its production was continued by the English, after they got possession of the island in 1655 (Gardner 1873, van Hall 1914). However, cacao cultivation was not successful in the following years with reports of unknown “blasts” frequently registered in the literature starting in 1664 (Motilal and Sreenivasan 2012). By 1670 there were around forty-seven plantations yielding about 94 tons and by 1772 there were zero plantations (Gardner 1873). This can be explained by a disease outbreak in the

1720s that affected severely cacao cultivation in the Caribbean, from Trinidad to Jamaica (Bartley 2005, Motilal and Sreenivasan 2012), historically referred to as the "blast" disease (Motilal and Sreenivasan 2012).

Several hypotheses to explain the "blast" have been proposed (Leiter and Harding 2004, Motilal and Sreenivasan 2012). Religious tithes and astronomical reasons (Joseph 1838, Millas 1968) are the less accepted ones; hurricanes, although no hurricanes seem to have hit Trinidad in those years (Millas 1968); and pathogens, *Phytophthora* spp. (steam canker and black pod rot) and *Ceratocystis cacaofunesta* (Ceratocystis wilt) being widely accepted (Newson 1976, Leiter and Harding 2004), are among the causes most discussed by researchers. Motilal and Sreenivasan (2012) proposed a model to explain the "blast" that combines the poor performance of the *Criollo* cacao under eroded soils, bad agronomic practices and generalized low temperatures caused by the Little Ice Age. However, no other tropical crops in Trinidad and in the Caribbean seem to have suffered low temperatures; conversely, sugarcane production experienced a sharp increase during the entire 18th century and the Caribbean became the top world supplier (Galloway 1989). Additionally, the coldest temperatures observed during the Little Ice Age took place between 1400 and 1700 (Mann et al. 2009), which does not overlap with the cacao "blast" times in the Caribbean. Thorough population genetics studies of *Phytophthora* spp. and *Ceratocystis cacaofunesta* will reveal the evolutionary and dissemination history of these pathogens and maybe provide hints into the real cause of the "blast" of the 18th century in the Caribbean.

The years following the "blast" were adverse in Trinidad: cacao production stopped, small pox spread out, and people of all classes started to leave the island (Dand 1997, Leiter and Harding 2004, Momsen and Richardson 2009). There are accounts that the plantation of a farmer who cultivated a "hardier" cacao but of lower quality survived the "blast" (Joseph 1838, Dand 1997). If this was the case this resistant cacao must have been of a *Forastero* genetic background (Dand 1997). Nevertheless, there are no accounts of *Forastero* introduction to the country until 1756, which were made by Capuchin missionaries with seeds from the Amazon basin (Bartley 2005). Thanks to this foreign introduction cacao started again being cultivated (Leiter and Harding 2004). During the 19th century

Venezuelan migrants from the Paria peninsula, known as the Cocoa Panyols, would lead the resurgence of cacao cultivation in the island (Moodie-Kublalsingh 1994). This golden era of cacao in Trinidad resulted in the natural development of *Trinitario* cacao. All the events described are consistent with the hypothesis of its genesis: *Criollo* cacao from the Paria peninsula naturally hybridized with an early introduced *Forastero* cacao (Ciferri 1949). It was demonstrated that this *Forastero* corresponds to the *Amelonado* genetic group, highly disseminated in the Amazon basin around the same time Capuchin missionaries took cacao seeds to Trinidad (Alden 1976, Motamayor et al. 2003, Bartley 2005).

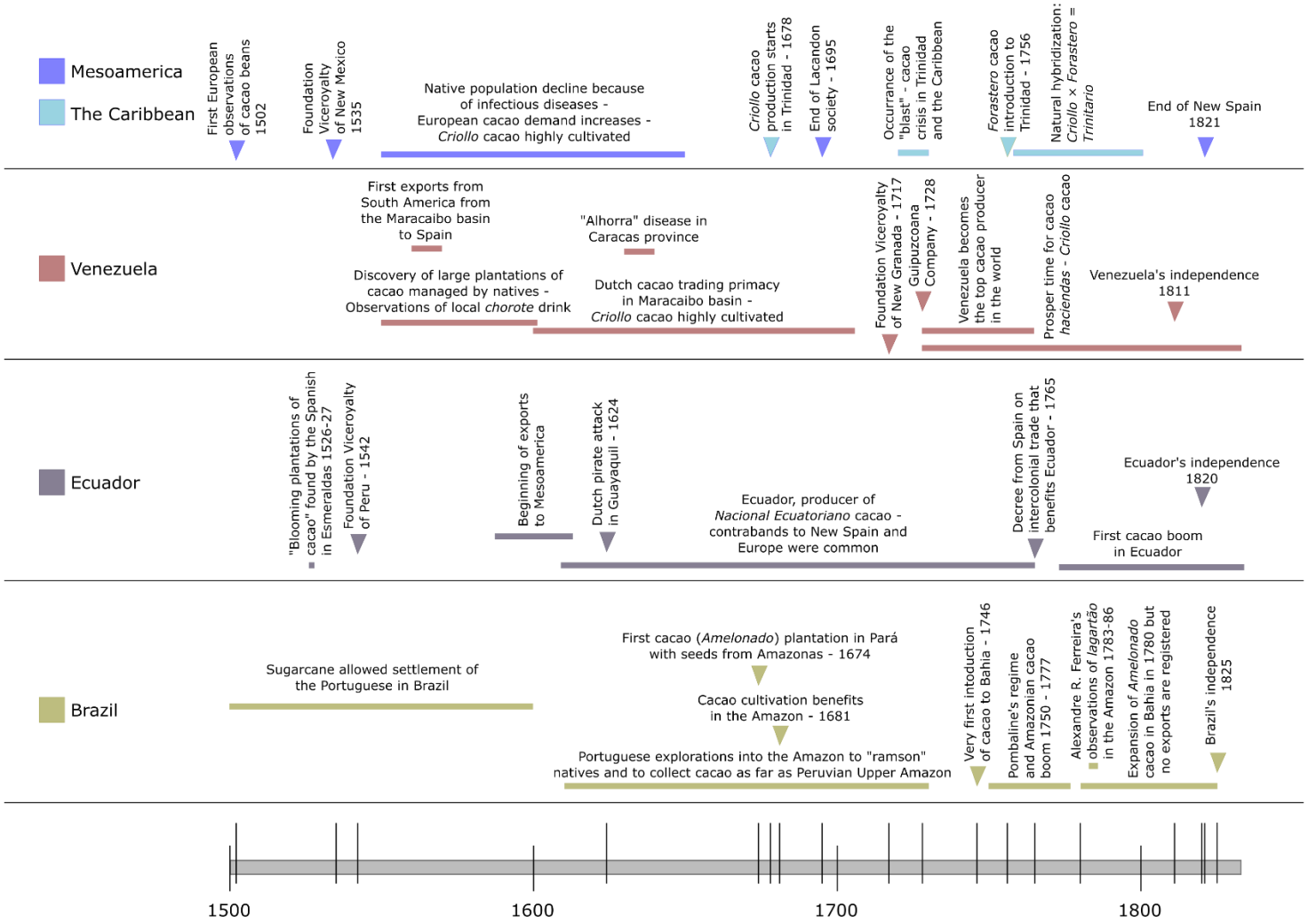


Figure 1.2 Main events regarding cacao cultivation in colonial times reviewed in the text (from 1500 to the 1820's). Bars represent a time span while arrows point to a specific year.

1.5.3 The republican era

Once the colonial times ended during the first decades of the 19th century, the now independent nations would experience different circumstances that had an impact on cacao cultivation. Here we will focus on Venezuela, the country which suffered the most from the consequences of independence, and Ecuador and Brazil, the top producers of this era in South America.

1.5.3.1 Venezuela

By the time Venezuela acquired independence from Spain in 1811, national cacao production was already surpassed by Ecuador and this figure never changed. By 1825 cacao of the *Criollo* genotype group was the most frequent, if not the only one, cultivated cacao in the country (Ciferri 1949, Reyes and Capriles de Reyes 2000); i.e., cacao from the same genetic background was leaving the door open for another disease epidemic. This indeed occurred at the beginning of the 19th century; some disease struck *Criollo* cacao in Tuy, spread throughout the Caracas province and virtually eliminated cacao from that area (Palma 1953, Reyes and Capriles de Reyes 2000). Unfortunately, the latter references do not specify which disease, and no other sources were found. As a consequence, between 1820 and 1830 *Trinitario* cacao, of lower quality than *Criollo*, was brought to the province of Caracas and expanded East, to the Paria peninsula (Pittier 1935, Palma 1953). Unfortunately, these introductions did not improve the overall situation during the rest of the century because of the aftermath of independence wars and the Federal War (1859-1863). Many cacao plantations were physically destroyed by combating troops and many others were either abandoned or replaced with the more-affordable-to-start coffee (Reyes and Capriles de Reyes 2000, Arroyo Abad 2013). Additionally, *hacienda* owners broke after the abolishment of slavery in 1854 (Delgado C. 2008, Arroyo Abad 2013). All these negative circumstances in Venezuela were exploited by Ecuador, which from 1817 to 1842 became the number one cacao world exporter (McCook 2002a).

At the beginning of the 20th century, the profitable petroleum industry caused a massive migration from rural areas and thus a decline in not just cacao production but the whole agricultural system (Malavé Mata 1974, Karl 1997, Quevedo C. 1998, Schiavoni and

Camacaro 2009). By 1937 WBD had caused the demise of nearly all plantations in the Northeastern regions of the country, and by 1941 a second major disease, frosty pod rot (FPR), appeared in the Zulia state, Northwestern Venezuela, which worsened the national situation (Müller 1941, Reyes and Capriles de Reyes 2000, Parra et al. 2009). Because of abandonment of plantations, other minor diseases such as Diplodia pod rot also became important (Müller 1941). To mitigate these negative impacts, in 1956, the government created the FNCC (National Cacao and Coffee Fund), to promote both crops and to manage their commercialization (Díaz Morales 2000b). In 1975, through an executive order, FNCC split into two independent bodies: FONCACAO (National Cacao Fund) and FONCAFE (National Coffee Fund). FONCACAO was the only institution allowed to collect and buy cacao beans from producers and to manage their exportation; i.e., the cacao trade became monopolized (Díaz Morales 2000b). FONCACAO collapsed in the 1990s because poor management, lack of promotion of added-value, delayed payment to producers, and migration, among other factors (Díaz Morales 2000a, Schiavoni and Camacaro 2009). During FONCACAO years, national cacao production and harvested areas experienced their lowest levels in their most current history (Figure 1.3). There were no resources to combat diseases; for example, in the region south to Maracaibo Lake, only 28% of producers took some action against these diseases which directly affected their yields (Portillo et al. 1995).

During the late 2000s, governmental measurements to favor cacao producers were taking place; this included promotion of micro-enterprise, low-interest credits, technical support, facilities, etc. (Schiavoni and Camacaro 2009), which translated in Venezuela to the highest national production by 2012 for the last sixty years (Figure 4). However, by that time there were some cacao areas, such as the Orinoco Delta in Eastern Venezuela, suffering from government indifference and high incidence of diseases (Rendiles et al. 2009). Farmers from Sucre revealed they do not apply disease management strategies because they do not have the resources (Lanz and Granado 2009). Consequently, cacao production nowadays seems to experience another drop (Figure 1.3).

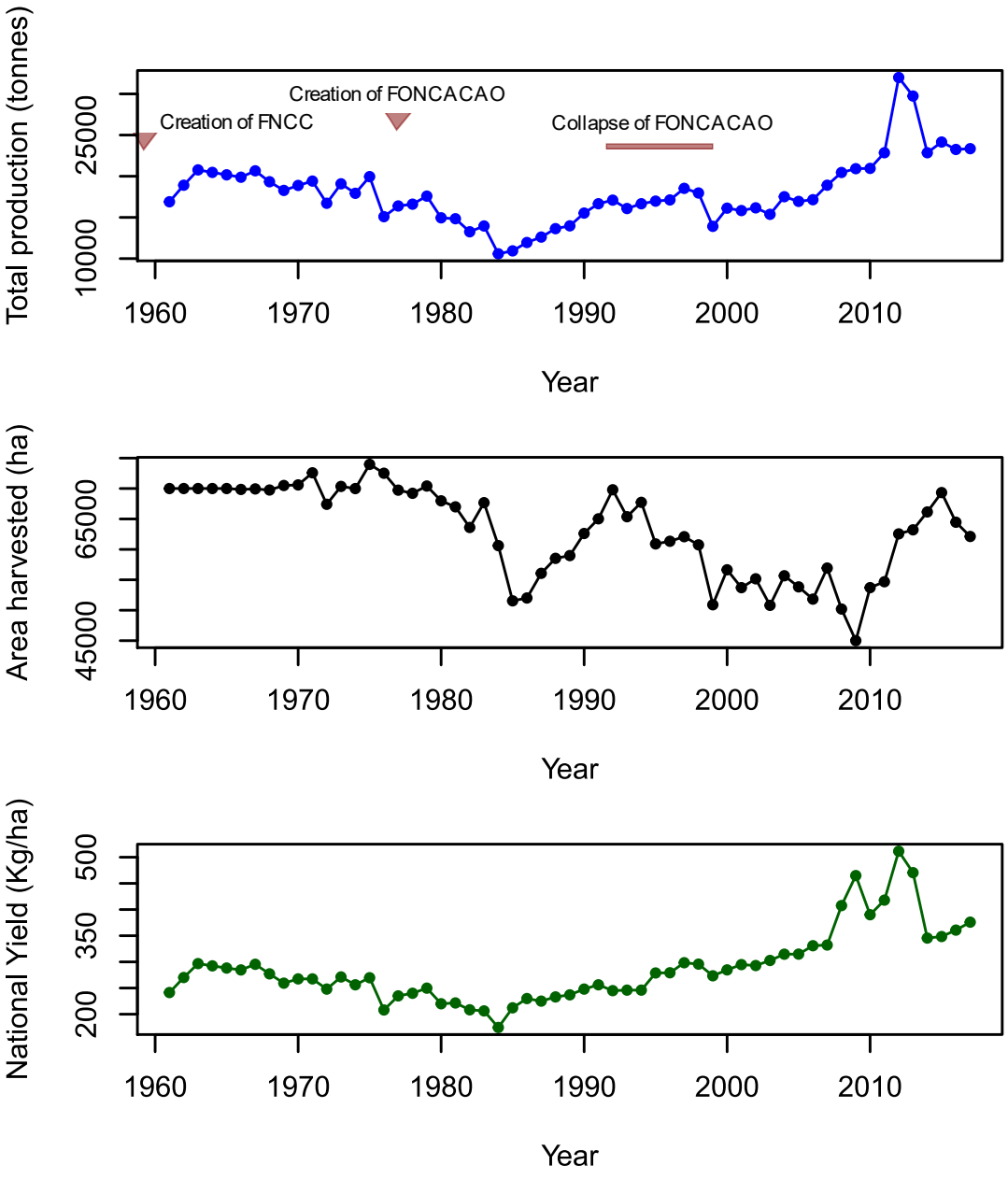


Figure 1.3 Decline of the harvested area of cacao in Venezuela because of the poor management and eventual collapse of FONCACAO. Data contains official, semi-official and estimated values from FAOSTAT (FAO 2018)

1.5.3.2 Ecuador

Unlike in Venezuela, Ecuador's independence in 1820 brought many economic benefits to the country. Guayaquil trading ships no longer had to transit by Callao so cacao commerce became fully liberated (Gondard 1986). Subsequently Ecuador was the top world producer of cacao until 1842, the year of the yellow fever epidemic, which reduced the Guayaquil population by half (Connor 1920, Gondard 1986). The following three decades were very difficult not just for the cacao sector but for the entire country.

During the second half of the 19th century, the second cacao boom in Ecuador took place (1880–1910). Many factors contributed to the country's economic resurgence (Hamerly 1978). During the 1840s chocolate candy was invented and released to the European market, and suddenly the global demand for cacao was drastically increased (Dand 1997). In 1851 and 1857, African slavery and the oppressive "indian tribute" imposed in the first republican years were fully abolished, respectively (van Aken 1981, Sattar 2007, Valencia Rodríguez 2007), and for the first time in Ecuador's history, labor people were well paid (McCook 2002a, Chiriboga 2013). From 1885 to 1905 the world's cacao production increased by 257% (from approx. 40,600 to 145,600 tons) (McCook 2002a). Enthusiasm among Ecuador's cacao *hacienda* owners increased so they started to take decisions to maximize income. They acquired more land to expand cultivation (Mckenzie 1994) and in 1880 introduced the promising *Trinitario* varieties from Trinidad and Venezuela (Erneholm 1948, McCook 2002a). The *Trinitario* material had high yields but didn't have the high quality beans of the *Nacional Ecuatoriano* (Erneholm 1948). *Hacienda* owners then decided to sacrifice quality for quantity by disseminating *Trinitario* cacao cultivation, which allowed them to accumulate large fortunes (McCook 2002a, Chiriboga 2013). The crisis situation in Venezuela (as previously described) also helped Ecuador to become the biggest world supplier of cacao of the century (McCook 2002a). Some studies even suggest that Ecuadorian cacao exports increased by 340% from the 1870s to 1914 (Clarence-Smith 2000). Unfortunately, soon the country's export incomes relied on the world cacao market, which put Ecuador into a vulnerable position (Pineo 1988). Additionally, sanitary practices were not widespread among growers, which would facilitate disastrous consequences in the following years (Pineo 1988).

Ecuador lead the world cacao production until 1905–1912 when it was reached by São Tomé and Príncipe, Ghana, and Brazil (Clarence-Smith 2000). However, the second cacao boom in Ecuador was brought down when FPR disease was first reported in Esmeraldas in 1911 and made its appearance in the form of an epidemic in 1917-18 in Los Ríos province and Guayas basin (Ciferri and Parodi 1933). FPR had already been seen in Santander, Colombia for the first time in 1817 and multiple times throughout cacao- producing regions of the country but the disease did not cause as large an impact in Colombia in the 19th century as it did in Ecuador in the 20th century (Phillips-Mora 2003). Additionally, in 1918 WBD was discovered in Ecuador in the Guayas basin (Pound 1938) and together with FPR caused the biggest decline of Ecuador's cacao production. National production went from 50,000 tons in 1915 to 20,000 in 1925 (Thorold 1975). This created an unexpected national economic crisis and subsequently social upheavals like the one in Guayaquil in 1922, among the worst in Ecuador's history (Pineo 1988, Henderson 1997). During these years, the cacao cultivation system was highly diminished and some growers even started to switch to other crops like bananas and coffee (Larrea and North 1997, Striffler 1999, Grimes 2009). The monetary incentives from the government to find solutions against cacao diseases led to the collection and identification of cacao germplasm from the coast of Ecuador with some levels of resistance to WBD; this germplasm is known as "Refractario" (Pound 1938). Researchers noticed that this "Refractario" material did not belong to the *Nacional Ecuatoriano* genetic group but to the *Trinitario*, which was introduced to the country since 1880 (Pound 1938, Erneholm 1948). Additionally, Pound (1938) found that highly resistant "Refractario" cacao was very similar to material he found throughout the region of the Napo river, in the Ecuadorian Amazon. He then proposed an earlier introduction of Amazonian cacao to the coast of Ecuador (Pound 1938).

Cacao collections along the Ecuadorian territory continued during the 1930s and 40s and have been maintained in the two biggest germplasm banks of the country, the "Estación Experimental Tropical Pichilingue" and "Centro de Cacao Aroma Tenguel" (Quiroz V. 1997, Loor et al. 2009). The former was originally a private cacao farm abandoned during the disease outbreak of the 1920s and became a possession of the Government for research on cacao and other tropical crops since 1942 (Cuvi 2009). The latter mainly contains

Nacional Ecuatoriano germplasm that started as private efforts from the United Fruit Company to preserve this fine flavored cacao (Loor et al. 2009); it currently belongs to the "Universidad Técnica Estatal de Quevedo" (Carranza Patiño et al. 2008, Haz Alvarez and Cabrera Vicuña 2010). Additionally, the work of a single individual, a young breeder named Homero Castro Zurita, had a significant impact on cacao production in Ecuador in the next decades. During the 1950s Castro Zurita performed expeditions to the Canelos valley in the Ecuadorian Amazon and incorporated cacao material into his own collection in Naranjal, known as the "Colección Castro Naranjal" or CCN (Crespo del Campo and Crespo Andía 1997); the CCN was also composed of the widespread ICS (*Trinitario*) and IMC (*Iquitos*) germplasm (Iwaro et al. 2003, Boza et al. 2014). During the 1960's one of his crosses (ICS-95 × IMC-67) × "Oriente 1" (from Canelos Valley) resulted in the generation of the clone CCN-51 (Boza et al. 2014).; although the passport information of "Oriente 1" was lost, the genetic composition of CCN-51 corroborates Castro Zurita's crosses: 45.5 % *Iquitos*, 22.2% *Criollo*, 21.5% *Amelonada*, 1.1% *Nacional Ecuatoriano* and the rest a mix of other *Forastero* genotypes (Boza et al. 2014). The beans from CCN-51 resulted in a quality lower than the *Nacional Ecuatoriano*'s (Afoakwa et al. 2008) but also had many agronomic advantages: it was early maturing, highly productive and showed high levels of disease tolerance (Crespo del Campo and Crespo Andía 1997). Subsequently, CCN-51 gradually gained farmers' preference despite of being considered "bulk cacao" and despite *Nacional Ecuatoriano* cacao is highly appreciated in the fine aroma international market and commands a premium price (Flores González 2007, Jimenez et al. 2018). Extensive cultivation of CCN-51 started to appear since 1985 (Crespo del Campo and Crespo Andía 1997). Recently, CCN-51 cacao exports have sharply increased from 5.8% of total national exports in 2005 (Flores González 2007) to 30% in 2015 (Moncayo R. 2016). This planting of CCN-51 has brought many inherent issues: from the conservation perspective, the unavoidable natural hybridization of *Nacional Ecuatoriano* and CCN-51 genotypes (and also *Trinitario* germplasm introduced since 1880) has put at risk the precious *Nacional Ecuatoriano* genetic integrity (Loor et al. 2009, Loor Solorzano et al. 2012); from the food industry perspective, adulterations of *Nacional Ecuatoriano* cacao bean cargos with CCN-51 beans are frequently reported which has generated the development of molecular and computational imaging methods to detect them (Herrmann

et al. 2014, Jimenez et al. 2018); and from the biodiversity perspective, CCN-51 is mainly cultivated as a monocrop requiring a lot of agronomic inputs while *Nacional Ecuatoriano* is produced under agroforestry systems (Bentley et al. 2004). Additionally, most of the fine aroma cultivated cacao are actually hybrids between *Nacional Ecuatoriano* and *Trinitario* that maintain fine aroma characteristics and only 1% is pure *Nacional Ecuatoriano* (Bentley et al. 2004). Nowadays, the national priority is to protect the *Nacional Ecuatoriano* germplasm (Flores González 2007, Loor Solorzano et al. 2012, Melo and Hollander 2013) as it represents 66% of world fine aroma cacao (Fountain and Huetz-Adams 2018). Moreover, CCN-51 reached other important fine aroma cacao producing countries like Peru. By 2011 there were 45,445 ha (53.6% of national total) under cultivation of exclusively CCN-51 cacao in this country (García Carrión 2010).

1.5.3.3 Brazil

During the first decades of the republican period the Lower Amazon (Maranhão and Pará) was still the main cacao-producing region in Brazil. By 1880, exports from there were twice as big as exports from Bahia: 3,121 vs 1,668 tons; however, this situation would drastically change by 1900, when exports from the Amazon were four times smaller: 3085 vs 13,131 tons (Walker 2007). Unlike in the Amazon, the main labor force in Bahia were African slaves, until 1888, year of the abolition of slavery (Martin 1933). Cacao then represented the door to economic independence for the now ex-slaves in Brazil; many would move to Bahia and encounter large uncultivated areas ideal for the cultivation of the highly demanded cacao (Mahony 2008). Additionally, since 1888 small farmers developed into large *haciendas* because of the abundance of poor available workers (Erneholm 1948, Walker 2007). Therefore, the production of cacao in Bahia started to rapidly expand: from 1895 to 1941-45 total production went from about 6,000 to 125,000 tons, i.e., a 1900% increase in fifty years! (Erneholm 1948). Brazil's highest ever production peak occurred in 1986 with 459,477 tons (Figure 1.4). The only genotype of the cacao widely cultivated in the history of Brazil has been *Amelonado*, introduced since 1746 to Bahia (See Colonial Era). There were some attempts to introduce and disseminate *Criollo* cacao (high quality) in Bahia in the 1920s but they were never successful (Erneholm 1948).

While Ecuador was dealing with a cacao crisis due to FPR and WBD, Brazil was actively competing with the increasing production in West African countries (where *Amelonado* cacao has been the extensively cultivated one; Aikpokpodion et al. 2009), mainly São Tome and Principe and Ghana in the first half, and Ivory Coast in the second half of the 20th century (Dand 1997, Leiter and Harding 2004). Bahia, where most of the cacao was produced in the country, was seemingly protected by the Amazonian natural barrier but concerns of potential introductions started to appear since the early 1980s because of the expansion of cultivation in the Amazon and frequent communications back and forth with Bahia (Rocha 1983). Quarantine controls were established but the imminent spread finally occurred in 1989 (Periera et al. 1990, Pereira et al. 1996). The arrival of WBD to Bahia caused a decline in national production that even nowadays it has not been able to recover (Figure 1.4).

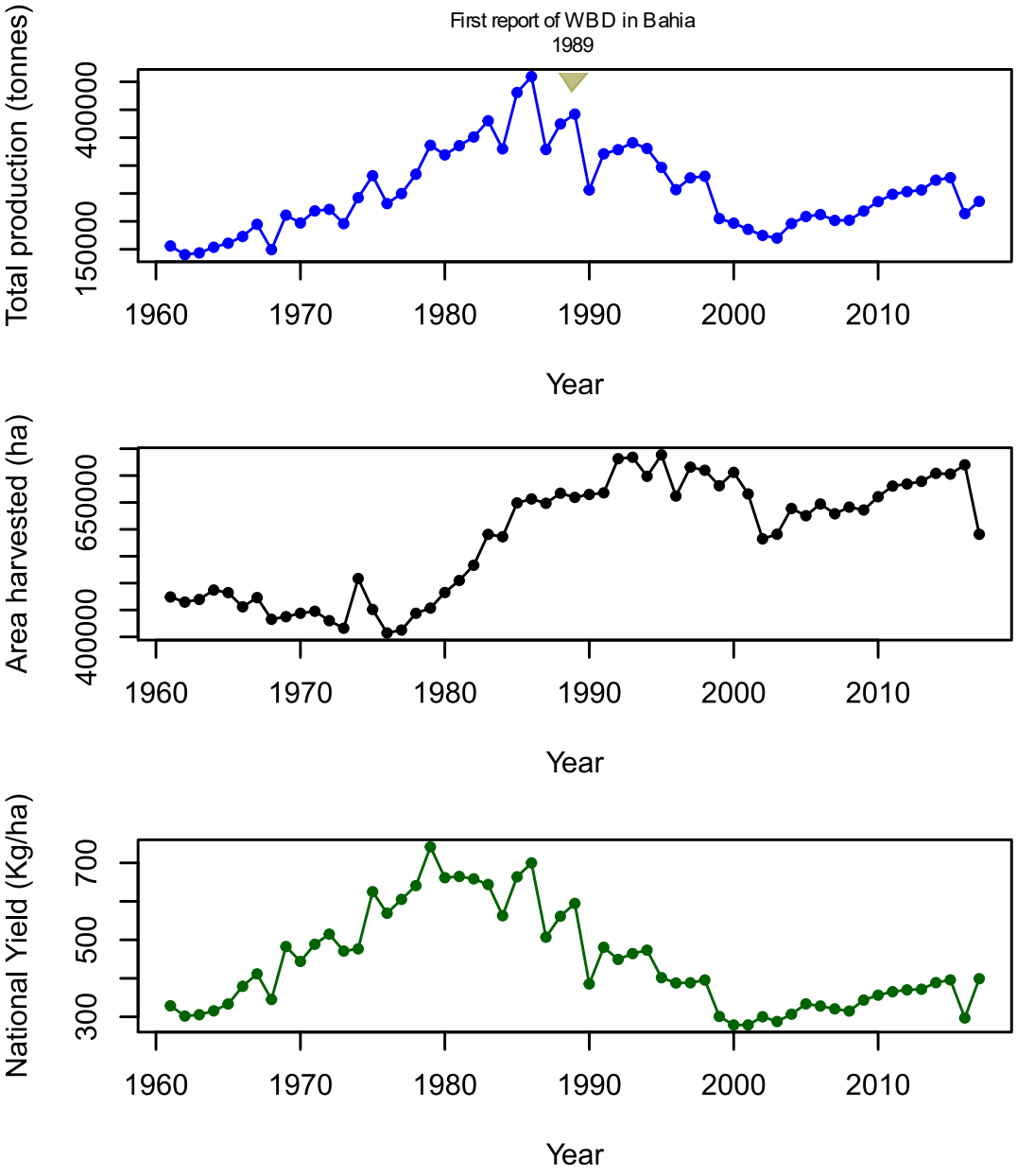


Figure 1.4 Decline of national cacao production in Brazil since the appearance of witches' broom disease (WBD) in Bahia. Data contains official, semi-official and estimated values from FAOSTAT (FAO 2018)

1.5.3.4 The Caribbean and other countries

Most of the Caribbean countries remained European colonies during the 19th and the first half of the 20th century. However, some important events regarding cacao production that occurred in this region would have an effect on the rest of the countries. In 1895, WBD was first observed in Suriname (Stahel 1915) and spread to other countries, being the most relevant Ecuador in 1918 (see Ecuador's section) and Trinidad in 1928 (Laker et al. 1988). Because of the latter invasion, the major expeditions into the Upper Amazon to search for disease resistant cacao germplasm took place (Pound 1938, 1943), which were led by F. J. Pound, researcher from the then Imperial College Station (ICS). This resulted in the creation of the International Cocoa Genebank, Trinidad (ICG,T). Additionally, other collecting expeditions were performed throughout the entire Amazonian region (Allen 1987, Zhang and Motilal 2016).

On the other hand, cacao production in Mesoamerica suffered because of FPR. This disease progressively took over plantations since the 1950s, starting with Panama in 1956 (Phillips-Mora and Wilkinson 2007). Currently, the disease can be found in all countries of Mesoamerica and in Jamaica, the only country in the Caribbean reported to have FPR (Johnson et al. 2017). Among the countries more affected by FPR has been Costa Rica, once the top producer of Mesoamerica at the beginning of the 20th century (Clarence-Smith 2000). The first exports from Costa Rica in the Republican era started in 1884 with the very modest amount of four tons (Clarence-Smith 2000). However, by 1977 its exports increased to 5,719 tons (Enríquez and Suárez 1978). Unfortunately, FPR arrived to Costa Rica in 1978 (Phillips-Mora and Wilkinson 2007) and caused a devastating decline in cacao production (Figure 1.5). Major breeding efforts in Latin America rely greatly on the germplasm collected during the expeditions of the 20th century. Unfortunately, resistant cacao cultivars to the major diseases, WBD and FPR have not been developed yet (Phillips-Mora et al. 2013).

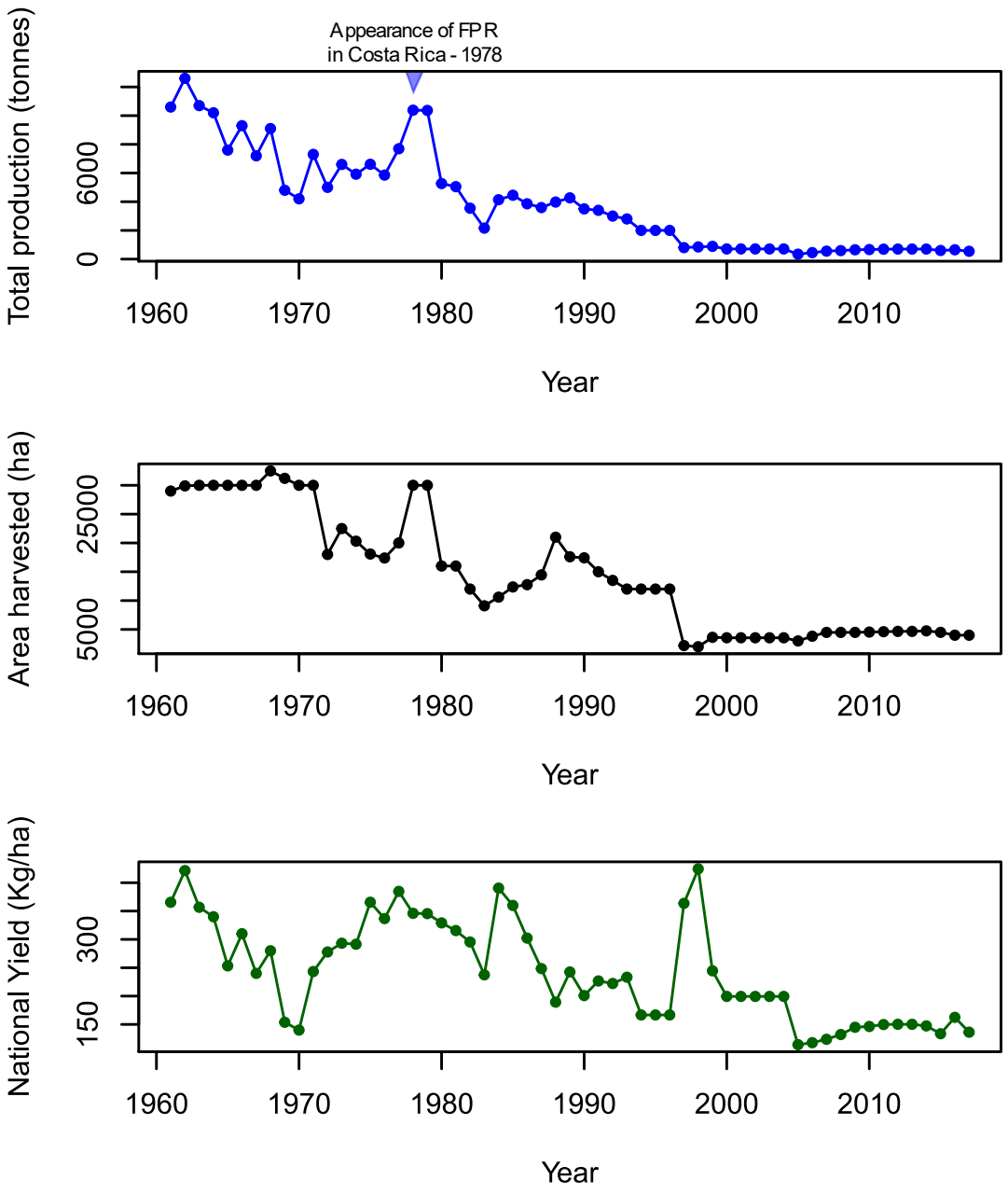


Figure 1.5 Decline of national cacao production in Costa Rica since the appearance of frosty pod rot (FPR) in the country. Data contains official, semi-official and estimated values from FAOSTAT (FAO 2018)

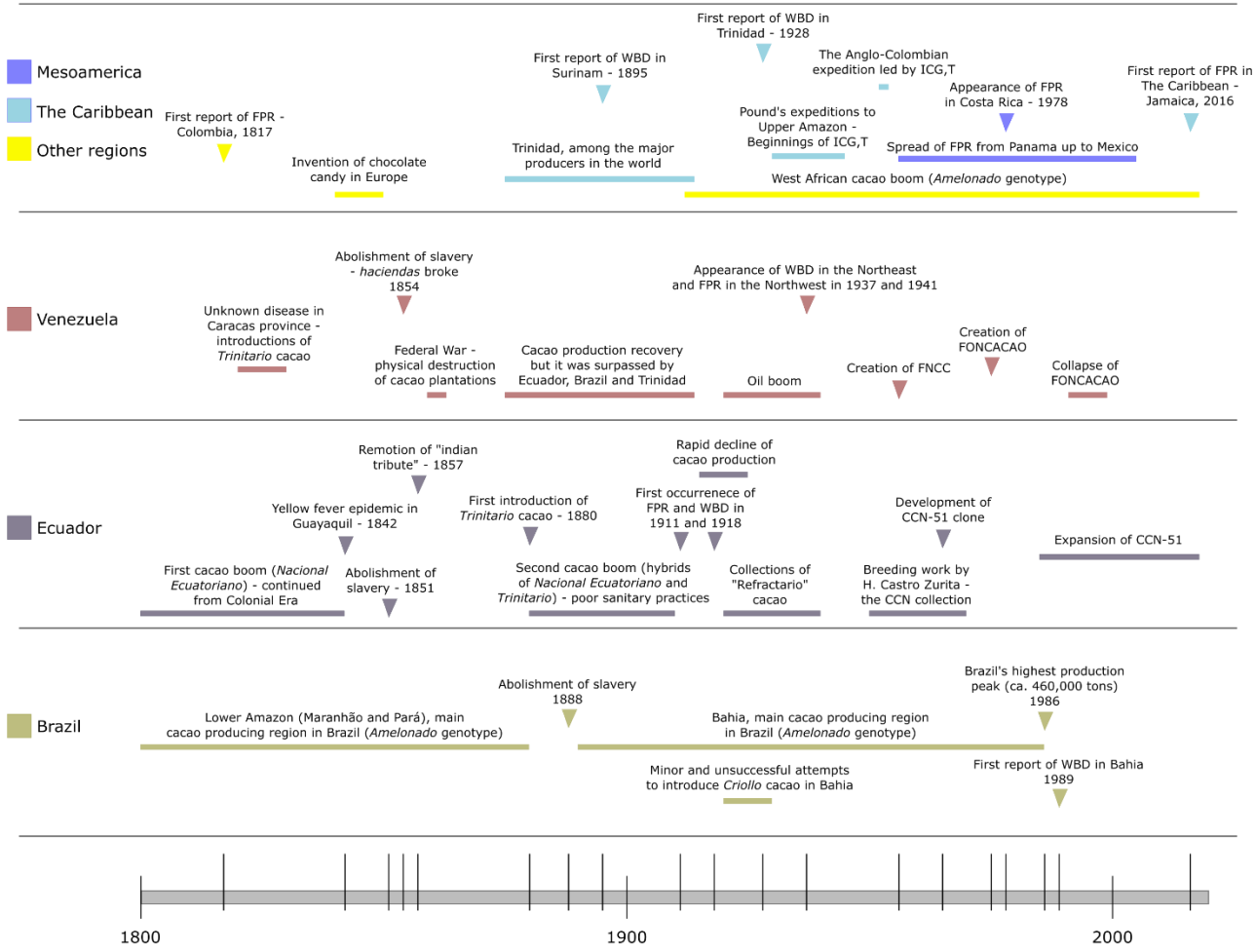


Figure 1.6 Main events regarding cacao cultivation in the Americas from 1800 to 2016.

1.6 Current situation of cacao cultivation

Currently, the majority of world cacao is produced out of its center of origin (Figure 1.7). West African and Southeast Asian countries produce approximately 67% and 17%, while countries in the Americas produce 16% of world cacao (Figure 1.7). Additionally, in the middle of the 19th century, around 95% of cacao produced in the world was of fine quality (Erneholm 1948); nowadays it is exactly the opposite, 95% of world cacao is of low quality and a scarce 5% is fine cacao (Melo and Hollander 2013). Based on the historical movements of germplasm, in absolutely all the largest producing countries the genetic background to resist disease epidemics is very low, even the ones located in its center of origin. This means that cacao diseases will continue to diminish cacao yields at least for the near future. Currently, between 80 to 90% of world cacao is produced by about six million small-holder farmers in plantations averaging 2-4 hectares (Beg et al. 2017, Wickramasuriya and Dunwell 2018). Therefore, the goal is that cacao breeding programs can find long-lasting solutions for cacao diseases and in this way avoiding giant crisis, as we have seen at least once per century, since colonial times.

World Region / Country	Production (tons)	Percentage
Africa	20,921,342	67.4
Asia	5,182,314	16.7
Brazil	1,759,245	5.7
South America except Brazil	2,192,898	7.1
Mexico, Central America and The Caribbean	963,644	3.1
Total	**31,019,443**	**100.0**

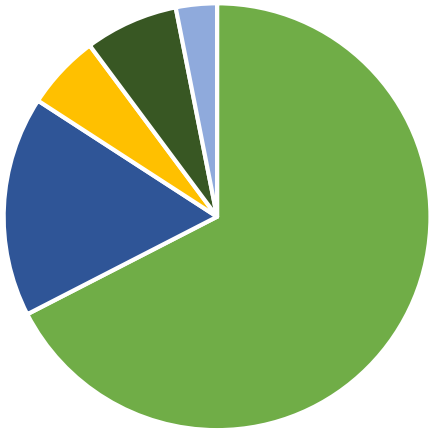


Figure 1.7 World cacao bean production from 2010 to 2016. Data contains official, semi-official and estimated values (FAO 2018)

CHAPTER 2. A REVISION OF THE CENTER OF ORIGIN OF *MONILIOPHTHORA RORERI*

2.1 Introduction

The center of origin of cacao (*Theobroma cacao* L.) is in the Upper Amazonian regions of Peru, Ecuador, Colombia and Brazil (Motamayor et al. 2002, 2008, Silva et al. 2010, Thomas et al. 2012, Osorio-Guarín et al. 2017); the earliest evidence for its use and domestication is in the Mayo-Chinchipe basin in the Upper Amazon of Ecuador (Zarrillo 2012, Valdez 2013, Zarrillo et al. 2018). In other words, the center of origin and domestication of cacao points to the same region in South America. Throughout history, the cacao crop has undergone an inter-continental dissemination in response to the global demand for chocolate (Chapter 1). However, its cultivation is limited to the usage of selected clone cultivars that have a narrow genetic basis (Phillips-Mora et al. 2009, 2013). This clonal propagation of cacao to areas outside its center of origin has decreased the genetic diversity of the crop which subsequently increased the risk of diseases and made the development of fully disease-resistant cultivars difficult (Brown and Hovmøller 2002). Among the most important cacao diseases in the Americas is frosty pod rot (FPR), caused by the fungus *Moniliophthora roreri*, which has been considered among the most threatening plant pathogens in the world since the 1970s (Thurston 1973). Before the 1950s, the geographical range of *M. roreri* was Colombia, Ecuador and Western Venezuela (Phillips-Mora 2003). In only 50 years, starting in 1956 with the first report in Panama, *M. roreri* disseminated across every Mesoamerican nation (Phillips-Mora et al. 2006b, 2006a, Phillips-Mora and Wilkinson 2007), and by 2016 it reached its first Caribbean victim, Jamaica (Johnson et al. 2017). The pathogen also spread South, reaching northern Peru in 1988, southern Peru in 1998 and Bolivia in 2012 (Hernández T. et al. 1990, Ríos-Ruiz and Rodríguez 1998, Phillips-Mora et al. 2015). The total cacao production from Western Africa, Southeast Asia and Brazil add up to almost 90% of the world's production (Chapter 1). Fortunately for world chocolate lovers, *M. roreri* is not present yet in these major cacao-producing regions.

Moniliophthora roreri has traditionally been considered to be indigenous to Ecuador (Briton-Jones 1934). However, recent molecular-based studies have found high levels of genetic diversity for *M. roreri* in Colombia (Phillips-Mora et al. 2007a, Jaimes et al. 2016). In fact, a recent study proposed that the center of origin of the fungus is the upper Magdalena Valley of Colombia (Ali et al. 2015). However, in this study Colombia was over represented in comparison to other countries (for example, 66 isolates from Colombia vs 11 from Ecuador and two from Peru), which could have biased the results. Therefore, a more complete geographic sampling is needed to fully resolve the center of diversity for this pathogen, especially given that the origin and earliest human use of the host took place in their Upper Amazonian regions, not the Magdalena Valley (Thomas et al. 2012, Zarrillo et al. 2018). In addition to *T. cacao*, *M. roreri* is able to infect other Malvaceae species within the genera *Theobroma* and *Herrania* (Phillips-Mora and Wilkinson 2007); however, the genetic diversity of *M. roreri* isolates coming from cacao wild relatives has not been evaluated in depth.

Finally, the invasive history of *M. roreri* tells us that its dissemination throughout Mesoamerica has been due to one or very few introductions of the same or almost identical genotypes in a clonal manner from South America (Phillips-Mora et al. 2007a, Ali et al. 2015, Díaz-Valderrama and Aime 2016a). Additionally, a second genotype seems to have invaded Peru in a similar manner (Ali et al. 2015, Díaz-Valderrama and Aime 2016a). The recent invasions of *M. roreri* into Jamaica in 2016 and Bolivia in 2012 (Phillips-Mora et al. 2015, Johnson et al. 2017) raise the question of whether these represent further incursions of the same genotypes that have been invading throughout the Americas, or represent the evolution of new invasive genotypes.

The objectives of this study are: 1) to re-examine the genetic diversity and center of origin of *M. roreri* by broadening the sampling of *M. roreri* to Ecuador, Peru and the Peruvian Upper Amazon, and by including samples from wild cacao relatives *T. bicolor* and *T. grandiflorum*; and 2) to determine the genotypes that invaded Jamaica and Bolivia.

2.2 Hypotheses

Based on the results from previous studies, the research hypotheses in this study are:

1. Magdalena valley in Colombia is the center of origin of *M. roreri*.
2. The genotypes that invaded Bolivia and Jamaica are the same invasive genotypes that have been spreading throughout the Americas since the 1950s.

2.3 Materials and Methods

2.3.1 Collection and isolation of *M. roreri* samples

A total of 228 samples of *M. roreri* were used in this study (APPENDIX A). The collections took place in Ecuador, Colombia, Peru, Bolivia, Costa Rica and Jamaica from 2015 to 2017. There were sixteen samples from Magdalena Valley in Colombia; forty-six samples from Ecuador (Guayas basin, coastal region), most of which came from an 850-acre cacao farm; fourteen from Maynas in the Peruvian Upper Amazon (a place where *M. roreri* was never sampled before); and one hundred and twenty-six across the main cacao-cultivating regions of Peru from North to South. Additionally, samples from the first reports of FPR in Mexico and Belize (Phillips-Mora et al. 2006a, 2006b) were included in the analysis, and the samples from Jamaica and Bolivia come from the same areas where the first report of FPR in these countries occurred (Phillips-Mora et al. 2015, Johnson et al. 2017). 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 (APPENDIX A). Samples from *T. cacao* came from the main three cacao genetic groups *Trinitario, Criollo* and *Forastero,* from Peruvian native varieties and hybrid cultivars (APPENDIX A; García Carrión 2010). Five isolates, including one from Panama, analyzed in a previous study were included here (Díaz-Valderrama and Aime 2016a).

2.3.2 Molecular characterization of *M. roreri* samples

The DNA extraction of samples is thoroughly described in Chapter 3. Sixteen simple sequence repeat (SSR) or microsatellite markers were used to molecularly characterize *M. roreri* samples (Díaz-Valderrama and Aime 2016a; Table 2.1). SSR loci were detected and primers flanking the SSR sequences were designed with program QDD 3.1.2b (Meglécz et

al. 2010, 2014) by screening an improved assembly of a previous version of a *M. roreri* genome (See Chapter 4 for details). The program identified 1,940 SSR loci from out of which sixteen were selected (Table 2.1). The selection criteria followed were: number of motif repeats, the scaffold size where loci were located and position in the scaffold. In this way, it is ensured they were as evenly distributed throughout the *M. roreri* genome as possible. Eleven SSR loci selected were used previously in a prior molecular study of *M. roreri* (Díaz-Valderrama and Aime 2016; Table 2.1). The genotyping of samples was performed via capillary electrophoresis with a modified "M13-tailed primer" method (Schuelke 2000) as previously done (Díaz-Valderrama and Aime 2016a, Koch and Aime 2018). To ensure replicability, PCR reactions on all the markers and on a subset of five samples in each region were performed three times. Also, in some cases reactions generated null alleles. Null alleles in SSR genotyping is frequent and might have an effect on genetic diversity estimations (Grünwald et al. 2017). Then, in all these cases PCRs were replicated three times with positive controls to make sure the legitimacy of the null allele. In addition to the SSRs, a dataset of 88 single-nucleotide polymorphisms (SNPs) of 172 *M. roreri* samples available from a previous study (Ali et al. 2015) were analyzed to compare results leading to identify the center of origin of *M. roreri*.

The SSR alleles were scored with Geneious 10.0.5 (Kearse et al. 2012). Most of the downstream analyses, unless specified, were done with the multi-function R package *poppr* for population genetics studies (Kamvar et al. 2014, 2015a). Data sets were arranged in GenAlEx format (Peakall and Smouse 2006, 2012), compatible with *poppr*. Because of the invasive history of *M. roreri* and to meet the objectives of the study, six genetic groups or "populations" were defined *a priori* of analyses based on geography: Ecuador, Colombia, Peru (which contains all samples collected in the country except in Maynas province), Maynas, Bolivia and Central America (which contained all samples from the Mesoamerican region plus samples from Jamaica). The order of analyses was as follows: identification of multilocus genotypes (MLGs), identification of multilocus lineages (MLLs) and clone correction, adjustment of alleles in the entire dataset based on the MLLs identified, evaluation of the genetic resolution power of markers, analysis of genetic structure, analyses of genetic diversity and relationships of samples, and spatial

autocorrelation based on geographical data (Figure 2.1). The same pipeline was applied to the SNP dataset by replacing nucleotides for numbers (A for 1; C = 2; G = 3; T = 4; a deletion = 5) as specified in the GenAlEx tutorial. The predefined groups in the SNP dataset were also based on geography, but since there was only one and two samples from Venezuela and Peru (Ali et al. 2015) they were grouped together with samples from Colombia and Bolivia, respectively. Also, this dataset only contained samples from Costa Rica in Central America. Then, for the SNP dataset there were four pre-defined groups: Costa Rica, Colombia/Venezuela, Ecuador and Peru/Bolivia.

2.3.2.1 Clone correction, MLLs and resolution of markers

Clone correction is the process of keeping one copy of each MLG in each location. However, it is possible that almost identical MLGs, i.e., MLGs that differ in very few alleles, may be part of the same MLL and that those allele differences are due to mutations during mitosis or even scoring errors, and not because of previous hypothetical sexual reproduction events. Since *M. roreri* is a clonal fungus (Díaz-Valderrama and Aime 2016a, 2016b), the identification of MLLs was based on the procedure established for population genetics analyses in clonal organisms (Arnaud-Haond et al. 2007a, 2007b). First, we examined the spectrum of genetic diversity (SGD) by plotting the frequency distribution of the genetic distances among MLGs (Rozenfeld et al. 2007); for this Nei genetic distances were used (Nei 1972). If at the low distance extreme of the SGD there is a small peak, i.e., the SGD has a bimodal rather than a unimodal distribution (Arnaud-Haond et al. 2007a), it is possible that MLGs having genetic distances at that threshold (the small distance peak) or lower could belong to the same MLL. To statistically corroborate the bimodality of the SGD, we performed the Hartigan's Dip test with R package *diptest*; a *p-value* less than 0.05 indicates the Dip value (D) of the SGD statistically differs from zero, i.e. SGD 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 belongs to the same MLL. To accomplish this, we identified and temporarily removed the loci for which there were allele differences among MLGs in each cluster. Then, without these allele-different loci, the probability that the now identical

MLGs arose from hypothetical independent sexual reproductive events or p_{sex} was calculated; a p_{sex} value less than 0.01 meant that those almost identical MLGs were likely derived from the same hypothetical sexual event; i.e., they were part of the same MLL (Arnaud-Haond et al. 2007a, 2007b). The differing alleles of samples in each MLL were then adjusted by replacing them with the most common allele in the MLL. At the end, each MLL would have a representative genotype that was used in further analysis. Then, the genetic resolution power of 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. 2007a, Kamvar et al. 2015a). These curves were generated by randomly sampling 1,000 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.

2.3.2.2 Genetic diversity

Several diversity indices were performed over the SSR and SNP datasets. Clonal richness (*R*) was estimated with the index proposed by Ellstrad and Roose (2019) with the adjustments of Dorken and Eckert (2001). The genetic variation in isolates in each region was quantified in terms of gene and genotypic diversity. For the former, the Nei diversity index (Nei 1978), while for the latter, the Shannon-Wiener (*H*), Simpson (λ), and Stoddart and Taylor (*G*) indexes (Simpson 1949, Hill 1973, Stoddart and Taylor 1988) were calculated. The distribution of the gene and genotypic diversity among regions was calculated with the evenness index E_5 which provides a ratio between abundant over rare alleles/genotypes (Ludwig and Reynolds 1988, Grünwald et al. 2003). All these values were calculated with both the entire (without clone correction) and the clone-corrected dataset. Similarly, the diversity and evenness indexes of individual SSR/SNP loci were calculated to determine the more informative loci.

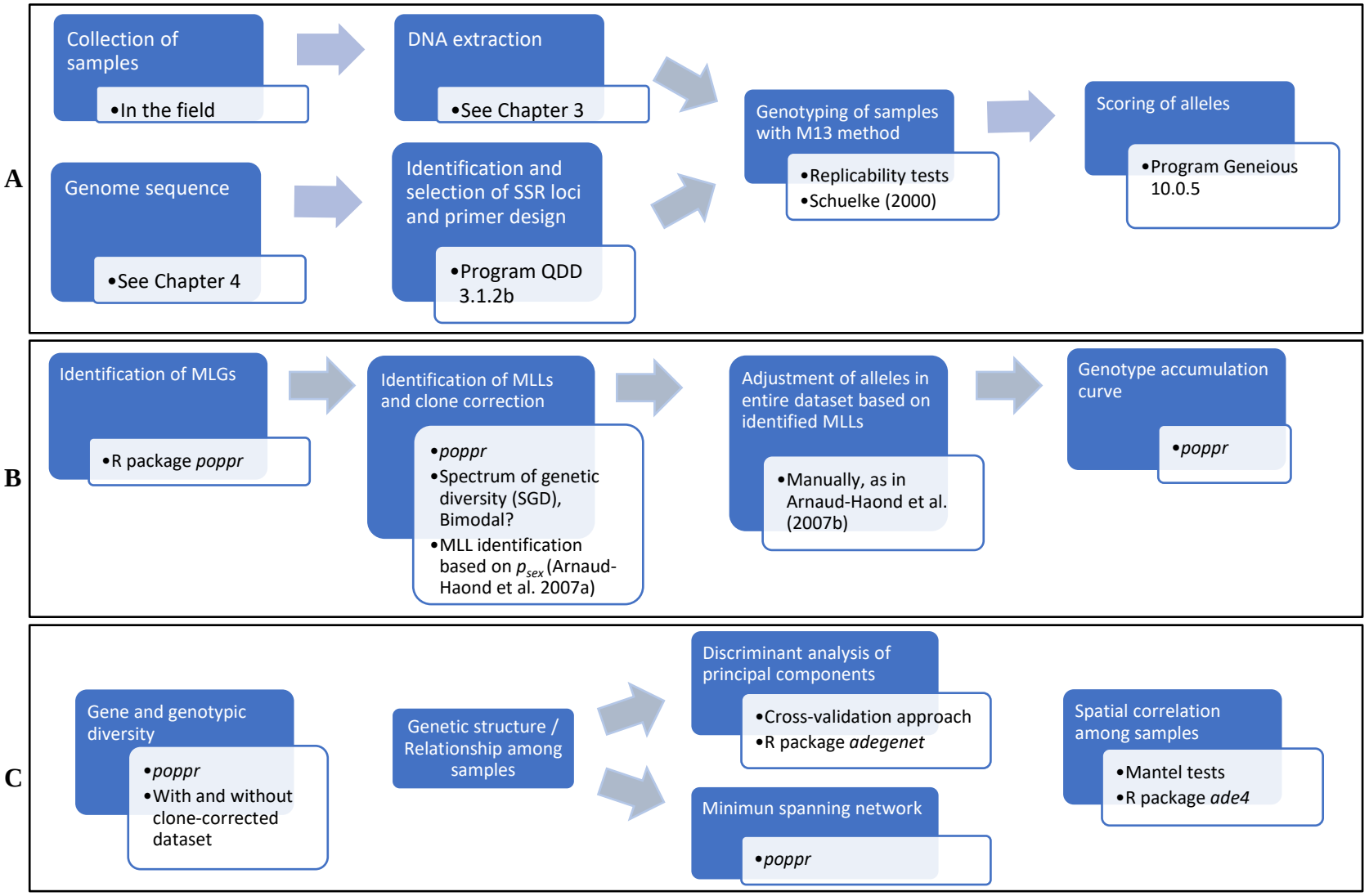


Figure 2.1 Overview of the methods and analyses performed in this study. **A)** Genotyping process. **B)** Clone correction. **C)** Analyses.

Table 2.1 Microsatellite markers used in this study

Locus*	Forward primer (5' to 3')	Reverse primer (5' to 3')	Motif	Improved genome assembly		
				Scaffold	Approx. position	Scaffold size
Mr_SSR1	CAATCCAAATCCCCCAAATC	TAGACTYGAGATCTGAAAGCAGGG	$(CTTCT)_{12}$	dna.fa_2	15000	297771
Mr_SSR4	GAAGAGGCATATAAGGACGTTGG	CAGGTGGATTCGGATAGTTTGTAT	$(TC)_{20}$	dna.fa_50	239000	509665
Mr_SSR9	ATCACCTCTTGCTACTTTCTTGCT	AAGATACAAAATGGATTAACTCG	$(CA)_{16}$	dna.fa_38	588000	698367
Mr_SSR12	CCAGTGCATGAGTAGGGATAAATA	GTTAGAAATGCTGCTAATGGGTCT	$(AG)_{16}$	dna.fa_94	21000	213359
Mr_SSR17	GCAGTCTAGCCATATCGTGTTGTA	GTATTTTACTAGGCTTGCTCTCGC	$(GTGTT)_{7}$	dna.fa_46	599000	601134
Mr_SSR18	AGTTTAAGTCTTGAGGTGAAGCGT	GAACAGTAGCGAAGAATCTAGGGA	$(GA)_{16}$	dna.fa_18	932000	1140650
Mr_SSR23	ATCGTATCTGTATGGTGGTTGTTG	GTGTGTCTTCGTTCTCTCGTTCTA	$(AG)_{16}$	dna.fa_23	75000	992592
Mr_SSR25	GAGCCTATATTCCCACATCCATAC	TGCTGACTGACTTCTTGCTATTTT	$(TGA)_{10}$	dna.fa_19	746000	1110851
Mr_SSR27	AAGAAGGTGAAGAAGAACAAGTGG	GAAATGGATATGGACAATGGGTAT	$(GA)_{15}$	dna.fa_4	337000	1556181
Mr_SSR28	CTTTATTCTTCACGACATGACACC	CGTCCGTATAAAAAGACTAGGCAG	$(TC)_{12}$	dna.fa_36	626000	713525
Mr_SSR30	GGGTTCATCTCCGAACTATCATAC	TCCATTCCAAGGGTATCTATCAAT	$(TC)_{12}$	dna.fa_71	233000	316685
Mr_SSR33	TTGTGCACAGAGCCAAATGC	CAGCACCGACACTGGGATTT	$(TC)_{18}$	dna.fa_27	745000	845985
Mr_SSR37	ACCTGAAAGAGCGGCAATGA	AGTTGGACGCTTCGATACCG	$(GA)_{11}$	dna.fa_44	176000	616059
Mr_SSR38	ACAGCCAAGAAGCCCAAAGA	GCCTCTGTGCTTTGCTTTCC	$(TG)_{11}$	dna.fa_70	39000	324791
Mr_SSR39	TGGTGCTGTGGTGAGATAGC	TCCAACTTCTCCAACCCTGC	$(GTC)_{10}$	dna.fa_4	1088000	1556181
Mr_SSR40	AGACGAGCACAGAAGACAGC	TGGTGGAGTGAAGGTGAAGC	$(AAC)_{10}$	dna.fa_16	443000	560533

* Markers Mr_SSR1-30 were previously used (Díaz-Valderrama and Aime 2016a), and were identified again in the QDD analysis with the improved genome assembly of *M. roreri* (Chapter 4)

2.3.2.3 Genetic structure, relationships and spatial correlation of samples

A discriminant analysis of principal components (DAPC) was performed to visualize the genetic structure among *M. roreri* samples and to investigate which markers/alleles contribute to the genetic differentiation among the pre-defined groups that define the samples, i.e., regions. The DAPC is a model-free multivariate statistical approach that does not rely on Hardy-Weinberg equilibrium or linkage disequilibrium assumptions; therefore, it is recommended for asexual organisms (Jombart et al. 2010). This analysis was performed with R package *adegenet* (Jombart 2008). The most appropriate number of principal components (PCs) retained in the discriminant analysis (DA) was calculated with the cross-validation approach as in Kamvar et al. (2015b), with a training set of 90% of the data and 1,000 replicates. The contribution of alleles/SNPs to the genetic differentiation of *M. roreri* groups was examined by inspecting the linear coefficients of the discriminant functions of the PCs (Jombart et al. 2010).

To examine the relationships and genetic distances among MLG/MLLs and to ultimately draw conclusions about the center of origin of *M. roreri*, a minimum spanning network analysis (MSN) was performed with *poppr*, which uses Prim's algorithm (Prim 1957) implemented in the R package *igraph* (Csárdi 2006). For this, a discrete dissimilarity matrix reflecting the number of differing alleles was used. 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 origin of genotypes located in the center of the network corresponds to the center of origin of *M. roreri* (Couch et al. 2005).

Finally, as another tool to explore the center of origin of *M. roreri*, Mantel tests (Mantel 1967) implemented in *ade4* R package (Dray and Dufour 2007) with 1,000 bootstrap replicates were performed. Nei distances and Euclidean distances between decimal degree geographical coordinates were used for these tests. Significant correlations ($p < 0.05$) between genetic and geographical distances are expected in the regions located in the center

of origin, while non-significant correlations are expected in regions outside the center of origin.

2.4 Results

2.4.1 SSR dataset

2.4.1.1 Clone correction, identification of MLGs and MLLs

Fifty-seven MLGs were initially found across the 228 samples analyzed before inspection for MLLs. Then, the SGD with Nei genetic distances of samples was examined with unique MLGs (clone-corrected dataset) and found to have a binomial distribution (D = 0.076; $p <$ 2.2 E-6) with a small peak in the beginning of the spectrum, setting up the threshold value for MLL identification at 0.13 (Figure 2.2). Therefore, MLGs that clustered together at a 0.13 Nei distance or lower were evaluated to see whether they belong to the same MLL via p_{sex} (Figures Figure 2.2 and Figure 2.3). Five MLLs were identified ($p_{sex} < 0.01$; Figure 2.3) and their differing alleles were adjusted (APPENDIX B). The final SSR dataset used in further analyses contained 50 total MLGs including the representative genotype of each of the five identified MLLs (Figure 2.3). One MLG from Central America was found in Ecuador (MLG_5) and the MLL found in Bolivia was observed repetitively among the great majority of Peruvian samples (MLL_5). All the 14 samples from Maynas province had MLG_17, but this was a unique genotype not seen in any other region (Table 2.2). Samples from *T. grandiflorum* and *T. bicolor* had the same MLG as samples from *T. cacao* (Table 2.2; APPENDIX A), except for samples JD_E13 and JD_E25 (from *T. grandiflorum*) which had unique MLGs (Table 2.2).

2.4.1.2 Null alleles and the genetic resolution power of markers

Eleven out of 16 markers had at least one sample having an allele not successfully amplified, i.e., a null allele (SSR40, 28, 30,1, 39,38,4,18,37,33,25). Also, twenty-one samples had at least one null allele: 19 from Ecuador, one from Colombia and one from Peru. Marker Mr_SSR4 was the one that had the highest number of samples with null alleles (seven samples), followed by Mr_SSR33 and Mr_SSR28 with six and five, respectively (APPENDIX B). Overall, there was a low rate of null alleles in the SSR dataset;

on average only 2.31 or 1.01% of samples had null alleles at each locus. Additionally, 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 2.4).

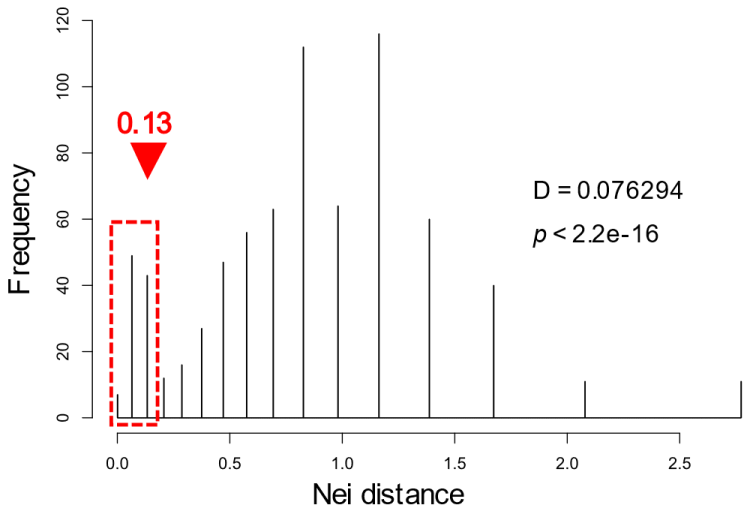


Figure 2.2 Spectrum of genetic diversity (SGD) of fifty-seven unique MLGs from 228 *Moniliophthora roreri* samples based on the sixteen-SSR data set of this study. D and *p* are the Hartigan's Dip value of unimodality and its probability, respectively. Red-dashed box highlights the region of the small distance distribution of the SGD up to the threshold value of 0.13.

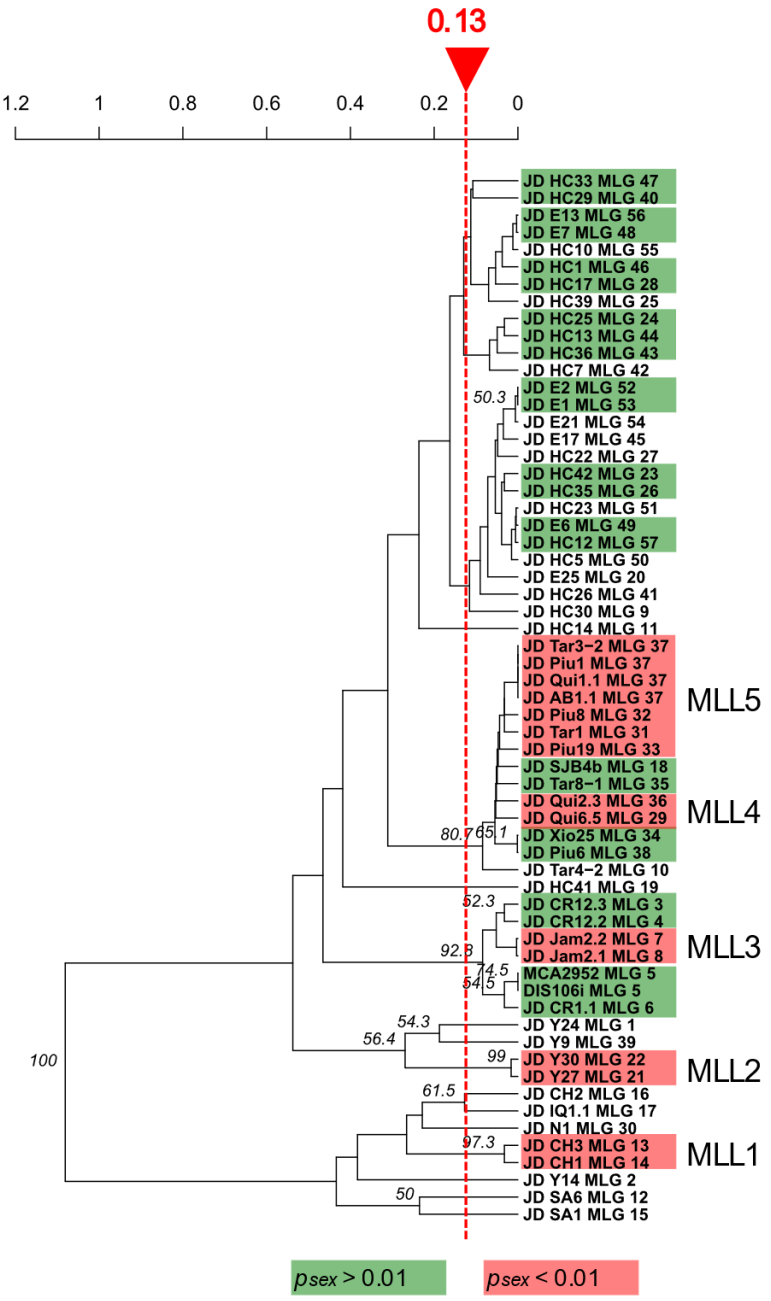


Figure 2.3 Nei-distance UPGMA dendrogram of the fifty-seven multilocus genotypes (MLGs) of *Moniliophthora roreri* encountered in the SSR analysis. Boxes indicate the clusters of samples under the Nei distance threshold of 0.13 that were examined to see whether they belong to the same multilocus lineage (MLL) via calculation of p_{sex}; if they do (red box) they were assigned to an MLL and their alleles were adjusted (See Materials and Methods). Tip labels contain the sample ID and the assigned MLG and MLL number; only bootstrap values greater than 50 are shown on nodes

Table 2.2 Final list of multilocus genotypes (MLGs) and multilocus lineages (MLLs) of *Moniliophthora roreri* identified in this study, with their region of origin and the number of samples (N) in each MLG/MLL. For full information of samples see APPENDIX A

Region	MLG / MLL	Samples	N
Central America	MLG_3	JD_CR12.3	1
	MLG_4	JD_CR12.2	1
	MLG_5	MCA2952; MCA2954; MCA2518	3
	MLG_6	JD_CR1.1; JD_CR1.2; JD_CR3.3; JD_CR4.1; JD_CR4.2; JD_CR6.1; JD_CR9.1; JD_CR9.3; JD_CR9.4	9
Jamaica	MLL_3	JD_Jam2.1; JD_Jam2.2; JD_Jam2.7	3
Colombia	MLG_1	JD_Y24	1
	MLG_2	JD_Y14; JD_Y17	2
	MLG_12	JD_SA6	1
	MLG_15	JD_SA1; JD_SA3	2
	MLG_16	JD_CH2	1
	MLG_30	JD_N1; JD_N2; JD_N3	3
	MLG_39	JD_Y9	1
	MLL_1	JD_CH1; JD_CH3	2
	MLL_2	JD_Y27; JD_Y27'; JD_Y30	3
Ecuador	MLG_5	DIS106i; Dis371.1.3	2
	MLG_9	JD_HC30	1
	MLG_11	JD_HC14	1
	MLG_19	JD_HC41	1
	MLG_20	JD_E25	1
	MLG_23	JD_HC42	1
	MLG_24	JD_HC25	1
	MLG_25	JD_HC39	1
	MLG_26	JD_HC35	1
	MLG_27	JD_HC22	1
	MLG_28	JD_HC17	1
	MLG_40	JD_HC29	1
	MLG_41	JD_HC26	1
	MLG_42	JD_HC7	1
	MLG_43	JD_HC36; JD_HC38; JD_HC40; JD_HC43; JD_HC44	5
	MLG_44	JD_HC13; JD_HC24; JD_HC31	3
	MLG_45	JD_E17; JD_E22	2
	MLG_46	JD_HC1; JD_HC45	2
	MLG_47	JD_HC33	1
	MLG_48	JD_E7; JD_E9; JD_E11; JD_HC27	4
	MLG_49	JD_E6; JD_HC32; JD_HC34; JD_HC37	4
	MLG_50	JD_HC5	1
	MLG_51	JD_HC23	1
	MLG_52	JD_E2; JD_HC21	1
	MLG_53	JD_E1; JD_HC28	2
	MLG_54	JD_E21	1
	MLG_55	JD_HC10	1
	MLG_56	JD_E13	1
	MLG_57	JD_HC12	1

Maynas	MLG_17	JD_IQ1.1; JD_IQ1.2; JD_IQ2.1; JD_IQ2.2; JD_IQ3; JD_IQ4; JD_IQ11.1; JD_IQ18.1; JD_IQ19.1; JD_IQ19.3; JD_IQ19.4; JD_IQ20; JD_IQ21.1; JD_IQ21.2	14
Peru	MLG_10	JD_Tar4-2	1
	MLG_18	JD_SJB4b	1
	MLG_34	JD_Xio25; JD_HU-10	2
	MLG_35	JD_Tar8-1; JD_Tar9-4; JD_Tar9-5	3
	MLG_38	JD_Piu6	1
	MLL_4	JD_Qui2.3; JD_Qui6.5; JD_Qui6.6	3
	MLL_5	JD_Piu1; JD_Piu5; JD_Piu8; JD_Piu11; JD_Piu12; JD_Piu15; JD_Piu17; JD_Piu19; JD_Piu20-1; JD_Piu20-2; JD_Piu21; JD_Piu22; JD_Piu24; JD_Piu25; JD_Piu26; JD_Piu28; JD_Piu29; JD_Piu31; JD_Piu32-1; JD_Piu32-3; JD_Piu33; JD_Piu34; JD_Piu35; JD_Piu36; JD_Piu37; JD_Piu38; JD_Ja2.1; JD_Ja2.2; JD_Ja3.1; JD_Ja3.2; JD_Ja5; JD_Ja6; JD_Ja7.1; JD_Ja7.2; JD_Ja8.1; JD_Ja8.2; JD_Ja9.1; JD_Ja9.2; JD_Ja10.1; JD_Ja10.2; JD_Ja11.1; JD_Ja11.2; JD_Ja12.1; JD_Ja13.1; JD_Ja14.1; JD_Tar1; JD_Tar3-2; JD_Tar4-3; JD_Tar5-2; JD_Tar6-1; JD_Tar6-2; JD_Tar7; JD_Tar11-2; JD_Qui1.1; JD_Qui2.1; JD_Qui2.2; JD_Qui3.1; JD_Qui3.2; JD_Qui3.3; JD_Qui5; JD_Qui6.1; JD_Qui6.2; JD_Qui7.1; JD_Qui7.2; JD_Qui7.5; JD_Qui8.1; JD_Qui8.2; JD_Qui9; JD_SJB1.1; JD_SJB1.2; JD_SJB2; JD_SJB3.1; JD_SJB3.2; JD_Xio1; JD_Xio2; JD_Xio4; JD_Xio6; JD_Xio8; JD_Xio10; JD_Xio12; JD_Xio15; JD_Xio17; JD_Xio19; JD_Xio21; JD_Xio22; JD_Xio23; JD_HU-01; JD_HU-03; JD_HU-04; JD_HU-05; JD_HU-06; JD_HU-09; JD_HU-11; JD_HU-13; JD_HU-14; JD_HU-15; JD_HU-18; JD_JU-34; JD_JU-35; JD_HU-17; JD_JU-36; JD_JU-37; JD_JU-38; JD_JU-39; JD_SM-19; JD_SM-23; JD_SM-24; JD_SM-25; JD_SM-26; JD_UC-28; JD_UC-29; JD_UC-30; JD_UC-31; JD_UC-32; JD_UC-33	115
Bolivia	MLL_5	JD_AB1.1; JD_AB2; JD_AB2.3; JD_AB3; JD_AB5.2; JD_AB6.1; JD_AB7.1; JD_AB8.1; JD_AB9.1	9

Samples and MLGs from hosts other than *T. cacao* are highlighted: green for *T. bicolor* and yellow for *T. grandiflorum*

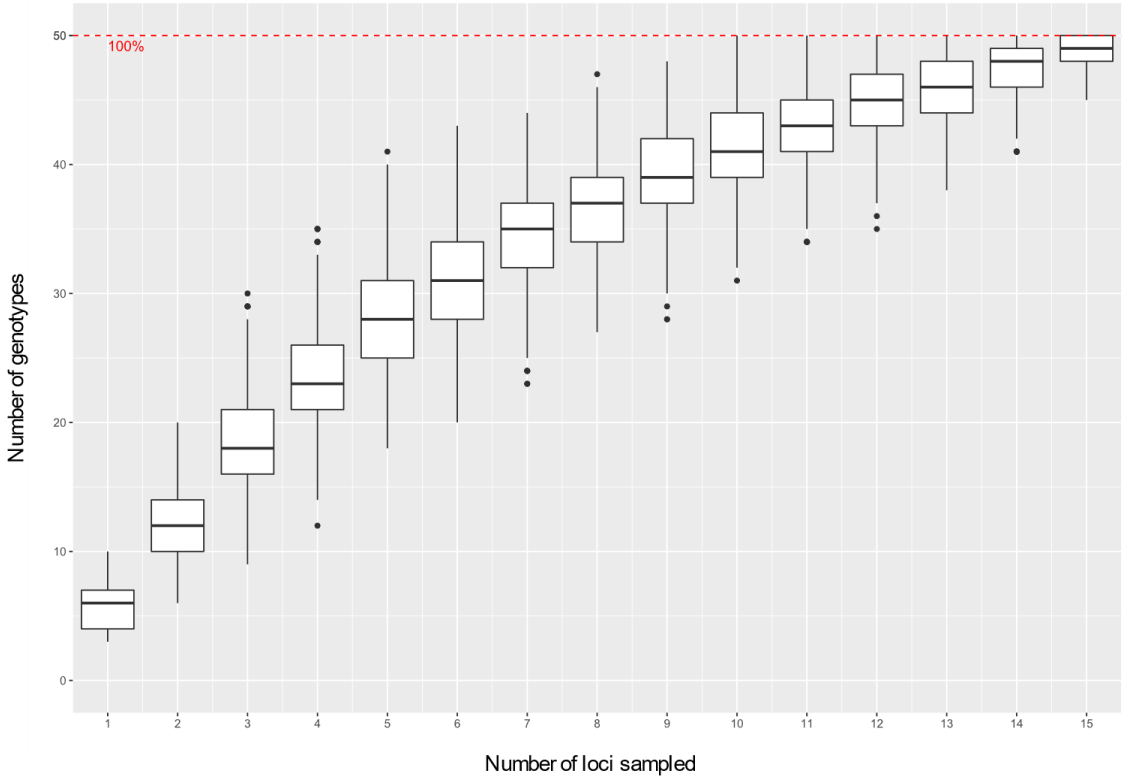


Figure 2.4 Genotype accumulation curve of the markers used in this study.

2.4.1.3 Genetic structure, diversity, relationship of samples and spatial correlation

Based on the cross-validation analysis, the first fifteen PCs were retained in the DAPC on the SSR dataset (conserving 93.2% variance). The genetic structure among *M. roreri* MLG/MLLs identified by the DAPC reduced the six pre-defined groups (regions) into three main genetic groups: samples from Colombia and Maynas; Peru and Bolivia; and Ecuador and Central America (Figure 2.5). Therefore, subsequent analyses were presented jointly for Colombia and Maynas, and Peru and Bolivia, unless specified. Central America and Ecuador samples were kept separate. Additionally, the SSR loci that highly contributed to the separation of genetic groups according to the first two PCs of the DAPC were Mr_SSR27, Mr_SSR28, Mr_SSR18 and Mr_SSR39 (Figure 2.5). In this way, all samples from Peru/Bolivia had alleles 202, 182, and 209 in Mr_SSR27, Mr_SSR28 and Mr_SSR18, respectively; all samples from Ecuador/Central America had allele 213 in Mr_SSR18; and all samples from Colombia and Maynas had allele 251 in Mr_SSR39 (Figure 2.5).

All SSR loci were polymorphic, especially in Ecuador and Colombia/Maynas (Table 2.3). The average numbers of alleles (*Na*) per locus found in these regions were 3.56 and 2.94, respectively, while in Central America and Peru/Bolivia *Na* were 1.44 and 1.25, respectively (Table 2.3). Similarly, Ecuador and Colombia/Maynas were the regions with the highest Nei gene diversity, and Central America and Peru/Bolivia, with the lowest (Table 2.3). Overall, the most polymorphic locus was Mr_SSR1 with ten alleles, followed by Mr_SSR18, Mr_SSR27 and Mr_SSR33 all with eight (Table 2.3). The locus with the highest Nei gene diversity and with the most evenly distributed alleles among samples was Mr_SSR4 (*Na* = 6; Nei = 0.79, E_5 = 0.84). In Colombia/Maynas, locus Mr_SSR25 was the most informative as it had the highest gene diversity and their alleles are among the most evenly distributed (Nei = 0.80; E_5 = 0.95). Also, the diversity of the Mr_SSR12, Mr_SSR23 and Mr_SSR25 loci in Colombia/Maynas reduced severely in all other regions (Table 2.3). In Ecuador, the most informative loci were Mr_SSR4 and Mr_SSR33 based on Nei gene diversity and *E5.*

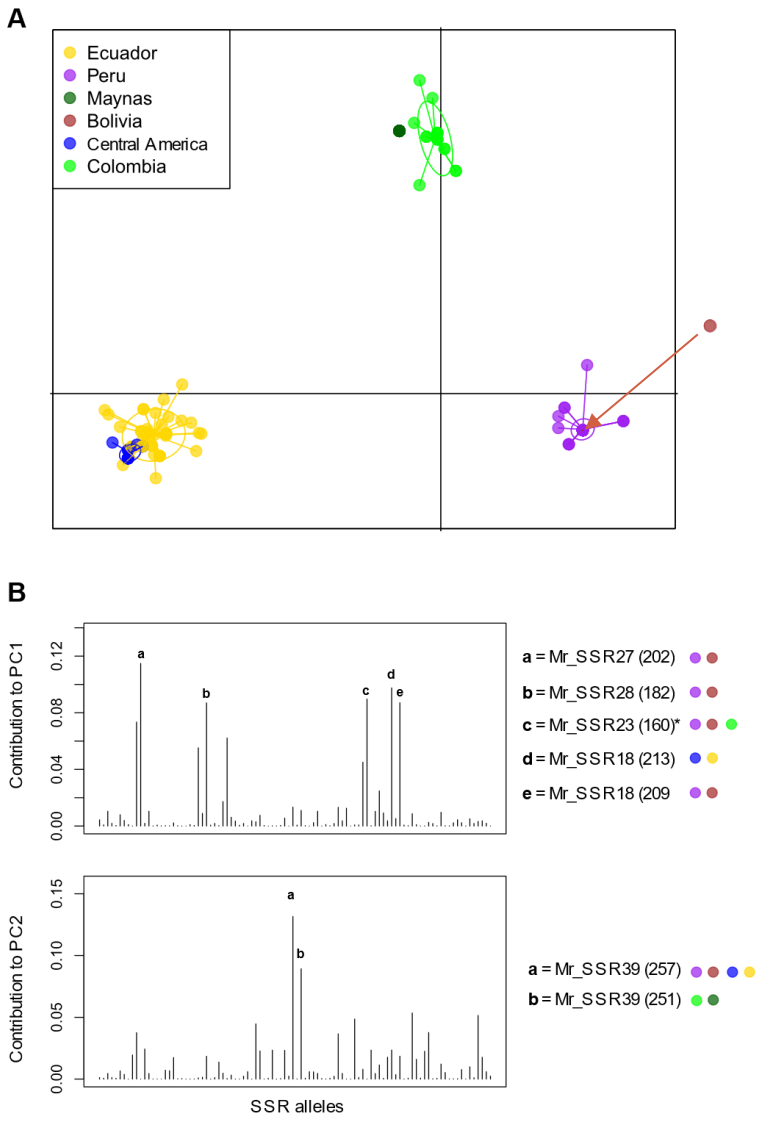


Figure 2.5 Discriminant Analysis of Principal Components (DAPC) of the entire SSR data set of *Moniliophthora roreri*. **A)** Scatterplot of the first two principal components (PCs) of the DAPC using region as pre-defined genetic groups (colors); each dot represents a sample and inertia ellipses, the 95% confidence clouds of each group; lines connect each sample to the center of its ellipse; all samples from Bolivia have the same MLG and overlap with samples from Peru, they are indicated with an arrow. **B)** Contributions of SSR alleles to the first (top) and second (bottom) PC of the DAPC. The highest-contributing alleles able to discriminate samples from different regions are labeled; the specific allele involved is in parentheses and the discriminated region(s) are color-coded based on the same legend as in A).

Table 2.3 **.** Number of alleles (*Na*), Nei gene diversity and Evenness (E_5) of the sixteen SSR loci used for the genetic analysis of 228 samples of *Moniliophthora roreri* (clone-corrected dataset).

Locus	All regions			Central America			Colombia and Maynas			Ecuador			Peru and Bolivia		
	Na	**Nei**	E_5	***Na***	**Nei**	E_5	***Na***	**Nei**	E_5	***Na***	**Nei**	E_5	***Na***	**Nei**	E_5
Mr_SSR1	10	0.78	0.68	1	—	—	4	0.73	0.84	6	0.71	0.77	2	0.25	0.61
Mr_SSR4	6	0.79	0.84	2	0.40	0.72	4	0.64	0.70	4	0.74	0.91	2	0.43	0.79
Mr_SSR9	3	0.48	0.75	1	—	—	2	0.47	0.86	2	0.41	0.83	2	0.25	0.61
Mr_SSR12	4	0.31	0.49	2	0.40	0.72	3	0.69	0.91	2	0.13	0.52	2	0.25	0.61
Mr_SSR17	5	0.37	0.5	1	—	—	3	0.38	0.58	3	0.20	0.48	2	0.25	0.61
Mr_SSR18	8	0.66	0.58	1	—	—	4	0.64	0.70	3	0.31	0.58	1	—	—
Mr_SSR23	4	0.59	0.71	1	—	—	3	0.73	0.98	2	0.07	0.44	1	—	—
Mr_SSR25	5	0.34	0.46	1	—	—	4	0.80	0.95	2	0.19	0.58	1	—	—
Mr_SSR27	8	0.77	0.74	1	—	—	3	0.60	0.81	5	0.46	0.56	1	—	—
Mr_SSR28	6	0.57	0.58	1	—	—	4	0.73	0.84	2	0.30	0.68	1	—	—
Mr_SSR30	7	0.70	0.70	2	0.60	0.96	3	0.62	0.8	4	0.41	0.56	1	—	—
Mr_SSR33	8	0.72	0.64	2	0.60	0.96	2	0.36	0.72	5	0.76	0.88	2	0.25	0.61
Mr_SSR37	4	0.31	0.52	1	—	—	2	0.53	0.96	3	0.20	0.48	1	—	—
Mr_SSR38	6	0.43	0.49	1	—	—	2	0.36	0.72	5	0.37	0.48	1	—	—
Mr_SSR39	5	0.42	0.54	1	—	—	2	0.36	0.72	3	0.13	0.43	2	0.25	0.61
Mr_SSR40	7	0.34	0.43	1	—	—	2	0.56	1.00	6	0.32	0.41	1	—	—
Mean	6	0.54	0.6	1.25	0.12	0.84	2.94	0.57	0.82	3.56	0.36	0.60	1.44	0.12	0.64

Clonal richness and evenness were more informative in the dataset without clone correction as expected (Figure 2.6A). Colombia and Ecuador are the regions with the highest *R* values: 0.53 and 0.62, respectively; i.e., in both regions more than half of the samples had a unique MLG. Also, the alleles found in Colombia and Ecuador were more evenly distributed compared to other regions, with E_5 values of 0.91 and 0.81, respectively (Figure 2.6A). Gene and genotypic diversity were more informative in the dataset with clone correction. Colombia/Maynas and Ecuador were the regions with the highest gene and genotypic diversity. However, Colombia/Maynas had the highest Nei gene diversity while Ecuador had the highest genotypic diversity (Figure 2.6A).

The MSN shows the relationship among *M. roreri* MLG/MLLs (Figure 2.6B). The samples from Ecuador and Colombia were in the center of the network but genotypes from Peru/Bolivia and Central America (including Jamaica) were connected only to the Ecuadorian portion of the network on two opposite sites. Conversely, the Maynas MLG is embedded within the Colombian network (Figure 2.6B). Genotypes from samples from other *Theobroma* species were interconnected to genotypes from *T. cacao* and do not seem to split into host-specific groups (Figure 2.6B).

The Mantel tests revealed significant correlations between genetic and geographical distances in Ecuador alone, Ecuador and Colombia combined, and Ecuador, Colombia and Maynas combined; analysis in Colombia alone was not significant ($p = 0.061$; Figure 2.7). There were also not significant correlations in Peru and Central America, and in Maynas alone and Bolivia the correlations were not calculated as only one MLG/MLL was present in those regions (Figure 2.7).

A

No clone correction	N	MLG	*R*	Nei	*H*	λ	*G*	E_5
Central America	17	5	0.25	0.07	1.28	0.65	2.86	0.72
Colombia	16	9	0.53	0.53	2.10	0.87	7.53	0.91
Maynas	14	1	0.00	0.00	0.00	0.00	1.00	NaN
Ecuador	46	29	0.62	0.31	3.19	0.95	19.96	0.81
Peru	126	7	0.05	0.01	0.44	0.17	1.20	0.36
Bolivia	9	1	0.00	0.00	0.00	0.00	1.00	NaN
Total	228	50	0.22	0.37	2.30	0.70	3.28	0.25
Clone correction	**N**	**MLG**	***R***	**Nei**	***H***	**λ**	***G***	**E_5**
Central America	5	5	1.00	0.13	1.61	0.80	5.00	1.00
Colombia	9	9	1.00	0.73	2.30	0.90	10.00	1.00
Maynas	1	1						
Ecuador	29	29	1.00	0.36	3.37	0.97	29.00	1.00
Peru	7	7	1.00	0.12	1.91	0.84	6.40	0.94
Bolivia	1	1						
Total	52	50	0.96	0.54	3.90	0.98	48.30	0.98

B

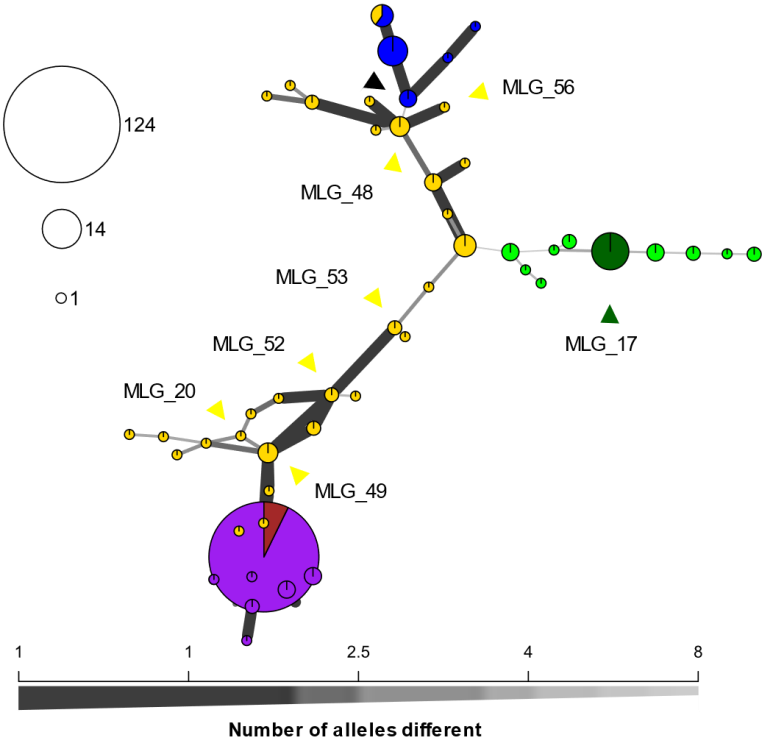


Figure 2.6 Genetic diversity and relationship among *Moniliophthora roreri* samples in each region with the SSR dataset. **A)** Diversity parameters and indexes found with the entire dataset (no clone correction) and clone-corrected dataset; the number of samples (N), number of MLGs, clonal richness (*R*), Nei gene diversity, Shannon-Wiener (*H*), Simpson (λ), Stoddart and Taylor (*G*) diversity indexes, and evenness (E_5) index are presented; NaN = not a number. **B)** Minimum spanning network of the entire dataset of *M. roreri*; nodes (circles) represent MLG/MLLs; node size, the number of samples; colors, regions; and connecting line widths and shading, relatedness (line lengths are arbitrary). Colors in A serve as legend for B. Colored arrows point to nodes containing at least one MLG from a host other than *Theobroma cacao* (Table 2.2); the black arrow point to MLG from Jamaica.

A

Region	Mantel correlation coefficient	*p* value
Central America	0.22	0.147
Colombia	0.17	0.061
Ecuador	0.48	0.006
Ecuador and Colombia	0.30	0.001
Ecuador, Colombia and Maynas	0.34	0.001
Ecuador, Colombia, Maynas, Peru	-0.03	0.428
Peru	-0.03	0.335
All regions	-0.02	0.455

Mantel tests in Maynas and Bolivia were not calculated because only one MLG was found in these regions

B

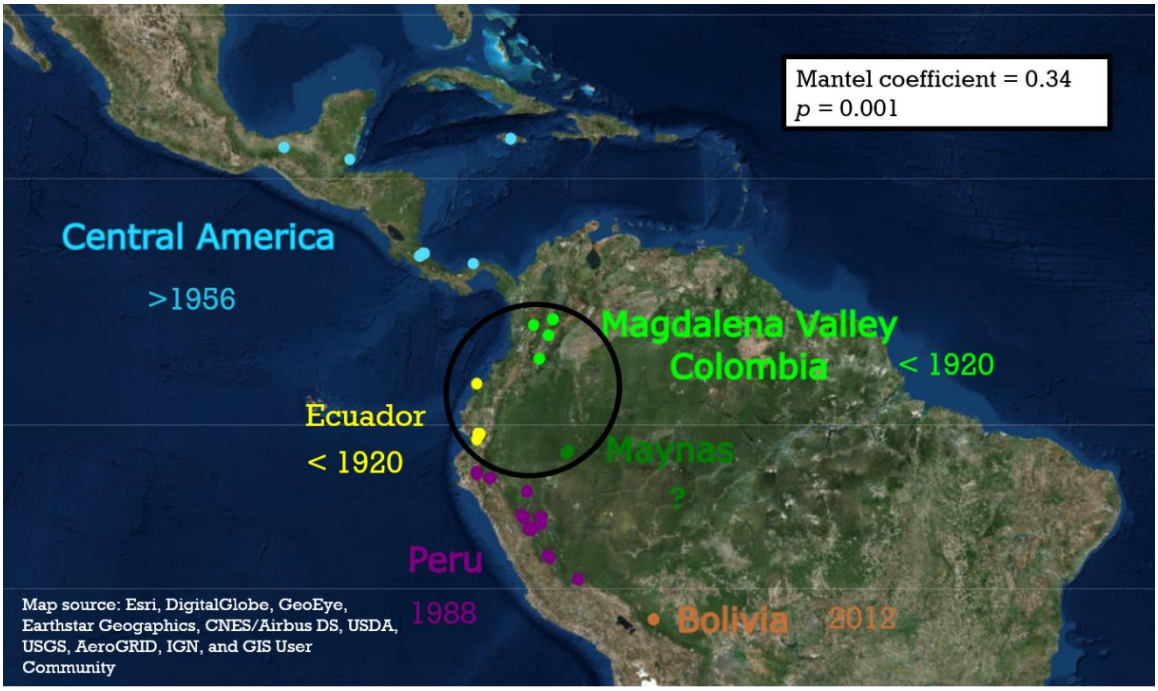


Figure 2.7 Mantel tests of *Moniliophthora roreri* samples to unveil its center of origin. **A)** Mantel correlation coefficients and probability values from individual and combined regions. **B)** Putative center of origin of *M. roreri* in Ecuador, Colombia and Maynas (Peruvian Upper Amazon) as predicted with the Mantel test. The map shows the location of the 228 samples used in this study. The years in which the first occurrence of *M. roreri* in each location are also indicated; this study reports for the first time the pathogen in Maynas but based on all the analyses the year of occurrence here must be earlier than 1988, when the first official report in Peru took place.

2.4.2 SNP dataset

The SNP dataset was also clone-corrected similarly to the SSR dataset (APPENDIX B and APPENDIX C). The SNP-SGD had a binomial distribution (D = 0.017; p = 8.27E-4) and the threshold value for MLL identification was 0.0365 (APPENDIX C). Then, eleven MLLs were identified (p_{sex} < 0.01). The final SNP dataset used in further analyses contained thirty-three MLG/MLLs (APPENDIX C). 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.

As determined by the cross-validation analysis, the first fifteen PCs were retained to calculate the discriminant functions of the DAPC on the SNP dataset (conserving 96.3% of variance). This DAPC showed an overlapping and wide genetic coverage among samples from Ecuador and Colombia compared to samples from Costa Rica and Peru/Bolivia (Figure 2.8A). However, Ecuador had a wider genetic coverage than Colombia, despite having many fewer samples (eleven vs sixty-seven; Figures Figure 2.8A and Figure 2.9A). MLLs from Costa Rica and Peru/Bolivia are depicted as single dots as there was only one MLL in each region (Figure 2.8A). SNPs that contribute more to the discriminant analysis along PC1 and PC2 were identified (Figure 2.8B). Some of these SNPs like “T” in 112_1_50358 and 251_4_7835, and “A” in 064_1_5323 could discriminate samples from Costa Rica, while a “C” in 064_1_5323 discriminates samples from Peru/Bolivia.

In the entire SNP dataset for Colombia and Ecuador, R values were 0.47 and 0.30, while E_5 were 0.76 and 0.66, respectively (results without clone correction; Figure 2.9A). Since there was only one MLL found in both Peru/Bolivia and Costa Rica, they have zero values for R (Figure 2.9A). 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 2.9A). Unlike the SSR dataset, the SNP-based MSN showed Colombian MLG/MLLs distributed along the entire network, while samples from other regions were embedded within the Colombian network (Figure 2.9B).

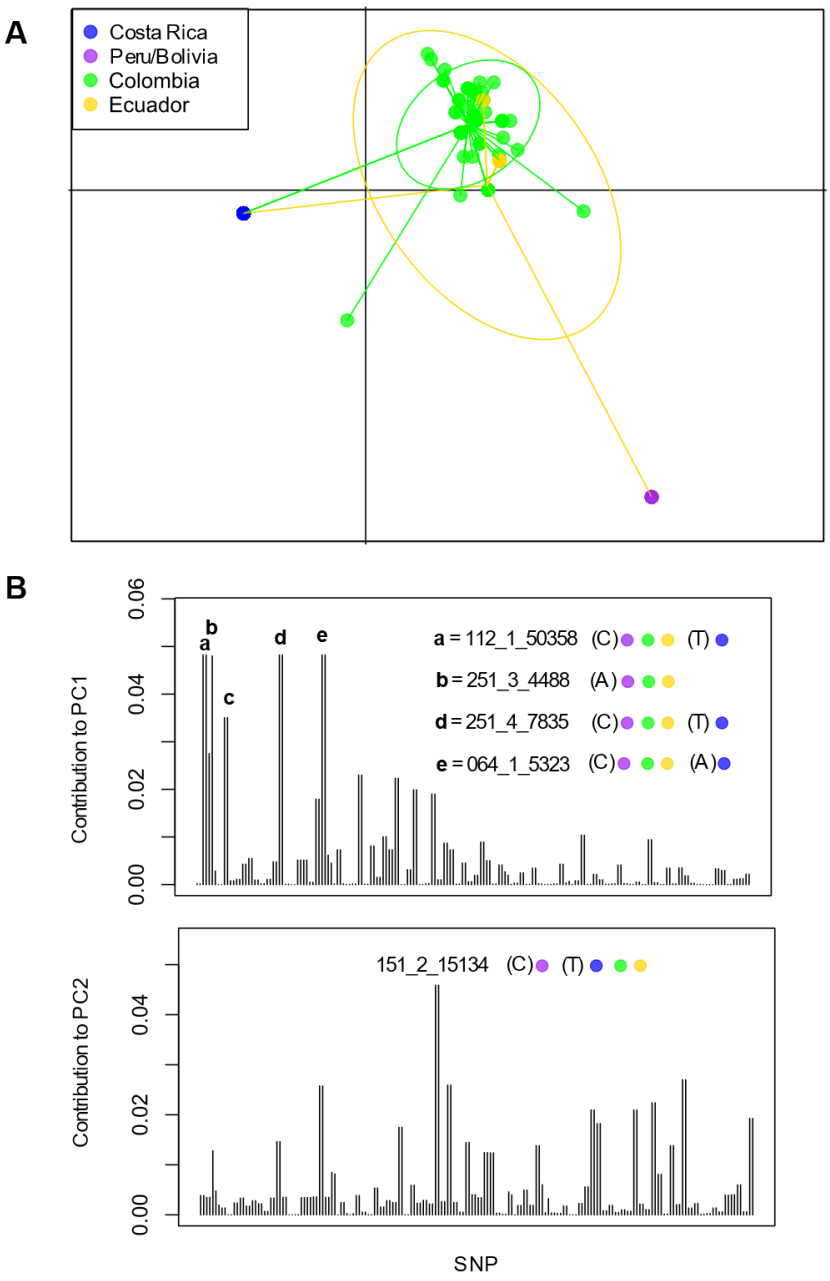


Figure 2.8 Discriminant Analysis of Principal Components (DAPC) of the SNP data set of *Moniliophthora roreri*. **A)** Scatterplot of the first two principal components of the DAPC using region as pre-defined genetic groups (colors); each dot represents an MLG and inertia ellipses, the 95% confidence clouds of each group; lines connect each MLG to the center of its ellipse. **B)** SNP contributions (coefficients) to the first (top) and second (bottom) principal component (PC) of the DAPC. The highest-contributing SNPs able to discriminate samples from different regions are labeled; the specific nucleotide involved is in parentheses and the discriminated region(s) are color-coded based on the legend in A.

A

No clone correction		N	MLG	*R*	Nei	*H*	λ	*G*	E5
	Costa Rica	86	1	0.00	0.00	0.00	0.00	1.00	NaN
	Colombia	67	32	0.47	0.34	3.20	0.95	18.94	0.76
	Ecuador	11	4	0.30	0.19	1.03	0.55	2.20	0.66
	Peru/Bolivia	8	1	0.00	0.00	0.00	0.00	1.00	NaN
	Total	172	33	0.19	0.30	2.18	0.72	3.54	0.32
Clone correction		**N**	**MLG**	***R***	**Nei**	***H***	**λ**	***G***	**E5**
	Costa Rica	1	1	NaN	NaN	0.00	0.00	1.00	NaN
	Colombia	32	32	1.00	0.36	3.47	0.97	32.00	1.00
	Ecuador	4	4	1.00	0.35	1.39	0.75	4.00	1.00
	Peru/Bolivia	1	1	NaN	NaN	0.00	0.00	1.00	NaN
	Total	38	33	0.86	0.37	3.44	0.97	28.90	0.92

B

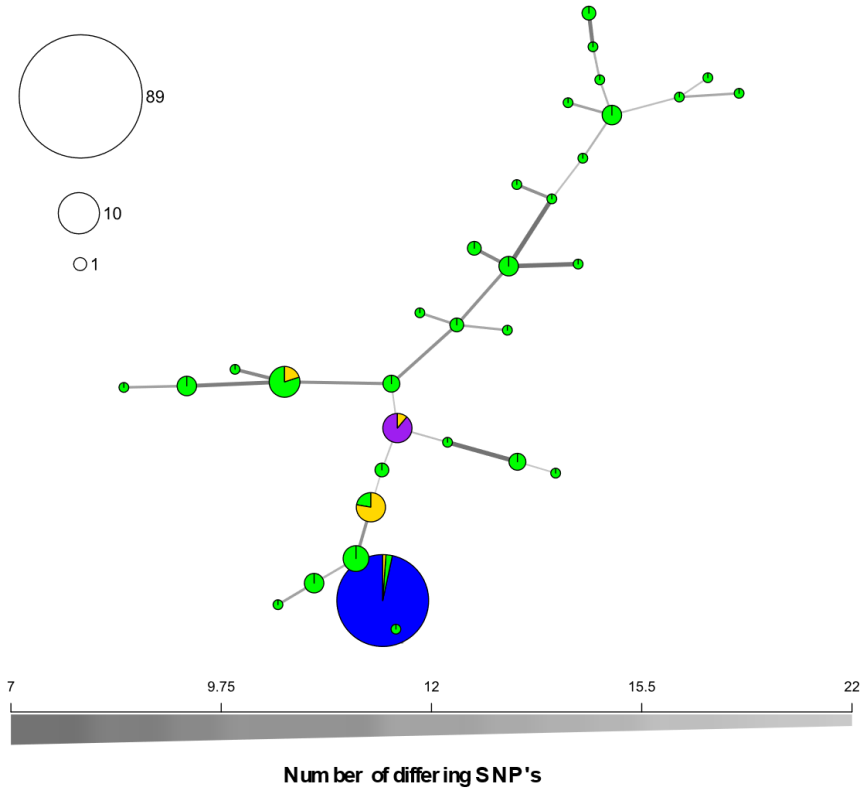


Figure 2.9 Genetic diversity and relatedness among *Moniliophthora roreri* MLGs with the SNP dataset. **A)** Diversity parameters and indexes found with the entire dataset (no clone correction) and clone-corrected dataset; the number of samples (N), number of MLGs, clonal richness (*R*), Nei gene diversity, Shannon-Wiener (*H*), Simpson (λ), Stoddart and Taylor (*G*) diversity indexes, and evenness (E_5) index are presented; NaN = not a number. **B)** Minimum spanning network of the entire dataset of *M. roreri*; nodes (circles) represent MLGs; node size, the number of samples; colors, regions; and connecting line widths and shading, relatedness (line lengths are arbitrary). Colors in A serve as the legend for B.

2.5 Discussion

2.5.1 On the center of origin

Moniliophthora roreri causes an invasive disease of cacao that has been increasing its range throughout the Americas via clonal reproduction (Díaz-Valderrama and Aime 2016a, 2016b). Currently, global cacao production is under threat from *M. roreri* (Chapter 1). There have been several attempts to investigate its genetic diversity and thus trace its center of origin (Table 2.4). Most of these have concluded that Ecuador and/or Colombia are the regions with the highest diversity in terms of clonal richness and Shannon-Wiener index (Table 2.4). Only one study found higher genetic diversity in Peru than in Ecuador (Table 2.4; Moreira 2006), but this is probably due to an artifact of the nature of the seven SSR markers they used, which were designed for analyses on *M. perniciosa* and as a result three of them were not polymorphic for *M. roreri* (Moreira 2006). When looking at the methodology followed in the studies on genetic diversity of *M. roreri* (Table 2.4), some issues arose. Except for Jaimes et al. (2016), these studies do not provide results with clone-corrected datasets. Thus, in this study we performed the genetic diversity analyses with and without clone correction of the generated SSR dataset as suggested for clonal organisms (Grünwald et al. 2017; Figures Figure 2.2 and Figure 2.3, and APPENDICES APPENDIX B and APPENDIX C). Furthermore, one study based on SNP markers concluded that the Magdalena Valley of Colombia is the center of origin of *M. roreri* (Ali et al. 2015). Unfortunately, it did not provide any diversity index measurement other than the number of MLGs found across the regions they sampled. Therefore, their SNP dataset was analyzed here, and the results were compared to the results obtained with the sixteen-SSR dataset (Figure 2.6A vs Figure 2.9A). When the clone correction is performed in the SNP dataset, the Nei gene diversity found in Ecuador and Colombia were almost the same (Figure 2.9A). Also, the SNP dataset revealed higher genotypic diversity in Colombia than Ecuador while the SSR analysis found the opposite, both regions being always the ones with higher genotypic diversity compared to Peru, Bolivia and Central America (Figures Figure 2.6A and Figure 2.9A). In either case, these results reject the hypothesis that *M. roreri* originated in the Magdalena valley in Colombia since the diversity clearly expands at least to Ecuador.

Table 2.4 Comparison of the genetic diversity of *M. roreri* reported across the literature

Study	Marker type	Number of markers	Region	N	*g*	*R*	*H*	*H*/ln(*g*)
Gutarra C. et al. (2013)	RAPD	14	Peru	21	3	0.10	0.39	0.35
Phillips-Mora et al. (2007)	AFLP/ ISSR	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
			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	0.59	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) *	SNP	88	Colombia	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

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). * Diversity indexes for data on Ali et al. (2015) were calculated in this study. H should be calculated based on clone-corrected datasets, but only Jaimes et al. (2016) and this study provide such a calculation

Furthermore, the center of origin of pathogens and epidemics can be examined by analyzing the correlation between genetic diversity and geographical distances as recently done in the oomycete pathogen *Phytophthora ramorum* (Kamvar et al. 2015b). If samples are taken in the center of origin, then as the geographical distance increases, the genetic diversity should also increase. Conversely, if samples are taken outside the center of origin their diversity should be reduced while geographical distances increase. Then, there will not be significant correlations with samples from outside the center of origin (Kamvar et al. 2015b). Based on this rationale, the spatial correlation analysis (Figure 2.7) supports that the center of origin of *M. roreri* covers Colombia (the Magdalena valley), Ecuador

(coastal region) and the Peruvian Upper Amazon (Maynas province). Knowing that the center of origin of *T. cacao* is in the Upper Amazonian regions from Ecuador, Peru, Colombia and Brazil (Motamayor et al. 2008, Thomas et al. 2012, Osorio-Guarín et al. 2017), it can be said that the center of origin of *M. roreri* proposed in this study does not fully match with the strict Amazonian origin of its host.

The coast of Ecuador has a long history of cultivation of cacao (Chapter 1) and agricultural systems are coincidently the major drivers for the emergence of plant pathogens in places outside of the center of origin of the host (Stukenbrock and McDonald 2008). It has been about 200 years since *M. roreri* was first observed (Phillips-Mora et al. 2007a) thus it is a fairly recently emerged pathogen. One of the hypotheses for the emergence of plant pathogens are host shifts or host jumps, when adaptation to a new host from wild relatives or even different species without necessarily having the same center of origin occurs (Stukenbrock and McDonald 2008). Some species of *Moniliophthora* are known to be common endophytes of grass roots from semi-arid ecosystems (Aime and Phillips-Mora 2005, Khidir et al. 2010). Additionally, *M. roreri* and its sister species, *M. perniciosa*, also a cacao pathogen, have acquired pathogenicity genes from oomycetes and bacteria through horizontal gene transfer (Tiburcio et al. 2010). Therefore, it may be possible that *M. roreri*, once cacao started being extensively cultivated outside its center of origin, acquired pathogenicity genes and underwent a host shift to cacao. Hypotheses on the genetic causes for *M. roreri* pathogenicity are explored in Chapter 4.

The genotypes found in *M. roreri* samples from *T. grandiflorum* were also found in commercial cultivars of *T. cacao* (Figure 2.6B). Most of the Ecuadorian samples were taken from a single farm. The outer perimeter of this farm was planted with *T. grandiflorum* trees whose fruits would get infected by *M. roreri*. 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 (Figure 2.6B), meaning there is no sign of host-specificity. Conversely, it speaks of the wide host range of the pathogen (Phillips-Mora and Wilkinson 2007), and its capacity to persist in cacao

plantations by means of infecting an alternate host, which serves as a source of primary inoculum for the next cacao fructification and production season.

2.5.2 On the invasive history of *M. roreri*

The genetic structure of *M. roreri*, as revealed by the DAPC on both the SNP and SSR datasets, shows that the samples collected in Peru (except for Maynas) and Bolivia (all belonging to the same MLL; Figure 2.3 APPENDIX B) correspond to a genetically different lineage (Figures Figure 2.5A and Figure 2.8A). This *M. roreri* lineage was probably the one introduced to the north of Peru in 1988 (Hernández T. et al. 1990), spread to the south of the country in 1998 (Ríos-Ruiz and Rodríguez 1998) and invaded Bolivia in 2012 (Phillips-Mora et al. 2015). Additionally, the DAPC analyses revealed that the lineage that invaded Central America and Jamaica from 1958 to 2016 (Phillips-Mora et al. 2007a, Johnson et al. 2017) is genetically different from the one from Peru and Bolivia. These results support previous results that found *M. roreri* strains with a single mating type disseminated across Central America and strains with a second mating type invaded Peru (Díaz-Valderrama and Aime 2016a). It will be very important to examine the mating type loci of the samples analyzed in this study (Chapter 3). The DAPC analyses also revealed the specific alleles and SNPs that contribute to the genetic differentiation of both invasive lineages (Figures Figure 2.5B and Figure 2.8B).

The MSN on the SSR dataset shows that the two invasive lineages (the Peru/Bolivian and the Central American) derive from the Ecuadorian network rather than the Colombian (Figure 2.6B). Conversely, the pattern observed in the MSN on the SNP dataset is not the same and in fact it looks like Colombia covers the great majority of the network (Figure 2.9B). This is probably due to the under-sampling in Ecuador compared to Colombia (Ali et al. 2015). Regardless, the SNP genotypes of both invasive lineages can be found in Ecuador while only the Central American genotype is present in Colombia (Figure 2.8B). As extra evidence for the close genetic relationship of Central American and Ecuadorian samples, Ecuador samples in another study grouped with isolates from Mexico (Maridueña-Zavala et al. 2016). All this has an implication on the origin of *M. roreri* as an epidemic. Being that the first known outbreak of the disease in coastal Ecuador at the

beginning of the 20th century, it is likely that the *M. roreri* lineages that invaded Central America and Peru since the 1950s come originally from this region.

2.5.3 SSRs vs SNPs markers

The use of SNP markers has exponentially increased during the last three decades but it does not guarantee better performance better than other markers like SSRs (Schlötterer 2004, Guichoux et al. 2011). SSR markers can have more alleles per locus and have a much faster evolution rate than SNPs, thus SSR mutations require shorter periods of time to accumulate in newly established populations, which makes them more suitable to explain recent movements of organisms (Morin et al. 2004, Schlötterer 2004, Guichoux et al. 2011). This is consistent with the results found in this study, in which Central America and Peru (regions invaded by *M. roreri* between 1958 and 2006; Phillips-Mora and Wilkinson 2007) had a clonal richness and genotypic diversity higher in SSRs than in SNPs (Figures Figure 2.6A and Figure 2.9A). Additionally, in the most recent invaded countries (Bolivia in 2012 and Jamaica in 2016; Phillips-Mora et al. 2015, Johnson et al. 2017) the clonal richness was zero even with SSR's (Figure 2.6); i.e., there has not been enough time for the invasive *M. roreri* genotypes to accumulate SSR mutations.

The overall clonal richness detected in the proposed center of origin (Figure 2.7) with SSRs was higher than with SNPs (Figure Figure 2.6 vs Figure 2.9). In other words, SSRs detected more MLGs per number of samples collected than SNPs. Additionally, the genetic structure identified by SSRs and SNPs showed distinct patterns (Figure Figure 2.5 vs Figure 2.8). Despite using a DAPC, which maximizes the variation between pre-defined groups (Jombart et al. 2010), the SNP dataset was not able to fully discriminate samples from Ecuador and Colombia, which also supports the idea that the center of origin of *M. roreri* goes beyond Colombia. These results are comparable to those from a similar study on *Armillaria cepistipes* (Tsykun et al. 2017). They investigated the genetic structure of this facultative forest pathogen, closely related to *M. roreri* as they both belong to the Marasmiineae (Dentinger et al. 2015), with seventeen SSR and twenty-four SNP markers (Tsykun et al. 2017). These authors found that SSR markers had a better resolution power than SNPs to discriminate the geographical origin of *A. cepistipes.*

In addition to the fact that SSRs tend to be more polymorphic than SNPs (Guichoux et al. 2011), SSRs also allow you incorporate null alleles in the analyses. Null alleles are generated because of mutations where primers anneal, and deletions or insertions of the locus, all leading to absence of amplicon during PCRs. Null alleles can be informative especially if they are restricted to samples from specific locations (Grünwald et al. 2017). In fact, the great majority of the samples with at least one null allele came from Ecuador (APPENDIX B). This reveals that Ecuador might hide even more genetic diversity that was not captured in this study. Definitely, strong evidence is provided to reject the hypothesis of an exclusive Colombian origin of *M. roreri*. Further genetic studies should include *M. roreri* samples from the Upper Amazon from Ecuador, Colombia, Brazil and even Peru, in regions like Madre de Dios.

CHAPTER 3. RAPID METHODS TO CHARACTERIZE MATING TYPE LOCUS ALLELES IN *M. RORERI*

3.1 Introduction

The earliest reports of Frosty Pod Rot (FPR) disease of cacao (*Theobroma cacao* L.), caused by the fungus *Moniliophthora roreri*, possibly date back to 1817 and 1851 in Colombia; a disease with similar symptomatology to FPR affecting cacao pods was reported in local newspapers in farms from Santander and Antioquia, respectively (Ancízar 1853, Parsons 1949, Phillips-Mora 2003, Phillips-Mora and Wilkinson 2007). Decades later, during the 1910s, the best-known outbreak of the disease took place in western Ecuador (Phillips-Mora 2003), which caused a reduction of 20% in total national production (Erneholm 1948) leading to complete abandonment of plantations from 1916 to 1920 (Ciferri and Parodi 1933). During the 1950s, *M. roreri* underwent a dramatic geographical expansion throughout Latin America at the expense of small cacao farmers. It currently threatens cacao production in major producing countries, like Brazil and Western African and Southeast Asian countries, all FPR-free at this time (Arévalo and Hernández 1990, Phillips-Mora et al. 2006a, 2006b, 2015, Phillips-Mora and Wilkinson 2007, Johnson et al. 2017).

Moniliophthora roreri belongs to the Marasmiaceae in the Agaricales (Agaricomycetes, Basidiomycota) (Aime and Phillips-Mora 2005). Mating in this fungal phylum is typically regulated by either one (bipolar mating system) or two unlinked (tetrapolar mating system) loci (Lee et al. 2010, Heitman et al. 2013). One locus, referred to as the *A* mating locus, contains genes that code for homeodomain (HD) transcription factors, while the *B* mating locus harbors genes encoding pheromone receptors and pheromone precursors (Lee et al. 2010). The combination of alleles at each mating locus makes up the mating type of the fungal individual (Kües 2015). Molecular characterization established that *M. roreri* possesses a tetrapolar arrangement of the mating loci, and two mating types, *A1B1* and *A2B2*, have been identified (Díaz-Valderrama and Aime 2016a). Only samples with mating type *A1B1* have been found in Central America and Mexico, while samples having both mating types were found in South America (Díaz-Valderrama and Aime 2016a). *In vitro*

culture studies have demonstrated significant differences in vegetative growth and sporulation between *A1B1* and *A2B2* mating types, which may imply different levels of aggressiveness in the field (Díaz-Valderrama 2014).

The recent invasions of *M. roreri* to Bolivia and Jamaica in 2012 and 2016, respectively (Phillips-Mora et al. 2015, Johnson et al. 2017), have turned on the agricultural alarms because of the ease of dissemination among farms and the terrible consequences that FPR brings to cacao farmers in these counties (Imaña 2015, The Gleaner 2016). However, the mating types of these invasive strains remain unknown. Thus, one of the objectives of this study is to determine the mating types of *M. roreri* that recently invaded Bolivia and Jamaica. Additionally, the center of origin of *M. roreri* goes beyond the Magdalena Valley of Colombia as previously proposed (Ali et al. 2015), and extends to Ecuador and the Peruvian Upper Amazon (Chapter 2). Therefore, another objective is to analyze the mating type diversity in samples collected in these areas.

Current sampling and diagnostic strategies for *M. roreri* rely on isolation of pure cultures and subsequent Sanger sequencing of rDNA, typically the internal transcribed spacer (ITS) region (Phillips-Mora 2003, Phillips-Mora et al. 2007b, González Figueroa and Roble Orellana 2014); this type of sequence provides a confident diagnostic tool up to the species level and has been used to categorize *M. roreri* into two groups, termed Orientalis and Occidentalis (Phillips-Mora et al. 2007b). However, there are no diagnostic tools to identify *M. roreri* mating types. Therefore, this study also describes a simple sampling method of *M. roreri* that does not require pure culture isolation and provides PCR-based markers to easily detect the mating type of a sample without the need of Sanger sequencing.

3.2 Hypotheses

Based on the invasive history of *M. roreri*, the research hypotheses of this study are:

1. The mating types that have invaded Bolivia and Jamaica are *A2B2* and *A1B1*, respectively.
2. The two previously characterized mating types are present in the center of origin of the pathogen, i.e., Ecuador, Colombia and the Peruvian Upper Amazon.

3.3 Materials and Methods

3.3.1 Collection and DNA extraction

The samples investigated here are the same used in Chapter 2 (APPENDIX A). Diseased cacao pods were collected in paper bags. GPS coordinates, date, host or cultivar and other relevant information about the sample, tree and plantation were recorded (APPENDIX A). Isolations took place on the same day of harvest and were performed under the most possible aseptic conditions. We directly 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. We placed the white stroma and internal tissue in Eppendorf tubes (Eppendorf, Hamburg, Germany) containing 600 µL of Nuclei Lysis Solution from the Wizard® Genomic Purification Kit (Promega, Madison, WI) for DNA extraction without the need to perform pure culture isolation (Figure 3.1; APPENDIX A). To compare DNA extraction performance, we still performed traditional *M. roreri* isolations in pure culture in some samples. This included surface sterilization of pods in 2.5% sodium hypochlorite solution for three minutes and a rinse in sterile water for two minutes (Phillips-Mora 2003, González Figueroa and Roble Orellana 2014). Pods were dissected, and necrotic internal tissue was placed on Potato Dextrose Agar (PDA) media plates. At any sign of contamination, subcultures were performed until pure cultures were obtained. Samples used in this study came from 207 *M. roreri*-infected pods from 185 trees (APPENDIX A). Then, two samples may have come from the same pod but from different locations within the pod depending on the type of material collected, i.e., internal necrotic tissue, white stroma or pure culture isolation (APPENDIX A); if same-pod samples were from the same type of material, these were taken from different places, e.g., seed vs pulp necrotic tissue.

The Wizard® Genomic Purification Kit for DNA extraction was followed according to manufacturer's procedures. To speed up the initial grinding step, we used 2mm Zirconia beads (BioSpec Products, Bartlesville, OK) in a Mini-Beadbeater-24 (BioSpec Products) for 5 minutes. Quantification and assessment of purity of DNA were performed by measuring the ratios of absorbance at 260 nm over 280 nm (A260/280) and 230 nm (A260/230), widely used indicators of purity of nucleic acids (Teare et al. 1997, Gallagher and Desjardins 2006), in a nanoDrop™ One spectrophotometer (ThermoFisher Scientific,

Waltham, MA). Most proteins have the strongest absorbances at 280 nm while other impurities like phenols and salts, at 230 nm (Teare et al. 1997); therefore, A260/280 and A260/230 ratio values lower than the thresholds of 1.8 and 2.0, respectively, indicate contamination of protein and/or other impurities (Thermo Fisher Scientific, Teare et al. 1997, Wilfinger et al. 1997). Concentrations of DNA used in further PCRs were always between 0.2 to 4.0 ng/µl.

3.3.2 Primers specific for mating type and PCR

The genomes of the two invasive mating types of *M. roreri* are available (Meinhardt et al. 2014; Chapter 4), and their mating type loci have been fully characterized (Díaz-Valderrama and Aime 2016a). Therefore, those sequences served as template to design primers specific for each mating type locus and allele sequence to be used in diagnostics of *M. roreri* mating types (Table 3.1). The primers were designed in such a way that diagnosis of mating type alleles was based on the presence or absence of the amplicon band in the agarose gel. Some of these primers were designed in the flanking regions of the *A* and *B* mating loci to discover, via Long Range PCR (Curran et al. 1996), unknown mating alleles in samples for which primers for *A1B1* and *A2B2* mating types did not generate amplicons after at least three PCR attempts. Primers were designed using the Primer-BLAST tool from the National Center for Biotechnology Information (NCBI) (Ye et al. 2012).

3.3.3 Discovery of new mating alleles

Long-Range PCR products (Table 3.1) were sequenced using the *WideSeq* strategy in a small portion of a lane in the Illumina MiSeq platform implemented in the Purdue Genomics Core (https://www.purdue.edu/hla/sites/genomics/wideseq-2/), as in Kijpornyongpan et al. (2019). This approach involved the construction of Illumina Nextera® libraries in which PCR products were fragmented and tagged with adapter sequences to obtain paired-end reads. These reads were subsequently mapped to the known *A1B1* and *A2B2* mating type sequences with program BBMap 37 (Bushnell 2014) and assembled with SPAdes 3.11 (Bankevich et al. 2012). The genes in the assembled DNA sequence were predicted with FGENESH (Solovyev et al. 2006) using *Coprinopsis cinerea*

specific gene-finding parameters. If predicted genes match, via BLASTP searches (Camacho et al. 2009), either an *A* or *B* mating gene, primers were designed to specifically amplify the newly discovered mating allele (Table 3.1). Known and discovered mating alleles were compared, at the amino acid level, to each other using the local similarity algorithm (SIM; Huang and Miller 1991) within the ExPASy server (Gasteiger et al. 2003). Alignments were visualized with the program LALNVIEW 3.0 (Duret et al. 1996). Multiple sequence alignments were performed with MUSCLE (Edgar 2004). Conserved domains (CD) were identified with CD-searches in the CD Database (Marchler-Bauer et al. 2011) of the National Center for Biotechnology Information (NCBI).

3.3.4 rDNA sequencing

The internal transcribed spacer rDNA regions 1 and 2 and the 5.8S ribosomal subunit (ITS), the 28S large ribosomal subunit (LSU) and the 18S small sub unit (SSU) was sequenced to confirm the presence of *M. roreri* in newly invaded countries (Aime and Phillips-Mora 2005, Phillips-Mora et al. 2006a, 2006b, 2015, Johnson et al. 2017). rDNA sequencing also served as positive controls for samples for which mating type primers were unable to amplify. PCR amplification and sequencing were performed as described previously (Aime and Phillips-Mora 2005). Sequences were edited with Sequencher 5.0 (Gene Codes Corp., Ann Arbor, Michigan). Sequences were compared with the non-redundant database from the NCBI by *blastn* searches (Camacho et al. 2009). The alignments were performed with MUSCLE on MEGA-X (Kumar et al. 2018). Phylogenetic analyses were performed through the web portal Cipres Science Gateway (Miller et al, 2012) with the Randomized Accelerated Maximum Likelihood (RAxML) program (Stamatakis 2006, 2014). The number of bootstraps was automatically determined by the program via the so-called rapid bootstrapping algorithm (Stamatakis et al. 2008). Sequences from other studies were also incorporated in these analyses (Table 3.2).

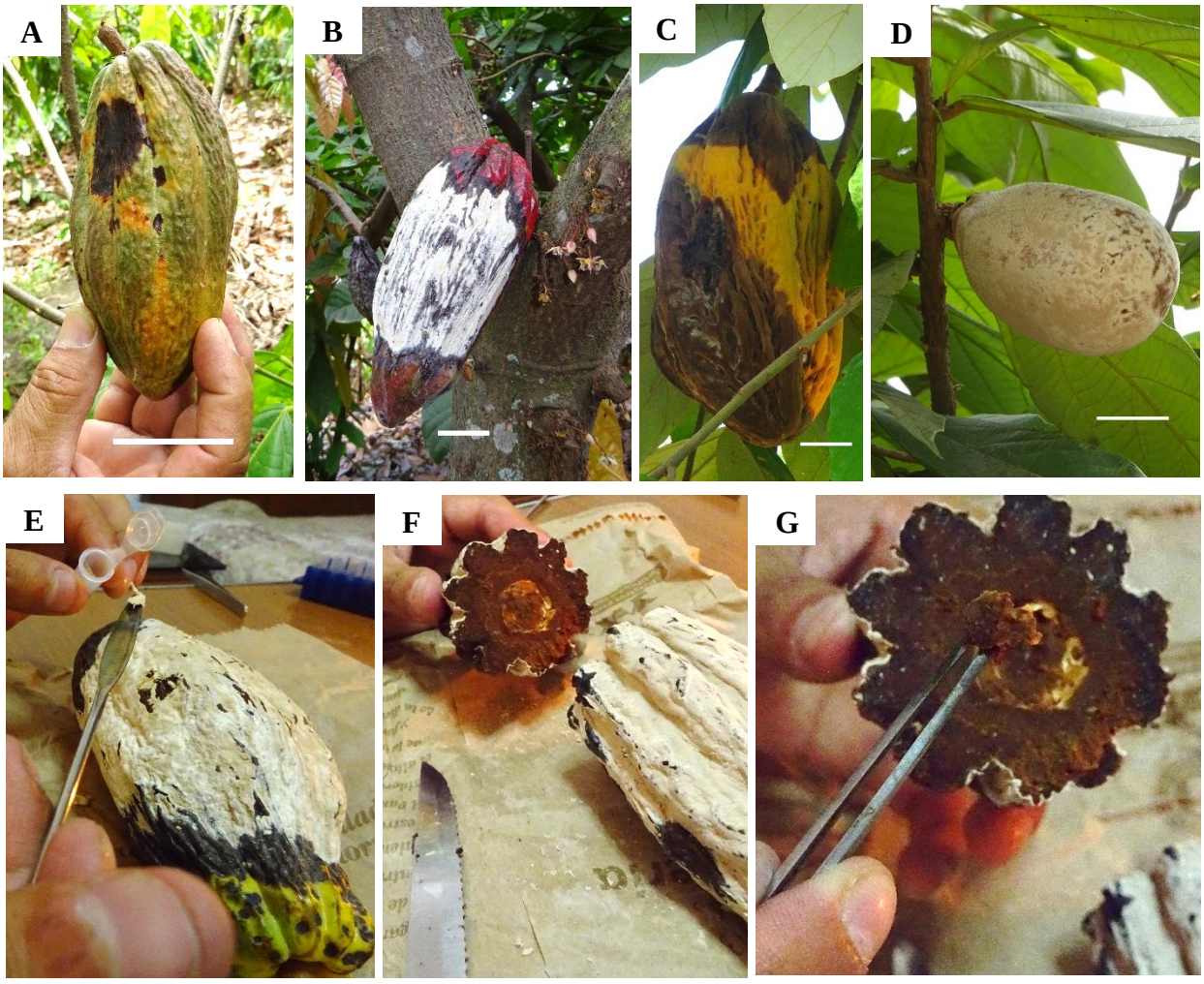


Figure 3.1 Direct collection method for *Moniliophthora roreri* without the need of pure culture isolation. **A)** Identification of brown spots on cacao pods, an early symptom of frosty pod rot. **B)** Identification of mummified cacao pods with white stroma on the surface, symptom and sign in an advanced stage of the disease. **C)** Symptoms (brown spots) and signs (white stroma) of FPR on *Theobroma bicolor* infected pod. **D)** Symptoms (mummification) and signs (white stroma) of FPR on *T. 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 place in Eppendorf tube with Nuclei Lysis solution. Scale bars = 5 cm.

Table 3.1 Specific primers to diagnose and identify the mating locus alleles in *Moniliophthora roreri*

Gene	Mating locus allele	Primer names and sequences (5' to 3')		PCR product size (bp)	Thermo-cycling conditions
Mr_HD1	*A1*	Mr_HD1_both_R	GGAAGAGTGATGGGCACAGA	1192	95ºC, 2 min; 35 cycles of 95ºC, 30 s 57ºC, 30 s 72ºC, 1min 45 s; 72 ºC, 5 min
		Mr_HD1_A1_F	AGTCTGCGGTGGACAATTTCA		
Mr_HD1	*A2*	Mr_HD1_both_R	GGAAGAGTGATGGGCACAGA	1166	
		Mr_HD1_A2_F	TATGAAGACCCAGCGCAAGT		
Mr_HD2	*A1*	Mr_HD2_Int_R	CTCTTCGTTCCTGCCTCGTT	1263	
		Mr_HD2_A1_R2	ATGGGTATTCCAACGGCCTCT		
Mr_HD2	*A2*	Mr_HD2_Int_R	CTCTTCGTTCCTGCCTCGTT	1275	
		Mr_HD2_A2_R2	ATGGGTATTCCGACGCTTCC		
STE3_Mr4	*B1*	Mr_Rec4_F2	CCCTCTGGAACCAAAGATTCTG	572	95ºC, 2 min; 35 cycles of 95ºC, 30 s 57ºC, 45 s 72ºC, 1min; 72 ºC, 5 min
		Mr_Rec4_R2	TGCACAGTCTGAGTAACGAGT		
STE3_Mr4	*B2*	Mr_R4_A2_F	ACATTGCGGTTCATCCCCAT	989	
		Mr_R4_A2_R	TAGATGAGCAAGCGTAGGCG		
Mr_Ph4	*B1*	Mr_Ph4_A1_F	CTTGCACGAAAGGCGAACAA	786	
		Mr_Ph4_A1_R	TTTATGTCGGAGGTGTGGGC		
Mr_Ph4	*B2*	Mr_Ph4_A2_F2	GGTGGACAAAAACTGGCGAC	622	
		Mr_Ph4_A2_R	GCAAAGGCACCTTACAGCTT		
Mr_HD1 / Mr_HD2	*A1* and *A3*	Mr_LR_A_F1[1]	CGAGAACCTTCCATACGACCTT	~ 7 kb (*A3*) - ~11 kb (*A1*)	95ºC, 2 min; 35 cycles of 95ºC, 30 s 57ºC, 45 s 72ºC, 9min; 72 ºC, 5 min
		Mr_LR_A_R1[1]	AGCTCTTTGGGTGTAAGAGCC		
STE3_Mr4 / Mr_Ph4	*B1* and *B2*	Mr_LR_B_F1[1,2]	GTCAGACGTACGACTCGAGAC	~ 6 kb (*B2*) - ~ 10 kb (*B1*)	
		Mr_LR_B_R1[1,2]	GAAATCACTACCGGGAAGGGT		

[1]Primers for Long Range PCR

[2]Primer combination unable to find other mating type alleles.

Table 3.2 GenBank accession numbers of rDNA of *Moniliophthora roreri* samples from other studies used in ITS, LSU and SSU phylogenetic analyses

Strain	**Country**	**ITS**	**LSU**	**SSU**
MCA2952	Mexico	DQ222923	DQ222924	—
MCA2997	Ecuador	From genome[1]	—	—
C21	Costa Rica	AY916746	AY916744	AY916745
MCA2954	Belize	DQ222927	DQ222928	—
MCA2953	Mexico	DQ222925	DQ222926	—
IMI506582	Jamaica	MF139030	—	—
IMI506584	Jamaica	MF139031	—	—
IMI506121	Jamaica	MF139032	—	—
IMI506123	Jamaica	MF139033	—	—
IMI506125	Jamaica	MF139034	—	—
C13	Costa Rica	JX315275	—	—
Co6	Colombia	JX315278	—	—
Co7	Colombia	JX315279	—	—
Co8	Colombia	JX315280	—	—
Co11	Colombia	JX315281	—	—
Co12	Colombia	JX315282	—	—
Co15	Colombia	JX315283	—	—
Co17	Colombia	JX315284	—	—
E16	Ecuador	JX315285	—	—
E18	Ecuador	JX315286	—	—
E32	Ecuador	JX315287	—	—
E43	Ecuador	JX315288	—	—
B1b	Bolivia	JX515287	JX515294	—
B2a	Bolivia	JX515288	JX515295	—
B3	Bolivia	JX515290	—	—
MCA2521	Ecuador	—	AY916750	MG717343
C22	Costa Rica	—	AY916749	—
MROCP	Mexico	—	—	KM998972

[1]ITS sequence was retrieved from genome sequence (Meinhardt et al. 2014)

3.4 Results

3.4.1 DNA extraction

The total DNA and the ratios of absorbance varied according to the nature of the sample (Figure 3.2). More than 75% of samples from internal necrotic tissue and white stroma (direct method; Figure 3.1) yielded more DNA than the median obtained from extractions of pure cultures of *M. roreri* (Figure 3.2A). More than half of necrotic tissue samples yielded DNA with A260/280 values below the threshold of 1.8 (Figure 3.2B), while DNA from more than half of white stroma samples and pure cultures had A260/280 values greater than 1.8 (Figure 3.2B). However, around 75% of all samples yielded DNA with A260/230 values less than the threshold of 2.0, including DNA extractions from pure cultures (Figure 3.2C); this situation was worse in samples from necrotic tissues, which nearly 100% of them had A260/230 ratios below 2.0 (Figure 3.2C).

3.4.2 Analysis of *A* mating alleles

Using the *WideSeq* approach we discovered a new *A* mating allele in samples from the municipality of Nilo, Cundinamarca in Colombia, called from now on allele *A3* (Figure 3.3; APPENDIX D). Gene predictions reveal it possesses two genes that encode for homeodomain transcription factors Mr_HD1.3 and Mr_HD2.3 with their respective homeobox fungal mating domains (APPENDIX D). A conserved domain from Mr_HD1.3 is identical to the one found in Mr_HD1.2 and it differs only in one amino acid to Mr_HD1.1 (APPENDIX D), while conserved domains from the three variants of Mr_HD2 are identical to each other (APPENDIX D). Identity ranges for Mr_HD1 variants go from 82.3% to 86.6% while for Mr_HD2 alleles, from 81.6% to 89.4% (Figure 3.3). In all cases the N-terminus of the transcription factors are more dissimilar to each other, and the C-termini are more similar (Figure 3.3; APPENDIX D). Then, diagnostic primers specific for allele *A3* were designed from this sequence (Table 3.1).

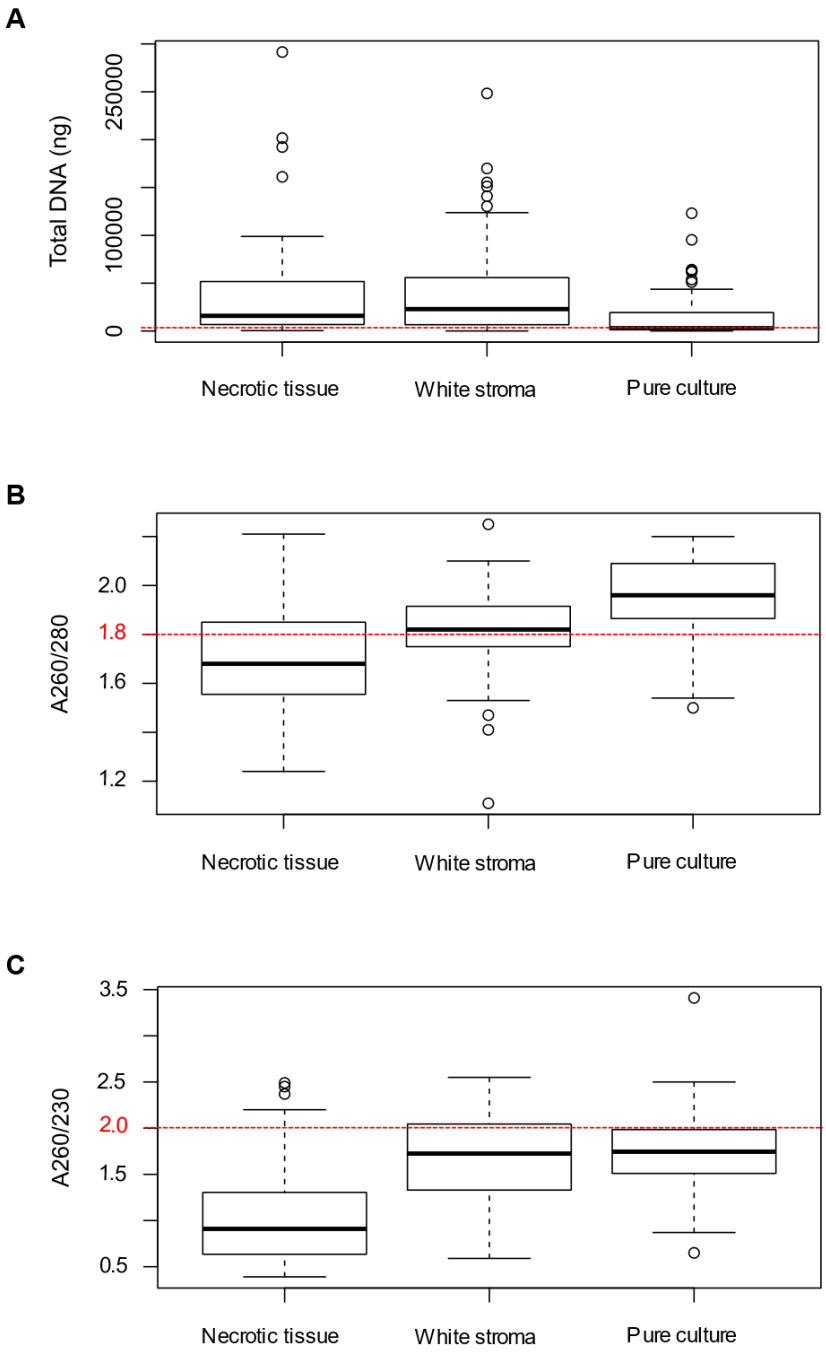


Figure 3.2 Spectrophotometer results of *Moniliophthora roreri* collections according to the type of sample: internal necrotic tissue (n = 55) and white stroma (n = 112) directly on DNA extraction buffer (see Materials and Methods and Figure 3.1), and pure culture DNA isolation (n = 56). **A)** Total DNA obtained (ng); red dotted line indicates de median of total DNA obtained from pure cultures. **B)** Ratio of absorbance at 260 over 280 nm; red dotted line indicates the expected ratio (1.8) for no-protein contaminated samples (Thermo Fisher Scientific). **C)** Ratio of absorbance at 260 over 230 nm; red dotted line indicates the minimum ratio (2.0) for "pure" nucleic acid samples

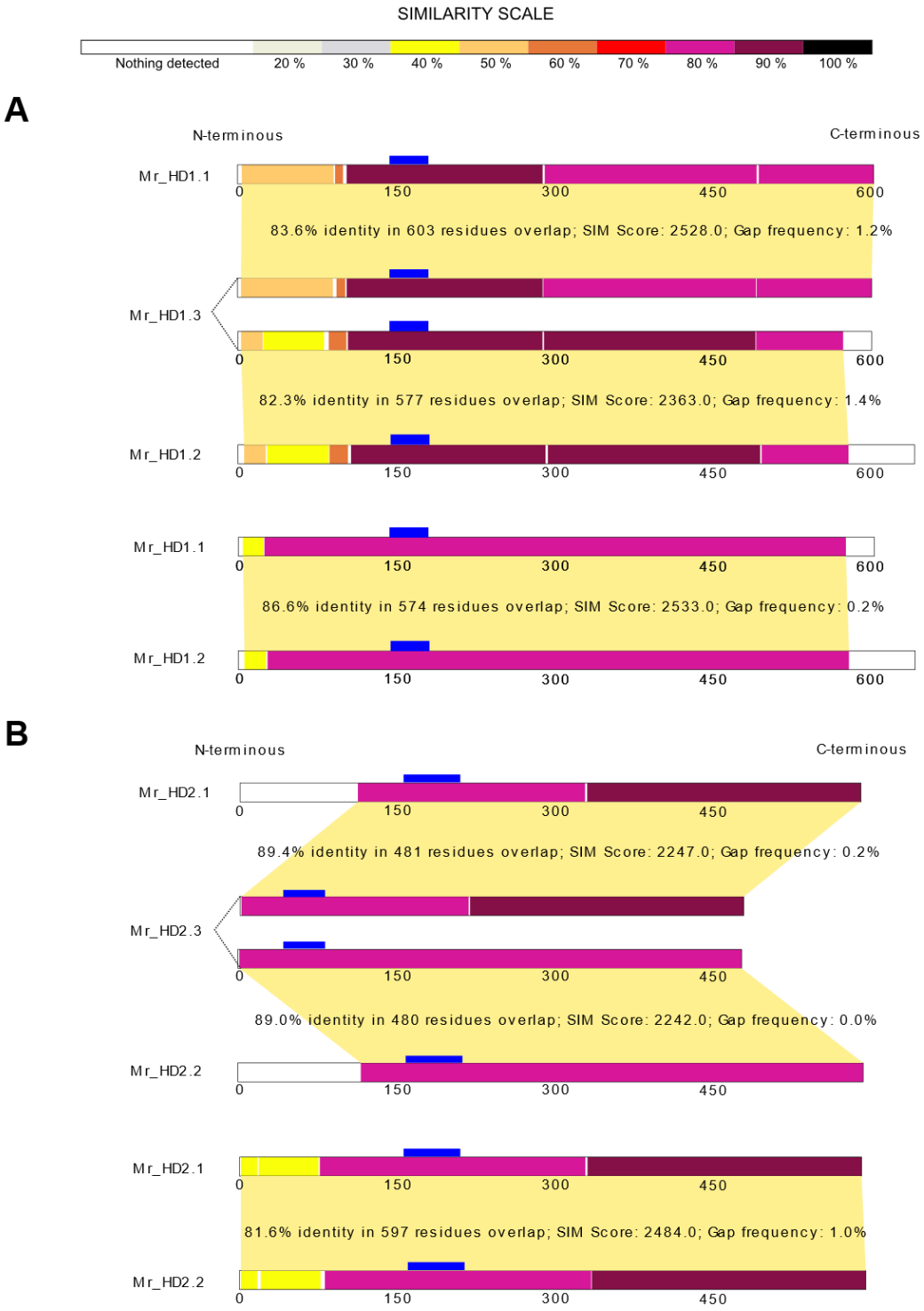


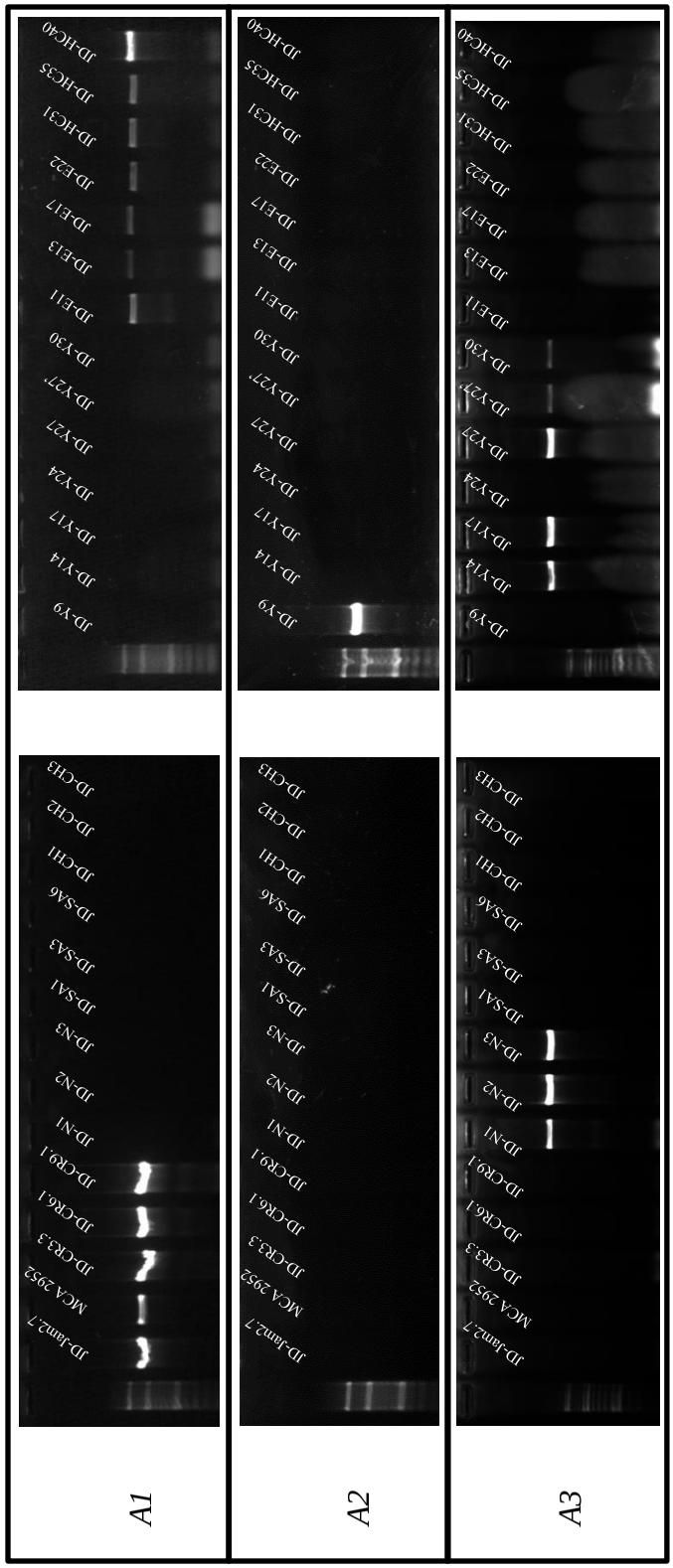
Figure 3.3 Graphical representation of local similarity alignments (SIM) at the amino acid level among the newly discovered *A3* and the known *A1* and *A2* mating alleles of *Moniliophthora roreri*. **A)** Alignments among HD1 proteins. **B)** Alignments among HD2 proteins. Blue boxes represent the homeobox conserved domains. Yellow shades joining HD variants represent the sequence overlap in alignment. Numbers indicate positions of amino acids. Statistics come from the output of SIM in each alignment.

3.4.3 Validation of specific primers and mating type analysis

The new specific primers determined the allele of each mating type locus in the majority of *M. roreri* samples (Figures Figure 3.4 and Figure 3.5;APPENDIX A). The mating type, or combination of alleles at each mating type locus, of samples from countries other than Ecuador, Colombia and in the Maynas province in Peru, were all determined with our primers. All isolates from Central America, Mexico and Jamaica have *A1B1* mating type, while isolates from Bolivia and Peru, except the ones from the Maynas province, possess *A2B2* mating type (Figure 3.6; APPENDIX A). Also, one sample (MCA2997) from Los Ríos in Ecuador has mating type *A2B2* (Figure 3.6).

Moniliophthora roreri in Ecuador and Colombia have the highest mating type diversity. Primers did not fully determine the mating type of all samples from these countries; in some cases, either the *A* or *B* mating allele was only determined (Figure 3.6; APPENDIX A). In Colombia, besides samples from Nilo, we found allele *A3* was present in samples from the municipality of Yacopí, also in Cundinamarca; this allele was not found outside Cundinamarca department (Figure 3.6). Also, the *B* allele for all Colombian samples was successfully identified as *B2*, except for sample JD-Y9 that harbors an undetermined *B* allele (Figure 3.6). Then, we determined that eight Colombian samples have the novel mating type *A3B2* (Figure 3.6). In the Guayas province in Ecuador we found fourteen samples harboring a new combination of mating alleles and thus a new mating type, *A1B2*; fifteen samples have an unknown *A* but known *B* mating allele; two samples with unknown mating alleles for *A* and *B* loci; and fifteen samples harboring the known *A1B1* mating type (Figure 3.6). The two samples from Esmeraldas province harbor mating type *A1B1* as well (Figure 3.6). Additionally, the *A* and *B* alleles in the samples from Maynas province, Peru remain unknown (Figure 3.6; APPENDIX A). Mating type *A2B1* was never observed in this study. In most cases where more than one sample per diseased pod were taken, the same mating type was always recovered, except for samples JD_E6 and JD_E7, which both come from samples of internal necrotic tissue but have *A1B2* and *A1B1* mating types, respectively (APPENDIX A). Finally, designed primers are effective to diagnose the two invasive genotypes based on mating type loci (*A1B1*, *A2B2*) and to detect the new *A* mating allele.

Figure 3.4 Gel photographs of PCR products amplified with primers designed for diagnostics of *A* mating alleles of *Moniliophthora roreri*. Primers used in here were specific for the *Mr_HD1* gene (Table 3.1). Samples are arranged vertically while *A* allele photographs, horizontally. The molecular ladder used was 100 bp O'RangeRuler™ (Thermo Fisher Scientific); then the most conspicuous bands in ladders are 1500, 1000 and 500 bp.


A1
JD-Jam2.7
MCA 2952
JD-CR3.3
JD-CR6.1
JD-CR9.1
JD-N1
JD-N2
JD-N3
JD-SA1
JD-SA3
JD-SA6
JD-CH1
JD-CH2
JD-CH3
A2
JD-Jam2.7
MCA 2952
JD-CR3.3
JD-CR6.1
JD-CR9.1
JD-N1
JD-N2
JD-N3
JD-SA1
JD-SA3
JD-SA6
JD-CH1
JD-CH2
JD-CH3
A3
JD-Jam2.7
MCA 2952
JD-CR3.3
JD-CR6.1
JD-CR9.1
JD-N1
JD-N2
JD-N3
JD-SA1
JD-SA3
JD-SA6
JD-CH1
JD-CH2
JD-CH3
A1
JD-HC43
JD-IO1.1
JD-IO4
JD-SUB3.1
JD-Ja10
JD3
JD-Tar1
JD-HU05
JD-OUT7.5
JD-Xio4
JD-AB1.1
JD-AB5.2
JD-AB9.1
Negative
A2
JD-HC43
JD-IO1.1
JD-IO4
JD-SUB3.1
JD-Ja10
JD3
JD-Tar1
JD-HU05
JD-OUT7.5
JD-Xio4
JD-AB1.1
JD-AB5.2
JD-AB9.1
Negative
A3
JD-HC43
JD-IO1.1
JD-IO4
JD-SUB3.1
JD-Ja10
JD3
JD-Tar1
JD-HU05
JD-OUT7.5
JD-Xio4
JD-AB1.1
JD-AB5.2
JD-AB9.1
Negative

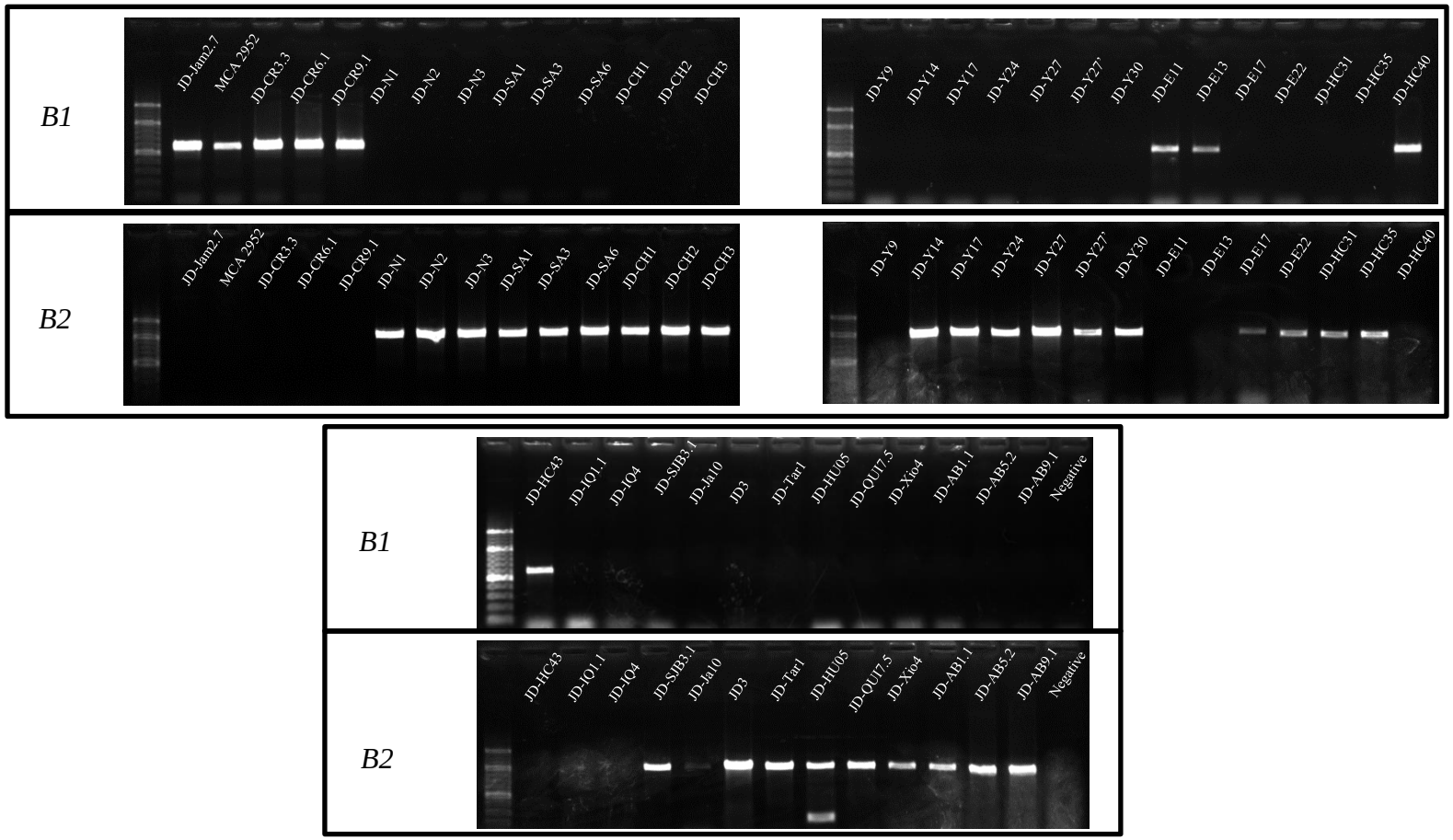


Figure 3.5 Gel photographs of PCR products amplified with primers designed for diagnostics of *B* mating alleles of *Moniliophthora roreri*. Primers used were specific for the *STE3_Mr4* gene (Table 3.1). Samples are arranged vertically while *B* allele photographs, horizontally. The molecular ladder used was 100 bp O'RangeRuler™ (Thermo Fisher Scientific); then the most conspicuous bands in the ladders are 1500, 1000 and 500 bp.

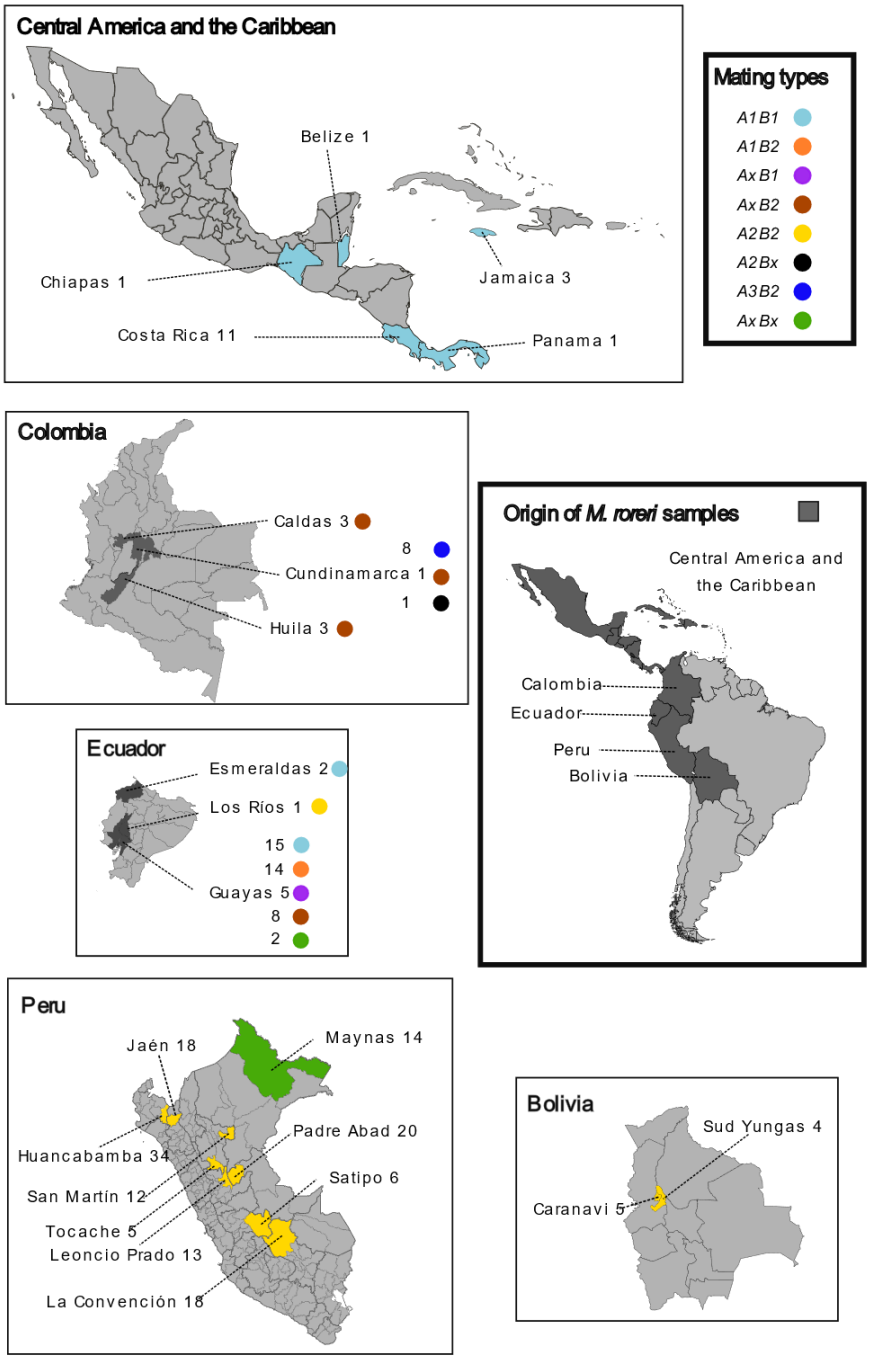
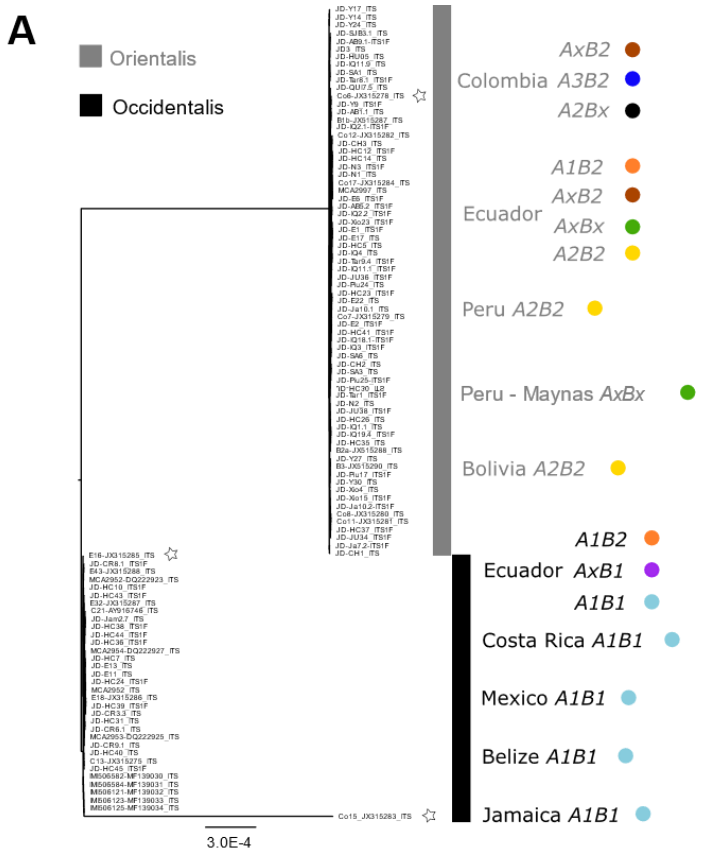


Figure 3.6 Geographical distribution of *Moniliophthora roreri* mating types. Mating type of the sample from Los Ríos, Ecuador was determined previously (Díaz-Valderrama and Aime 2016a) by looking at its genome (Meinhardt et al. 2014). The political division (provinces for Ecuador, Peru and Bolivia; department for Colombia; and country for the others) from where samples come are colored according to the mating type. If more than one mating type is present, colored dots are located next to the name of the division; number of samples is also specified. Map scales are not necessarily equal among zoom-in maps and boundaries are approximate. Maps modified from https://commons.wikimedia.org

3.4.4 rDNA sequence analyses

Isolates carrying undetermined mating alleles were confirmed to be *M. roreri* with rDNA sequencing. Phylogenetic analysis of ITS sequences splits isolates into the two previously circumscribed *M. roreri* groups, "Orientalis" and "Occidentalis", and no new groups were identified despite a few unique single nucleotide polymorphisms (SNPs) detected in some samples and mating types (Figure 3.7A). All Central American and Jamaican isolates were part of the Occidentalis group, while samples from Colombia, Peru and Bolivia were part of the Orientalis group (Figure 3.7A). Isolates from Ecuador were placed indistinctly in both groups (Figure 3.7A). Even though three different mating types are in the Occidentalis group, all members, except for isolate E16 from previous study (Phillips-Mora et al. 2007b), had 100% identical ITS sequences (Figure 3.7B). ITS sequences from samples within the Orientalis group have more DNA variation even though they had the same mating type. For example, isolates with the mating type *A3B2* had different ITS-SNP profiles (Figure 3.7B). Similar results were found with the LSU sequence analysis, however there was no DNA variation among SSU sequences of mating types from the Orientalis group (APPENDIX E).

Moreover, ITS sequencing from internal tissue or white stroma from some pods, even some that were positive for *M. roreri* like JD_Ja7.1 and JD_E25 (APPENDIX A), revealed the presence of other fungi associated with advanced stages of FPR (Table 3.3). Members of the Hypocreales were commonly recovered, especially species of *Acremonium* and *Fusarium* (Table 3.3). Almost all recoveries belonged to the Ascomycota, except for sample JD_IQ12.3, for which *M. perniciosa*, the sister species of *M. roreri* and cacao pathogen too, was obtained in ITS sequencing (Table 3.3), likely due to co-infection. *Blastn* top matches in all cases were endophytes or putative plant pathogens (or known plant pathogens in the case of *M. perniciosa*; Table 3.3).



B

Mating type	ITS group	ITS1 (6-232) 47	48	51	115	116	400	ITS2 (394-660) 435	451	531	582	
A2B2	Orientalis	-	-	G	-	G	A	-	-	C	-	
A1B2	Orientalis	-	-	G	-	G	A	-	-	C	-	
AxB2	Orientalis	-	-	G	-	G	A	-	-	C	-	
AxBx	Orientalis	-	-	G	-	G	A	-	-	C	-	
A3B2	Orientalis	-	-	G	-	G	A	-	-	C	-	
*A3B2**	Orientalis	G	G	G	-	-	A	-	-	C	-	
A2Bx	Orientalis	-	-	G	G	G	A	G	-	C	-	
Co6 JX315278 ☆	Orientalis	-	-	G	-	G	A	G	-	C	-	
A1B1	Occidentalis	-	-	G	-	-	-	-	T	T	G	
AxB1	Occidentalis	-	-	G	-	-	-	-	T	T	G	
A1B2	Occidentalis	-	-	G	-	-	-	-	T	T	G	
E16 JX315285 ★	Occidentalis	-	-	G	-	G	A	-	-	T	-	
Co15 JX315283 ★	Occidentalis	-	-	A	-	G	-	-	T	T	G	

Figure 3.7 ITS phylogenetic analysis, sequence profiling and association analysis of *Moniliophthora roreri* mating types. **A)** Maximum likelihood phylogenetic tree using ITS sequences of samples from all the countries and mating types found in this study. **B)** SNP comparison among mating types and the ITS groups Orientalis and Occidentalis; some samples with the same mating types have different ITS-SNP profile; *A3B2* mating type samples marked with asterisk having that ITS-SNP profile are JD-Y14, JD-Y17, JD-Y24; numbers indicate the relative position of SNPs in the alignment used for the phylogenetic tree (APPENDIX B); no SNPs were identified in the 5.8S rDNA region; stars indicate samples E16, Co6 and Co15 for which ITS sequences were retrieved from GenBank (accession numbers shown), then it was not possible to determine their mating type.

Table 3.3 *Blastn* top matches (as of Oct. 2018) of ITS sequences from samples collected with the direct method

Sample	Type of sample	Province	Country	Blastn top match	Query cover	E value	Ident	Accession	Origin	Host
JD_E25	NL - white stroma	Guayas	Ecuador	*Penicillium* sp.	100%	0	100%	KX953563	Mexico	Endophyte - *Vanilla planifolia*
JD_Piu18	NL - white stroma	Morropón	Peru	*Fusarium solani*	100%	0	100%	MG751209	Brazil	Stem endophyte - *Hevea braziliensis*
JD_Piu27-1	NL - white stroma	Huancabamba	Peru	Hypocreales sp.	99%	0	96%	KF435924	Panama	Leaf endophyte - *Poulsenia armata*[1]
JD_Piu30	NL - tissue	Huancabamba	Peru	*Diaporthe melonis*	100%	0	100%	MH465228	Mexico	Putative pathogen - *Heliconia sp.*
JD_Ja1	NL - white stroma	Jaén	Peru	*Acremonium* sp.	99%	0	98%	FR682361	South Africa	Putative pathogen – grapevine
JD_Ja7.1	NL - white stroma	Jaén	Peru	*Acremonium* sp.	97%	0	99%	EF042103	South Africa	Putative pathogen – grapevine
JD_Ja7.3	NL - white stroma	Jaén	Peru	Hypocreales sp.	82%	0	95%	KF435924	Panama	Leaf endophyte - *Poulsenia armata*[1]
JD_IQ1.2	NL - white stroma	Maynas	Peru	*Acremonium* sp.	97%	0	99%	KF435993	Panama	Leaf endophyte - *Rhizophora mangle*[1]
JD_IQ12.2	Isolation on PDA	Maynas	Peru	*Fusarium* sp.	100%	0	98%	KR350652	Mexico	Putative pathogen - *Laelia sp.*
JD_IQ12.3	NL - tissue	Maynas	Peru	*Moniliophthora perniciosa*	100%	0	100%	AY216468	Brazil	Pathogen - *Theobroma grandiflorum*
JD_IQ19.1	NL - white stroma	Maynas	Peru	*Bioneactraceae* sp.	75%	0	98%	MH267846	Peru	Inner bark endophyte - *Hevea pauciflora*

[1]These matches come from Higginbotham et al. (2013); other matches were available in GenBank but there were only unpublished studies associated with them.

3.5 Discussion

3.5.1 Direct sampling method

Pure culture isolation of *M. roreri* has traditionally been the first step to study this pathogen in the laboratory (Phillips-Mora 2003). This allows the extraction of good-quality DNA from fresh mycelia for molecular analyses and, with proper maintenance of pure culture isolates, it allows DNA re-extraction in the long term. However, the process of isolation, especially if sampling occurs in remote areas, can be a limiting factor because of potential contamination during isolation and transport. To mitigate this issue, in addition to attempting to isolate pure cultures, necrotic internal tissue and external white stroma if present, from FPR-diseased cacao pods were collected directly into DNA extraction buffer (Figure 3.1). DNA extraction from pure culture isolates tends to yield less contaminated DNA (Figure 3.2B and C) but in most cases it yields much less DNA than extractions from internal necrotic tissue and white stroma directly collected (Figure 3.2A). Therefore, the direct method of sampling ensured workable and high amounts of DNA for identification and mating type diagnosis of *M. roreri* (Figure 3.2). Moreover, most mummified pods in the field are so desiccated that performing pure culture isolation is impossible and the collection of white stroma directly into DNA extraction buffer may be the only way to sample. Additionally, white stroma or internal necrotic tissue sampled directly on DNA extraction buffer-containing Eppendorf tubes are easy to transport. Then, the proposed direct method is an excellent alternative when only mummified pods are commonly present in a plantation and when sampling occurs in remote areas, which is where wild cacao relatives (*Theobroma* spp.) are located (Thomas et al. 2012). However, whenever it is possible, isolation of *M. roreri* in pure culture is still recommended.

When sampling internal tissue, external DNA from other sources, such as other fungi associated with the pathogen in advanced stages of the disease, is likely to be present. Some fungi associated with advanced stages of FPR in this study have been previously reported to be endophytes or putative plant pathogens (Table 3.3). It has been shown that endophytes may become saprotrophs once the host dies, e.g., *Fusarium* spp. (Promputtha et al. 2007). Considering that all these samples also come from either necrotic tissue or white stroma, it is possible that the fungal taxa found in necrotic pods function as endophytes within healthy

cacao pods. This is supported by the fact that *Fusarium* sp. and *Acremonium* sp., frequently found in this study (Table 3.3), are listed as easily recoverable endophytes from cacao branches (Rubini et al. 2005). A more thorough investigation regarding cacao pods endophytes and *M. roreri*-associated fungi is needed to fully understand the interaction of these fungi in cacao.

3.5.2 Characterization of mating loci

The mating loci, *A* and *B*, of *M. roreri* are arranged in a tetrapolar manner, i.e., they are unlinked, and only *A1*, *A2*, *B1* and *B2* alleles had been previously characterized (Díaz-Valderrama and Aime 2016a). Only clonal isolates with mating type *A1B1* were responsible for the spread of FPR throughout Central America and more recently in Jamaica, as shown here (Figure 3.6), while mating type *A2B2* was responsible for the spread in major cacao-growing areas of Peru and Bolivia; this confirms a clonal dissemination of the pathogen throughout Latin America (Figure 3.6; Phillips-Mora et al. 2007b, Díaz-Valderrama and Aime 2016a). Even though Peru and Bolivia only have clonal populations with *A2B2* mating type, allele *A2* was not found anywhere else in this study. It is known that isolate MCA 2997 from Los Ríos, Ecuador harbors mating type *A2B2* (Meinhardt et al. 2014, Díaz-Valderrama and Aime 2016a), suggesting Ecuador has both mating types in its territory (Figure 3.6). Moreover, at least one undetermined mating type in *M. roreri* was detected in all samples from Maynas province in Peru (Figures Figure 3.4 and Figure 3.5), suggesting that more areas in Peru need to be explored to capture the whole mating type diversity within the country. On the other hand, a new allele *A3* and mating type *A3B2* was discovered in Cundinamarca, Colombia, as well as mating types never observed before like *A1B2* were found in Ecuador. These results are consistent with results found in Chapter 2: the higher number of genotypes observed in Colombia and Ecuador compared to Central America, Peru and Bolivia in previous studies (Chapter 2; Ali et al. 2015).

Moniliophthora roreri is only found in nature as a haploid and not as a sexual fruiting body or dikaryotic hyphae, which is the product of compatible mating between two haploid hyphae in the Basidiomycota (Brown and Casselton 2001), as has been reported (Díaz-

Valderrama and Aime 2016b, Bailey et al. 2018a). Interestingly, samples JD_E6 and JD_7 have different mating types (*A1B2* and *A1B1*, respectively) but come from the same pod. This means that different mating types may co-exist in the same host. In this specific case, mating would never occur, because in tetrapolar species it only occurs between haploid hyphae with distinct mating alleles at each locus (Brown and Casselton 2001). It will be interesting to find seemingly compatible mating types within the same pod. If this observation happens and dikaryotic hyphae is still not found, the lack of sexual reproduction in *M. roreri* may be governed by factors other than the mating type loci.

3.5.3 Primers for diagnostics

This study provides a list of primers to be used in diagnosis of *M. roreri* mating types (Table 3.1). These primers effectively discriminate between all *A* and *B* mating alleles found in this study without the need of Sanger sequencing (Figures Figure 3.4 and Figure 3.5). They allow the rapid detection of *M. roreri* mating types in areas were the disease has not been observed and to conduct mating type distribution analyses in places with a long history of FPR. It is important to note that there is at least one more allele in each mating locus which our primers are not able to capture (Figure 3.6). This is mainly because mating alleles vary greatly. For example, percent identities at the amino acid level of *A* mating alleles from the model mushrooms *Coprinopsis cinerea* and *Schizophyllum commune* range from 42% to 72% (Stankis et al. 1992, Badrane and May 1999). This has kept the design of primers to discover new mating alleles from being completely successful. Further genomic approaches will help to determine the sequences of these hypervariable regions.

3.5.4 rDNA sequence analysis

The genetic diversity of *M. roreri* is mainly dictated by the mating type of the sample (Díaz-Valderrama and Aime 2016a). Unfortunately, this diversity is not fully captured by rDNA sequences and it only divides *M. roreri* into two groups: Orientalis and Occidentalis (Phillips-Mora et al. 2007b). The rDNA phylogenies in this study show these groups do not have a correlative relationship with the mating type of the sample. For example, samples with the same mating type may have different ITS sequences and even belong to either rDNA group, regardless of origin (Figure 3.7). The only main conclusion that can be

drawn is, within our samples, allele *B1* is only found in samples from the "Occidentalis" group (Figure 3.7). Therefore, rDNA sequencing does not properly distinguish *M. roreri* isolates according to their mating type (Figure 3.7; Table 3.1). It is only good to diagnose the presence of the fungus up to the species level.

3.5.5 Final remarks

Before the 1950s, FPR was confined to Ecuador, Colombia and Western Venezuela (Phillips-Mora et al. 2007b). This is consistent with the fact that, in samples from Guayas, Ecuador and Cundinamarca, Colombia, we found new combinations of mating alleles and thus mating types that have never been observed in previous studies (Díaz-Valderrama and Aime 2016a). Since 1956 FPR started a geographical expansion throughout other countries in South America, all of Central America, and very recently to the Caribbean, specifically in Jamaica in 2016 (Phillips-Mora and Wilkinson 2007, Phillips-Mora et al. 2015, Johnson et al. 2017). With our primers we were able to successfully determine the mating types of all the samples from countries that are suffering from the disease after the 1950s (Figure 3.6): southwards, Perú and Bolivia (*A2B2*); northwards, Panamá, Costa Rica, Mexico, Belize and Jamaica (*A1B1*).

Moreover, we are reporting for the first time the presence of *M. roreri* in the Maynas province in Peru. Unfortunately, our primers were unable to determine the mating type of these samples (Figures Figure 3.4 and Figure 3.5), even though their ITS sequences are 100% identical to the rest of *M. roreri* samples belonging to the "Orientalis" group (Figure 3.7). Interestingly, these samples not only come from cultivated cacao but from wild relatives like the Macambo tree (*Theobroma bicolor*). Since the center of origin of *Theobroma* spp. is in the Peruvian Upper Amazon, the region where the Maynas province is located, we hypothesize that *M. roreri* was present in Perú (specifically in the Peruvian Upper Amazon) before the first official report of FPR in Perú in 1988 in the Utcubamba province, Amazonas region (Arévalo and Hernández 1990, Hernández T. et al. 1990). Based on the discovery of new mating alleles we conclude that the geographical areas harboring more mating types are Ecuador, Colombia and the Peruvian Upper Amazon (Maynas province), which is congruent with the findings in Chapter 2.

CHAPTER 4. EFFECTOROME AND CAZYOME COMPARISON IN THE MARASMIINEAE

4.1 Introduction

Moniliophthora (Agaricomycetes, Agaricales, Marasmiineae, Marasmiaceae) is an enigmatic genus within an order comprised mainly of saprotrophic and ectomycorrhizal fungi (Matheny et al. 2006, Moore et al. 2011, Dentinger et al. 2015). At present, there are seven formally described and one non-described species of *Moniliophthora* from tropical-humid to subtropical-semiarid ecosystems (Table 4.1). Most *Moniliophthora* spp. are assumed to be saprotrophic in humid forests and one is a root endophyte of grasses (Table 4.1). However, the most anthropogenically important species are *M. roreri* (Cif.) H.C. Evans, Stalpers, Samson & Benny and *M. perniciosa* (Stahel) Aime & Phillips-Mora as they are major cacao pathogens in the Americas, which also severely threaten worldwide production (Table 4.1; Chapter 1). An intriguing aspect is that pathogenicity is not very common in the Agaricales, and proportionately few species are plant pathogens, most of which, like *Moniliophthora*, belong to the Marasmiineae (Table 4.2). Both *M. roreri* and *M. perniciosa* are considered hemibiotrophic pathogens (although for *M. roreri* it has not been fully proved) meaning that during cacao infection they obtain nutrients from living cells, a.k.a. the biotrophic phase, and in advanced disease stages they start killing host cells and acquire nutrients from them, a.k.a. the necrotrophic phase (Mondego et al. 2008, Bailey et al. 2013). In general, the success of fungal pathogens during plant infection depend greatly on the biochemical compounds they produce upon interaction with their host. Therefore, they have evolved a complex repertoire of enzymes and small secreted proteins (SSP) that allow them to overcome a host defense response (Zhao et al. 2013, Kim et al. 2016).

Among the most important enzymes involved in host-microbe interactions are carbohydrate-active (CAZy) and lignin degradation enzymes. CAZymes are required for the breakdown and/or synthesis of multiple carbohydrate and glycoconjugate biopolymers (Cantarel et al. 2009, Zhao et al. 2013). They have been thoroughly databased (www.cazy.org; Lombard et al. 2013) and broadly grouped into classes based on their

catalytic activity: glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and carbohydrate-binding modules (CBMs). Some enzymes, originally classified as GHs and CBMs, have been shown to be lytic polysaccharide monooxygenases (LPMOs) instead (Vaaje-Kolstad et al. 2005, Forsberg et al. 2011, Quinlana et al. 2011), which act directly on polysaccharide chains, like cellulose (Hemsworth et al. 2015). Cellulose and lignin are intimately linked structural components of plant cell walls; their degradation involves different enzymatic reactions that occur in an orchestrated fashion. Therefore the CAZy database curators incorporated a new lignocellulosic CAZy class termed "Auxiliary Activities" (AA) that groups LPMOs and redox enzymes involved in degradation of lignin (Levasseur et al. 2013, Lombard et al. 2013).

Another important set of biochemical compounds that influences the success of plant-parasitic fungi is effector proteins. They are SSPs that target the plant immune system either leading to a successful colonization or triggering a defense response from the host (Presti et al. 2015, Kim et al. 2016). In contrast to other SSPs, effector proteins are short, usually < 300 amino acids, with low molecular weight, are rich in cysteine and tend to have less serine and tryptophan content (Sperschneider et al. 2016, Toro and Brachmann 2016). Other characteristics of effector proteins are that they may contain repetitive amino acid motifs, are transported to the host nucleus and thus have a nuclear localization signal (NLS), and tend to occupy genomic regions with high content of repetitive DNA (Jones and Dangl 2006, Raffaele et al. 2010, Saunders et al. 2012).

CAZyme families and effector proteins from pathogenic fungi vary according to the type of interaction with their host. Necrotrophic and hemibiotrophic fungi tend to have more CAZymes than biotrophic ones since the latter acquire their nutrients from living cells during the entire life cycle (Mendgen and Hahn 2002, Duplessis et al. 2011, Zhao et al. 2013). Conversely, biotrophic pathogens tend to have more effector proteins than necrotrophic and hemibiotrophic pathogens (Kim et al. 2016). The question here then is whether *Moniliophthora* spp. have evolved a specialized set of CAZymes or effector proteins that make them successful plant pathogens. It has been noted that *M. perniciosa*

has a reduced CAZy family profile compared to other hemibiotrophic pathogens (Zhao et al. 2013), while the full CAZy profile in *M. roreri* remains poorly characterized. Furthermore, some effector-like proteins from *M. perniciosa* have been demonstrated to participate directly in cacao infection (Fiorin et al. 2018), and the effectoromes from *M. perniciosa* and *M. roreri* have been compared (Barbosa et al. 2018). However, a CAZyme and effector protein repertoire screening within a Marasmiineae framework, which could reveal new insights into the emergence of these cacao pathogens, has not been performed. Therefore, the aim of this study is to examine the CAZyme and effector protein profiles from Marasmiineae species genomes to find hints into the emergence of pathogenicity within the suborder.

Table 4.1 Summary of relevant information of *Moniliophthora* species.

Species	**Range**	**Habitat**	**Role**	**Reference**
M. aurantiaca	Samoan Islands	Wood debris in littoral forests	Saprotroph	(Kropp and Albee-Scott 2012)
M. conchata	South Korea, Japan	Dead twigs of *Trachelospermum asiaticum*	Saprotroph	(Takahashi 2002, Antonín et al. 2014)
M. marginata	Malaysia	Montane cloud forest	Saprotroph	(Kerekes and Desjardin 2009)
M. canescens	Malaysia, Japan	Dead fallen twigs of broad-leaved dicots on primary forests	Saprotroph (formation of rhizomorphs)	(Corner 1996, Kerekes and Desjardin 2009)
M. nigrilineata	Singapore	Directly from substrate	Saprotroph	(Kerekes and Desjardin 2009)
M. perniciosa	The Americas	All aerial parts of Malvaceae, Solanaceae and Bignonaceae (Agricultural settings and Amazonian forests)	Pathogen	(Teixeira et al. 2015; Chapter 1)
M. roreri	The Americas except Brazil	Only fruits of *Theobroma and Herrania* species (Agricultural settings and Amazonian forests)	Pathogen	(Bailey et al. 2018; Chapter 1)
Moniliophthora sp.	Southern USA	Semiarid grassland (commonly isolated from *Bouteloua gracilis* and *Sporobolus cryptandrus*)	Root endophyte	(Aime and Phillips-Mora 2005, Khidir et al. 2010)

Table 4.2 Plant pathogenic Agaricales other than *Moniliophthora roreri* and *M. perniciosa*

Species	Family	Suborder	Host	Geographical range	Reference
Coprinopsis psychromorbida	Psathyrellaceae	Agaricineae	Winter cereals and grasses	North America	(Redhead and Traquair 1981, Gaudet and Bhalla 1988)
Armillaria spp.	Physalacriaceae	Marasmiineae	Hardwood trees and shrubs	Worldwide	(Coetzee et al. 2011)
Mycaureola dilseae	Physalacriaceae	Marasmiineae	Marine red alga *Dilsea carnosa*	Northern European coast	(Porter and Farnham 1986, Binder et al. 2006)
Mycena citricolor	Mycenaceae	Marasmiineae	Coffee (*Coffea arabica*)	Tropical America	(Avelino et al. 2007)
Marasmiellus palmivorus	Omphalotaceae	Marasmiineae	Coconut (*Cocos nucifera)* and oil palm (*Elaeis guineensis*)	Southeast Asia	(Pong et al. 2012, Almaliky et al. 2013)
Marasmius puerariae	Marasmiaceae	Marasmiineae	Kudzu (*Pueraria montana*)	Taiwan	(Kirschner et al. 2013)
Marasmius cyphella	Marasmiaceae	Marasmiineae	*Hevea* spp.	Malaya and West Africa	(Dennis and Reid 1957, Antonín 2013)
Marasmius scandens	Marasmiaceae	Marasmiineae	Cacao (*Theobroma cacao*)	West Africa	(Dennis and Reid 1957, Akrofi et al. 2016)
Marasmius pulcher	Marasmiaceae	Marasmiineae	Tea (*Camellia sinensis*)	Sri Lanka and West Africa	(Dennis and Reid 1957, Adedeji 2006)
Marasmius cymatelloides	Marasmiaceae	Marasmiineae	*Baphia sp.*	Sierra Leona	(Dennis and Reid 1957)
Marasmius graminum	Marasmiaceae	Marasmiineae	*Stenotaphrum secundatum* and *Cynodon dactylon*	USA and Australia	(Baird et al. 1992, Vinnere et al. 2005)
Marasmius rotula	Marasmiaceae	Marasmiineae	Bermudagrass (*Cynodon dactylon*)	USA	(Baird et al. 1992)
Crinipellis siparunae	Marasmiaceae	Marasmiineae	Living branches of *Siparuna lindeni*	Brazil, but collected in the St. Petersburg Botanical Garden	(Singer 1942, Pegler 1978)
**Cri. pseudostipitaria*	Marasmiaceae	Marasmiineae	Grasses and cereal plants	Tropical America, Asia and Europe	(Singer 1942)
Cri. stipitaria	Marasmiaceae	Marasmiineae	Grasses and cereal plants	Temperate North America, Europe and Asia	(Singer 1942)
Typhula spp.	Typhulaceae	Pleurotineae	Winter cereals and grasses	North America, Northern Europe, East Asia	(Hsiang et al. 1999, Kirschner et al. 2013)

**Crinipellis pseudostipitaria* var. *mesites* saprotrophic growing on grass debris from Veracruz, Mexico (Bandala et al. 2012).

4.2 Hypothesis

Pathogenic *Moniliophthora* spp. differ from other Marasmiineae species in their effector protein and CAZyme repertoires.

4.3 Materials and Methods

4.3.1 *Moniliophthora roreri* genome and transcriptome

The genome of *M. roreri* MCA2952, harboring invasive mating types *A1B1* and responsible for the first recorded incidence of the fungus in Mexico (Phillips-Mora et al. 2006b, Díaz-Valderrama and Aime 2016a), was sequenced and assembled. Approximately 1.5µg of genomic DNA was extracted from 28 day-old mycelial/spore tissue grown on potato dextrose agar (PDA) using the Promega Wizard® Purification Kit (Promega Corp., Madison, Wisconsin). Complementarily, the transcriptomes from *M. roreri* isolates having both invasive mating types, including MCA2952 (Table 4.3), were sequenced to have transcriptional evidence for genome annotation. Fourteen-day old cultures on V8 media (20% w/ v V8 juice, 0.1% w/v asparagine, 2.0% w/v maltose, 1.8% w/v agar) were grown for 28 days on PDA. Total RNA from each sample was extracted with the E.Z.N.A.® Fungal RNA Kit (Omega Bio-Tek, Norcross, GA, USA) following the manufacturer's instructions. For DNA, one paired-end and mate pair libraries were generated, while for RNA, paired-end RNA poly A+ libraries were constructed (all libraries contained ~100 bp-long reads) using the TrueSeq Stranded mRNA Kit (Illumina, Inc., San Diego, CA). Both DNA and RNA were sequenced on an Illumina HiSeq 2500 (Illumina, Inc., San Diego, CA) platform at the Purdue University Genomics Core Facility.

4.3.1.1 Genome and transcriptome assembly

The first step was the quality control of Illumina reads to be used in the assembly. DNA and mRNA raw reads containing TrueSeq barcodes were trimmed using TRIMMOMATIC 0.32 (Bolger et al. 2014) using default parameters. Trimmed reads were furthered filtered with Bowtie2 2.2.9 using the `–very-sensitive-local` parameter (Langmead and Salzberg 2012) to remove reads of potential contaminants like the bacteriophage PhiX-174

genome used as control in Illumina sequencing (Mukherjee et al. 2015), and the *M. roreri* mitochondrial genome (Costa et al. 2012). Quality check of reads before and after trimming/filtering was performed with FASTQC v.0.11.7 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

After quality control, two independent genome assemblies for MCA2952 were produced (Table 4.4): one with SPAdes 3.10.1 (Bankevich et al. 2012) using the read error correction algorithm (Nikolenko et al. 2013), and the other with Meraculous 2.0.5 (Chapman et al. 2011, 2016). The overall assembly pipeline was as follows: 1) initial assembly; 2) iterative runs of alternative scaffolding and gap closing with SSPACE 3.0 (Boetzer et al. 2011) and GapFiller 1.11 (Boetzer and Pirovano 2012) using both paired-end and mate-pair read information; 3) inspection and breaking of assemblies in possible misassembled positions with REAPR 1.0.16 (Hunt et al. 2013) between scaffolding runs; and 4) a final quality assessment and assembly breaking with QUAST 3.2 (Gurevich et al. 2013). At the end, the improved SPAdes assembly was chosen based on the very few possible misassembled positions after a final QUAST run (Table 4.4). Additionally, RNA filtered reads from isolate MCA2952 were *de novo* assembled using Trinity r20150110beta (Grabherr et al. 2011) with default parameters for paired-end reads.

4.3.2 *M. roreri* MCA2952 genome annotation

The MCA2952 genome was annotated with MAKER 2.31.8-openmpi-1.6.5 (Campbell et al. 2014a), which integrates an initial step of masking repetitive and low-complexity DNA regions with RepeatMasker (Smit et al. 2015), alignments of mRNA evidence to assembly with BLAST (Camacho et al. 2009) and Exonerate (Slater and Birney 2005), and *ab initio* gene prediction with SNAP (Korf 2004), AUGUSTUS (Stanke et al. 2006) and GeneMark (Ter-Hovhannisyan et al. 2008) in its pipeline. The initial SNAP training was performed with the Trinity assembly of MCA2952 reads as mRNA evidence, followed by two runs with the SNAP species parameter/hidden Markov model (HMM) output files from the previous training step. A specific species parameter file was created for *M. roreri* MCA2952 to train AUGUSTUS. Complementarily, GenMark was trained with the MCA2952 genome assembly with the parameters set for eukaryotes and fungal species. A

final MAKER run was performed with the SNAP, AUGUSTUS and GeneMark evidence. Functional InterPro and Pfam domains (Finn et al. 2016, 2017) and Gene Ontology (GO) terms (Ashburner et al. 2000, Gene Ontology Consortium 2017) of predicted proteins were inferred with InterProScan 5.26-65.0 (Jones et al. 2014). MAKER annotation generates three annotations: the *standard*, *default* and *max* annotations (Campbell et al. 2014a). The MAKER *standard* annotation was selected as the final annotation because of its accuracy over the *default* and *max* (Holt and Yandell 2011, Campbell et al. 2014b), which was assessed based on the number of functional domain-containing proteins and the Annotation Edit Distance (AED) values of predicted proteins (Eilbeck et al. 2009, Campbell et al. 2014a).

4.3.3 Effectorome prediction

The genomes of eleven Marasmiineae species were analyzed for comparison (Table 4.5). Two genomes from species belonging to the Agaricineae and Schizophyllineae were incorporated as outgroups (Table 4.5) as per Dentinger et al. (2015). Most of these genomes were generated through the 1000 Fungal Genomes Project by the US Department of Energy Joint Genome Institute (JGI) and are available in the JGI MycoCosm portal (Grigoriev et al. 2014). The filtered protein models of each species were used for analyses. The predicted proteomes of the two invasive *M. roreri* genotypes, MCA2952 (mating type *A1B1*; *standard* MAKER annotation generated in this study) and MCA2997 (mating type *A2B2*; Meinhardt et al. 2014) and *M. perniciosa* were also used (Mondego et al. 2008). InterProScan 5.26-65.0 analysis was performed simultaneously in all genomes to avoid potential version-biases in previous annotations of those genomes.

Table 4.3 Isolates of *Moniliophthora roreri* used for transcriptome sequencing

Mating type	Isolate ID	CBS accession number	Country	Host	Collection Year
A1B1	MCA2504	CBS 138626	San Carlos, Costa Rica	*T. cacao*	1999
	MCA2952	CBS 138632	Pichucalco, Mexico	*T. cacao*	2005
	MCA2974	—	Sucumbíos, Ecuador	*T. cacao*	2005
A2B2	JD 5	CBS 138634	Huánuco, Peru	*T. cacao*	2012
	JD 6	—	Huánuco, Peru	*T. cacao*	2012
	JD 8	CBS 138635	Cusco, Peru	*T. cacao*	2012

Table 4.4 Assembly of *Monliophthora roreri* MCA2952 and statistics of assemblies

Assembler	Parameters initial run	Unimproved assembly statistics	Improvement	
			Steps and Software	**Stats of improved assembly**
*SPADes	Read error correction kmers: 21, 33 and 55	Total number of scaffolds = 26511 Sum (bp) = 57416220 Total number of N's = 561913 Sum (bp) no N's = 56854307 GC Content = 46.40% Max scaffold size = 2742930 Min scaffold size = 56 Average scaffold size = 2165 N25 = 535394 N50 = 236985	1. 12 iterations of gap filling with GapFiller 2. SSPACE + 20 iterations GapFiller 3. SSPACE + 18 iterations GapFiller 4. Inspection and breaking with REAPR 5. SSPACE + 20 iterations GapFiller 6. SSPACE + 6 iterations GapFiller 7. 15 iterations SSPACE + 15 iterations GapFiller 8. Quast inspection and breaking	Total number of scaffolds = 257 Sum (bp) = 59685079 Total number of N's = 2831 Sum (bp) no N's = 59682248 GC Content = 46.16% Max scaffold size = 3205325 Min scaffold size = 502 Average scaffold size = 232238 N50 = 667426 N75 = 372933
Meracolous	default parameters	Total number of contigs = 1387 Sum (bp) = 56422423 Total number of N's = 322517 Sum (bp) no N's = 56099906 GC Content = 46.50% Max contig size = 1115457 Min contig size = 195 Average contig size = 40679 N25 = 302121 N50 = 142304	1. 16 iterations of scaffolding with SSPACE 2. 11 iterations of gap filling with GapFiller 3. Quast inspection and breaking	Total number of scaffolds = 434 Sum (bp) = 56756038 Total number of N's = 1648 Sum (bp) no N's = 56754390 GC Content = 46.46% Max scaffold size = 2031223 Min scaffold size = 530 Average scaffold size = 130774 N50 = 325073 N75 = 152194

*Assembly selected for annotation and further analyses.

Table 4.5 Agaricales genomes used for the effectorome and CAZyome analyses

Genome	Strain	Abbreviation	Family	Suborder	Accession / ID	Source
Coprinopsis cinerea	Okayama 7 #130	Copci	Psathyrellaceae	Agaricineae	AACS00000000[a]	(Stajich et al. 2010)
Guyanagaster necrorhizus	MCA3950	Guyne	Physalacriaceae	Marasmiineae	1019625[b]	Permission
Armillaria mellea	DSM 3731	Armme	Physalacriaceae	Marasmiineae	ERP000894[c]	(Collins et al. 2013)
Cylindrobasidium torrendii	HHB-15055	Cylto	Physalacriaceae	Marasmiineae	1016295[b]	(Floudas et al. 2015)
Gymnopus luxurians	FD-317	Gymlu	Omphalotaceae	Marasmiineae	403665[b]	(Kohler et al. 2015)
Dendrodontia bispora	CBS 962.96	Denbi	Marasmiaceae	Marasmiineae	1016303[b]	(Varga et al. 2019)
Omphalotus olearius	VT-653.13	Ompol	Omphalotaceae	Marasmiineae	AHIW00000000[a]	(Wawrzyn et al. 2013)
Marasmius fiardii	PR-910	Marfi	Marasmiaceae	Marasmiineae	1016747[b]	Permission
Moniliophthora perniciosa	FA553	MpFA553	Marasmiaceae	Marasmiineae	ABRE01000000[a]	(Mondego et al. 2008)
Moniliophthora roreri	MCA2997 (mating type *A2B2*)	MCA2997	Marasmiaceae	Marasmiineae	AWSO01000000[a]	(Meinhardt et al. 2014)
Moniliophthora roreri	MCA2952 (mating type *A1B1*)	MCA2952	Marasmiaceae	Marasmiineae	Improved version of LATX00000000[a]	This study
Mycena alexandri	CBHHK200	Mycale	Mycenaceae	Marasmiineae	1146206[b]	Permission
Fistulina hepatica	ATCC 64428	Fishe	Schizophyllaceae	Schizophyllineae	405398[b]	(Floudas et al. 2015)

[a] GenBank accession number; [b] JGI Project ID; [c] European Nucleotide Archive accession number (http://www.ebi.ac.uk/ena/

Signal peptides in all predicted proteins were identified with SignalP 4.1c (Petersen et al. 2011) and Phobius 1.01 (Käll et al. 2004, 2005) as they are the best signal peptide predictors for eukaryotes (Petersen et al. 2011). Phobius also predicts the number of transmembrane domains (TM) in proteins; only proteins that had zero or one TM were considered. Signal peptide-containing proteins were considered to conform the secretome of each species if they were detected in both programs runs. Then, the effectorome was predicted with EffectorP 1.0 and 2.0 (Sperschneider et al. 2016, 2018). EffectorP identifies effectors based on protein size, molecular weight, charge and cysteine, serine and tryptophan content (Sperschneider et al. 2016); to accomplish this it integrates the peptide statistics utility from EMBOSS (Rice et al. 2000) and the machine learning algorithms from WEKA (Hall et al. 2009). The combination of the results from both EffectorP versions generates an effectorome with low false positive rate (Sperschneider et al. 2018). As additional evidence, tandem amino acid repeats and NLS in proteins were searched with programs T-REKS (Jorda and Kajava 2009) and NLStradamus 1.8 (Ba et al. 2009), respectively. The putative effectors containing both repeats and NLS were individually inspected, and *blastp* searches against the non-redundant (nr) NCBI database excluding searches in the same genus were performed; EffectorP 1.0 and 2.0 runs over the top matches were further performed.

4.3.4 CAZYome

The CAZyomes from the evaluated species (Table 4.5) were predicted from the same proteomes used to identify effector proteins. All 585 HMM profiles of CAZy families available in the database for automated CAZ annotation, dbCAN release 6.0 (Yin et al. 2012) were retrieved. Then, HMM searches were performed with program HMMER 3.1b2 (Eddy 2011, Mistry et al. 2013). The output was parsed with the custom *hmmscan-parser* script available in the dbCAN database. E-values and coverage thresholds were 1E-17 and 0.45 as suggested for fungi (Yin et al. 2012).

4.3.5 Evolution of effector and CAZyme families

The GO term counts of predicted effector proteins belonging to any of the three GO categories: cellular component, molecular function and biological process (Ashburner et

al. 2000, Gene Ontology Consortium 2017) were performed with WEGO 2.0 (Ye et al. 2018). The WEGO output at the level "6" of annotation served as input for CAFE 4.2 (De Bie et al. 2006, Han et al. 2013), program used to evaluate the significance of expansion and contraction of effector GO terms during evolution. The second input for CAFE was the phylogenetic tree of species under evaluation in *newick* format (Olsen 1990). This was generated with a dataset, concatenated with SeaView 4.7 (Gouy et al. 2010), of ortholog proteins of 1590 and 5645 (JGI protein IDs of *Corpinopsis cinerea*), two of the twenty-seven most phylogenetic informative proteins in the Agaricales (Dentinger et al. 2015). Phylogeny was constructed following Chapter 3's guidelines with default parameters for proteins; the resulting phylogeny had the same topology depicted in Dentinger et al. (2015). CAFE requires an ultrametric phylogeny with branch lengths in integers. Then, the branch lengths of the resulting phylogeny were proportionally transformed with FigTree 1.4.3 (Rambaut 2006) and the phylogeny itself checked to be ultrametric with R package *phytools* (Revell 2012). As another method to transform the phylogeny, software r8s 1.81 (Sanderson 2003) was used with a calibration point of 90 million years for the emergence of Marasmiineae based on the fossil record (Hibbett et al. 1997). The CAFE threshold of significant rapid evolution of GO terms and CAZy families was $p = 0.01$. CAFE analysis with the FigTree-transformed phylogeny was more conservative, thus these results are presented and discussed; CAFE results with r8s-transformed phylogeny are still provided (APPENDIX B).

4.3.6 Synteny analysis of rapid-evolving genes

Synteny analysis of genes having rapidly evolving GOs across the evaluated taxa (Table 4.5) was performed with a custom program SyLOCAL specifically written for this study (APPENDIX F). The core of the program is in Perl and the extension to run it simultaneously on multiple genomes, in Python. Overall, SyLOCAL takes as input a *fasta* file containing the coding sequences (CDS) of the genomic region containing the comparing cluster of genes (query). Additionally, it takes the filtered gene catalogue in *GFF* format and the CDS *fasta* file of each genome to compare against (subject), downloaded from Mycocosm. The script performs local *tblastx* searches to identify the ortholog genes in the subject. It then collects the scaffold, start and end positions and strand

direction from the *GFF* file of each orthologue gene. Finally, it identifies the contig/scaffold/chromosome that contains more matches as it is the most syntenic one, for plotting. Synteny plots are generated with R package *genoPlotR* (Guy et al. 2010), invoked directly from the script with the module Statistics::R. The output synteny plot is in scalable vector graphics (*svg*) format (Quint 2003). Intermediate files are generated so the user can check the analysis performed. Also, package *ade4* is required for the script to work (Dray and Dufour 2007). SyLOCAL runs on macOS but with the Cygwin toolkit (Racine 2000) it runs on the Windows operating system.

4.3.7 Transcriptomic comparison of *M. roreri* mating types

A transcriptome comparison was performed using genomes from both *M. roreri* mating types, i.e., MCA2952 and MCA2997, and mRNA sequencing generated here (Table 4.3). First, filtered mRNA reads were aligned with Tophat 2.1.1 (Trapnell et al. 2009) using default alignment parameters. These alignments were used to feed `cufflinks` and `cuffmerge`, both tools of the program Cufflinks 2.2.1-t1 (Trapnell et al. 2010), to generate a transcriptome for each sample and one master transcriptome for all the samples, respectively. For `cufflinks` options `–compatible-hits-norm` and `–multi-read-correct were set`, and for `cuffmerge` the genome annotations and their *GFF* files were used as reference. Then, using HTSeq 0.7.0 (Anders et al. 2015) the number of reads that align to each gene in the annotation were counted using the master transcriptome from `cuffmerge` and the output alignments from Tophat. Gene expression in both invasive strains was compared with R package *DESeq2* (Love et al. 2014). Since both mating type genomes are highly divergent in the *A* and *B* mating loci (Díaz-Valderrama and Aime 2016a), HTSeq counts of reads from ortholog genes that aligned to the genome with their respective mating type were used for analysis. Ortholog protein coding genes between both genomes were identified via bi-directional `blastp` searches with BLAST 2.7.1+ with an E-value cutoff of E-10. Only genes that reciprocally appeared as top hits were considered orthologs and were used for *DESeq2* analysis. For this, read counts were normalized and outlier detection by Cook's distance (Cook and Dennis 1977) was performed prior to analysis.

4.4 Results

4.4.1 Genome sequencing and annotation

The SPADes assembly of *M. roreri* MCA2952 had better genomic characteristics than the Meracolous assembly and QUAST breaking analysis results into an assembly with very low possible misassembled scaffolds (Table 4.4); thus, the SPADes assembly, from now on referred to as "the MCA2952 genome", was selected for annotation, and effectorome and CAZyome analyses (APPENDIX B). Compared to the publicly available MCA2997 genome, the MCA2952 genome produced 257 scaffolds spanning 59.69 Mb (Table 4.6). Additionally, the MCA2952 genome is the most complete *Moniliophthora* genome generated so far (Table 4.6). The genome annotation of isolate MCA2952 generated 17,150 predicted proteins, out of which 15,494 (90.34 %) have an AED value less than 0.5. Also, 8,673 (50.57 %) proteins have at least one recognizable IPR/Pfam functional domain or GO term.

4.4.2 Effectorome prediction

Overall, *Mycena alexandri* (Mycenaceae) and Marasmiaceae species have more predicted effector proteins than members of the Physalacriaceae; within the Omphalotaceae, *Gymnopus luxurians* has a similar number of effectors as Marasmiaceae species, while *Omphalotus olearius* has the lowest number of predicted effectors (Figure 4.1). Twenty-three effectors with both amino acid repeats and NLS were identified among the thirteen Agaricales genomes examined. *Moniliophthora roreri* MCA2997 and MCA2952 have the highest numbers of effectors with those characteristics with four and six, respectively (Figure 4.1). The pathogen *Armillaria mellea* had three effectors while the cacao pathogen *M. perniciosa* had only one (Figure 4.1).

Most effectors with repeats and NLS did not have a recognizable protein domain and/or an NCBI match (APPENDIX G). Two ortholog effectors from both *M. roreri* genomes had a Ricin B-lectin domain (IPR000772), a third putative effector from MCA2997 genome belongs to the GH superfamily (IPR017853), and the predicted effector from *Mar. oreades* belongs to the Protein PriA (IPR038955) (APPENDIX G). Ricin B-lectin effectors from *M. roreri* do not have fungal matches in NCBI and only one has a high E-value match with

a putative acetyl-hydrolase from *Clostridium* spp. Predicted effectors from *A. mellea* do not have recognizable protein domains but one has as its only NCBI match an effector from the bacteria *Terribacillus halophilus*. The putative effector from *M. perniciosa* does not have a recognizable protein domain nor did it have any significant NCBI matches (APPENDIX G).

4.4.3 CAFE analysis on effectors

The GO analysis on the effectoromes from the thirteen Agaricales genomes revealed that only 17% of effectors (345/2018) have at least one known GO term (Table 4.7). CAFE analysis found that three GO terms underwent significant expansion or contraction in plant pathogenic species during evolution of Marasmiineae (Figure 4.2). Effectors with fungal-type cell wall ontology (GO:0009277) have significantly expanded in *M. roreri* during its divergence from *M. perniciosa*; other lineages that underwent a significant expansion of effectors with this ontology are *Cop. cinerea*, *G. necrorhizus* and *D. bispora*, while Omphalotaceae species underwent a significant contraction instead, i.e. they have less effectors with fungal-type cell wall ontology (Figure 4.2). Individual inspection in all species shows that these effectors are all hydrophobins (Table 4.8 and APPENDIX B). Most of hydrophobins in both genomes of *M. roreri* are clustered together in the same genomic region and are syntenic, which does not happen when it is compared to the other Agaricales genomes, even within Marasmiaceae (Figure 4.3).

Table 4.6 Summary statistics of the *Moniliophthora roreri* genome generated in this study and other *Moniliophthora* genomes produced in previous research

Species	Strain code	Assembled genome size (Mb)	N50 (kb)	Longest scaffold (kb)	Number of scaffolds/ contigs	GC %	Proteome predicted	Reference
M. roreri	MCA2952	59.7	667	3,205	257	46.2	17054	This study
M. roreri	MCA2997	52.2	48	571	3280	46.9	17910	(Meinhardt et al. 2014)
M. roreri	MrPeru	45.2	56	530	2994	47.8	14,154	(Barbosa et al. 2018)
*M. perniciosa**	MpFA553	18.4	1.5	24	14868	48.0	13,560	(Mondego et al. 2008)
M. perniciosa	Mp4145	47.0	87	910	2676	47.7	14,210	(Barbosa et al. 2018)
M. perniciosa	Mp1441	46.3	90	910	2100	47.7	13,404	
M. perniciosa	Mp4124	45.5	90	910	1967	47.8	12,188	
M. perniciosa	Mp178	43.9	92	910	1526	48.0	11,203	
M. perniciosa	Mp4071	44.4	92	910	1689	47.9	11,474	

*The statistics presented on this genome are based on data currently available in JGI and it differs from what is reported in Mondego et al. (2008) mainly on the assembly size (18.4 Mb vs 26.7 Mb), the number of predicted proteins is almost the same (13,560 vs 13,640)

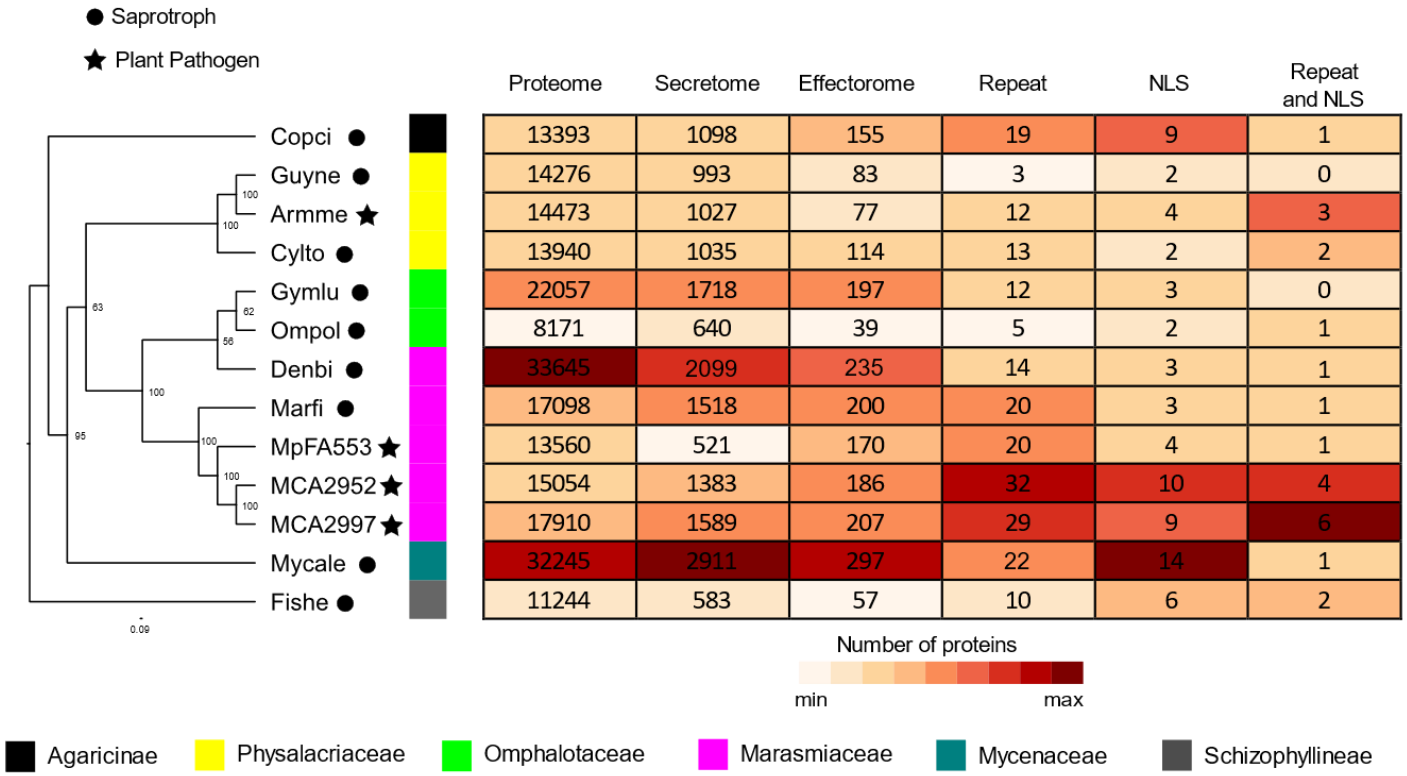


Figure 4.1 Summary of the number of proteins conforming the proteome, secretome and effectorome of genomes evaluated. Repeat = number of effector proteins with amino acid repeats; NLS = Number of effectors with a nuclear localization signal; Repeat and NLS = Number of effectors with both amino acid repeats and NLS. Heat maps are scaled for each column; min = lowest value and max = highest value in each column. Species are phylogenetically organized based on the concatenated dataset of two phylogenetically informative proteins (see Materials and Methods); node labels are bootstrap values obtained in the maximum likelihood analysis before branch length transformation with FigTree. The saprotrophic and pathogenic nature of species are also depicted. Genome code meanings in (Table 4.5)

Table 4.7 Summary of gene ontology (GO) analysis of the thirteen Agaricales effectoromes

Species	Genome abbreviation	Total predicted effectors	Effectors with GO terms	GO term category		
				Biological Process	Cellular Component	Molecular Function
Coprinopsis cinerea	Copci	155	40	13	25	39
Guyanagaster necrorhizus	Guyne	83	12	1	11	11
Armillaria mellea	Armme	77	22	11	9	21
Cylindrobasidium torrendii	Cylto	114	15	3	7	12
Gymnopus luxurians	Gymlu	197	23	9	8	20
Dendrodontia bispora	Ompol	39	8	1	5	6
Omphalotus olearius	Denbi	235	37	9	18	35
Marasmius fiardii	Marfi	200	24	11	11	21
Moniliophthora perniciosa	MpFA553	170	59	36	12	52
Moniliophthora roreri	MCA2952	186	37	12	21	33
Moniliophthora roreri	MCA2997	207	43	12	24	37
Mycena alexandri	Mycale	297	21	10	5	18
Fistulina hepatica	Fishe	57	4	2	2	3
All		2,018	345	130	158	308

Moniliophthora perniciosa is the only taxon in our sampling to evolve nine effector proteins with a flavin adenine dinucleotide (FAD) binding ontology (GO:0050660) compared to two or fewer effectors in the other genomes analyzed (Figure 4.2). Also, there is a significant contraction of effectors with this GO term in *M. roreri* MCA2997 compared to MCA2952 (Figure 4.2). Further inspection reveals that these effectors belong mainly to the glucose-methanol-choline (GMC) oxidoreductase and FAD binding domain families, but only *M. perniciosa* and *M. roreri* MCA2952 contain GO:0050660 effectors in both protein families (

Table 4.9 and APPENDIX B). *Armillaria mellea* has undergone a significant expansion of effectors with a transition metal ion binding ontology (GO:0046914); in fact, none of the other Physalacriaceae species have effectors with this GO term (Figure 4.2). Also, Marasmiaceae species have more of these effectors than Omphalotaceae but this expansion is not significant (Figure 4.2). Inspection of these effectors reveals that they belong to the cytochrome P450 (CYP) and multicopper oxidase super families (Table 4.10).

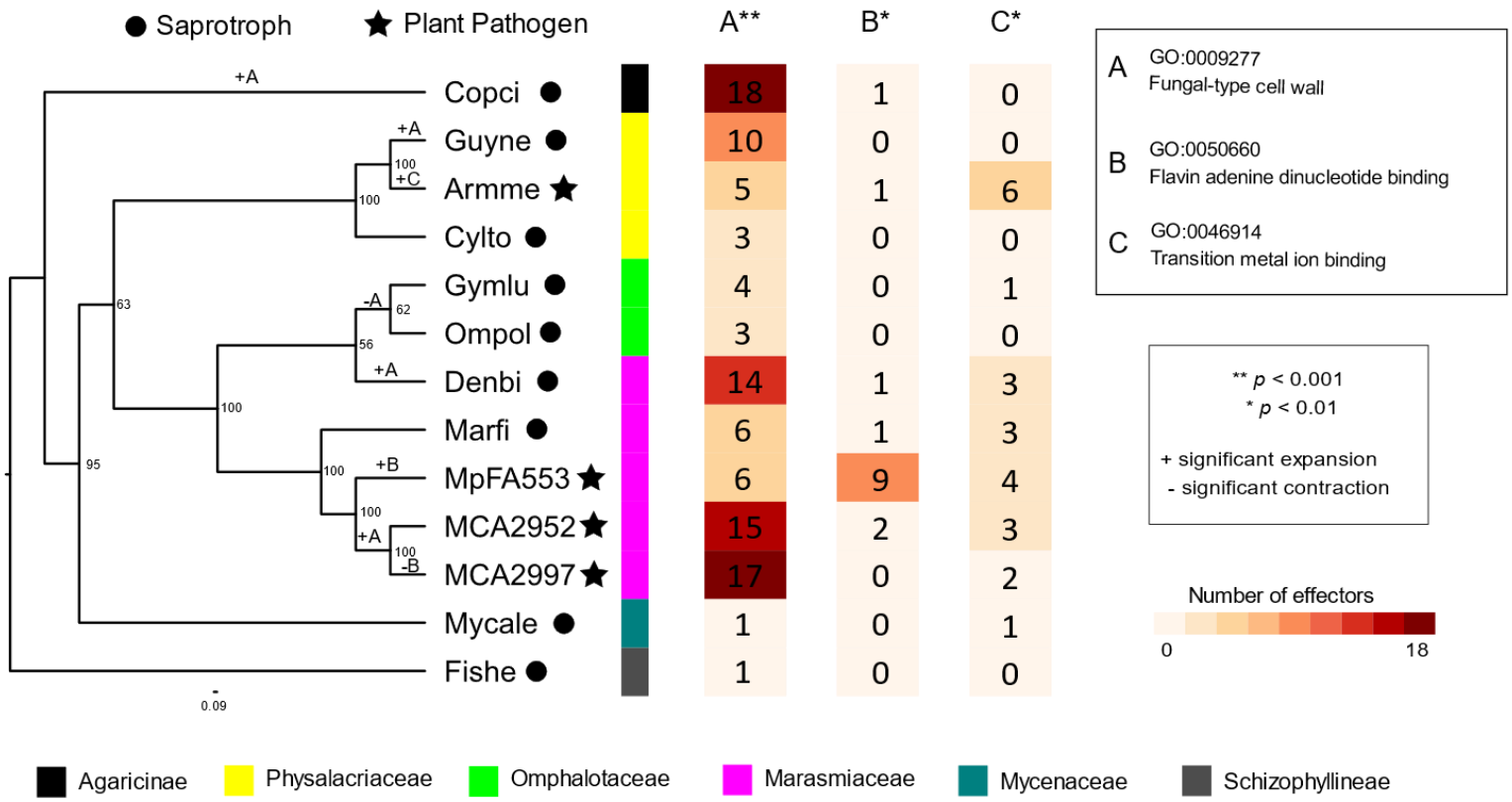
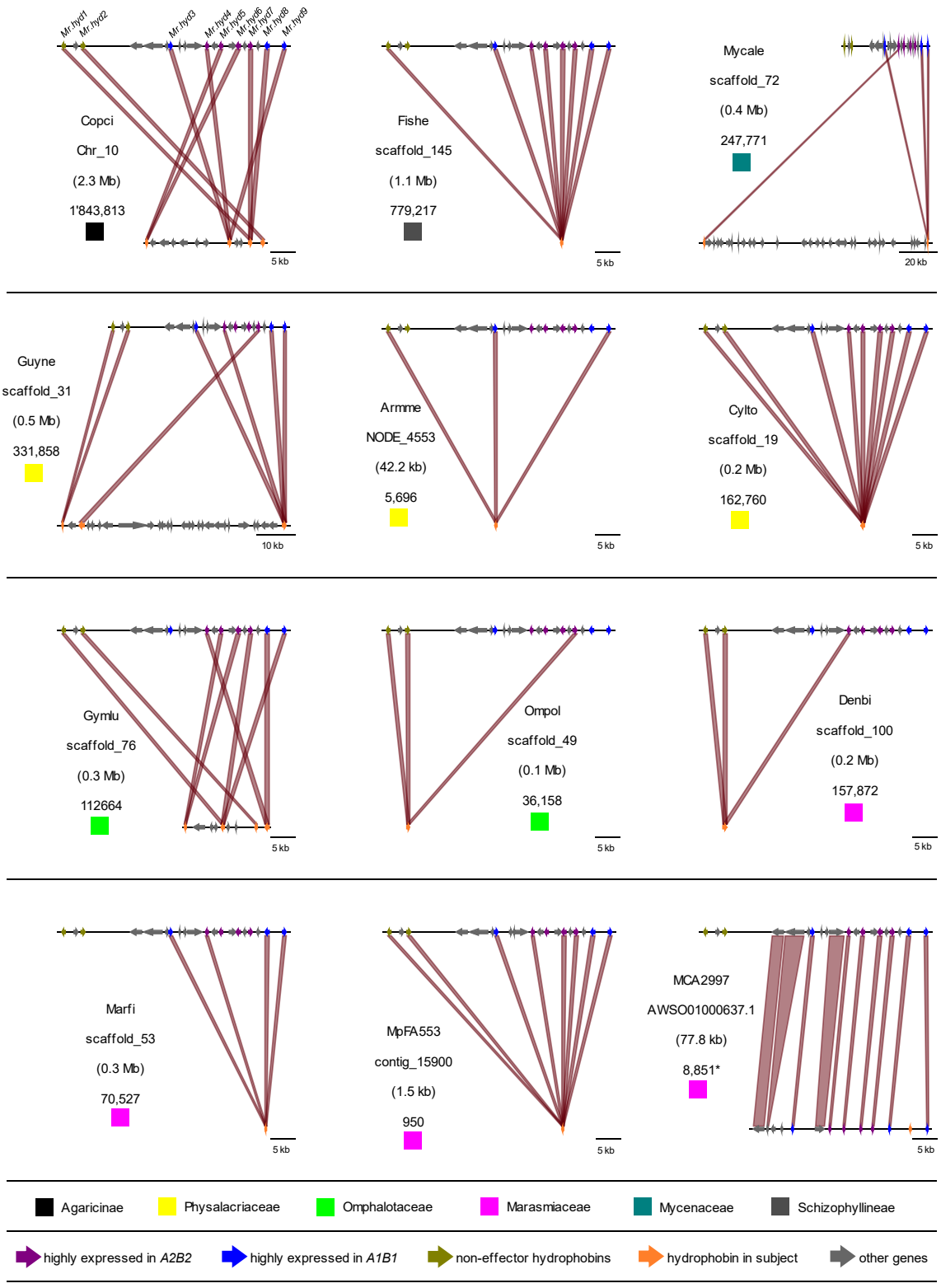


Figure 4.2 Rapidly-evolving gene ontology (GO) terms from effector proteins of evaluated genomes as detected by the family-wide probability values (*p*) from CAFE. Species are phylogenetically organized based on the concatenated dataset of two phylogenetically informative proteins (see Materials and Methods); node labels are bootstrap values obtained in the maximum likelihood analysis before branch length transformation with FigTree; plus (+) and minus (-) symbols on branches indicate the significant expansion or contraction, respectively, of a GO term (A, B or C) on that branch (See also APPENDIX B). The saprotrophic and pathogenic nature of species are also depicted. Genome code meanings in (Table 4.5).

Table 4.8 Annotation of predicted effectors from *Moniliophthora roreri* with fungal-type cell wall ontology that are significantly expanded in this species

Genome	#	Effector (ID / Accession Number)	Given name	InterProScan Annotation
M. roreri MCA2952	1	maker-dna.fa_109-augustus-gene-0.126-mRNA-1_1	*Mr.hyd3.1*	• PF01185 Fungal hydrophobin • IPR001338 Hydrophobin • GO:0005199|GO:0009277 Stuctural constitutent of cell wall | fungal-type cell wall
	2	maker-dna.fa_109-augustus-gene-0.143-mRNA-1_1	*Mr.hyd4.1*	
	3	maker-dna.fa_109-augustus-gene-0.145-mRNA-1_1	*Mr.hyd6.1*	
	4	maker-dna.fa_109-augustus-gene-0.146-mRNA-1_1	*Mr.hyd8.1*	
	5	maker-dna.fa_109-augustus-gene-0.144-mRNA-1_1	*Mr.hyd5.1*	
	6	augustus_masked-dna.fa_38-processed-gene-5.257-mRNA-1_1		
	7	maker-dna.fa_109-augustus-gene-0.129-mRNA-1_1	*Mr.hyd7.1*	
	8	augustus_masked-dna.fa_109-processed-gene-0.6-mRNA-1_1	*Mr.hyd9.1*	
	9	maker-dna.fa_121-augustus-gene-0.70-mRNA-1_1		
	10	maker-dna.fa_121-augustus-gene-0.71-mRNA-1_1		
	11	maker-dna.fa_160-augustus-gene-0.16-mRNA-1_1		
	12	maker-dna.fa_160-augustus-gene-0.17-mRNA-1_1		
	13	maker-dna.fa_38-augustus-gene-5.167-mRNA-1_1		
	14	augustus_masked-dna.fa_77-processed-gene-1.185-mRNA-1_1		
	15	maker-dna.fa_44-augustus-gene-5.97-mRNA-1_1		
M. roreri MCA2997	1	ESK81797.1		• PF01185 Fungal hydrophobin • IPR001338 Hydrophobin • GO:0005199|GO:0009277 • Stuctural constitutent of cell wall | fungal-type cell wall
	2	ESK83024.1		
	3	ESK84729.1		
	4	ESK84730.1		
	5	ESK87626.1		
	6	ESK87979.1		
	7	ESK88576.1	*Mr.hyd9.2*	
	8	ESK88577.1	*Mr.hyd10.2*	
	9	ESK88578.1	*Mr.hyd8.2*	
	10	ESK88579.1	*Mr.hyd7.2*	
	11	ESK88580.1	*Mr.hyd6.2*	
	12	ESK88581.1	*Mr.hyd5.2*	
	13	ESK88582.1	*Mr.hyd4.2*	
	14	ESK88584.1	*Mr.hyd3.2*	
	15	ESK89592.1		
	16	ESK90060.1		
	17	ESK95426.1		

Figure 4.3 Synteny analysis of the hydrophobin effector cassette of *Moniliophthora roreri*. The genomic region of scaffold dna.fa_109 of *M. roreri* MCA2952 (query) containing nine hydrophobins (*Mr.hyd1-9*), out of which seven were predicted as effectors (Table 4.8), was compared to all the other Agaricales genomes (subjects) with SyLOCAL. Ortholog genes from query and subject are interconnected with red shades. In each plot it is indicated: the subject genome, the contig/scaffold/chromosome with the highest number of *tblastx* matches and its length in parentheses, and the position at which the graphic starts in that sequence. Since the genome of *M. roreri* MCA2997 is not on JGI and the annotation feature file available is in *GFF3* format, not in *GFF* as the others, the synteny plot with MCA2997 was done manually following the SyLOCAL overall pipeline. All plots show sequences from the 5' to 3' direction, except for the plot of MCA2997 which is from the 3' to 5' for graphical purposes (*). Output files of SyLOCAL runs can be found in APPENDIX B. Blue and purple arrows indicate whether they are highly expressed in the *A1B1* or *A2B2* mating type group, respectively. The family or suborder to which they belong is also indicated.


Mr.hyd1
Mr.hyd2
Mr.hyd3
Mr.hyd4
Mr.hyd5
Mr.hyd6
Mr.hyd7
Mr.hyd8
Mr.hyd9
Copci
Chr_10
(2.3 Mb)
1'843,813
5 kb
Fishe
scaffold_145
(1.1 Mb)
779,217
5 kb
Mycale
scaffold_72
(0.4 Mb)
247,771
20 kb
Guyne
scaffold_31
(0.5 Mb)
331,858
10 kb
Armme
NODE_4553
(42.2 kb)
5,696
5 kb
Cylto
scaffold_19
(0.2 Mb)
162,760
5 kb
Gymlu
scaffold_76
(0.3 Mb)
112664
5 kb
Ompol
scaffold_49
(0.1 Mb)
36,158
5 kb
Denbi
scaffold_100
(0.2 Mb)
157,872
5 kb
Marfi
scaffold_53
(0.3 Mb)
70,527
5 kb
MpFA553
contig_15900
(1.5 kb)
950
5 kb
MCA2997
AWSO01000637.1
(77.8 kb)
8,851*
5 kb
Agaricinae
Physalacriaceae
Omphalotaceae
Marasmiaceae
Mycenaceae
Schizophyllineae
highly expressed in A2B2
highly expressed in A1B1
non-effector hydrophobins
hydrophobin in subject
other genes

4.4.4 CAFE analysis on CAZymes

Out of the 585 CAZyme HMMs used, 180 CAZy families were identified in the genomes of thirteen Agaricales species evaluated (APPENDIX B). From this, nineteen families underwent significant expansions and/or contractions throughout their evolution (Figure 4.4). Overall, Marasmiineae pathogens have a similar CAZyme family profile to most of their saprotrophic counterparts. However, *M. perniciosa* displays significant contractions in fourteen out of the nineteen rapidly evolving CAZyme families compared to other Marasmiaceae species (Figure 4.4, APPENDIX B). *Moniliophthora perniciosa* did not experience evolutionary contractions in the CBM13, CBM 67, CE5, GH128 and PL3_2 families (Figure 4.4). Other extreme contractions are the ones undergone by *O. olearius* and *Cylindrobasidium torrendii* in the Omphalotaceae and the Physalacriaceae, respectively (Figure 4.4). *Moniliophthora roreri*, like most Marasmiineae species, kept its repertoire of AA3_2 enzymes during evolution (Figure 4.4).

4.4.5 Transcriptomic comparison in *M. roreri*

The transcriptomic comparison of *M. roreri* strains of both invasive mating types was performed based on a standardized set of ortholog genes obtained via a bidirectional *blastp*. This gene set consisted of 11,183 genes that appeared reciprocally as top hits in both *M. roreri* genomes (APPENDIX B). The expression analysis shows several genes differentially expressed genes depending on the mating type group, *A1B1* or *A2B2* (Figure 4.5). Among the top expressed genes in all isolates regardless of their mating type included hydrophobins (Table 4.11) that were already identified as highly evolving effectors by CAFE (Table 4.8) and that were found to be syntenic only within *M. roreri* (Figure 4.3). Further inspection of these highly expressed hydrophobins showed that isolates having mating type *A1B1* have a significantly higher expression of hydrophobins *Mr.hyd3*, *Mr.hyd8* and *Mr.hyd9*, while isolates with mating type *A2B2* have a significantly higher expression of hydrophobins *Mr.hyd4*, *Mr.hyd5*, *Mr.hyd6* and *Mr.hyd7* (Table 4.11 and Figure 4.3).

Table 4.9 Annotation of putative effectors from *Moniliophthora perniciosa* MpFA553 with flavin adenine dinucleotide (FAD) binding ontology that were significantly expanded in this species

#	Protein ID	InterProScan Annotation		
1	MPER_05504	• PF00732 GMC oxidoreductase	• IPR000172 Glucose-methanol-choline oxidoreductase, N-terminal	• GO:0016614 | GO:0050660 | GO:0055114 Oxidoreductase activity, acting on CH-OH group of donors | Flavin adenine dinucleotide binding | Oxidation-reductase process
2	MPER_09430			
3	MPER_00552			
4	MPER_00581			
5	MPER_11948	• PF01565 FAD binding domain	• IPR006094 FAD linked oxidase, N-terminal	
6	MPER_00602			
7	MPER_01152			
8	MPER_04394			
9	MPER_03223	• PF01266 FAD dependent oxidoreductase	• IPR006076 FAD dependent oxidoreductase	• GO:0016491 | GO:0055114 Oxidoreductase activity | Oxidation-reductase process

Table 4.10 Annotation of putative effectors from *Armillaria mellea* with transition metal ion binding ontology that were significantly expanded in this species

#	JGI Protein ID	InterProScan Annotation		
1	6722	• PF00067 Cytochrome P450	• IPR001128 Cytochrome P450	• GO:0005506 | GO:0016705 | GO:0020037 | GO:0055114 Iron ion binding | Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen | Oxidation reduction process
2	4524			
3	7431	• PF07732 Multicopper oxidase	• IPR011707 Multicopper oxidase, type 3	• GO:0005507 Copper ion binding
4	8465			
5	7476	• No Pfam domains detected	• IPR015889 Intradiol ring-cleavage dioxygenase, core	• GO:0005506 | GO:0016702 | GO:0055114 Iron ion binding | Oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen | Oxidation reduction process
6	7477			

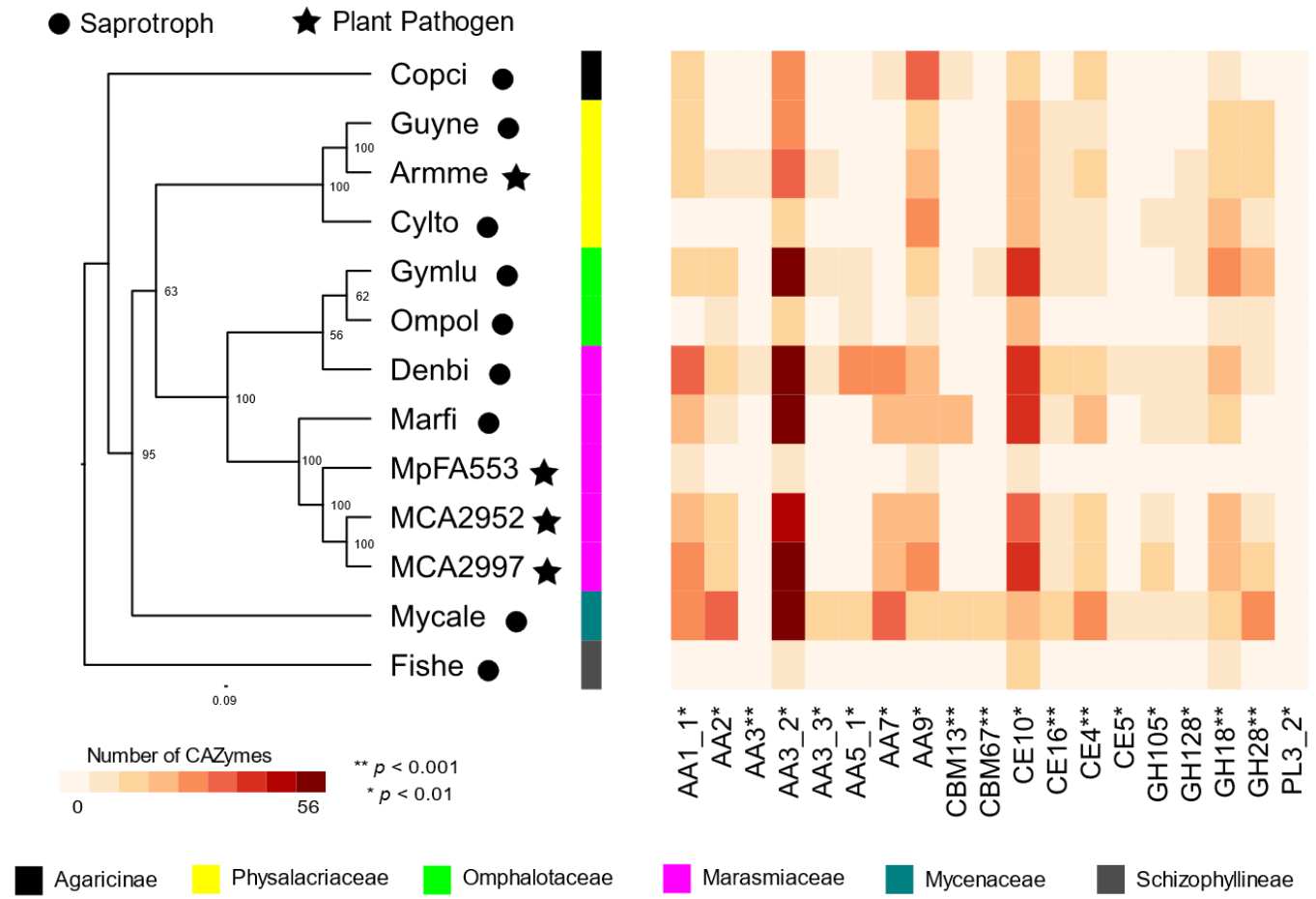


Figure 4.4 Rapidly-evolving CAZy families of evaluated genomes as detected by the family-wide probability values (*p*) from CAFE. Species are phylogenetically organized based on the concatenated dataset of two phylogenetically informative proteins (see Materials and Methods); node labels are bootstrap values obtained in the maximum likelihood analysis before branch length transformation with FigTree; see APPENDIX B for information on non-significant families. The saprotrophic and pathogenic nature of species are also depicted. Genome code meanings in (Table 4.5)

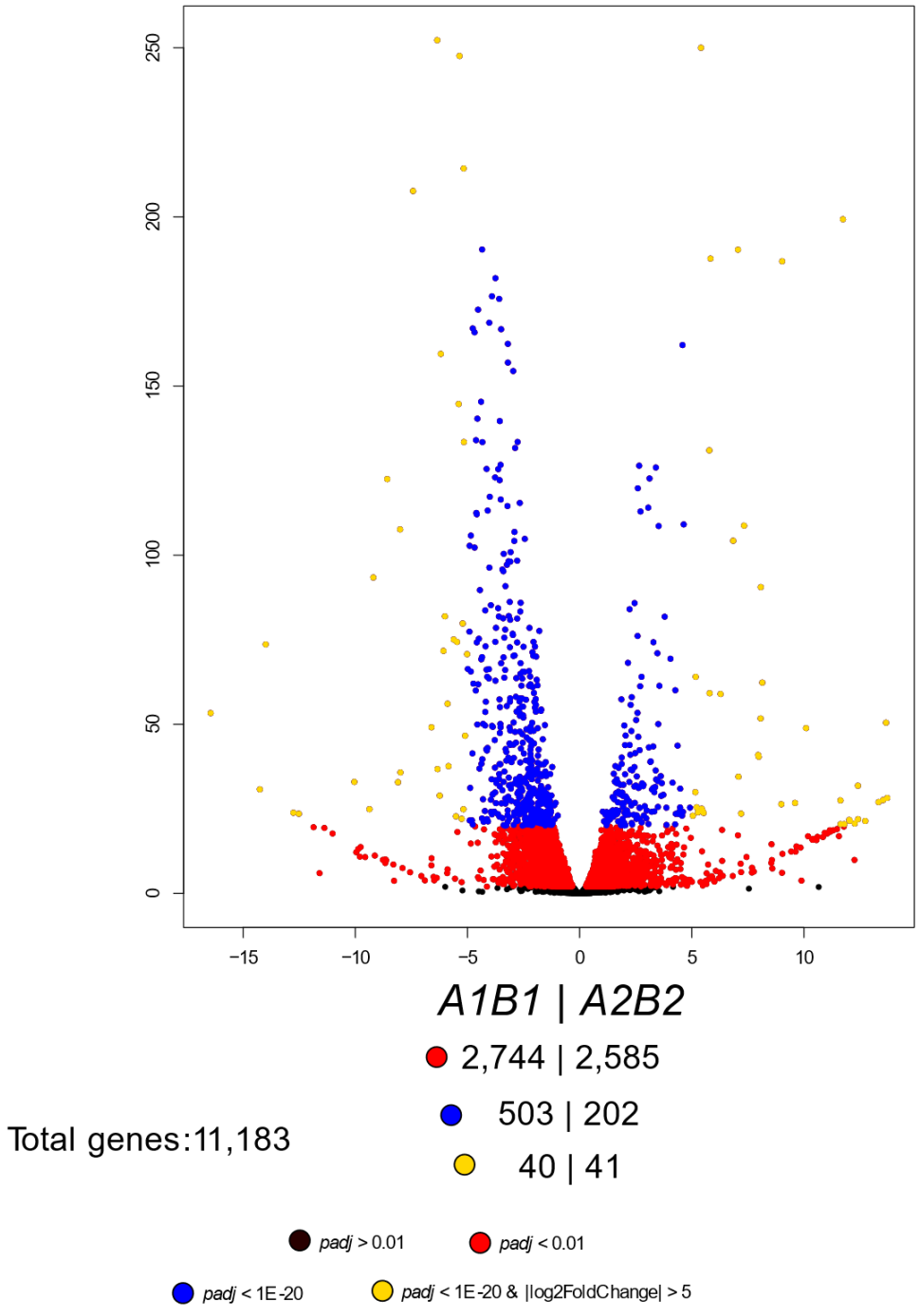


Figure 4.5 Volcano plot of the differential expression analyses on six isolates of *Moniliophthora roreri* (three each of two invasive mating types, *A1B1* and *A2B2*; Table 4.3). Each dot represents a gene expressed in each mating type group. The X-axis displays the $\log_2$ of the Fold Change of expression in terms of number of reads while the Y-axis displays the $-\log_{10}$ of the adjusted probability value (*padj*), as calculated by *DESeq2*. The numbers in color legends are the number of genes expressed in each mating type group. For full information see APPENDIX B.

Table 4.11 Top 25 genes with high expression on cultures of both, *A1B1* and *A2B2*, mating type groups, shown in decreasing order of expression

[1]Gene name MCA2952	[2]Scaffold MCA2952	[3]Gene name MCA2997	[4]Scaffold MCA2997	[5]Annotation	[6]Given name	[7]baseMean	[8]*log2 FC*	[9]*padj*	[10]Mat. type
augustus_masked-dna.fa_109-processed-gene-0.6-mRNA-1	dna.fa_109	ESK88576	AWSO01000637	hydrophobin 2	*Mr.hyd9*	550014.4	-4.46	NA	*A1B1*
maker-dna.fa_109-augustus-gene-0.126-mRNA-1	dna.fa_109	ESK88584	AWSO01000637	hydrophobin 2	*Mr.hyd3*	549979.3	-3.74	7.93E-16	*A1B1*
maker-dna.fa_54-augustus-gene-1.126-mRNA-1	dna.fa_54	ESK85505	AWSO01001060	circumsporozoite protein variant vk210	—	252041.5	-0.08	8.07E-01	Both
maker-dna.fa_109-augustus-gene-0.144-mRNA-1	dna.fa_109	ESK88581	AWSO01000637	hydrophobin	*Mr.hyd5*	250878.8	4.30	2.61E-18	*A2B2*
maker-dna.fa_4-augustus-gene-4.117-mRNA-1	dna.fa_4	ESK94583	AWSO01000127	12 kda heat shock protein (glucose and lipid-regulated protein)	—	245340.6	7.76	3.49E-07	*A2B2*
genemark-dna.fa_1-processed-gene-15.213-mRNA-1	dna.fa_1	ESK98373	AWSO01000002	polyubiquitin	—	195953.0	0.83	1.67E-04	*A2B2*
maker-dna.fa_16-augustus-gene-4.277-mRNA-1|maker-dna.fa_16-snap-gene-4.72-mRNA-1	dna.fa_16	ESK96617| ESK96618	AWSO01000045	pro41 protein	—	185285.2	-0.82	8.12E-06	*A1B1*
maker-dna.fa_109-augustus-gene-0.146-mRNA-1	dna.fa_109	ESK88578	AWSO01000637	hydrophobin 2	*Mr.hyd8*	155961.6	-9.02	NA	*A1B1*

Table 4.11 Cont.

[1]Gene name MCA2952	[2]Scaffold MCA2952	[3]Gene name MCA2997	[4]Scaffold MCA2997	[5]Annotation	[6]Given name	[7]baseMean	[8]*log2 FC*	[9]*padj*	[10]Mat. type
maker-dna.fa_23-augustus-gene-7.165-mRNA-1|maker-dna.fa_23-augustus-gene-7.166-mRNA-1	dna.fa_23	ESK92957| ESK92958	AWSO01000231	glycine-rich RNA binding protein	—	125617.1	-0.57	1.04E-06	*A1B1*
maker-dna.fa_4-augustus-gene-9.137-mRNA-1	dna.fa_4	ESK96936	AWSO01000038	extracellular serine-rich	—	124789.8	-0.88	5.33E-06	*A1B1*
maker-dna.fa_111-augustus-gene-0.161-mRNA-1	dna.fa_111	ESK81379	AWSO01002508	serine protease inhibitor	—	122664.7	-5.90	1.04E-06	*A1B1*
maker-dna.fa_22-augustus-gene-7.42-mRNA-1	dna.fa_22	ESK87056	AWSO01000847	peptidyl-prolyl cis-trans isomerase	—	118553.8	0.15	2.37E-01	Both
maker-dna.fa_109-augustus-gene-0.145-mRNA-1	dna.fa_109	ESK88580	AWSO01000637	hydrophobin 2	*Mr.hyd6*	114163.4	2.89	3.42E-03	*A2B2*
maker-dna.fa_24-augustus-gene-8.16-mRNA-1	dna.fa_24	ESK94609	AWSO01000125	hypothetical protein	—	110788.3	-2.06	3.27E-13	*A1B1*
maker-dna.fa_1-snap-gene-5.163-mRNA-1|snap_masked-dna.fa_1-processed-gene-5.12-mRNA-1|maker-dna.fa_1-augustus-gene-5.90-mRNA-1	dna.fa_1	ESK98121| ESK98122| ESK98123	AWSO01000010	hypothetical protein	—	110271.4	-0.51	1.16E-02	Both
genemark-dna.fa_1-processed-gene-14.199-mRNA-1	dna.fa_1	ESK98327	AWSO01000002	hypothetical protein	—	110099.3	-1.20	5.40E-12	*A1B1*
maker-dna.fa_18-augustus-gene-10.260-mRNA-1	dna.fa_18	ESK90832	AWSO01000402	heme peroxidase	—	107993.5	-0.57	3.09E-01	Both

Table 4.11 Cont.

[1]Gene name MCA2952	[2]Scaffold MCA2952	[3]Gene name MCA2997	[4]Scaffold MCA2997	[5]Annotation	[6]Given name	[7]baseMean	[8]*log2 FC*	[9]*padj*	[10]Mat. type
maker-dna.fa_51-augustus-gene-1.46-mRNA-1	dna.fa_51	ESK89835	AWSO01000510	hypothetical protein	—	99711.2	-0.26	3.77E-01	Both
maker-dna.fa_49-augustus-gene-4.187-mRNA-1	dna.fa_49	ESK89445	AWSO01000560	manganese superoxide dismutase	—	97217.0	0.48	4.27E-02	Both
maker-dna.fa_31-snap-gene-4.35-mRNA-1	dna.fa_31	ESK88546	AWSO01000639	aldehyde dehydrogenase	—	96914.1	-0.10	6.43E-01	Both
maker-dna.fa_109-augustus-gene-0.143-mRNA-1	dna.fa_109	ESK88582	AWSO01000637	hydrophobin	*Mr.hyd4*	93375.0	4.26	3.78E-21	*A2B2*
maker-dna.fa_4-augustus-gene-1.36-mRNA-1	dna.fa_4	ESK91624	AWSO01000333	hypothetical protein	—	91848.9	-1.94	1.68E-25	*A1B1*
maker-dna.fa_21-snap-gene-6.76-mRNA-1	dna.fa_21	ESK92232	AWSO01000299	vacuolar ATPase 98 kda	—	88728.3	-4.68	5.71E-103	*A1B1*
maker-dna.fa_6-augustus-gene-20.121-mRNA-1	dna.fa_6	ESK84368	AWSO01001310	putative aldo-keto reductase	—	86958.2	2.53	9.97E-18	*A2B2*
maker-dna.fa_6-augustus-gene-10.104-mRNA-1	dna.fa_6	ESK88777	AWSO01000619	hypothetical protein	—	86439.4	0.66	5.80E-01	Both

Table 4.11 Cont.

[1]Gene name MCA2952	[2]Scaffold MCA2952	[3]Gene name MCA2997	[4]Scaffold MCA2997	[5]Annotation	[6]Given name	[7]baseMean	[8]*log2 FC*	[9]*padj*	[10]Mat. type
maker-dna.fa_16-augustus-gene-3.160-mRNA-1	dna.fa_16	ESK96614	AWSO01000045	short-chain dehydrogenase reductase sdr	—	83027.9	0.55	1.26E-02	Both

[1] Gene name based on the MCA2952 MAKER annotation

[2] Scaffold where gene is in the MCA2952 genome

[3] GenBank accession number associated with gene in the MCA2997 genome; genes in the same row are orthologs.

[4] GenBank accession number of the scaffold where the gene is in the MCA2997 genome.

[5] Annotation is based on the MCA2997 annotation provided in NCBI.

[6] If gene is a hydrophobin, the given name from (Table 4.8) is displayed.

[7] Mean of the normalized read counts of all the samples calculated by *DESeq2*; only genes .

[8] The $\log_2 Fold\ Change$ of normalized read count means of each mating type group calculated by *DESeq2*.

[9] Adjusted probability value calculated by *DESeq2*.

[10] Mating type in which there was a high expression of that gene; if $log2FC < 0$ and $padj < 0.01$, the gene had a significantly higher expression in mating type *A1B1*; if $log2FC > 0$ and $padj < 0.01$, the gene had a significantly higher expression in mating type *A2B2*; if $padj > 0.01$, the gene was expressed similarly in both mating type groups; *padj* for hydrophobins *Mr.hyd9* and *Mr.hyd8* were not calculated as there were outliers detected by Cook's distance calculations but looking at individual read counts it is clear that these two hydrophobins are highly expressed in mating type *A1B1* (APPENDIX H).

4.5 Discussion

4.5.1 Genome assembly and annotation

Moniliophthora roreri MCA2952 has an assembled genome size of 59.7 Mb distributed in 257 scaffolds, 7.5 Mb larger and about 3,000 scaffolds less than the *M. roreri* MCA2997 genome (Table 4.6). Criteria for a well annotated genome include having AED values of less than 0.5 for 90% of predicted proteins, and a recognizable functional domain in at least 50% of predicted proteins (Campbell et al. 2014a). Our annotation of the *M. roreri* MCA2952 genome meet these criteria (APPENDIX B). Additionally, the big difference in genomic statistics indices like N50 and final number of scaffolds (Table 4.6), makes it the most complete genome generated thus far for a member of this genus.

4.5.2 Pathogenicity within Marasmiineae

The ecological specialization of fungi determines the family of enzymes and effector proteins they utilize. Lignin-degrading or white rot fungi, for example, have a specialized set of CAZy and peroxidases that brown rot fungi do not (Floudas et al. 2012). CAZymes are also important players in host-microbe interactions as pathogenic fungi in general tend to have higher numbers of CAZymes than saprotrophs (Zhao et al. 2013). In the following sections hypotheses on the emergence of pathogenicity within the Marasmiineae based on the results of effectors and CAZymes found in this study will be provided.

4.5.2.1 Pathogenicity in *M. roreri*

4.5.2.1.1 Hydrophobins

The cuticle is a hydrophobic network composed of cutin and waxes (Koch and Ensikat 2008) and constitutes the first point of interaction between a pathogen and its plant host, thus it is the most basal form of protection that plants have against pathogens (Serrano et al. 2014). Cacao pods have an epidermal wax layer content that can vary from 12 to 60 µg/cm^2 depending on the genotype (Nyadanu et al. 2012b, 2012a). Also, a wax synthase has been recognized as a passive defense-associated protein in cacao (Jones et al. 2002). Genotypes with a higher wax content on pods show higher levels of tolerance to *Phytophthora* spp. infections; this tolerance significantly diminishes when outer wax is

physically removed (Nyadanu et al. 2012b). Therefore, this hydrophobic structural component of cacao pods provides a natural protection, that pathogens like *M. roreri* need to overcome to be successful pathogens.

Cacao pods are the only known substrate on which *M. roreri* is able to thrive; i.e., the life cycle of this pathogen occurs entirely within cacao pods (Bailey et al. 2018b). Therefore, *M. roreri* must have evolved effective mechanisms to: 1) disseminate between cacao pods; and 2) to penetrate and colonize the wax-coated cacao pod. It has been shown that *M. roreri* can produce up to 44 spores/cm^2 or 7 billion spores in a mature cacao pod (Campuzano 1976), and in this study it is shown that hydrophobins are potential effectors that have expanded during the evolution of *M. roreri* (Figures Figure 4.2 and Figure 4.3). Hydrophobins are fungal SSPs that play major biological roles, e.g., in production of aerial structures like spores and attachment to hydrophobic surfaces (Wösten 2001, Kulkarni et al. 2017). The hydrophobins in *M. roreri* have already been identified and found to be differentially expressed during the biotrophic and necrotrophic phases of infection (Bailey et al. 2013, Meinhardt et al. 2014). However, the results presented here, i.e.: 1) hydrophobins are significantly expanded effectors within the Marasmiaceae; 2) that they are uniquely rearranged in a specific genomic region conforming a hydrophobin-effector cassette (Figure 4.3); 3) and that they are the top expressed genes in *M. roreri* cultures, set hydrophobins as likely major virulence factors and determinant players in spore production, thus in aerial dissemination.

4.5.2.1.2 CAZy repertoire of *M. roreri*

Carbohydrates and lignin account for 32 and 21%, respectively, of the dried composition of cacao pod husks (Vriesmann et al. 2011). Therefore, for *M. roreri* to successfully colonize cacao pods, especially during the necrotrophic phase, it needs a full set of enzymes capable of degrading such compounds. This is exactly what was found: *M. roreri* has not lost CAZymes during evolution compared to most Marasmiineae species examined (Figure 4.4). The family with the highest number of proteins in *M. roreri* is AA3_2; it also has a high number of laccases (AA1_1; Figure 4.4). This family comprises aryl-alcohol oxidases, glucose oxidases and laccases which are commonly found in white-rotting fungi and

participate in lignin degradation (Levasseur et al. 2013). Due to the high lignin content of cacao pod husks, a vast AA3_2 enzyme armament is suitable for *M. roreri.* Additionally, it has been shown that the fructose content in cacao pods decreases as FPR progresses (Bailey et al. 2013). This is probably the result of the carbohydrate assimilation capacity of the pathogen thanks to its CAZy repertoire, which is as abundant as it is in other Marasmiineae species.

4.5.2.1.3 Ricin B lectin effector

Moniliophthora roreri unlike its other examined Marasmiineae counterparts has genes that code for effectors with amino acid repeats, NLS and a Ricin B-lectin domain (APPENDIX G). Lectins are ubiquitous proteins in microbes that recognize carbohydrates and glycoconjugates with high a levels of specificity, facilitating the anchorage of the pathogen (Khan and Khan 2011, Varrot et al. 2013). Therefore, one of their most important functions in symbiotic microbes is host recognition, as shown in bacterial plant-mutualistic *Frankia alni* and pathogenic *Escherichia coli* which, prior to symbiosis establishment, recognize and bind to polyglucosamines from *Alnus* spp. roots and sugars like D-mannose from uroepithelial cells, respectively (Kau et al. 2005, Pujic et al. 2012). In fungi, sugar-specific lectins are also widespread. For example, fucose-binding lectins have been reported in the rice pathogen *Aspergillus oryzae* and the saprotrophs *Aleuria aurantia* and *Rhizopus stolonifer*, while specific galactose polysaccharide-binding lectins occur in *Marasmius oreades*; all have multiple medicinal, serological and biotechnological applications (Fukumori et al. 1989, Winter et al. 2002, Oda et al. 2003, Matsumura et al. 2007). In fungal pathogenicity, chitin-binding lectin effectors from the tomato pathogen *Passalora fulva* are needed to sequester its own chitin molecules to avoid host recognition upon infection and have been shown to enhance virulence of *Fusarium oxysporum* on tomato (van den Burg et al. 2007, Bolton et al. 2008, de Jonge et al. 2010). Lectin effectors from *Magnaporthe oryzae* and *Zymoseptoria tritici* have similar protection roles for the pathogens during disease establishment in rice and wheat, respectively (Marshall et al. 2011, Mentlak et al. 2012). Even a lectin from the root endophyte *Piriformospora indica* has the potential to initiate pathogenic interactions by avoiding β-glucan-triggered immunity (Wawra et al. 2016).

Ricin-B domain effectors were originally described from a toxin produced by the human pathogen *Clostridium difficile* (Teneberg et al. 1990, Kuehne et al. 2010). They typically bind galactose-containing glycan receptors (Rutenber and Robertus 1991, De Schutter and Van Damme 2015) and can lead to cell death upon infection (Chen et al. 2012, Hasan et al. 2015). Ricin B-domain containing proteins are associated with enzymes that degrade sugars (Kruger et al. 2002). Monosaccharide composition analysis of dried cacao pod husks revealed that galactose alone accounts for about 3.2% of total weight (Vriesmann et al. 2011). Therefore, the Ricin B lectin effectors from *M. roreri* might recognize some galactose-containing receptors inside cacao pods and trigger development of disease.

4.5.2.2 Pathogenicity in *M. perniciosa*

4.5.2.2.1 The CAZy repertoire is contracted in *M. perniciosa*

Overall, the hemibiotrophic pathogen *M. perniciosa* has a reduced repertoire of CAZymes compared to other Marasmiineae species (Figure 4.4). This might have something to do with its life cycle and host range (Teixeira et al. 2015). Unlike *M. roreri, M. perniciosa* infects not only cacao, but plants in the Solanaceae and Bignoniaceae families, and infects not only the fruits but virtually all aerial parts of the host at any stage of development (Griffith and Hedger 1994, Teixeira et al. 2015). Even more, *M. perniciosa* differs from other hemibiotrophic fungi in that its biotrophic phase is unusually long, lasting more than 60 days (Barau et al. 2015). During this phase, monokaryotic hyphae feed off sugars from the apoplast without entering the plant cell (Frias et al. 1991). Biotrophic and endophytic fungi in general have a reduced repertoire of plant cell wall degradation enzymes (Duplessis et al. 2011, Gazis et al. 2016). The long-lasting biotrophic phase of *M. perniciosa* may explain then the contraction of its CAZyme arsenal.

It has been shown that the successful establishment of the biotrophic phase of *M. perniciosa* greatly depends on a specific chitinase-like effector, aka MpChi (Fiorin et al. 2018). It belongs to the GH18 CAZy family but during evolution has lost its hydrolytic activity. However, it still binds to chitin fragments within the plant and in this way the fungus avoids host recognition; i.e., the pathogen repurposed the function of this enzyme to be useful in pathogenicity (Fiorin et al. 2018). When looking at the GH18 number of proteins, *M.*

perniciosa has only five while *M. roreri* has more than twenty (Figure 4.4; APPENDIX B). Despite this substantial contraction in GH18 protein family, *M. perniciosa* kept few GH18 proteins and, at least one of them (MpChi), was modified during evolution to develop a function in virulence. This is consistent with the neofunctionalization of enzymes for the evolution of effectors previously proposed (Fiorin et al. 2018).

4.5.2.2.2 FAD binding domain- containing effectors

Sugar scarcity in the apoplast is one of the main triggers of the beginning of the necrotrophic phase in *M. perniciosa* (Barau et al. 2015). When this happens, dikaryotic hyphae develop and start to grow inside plant cells where the expression of several necrosis-related effectors like *MpNEP2* are upregulated (Barau et al. 2015). Here, it is shown that *M. perniciosa* experienced an evolutionary expansion of effectors that have either GMC oxidoreductase, FAD linked oxidases or a FAD-dependent oxidoreductase functional domain (

Table 4.9). In *M. roreri*, it has been found that enzymes in these families are highly expressed during the necrotrophic phase of FPR (Meinhardt et al. 2014). It is likely that these *M. perniciosa* effectors are also upregulated in the necrotrophic phase of WBD, in a similar fashion as *MpNEP2* (Barau et al. 2015). GMC and FAD dependent oxidoreductases are known to the generate H_2O_2 during the degradation of plant cell wall components and are widely present in both white and brown rotting fungi (Ferreira et al. 2015). Even more, a FAD-binding virulence factor from the human pathogen *Mycoplasma mycoides* is directly involved in the production and transport of H_2O_2 into host cells leading to cell death (Pilo et al. 2005). Therefore, the predicted effectors containing GMC and FAD-dependent oxidoreductase domains may be involved in host cell death during the necrotrophic phase of WBD by generating toxic H_2O_2.

4.5.2.3 Pathogenicity in *A. mellea*

4.5.2.3.1 The role of CYP, multicopper oxidase and other effectors

The effectorome analysis in Marasmiineae species revealed that the major forest necrotrophic pathogen *A. mellea*, has undergone a significant expansion of CYP and multicopper oxidase effectors during its divergence from other Physalacriaceae species (Figure 4.2; Table 4.10). CYP superfamily-belonging proteins act directly on toxic plant chemicals, like residual phenolic compounds derived from lignin degradation (MacDonald et al. 2011). CYPs have been analyzed in the proteome of *A. mellea* and compared to other saprotrophic and biotrophic plant pathogens in the Basidiomycota (Qhanya et al. 2015). They found expansion of specific CYP families in *A. mellea*, just as the results presented here at the effectorome level. One *A. mellea* effector (JGI Protein ID 6722; Table 4.10) belongs to family CYP5136 (Qhanya et al. 2015), which is one of those lineage-specific expanded CYP families. Another CYP-genomic comparison among saprotrophic Agaricomycetes found that *Agaricus bisporus*, the only Agaricales species evaluated in that study, does not have CYP5136 proteins, unlike Polyporales and Boletates genomes (Syed et al. 2014). Also, CYP5136 monooxygenases conform the virulence protein repertoire of the Russulales tree pathogen *Heterobasidion occidentale* (Liu et al. 2018). CYP5136 proteins have been shown to oxidize toxic aromatic hydrocarbons and alkylphenols (Syed et al. 2011). Therefore, it is likely that CYP effectors of the CYP5136 family are necessary in the necrotrophic interactions between *A. mellea* and its host.

The other protein family of effectors from *A. mellea* that are rapidly evolving within the Physalacriaceae are the multicopper oxidases (Table 4.10). These are classified as AA1 in the CAZy database (Levasseur et al. 2013). Laccases or AA1_1 are involved in the production of melanin, another virulence factor that contributes to the protection of the fungal pathogen from host-elicited defense compounds during infection (Coman et al. 2013, Kaur et al. 2019). *Armillaria* rhizomorphs, the main infectious structure of the pathogen, have a melanized outer cortex that accumulates high concentrations of iron and copper ions, among other metals (Rizzo and Blanchette 1992). Precisely iron and copper are required in the formation of CYP and multicopper oxidases, the rapidly evolving lineage-specific family of effectors from *A. mellea* (Table 4.10). Interestingly, all analyzed species in the

Marasmiaceae, one of the few families in the Agaricales that contains several documented plant pathogens (Table 4.2), have more effectors with a transition metal ion binding ontology than other families (Figure 4.2). Like *Armillaria*, several Marasmiaceae produce rhizomorphs (Table 4.2 and references therein). Possibly, concentration of metal ions in the outer cortex of rhizomorphs is also the infectious strategy for many plant-pathogenic Marasmiaceae.

Furthermore, there is a plethora of effectors with unknown functionalities (Figure 4.1; APPENDIX B). In *A. mellea* there are three effectors that contains amino acid repeats and NLS but they do not have a recognizable functional domain or ontology (Figure 4.1; APPENDIX G).However, one of these has as its only *blastp* match an effector protein from *Terribacillus halophilus* (APPENDIX G). *Terribacillus* spp. is a soil-habiting bacterial genus found in Japan, China and Mexico (An et al. 2007, Lu et al. 2015, Dent and Del Castillo 2018), i.e., it has a wide geographical distribution just as *Armillaria* spp. Also, it can establish endophytic and mutualistic associations with vegetables and trees, and parasitic interactions against plant-pathogenic fungi (Lu et al. 2015, Nithya and Babu 2017, Dent and Del Castillo 2018). Therefore, it is reasonable to hypothesize that an *Armillaria* sp. effector derives from this bacterial genus.

4.5.3 Final remarks

One of the objectives of this study was to identify effector repertoire differences among thirteen total pathogenic and non-pathogenic Marasmiineae species. To accomplish this, two software were independently run to predict secretomes (SignalP and Phobius); two other independent software runs were used to predict effectoromes (EffectorP 1.0 and EffectorP 2.0); and programs T-REKS and NLStradamus were used to predict whether a protein has amino acid repeats and NLS, respectively. If a protein was positive in all software runs, i.e., they passed all of these *in silico* filters, it would appear in the last column of Figure 4.1 and was annotated in APPENDIX G. Some of the discussions presented are based on these selected effectors as they are highly likely to interact with the host and be determinant virulence factors. For example, the fact that the ricin B-domain lectin effectors

passed all software filters in both *M. roreri* genomes suggests a likely involvement in pathogenicity.

Additionally, the thirteen effectoromes were subjected to an analysis of GO evolution with CAFE. Interestingly, the three pathogenic species evaluated (*M. roreri*, *M. perniciosa* and *A. mellea*) had effectors with GOs evolving rapidly. The most outstanding ones are the hydrophobins from *M. roreri* (*Mr.hyd3-9*), which not only are predicted as effectors but are arranged in the same genomic region and seem to be the result of gene duplication events during evolution (Figure 4.3). Even more, these hydrophobins are among the top-expressed genes in sporulating cultures of *M. roreri*. It is concluded then that *M. roreri* has taken advantage of the functionality of hydrophobins, so it can establish successful pathogenic interactions and disseminate via production of spores. Furthermore, the evolution of CAZyme families in the thirteen genomes was also analyzed. Its most outstanding finding was the significant contraction of most CAZyme families in *M. perniciosa.*

Finally, this study has two important contributions to the scientific community. First it makes available the most complete genome of a *Moniliophthora* species (*M. roreri* MCA2952) and second, it presents program SyLOCAL to analyze synteny of genes between two genomes. The synteny analysis of hydrophobins shows its functionality. SyLOCAL scripts, example files and specifications can be found in APPENDIX F.

APPENDIX A

Table A.1 Sample information of *Moniliophthora roreri* collections

Samples	Country	City / town	Province	GPS X	GPS Y	Isolation Date	Type of Sample	Host	MLL or MLG	Pod	Tree	Mating type
MCA2954	Belize	Maya Mopán	Stann Creek	16.66	-88.52	Sep-04	Pure culture	Unknown	MLG_5	lf	JK	A1B1
JD_AB1.1	Bolivia	Sapecho	Sud Yungas, La Paz	-15.56	-67.32	Oct 31 2016	NL - white stroma	Trinitario	MLL_5	dq	CJ	A2B2
JD_AB2	Bolivia	Sapecho	Sud Yungas, La Paz	-15.56	-67.33	Oct 31 2016	NL - white stroma	SCA6 x ICS6	MLL_5	dr	CK	A2B2
JD_AB2.3	Bolivia	Sapecho	Sud Yungas, La Paz	-15.56	-67.33	Oct 31 2016	NL - white stroma	SCA6 x ICS6	MLL_5	dt	CK	A2B2
JD_AB3	Bolivia	Sapecho	Sud Yungas, La Paz	-15.56	-67.33	Oct 31 2016	NL - white stroma	Trinitario	MLL_5	du	CL	A2B2
JD_AB5.2	Bolivia	Alto Beni	Caranavi, La Paz	-15.58	-67.31	Oct 31 2016	NL - white stroma	ICS8 x SCA6	MLL_5	dx	CN	A2B2
JD_AB6.1	Bolivia	Alto Beni	Caranavi, La Paz	-15.58	-67.31	Oct 31 2016	NL - white stroma	Híbrido	MLL_5	dy	CO	A2B2
JD_AB7.1	Bolivia	Alto Beni	Caranavi, La Paz	-15.58	-67.31	Oct 31 2016	NL - white stroma	IMC x ICS 111	MLL_5	dz	CP	A2B2
JD_AB8.1	Bolivia	Alto Beni	Caranavi, La Paz	-15.58	-67.31	Oct 31 2016	NL - white stroma	Trinitario	MLL_5	eb	CQ	A2B2
JD_AB9.1	Bolivia	Alto Beni	Caranavi, La Paz	-15.58	-67.31	Oct 31 2016	NL - white stroma	ICS 1 x IMC 67	MLL_5	ec	CR	A2B2
JD_Y9	Colombia	Yacopí	Cundinamarca	5.46	-74.36	Julio 2017	Pure culture	Unknown	MLG_39	im	GQ	A2Bx
JD_Y14	Colombia	Yacopí	Cundinamarca	5.44	-74.35	Julio 2017	Pure culture	Unknown	MLG_2	in	GR	A3B2
JD_Y17	Colombia	Yacopí	Cundinamarca	5.44	-74.35	Julio 2017	Pure culture	Unknown	MLG_2	io	GS	A3B2
JD_Y24	Colombia	Yacopí	Cundinamarca	5.45	-74.34	Julio 2017	Pure culture	Unknown	MLG_1	ip	GT	AxB2
JD_Y27	Colombia	Yacopí	Cundinamarca	5.49	-74.33	Julio 2017	Pure culture	Unknown	MLL_2	iq	GU	A3B2
JD_Y27'	Colombia	Yacopí	Cundinamarca	5.49	-74.33	Julio 2017	Pure culture	Unknown	MLL_2	iq	GU	A3B2
JD_Y30	Colombia	Yacopí	Cundinamarca	5.49	-74.33	Julio 2017	Pure culture	Unknown	MLL_2	ir	GV	A3B2
JD_N1	Colombia	Nilo	Cundinamarca	4.33	-74.66	Julio 2017	Pure culture	Unknown	MLG_30	is	GW	A3B2
JD_N2	Colombia	Nilo	Cundinamarca	4.33	-74.66	Julio 2017	Pure culture	Unknown	MLG_30	it	GX	A3B2
JD_N3	Colombia	Nilo	Cundinamarca	4.35	-74.65	Julio 2017	Pure culture	Unknown	MLG_30	iu	GY	A3B2
JD_SA1	Colombia	Palestina	Caldas	5.07	-75.69	Julio 2017	Pure culture	Unknown	MLG_15	iv	GZ	AxB2
JD_SA3	Colombia	Palestina	Caldas	5.07	-75.69	Julio 2017	Pure culture	Unknown	MLG_15	iw	HA	AxB2
JD_SA6	Colombia	Palestina	Caldas	5.07	-75.69	Julio 2017	Pure culture	Unknown	MLG_12	ix	HB	AxB2
JD_CH1	Colombia	Campoalagre	Huila	2.69	-75.30	Julio 2017	Pure culture	Unknown	MLL_1	iy	HC	AxB2
JD_CH2	Colombia	Campoalagre	Huila	2.69	-75.31	Julio 2017	Pure culture	Unknown	MLG_16	iz	HD	AxB2
JD_CH3	Colombia	Campoalagre	Huila	2.69	-75.31	Julio 2017	Pure culture	Unknown	MLL_1	ja	HE	AxB2
JD_CR1.1	Costa Rica	Matina	Limon	10.08	-83.29	April 12 2017	NL - white stroma	Unknown	MLG_6	gc	EU	A1B1
JD_CR1.2	Costa Rica	Matina	Limon	10.08	-83.29	April 12 2017	NL - white stroma	Unknown	MLG_6	gc	EU	A1B1
JD_CR3.3	Costa Rica	Matina	Limon	10.10	-83.38	April 12 2017	NL - white stroma	Unknown	MLG_6	ge	EW	A1B1
JD_CR4.1	Costa Rica	Matina	Limon	10.10	-83.38	April 12 2017	NL - white stroma	Unknown	MLG_6	gf	EX	A1B1
JD_CR4.2	Costa Rica	Matina	Limon	10.10	-83.38	April 12 2017	NL - tissue	Unknown	MLG_6	gf	EX	A1B1
JD_CR6.1	Costa Rica	Turrialba	Cartago	9.90	-83.66	April 12 2017	NL - white stroma	Unknown	MLG_6	gi1	EZ	A1B1
JD_CR9.1	Costa Rica	Matina	Limon	10.10	-83.38	April 12 2017	NL - white stroma	Unknown	MLG_6	gl	FC	A1B1
JD_CR9.3	Costa Rica	Matina	Limon	10.10	-83.38	April 12 2017	NL - tissue	Unknown	MLG_6	gl	FC	A1B1
JD_CR9.4	Costa Rica	Matina	Limon	10.10	-83.38	April 12 2017	Pure culture	Unknown	MLG_6	gl	FC	A1B1
JD_CR12.2	Costa Rica	Turrialba	Cartago	9.90	-83.66	April 12 2017	NL - tissue	Unknown	MLG_4	go	FF	A1B1
JD_CR12.3	Costa Rica	Turrialba	Cartago	9.90	-83.66	April 12 2017	Pure culture	Unknown	MLG_3	go	FF	A1B1

Table A.1 Cont.

Samples	Country	City / town	Province	GPS X	GPS Y	Isolation Date	Type of Sample	Host	MLL or MLG	Pod	Tree	Mating type
JD_E1	Ecuador	Naranjal	Guayas	-2.55	-79.52	Oct 03 2016	NL - tissue	*T. grandiflorum*	MLG_53	a	A	AxB2
JD_E2	Ecuador	Naranjal	Guayas	-2.55	-79.52	Oct 03 2016	NL - tissue	*T. grandiflorum*	MLG_52	a	A	AxB2
JD_E25	Ecuador	Naranjal	Guayas	-2.55	-79.52	Oct 03 2016	NL - white stroma	*T. grandiflorum*	MLG_20	a	A	AxB2
JD_E6	Ecuador	Naranjal	Guayas	-2.55	-79.52	Oct 03 2016	NL - tissue	*T. grandiflorum*	MLG_49	b	B	A1B2
JD_E7	Ecuador	Naranjal	Guayas	-2.55	-79.52	Oct 03 2016	NL - tissue	*T. grandiflorum*	MLG_48	b	B	A1B1
JD_E9	Ecuador	Naranjal	Guayas	-2.55	-79.52	Oct 03 2016	Pure culture	*T. grandiflorum*	MLG_48	b	B	A1B1
JD_E11	Ecuador	Naranjal	Guayas	-2.55	-79.52	Oct 03 2016	NL - tissue	*T. grandiflorum*	MLG_48	c	B	A1B1
JD_E13	Ecuador	Naranjal	Guayas	-2.55	-79.52	Oct 03 2016	NL - tissue	*T. grandiflorum*	MLG_56	d	C	A1B1
JD_E17	Ecuador	Naranjal	Guayas	-2.56	-79.52	Oct 03 2016	Pure culture	CCN-51	MLG_45	e	D	A1B2
JD_E21	Ecuador	Balao	Guayas	-2.95	-79.65	Oct 03 2016	Pure culture	Nacional	MLG_54	g	F	AxBx
JD_E22	Ecuador	Balao	Guayas	-2.96	-79.65	Oct 03 2016	NL - white stroma	Nacional	MLG_45	h	G	A1B2
JD_HC1	Ecuador	Naranjal	Guayas	-2.56	-79.51	May-2016	NL - tissue	CCN-51	MLG_46	ej	CY	AxB1
JD_HC5	Ecuador	Naranjal	Guayas	-2.55	-79.51	May-2016	NL - tissue	CCN-51	MLG_50	en	DC	AxB2
JD_HC7	Ecuador	Naranjal	Guayas	-2.56	-79.51	May-2016	NL - tissue	CCN-51	MLG_42	ep	DE	AxB1
JD_HC10	Ecuador	Naranjal	Guayas	-2.56	-79.51	May-2016	NL - tissue	CCN-51	MLG_55	es	DH	A1B1
JD_HC12	Ecuador	Naranjal	Guayas	-2.56	-79.51	May-2016	NL - tissue	CCN-51	MLG_57	eu	DJ	A1B2
JD_HC13	Ecuador	Naranjal	Guayas	-2.57	-79.51	May-2016	NL - tissue	CCN-51	MLG_44	ev	DK	A1B1
JD_HC14	Ecuador	Naranjal	Guayas	-2.54	-79.52	May-2016	NL - tissue	CCN-51	MLG_11	ew	DL	AxB2
JD_HC17	Ecuador	Naranjal	Guayas	-2.56	-79.51	May-2016	NL - tissue	CCN-51	MLG_28	ez	DQ	AxB1
JD_HC21	Ecuador	Naranjal	Guayas	-2.57	-79.51	May-2016	NL - tissue	CCN-51	MLG_52	fd	DU	A1B2
JD_HC22	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_27	fe	DV	A1B2
JD_HC23	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_51	ff	DW	AxB2
JD_HC24	Ecuador	Naranjal	Guayas	-2.56	-79.51	May-2016	NL - tissue	CCN-51	MLG_44	fg	DX	A1B1
JD_HC25	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_24	fh	DY	A1B1
JD_HC26	Ecuador	Naranjal	Guayas	-2.55	-79.52	May-2016	NL - tissue	CCN-51	MLG_41	fi	DZ	AxB2
JD_HC27	Ecuador	Naranjal	Guayas	-2.55	-79.52	May-2016	NL - tissue	CCN-51	MLG_48	fj	EA	A1B1
JD_HC28	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_53	fk	EB	A1B2
JD_HC29	Ecuador	Naranjal	Guayas	-2.55	-79.52	May-2016	NL - tissue	CCN-51	MLG_40	fl	EC	A1B1
JD_HC30	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_9	fm	ED	AxBx
JD_HC31	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_44	fn	EF	A1B2
JD_HC32	Ecuador	Naranjal	Guayas	-2.57	-79.51	May-2016	NL - tissue	CCN-51	MLG_49	fo	EG	A1B2
JD_HC33	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_47	fp	EH	A1B1
JD_HC34	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_49	fq	EI	A1B2
JD_HC35	Ecuador	Naranjal	Guayas	-2.55	-79.52	May-2016	NL - tissue	CCN-51	MLG_26	fr	EJ	A1B2
JD_HC36	Ecuador	Naranjal	Guayas	-2.55	-79.52	May-2016	NL - tissue	CCN-51	MLG_43	fs	EK	A1B2
JD_HC37	Ecuador	Naranjal	Guayas	-2.55	-79.52	May-2016	NL - tissue	CCN-51	MLG_49	ft	EL	A1B2
JD_HC38	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_43	fu	EM	A1B1
JD_HC39	Ecuador	Naranjal	Guayas	-2.55	-79.52	May-2016	NL - tissue	CCN-51	MLG_25	fv	EN	AxB1
JD_HC40	Ecuador	Naranjal	Guayas	-2.57	-79.51	May-2016	NL - tissue	CCN-51	MLG_43	fw	EO	A1B1
JD_HC41	Ecuador	Naranjal	Guayas	-2.57	-79.51	May-2016	NL - tissue	CCN-51	MLG_19	fx	EP	AxB2
JD_HC42	Ecuador	Naranjal	Guayas	-2.57	-79.51	May-2016	NL - tissue	CCN-51	MLG_23	fy	EQ	A1B2
JD_HC43	Ecuador	Naranjal	Guayas	-2.56	-79.52	May-2016	NL - tissue	CCN-51	MLG_43	fz	ER	A1B1
JD_HC44	Ecuador	Naranjal	Guayas	-2.56	-79.51	May-2016	NL - tissue	CCN-51	MLG_43	ga	ES	A1B1

Table A.1 Cont.

Samples	Country	City / town	Province	GPS X	GPS Y	Isolation Date	Type of Sample	Host	MLL or MLG	Pod	Tree	Mating type
JD_HC45	Ecuador	Naranjal	Guayas	-2.55	-79.52	May-2016	NL - tissue	CCN-51	MLG_46	gb	ET	AxB1
DIS106i	Ecuador	Esmeraldas	Esmeraldas	0.97	-79.66	1999	Pure culture	Unknown	MLG_5	lh	JM	A1B1
Dis371.1.3	Ecuador	Esmeraldas	Esmeraldas	0.97	-79.66	2003	Pure culture	Unknown	MLG_5	li	JN	A1B1
JD_IQ1.1	Iquitos	Iquitos	Maynas, Loreto	-3.76	-73.27	Oct 17 2016	NL - white stroma	Criollo	MLG_17	cg	BL	AxBx
JD_IQ1.2	Iquitos	Iquitos	Maynas, Loreto	-3.76	-73.27	Oct 17 2016	NL - white stroma	Criollo	MLG_17	ch	BL	AxBx
JD_IQ2.1	Iquitos	Iquitos	Maynas, Loreto	-3.76	-73.27	Oct 17 2016	NL - white stroma	Criollo	MLG_17	ci	BM	AxBx
JD_IQ2.2	Iquitos	Iquitos	Maynas, Loreto	-3.76	-73.27	Oct 17 2016	NL - white stroma	Criollo	MLG_17	cj	BN	AxBx
JD_IQ3	Iquitos	Iquitos	Maynas, Loreto	-3.76	-73.27	Oct 17 2016	NL - white stroma	Criollo	MLG_17	ck	BO	AxBx
JD_IQ4	Iquitos	Iquitos	Maynas, Loreto	-3.76	-73.27	Oct 17 2016	NL - white stroma	Criollo	MLG_17	cl	BP	AxBx
JD_IQ11.1	Iquitos	Iquitos	Maynas, Loreto	-3.96	-73.43	Oct 18 2016	NL - white stroma	*T. bicolor*	MLG_17	cm	BQ	AxBx
JD_IQ18.1	Iquitos	Belén	Maynas, Loreto	-3.78	-73.24	Oct 19 2016	NL - white stroma	Criollo	MLG_17	co	BS	AxBx
JD_IQ19.1	Iquitos	Belén	Maynas, Loreto	-3.78	-73.24	Oct 19 2016	NL - white stroma	Criollo	MLG_17	cp	BT	AxBx
JD_IQ19.3	Iquitos	Belén	Maynas, Loreto	-3.78	-73.24	Oct 19 2016	NL - white stroma	Criollo	MLG_17	cq	BT	AxBx
JD_IQ19.4	Iquitos	Belén	Maynas, Loreto	-3.78	-73.24	Oct 19 2016	NL - white stroma	Criollo	MLG_17	cr	BT	AxBx
JD_IQ20	Iquitos	Belén	Maynas, Loreto	-3.78	-73.24	Oct 19 2016	NL - white stroma	Criollo	MLG_17	ct	BU	AxBx
JD_IQ21.1	Iquitos	Belén	Maynas, Loreto	-3.78	-73.24	Oct 19 2016	NL - white stroma	Criollo	MLG_17	cu	BV	AxBx
JD_IQ21.2	Iquitos	Belén	Maynas, Loreto	-3.78	-73.24	Oct 19 2016	NL - white stroma	Criollo	MLG_17	cv	BW	AxBx
JD_Jam2.1	Jamaica	Crooked River	Clarendon, Middlesex	18.14	-77.31	Dec 06 2016	NL - white stroma	Unknown	MLL_3	ee	CT	A1B1
JD_Jam2.2	Jamaica	Crooked River	Clarendon, Middlesex	18.14	-77.31	Dec 06 2016	NL - tissue	Unknown	MLL_3	ee	CT	A1B1
JD_Jam2.7	Jamaica	Crooked River	Clarendon, Middlesex	18.14	-77.31	Dec 06 2016	Pure culture	Unknown	MLL_3	ee	CT	A1B1
MCA2952	Mexico	Pichucalco	Chiapas	17.51	-93.13	Mar-2005	Pure culture	Unknown	MLG_5	le	JI	A1B1
MCA2518	Panama	Colón	Colón	9.36	-79.90	1999	Pure culture	Unknown	MLG_5	lg	JL	A1B1
JD_Piu1	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Oct 05 2016	NL - white stroma	Trinitario	MLL_5	m	K	A2B2
JD_Piu5	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Oct 05 2016	NL - white stroma	Blanco Piurano	MLL_5	o	M	A2B2
JD_Piu6	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Oct 05 2016	NL - white stroma	Trinitario	MLG_38	p	N	A2B2
JD_Piu8	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Oct 05 2016	NL - white stroma	Blanco Piurano	MLL_5	q	O	A2B2
JD_Piu11	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Oct 05 2016	NL - white stroma	Blanco Piurano	MLL_5	s	Q	A2B2
JD_Piu12	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Oct 05 2016	NL - white stroma	Marañon	MLL_5	t	R	A2B2
JD_Piu15	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Oct 05 2016	NL - white stroma	Blanco Piurano	MLL_5	v	S	A2B2
JD_Piu17	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Oct 05 2016	NL - white stroma	Blanco Piurano	MLL_5	x	U	A2B2
JD_Piu19	Peru	Lalaquiz	Huancabamba, Piura	-5.24	-79.66	Oct 06 2016	NL - white stroma	Trinitario	MLL_5	y	V	A2B2
JD_Piu20-1	Peru	Lalaquiz	Huancabamba, Piura	-5.24	-79.66	Oct 06 2016	NL - white stroma	Blanco Piurano	MLL_5	z	W	A2B2
JD_Piu20-2	Peru	Lalaquiz	Huancabamba, Piura	-5.24	-79.66	Oct 06 2016	NL - white stroma	Blanco Piurano	MLL_5	aa	W	A2B2
JD_Piu21	Peru	Lalaquiz	Huancabamba, Piura	-5.24	-79.66	Oct 06 2016	NL - white stroma	Trinitario	MLL_5	ab	X	A2B2
JD_Piu22	Peru	Lalaquiz	Huancabamba, Piura	-5.24	-79.66	Oct 06 2016	NL - white stroma	Trinitario	MLL_5	ac	Y	A2B2
JD_Piu24	Peru	Lalaquiz	Huancabamba, Piura	-5.25	-79.67	Oct 06 2016	NL - white stroma	Trinitario	MLL_5	ae	AA	A2B2
JD_Piu25	Peru	Lalaquiz	Huancabamba, Piura	-5.25	-79.67	Oct 06 2016	NL - white stroma	CCN-51	MLL_5	af	AB	A2B2
JD_Piu26	Peru	Lalaquiz	Huancabamba, Piura	-5.25	-79.67	Oct 06 2016	NL - white stroma	Blanco Piurano	MLL_5	ag	AC	A2B2
JD_Piu28	Peru	Lalaquiz	Huancabamba, Piura	-5.25	-79.68	Oct 06 2016	NL - white stroma	Criollo	MLL_5	aj	AE	A2B2
JD_Piu29	Peru	Lalaquiz	Huancabamba, Piura	-5.25	-79.68	Oct 06 2016	NL - white stroma	Criollo	MLL_5	ak	AF	A2B2
JD_Piu31	Peru	Lalaquiz	Huancabamba, Piura	-5.25	-79.68	Oct 06 2016	NL - white stroma	Blanco Piurano	MLL_5	am	AH	A2B2
JD_Piu32-1	Peru	Canchaque	Huancabamba, Piura	-5.30	-79.66	Oct 06 2016	NL - white stroma	Blanco Piurano	MLL_5	an	AI	A2B2
JD_Piu32-3	Peru	Canchaque	Huancabamba, Piura	-5.30	-79.66	Oct 06 2016	NL - white stroma	Blanco Piurano	MLL_5	ao	AI	A2B2

Table A.1 Cont.

Samples	Country	City / town	Province	GPS X	GPS Y	Isolation Date	Type of Sample	Host	MLL or MLG	Pod	Tree	Mating type
JD_Piu33	Peru	Canchaque	Huancabamba, Piura	-5.30	-79.66	Oct 06 2016	NL - white stroma	Unknown	MLL_5	ap	AJ	A2B2
JD_Piu34	Peru	Canchaque	Huancabamba, Piura	-5.41	-79.64	Oct 07 2016	NL - white stroma	Blanco Piurano	MLL_5	aq	AK	A2B2
JD_Piu35	Peru	Canchaque	Huancabamba, Piura	-5.41	-79.64	Oct 07 2016	NL - white stroma	Blanco Piurano	MLL_5	ar	AL	A2B2
JD_Piu36	Peru	Canchaque	Huancabamba, Piura	-5.41	-79.64	Oct 07 2016	NL - white stroma	Blanco Piurano	MLL_5	ar	AL	A2B2
JD_Piu37	Peru	Canchaque	Huancabamba, Piura	-5.41	-79.64	Oct 07 2016	NL - white stroma	Blanco Piurano	MLL_5	as	AM	A2B2
JD_Piu38	Peru	Canchaque	Huancabamba, Piura	-5.41	-79.64	Oct 07 2016	NL - white stroma	Blanco Piurano	MLL_5	at	AN	A2B2
JD_Ja2.1	Peru	Yanuyacu Bajo	Jaén, Cajamarca	-5.68	-78.77	Oct 09 2016	NL - white stroma	Criollo	MLL_5	av	AN	A2B2
JD_Ja2.2	Peru	Yanuyacu Bajo	Jaén, Cajamarca	-5.68	-78.77	Oct 09 2016	NL - white stroma	Criollo	MLL_5	aw	AN	A2B2
JD_Ja3.1	Peru	Yanuyacu Bajo	Jaén, Cajamarca	-5.68	-78.77	Oct 09 2016	NL - white stroma	Nacional	MLL_5	ax	AO	A2B2
JD_Ja3.2	Peru	Yanuyacu Bajo	Jaén, Cajamarca	-5.68	-78.77	Oct 09 2016	Pure culture	Nacional	MLL_5	ay	AO	A2B2
JD_Ja5	Peru	Yanuyacu Bajo	Jaén, Cajamarca	-5.68	-78.77	Oct 09 2016	NL - white stroma	Nacional	MLL_5	ba	AQ	A2B2
JD_Ja6	Peru	Canchaque	Huancabamba, Piura	-5.41	-79.64	Oct 09 2016	NL - white stroma	Unknown	MLL_5	bb	AR	A2B2
JD_Ja7.1	Peru	Yanuyacu Bajo	Jaén, Cajamarca	-5.68	-78.77	Oct 09 2016	NL - white stroma	Unknown	MLL_5	bc	AS	A2B2
JD_Ja7.2	Peru	Yanuyacu Bajo	Jaén, Cajamarca	-5.68	-78.77	Oct 09 2016	NL - white stroma	Unknown	MLL_5	bd	AS	A2B2
JD_Ja8.1	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 09 2016	NL - white stroma	Unknown	MLL_5	bf	AT	A2B2
JD_Ja8.2	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 09 2016	NL - white stroma	Unknown	MLL_5	bg	AT	A2B2
JD_Ja9.1	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 09 2016	NL - white stroma	Unknown	MLL_5	bh	AU	A2B2
JD_Ja9.2	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 09 2016	Pure culture	Unknown	MLL_5	bh	AU	A2B2
JD_Ja10.1	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 09 2016	NL - white stroma	Criollo	MLL_5	bi	AV	A2B2
JD_Ja10.2	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 09 2016	NL - white stroma	Criollo	MLL_5	bj	AV	A2B2
JD_Ja11.1	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 10 2016	NL - white stroma	Nacional	MLL_5	bk	AW	A2B2
JD_Ja11.2	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 10 2016	NL - white stroma	Nacional	MLL_5	bl	AW	A2B2
JD_Ja12.1	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 10 2016	NL - white stroma	Nacional	MLL_5	bm	AX	A2B2
JD_Ja13.1	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 10 2016	NL - white stroma	Nacional	MLL_5	bm	AX	A2B2
JD_Ja14.1	Peru	Bellavista	Jaén, Cajamarca	-5.65	-78.71	Oct 10 2016	NL - white stroma	Nacional	MLL_5	bo	AZ	A2B2
JD_Tar1	Peru	Chazuta	San Martín, San Martín	-6.60	-76.15	Oct 15 2016	NL - white stroma	CCN-51	MLL_5	bq	BA	A2B2
JD_Tar3-2	Peru	Chazuta	San Martín, San Martín	-6.60	-76.15	Oct 15 2016	NL - tissue	CCN-51	MLL_5	br	BB	A2B2
JD_Tar4-2	Peru	Chazuta	San Martín, San Martín	-6.61	-76.17	Oct 15 2016	NL - white stroma	CCN-51	MLG_10	bt	BC	A2B2
JD_Tar4-3	Peru	Chazuta	San Martín, San Martín	-6.61	-76.17	Oct 15 2016	NL - white stroma	CCN-51	MLL_5	bu	BC	A2B2
JD_Tar5-2	Peru	Chazuta	San Martín, San Martín	-6.61	-76.15	Oct 15 2016	NL - white stroma	CCN-51	MLL_5	bw	BD	A2B2
JD_Tar6-1	Peru	Chazuta	San Martín, San Martín	-6.58	-76.14	Oct 15 2016	NL - white stroma	CCN-51	MLL_5	by	BE	A2B2
JD_Tar6-2	Peru	Chazuta	San Martín, San Martín	-6.58	-76.14	Oct 15 2016	NL - white stroma	CCN-51	MLL_5	bz	BE	A2B2
JD_Tar7	Peru	Chazuta	San Martín, San Martín	-6.59	-76.14	Oct 15 2016	NL - white stroma	CCN-51	MLL_5	ca	BF	A2B2
JD_Tar8-1	Peru	Chazuta	San Martín, San Martín	-6.58	-76.14	Oct 15 2016	NL - tissue	CCN-51	MLG_35	cb	BG	A2B2
JD_Tar9-4	Peru	Chazuta	San Martín, San Martín	-6.58	-76.14	Oct 15 2016	NL - white stroma	CCN-51	MLG_35	ce	BH	A2B2
JD_Tar9-5	Peru	Chazuta	San Martín, San Martín	-6.58	-76.14	Oct 15 2016	Pure culture	CCN-51	MLG_35	ce	BH	A2B2
JD_Tar11-2	Peru	Chazuta	San Martín, San Martín	-6.58	-76.15	Oct 15 2016	NL - tissue	Criollo	MLL_5	cf	BJ	A2B2
JD_Qui1.1	Peru	Echarati	La Convención, Cusco	-12.77	-72.59	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	cy	BZ	A2B2
JD_Qui2.1	Peru	Echarati	La Convención, Cusco	-12.77	-72.59	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	cz	CA	A2B2
JD_Qui2.2	Peru	Echarati	La Convención, Cusco	-12.77	-72.59	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	da	CA	A2B2
JD_Qui2.3	Peru	Echarati	La Convención, Cusco	-12.77	-72.59	Oct 26 2016	NL - white stroma	Trinitario	MLL_4	db	CA	A2B2
JD_Qui3.1	Peru	Echarati	La Convención, Cusco	-12.77	-72.59	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	dc	CB	A2B2
JD_Qui3.2	Peru	Echarati	La Convención, Cusco	-12.77	-72.59	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	dd	CB	A2B2

Table A.1 Cont.

Samples	Country	City / town	Province	GPS X	GPS Y	Isolation Date	Type of Sample	Host	MLL or MLG	Pod	Tree	Mating type
JD_Qui3.3	Peru	Echarati	La Convención, Cusco	-12.77	-72.59	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	de	CB	A2B2
JD_Qui5	Peru	Echarati	La Convención, Cusco	-12.77	-72.58	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	dg	CD	A2B2
JD_Qui6.1	Peru	Echarati	La Convención, Cusco	-12.77	-72.58	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	dh	CE	A2B2
JD_Qui6.2	Peru	Echarati	La Convención, Cusco	-12.77	-72.58	Oct 26 2016	NL - tissue	Trinitario	MLL_5	dh	CE	A2B2
JD_Qui6.5	Peru	Echarati	La Convención, Cusco	-12.77	-72.58	Oct 26 2016	NL - white stroma	Trinitario	MLL_4	di	CE	A2B2
JD_Qui6.6	Peru	Echarati	La Convención, Cusco	-12.77	-72.58	Oct 26 2016	NL - white stroma	Trinitario	MLL_4	dj	CE	A2B2
JD_Qui7.1	Peru	Echarati	La Convención, Cusco	-12.77	-72.58	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	dk	CF	A2B2
JD_Qui7.2	Peru	Echarati	La Convención, Cusco	-12.77	-72.58	Oct 26 2016	NL - white stroma	Trinitario	MLL_5	dl	CF	A2B2
JD_Qui7.5	Peru	Echarati	La Convención, Cusco	-12.77	-72.58	Oct 26 2016	Pure culture	Trinitario	MLL_5	dl	CF	A2B2
JD_Qui8.1	Peru	Echarati	La Convención, Cusco	-12.77	-75.58	Oct 26 2016	NL - white stroma	Unknown	MLL_5	dm	CG	A2B2
JD_Qui8.2	Peru	Echarati	La Convención, Cusco	-12.77	-75.58	Oct 26 2016	NL - white stroma	Unknown	MLL_5	dn	CG	A2B2
JD_Qui9	Peru	Echarati	La Convención, Cusco	-12.77	-72.58	Oct 26 2016	NL - white stroma	Unknown	MLL_5	do	CH	A2B2
JD_SJB1.1	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Aug 7 2017	NL - white stroma	Blanco Piurano	MLL_5	gp	FG	A2B2
JD_SJB1.2	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Aug 7 2017	NL - white stroma	Blanco Piurano	MLL_5	gp	FG	A2B2
JD_SJB2	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Aug 7 2017	NL - white stroma	Criollo	MLL_5	gq	FH	A2B2
JD_SJB3.1	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Aug 7 2017	NL - white stroma	Criollo	MLL_5	gr	FI	A2B2
JD_SJB3.2	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Aug 7 2017	NL - white stroma	Criollo	MLL_5	gs	FI	A2B2
JD_SJB4b	Peru	Lalaquiz	Huancabamba, Piura	-5.26	-79.68	Aug 7 2017	NL - white stroma	Criollo	MLG_18	gt	FJ	A2B2
JD_Xio1	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - white stroma	CCN-51	MLL_5	jb	HF	A2B2
JD_Xio2	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - white stroma	CCN-51	MLL_5	jc	HG	A2B2
JD_Xio4	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - tissue	CCN-51	MLL_5	jd	HH	A2B2
JD_Xio6	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - tissue	Nativo	MLL_5	je	HI	A2B2
JD_Xio8	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - white stroma	Nativo	MLL_5	jf	HJ	A2B2
JD_Xio10	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - white stroma	Nativo	MLL_5	jh	HL	A2B2
JD_Xio12	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - white stroma	Nativo	MLL_5	ji	HM	A2B2
JD_Xio15	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - white stroma	Nativo	MLL_5	jk	HO	A2B2
JD_Xio17	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - tissue	Nativo	MLL_5	jl	HP	A2B2
JD_Xio19	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - tissue	Nativo	MLL_5	jm	HQ	A2B2
JD_Xio21	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - tissue	Nativo	MLL_5	jn	HR	A2B2
JD_Xio22	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - tissue	Nativo	MLL_5	jn	HR	A2B2
JD_Xio23	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - tissue	Nativo	MLL_5	jo	HS	A2B2
JD_Xio25	Peru	Curimaná	Padre Abad	-8.41	-75.23	9/6/2017	NL - tissue	Nativo	MLG_34	jp	HT	A2B2
JD_HU-01	Peru	Alto San Juan	Leoncio Prado	-9.32	-75.85	Jul-2015	Pure culture	Trinitario	MLL_5	jq	HU	A2B2
JD_HU-03	Peru	Inkari	Leoncio Prado	-9.25	-75.97	Jul-2015	Pure culture	Forastero	MLL_5	js	HW	A2B2
JD_HU-04	Peru	Vista Alegre	Leoncio Prado	-9.28	-75.95	Jul-2015	Pure culture	Forastero	MLL_5	jt	HX	A2B2
JD_HU-05	Peru	Merced de Locro	Leoncio Prado	-9.15	-76.05	Jul-2015	Pure culture	Trinitario	MLL_5	ju	HY	A2B2
JD_HU-06	Peru	Venenillo	Leoncio Prado	-9.08	-76.08	Jul-2015	Pure culture	Trinitario	MLL_5	jv	HZ	A2B2
JD_HU-09	Peru	Las Vegas	Leoncio Prado	-9.18	-75.90	Jul-2015	Pure culture	Trinitario	MLL_5	jy	IC	A2B2
JD_HU-10	Peru	La Victoria	Leoncio Prado	-9.17	-75.93	Jul-2015	Pure culture	Trinitario	MLG_34	jz	ID	A2B2
JD_HU-11	Peru	Lota	Leoncio Prado	-9.30	-76.07	Jul-2015	Pure culture	Trinitario	MLL_5	ka	IE	A2B2
JD_HU-13	Peru	Tulumayo	Leoncio Prado	-9.12	-76.03	Jul-2015	Pure culture	Criollo	MLL_5	kc	IG	A2B2
JD_HU-14	Peru	Arabe	Leoncio Prado	-9.05	-76.05	Jul-2015	Pure culture	Forastero	MLL_5	kd	IH	A2B2
JD_HU-15	Peru	Sai Pai	Leoncio Prado	-9.08	-76.00	Jul-2015	Pure culture	Trinitario	MLL_5	ke	II	A2B2

Table A.1 Cont.

Samples	Country	City / town	Province	GPS X	GPS Y	Isolation Date	Type of Sample	Host	MLL or MLG	Pod	Tree	Mating type
JD_HU-18	Peru	Paraiso	Leoncio Prado	-8.49	-76.39	Jul-2015	Pure culture	Trinitario	MLL_5	kh	IL	A2B2
JD_JU-34	Peru	Kapirushari	Satipo	-11.29	-74.55	Aug-2015	Pure culture	Forastero	MLL_5	ki	IM	A2B2
JD_JU-35	Peru	Los Angeles de Ubiriki	Satipo	-11.24	-74.67	Aug-2015	Pure culture	Criollo	MLL_5	kj	IN	A2B2
JD_HU-17	Peru	Marona	Leoncio Prado	-9.23	-75.95	Aug-2015	Pure culture	Forastero	MLL_5	kk	IK	A2B2
JD_JU-36	Peru	Rio Negro	Satipo	-11.19	-74.66	Aug-2015	Pure culture	Forastero	MLL_5	kk	IO	A2B2
JD_JU-37	Peru	Villa Kapiri	Satipo	-11.13	-74.67	Aug-2015	Pure culture	Criollo	MLL_5	kl	IP	A2B2
JD_JU-38	Peru	Union Capiri	Satipo	-11.09	-74.70	Aug-2015	Pure culture	Criollo	MLL_5	km	IQ	A2B2
JD_JU-39	Peru	San Juan de Cheni	Satipo	-11.09	-74.74	Aug-2015	Pure culture	Criollo	MLL_5	kn	IR	A2B2
JD_SM-19	Peru	Tocache	Tocache	-8.20	-76.56	Jul-2015	Pure culture	Trinitario	MLL_5	kp	IT	A2B2
JD_SM-23	Peru	Buenos Aires	Tocache	-8.40	-76.45	Jul-2015	Pure culture	Forastero	MLL_5	kt	IX	A2B2
JD_SM-24	Peru	Uchiza	Tocache	-8.44	-76.46	Jul-2015	Pure culture	Forastero	MLL_5	ku	IY	A2B2
JD_SM-25	Peru	Union Cadena	Tocache	-8.35	-76.40	Jul-2015	Pure culture	Forastero	MLL_5	kv	IZ	A2B2
JD_SM-26	Peru	Santa Lucia	Tocache	-8.37	-76.32	Jul-2015	Pure culture	Trinitario	MLL_5	kw	JA	A2B2
JD_UC-28	Peru	Curimaná	Padre Abad	-8.44	-75.16	Aug-2015	Pure culture	Forastero	MLL_5	ky	JC	A2B2
JD_UC-29	Peru	Malvinas	Padre Abad	-8.41	-75.10	Aug-2015	Pure culture	Forastero	MLL_5	kz	JD	A2B2
JD_UC-30	Peru	Nueva Meriba	Padre Abad	-8.49	-75.11	Aug-2015	Pure culture	Trinitario	MLL_5	la	JE	A2B2
JD_UC-31	Peru	Nuevo Huánuco	Padre Abad	-8.89	-75.21	Aug-2015	Pure culture	Trinitario	MLL_5	lb	JF	A2B2
JD_UC-32	Peru	Los Vencedores	Padre Abad	-8.91	-75.20	Aug-2015	Pure culture	Forastero	MLL_5	lc	JG	A2B2
JD_UC-33	Peru	Villa El Salvador	Padre Abad	-8.85	-75.17	Aug-2015	Pure culture	Trinitario	MLL_5	ld	JH	A2B2

GPS data based on the WGS1984 UTM coordinate system
Host: If host is *T. cacao*, then the genotype/cultivar or "unknown" is specified
MLL or MLG: Multilocus lineage or Multilocus genotype of sample as determined in Chapter 2.
Type of Sample: Whether the internal necrotic tissue or white stroma of the infected cacao pod was collected in Nuclei Lysis solution or if it was isolated in Pure culture (See Chapter 3 for details)
Pod: Samples with same lowercase letter(s) come from the same pod
Tree: Samples with same uppercase letter(s) come from the same tree.
Mating type: Results determined in Chapter 3.

APPENDIX B

Various data described are available online, in file “APPENDICES B AND F.zip”:

1. For Chapter 2:
    a. Identification of MLLs in the SSR and SNP datasets.
    b. Summary of null alleles for SSR.
2. For Chapter 3:
    a. rDNA alignments of samples.
3. For Chapter 4:
    a. MAKER annotation of *Moniliophthora roreri* MCA2952 genome.
    b. Outputs from program runs to predict effectors.
    c. HMMER output to identify CAZy families.
    d. Alignment of the two-protein dataset to build phylogenetic tee of species analyzed.
    e. Inputs and outputs of CAFE analysis on effectoromes and CAZyomes.
    f. Protein IDs of effectors with a fungal-type cell wall GO from all the genomes analyzed.
    g. Protein IDs of effectors with a flavin adenine dinucleotide binding ontology from all the genomes analyzed.
    h. Protein IDs of effectors with a transition metal ion ontology from all the genomes analyzed.
    i. SyLOCAL output files of all the runs performed.
    j. DESeq2 results of 11,183 ortholog genes of MCA2952 and MCA2997 genomes

APPENDIX C

Figure C.1 Spectrum of genetic diversity (SGD). D and *p* are the Hartigan's Dip value of unimodality and its probability, respectively. Red-dashed line highlights points to the threshold value of 0.13.

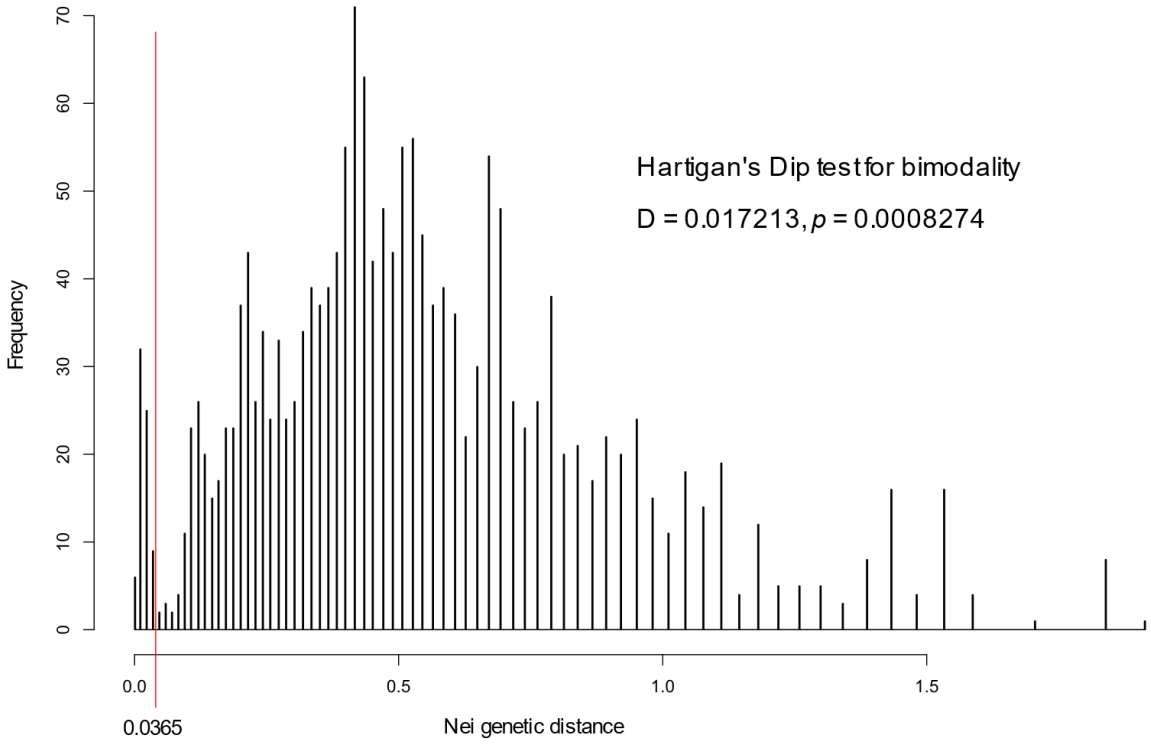
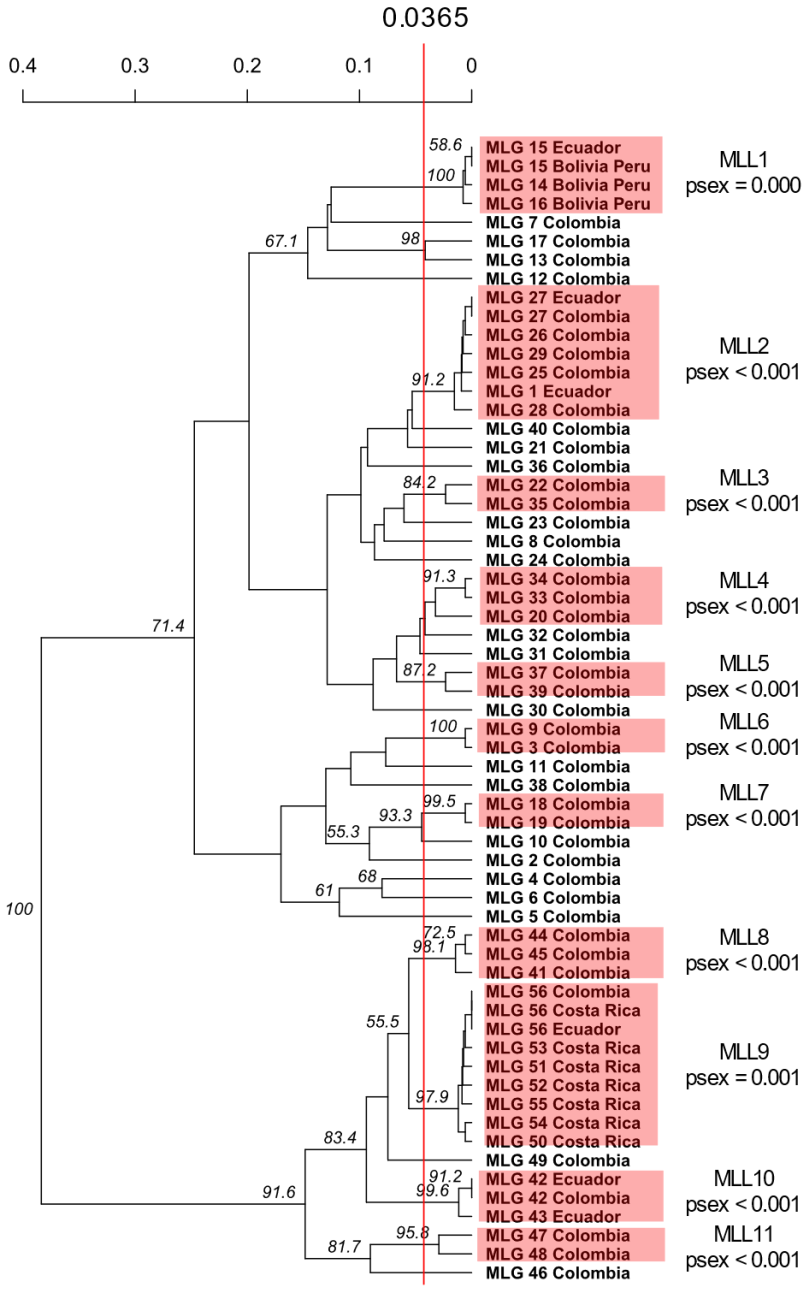


Figure C.2 Nei-distance UPGMA dendrogram of the thirty-three MLGs encountered in the SNP dataset. Red boxes indicate the clusters of samples below the threshold value of 0.0365 (red line) examined with p_{sex} (all were significant thus all were assigned to MLLs and their alleles were adjusted; see Materials and Methods). Tips are labeled with the MLG or MLL assigned for the SNP dataset and sample's origin (APPENDIX B). Bootstrap values greater than 50 are shown.



APPENDIX D

Figure D.1 Multiple sequence alignment of homeodomain transcription factors HD1 from *Moniliophthora roreri*. Black box represents the homeobox conserved domain. Numbers are the positions of amino acids in the alignment. Color of amino acids are based on their biochemical properties according to Mega X (Kumar et al. 2018)

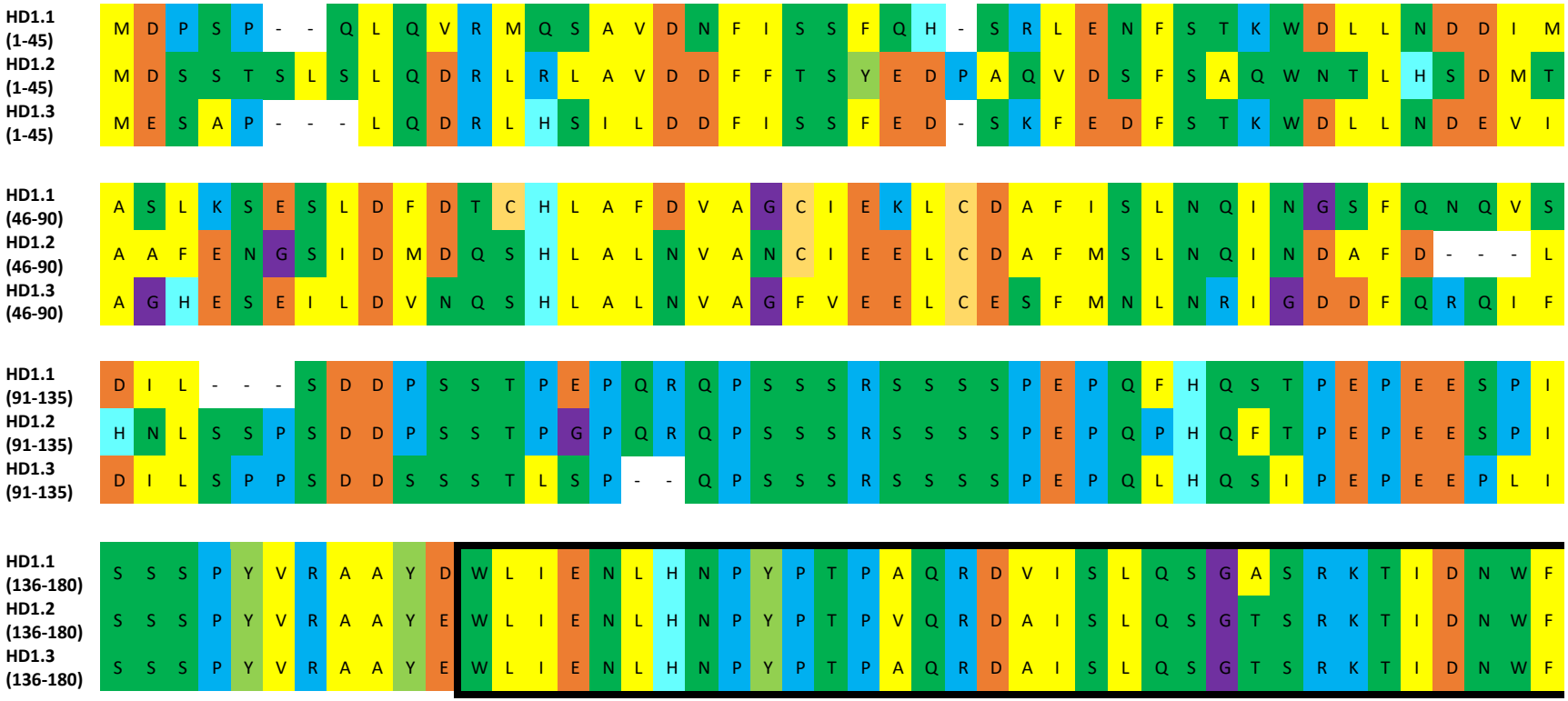


Figure D.1 Cont.

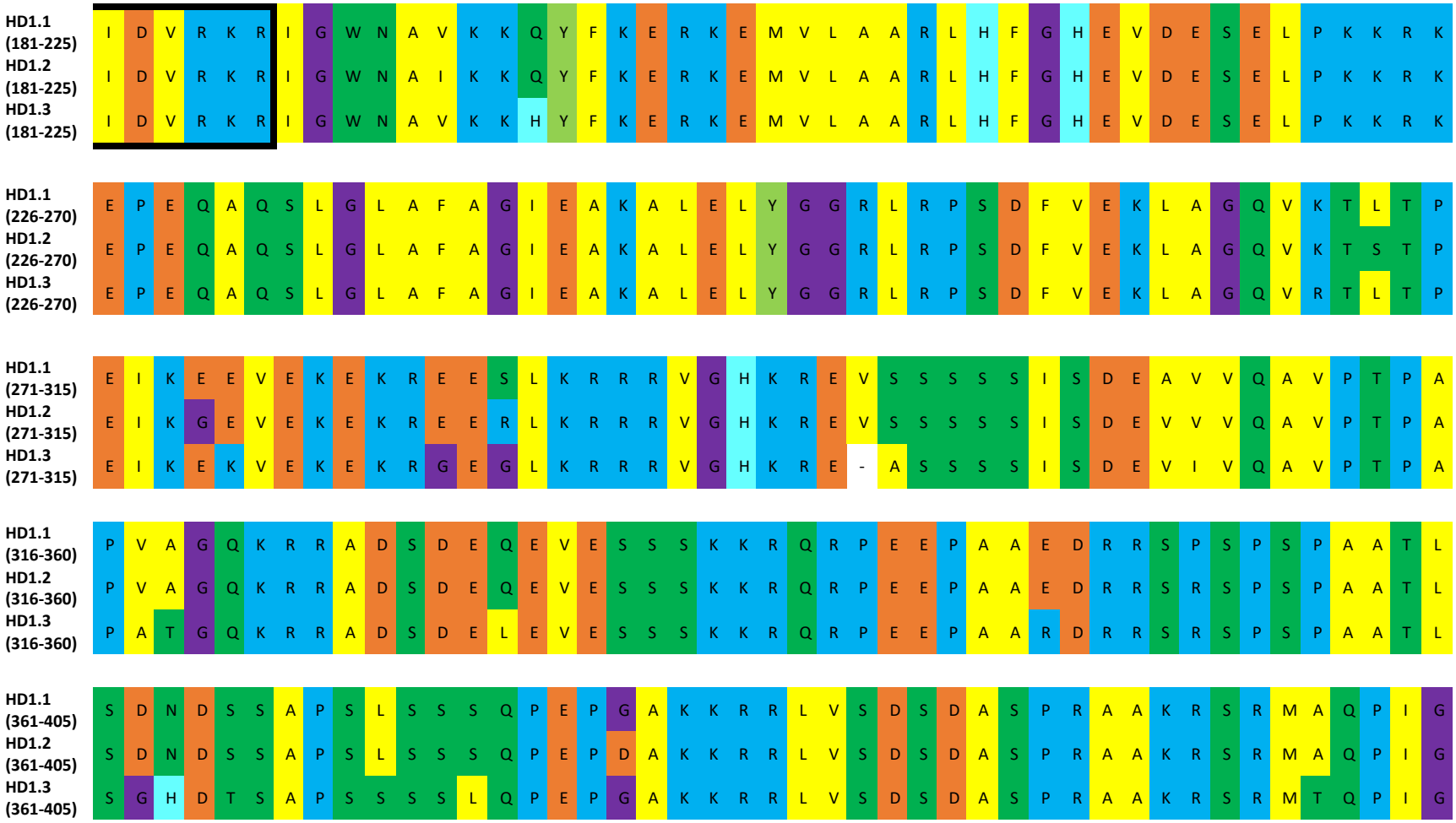


Figure D.1 Cont.

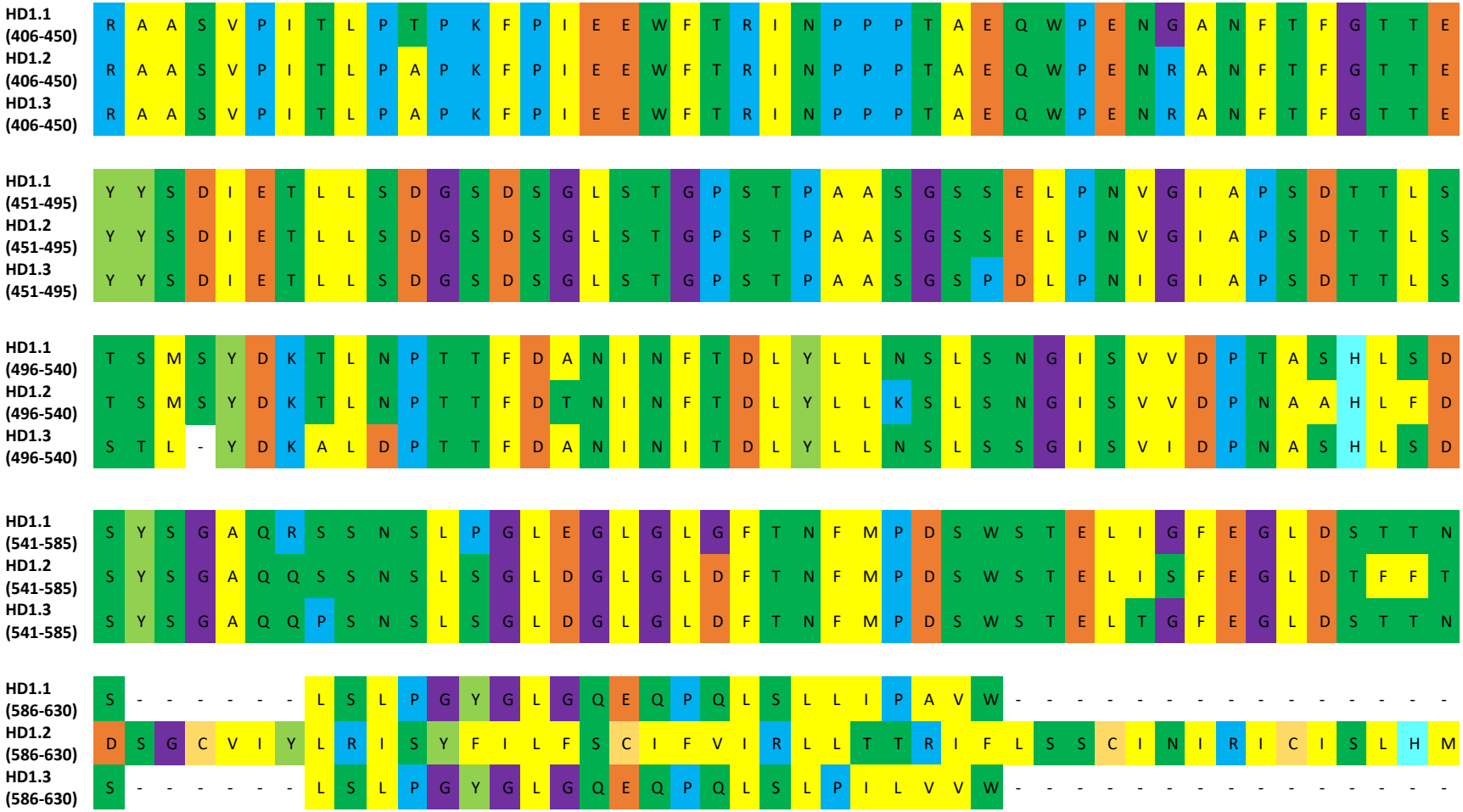


Figure D.1 Cont.

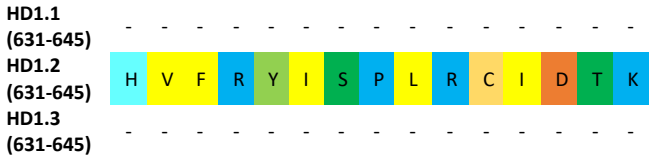


Figure D.2 Multiple sequence alignment of homeodomain transcription factors HD2 from *Moniliophthora roreri*. Black box represents the homeobox conserved domain. Numbers are the positions of amino acids in the alignment. Color of amino acids are based on their biochemical properties according to Mega X (Kumar et al. 2018)

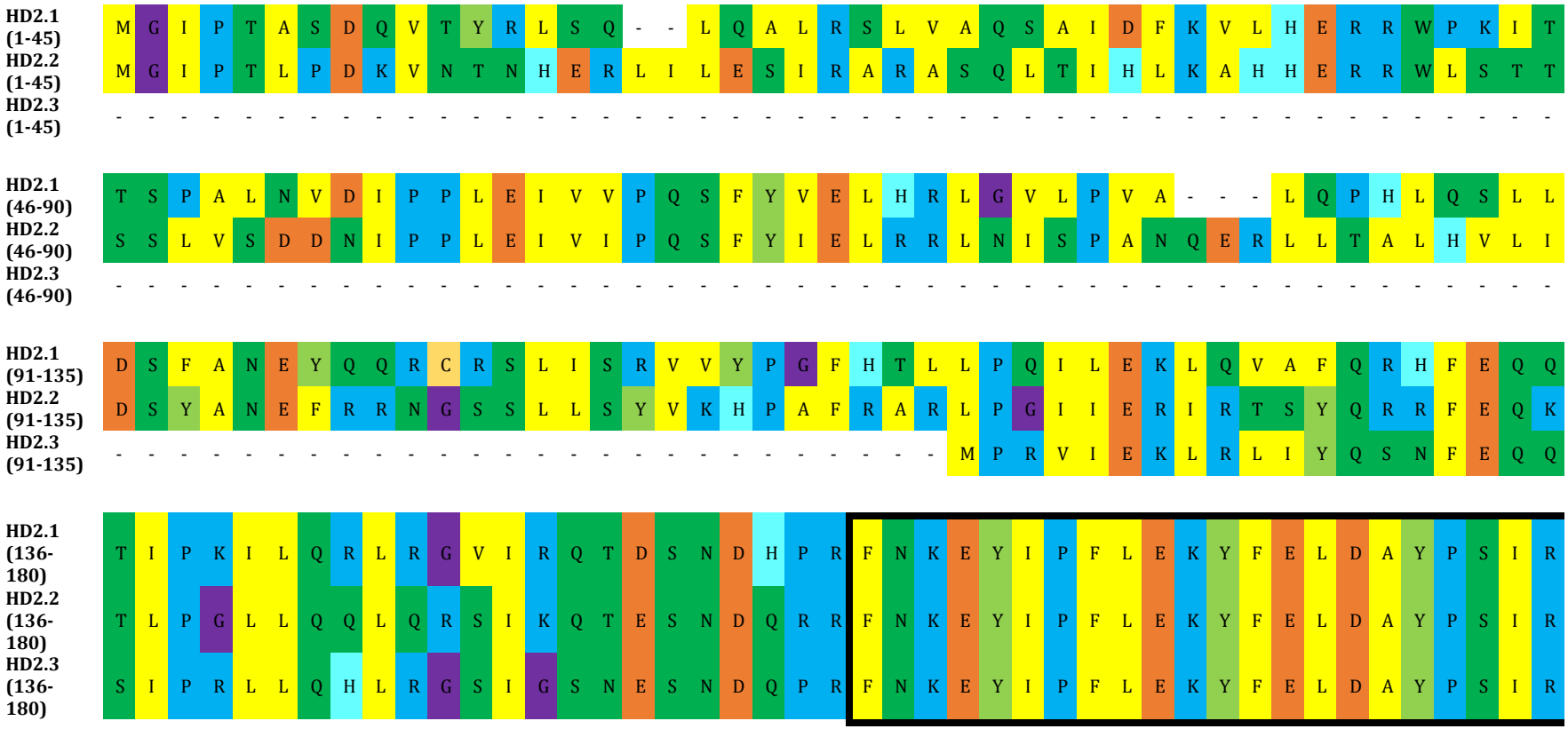


Figure D.2 Cont.

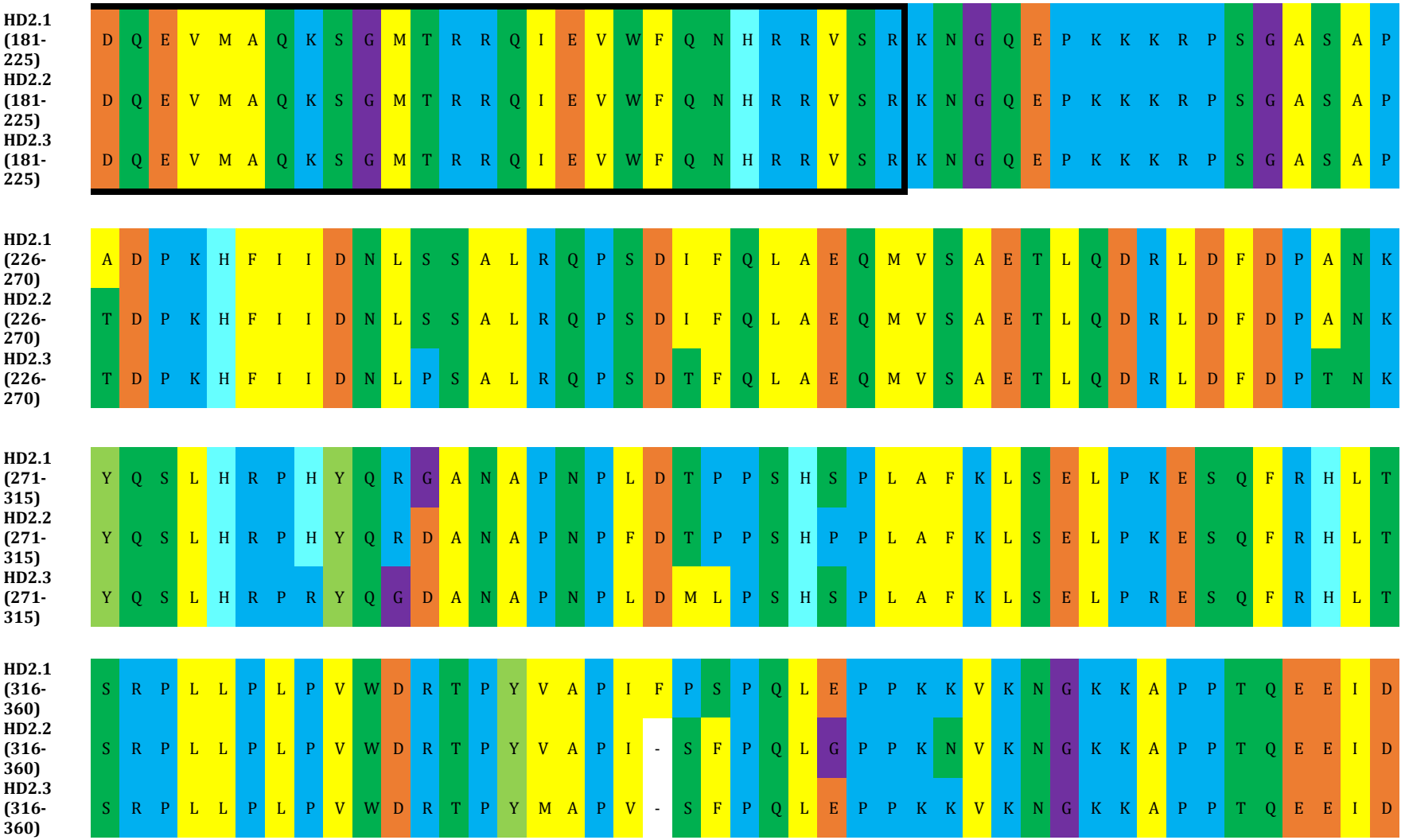


Figure D.2 Cont.

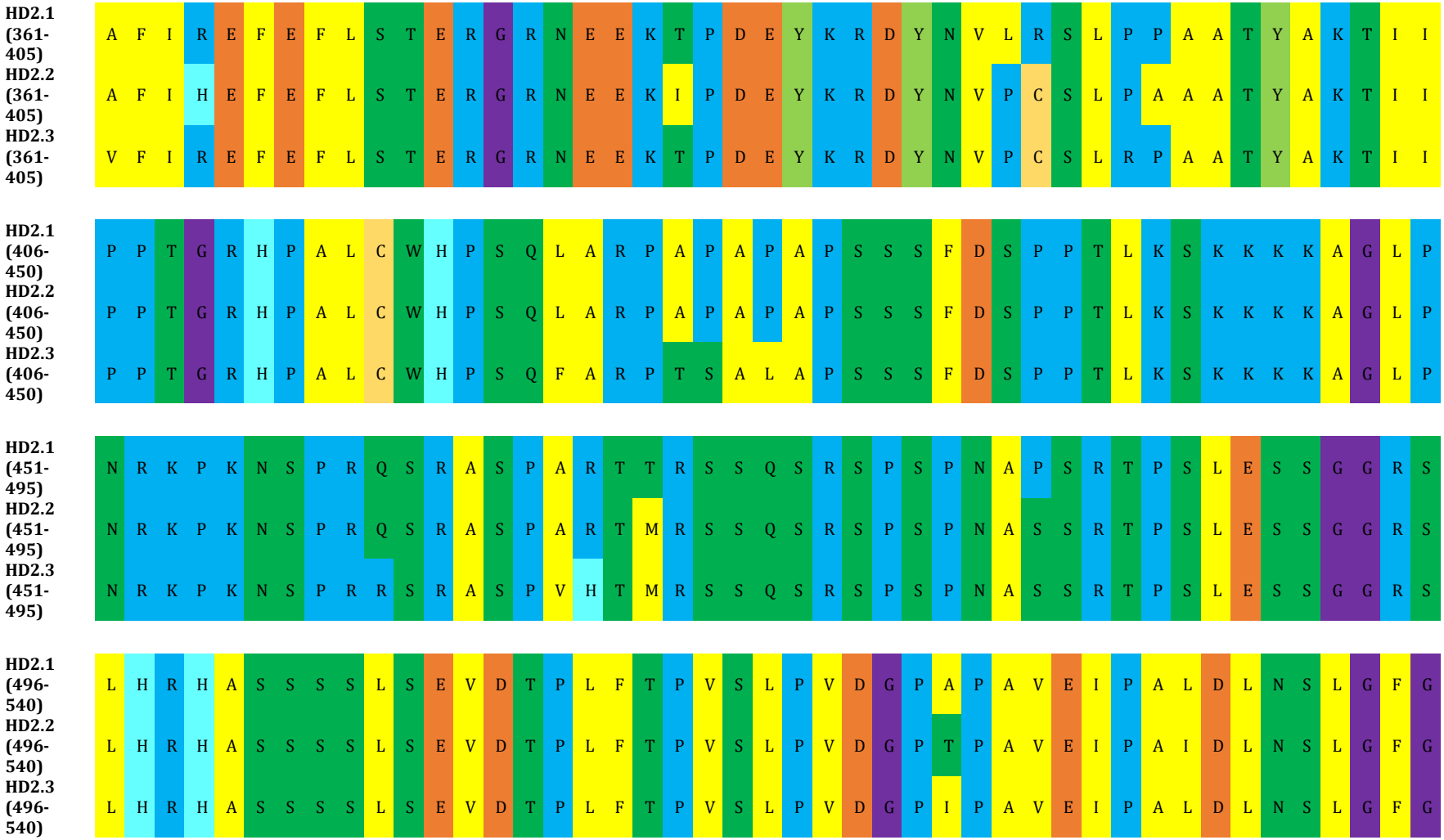
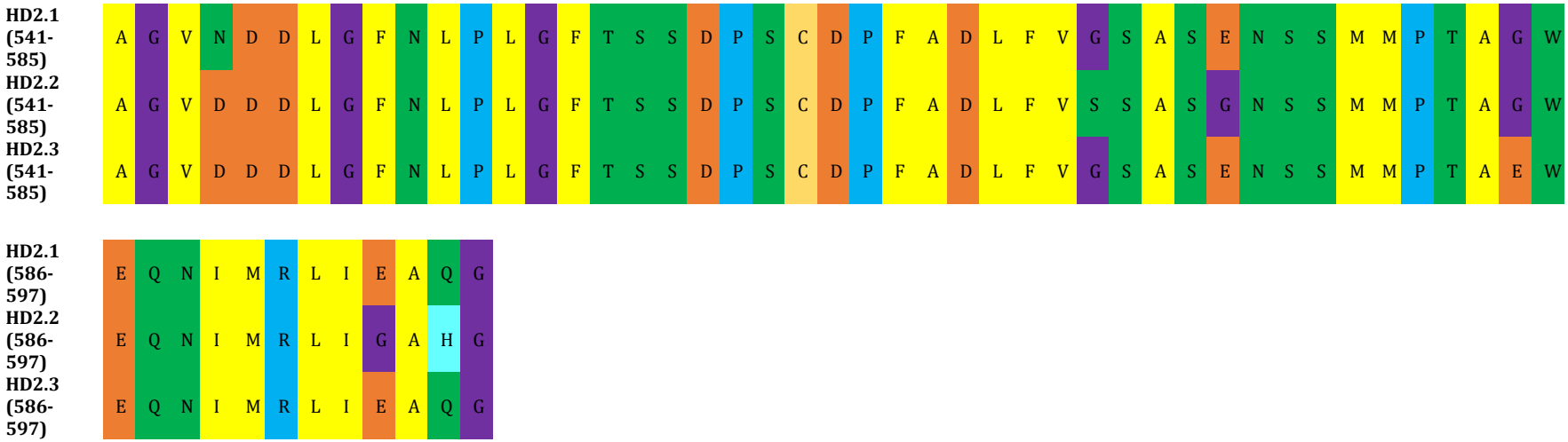
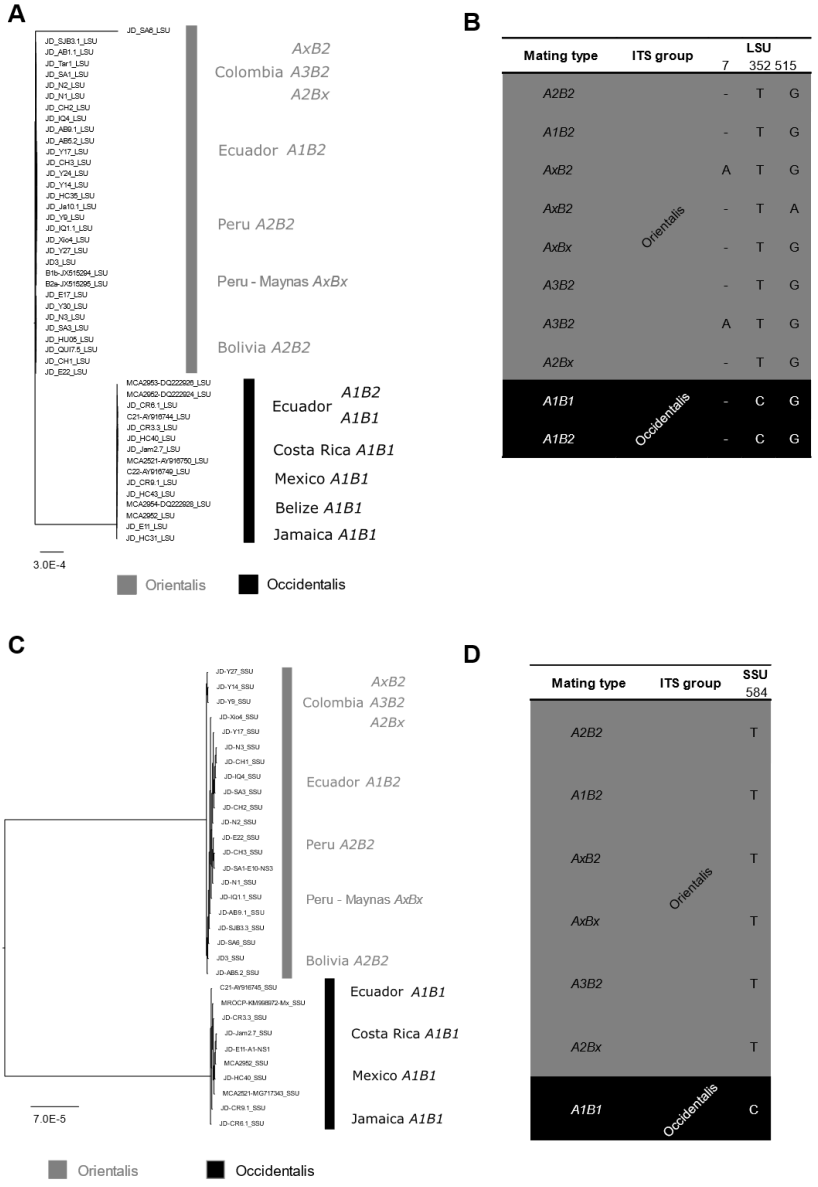


Figure D.2 Cont.



APPENDIX E

Figure E.1 LSU and SSU sequence analysis of *Moniliophthora roreri* samples. **A)** and **C)** Maximum likelihood phylogenetic tree using LSU and SSU sequences, respectively, of samples from all the countries and mating types found in this study, except for *AxB1* from the ITS group Occidentalis for both LSU and SSU, and *A1B2* for SSU analysis. **B)** and **D)** SNP comparison among mating types and the ITS groups Orientalis and Occidentalis; numbers indicate the relative position of SNPs in the alignment used for LSU and SSU phylogenies (APPENDIX B).



APPENDIX F

Program SyLOCAL and other accessory files are available online, in file "APPENDICES B AND F.zip":

1. The core Perl script of the program, SyLOCAL_v3.pl.
2. The Python extension run_multiple_SyLOCAL.py to run multiple genomes at the time.
3. A README file (Please read if you want to use this program).
4. Example of input files:
    a. Query: *Fasta* file containing the CDS of a portion of scaffold dna.fa_109 of assembly of *M. roreri* MCA2952, which includes the hydrophobins described in Chapter 4
    b. Subject (as available in Mycocosm; the purpose of including these files is just to show the functionality of SyLOCAL):
        i. *GFF* files of the filtered gene catalogues of *Coprinopsis cinerea* and *Fistulina hepatica* (Table 4.5).
        ii. *Fasta* files containing the CDS of all filtered gene models of *C. cinerea* and *F. hepatica.*

APPENDIX G

Table G.1 Information and downstream analyses of the predicted effector proteins with amino acid repeats and nuclear localization signal (NLS) in the thirteen Agaricales genomes analyzed

G	Prot. ID	Annotation	T-REKS	Hit	QC	E	Pid	GBN	EffectorP 1.0 Y or N	EffectorP 1.0 *p*	EffectorP 2.0 Y, N or U	EffectorP 2.0 *p*
MCA2952	maker-dna.fa_21-augustus-gene-7.52-mRNA-1	NA	32 to 57 - Psim:0.7 KRG-LKQLKL KRGDPNPL-L KRDDVKPK--	hypothetical protein LENED_004165 [Lentinula edodes]	98%	3.00 E-12	57.95 %	GAW02505.1	Y	0.98	N	0.61
	maker-dna.fa_22-augustus-gene-1.70-mRNA-1	NA	38 to 56 - Psim:0.75 NAGA-ANPAN NAGSGAGAAG	hypothetical protein Hypma_015100 [Hypsizygus marmoreus]	80%	5.00 E-93	62.82 %	RDB29153.1	N	1	N	0.84
	maker-dna.fa_65-augustus-gene-0.1-mRNA-1	NA	32 to 45 - Psim:0.79 KKSKGDP KKGKGGD					NA				
	maker-dna.fa_65-augustus-gene-0.10-mRNA-1	IPR000772 Ricin B, lectin domain	32 to 45 - Psim:0.79 KKSKGDP KKGKGGD					NA				
MCA2997	ESK81151.1	NA	164 to 177 - Psim:0.86 GKKGSKH GKKGGKD ********** 67 to 84 - Psim:0.67 KHCTKEEYN KHMKDHEES					NA				
	ESK82669.1	NA	32 to 57 - Psim:0.7 KRG-LKQLKL KRGDPNPL-L KRDDVKPK--	Extracellular membrane protein, CFEM domain protein [*Ascosphaera apis* ARSEF 7405]	29%	0.5	43.33 %	KZZ89857.1	N	1	N	0.99
	ESK88269.1	NA	32 to 57 - Psim:0.7 KRG-LKQLKL KRGDPNPL-L KRDDVKPK--	hypothetical protein LENED_004165 [*Lentinula edodes*]	98%	3.00 E-12	57.95 %	GAW02505.1	Y	0.98	N	0.61
	ESK88962.1	NA	38 to 56 - Psim:0.75 NAGA-ANPAN NAGSGAGAAG	hypothetical protein Hypma_015100 [*Hypsizygus marmoreus*]	80%	6.00 E-93	62.39 %	RDB29153.1	N	1	N	0.84

Table G.1 Cont.

G	Prot. ID	Annotation	T-REKS	Hit	QC	E	Pid	GBN	EffectorP 1.0 Y or N	EffectorP 1.0 *p*	EffectorP 2.0 Y, N or U	EffectorP 2.0 *p*
	ESK89252.1	IPR000772 Ricin B, lectin domain	32 to 45 - Psim:0.71 KKSKVDP KKGKGGD	putative acetyl-hydrolase [*Clostridium* sp. CAG:590]	45%	9.5	30.77 %	CCX85662.1	N	1	N	0.65
MCA2997	ESK97420.1	PF11790 Glycosyl hydrolase catalytic core IPR017853 Glycoside hydrolase superfamily IPR024655 Uncharacterized protein family, glycosyl hydrolase catalytic domain	136 to 149 - Psim:1.0 N(14) 123 to 136 - Psim:0.93 TN TN AN TN TN TN TN 88 to 113 - Psim:0.71 NNN-DNG NNG-DNG NNNSNNN NNEGNND	glycoside hydrolase [*Gloeophyllum trabeum* ATCC 11539]	78%	1.00 E-109	57.84 %	XP_007862538.1	N	1	N	0.84
MpFA553	MPER_13271	NA	159 to 174 - Psim:0.75 KKG- KKV- KKVR KKE- EKH-					NA				
Marfi	22262	IPR038955 Protein PriA	123 to 138 - Psim:0.75 PSSH PSGV PSSV PSGY	hypothetical protein WG66_4305 [*Moniliophthora roreri*]	100 %	3.00 E-83	56.60 %	KTB43134.1	Y	0.99	Y	0.83
Arme	2872	NA	29 to 48 - Psim:0.67 KGEH-- KGAKHV KGNHH- KGRDN-	hypothetical protein [*Terribacillus halophilus*]	26%	2.6	39.22 %	WP_093727630.1	Y	0.74	Y	0.55
	9947	NA	729 to 48 - Psim:0.67 KGEH-- KGAKHV KGNHH- KGRDN-					NA				

Table G.1 Cont.

G	Prot. ID	Annotation	T-REKS	Hit	QC	E	Pid	GBN	EffectorP 1.0 Y or N	EffectorP 1.0 *p*	EffectorP 2.0 Y, N or U	EffectorP 2.0 *p*
Arme	9973	NA	731 to 45 - Psim:0.68 7 -KGEH -KG-E HK--- VKH-- VKG-- 104 to 243 - Psim:0.69 -TWEPVFENKTXXXXKQCPSEVHKDCPCL- KDSECGFKCPQQWPV-TNC -TWEPVFENKKGWKDWKSGKYTPITVGYVNKDT- FELDCKNLCEAHEKC YSCQ-AFSXXXXWKDWKSGKYTPITVGYVNKDT- FELDCKNLCEAHEKC	hypothetical protein BBJ28_00015972 [*Nothophytophthora* sp. Chile5]	29%	0.59	29.47 %	RLN77817.1	Y	1	N	0.99
Mycale	328926	NA	133 to 147 - Psim:0.65 -GKK -GKD -GQ- NGKK NGD-					NA				
Ompol	1444	NA	224 to 243 - Psim:0.65 SGAVS SGDEL SGGET SGTEG	hypothetical protein ARMSODRAFT_949800 [*Armillaria solidipes*]	100 %	1.00 E-104	59.54 %	PBK75537.1	N	0.52	U	0.53
Copci	2390	NA	95 to 123 - Psim:0.75 KP-TT--- KPSTT--- KP-AA--Q -P-TTPAK KP-DN--- KP-KT--R ********** 141 to 161 DFDIFE-R DFDI-EGR DFE-FELE	hypothetical protein FPSE_01498 [*Fusarium pseudograminearum* CS3096]	33%	0.22	46.15 %	XP_009252893.1	N	1	N	0.97

Table G.1 Cont

G	Prot. ID	Annotation	T-REKS	Hit	QC	E	Pid	GBN	EffectorP 1.0 Y or N	EffectorP 1.0 *p*	EffectorP 2.0 Y, N or U	EffectorP 2.0 *p*
Fishe	56629	NA	94 to 110 - Psim:0.75 NNQGQ NNNS- NNNN- NNQN- ********** 77 to 91 - Psim:0.67 KAEDKH KA--VE KA-GKQ					NA				
	62016	NA	74 to 93 - Psim:0.77 KQNKSPAQPPQ KQG-TPAQP-H	flagellar motor protein MotB [*Nitratireductor pacificus*]	30%	0.2	35.19 %	WP_008597843.1	N	1	N	0.99
Denbi	922264	NA	69 to 91 - Psim:0.67 EEGR-KAALA EEGR-RK--- EEEEERI--K					NA				
Cylto	246140	NA	104 to 124 - Psim:0.74 AKK-AEE-- AKKLAER-A AKA-TEAV-	hypothetical protein DICSQDRAFT_170686 [*Dichomitus squalens* LYAD-421 SS1]	47%	0.031	30.56 %	XP_007366361.1	N	1	N	0.99
	423322	NA	224 to 238 - Psim:0.73 SGDEL SGGET SGTEG	hypothetical protein ARMSODRAFT_949800 [*Armillaria solidipes*]	99%	1.00E-116	65.86 %	PBK75537.1	N	0.52	U	0.53

G = Genome

Prot ID = Protein ID of predicted effector

Annotation = InterProScan Annotation (IPR/PF/GO)

T-REKS = Repeats found by T-REKS; Psim = Percentage of similarity; numbers are the positions in the protein where the repeat starts and ends.

Hit = Top non-generic *blastp* hit

QC = Query coverage

E = E value

Pid = Percent identity

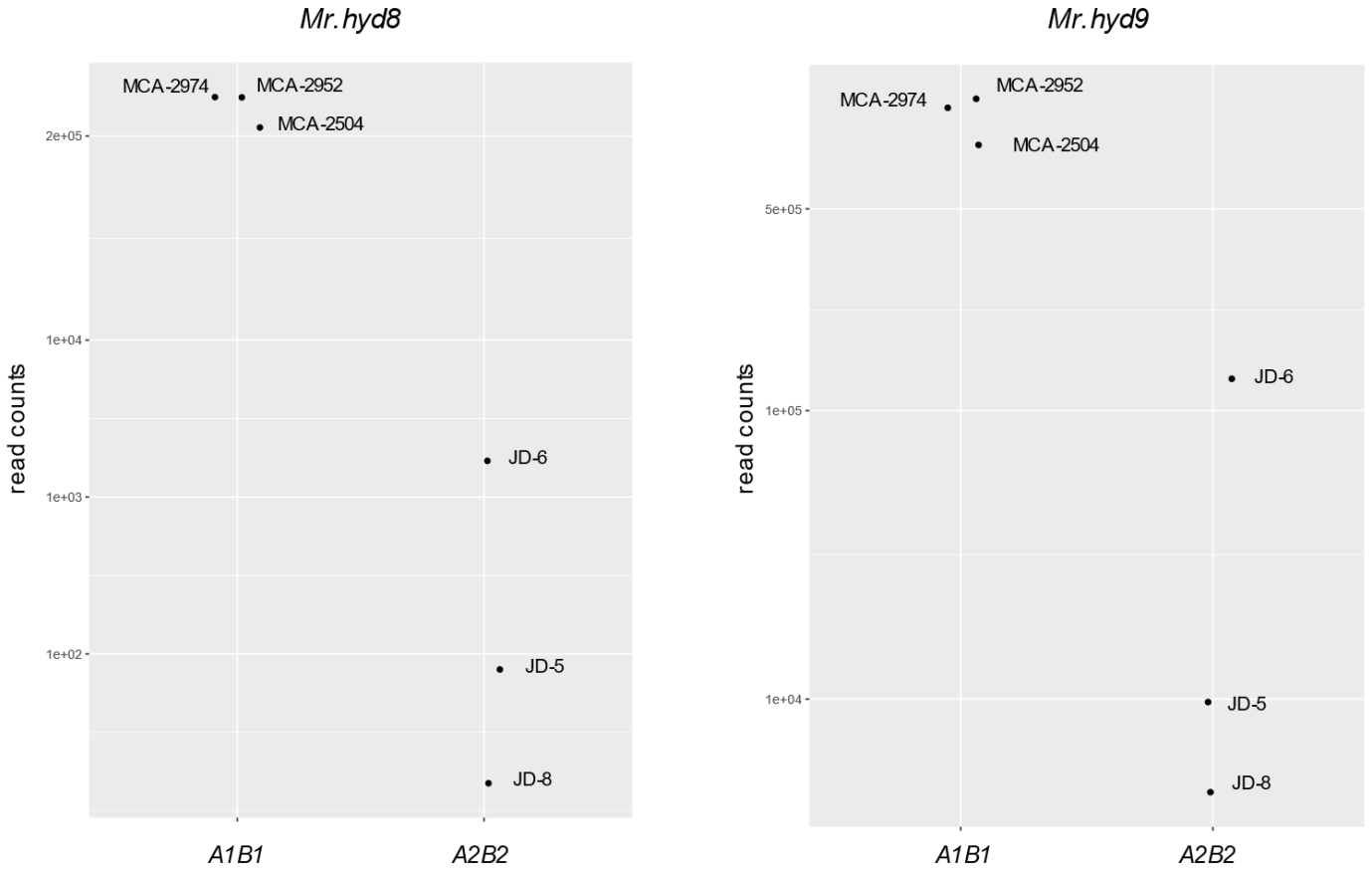
GBN = GenBank accession number

EffectorP 1.0 = EffectorP 1.0 analysis of *blastp* match; Y = match is an effector; N = match is not an effector; *p* = probability

Effector P 2.0 = EffectorP 2.0 analysis of *blastp* match; Y = match is an effector; N = match is not an effector; U = match unlikely to be an effector; *p* = probability

APPENDIX H

Figure H.1 Plots of the read counts for genes *Mr.hyd8* and *Mr.hyd9* of *Moniliophthora roreri* samples (Table 4.3) for which *DESeq2* did not calculate a *padj* value because of outliers detected by Cook's distance.



REFERENCES

Adedeji AR. 2006. Thread blight disease of tea [*Camellia sinensis* (L.) O. Kuntze] caused by *Marasmius pulcher* (Berk & Br.) Petch in the South Western Nigeria. African Sci. 7:107–112.

Afoakwa EO, Paterson A, Fowler M, Ryan A. 2008. Flavor Formation and Character in Cocoa and Chocolate: A Critical Review. Crit Rev Food Sci Nutr. 48:840–857, doi:10.1080/10408390701719272.

Aikpokpodion PO, Motamayor JC, Adetimirin VO, Adu-Ampomah Y, Ingelbrecht I, Eskes AB, Schnell RJ, Kolesnikova-Allen M. 2009. Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. Tree Genet Genomes. 5:699–711, doi:10.1007/s11295-009-0221-1.

Aime MC, 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, doi:10.3852/mycologia.97.5.1012.

Aken M van. 1981. The Lingering Death of Indian Tribute in Ecuador. Hisp Am Hist Rev. 61:429–459, doi:10.2307/2513393.

Akrofi AY, Amoako-Atta I, Acheampong K, Assuah MK, Melnick RL. 2016. Fruit and Canopy Pathogens of Unknown Potential Risk. In: B.A. Bailey, and L.W. Meinhardt, eds. Cacao Diseases: A History of Old Enemies and New Encounters. Springer. pp. 361–382.

Alden D. 1976. The Significance of Cacao Production in the Amazon Region during the Late Colonial Period: An Essay in Comparative Economic History. Proc Am Philos Soc. 120:103–135.

Ali SS, Shao J, Strem MD, Phillips-Mora W, Zhang D, Meinhardt LW, Bailey B a. 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. Front Microbiol. 6:850, doi:10.3389/fmicb.2015.00850.

Allen JA. 1987. Recolecciones de cacao silvestre de la region amazónica ecuatoriana.- Proyecto Amazonas, convenio Iniap-CCCA reporte final. Comun Técnica - INIAP, Ecuador. 15:119.

Allen JB. 1988. Geographical variation and population biology ind wild Theobroma cacao. University of Edinburgh.

Almaliky BSA, Abidin MZ, Kader J, Wong MY. 2013. First Report of *Marasmiellus palmivorus* Causing Post-Emergence Damping Off on Coconut Seedlings in Malaysia. Plant Dis. 97:143.

An S-Y, Asahara M, Goto K, Kasai H, Yokota A. 2007. *Terribacillus saccharophilus* gen. nov., sp. nov. and *Terribacillus halophilus* sp. nov., spore-forming bacteria isolated from field soil in Japan. Int J Syst Evol Microbiol. 57:51–55, doi:10.1099/ijs.0.64340-0.

Ancízar M. 1853. Peregrinación de Alpha: por las provincias del norte de Nueva Granada, en 1850 i 51. Bogotá, Colombia: Bogotá: Echeverria Hermanos. 524 p.

Anders S, Pyl PT, Huber W. 2015. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics. 31:166–169, doi:10.1093/bioinformatics/btu638.

Antonín V. 2013. Supplements to the Monograph of Tropical African Species of *Marasmius* (Basidiomycota, Marasmiaceae). Cryptogam Mycol. 34:113–135, doi:10.7872/crym.v34.iss2.2013.113.

Antonín V, Ryoo R, Ka K-H, Sou H-D. 2014. Three new species of *Crinipellis* and one new variety of *Moniliophthora* (Basidiomycota, Marasmiaceae) described from the Republic of Korea. Phytotaxa. 170:86–102.

Arévalo E, Hernández TA. 1990. La moniliasis, una nueva enfermedad del cacao en el Perú. Fitopatol. 25:4.

Arnaud-Haond S, Duarte CM, Alberto F, Serrão EA. 2007a. Standardizing methods to address clonality in population studies. Mol Ecol. 16:5115–5139, doi:10.1111/j.1365-294X.2007.03535.x.

Arnaud-Haond S, Migliaccio M, Diaz-Almela E, Teixeira S, Vliet MS Van De, Alberto F, Procaccini G, Duarte CM, Serrão EA. 2007b. Vicariance patterns in the Mediterranean Sea: East-west cleavage and low dispersal in the endemic seagrass Posidonia oceanica. J Biogeogr. 34:963–976, doi:10.1111/j.1365-2699.2006.01671.x.

Arroyo Abad L. 2013. Inestabilidad, costo de vida y salarios reales en Venezuela en el Siglo XIX. América Lat En La Hist Económica. 20:114–137.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. Nat Genet. 25:25–29, doi:10.1038/75556.

Avelino J, Cabut S, Barboza B, Barquero M, Alfaro R, Esquivel C, Durand J-F, Cilas C. 2007. Topography and Crop Management Are Key Factors for the Development of American Leaf Spot Epidemics on Coffee in Costa Rica. Phytopathology. 97:1532–1542, doi:10.1094/phyto-97-12-1532.

Ba ANN, Pogoutse A, Provart N, Moses AM. 2009. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. BMC Bioinformatics. 10:202, doi:10.1186/1471-2105-10-202.

Badrane H, May G. 1999. The Divergence-Homogenization Duality in the Evolution of the b1 Mating Type Gene of Coprinus cinereus. Mol Biol Evol. 16:975–986, doi:10.1093/oxfordjournals.molbev.a026187.

Bailey BA, Ali SS, Strem MD, Meinhardt LW. 2018a. Morphological variants of *Moniliophthora roreri* on artificial media and the biotroph/necrotroph shift. Fungal Biol. 122:701–716, doi:10.1016/j.funbio.2018.03.003.

Bailey BA, Crozier J, Sicher RC, Strem MD, Melnick R, Carazzolle MF, Costa GGL, Pereira GAG, Zhang D, Maximova S, Guiltinan M, Meinhardt L. 2013. Dynamic changes in pod and fungal physiology associated with the shift from biotrophy to necrotrophy during the infection of *Theobroma cacao* by *Moniliophthora roreri*. Physiol Mol Plant Pathol. 81:84–96, doi:10.1016/j.pmpp.2012.11.005.

Bailey BA, Evans HC, Phillips-Mora W, Ali SS, Meinhardt LW. 2018b. <i>Moniliophthora roreri<i>, causal agent of cacao frosty pod rot. Mol Plant Pathol. 19:1580–1594, doi:10.1111/mpp.12648.

Baird RE, Wilson JP, Sumner DR. 1992. Identity and Pathogenicity of Two *Marasmius* Species from the Sterile White Basidiomycete Complex. Plant Dis. 76:244–247.

Baker RED, Cope FW, Holliday PC, Bartley BG, Taylor DJ. 1954. The Anglo-Colombian cacao collecting expedition. In: Annual Report Cacao Research Imperial College Tropical Agriculture 1954. St. Augustine, Trinidad.: The Imperial College of Agriculture. pp. 8–29.

Bandala VM, Ryoo R, Montoya L, Ka K-H. 2012. New species and new records of *Crinipellis* from tropical and subtropical forests of the east coast of Mexico. Mycologia. 104:733–745, doi:10.3852/11-223.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 19:455–477, doi:10.1089/cmb.2012.0021.

Barau J, Grandis A, Carvalho VMDA, Teixeira GS, Zaparoli GHA, Rio MCS Do, Rincones J, Buckeridge MS, Pereira GAG. 2015. Apoplastic and intracellular plant sugars regulate developmental transitions in witches' broom disease of cacao. J Exp Bot. 66:1325–1337, doi:10.1093/jxb/eru485.

Barbosa CS, Fonseca RR da, Batista TM, Barreto MA, Argolo CS, Carvalho MR de, Amaral DOJ do, Silva EM de A, Arévalo-Gardini E, Solis Hidalgo K, Franco GR, Pirovani CP, Micheli F, Gramacho KP. 2018. Genome sequence and effectorome of *Moniliophthora perniciosa* and *Moniliophthora roreri* subpopulations. BMC Genomics. 19:509, doi:10.1186/s12864-018-4875-7.

Barham BL, Coomes OT. 1994. Reinterpreting the Amazon Rubber Boom: Investment, the State, and Dutch Disease. Lat Am Res Rev. 29:73–109.

Barrau J. 1979. Sur l'origine du cacaoyer, *Theobroma cacao* Linné, Sterculiacées. J d'agriculture Tradit Bot Appliquée. 26:171–180.

Bartley BGD. 2005. The genetic diversity of cacao and its utilization. London: CABI. 341 p.

Beg MS, Ahmad S, Jan K, Bashir K. 2017. Status, supply chain and processing of cocoa - A review. Trends Food Sci Technol. 66:108–116, doi:10.1016/j.tifs.2017.06.007.

Bekele FL. 2004. The History of Cocoa Production in Trinidad and Tobago. In: Proceedings of the APASTT Seminar – Exhibition Entitled Re-Vitalisation of the Trinidad & Tobago Cocoa Industry. St. Augustine, Trinidad. pp. 4–12.

Bennett AB. 2003. Out of the Amazon: *Theobroma cacao* enters the genomic era. Trends Plant Sci. 8:561–563.

Bentley JW, Boa E, Stonehouse J. 2004. Neighbor trees: shade, intercropping, and cacao in Ecuador. Hum Ecol. 32:241–270, doi:10.1023/B:HUEC.0000019759.46526.4d.

Bergmann JF. 1969. The distribution of cacao cultivation in pre-columbian America. Ann Assoc Am Geogr. 59:85–96, doi:10.1111/j.1467-8306.1969.tb00659.x.

Bie T De, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. Bioinformatics. 22:1269–1271, doi:10.1093/bioinformatics/btl097.

Binder M, Hibbett DS, Wang Z, Farnham WF. 2006. Evolutionary relationships of *Mycaureola dilseae* (Agaricales), a basidiomycete pathogen of a subtidal rhodophyte. Am J Bot. 93:547–556, doi:10.3732/ajb.93.4.547.

Bletter N, Daly DC. 2006. Cacao and Its Relatives in South America An Overview of Taxonomy, Ecology, Biogeography, Chemistry, and Ethnobotany. In: Chocolate in Mesoamerica: A Cultural History of Cacao. Gainesville: University Press of Florida. pp. 29–68.

Boetzer M, Henkel C V, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 27:578–579, doi:10.1093/bioinformatics/btq683.

Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. Genome Biol. 13:R56, doi:https://doi.org/10.1186/gb-2012-13-6-r56.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120, doi:10.1093/bioinformatics/btu170.

Bolton MD, Esse HP van, Vossen JH, Jonge R de, Stergiopoulos I, Stulemeijer IJE, Berg GCM van den, Borrás-Hidalgo O, Dekker HL, Koster CG de, Wit PJGM de, Joosten MHAJ, Thomma BPHJ. 2008. The novel *Cladosporium fulvum* lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species. Mol Microbiol. 69:119–136, doi:10.1111/j.1365-2958.2008.06270.x.

Bonavia D. 1994. Arte e Historia del Perú Antiguo: Colección Enrico Poli Bianchi. Arequipa, Peru: Banco del Sur. 338 p.

Boza EJ, Motamayor JC, Amores FM, Cedeño-Amador S, Tondo CL, Livingstone DS, Schnell RJ, Gutiérrez OA. 2014. Genetic Characterization of the Cacao Cultivar CCN 51: Its Impact and Significance on Global Cacao Improvement and Production. J Am Soc Hortic Sci. 139:219–229, doi:10.21273/JASHS.139.2.219.

Briton-Jones HR. 1934. The diseases and curing of cacao. London: MacMillan. 161 p.

Brown AJ, Casselton LA. 2001. Mating in mushrooms: increasing the chances but prolonging the affair. Trends Genet. 17:393–400, doi:10.1016/S0168-9525(01)02343-5.

Brown JKM, Hovmøller MS. 2002. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. Science (80-). 297:537–541, doi:10.1126/science.1072678.

Bryant SK. 2006. Finding gold, forming slavery: the creation of a classic slave society, Popayán, 1600-1700. Americas (Engl Ed). 63:81–112, doi:10.1017/S0003161500062532.

Burg HA van den, Harrison SJ, Joosten MHAJ, Vervoort J, Wit PJGM de. 2007. *Cladosporium fulvum* Avr4 Protects Fungal Cell Walls Against Hydrolysis by Plant Chitinases Accumulating During Infection. Mol Plant-Microbe Interact. 19:1420–1430, doi:10.1094/mpmi-19-1420.

Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner (No. LBNL-7065E).

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421, doi:10.1186/1471-2105-10-421.

Campbell MS, Holt C, Moore B, Yandell M. 2014a. Genome Annotation and Curation Using MAKER and MAKER-P. 4-11 p.

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, Ware D, Shiu S-H, Childs KL, Sun Y, Jiang N, Yandell M. 2014b. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. Plant Physiol. 164:513–524, doi:10.1104/pp.113.230144.

Campuzano H. 1976. Fluctuación de poblaciones de esporas de *Monilia roreri* Cif. & Par. y viabilidad durante un ciclo completo de afección. Not Fitopatológicas. 5:107.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 37:D233–D238, doi:10.1093/nar/gkn663.

Cárdenas Vega ME. 2017. Creación de una Ruta del Cacao Ecuatoriano en Base a sus Propiedades Organolépticas que Atraviese las Provincias de Esmeraldas, Guayas, Manabí y El Oro. Universidad de las Américas.

Carmen D. 2005. Cultural Itinerary: The Route of the Cacao: Trade of the Cacao in Venezuela, Transformation of a Territory. In: 15th ICOMOS General Assembly and International Symposium: "Monuments and Sites in Their Setting - Conserving Cultural Heritage in Changing Townscapes and Landscapes." Xi'an, China. p.

Carranza Patiño MS, Motte E, Cedeño V, Cevallos Falquez OF, Saucedo Aguiar SG, Canchignia Martínez HF. 2008. Estudio de la diversidad genética de 20 accesiones de cacao (*Theobroma cacao* L.) mediante AP-PCR de la colección del Centro dl Cacao d Aroma Tenguel en la finca experimental La Buseta. Cienc y Tecnol. 1:1–5, doi:10.18779/cyt.v1i1.94.

Caso Barrera L, Aliphat Fernández M. 2006. Cacao, vanilla and annatto: three production and exchange systems in the Southern Maya lowlands, XVI-XVII centuries. J Lat Am Geogr. 5:29–52.

Chambouleyron R. 2014. Cacao, Bark-Clove and Agriculture in the Portuguese Amazon Region in the Seventeenth and Early Eighteenth Century. Luso-Braz Rev. 51:1–35, doi:10.1353/lbr.2014.0012.

Chapman JA, Ho I, Sunkara S, Luo S, Schroth GP, Rokhsar DS. 2011. Meraculous: *De Novo* Genome Assembly with Short Paired-End Reads. PLoS One. 6:e23501, doi:10.1371/journal.pone.0023501.

Chapman JA, Ho IY, Goltsman E, Rokhsar DS. 2016. Meraculous2: fast accurate short-read assembly of large polymorphic genomes. ArXiv Prepr ArXiv160801031.

Cheesman EE. 1944. Notes on the nomenclature, classification and possible relationships of cacao populations. Trop Agric. 21:144–159.

Chen S, Songkumarn P, Venu RC, Gowda M, Bellizzi M, Hu J, Liu W, Ebbole D, Meyers B, Mitchell T, Wang G-L. 2012. Identification and Characterization of In planta–Expressed Secreted Effector Proteins from *Magnaporthe oryzae* That Induce Cell Death in Rice. Mol Plant-Microbe Interact. 26:191–202, doi:10.1094/mpmi-05-12-0117-r.

Chiriboga M. 2013. Jornaleros, grandes propietarios y exportación cacaotera, 1790-1925. Quito, Ecuador: Universidad Andina Simón Bolívar / Corporación Editora Nacional. 424 p.

Church WB, Hagen A von. 2008. Chachapoyas: Cultural Development at an Andean Cloud Forest Crossroads. In: H. Silverman, and W.H. Isbell, eds. The Handbook of South American Archeology. New York, USA: Springer. pp. 903–923.

Ciferri R. 1949. Early "Criollo" Cacao in Surinam and the Origin of "Forasteros" of Trinidad and Venezuela. Nature. 163:953–953, doi:10.1038/163953a0.

Ciferri R, Ciferri F. 1957. The evolution of cultivated cacao. Evolution (N Y). 11:381–397, doi:10.2307/2406059.

Ciferri R, Parodi E. 1933. Descrizione del fungo che causa la Moniliasi del cacao. Phytopathol Zeitschrift. 6:539–542.

Clarence-Smith WG. 2000. Cocoa and Chocolate, 1765-1914. New York, NY: Routledge. 319 p.

Clayton LA. 1974. Local Initiative and Finance in Defense of the Viceroyalty of Peru: The Development of Self-Reliance. Hisp Am Hist Rev. 54:284–304, doi:10.2307/2512570.

Clayton LA. 1975. Trade and Navigation in the Seventeenth-Century Viceroyalty of Peru. J Lat Am Stud. 7:1–21, doi:10.1017/S0022216X00016631.

Clement CR, Cristo-Araújo M de, D'Eeckenbrugge GC, Pereira AA, Picanço-Rodrigues D. 2010. Origin and Domestication of Native Amazonian Crops. Diversity. 2:72–106, doi:10.3390/d2010072.

Coetzee MPA, Bloomer P, Wingfield MJ, Wingfield BD. 2011. Paleogene radiation of a plant pathogenic mushroom. PLoS One. 6:e28545, doi:10.1371/journal.pone.0028545.

Collins C, Keane TM, Turner DJ, Kee GO, Fitzpatrick DA, Doyle S. 2013. Genomic and Proteomic Dissection of the Ubiquitous Plant Pathogen, *Armillaria mellea*: Toward a New Infection Model System. J Proteome Res. 12:2552–2570, doi:10.1021/pr301131t.

Coman C, Mot AC, Gal E, Parvu M, Silaghi-Dumitrescu R. 2013. Laccase is upregulated via stress pathways in the phytopathogenic fungus *Sclerotinia sclerotiorum*. Fungal Biol. 117:528–539, doi:10.1016/j.funbio.2013.05.005.

Connor ME. 1920. Yellow fever control in Ecuador: preliminary report. J Am Med Assoc. 74:650–651, doi:10.1001/jama.1920.02620100010004.

Cook RD, Dennis R. 1977. Detection of influential observation in linear regression. Technometrics. 19:15–18, doi:10.1080/00401706.1977.10489493.

Corner EJH. 1996. The agaric genera Marasmius, Chaetocalathus, Crinipellis, Heimiomyces, Resupinatus, Xerula, and Xerulina in Malesia. Berlin: J. Cramer. 175 p.

Costa GGL, Cabrera OG, Tiburcio RA, Medrano FJ, Carazzolle MF, Thomazella DPT, Schuster SC, Carlson JE, Guiltinan MJ, Bailey BA, Mieczkowski P, Pereira GAG, Meinhardt LW. 2012. The mitochondrial genome of *Moniliophthora roreri*, the frosty pod rot pathogen of cacao. Fungal Biol. 116:551–562, doi:10.1016/j.funbio.2012.01.008.

Couch BC, Fudal I, Lebrun M-H, Tharreau D, Valent B, Kim P van, Nottéghem J-L, Kohn LM. 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, doi:10.1534/genetics.105.041780.

Crespo del Campo E, Crespo Andía F. 1997. Cultivo y beneficio del cacao CCN51. Quito, Ecuador: Editorial el Conejo. 136 p.

Csárdi G. 2006. The igraph software package for complex network research. InterJournal, Complex Syst. 1695:1–9.

Cuatrecasas J. 1964. Cacao and its allies: a taxonomic revision of the genus *Theobroma*. Contrib from United States Natl Herb. 35:379–614.

Curran MD, Williams F, Earle JAP, Rima BK, Dam MG van, Bunce M, Middleton D. 1996. Long-Range PCR amplification as an alternative strategy for characterizing novel HLA-B alleles. Eur J Immunogenet. 23:297–309, doi:10.1111/j.1744-313X.1996.tb00125.x.

Cuvi N. 2009. Las semillas del imperialismo agrícola estadounidense en el Ecuador. Procesos Rev Ecuatoriana Hist. 30:69–98, doi:10.29078/rp.v1i30.125.

Dand R. 1997. History and origins of the international cocoa trade. In: The International Cocoa Trade. New York, USA: John Wiley & Sons. pp. 1–22.

Delgado C. AA. 2008. Los productores de cacao en Venezuela: de la esclavitud al cooperativismo. Obs Labor Rev Venez. 1:101–125.

Dennis RWG, Reid DA. 1957. Some Marasmioid Fungi Allegedly Parasitic on Leaves and Twigs in the Tropics. Kew Bull. 12:287–292, doi:10.2307/4114423.

Dent D, Castillo MI Del. 2018. Agricultural Methods. US Pat Appl No 15/327,001.

Dentinger BTM, Gaya E, O'Brien H, Suz LM, Lachlan R, Díaz-Valderrama JR, Koch RA, Aime MC. 2015. Tales from the crypt: genome mining from fungarium specimens improves resolution of the mushroom tree of life. Biol J Linn Soc. 117:11–32, doi:10.1111/bij.12553.

Díaz-Valderrama JR. 2014. A multi-faceted approach for determining the reproductive biology of the causal agent of frosty pod rot of cacao *Moniliopntnora roreri*. Purdue University.

Díaz-Valderrama JR, Aime MC. 2016a. The cacao pathogen *Moniliophthora roreri* (Marasmiaceae) possesses biallelic *A* and *B* mating loci but reproduces clonally. Heredity (Edinb). 116:491–501, doi:10.1038/hdy.2016.5.

Díaz-Valderrama JR, Aime MC. 2016b. The cacao pathogen *Moniliophthora roreri* (Marasmiaceae) produces rhexolytic thallic conidia and their size is influenced by nuclear condition. Mycoscience. 57:208–216, doi:10.1016/j.myc.2016.01.004.

Díaz Morales KM. 2000a. El circuito cacaotero en Venezuela: un análisis del sistema de comercialización (1975-1998). Universidad de los Andes, Mérida, Venezuela.

Díaz Morales KM. 2000b. La comercialización del cacao en Venezuela: un análisis antes y después de la apertura comercial 1975-1998. Agroalimentaria. 11:33–46.

Dorken ME, Eckert CG. 2001. Severely reduced sexual reproduction in northern populations of a clonal plant, Decodon verticillatus (Lythraceae). J Ecol. 89:339–350, doi:10.1046/j.1365-2745.2001.00558.x.

Dray S, Dufour A-B. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. J Stat Softw. 22:1–20.

Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL, Chiu R, Coutinho PM, Feau N, Field M, Frey P, Gelhaye E, Goldberg J, Grabherr MG, Kodira CD, Kohlera A, Kües U, Lindquist EA, Lucas SM, Mago R, Mauceli E, Morin E, Murat C, Pangilinan JL, Park R, et al. 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci U S A. 108:9166–9171, doi:10.1073/pnas.1019315108.

Duret L, Gasteiger E, Perrière G. 1996. LALNVIEW: a graphical viewer for pairwise sequence alignments. Bioinformatics. 12:507–510, doi:10.1093/bioinformatics/12.6.507.

Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol. 7:e1002195, doi:10.1371/journal.pcbi.1002195.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797, doi:10.1093/nar/gkh340.

Edmundson G. 1922. Journal of the Travels and Labours of Father Samuel Fritz in the River of the Amazons Between 1686 and 1723. London: Cambridge University Press. 164 p.

Eilbeck K, Moore B, Holt C, Yandell M. 2009. Quantitative measures for the management and comparison of annotated genomes. BMC Bioinformatics. 10:1–15, doi:10.1186/1471-2105-10-67.

Ellstrad NC, Roose ML. 1987. Patterns of Genotypic Diversity in Clonal Plant Species. Am J Bot. 74:123–131, doi:10.1002/j.1537-2197.1987.tb08586.x.

Enríquez GA, Suárez C. 1978. Monilia disease of cocoa in Costa Rica. Turrialba. 28:339–340.

Erneholm I. 1948. Cacao Production of South America: Historical Development and Present Geographical Distribution. Gothenburg, Sweden: C.R. HOLMQVISTS BOKTRYCKERI AB. 279 p.

Fagan HJ. 1984. An Assessment of Pathological Research on Cocoa in Jamaica from 1950 to 1980 and Current Research Priorities. Trop Pest Manag. 30:430–439, doi:10.1080/09670878409370918.

FAO. 2018. FAOSTAT.

Fernández de Piedrahíta L. 1881. Historia General de las Conquistas del Nuevo Reino de Granada. Bogotá, Colombia: Medardo Rivas. 412 p.

Ferreira AR. 1786. Diário da viagem filosófica pela capitania de São José do Rio negro com a informação do estado presente dos estabelecimentos portugueses na sobredita capitania, desde a vila capital de Barcelos até a fortaleza da barra do dito rio. Barcelos, Brazil. 125 p.

Ferreira P, Carro J, Serrano A, Martinez AT. 2015. A survey of genes encoding H_2O_2-producing GMC oxidoreductases in 10 Polyporales genomes. Mycologia. 107:1105–1119, doi:10.3852/15-027.

Ferry RJ. 1981. Encomienda, African Slavery, and Agriculture in Seventeenth-Century Caracas. Hisp Am Hist Rev. 61:609–635, doi:10.2307/2514606.

Ferry RJ. 1989. The Colonial Elite of Early Caracas: Formation and Crisis 1567 - 1767. Berkeley: University of California Press. 341 p.

Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, et al. 2017. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 45:D190–D199, doi:10.1093/nar/gkw1107.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44:D279–D285, doi:10.1093/nar/gkv1344.

Fiorin GL, Sanchéz-Vallet A, Thomazella DP de T, Prado PFV do, Nascimento LC do, Figueira AV de O, Thomma BPHJ, Pereira GAG, Teixeira PJPL. 2018. Suppression of Plant Immunity by Fungal Chitinase-like Effectors. Curr Biol. 28:3023–3030, doi:10.1016/j.cub.2018.07.055.

Flores González M. 2007. La protección jurídica para el cacao fino y de aroma del Ecuador. Quito, Ecuador: Universidad Andina Simón Bolívar / Corporación Editora Nacional. 159 p p.

Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, Vries RP de, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Górecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, et al. 2012. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. Science (80-). 336:1715–1719, doi:10.1126/science.1221748.

Floudas D, Held BW, Riley R, Nagy LG, Koehler G, Ransdell AS, Younus H, Chow J, Chiniquy J, Lipzen A, Tritt A, Sun H, Haridas S, Labutti K, Ohm RA, Kües U, Blanchette RA, Grigoriev I V, Minto RE, Hibbett DS. 2015. Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. Fungal Genet Biol. 76:78–92, doi:10.1016/j.fgb.2015.02.002.

Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunæs AC, Stenstrøm Y, MacKenzie A, Sørlie M, Horn SJ, Eijsink VGH. 2011. Cleavage of cellulose by a CBM33 protein. Protein Sci. 20:1479–1483, doi:10.1002/pro.689.

Fountain A, Huetz-Adams F. 2018. Cocoa Barometer. Barometer Consortium. 72 p.

Frias G, Purdy L, Schmidt R. 1991. Infection biology of *Crinipellis perniciosa* on vegetative flushes of cacao. Plant Dis. 75:552–556.

Friedrich KM, Nemec S, Czerny C, Fischer H, Plischke S, Gahleitner A, Viola TB, Imhof H, Seidler H, Guillen S. 2010. The story of 12 Chachapoyan mummies through multidetector computed tomography. Eur J Radiol. 76:143–150, doi:10.1016/j.ejrad.2009.07.009.

Fukumori F, Takeuchi N, Hagiwara T, Ito K, Kochibe N, Kobata A, Nagata Y. 1989. Cloning and expression of a functional fucose-specific lectin from an orange peel mushroom, *Aleuria aurantia*. FEBS Lett. 250:153–156, doi:10.1016/0014-5793(89)80709-4.

Gallagher S, Desjardins PR. 2006. Quantitation of DNA and RNA with Absorption and Fluorescence Spectroscopy. Curr Protoc Mol Biol. 76:A.3D.1-A.3D.21, doi:10.1101/pdb.ip47.

Galloway JH. 1989. The Sugar Cane Industry: An Historical Geography from Its Origins to 1914. New York: Cambridge University Press. 271 p.

García Carrión LF. 2010. Catálogo de cultivares de cacao del Perú. Lima, Peru: Ministerio de Agricultura y Riego. 108 p.

Gardner WJ. 1873. A history of Jamaica from its discovery by Christopher Columbus to the present. London: Elliot Stock. 512 p.

Gasco JL. 1987. Cacao and the economic integration of native society in colonial Soconusco, New Spain. University of California, Santa Barbara.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31:3784–3788, doi:10.1093/nar/gkg563.

Gaudet DA, Bhalla MK. 1988. Survey for snow mold diseases of winter cereals in central and northern Alberta , 1983-87. Can Plant Dis Surv. 68:15-.

Gazis R, Kuo A, Riley R, LaButti K, Lipzen A, Lin J, Amirebrahimi M, Hesse CN, Spatafora JW, Henrissat B, Hainaut M, Grigoriev I V., Hibbett DS. 2016. The genome of Xylona heveae provides a window into fungal endophytism. Fungal Biol. 120:26–42, doi:10.1016/j.funbio.2015.10.002.

Gene Ontology Consortium. 2017. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. 45:D331–D338, doi:10.1093/nar/gkw1108.

Glenboski LL. 1983. The ethnobotany of Tukuna indians Amazonas, Colombia. Bogotá, Colombia: Universidad Nacional de Colombia. 92 p.

Gondard P. 1986. Cambios históricos en el aprovechamiento del medio natural ecuatoriano: papel de la demanda social. Cultura. 24:567–577.

González Figueroa AB, Roble Orellana AD. 2014. Aislamiento y Caracterización del hongo *Moniliophthora roreri* (Monilia) en Frutos de *Theobroma cacao* L. (Cacao) del Cultivar San José del Real de la Carrera, Usultán. Universidad de El Salvador.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 27:221–224, doi:10.1093/molbev/msp259.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma F di, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29:644–652, doi:10.1038/nbt.1883.

Griffith GW, Hedger JN. 1994. The breeding biology of Crinipellis perniciosa. Heredity (Edinb). 72:278–289, doi:10.1038/hdy.1994.38.

Grigoriev I V., Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm portal: Gearing up for 1000 fungal genomes. Nucleic Acids Res. 42:699–704, doi:10.1093/nar/gkt1183.

Grimes J. 2009. Rediscovering the Cacao in Ecuador's Upper Napo River Valley. Focus Geogr. 51:23–30, doi:10.1111/j.1949-8535.2009.tb00237.x.

Grisales Ortega SP, Kafuri LA. 2007. Análisis de variabilidad genética en *Moniliophthora roreri* con AP-PCR y RAPD en Antioquia, Colombia. Rev Colomb Biotecnol. IX:15–32.

Grünwald NJ, Everhart SE, Knaus BJ, Kamvar ZN. 2017. Best Practices for Population Genetic Analyses. Phytopathology. 107:1000–1010, doi:10.1094/PHYTO-12-16-0425-RVW.

Grünwald NJ, Goodwin SB, Milgroom MG, Fry WE. 2003. Analysis of Genotypic Diversity Data for Populations of Microorganisms. Phytopathology. 93:738–746, doi:10.1094/PHYTO.2003.93.6.738.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ. 2011. Current trends in microsatellite genotyping. Mol Ecol Resour. 11:591–611, doi:10.1111/j.1755-0998.2011.03014.x.

Guimac Cedillo LY. 2017. Caracterización fisicoquímica y organoléptica del cacao criollo nativo (*Theobroma cacao* L.) de las parcelas cacaoteras de Amazonas APROCAM. Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29:1072–1075, doi:10.1093/bioinformatics/btt086.

Gutarra C. BA, Silva MJ, Márquez KJ, León B. 2013. Análisis de la diversidad genética de 21 aislamientos del hongo *Moniliophthora roreri* basado en marcadores RAPD. Apunt Cienc y Soc. 3:100–110.

Guy L, Roat Kultima J, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 26:2334–2335, doi:10.1093/bioinformatics/btq413.

Hall CJJ van. 1914. Cocoa. London: MacMillan and Co. 515 p.

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. 2009. The WEKA data mining software: An update. ACM SIGKDD Explor Newsl. 11:10–18.

Hamerly MT. 1978. Quantifying the Nineteenth Century: The Ministry Reports and Gazettes of Ecuador as Quantitative Sources. Lat Am Res Rev. 13:138–156.

Han M V., Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol. 30:1987–1997, doi:10.1093/molbev/mst100.

Hartigan PM. 1985. Algorithm AS 217: Computation of the Dip Statistic to Test for Unimodality. J R Stat Soc Ser C (Applied Stat. 34:320–325, doi:10.2307/2347485.

Hasan I, Sugawara S, Fujii Y, Koide Y, Terada D, Iimura N, Fujiwara T, Takahashi KG, Kojima N, Rajia S, Kawsar SMA, Kanaly RA, Uchiyama H, Hosono M, Ogawa Y, Fujita H, Hamako J, Matsui T, Ozeki Y. 2015. MytiLec, a Mussel R-Type Lectin, Interacts with Surface Glycan Gb3 on Burkitt's Lymphoma Cells to Trigger Apoptosis through Multiple Pathways. Mar Drugs. 13:7377–7389, doi:10.3390/md13127071.

Haz Alvarez M, Cabrera Vicuña T. 2010. Informe de Labores 2009-2010. Quevedo, Ecuador: Universidad Técnica Estatal de Quevedo. 73 p.

Hebbar PK. 2007. Cacao Diseases: A Global Perspective from an Industry Point of View. Phytopathology. 97:1658–1663, doi:10.1094/PHYTO-97-12-1658.

Heitman J, Sun S, James TY. 2013. Evolution of fungal sexual reproduction. Mycologia. 105:1–27, doi:10.3852/12-253.

Hemming J. 2008. Tree of rivers : The story of the Amazon. New York: Thames & Hudson. 368 p.

Hemsworth GR, Johnston EM, Davies GJ, Walton PH. 2015. Lytic Polysaccharide Monooxygenases in Biomass Conversion. Trends Biotechnol. 33:747–761, doi:10.1016/j.tibtech.2015.09.006.

Henderson P. 1997. Cocoa, Finance and the State in Ecuador, 1895-1925. Bull Lat Am Res. 16:169–186, doi:10.1016/S0261-3050(96)00011-3.

Hernández T. TA, Aranzazu H. F, Arévalo G. E, Rios R. R. 1990. La moniliasis del cacao en el Perú. Agrotrópica (Brasil). 2:56–58.

Herrmann L, Haase I, Blauhut M, Barz N, Fischer M. 2014. DNA-based differentiation of the ecuadorian cocoa types CCN-51 and Arriba based on sequence differences in the chloroplast genome. J Agric Food Chem. 62:12118–12127, doi:10.1021/jf504258w.

Hibbett DS, Grimaldi D, Donoghue MJ. 1997. Fossil mushrooms from Miocene and Cretaceous ambers and the evolution of Homobasidiomycetes. Am J Bot. 84:981–991, doi:10.2307/2446289.

Higginbotham SJ, Arnold AE, Ibañez A, Spadafora C, Coley PD, Kursar TA. 2013. Bioactivity of Fungal Endophytes as a Function of Endophyte Taxonomy and the Taxonomy and Distribution of Their Host Plants. PLoS One. 8:e73192, doi:10.1371/journal.pone.0073192.

Hill MO. 1973. Diversity and Evenness: A Unifying Notation and Its Consequences. Ecology. 54:427–432, doi:10.2307/1934352.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 12:491, doi:10.1186/1471-2105-12-491.

Hsiang T, Matsumoto N, Millet SM. 1999. Biology and Management of Typhula Snow Molds of Turfgrass. Plant Dis. 83:788–798.

Huang X, Miller W. 1991. A time-efficient, linear-space local similarity algorithm. Adv Appl Math. 12:337–357, doi:10.1016/0196-8858(91)90017-D.

Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD. 2013. REAPR : a universal tool for genome assembly evaluation. Genome Biol. 14:R47, doi:10.1186/gb-2013-14-5-r47.

Hurst WJ, Tarka Jr SM, Powis TG, Valdez Jr F, Hester TR. 2002. Cacao usage by the earliest Maya civilization. Nature. 418:289–290, doi:10.1038/418289a.

Imaña G. 2015. Una plaga afecta al 65% de los cultivos de cacao de Alto Beni. La Razón - Boliv.

INDECOPI. 2016. El Indecopi entrega décima Denominación de Origen Cacao Amazonas Perú que contribuirá al desarrollo económico de más de 1,200 productores. Lima, Peru. 2 p.

Israel J. 1989. Dutch primacy in world trade, 1585-1740. Oxford, UK: Clerendon Press. 462 p.

Iwaro AD, Bekele FL, Butler DR. 2003. Evaluation and utilisation of cacao (*Theobroma cacao* L.) germplasm at the International Cocoa Genebank, Trinidad. Euphytica. 130:207–221, doi:10.1023/A:102285513.

Jaimes YY, Gonzalez C, Rojas J, Cornejo OE, Mideros MF, Restrepo S, Cilas C, Furtado EL. 2016. Geographic Differentiation and Population Genetic Structure of *Moniliophthora roreri* in the Principal Cocoa Production Areas in Colombia. Plant Dis. 100:1548–1558, doi:10.1094/PDIS-12-15-1498-RE.

Jimenez JC, Amores FM, Solórzano EG, Rodríguez GA, Mantia A La, Blasi P, Loor RG. 2018. Differentiation of Ecuadorian National and CCN-51 cocoa beans and their mixtures by computer vision. J Sci Food Agric. 98:2824–2829, doi:10.1002/jsfa.8790.

Johnson ES, Rutherford MA, Edgington S, Flood J, Crozier J, Cafá G, Buddie AG, Offord L, Elliott SM, Christie K V. 2017. First report of *Moniliophthora roreri* causing frosty pod rot on *Theobroma cacao* in Jamaica. New Dis Reports. 36:2–2, doi:10.5197/j.2044-0588.2017.036.002.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24:1403–1405, doi:10.1093/bioinformatics/btn129.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11:94, doi:doi:10.1186/1471-2156-11-94.

Jones J, Dangl J. 2006. The plant immune system. Nature. 444:323–329, doi:10.1038/nature05286.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236–1240, doi:10.1093/bioinformatics/btu031.

Jones PG, Allaway D, Gilmour DM, Harris C, Rankin D, Retzel ER, Jones CA. 2002. Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. Planta. 216:255–264, doi:10.1007/s00425-002-0882-6.

Jonge R de, Esse HP Van, Kombrink A, Shinya T, Desaki Y, Bours R, Krol S van der, Shibuya N, Joosten MHAJ, Thomma BPHJ. 2010. Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants. Science (80-). 329:953–955, doi:10.1126/science.1190859.

Jorda J, Kajava A V. 2009. T-REKS: identification of Tandem REpeats in sequences with a K-meanS based algorithm. Bioinformatics. 25:2632–2638, doi:10.1093/bioinformatics/btp482.

Joseph EL. 1838. History of Trinidad. London: F. Cass. 272 p.

Käll L, Krogh A, Sonnhammer ELL. 2004. A Combined Transmembrane Topology and Signal Peptide Prediction Method. J Mol Biol. 338:1027–1036, doi:10.1016/j.jmb.2004.03.016.

Käll L, Krogh A, Sonnhammer ELL. 2005. An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics. 21:i251–i257, doi:10.1093/bioinformatics/bti1014.

Kamvar ZN, Brooks JC, Grünwald NJ. 2015a. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet. 6:1–10, doi:10.3389/fgene.2015.00208.

Kamvar ZN, Larsen MM, Kanaskie AM, Hansen EM, Grünwald NJ. 2015b. Spatial and Temporal Analysis of Populations of the Sudden Oak Death Pathogen in Oregon Forests. Phytopathology. 105:982–987, doi:10.1094/PHYTO-12-14-0350-FI.

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2:e281, doi:10.7717/peerj.281.

Karl TL. 1997. The paradox of plenty: oil booms and petro-states. University of California Press. 342 p.

Kau AL, Hunstad DA, Hultgren SJ. 2005. Interaction of uropathogenic *Escherichia coli* with host uroepithelium. Curr Opin Microbiol. 8:54–59, doi:10.1016/j.mib.2004.12.001.

Kaur K, Sharma A, Capalash N, Sharma P. 2019. Multicopper oxidases: Biocatalysts in microbial pathogenesis and stress management. Microbiol Res. 222:1–13, doi:10.1016/j.micres.2019.02.007.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649, doi:10.1093/bioinformatics/bts199.

Keen B. 1992. The life of the admiral Christopher by his son Ferdinand. New Brunswick, New Jersey: Ritgers University. 316 p.

Kerekes JF, Desjardin DE. 2009. A monograph of the genera *Crinipellis* and *Moniliophthora* from Southeast Asia including a molecular phylogeny of the nrITS region. Fungal Divers. 37:101–152.

Khan F, Khan MI. 2011. Fungal Lectins: Current Molecular and Biochemical Properties. Int J Biol Chem. 5:1–15.

Khidir HH, Eudy DM, Porras-Alfaro a., Herrera J, Natvig DO, Sinsabaugh RL. 2010. A general suite of fungal endophytes dominate the roots of two dominant grasses in a semiarid grassland. J Arid Environ. 74:35–42, doi:10.1016/j.jaridenv.2009.07.014.

Kiemen MC. 1948. The Indian Policy of Portugal in America, with Special Reference to the Old State Maranhão, 1500-1755. Americas (Engl Ed). 5:131–171.

Kijpornyongpan T, Urbina H, Suh SO, Luangsa-ard J, Aime MC, Blackwell M. 2019. The Suhomyces clade: from single isolate to multiple species to disintegrating sex loci. FEMS Yeast Res. 19, doi:10.1093/femsyr/foy125.

Kim K-T, Jeon J, Choi J, Cheong K, Song H, Choi G, Kang S, Lee Y-H. 2016. Kingdom-Wide Analysis of Fungal Small Secreted Proteins (SSPs) Reveals their Potential Role in Host Association. Front Plant Sci. 7:186, doi:10.3389/fpls.2016.00186.

Kirschner R, Lee IS, Chen C-J. 2013. Ovularia puerariae Sawada is the hyphomycetous anamorph of a new Marasmius species on living leaves of kudzu (Pueraria montana, Fabaceae). Mycologia. 105:781–792, doi:10.3852/12-285.

Knapp AW. 1920. Cocoa and chocolate: their history from plantation to consumer. London: Chapman and Hall, Ltd. 210 p.

Koch K, Ensikat H-J. 2008. The hydrophobic coatings of plant surfaces: Epicuticular wax crystals and their morphologies, crystallinity and molecular self-assembly. Micron. 39:759–772, doi:10.1016/j.micron.2007.11.010.

Koch RA, Aime MC. 2018. Population structure of *Guyanagaster necrorhizus* supports termite dispersal for this enigmatic fungus. Mol Ecol. 27:2667–2679, doi:10.1111/mec.14710.

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Choi C, Cichocki N, Clum A, Colpaert J, Copeland A, Costa MD, Doré J, Floudas D, Gay G, Girlanda M, Henrissat B, Herrmann S, Hess J, Högberg N, Johansson T, Khouja H-R, LaButti K, Lahrmann U, Levasseur A, Lindquist EA, Lipzen A, Marmeisse R, et al. 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nat Genet. 47:410–415, doi:10.1038/ng.3223.

Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics. 5:59, doi:10.1186/1471-2105-5-59.

Kropp BR, Albee-Scott S. 2012. *Moniliophthora aurantiaca* sp. nov., a Polynesian species occurring in littoral forests. Mycotaxon. 120:493–503, doi:10.5248/120.493.

Kruger RP, Winter HC, Simonson-Leff N, Stuckey JA, Goldstein IJ, Dixon JE. 2002. Cloning, Expression, and Characterization of the Galα1,3Gal High Affinity Lectin from the Mushroom *Marasmius oreades*. J Biol Chem. 277:15002–15005, doi:10.1074/jbc.m200165200.

Kuehne SA, Cartman ST, Heap JT, Kelly ML, Cockayne A, Minton NP. 2010. The role of toxin A and toxin B in *Clostridium difficile* infection. Nature. 467:711–713, doi:10.1038/nature09397.

Kües U. 2015. From two to many: Multiple mating types in Basidiomycetes. Fungal Biol Rev. 29:126–166, doi:10.1016/j.fbr.2015.11.001.

Kulkarni S, Nene S, Joshi K. 2017. Production of Hydrophobins from fungi. Process Biochem. 61:1–11, doi:10.1016/j.procbio.2017.06.012.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 35:1547–1549, doi:10.1093/molbev/msy096.

Laker HA, Sreenivasan TN, Kumar DR. 1988. Present status of witches' broom disease of cocoa in Trinidad. Trop Pest Manag. 34:318–323, doi:10.1080/09670878809371264.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359, doi:10.1038/nmeth.1923.

Lanz O, Granado Y. 2009. Diagnóstico agrosocioeconómico del sector cacao (*Theobroma cacao* L.) en Yaguaraparo, Municipio Cajigal, estado Sucre, Venezuela. Rev UDO Agrícola. 9:425–435.

Larrea C, North LL. 1997. Ecuador: Adjustment Policy Impacts on Truncated Development and Democratisation. Third World Q. 18:913–934.

Laurent V, Risterucci AM, Lanaud C. 1994. Genetic diversity in cocoa revealed by cDNA probes. Theor Appl Genet. 88:193–198, doi:10.1007/BF00225897.

Lavaina Cuetos ML. 2014. “Señor, por amor de Dios, no me coma”: La piratería en el Pacífico según Francisco Requena. Derroteros La Mar Del Sur. 22:100–109.

Lee SC, Ni M, Li W, Shertz C, Heitman J. 2010. The Evolution of Sex: a Perspective from the Fungal Kingdom. Microbiol Mol Biol Rev. 74:298–340, doi:10.1128/MMBR.00005-10.

Leiter J, Harding S. 2004. Trinidad, Brazil, and Ghana: three melting moments in the history of cocoa. J Rural Stud. 20:113–130, doi:10.1016/S0743-0167(03)00034-2.

Lerceteau E, Robert T, Pétiard V, Crouzillat D. 1997. Evaluation of the extent of genetic variability among *Theobroma cacao* L. accessions using RAPD and RFLP markers. Theor Appl Genet. 95:10–19, doi:10.1007/s001220050.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels. 6:1–14, doi:10.1186/1754-6834-6-41.

Liu J-J, Shamoun SF, Leal I, Kowbel R, Sumampong G, Zamany A. 2018. Characterization of *Heterobasidion occidentale* transcriptomes reveals candidate genes and DNA polymorphisms for virulence variations. Microb Biotechnol. 11:537–550, doi:10.1111/1751-7915.13259.

Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. 2013. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42:D490–D495, doi:10.1093/nar/gkt1178.

Loor RG, Risterucci a. M, Courtois B, Fouet O, Jeanneau M, Rosenquist E, Amores F, Vasco A, Medina M, Lanaud C. 2009. Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador. Tree Genet Genomes. 5:421–433, doi:10.1007/s11295-008-0196-3.

Loor Solorzano RG, Fouet O, Lemainque A, Pavek S, Boccara M, Argout X, Amores F, Courtois B, Risterucci AM, Lanaud C. 2012. Insight into the Wild Origin , Migration and Domestication History of the Fine Flavour Nacional Theobroma cacao L. Variety from Ecuador. PLoS One. 7:e48438, doi:10.1371/journal.pone.0048438.

López de Velasco J. 1574. Geografía y Descripción Universal de las Indias. Madrid, Spain: Real Academia de la Historia. 808 p.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550, doi:10.1186/s13059-014-0550-8.

Lovell WG. 1992. "Heavy Shadows and Black Night": Disease and Depopulation in Colonial Spanish America. Ann Assoc Am Geogr. 82:426–443, doi:10.1111/j.1467-8306.1992.tb01968.x.

Lovell WG, Lutz CH. 1992. The historical demography of colonial Central America. In: Yearbook (Conference of Latin Americanist Geographers). pp. 127–138.

Lu P, Lei M, Xiao F, Zhang L, Wang Y. 2015. Complete Genome Sequence of *Terribacillus aidingensis* Strain MP602, a Moderately Halophilic Bacterium Isolated from *Cryptomeria fortunei* in Tianmu Mountain in China. Genome Announc. 3:e00126-15, doi:10.1128/genomeA.00126-15.Copyright.

Ludwig JA, Reynolds JF. 1988. Statistical ecology: a primer on methods and computing. New York, USA: John Wiley and Sons. 337 p.

MacDonald J, Doering M, Canam T, Gong Y, Guttman DS, Campbell MM, Master ER. 2011. Transcriptomic Responses of the Softwood-Degrading White-Rot Fungus Phanerochaete carnosa during Growth on Coniferous and Deciduous Wood. Appl Environ Microbiol. 77:3211–3218, doi:10.1128/AEM.02490-10.

Mahony MA. 2008. Creativity under Constraint: Enslaved Afro-Brazilian Families in Brazil's Cacao Area, 1870- 1890. J Soc Hist. 41:633–666.

Malavé Mata H. 1974. Formación histórica del antidesarrollo de Venezuela. Havana, Cuba: Casa de las Américas. 274 p.

Mann ME, Zhang Z, Rutherford S, Bradley RS, Hughes MK, Shindell D, Ammann C, Faluvegi G, Ni F. 2009. Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly. Science (80-). 326:1256–1261, doi:10.1126/science.1177303.

Mantel N. 1967. The Detection of Disease Clustering and a Generalized Regression Approach. Cancer Res. 27:209–220.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko M V, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39:D225–D229, doi:10.1093/nar/gkq1189.

Maridueña-Zavala MG, Villavicencio-Vásquez ME, Cevallos-Cevallos JM, Peralta EL. 2016. Molecular and morphological characterization of *Moniliophthora roreri* isolates from cacao in Ecuador. Can J Plant Pathol. 38:460–469, doi:10.1080/07060661.2016.1261372.

Marshall R, Kombrink A, Motteram J, Loza-Reyes E, Lucas J, Hammond-Kosack KE, Thomma BPHJ, Rudd JJ. 2011. Analysis of Two in Planta Expressed LysM Effector Homologs from the Fungus Mycosphaerella graminicola Reveals Novel Functional Properties and Varying Contributions to Virulence on Wheat. Plant Physiol. 156:756–769, doi:10.1104/pp.111.176347.

Martin PA. 1933. Slavery and Abolition in Brazil. Hisp Am Hist Rev. 13:151–196, doi:10.2307/2506690.

Matheny PB, Curtis JM, Hofstetter V, Aime MC, Moncalvo J-M, Ge Z-W, Yang Z-L, Slot JC, Ammirati JF, Baroni TJ, Bougher NL, Hughes KW, Lodge DJ, Kerrigan RW, Seidl MT, Aanen DK, DeNitis M, Daniele GM, Desjardin DE, Kropp BR, Norvell LL, Parker A, Vellinga EC, Vilgalys R, Hibbett DS. 2006. Major clades of Agaricales: a multilocus phylogenetic overview. Mycologia. 98:982–995, doi:10.1080/15572536.2006.11832627.

Matson PA, Parton WJ, Power AG, Swift MJ. 1997. Agricultural Intensification and Ecosystem Properties. Science (80-). 277:504–509, doi:10.1126/science.277.5325.504.

Matsumura K, Wang X, Hata Y, Shigeta M, Mizuno-Horikawa Y, Higashida K, Ishida H, Miyoshi E, Gu J, Taniguchi N, Yamamoto K. 2007. Carbohydrate Binding Specificity of a Fucose-specific Lectin from *Aspergillus oryzae* A NOVEL PROBE FOR CORE FUCOSE. J Biol Chem. 282:15700–15708, doi:10.1074/jbc.m701195200.

McCook S. 2002a. Las epidemias liberales: Agricultura, ambiente y globalización en Ecuador (1790-1930). In: B. García Martínez, and M. del R. Prieto, eds. Estudios Sobre Historia y Ambiente En América II: Norteamérica, Sudamérica y El Pacífico. Mexico: Instituto Panamericano de Geografía e Historia. pp. 223–246.

McCook S. 2002b. States of Nature: Science, Agriculture and Environment in the Spanish Caribbean, 1760 - 1940. Austin: University of Texas Press. 203 p.

Mckenzie M. 1994. La política y la gestión rural: la experiencia del Ecuador. Quito, Ecuador: FLACSO. 368 p.

McNeil CL. 2006. Chocolate in Mesoamerica: A Cultural History of Cacao. Gainesville: University Press of Florida. 542 p.

Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF. 2010. QDD: A user-friendly program to select microsatellite markers and design primers from large sequencing projects. Bioinformatics. 26:403–404, doi:10.1093/bioinformatics/btp670.

Meglécz E, Pech N, Gilles A, Dubut V, Hingamp P, Trilles A, Grenier R, Martin JF. 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. Mol Ecol Resour. 14:1302–1313, doi:10.1111/1755-0998.12271.

Meinhardt LW, Lacerda Costa GG, Thomazella DPT, Teixeira PJPL, Carazzolle MF, Schuster SC, Carlson JE, Guiltinan MJ, Mieczkowski P, Farmer A, Ramaraj T, Crozier J, Davis RE, Shao J, Melnick RL, Pereira GAG, Bailey BA. 2014. Genome and secretome analysis of the hemibiotrophic fungal pathogen, *Moniliophthora roreri*, which causes frosty pod rot disease of cacao: mechanisms of the biotrophic and necrotrophic phases. BMC Genomics. 15:164–189, doi:10.1186/1471-2164-15-164.

Melo CJ, Hollander GM. 2013. Unsustainable development: Alternative food networks and the Ecuadorian Federation of Cocoa Producers, 1995-2010. J Rural Stud. 32:251–263, doi:10.1016/j.jrurstud.2013.07.004.

Méndez Ramírez LA. 2015. El cultivo del cacao venezolano a partir de Maruma. Hist Caribe. X:69–101, doi:10.15648/hc.27.2015.3.

Mendgen K, Hahn M. 2002. Plant infection and the establishment of fungal biotrophy. Trends Plant Sci. 7:352–356, doi:10.1016/S1360-1385(02)02297-5.

Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, Terauchi R, Nishizawa Y, Shibuya N, Thomma BPHJ, Talbot NJ. 2012. Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. Plant Cell. 24:322–335, doi:10.1105/tpc.111.092957.

Millas JC. 1968. Hurricanes of the Caribbean and adjacent regions, 1492–1800. Cambridge, MA: Academy of the Arts and Sciences of the Americas. 328 p.

Miller RP, Nair PKR. 2006. Indigenous agroforestry systems in Amazonia: From prehistory to today. Agrofor Syst. 66:151–164, doi:10.1007/s10457-005-6074-1.

Miller RP, Penn JW, Leeuwen J van. 2006. Amazonian homegarderns: their ethnohistory and potential contribution to agroforestry development. In: B.M. Kuma, and P.K.R. Nair, eds. Tropical Homegardens: A Time-Tested Example of Sustainable Agroforestr. Dordrecht, The Netherlands: Springer. pp. 43–60.

Millon RF. 1955. Trade, Tree Cultivation, and the Development of Private Property in Land. Am Anthropol. 57:698–712.

Miranda F. 1962. Wild Cacao in the Lacandona Forest, Chiapas, Mexico. Cacao (Turrialba). 7:7.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41:e121, doi:10.1093/nar/gkt263.

Momsen JH, Richardson P. 2009. Caribbean Cocoa: Planting and Production. In: L.E. Grivetti, and H.Y. Shapiro, eds. Chocolate: History, Culture, and Heritage. New Jersey: John Wiley & Sons. pp. 481–491.

Moncayo R. R. 2016. Exportación Ecuatoriana de Cacao 2015. Guayaquil, Ecuador. 6 p.

Mondego JMC, Carazzolle MF, Costa GGL, Formighieri EF, Parizzi LP, Rincones J, Cotomacci C, Carraro DM, Cunha AF, Carrer H, Vidal RO, Estrela RC, García O, Thomazella DPT, Oliveira B V de, Pires AB, Rio MCS, Araújo MRR, Moraes MH de, Castro LAB, Gramacho KP, Gonçalves MS, Neto JPM, Neto AG, Barbosa L V, Guiltinan MJ, Bailey BA, Meinhardt LW, Cascardo JCM, et al. 2008. A genome survey of Moniliophthora perniciosa gives new insights into Witches' Broom Disease of cacao. BMC Genomics. 9:548, doi:10.1186/1471-2164-9-548.

Moodie-Kublalsingh S. 1994. The Cocoa Panyols of Trinidad: An Oral Record. London: British Academic Press. 242 p.

Moore D, Robson GD, Trinci APJ. 2011. 21st Century Guidebook to Fungi. Cambridge, UK: Cambridge University Press. 627 p.

Mora Urpi J. 1958. Notas sobre el posible origen y la variabilidad del cacao cultivado en America Tropical. Turrialba. 8:34–43.

Moreira RFC. 2006. Estructura Genética de Populações de *Crinipellis perniciosa* e *Moniliophthora roreri* utilizando marcadores RAPD e SSR. Universidade Estadual Paulista.

Morin PA, Luikart G, Wayne RK, The SNP workshop group. 2004. SNPs in ecology , evolution and conservation. Trends Ecol Evol. 19:208–216, doi:10.1016/j.tree.2004.01.009.

Motamayor JC, Lachenaud P, Silva e Mota JW da, Loor R, Kuhn DN, Brown JS, Schnell RJ. 2008. Geographic and Genetic Population Differentiation of the Amazonian Chocolate Tree (*Theobroma cacao* L). PLoS One. 3:e3311, doi:10.1371/journal.pone.0003311.

Motamayor JC, Risterucci AM, Heath M, Lanaud C. 2003. Cacao domestication II: progenitor germplasm of the Trinitario cacao cultivar. Heredity (Edinb). 91:322–330, doi:10.1038/sj.hdy.6800298.

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C. 2002. Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity (Edinb). 89:380–386, doi:10.1038/sj.hdy.6800156.

Motilal LA, Sreenivasan TN. 2012. Revisiting 1727: Crop Failure Leads to the Birth of Trinitario Cacao. J Crop Improv. 26:599–626, doi:10.1080/15427528.2012.663734.

Mukherjee S, Huntemann M, Ivanova N, Kyrpides NC, Pati A. 2015. Large-scale contamination of microbial isolate genomes by illumina Phix control. Stand Genomic Sci. 10:1–4, doi:10.1186/1944-3277-10-18.

Müller AS. 1941. El reconocimiento de las enfermedades de las plantas cultivadas de Venezuela. Boletín La Soc Venez Ciencias Nat. 8:99–113.

Nei M. 1972. Genetic Distance between Populations. Am Nat. 106:283–292, doi:10.1086/282771.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89:583–590.

Newson LA. 1976. Aboriginal and Spanish Colonial Trinidad: A study in culture contact. London: Academic Press. 344 p.

Nikolenko SI, Korobeynikov AI, Alekseyev MA. 2013. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics. 14:S7, doi:10.1186/1471-2164-14-S1-S7.

Nithya A, Babu S. 2017. Prevalence of plant beneficial and human pathogenic bacteria isolated from salad vegetables in India. BMC Microbiol. 17:64, doi:10.1186/s12866-017-0974-x.

Nyadanu D, Akromah R, Adomako B, Kwoseh C, Dzahini-Obiatey H, Lowor ST, Akrofi AY, Assuah MK. 2012a. Host Plant Resistance to *Phytophthora* Pod Rot in Cacao (*Theobroma cacao* L.): The Role of Epicuticular Wax on Pod and Leaf surfaces. Int J Bot. 8:13–21, doi:10.3923/ijb.2012.13.21.

Nyadanu D, Akromah R, Adomako B, Kwoseh C, Lowor ST, Dzahini-Obiatey H, Akrofi AY, Ansah FO, Assuah MK. 2012b. Morpho-physiological Mechanisms of Resistance to Black Pod Disease in Cacao (*Theobroma cacao* L.). Int J Plant Breed Genet. 6:54–68, doi:10.3923/ijpbg.2012.54.68.

Oda Y, Senaha T, Matsuno Y, Nakajima K, Naka R, Kinoshita M, Honda E, Furuta I, Kakehi K. 2003. A new fungal lectin recognizing α(1-6)-linked fucose in the *N*-glycan. J Biol Chem. 278:32439–32447, doi:10.1074/jbc.M305181200.

Ogata N. 2002. Studies of Mesoamerican Tropical Trees: Trees o f the Maya Region and a Case of Study on the Ethnobotany and Phylogeography of Cacao (*Theobroma cacao* L.). University of California Riverside.

Ogata N, Gómez-Pompa A, Taube KA. 2006. The Domestication and Distribution of *Theobroma cacao* L. in the Neotropics. In: C.L. McNeil, ed. Chocolate in Mesoamerica: A Cultural History of Cacao. Gainesville: University Press of Florida. pp. 70–89.

Olavarriaga PJ de. 1722. Instrucción general y particular del estado presente de la provincia de Venezuela en los años de 1720 y 1721. Venezuela: Juan Amador Courten. 237 p.

Olsen G. 1990. “Newick’s 8:45” Tree Format Standard.

Osorio-Guarín JA, Berdugo-Cely J, Coronado RA, Zapata YP, Quintero C, Gallego-Sánchez G, Yockteng R. 2017. Colombia a Source of Cacao Genetic Diversity As Revealed by the Population Structure Analysis of Germplasm Bank of *Theobroma cacao* L. Front Plant Sci. 8:1994, doi:10.3389/fpls.2017.01994.

Palma M. 1953. El cultivo del cacaotero en la Región Central. El Agric Venez. 18:28–33.

Parra D, Pérez Martínez S, Sosa D, Rumbos R, Gutiérrez B, Moya A. 2009. Avances en las investigaciones venezolanas sobre enfermedades del cacao. Rev Estud Transdiscipl. 1:56–75.

Parsons JJ. 1949. Antioqueño colonization in Western Colombia. University of California Press. 233 p.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol. 6:288–295, doi:10.1111/j.1471-8286.2005.01155.x.

Peakall R, Smouse PE. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics. 28:2537–2539, doi:10.1093/bioinformatics/bts460.

Pegler DN. 1978. *Crinipellis perniciosa* (Agaricales). Kew Bull. 32:731–736.

Pereira JL, Almeida LCC de, Santos SM. 1996. Witches’ broom disease of cocoa in Bahia: attempts at eradication and containment. Crop Prot. 15:743–752.

Periera JL, Ram A, Figueiredo JM de, Almeida LCC de. 1990. The first occurrence of witches’ broom disease in the principal cocoa growing region of Brazil. Trop Agric. 67:188–189.

Petersen TN, Brunak S, Heijne G von, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8:785–786, doi:10.1038/nmeth.1701.

Phillips-Mora W. 2003. Origin, biogeography, genetic diversity and taxonomic affinities of the cacao (*Theobroma cacao* L.) fungus Moniliophthora roreri (Cif.) Evans et al. as determined using molecular, phytopathological and morpho-physiological evidence. University of Reading, UK.

Phillips-Mora W, Aime MC, Wilkinson MJ. 2007a. Biodiversity and biogeography of the cacao (*Theobroma cacao*) pathogen *Moniliophthora roreri* in tropical America. Plant Pathol. 56:911–922, doi:10.1111/j.1365-3059.2007.01646.x.

Phillips-Mora W, Aime MC, Wilkinson MJ. 2007b. Biodiversity and biogeography of the cacao (*Theobroma cacao*) pathogen Moniliophthora roreri in tropical America. Plant Pathol. 56:911–922, doi:10.1111/j.1365-3059.2007.01646.x.

Phillips-Mora W, Arciniegas-Leal A, Mata-Quirós A, Motamayor-Arias JC. 2013. Catalogue of cacao clones selected by CATIE for commercial plantings. Turrialba, Costa Rica. 68 p.

Phillips-Mora W, Baqueros F, Melnick RL, Bailey BA. 2015. First report of frosty pod rot caused by *Moniliophthora roreri* on cacao in Bolivia. New Dis Reports. 31:29, doi:10.1111/j.1365-3059.2006.01378.x.

Phillips-Mora W, Castillo J, Arciniegas A, Mata A, Sánchez A, Leandro M, Astorga C, Motamayor J, Guyton B, Seguine E, Schnell R. 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; 16-21th November 2009. Bali, Indonesia. pp. 93–99.

Phillips-Mora W, Cawich J, Garnett W, Aime MC. 2006a. First report of frosty pod rot (moniliasis disease) caused by *Moniliophthora roreri* on cacao in Belize. Plant Pathol. 55:584, doi:10.1111/j.1365-3059.2006.01378.x.

Phillips-Mora W, Coutiño A, Ortiz CF, López AP, Hernández J, Aime MC. 2006b. First report of *Moniliophthora roreri* causing frosty pod rot (moniliasis disease) of cacao in Mexico. Plant Pathol. 55:584.

Phillips-Mora W, Wilkinson MJ. 2007. Frosty Pod of Cacao: A Disease with a Limited Geographic Range but Unlimited Potential for Damage. Phytopathology. 97:1644–1647, doi:10.1094/PHYTO-97-12-1644.

Pilo P, Vilei EM, Peterhans E, Bonvin-Klotz L, Stoffel MH, Dobbelaere D, Frey J. 2005. A Metabolic Enzyme as a Primary Virulence Factor of Mycoplasma mycoides subsp. mycoides Small Colony. J Bacteriol. 187:6824–6831, doi:10.1128/jb.187.19.6824-6831.2005.

Pineo RF. 1988. Reinterpreting Labor Militancy: The Collapse of the Cacao Economy and the General Strike of 1922 in Guayaquil, Ecuador. Hisp Am Hist Rev. 68:707–736.

Pittier H. 1935. Degeneration of cacao through natural hybridization. Heredity (Edinb). 26:385–390.

Pong VM, Abidin MA, Almaliky BSA, Kadir J, Wong MY. 2012. Isolation, Fruiting and Pathogenicity of *Marasmiellus palmivorus* (Sharples) Desjardin (comb. prov.) in Oil Palm Plantations in West Malaysia. Pertanika J Trop Agric Sci. 35:37.

Porter D, Farnham WF. 1986. *Mycaureola dilseae*, a marine basidiomycete parasite of the red alga, *Dilsea carnosa*. Trans Br Mycol Soc. 87:575–582, doi:10.1016/S0007-1536(86)80098-5.

Portillo E, Martínez E, Araujo F, Parra R, Esparza D. 1995. Diagnóstico Técnico-Agronómico para el Cultivo Cacao (*Theobroma cacao* L.) en el Sur del Lago de Maracaibo. Rev Agron. 12:151–166.

Pound FJ. 1938. Cacao and witchbroom disease (*Marasmius perniciosus*) of South America with notes on other species of *Theobroma*. Trinidad and Tobago: Vuille's Printerie. 59 p.

Pound FJ. 1943. Cacao and witches' broom disease (*Marasmius perniciosus*): Report on a recent visit to the Amazon territory of Peru, September, 1942.

Powis TG, Cyphers A, Gaikwad NW, Grivetti L, Cheong K. 2011. Cacao use and the San Lorenzo Olmec. Proc Natl Acad Sci. 108:8595–8600, doi:10.1073/pnas.1100620108.

Powis TG, Hurst WJ, Rodríguez M del C, Ortíz CP, Blake M, Cheetham D, Coe MD, Hodgson JG. 2008. The Origins of Cacao Use in Mesoamerica. Mexicon. 30:35–38.

Powis TG, Valdez Jr. F, Hester TR, Hurst WJ, Tarka SM. 2002. Spouted vessels and cacao use among the Preclassic Maya. Lat Am Antiq. 13:85–106.

Prado Júnior C. 1967. The colonial background of modern Brazil. Berkeley: University of California Press. 535 p.

Prescott WH. 1847. History of the Conquest of Peru. New York, USA: Heritage Press.

Presti L Lo, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. 2015. Fungal Effectors and Plant Susceptibility. Annu Rev Plant Biol. 66:513–545, doi:10.1146/annurev-arplant-043014-114623.

Prim RC. 1957. Shortest Connection Networks And Some Generalizations. Bell Syst Tech J. 36:1389–1401.

Promputtha I, Lumyong S, Dhanasekaran V, McKenzie EHC, Hyde KD, Jeewon R. 2007. A phylogenetic evaluation of whether endophytes become saprotrophs at host senescence. Microb Ecol. 53:579–590, doi:10.1007/s00248-006-9117-x.

Pujic P, Fournier P, Alloisio N, Hay A-E, Maréchal J, Anchisi S, Normand P. 2012. Lectin genes in the *Frankia alni* genome. Arch Microbiol. 194:47–56, doi:10.1007/s00203-011-0770-1.

Qhanya LB, Matowane G, Chen W, Sun Y, Letsimo EM, Parvez M, Yu J-H, Mashele SS, Syed K. 2015. Genome-Wide Annotation and Comparative Analysis of Cytochrome P450 Monooxygenases in Basidiomycete Biotrophic Plant Pathogens. PLoS One. 10:e0142100, doi:10.1371/journal.pone.0142100.

Quevedo C. RI. 1998. Venezuela: un perfil general - La tenencia de la tierra y puntos algidos de su mercado. Agroalimentaria. 6:59–66.

Quinlana RJ, Sweeneya MD, Leggiob L Lo, Ottenb H, Poulsenb J-CN, Johansenc KS, Kroghc KBRM, Jørgensenc CI, Tovborgc M, Anthonsenc A, Tryfonad T, Walterc CP, Dupreed P, Xua F, Daviese GJ, Walton PH. 2011. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. Proc Natl Acad Sci. 108:15079–15084, doi:10.1073/pnas.1105776108.

Quiñones M, Espinoza E, Yovera F, Cuchilla Y, Castro D. 2018. Identificación, georreferenciación y caracterización morfológica de árboles superiores de *Theobroma cacao* L. 1753 cultivar Cacao Blanco de Piura, Peru. Biol. 16:105–117.

Quint A. 2003. Scalable Vector Graphics. IEEE Multimed. 10:99–102.

Quiroz V. J. 1997. Recolección de genotipos y establecimiento de un banco de germoplasma de cacao Nacional en Ecuador. Quevedo, Ecuador: INIAP - Estación Experimental Pichilingue. 12 p.

Racine J. 2000. The Cygwin tools: a GNU toolkit for windows. J Appl Econom. 15:331–341.

Raffaele S, Win J, Cano LM, Kamoun S. 2010. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. BMC Genomics. 11:637, doi:10.1186/1471-2164-11-637.

Rambaut A. 2006. FigTree: Tree Figure Drawing Tool. http://tree.bio.ed.ac.uk/.

Redhead SA, Traquair J. 1981. *Coprinus* sect. Herbicolae from Canada, notes on extralimital taxa, the taxonomic position of a low temperature Basidiomycete forage crop pathogen from western Canada. Mycotaxon. 13:373–404.

Rendiles E, Dimas A, Montero L. 2009. Estudio preliminar sobre el cultivo de cacao (*Theobroma cacao* L.) en el municipio Tucupita del estado Delta Amacuro, Venezuela. Rev UDO Agrícola. 9:268–272.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 3:217–223, doi:10.1111/j.2041-210X.2011.00169.x.

Reyes H, Capriles de Reyes L. 2000. El Cacao en Venezuela: Moderna Tecnología para su Cultivo. Caracas, Venezuela: Chocolates El Rey, C.A. 272 p.

Rice P, Longden I, Bleasby A. 2000. The European Molecular Biology Open Software Suite EMBOSS : The European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.

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.

Rizzo DM, Blanchette RA. 1992. Biosorption of metal ions by *Armillaria rhizomorphs*. Can J Bot. 70:1515–1520.

Rocha HM. 1983. The ecology of *Crinipellis perniciosa* (Stahel) Singer in Witches' brooms on cocoa (*Theobroma cacao* L.). University of London.

Rojas R, Rodríguez C, Ruiz C, Portales R, Neyra E, Patel K, Mogrovejo J, Salazar G, Hurtado J. 2017. Cacao Chuncho del Cusco. Lima, Peru: Universidad Peruana Cayetano Heredia. 116 p.

Rozenfeld AF, Arnaud-Haond S, Hernández-García E, Eguíluz VM, Matías MA, Serrao E, Duarte CM. 2007. Spectrum of genetic diversity and networks of clonal organisms. J R Soc Interface. 4:1093–1102, doi:10.1098/rsif.2007.0230.

Rubini MR, Silva-Ribeiro RT, Pomella AWV, Maki CS, Araújo WL, Santos DR Dos, Azevedo JL. 2005. Diversity of endophytic fungal community of cacao (*Theobroma cacao* L.) and biological control of *Crinipellis perniciosa*, causal agent of Witches' Broom Disease. Int J Biol Sci. 1:24–33, doi:10.7150/ijbs.1.24.

Ruiz Estrada A. 2017. Deslindes étnicos en la historia de Amazonas, Perú. Boletín Arqueol PUCP. 23:41–56.

Rutenber E, Robertus JD. 1991. Structure of Ricin B-Chain at 2.5 Å Resolution. Proteins Struct Funct Genet. 269:260–269.

Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 19:301–302.

Sattar A. 2007. ¿Indígena o Ciudadano? Republican Laws and Highland Indian Communities in Ecuador, 1820-1857. In: K. Clark, and M. Becker, eds. Highlan Indians and the State of Modern Ecuador. Pittsburgh, PA: University of Pittsburgh Press. pp. 22–36.

Sauer JD. 1993. Historical Geography of Crop Plants: A Select Roster. Boca Raton, FLorida: CRC Press. 309 p.

Saunders DGO, Win J, Cano LM, Szabo LJ, Kamoun S, Raffaele S. 2012. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. PLoS One. 7:e29847, doi:10.1371/journal.pone.0029847.

Schiavoni C, Camacaro W. 2009. The Venezuelan Effort to Build a New Food and Agriculture System. Mon Rev. 61:129–141, doi:10.14452/MR-061-03-2009-07.

Schlötterer C. 2004. The evolution of molecular markers — just a matter of fashion? Nat Rev Genet. 5:63–69.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments: A poor man's approach to genotyping for research and high-throughput diagnostics. Nat Biotechnol. 18:233–234.

Schultes RE. 1984. Amazonian cultigens and their northward and westward migration in Pre-Colombian times. Pap Peabody Museum Archeol Ethnol. 76:19–38.

Schutter K De, Damme EJM Van. 2015. Protein-Carbohydrate Interactions as Part of Plant Defense and Animal Immunity. Molecules. 20:9029–9053, doi:10.3390/molecules20059029.

Sereno ML, Albuquerque PSB, Vencovsky R, Figueira A. 2006. Genetic Diversity and Natural Population Structure of Cacao (*Theobroma cacao* L.) from the Brazilian Amazon Evaluated by Microsatellite Markers. Conserv Genet. 7:13–24, doi:10.1007/s10592-005-7568-0.

Serrano M, Coluccia F, Torres M, L'Haridon F, Métraux J. 2014. The cuticle and plant defense to pathogens. Front Plant Sci. 5:274, doi:10.3389/fpls.2014.00274.

Shephard. 1932. The cacao industry of Trinidad, some economic aspects Part III: History of the industry up to 1870. Trop Agric. 9:95–100.

Silva CRS, Albuquerque PSB, Ervedosa FR, Mota JWS, Figueira A, Sebbenn AM. 2010. Understanding the genetic diversity, spatial genetic structure and mating system at the hierarchical levels of fruits and individuals of a continuous *Theobroma cacao* population from the Brazilian Amazon. Heredity (Edinb). 106:973–985, doi:10.1038/hdy.2010.145.

Silva P. 1987. Cacau e lagartão ou vassoura-de-bruxa: registros efetuados por Alexandre Rodrigues Ferreira nos anos de 1785 a 1787 na Amazonia. Bol Técnico. 146:3–21.

Simon P. 1627. Noticias Historiales de las Conquistas de Tierra Firme en las Indias Occidentales. Bogotá, Colombia. 425 p.

Simpson EH. 1949. Measurement of Diversity. Nature. 163:688.

Singer R. 1942. A monographic study of the genera *Crinipellis* and *Chaetocalathus*. Lilloa. 8:503.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 6:1–11, doi:10.1186/1471-2105-6-31.

Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Smith NJH. 1999. The Amazon River forest: a natural history of plants, animals, and people. New York, USA: Oxford University Press. 208 p.

Solovyev V, Kosarev P, Seledsov I, Vorobyev D. 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 7:S10, doi:10.1186/gb-2006-7-s1-s10.

Soria V. J. 1970. Principal Varieties of Cocoa Cultivated in Tropical America. Cocoa Grow Bull. 15:13–21.

Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol Plant Pathol. 19:2094–2110, doi:10.1111/mpp.12682.

Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM, Taylor JM. 2016. EFFECTORP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210:743–761.

Stahel G. 1915. *Marasmius perniciosus* nov. spec. Dept Landbouw Suriname Bull. 33:1–26.

Stajich JE, Wilke SK, Ahrén D, Au CH, Birren BW, Borodovsky M, Burns C, Canbäck B, Casselton L a, Cheng CK, Deng J, Dietrich FS, Fargo DC, Farman ML, Gathman AC, Goldberg J, Guigó R, Hoegger PJ, Hooker JB, Huggins A, James TY, Kamada T, Kilaru S, Kodira C, Kües U, Kupfer D, Kwan HS, Lomsadze A, Li W, et al. 2010. Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). Proc Natl Acad Sci U S A. 107:11889–11894, doi:10.1073/pnas.1003391107.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690, doi:10.1093/bioinformatics/btl446.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313, doi:10.1093/bioinformatics/btu033.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst Biol. 57:758–771, doi:10.1080/10635150802429642.

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 7:1–11, doi:10.1186/1471-2105-7-62.

Stankis MM, Specht CA, Yang H, Giasson L, Ullrich RC, Novotny CP. 1992. The *Aα* mating locus of *Schizophyllum commune* encodes two dissimilar multiallelic homeodomain proteins. Proc Natl Acad Sci U S A. 89:7169–7173.

Stoddart JA, Taylor JF. 1988. Genotypic Diversity: Estimation and Prediction in Samples. Genetics. 118:705–711.

Striffler S. 1999. Wedded to work: Class struggles and gendered identities in the restructuring of the Ecuadorian Banana Industry. Identities Glob Stud Cult Power. 6:91–120, doi:10.1080/1070289X.1999.9962637.

Stukenbrock EH, McDonald B a. 2008. The Origins of Plant Pathogens in Agro-Ecosystems. Annu Rev Phytopathol. 46:75–100, doi:10.1146/annurev.phyto.010708.154114.

Suárez Capello C, Moreira Duque M, Vera Barahona J. 1993. Manual del Cultivo de Cacao. Quevedo, Ecuador: INIAP - Estación Experimental Pichilingue. 135 p.

Subrahmanyam S. 2012. The Portuguese Empire in Asia, 1500-1700: A Political and Economic History. Chichester, UK: John Wiley & Sons. 340 p.

Syed K, Porollo A, Lam YW, Yadav JS. 2011. A Fungal P450 (CYP5136A3) Capable of Oxidizing Polycyclic Aromatic Hydrocarbons and Endocrine Disrupting Alkylphenols : Role of Trp129 and Leu324. PLoS One. 6:e28286, doi:10.1371/journal.pone.0028286.

Syed K, Shale K, Pagadala NS, Tuszynski J. 2014. Systematic Identification and Evolutionary Analysis of Catalytically Versatile Cytochrome P450 Monooxygenase Families Enriched in Model Basidiomycete Fungi. PLoS One. 9:e86683, doi:10.1371/journal.pone.0086683.

Takahashi H. 2002. Four new species of *Crinipellis* and *Marasmius* in eastern Honshu, Japan. Mycoscience. 43:343–350, doi:10.1007/S102670200050.

Teare JM, Islam R, Flanagan R, Gallagher S, Davies MG, Grabau C. 1997. Measurement of nucleic acid concentrations using the DyNA Quant(TM) and the GeneQuant(TM). Biotechniques. 22:1170–1174.

Teixeira PJPL, Toledo Thomazella DP de, Pereira GAG. 2015. Time for Chocolate: Current Understanding and New Perspectives on Cacao Witches' Broom Disease Research. PLOS Pathog. 11:e1005130, doi:10.1371/journal.ppat.1005130.

Teneberg S, Lönnroth I, Torres López JF, Galili U, Halvarsson MÖ, Ångström J, Karlsson K-A. 1990. Molecular mimicry in the recognition of glycosphingolipids by Galα3Galβ4GlcNAcβ-binding Clostridium difficile toxin A, human natural anti α-galactosyl IgG and the monoclonal antibody Gal-13: characterization of a binding-active human glycosphingoli, non i. Glycobiology. 6:599–609.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18:1979–90, doi:10.1101/gr.081612.108.

The Gleaner. 2016. Millions to Battle Frosty Pod Rot. Glean.

Thermo Fisher Scientific. Assessment of Nucleic Acid Purity. 2, doi:10.7860/JCDR/2015/11821.5896.

Thomas E, Zonneveld M van, Loo J, Hodgkin T, Galluzzi G, Etten J van. 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, doi:10.1371/journal.pone.0047676.

Thorold CA. 1975. Disease of cocoa. Oxford, UK: Clarendon. 423 p.

Thurston UD. 1973. Threatening Plant Diseases. Annu Rev Phytopathol. 11:27–52, doi:10.1146/annurev.py.11.090173.000331.

Tiburcio RA, Costa GGL, Carazzolle MF, Mondego JMC, Schuster SC, Carlson JE, Guiltinan MJ, Bailey B a, Mieczkowski P, Meinhardt LW, Pereira GAG. 2010. Genes acquired by horizontal transfer are potentially involved in the evolution of phytopathogenicity in *Moniliophthora perniciosa* and *Moniliophthora roreri*, two of the major pathogens of cacao. J Mol Evol. 70:85–97, doi:10.1007/s00239-009-9311-9.

Toro KS, Brachmann A. 2016. The effector candidate repertoire of the arbuscular mycorrhizal fungus Rhizophagus clarus. BMC Genomics. 17:101, doi:10.1186/s12864-016-2422-y.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics. 25:1105–1111, doi:10.1093/bioinformatics/btp120.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Baren MJ Van, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 28:511–515, doi:10.1038/nbt.1621.

Tsykun T, Rellstab C, Dutech C, Sipos G, Prospero S. 2017. Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus *Armillaria cepistipes*. Heredity (Edinb). 119:371–380, doi:10.1038/hdy.2017.48.

Turner BLI. 1974. Prehistoric Intensive Agriculture in the Mayan Lowlands. Science (80-). 185:118–124.

Vaaje-Kolstad G, Horn SJ, Aalten DMF van, Synstad B, Eijsink VGH. 2005. The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. J Biol Chem. 280:28492–28497, doi:10.1074/jbc.M504468200.

Valdez F. 2013. Mayo Chinchipe: Hacia un replanteamiento del origen de las sociedades complejas en la Civilización Andina. In: F. Valdez, ed. Arqueología Amazónica: Las Civilizaciones Ocultas Del Bosque Tropical. Quito, Ecuador: Abya-Yala. pp. 107–153.

Valencia Rodríguez L. 2007. El Ecuador y la discriminación racial. AFESE Ecuador. 46:39–57.

Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllősi GJ, Szarkándi JG, Papp V, Albert L, Andreopoulos W, Angelini C, Antonín V, Barry KW, Bougher NL, Buchanan P, Buyck B, Bense V, Catcheside P, Chovatia M, Cooper J, Dämon W, Desjardin D, Finy P, Geml J, Haridas S, Hughes K, Justo A, Karasiński D, et al. 2019. Megaphylogeny resolves global patterns of mushroom evolution. Nat Ecol Evol. doi:10.1038/s41559-019-0834-1.

Varrot A, Basheer SM, Imberty A. 2013. Fungal lectins: structure, function and potential applications. Curr Opin Struct Biol. 23:678–685, doi:10.1016/j.sbi.2013.07.007.

Venturini OL. 1983. Geografía de la región de los Andes venezolanos. Caracas, Venezuela: Ariel/Seix Barral Venezolana. 287 p.

Viera JT. 1942. Lagartão ou vassoura de bruxa. Bol Da Soc Bras Agron. 5:393–400.

Vinnere O, Fatehi J, Sivasithamparam K, Gerhardson B. 2005. A new plant pathogenic sterile white basidiomycete from Australia. Eur J Immunogenet. 112:63–77, doi:10.1007/s10658-005-2191-y.

Vriesmann LC, Amboni RD de MC, Petkowicz CL de O. 2011. Cacao pod husks (*Theobroma cacao* L.): Composition and hot-water-soluble pectins. Ind Crops Prod. 34:1173–1181, doi:10.1016/j.indcrop.2011.04.004.

Wagner E. 1987. Archeologia de los Andes Venezolanos. Actual Investig. 14:5–15.

Walker T. 2007. Slave Labor and Chocolate in Brazil: The Culture of Cacao Plantations in Amazonia and Bahia (17th-19th Centuries). Food Foodways. 15:75–106, doi:10.1080/07409710701260214.

Warren JM. 1994. Isozyme variation in a number of populations of *Theobroma cacao* L. obtained through various sampling regimes. Euphytica. 72:121–126, doi:10.1007/BF00023780.

Wawra S, Fesel P, Widmer H, Timm M, Seibel J, Leson L, Kesseler L, Nostadt R, Hilbert M, Langen G, Zuccaro A. 2016. The fungal-specific β-glucan-binding lectin FGB1 alters cell-wall composition and suppresses glucan-triggered immunity in plants. Nat Commun. 7:13188, doi:10.1038/ncomms13188.

Wawrzyn GT, Quin MB, Choudhary S, López-Gallego F, Schmidt-Dannert C. 2013. Draft Genome of *Omphalotus olearius* Provides a Predictive Framework for Sesquiterpenoid Natural Product Biosynthesis in Basidiomycota. Chem Biol. 19:772–783, doi:10.1016/j.chembiol.2012.05.012.Draft.

Whitkus R, La Cruz M De, Mota-Bravo L, Gómez-Pompa A. 1998. Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. Theor Appl Genet. 96:621–627, doi:10.1007/s001220050780.

Whitmore TM, Turner BLI. 1992. Landscapes of Cultivation in Mesoamerica on the Eve of the Conquest. Ann Assoc Am Geogr. 82:402–425.

Wickramasuriya AM, Dunwell JM. 2018. Cacao biotechnology: current status and future prospects. Plant Biotechnol J. 16:4–17, doi:10.1111/pbi.12848.

Wilfinger WW, Mackey K, Chomczynski P. 1997. Effect of pH and Ionic Strength on the Spectro-photometric Assessment of Nucleic Acid Purity.

Winter HC, Mostafapour K, Goldstein IJ. 2002. The mushroom *Marasmius oreades* lectin is a blood group type B agglutinin that recognizes the Galα1,3Gal and Galα1,3Gal,β1,4GlcNAc porcine xenotransplantation epitopes with high affinity. J Biol Chem. 277:14996–15001, doi:10.1074/jbc.M200161200.

Wolters B. 1999. Verbreitung amerikanischer Nutzpflanzen auf präkolumbischen Seewegen durch Indianer. Düsseldorf. 19 p.

Wösten HAB. 2001. Hydrophobins: Multipurpose Proteins. Annu Rev Microbiol. 55:625–646.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 13:134, doi:10.1186/1471-2105-13-134.

Ye J, Zhang Y, Cui H, Liu J, Wu Y, Cheng Y, Xu H, Huang X, Li S, Zhou A, Zhang X, Bolund L, Chen Q, Wang J, Yang H, Fang L, Shi C. 2018. WEGO 2.0: a web tool for analyzing and plotting GO annotations , 2018 update. Nucleic Acids Res. 46:W71–W75, doi:10.1093/nar/gky400.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40:W445–W451, doi:10.1093/nar/gks479.

Young AM. 1994. The chocolate tree: a natural history of cacao. Washington, DC, USA: Smithsonian Institution Press. 215 p.

Zarrillo S. 2012. Human Adaptation, Food Production, and Cultural Interaction during the Formative Period in HIghland Ecuador. University of Calgary.

Zarrillo S, Gaikwad N, Lanaud C, Powis T, Viot C, Lesur I, Fouet O, Argout X, Guichoux E, Salin F, Solorzano RL, Bouchez O, Vignes H, Severts P, Hurtado J, Yepez A, Grivetti L, Blake M, Valdez F. 2018. The use and domestication of Theobroma cacao during the mid-Holocene in the upper Amazon. Nat Ecol Evol. 2:1879–1888, doi:10.1038/s41559-018-0697-x.

Zhang D, Martínez WJ, Johnson ES, Somarriba E, Phillips-Mora W, Astorga C, Mischke S, Meinhardt LW. 2012. Genetic diversity and spatial structure in a new distinct Theobroma cacao L. population in Bolivia. Genet Resour Crop Evol. 59:239–252, doi:10.1007/s10722-011-9680-y.

Zhang D, Motilal LA. 2016. Origin, Dispersal, and Current Global Distribution of Cacao Genetic Diversity. In: B.A. Bailey, and L.W. Meinhardt, eds. Cacao Diseases. Springer. pp. 3–32.

Zhao Z, Liu H, Wang C, Xu JR. 2013. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 14:274, doi:10.1186/1471-2164-15-6.