



**UNIVERSIDAD DE INVESTIGACIÓN DE  
TECNOLOGÍA EXPERIMENTAL YACHAY**

**Escuela de Ciencias Biológicas e Ingeniería**

**TÍTULO: Unraveling the phylogenomics, phylogeography, and  
demographic history of *Ralstonia solanacearum* phylotype II that  
infects *Musa spp.* across America.**

Trabajo de integración curricular presentado como requisito para la  
obtención del título de Ingeniera Biomédica

**Autor:**

Shirley Guadalupe Duchi Chimbo

**Tutor:**

José Antonio Castillo Morales

Urcuquí, Noviembre de 2024

# Autoría

Yo, **Shirley Guadalupe Duchi Chimbo**, con cédula de identidad 1725509473, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autor/a del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, Noviembre 2024.

---

Shirley Guadalupe Duchi Chimbo

CI: 1725509473

# Autorización de publicación

Yo, **Shirley Guadalupe Duchi Chimbo**, con cédula de identidad 1725509473, cedo a la Universidad de Investigación de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación

Urququí, Noviembre 2024.

---

Shirley Guadalupe Duchi Chimbo

CI: 1725509473

# Dedicatoria

A mi hermano Dominic, por ser mi mayor motivación y darle sentido a mi existencia.

A mis dos madres, Inés y Carmen, por ser mi más grande admiración.

A mi padre Augusto, por el apoyo que me ofreció sin necesidad de justificaciones.

A mi hermana Emily, por darme una razón más para ser fuerte y superarme.

A mi hermana de corazón, Camila, por ser mi mayor confidente.

A toda mi familia, por su infinito amor y constante apoyo.

A mi alma gemela, CATO, por alentarme a crecer tanto individualmente como a su lado.

A mis amigos, cuya amistad ha sido un refugio de paz y alegría en mi vida.

Shirley Guadalupe Duchi Chimbo

# Resumen

*Ralstonia solanacearum* filotipo II (Rs-pII) es el agente causante de la marchitez bacteriana en especies de *Musa* (por ejemplo, plátano y banano), comúnmente conocida como Moko. Esta bacteria fitopatógena provoca las consecuencias más devastadoras en los cultivos de plátano y banano en todo el continente americano, generando preocupaciones globales sobre su impacto en la seguridad alimentaria. La investigación genómica sobre Rs-pII es escasa. Una clara comprensión de la dinámica evolutiva de Rs-pII, causante de la enfermedad de Moko, es crucial para adoptar estratégicamente las medidas de prevención y control biológico más eficaces pero asequibles para salvaguardar la seguridad alimentaria a nivel mundial. En este estudio, se realizó una reconstrucción filogenómica, historia demográfica, e inferencia filogeográfica a partir de datos genómicos para descifrar la dinámica evolutiva de Rs-pII causante de la enfermedad de Moko. Las secuencias genómicas fueron descargadas desde la base de datos NCBI. La base de datos genómica comprende cepas de Rs-pII que infectan especies de *Musa*, muestreadas entre 1959 y 2015 en diversas ubicaciones geográficas de América. El árbol filogenético, la demografía histórica y la dinámica espacial de la bacteria fitopatógena se estimaron utilizando el software BEAST y programas relacionados. Los resultados confirman la clasificación de las cepas de Rs-pII en filotipos IIA y IIB. La historia demográfica de Rs-pII causante de la enfermedad de Moko se describe como constante durante el período histórico estudiado. Las epidemias de Rs-pII están ampliamente distribuidas en América, particularmente en América del Sur, a través de una compleja red de rutas de diseminación. En consecuencia, Brasil y Colombia destacan como puntos críticos clave para la propagación de las epidemias de la enfermedad de Moko. Adicionalmente, se encontró respaldo filogenómico y filogeográfico que sustenta el origen brasileño de las cepas de Rs-pII.

## **Palabras Clave:**

Filogenómica, Filogeografía, Dinámica evolutiva, Distribución geográfica, *Ralstonia solanacearum* filotipo II, Especies de *Musa*, América

# Abstract

*Ralstonia solanacearum* phylotype II (Rs-pII) is the causative agent of bacterial wilt in *Musa* species (e.g., banana and plantain), commonly called Moko. Giving rise to the most devastating consequences in banana and plantain crops throughout the American continent, this phytopathogenic bacteria raises global concerns about its impact on food security. The genomic research on Rs-pII is scarce. A clear understanding of the evolutionary dynamics of Moko disease-causing Rs-pII is crucial to strategically adopt the most efficient yet cost-effective prevention and control measures to safeguard global food safety. In this research, I performed phylogenomic reconstruction, demographic history, and phylogeographic inference from genomic data to unravel the evolutionary dynamics of Moko disease-causing Rs-pII. The genome sequences were downloaded from the NCBI database. The genomic database comprises *Musa*-infecting Rs-pII strains sampled at different geographic locations across America and dates between 1959 and 2015. The phylogenetic tree, historical demography, and spatial dynamics of the phytopathogenic bacteria were estimated using BEAST software and related programs. Results confirm the classification of Rs-pII strains into phylotypes IIA and IIB. The demographic history of the Moko disease-causing Rs-pII is described as constant over the studied historical period. Furthermore, the Rs-pII epidemics are widespread across America, particularly in South America, throughout a complex network of dissemination pathways. Accordingly, Brazil and Colombia stand out as key hotspots for Moko disease epidemics. Furthermore, I found phylogenomic and phylogeographic support for a Brazilian origin of the Rs-pII strains.

**Keywords:**

Phylogenomics, Phylogeography, Evolutionary dynamics, Geographic distribution, *Ralstonia solanacearum* phylotype II, *Musa* species, America

# Contents

<b>Dedicatoria</b>	<b>iii</b>
<b>Resumen</b>	<b>iv</b>
<b>Abstract</b>	<b>v</b>
<b>Contents</b>	<b>vi</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Figures</b>	<b>ix</b>
<b>Abbreviations</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Theoretical framework . . . . .	1
1.1.1 <i>Ralstonia solanacearum</i> phylotype II: Strains of the Rs species complex infecting banana and plantain . . . . .	1
1.1.2 Phylogenomics . . . . .	2
1.1.3 Phylogenomics based on Bayesian analysis . . . . .	3
1.1.4 Bayes factor . . . . .	4
1.1.5 Demographic history . . . . .	5
1.1.6 Phylogeography . . . . .	6
1.2 Problem statement . . . . .	6

1.3	Objectives . . . . .	8
1.3.1	General Objective . . . . .	8
1.3.2	Specific Objectives . . . . .	8
<b>2</b>	<b>Methodology</b>	<b>9</b>
2.1	Sequence data and LCBs-based genome alignment . . . . .	9
2.2	Nucleotide substitution model estimation . . . . .	12
2.3	Bayes factor comparison for molecular clock model selection . . . . .	12
2.4	Evolutionary reconstruction . . . . .	13
2.5	Discrete Phylogeographic reconstruction . . . . .	14
<b>3</b>	<b>Results</b>	<b>17</b>
3.1	Phylogenomic reconstruction . . . . .	17
3.2	Demographic history . . . . .	22
3.3	Discrete phylogeographic reconstruction . . . . .	23
<b>4</b>	<b>Discussion</b>	<b>26</b>
<b>5</b>	<b>Conclusions</b>	<b>32</b>
<b>6</b>	<b>Recommendations and future research</b>	<b>34</b>

# List of Tables

2.1	Genomic dataset of 33 Rs-pII strains. . . . .	10
2.2	Genomic properties of the Rs-pII strains under study. . . . .	11
2.3	Sampling location information of the Rs-pII strains under study. . . . .	16
3.1	Overview of Substitution Model Assessment. . . . .	18
3.2	Molecular clock model comparison based on the Bayes factor estimated from the marginal likelihoods. . . . .	18

# List of Figures

- 3.1 Bayesian maximum clade credibility tree for the 33 strains belonging to Rs-pII based upon concatenated nucleotide sequences of 5000 bp long LCBs. The tree is rooted at the midpoint, and branch lengths are cladogram-like. Values assigned to internal nodes indicate the posterior probability (pp) supporting the accuracy of the clade according to the model. The scale bar provides a reference for the number of nucleotide substitutions per site. The two major clusters, phylotypes IIA and IIB, are highlighted in blue and green, respectively. . . . . 19
- 3.2 Bayesian maximum clade credibility tree inferred for the phylogeographic history of *Ralstonia solanacearum* phylotype II. The tree is rooted at the midpoint, and branch lengths are cladogram-like. Values assigned to internal nodes represent the posterior probability. The scale bar provides a reference for the number of nucleotide substitutions per site. The two major clusters, phylotypes IIA and IIB, are highlighted in blue and green, respectively. . . . . 20
- 3.3 Comparison of the phylogenetic trees for the studied Rs-pII strains based on genetic data (left tree) and both genetic and geographic data (right tree). The blue-to-yellow color gradient represents the similarity of the best-matching branches between the two trees. The scale bar indicates nucleotide substitutions per site. . . 21

3.4	<p>Bayesian Skyline plot inferred from the multiple alignment of 33 Rs-pII genomes. The plot illustrates effective population size (<math>N_e</math>) as a function of time. Time reported in the x-axis is constrained by the sampling dates of the Rs-pII strains, spanning from 1959 to 2015. The solid blue line show the mean population size, while the blue shaded areas indicate the credibility interval determined by the 95% highest posterior probability density (HPD) interval. . . . .</p>	23
3.5	<p>Phylogeography of the <i>Ralstonia solanacearum</i> phylotype II infecting Musa species across America. <b>a,b,c</b>, Spatial dynamics of Rs-pII strains spreading across America. The color intensity of the circles and arrows represents the order of the transmission patterns, and the arrows indicate the direction of the migration. The red line outlines the geographical extent of the bacterial spread covered by the phylogeographic analysis.<b>d</b>, Geographic distribution of the spreading centers of the plant bacterium pathogenic to Musa species. The number of circles indicates the number of dispersal events originating from each respective location.</p>	25

# Abbreviations

**Rs-pII** *Ralstonia solanacearum* phylotype II

**RSSC** *Ralstonia solanacearum* species complex

**LCBs** Locally collinear blocks

**LCBs-bGA** Locally collinear blocks-based genome alignment

**MCMC** Markov chain Monte Carlo



# Chapter 1

## Introduction

### 1.1 Theoretical framework

#### 1.1.1 *Ralstonia solanacearum* phylotype II: Strains of the Rs species complex infecting banana and plantain

*Ralstonia solanacearum* phylotype II (Rs-pII) is a subgroup of aerobic gram-negative betaproteobacteria that belong to the *Ralstonia solanacearum* species complex, a cluster of soil-borne phytopathogens that infect a plethora of over 200 plant species, including important food crops such as those from the Solanaceae (tomato and potato), Cucurbitaceae (cucumber), Fabaceae (bean), Asteraceae (lettuce) families, among others, and causing them lethal diseases regarding vascular wilt, tissue necrosis, and root rot (Álvarez et al., 2010; Hayward, 1994; Safni et al., 2014).

The *R. solanacearum* strains that attack *Musa* species, such as banana (*Musa acuminata*) and plantain (*Musa x paradisiaca*), belong to phylotype II (Fegan and Prior, 2006). This pathogenic variant causes Moko disease, a highly destructive and persistent vascular infection that, upon invading the *Musa* plant through wounds or inherent root openings, colonizes the xylem vessels and triggers the lethal wilting of the plant, ultimately resulting in their collapse (Leonard et al., 2017; Silva et al., 2000; Woods, 1984). The pathogenicity of *R. solanacearum* Moko disease-causing strains relies on multiple secretion systems responsible for

exporting large repertoires of virulence effectors (Hikichi et al., 2007; Valls et al., 2006). These infection mechanisms regulated by gene expression and host plant metabolites collectively promote an effective systemic colonization and progression of the disease alongside the manifestation of its characteristic symptoms: fruit premature ripening, leaf yellowing, vascular and fruit discoloration, stem lesions, growth stunting, necrosis, oozing, and root and fruit rotting (de Pedro-Jové et al., 2021; Landry et al., 2020; Shen et al., 2020).

Historical reports indicate that the initial outbreak of phylotype II took place in Caribbean Central America, in Trinidad, during the 1890s and has since then spread widely and rapidly across the Americas, including Belize, Brazil, Colombia, Costa Rica, Ecuador, El Salvador, Grenada, Grenadines, Guatemala, Guyana, Honduras, Jamaica, Mexico, Nicaragua, Panama, Peru, Suriname, Trinidad and Tobago, and Venezuela. Therefore, Rs-pII prevails in tropical, subtropical, and temperate regions where temperatures range between 30 and 35 degrees Celsius (Ploetz, 2008; Tripathi et al., 2022). Some strains of phylotype II have also been reported in Asian countries such as the Philippines and Malaysia (Drenth and Kema, 2021; Horita et al., 2014; Zulperi and Sijam, 2014). These Rs strains are disseminated through insect vectors, farm animals, infested soil, contaminated planting material or water, colonized rhizospheres of non-host plants, or diseased plant debris that remains contagious for six months after bacterial infection (Kago et al., 2017).

### **1.1.2 Phylogenomics**

Following Darwin's publication of "The Origin of Species" in 1849, phylogenetics emerged as a discipline primarily focused on categorizing living organisms by grouping all descendants of a common ancestor into distinct clades. Its aim was to elucidate common attributes shared among members within each group and to deduce ancestral traits by examining observable features in extant organisms (Sleator, 2011).

The recent expansion in molecular sequence databases and tree-building methods has propelled the evolution of phylogenetics under a concept and scope that solidifies its importance across all branches of biology. Unlike conventional morphology-based approaches,

phylogenomics employs tree-building methods based on character or distance that analyze morphological data as well as genomic profiles (gene content, presence of orthologous genes, and gene organization) derived from alignments of nucleic acid- or amino acid- sequences, or both, to infer the distribution of character states (patterns of traits occurrence) and degree of similarity between organisms, ultimately constructing the evolutionary history of a species or group of organisms (Philippe et al., 2005; Semple and Steel, 2003). Besides visually representing genealogical relationships, phylogenomics is fundamental to organizing metagenomic sequences, providing insights into population histories and epidemiological dynamics, identifying genes and genomic elements, describing recent and ancient genomes, and reconstructing ancestral genomes (Wiley and Lieberman, 2011; Yang and Rannala, 2012).

Moreover, one of the most significant strengths of modern phylogenomics is that it reflects the influence of different mutational events in building evolutionary processes and relationships. That is, besides vertical gene transfer, phylogenomics analysis contemplates lateral gene transfer as a key factor in framing evolution, particularly in prokaryotic species where lateral gene transfer occurs at higher frequencies than spontaneous mutations (Lopez and Baptiste, 2009; Zhi et al., 2012).

### **1.1.3 Phylogenomics based on Bayesian analysis**

Bayesian analysis is a modern statistical framework for inferring unknown parameters in a model or evaluating hypotheses based on observed data and prior knowledge. For its purpose, the Bayesian relies on the Bayes theorem. To draw an inference about the value of a parameter from the data, the framework considers the parameter random and uses the Bayes formula that incorporates a model, a likelihood function, and a prior distribution, which is a representation of parameter uncertainty derived before observing the data from subjective beliefs or limited/no prior information, to combine prior knowledge about the parameter with observed data and calculate the conditional probability density of the parameter given the data. The conditional density refers to the posterior distribution, an updated indicator of uncertainty associated with the parameter estimated in the context of the new evidence. In the case of hypothesis testing, the

posterior odds of each relevant hypothesis are used for their comparison. (Florens et al., 2019; Ghosh et al., 2006; Iversen, 1984)

The Bayesian approach is advantageous for several real-life applications. It provides explicit solutions to problems in statistical inference that arise from high-dimensional data and complex decision scenarios, effectively handles prior knowledge and partial prior knowledge, including constraints, is grounded in sets of axioms to avoid paradoxes related to classical statistics, and validated through different techniques such as cross and frequentist validations (Kaplan and Depaoli, 2013; O'Hagan, 2004). In addition to all its benefits, the recent surge in the Bayesian framework owes much particularly to the widespread adoption of computational advancements such as Markov chain Monte Carlo, Patch, and Stepping-stone sampling methods Baele et al., 2016; Larget and Simon, 1999.

Since its introduction in the nineties, Bayesian analysis has been a cornerstone of bioinformatics (Nascimento et al., 2017). It offers a versatile and systematic approach encompassing stochastic and deterministic probabilistic models that aid in inference and decision-making regarding complex evolutionary phenomena using molecular sequence data and prior knowledge (Huelsenbeck et al., 2001; Rudge, 1998; Wilkinson, 2007). In the study of *Ralstonia solanacearum* phylotype II infecting *Musa* species throughout America, Bayesian analysis presents a robust and comprehensive framework that integrates prior knowledge and nucleotide sequences datasets on probabilities and uncertainties through probabilistic models and computational algorithms to gain reliable and deeper insights into the evolutionary and population processes underlying the dynamics of this plant pathogen, aiding the exploration of intricate biological networks, discovery of hidden patterns, and causal relationships.

#### **1.1.4 Bayes factor**

In Bayesian statistics, the Bayes factor is a pivotal method for hypothesis testing and comparing model fit for model choice. Unlike the p-value significance testing in frequentist statistics, the Bayes factor offers measurable evidence supporting the null hypothesis rather than solely the ability to reject it (Schmalz et al., 2023). The Bayes factor provides a quantitative

assessment of the strength of the evidence in favor of one hypothesis or model compared to others, enabling the selection of the model or hypothesis with the highest posterior probability from a pool of candidate models or hypotheses (Kass, 1993).

To calculate the Bayes factor, we need to compute the marginal likelihood critical. The marginal likelihood represents the evidence the data provides for a specific model (Ando, 2009). Essentially, it is the expected likelihood function considering the prior distribution, i.e., a measure of the average fit of a model to the observed data (Holder et al., 2014). Although estimating marginal likelihood accurately is computationally intensive, it is effectively addressed using numerical integration-based methods such as Stepping-Stone or Path sampling (Fan et al., 2011; Friel and Pettitt, 2008).

### **1.1.5 Demographic history**

Demographic history entails studying the fluctuations in the size of past populations and the structure and development they suffer over time, yielding significant perspectives for assessing population-genetic mechanisms and factors driving past population dynamics, interpreting interactions between populations and their environment, and tracing spreading patterns (Drummond et al., 2005; Ho and Shapiro, 2011).

Reconstructing population history involves scrutinizing the signature it has left in the genetic data of present-day representatives (Grant, 2015; Ho and Shapiro, 2011). Thus, nonparametric, semi-parametric, and parametric methods use nucleotide sequence data to estimate historical trends in population size (Mather et al., 2020). Among them, Bayesian skyline analysis is a parametric method for inferring population demographics that displays a skyline plot illustrating the historical changes in effective population size as a function of time (Salmona et al., 2019). This method applies the Bayesian framework for estimating historical effective population size that defines the demographic trajectory, providing accurate descriptions of intricate demographic scenarios (Navascués et al., 2017).

### 1.1.6 Phylogeography

Phylogeography is an interdisciplinary field that stems from the convergence of micro-level, including population genetics and demography, and macro-level, such as historical geography, paleontology, and phylogeny, evolutionary disciplines to explore relationships between phylogenetics and geography (Avice et al., 1987). Hence, this biogeography subfield studies principles and mechanisms underlying the geographical distributions of genealogical lineages (Avice, 2000). Thus, this enables the geographical illustration of the spatial relationships inherent to the gene genealogies within or among species, aiding in deducing the space-time dimensions of evolutionary history, the influence of ecological pressures, and the geospatial distribution of genetic diversity (Avice, 2009; Emerson et al., 2011).

The development of DNA sequencing technology that aroused the phylogenomics revolution also ushered in this field in the 1980s. Supported by Bayesian statistical inference, phylogeography reconstructs a reliable phylogeographic history of a species by analyzing the genealogical information expressed on independent loci from numerous genome sequences (Beaumont et al., 2002; Emerson and Hewitt, 2005).

## 1.2 Problem statement

*Ralstonia solanacearum* phylotype II causing bacterial wilt in bananas and plantains is a global concern for food safety by particularly impacting countries in the Americas that contribute not only to over 90% of global exports of bananas but also to the domestic consumption of this staple food (Food and Agriculture Organization of the United Nations (FAO), 2022). This plant pest, which adversely affects the health of *Musa* species and the development of their fruits, has the potential to compromise up to 100% of production (J. A. Alvarez et al., 2008), representing a significant threat to the quality and supply of bananas and plantains in the Americas. The profound gaps in understanding the evolutionary dynamics of Rs-pII hinder the effective implementation of management strategies specific to the biological control of Moko disease (Ailloud et al., 2015; Aoun et al., 2023; Beutler et al., 2023; French,

Sequeira, et al., 1970; N'guessan et al., 2013; Pardo et al., 2019; Vailliau and Genin, 2023). Consequently, the prevention and control methods for the entry, establishment, and spread of the bacterial phytopathogen commonly employed thus far, such as the use of non-selective herbicide and bactericide chemicals, have been largely unsuccessful and costly, while additionally involving substantial risks to human welfare and ecological integrity (Agrocalidad, 2022; Alengebawy et al., 2021; E. Alvarez et al., 2013; Peillex and Pelletier, 2020).

Beyond the scarcity of literature interested in addressing the evolution of the Rs-pII exhibiting pathogenicity to *Musa* species, the predominant focus of research involves isolated analyses of Rs-pII strains sourced exclusively from individual American countries (Albuquerque et al., 2014; Ramírez et al., 2020). Relatively few studies have undertaken research integrating Moko disease-causing Rs-pII strains isolated at least from most of the regions in the Americas where the bacteria have been detected: South America, Central America, North America, and the Caribbean. Thus, their findings are inadequate in providing a deeper insight of the pathogen and its evolutionary dynamics across the American continent. A comprehensive understanding of how Rs-pII evolved spatially and temporally in America is challenging but decisive to strategically adopt the most efficient yet cost-effective prevention and control methods to safeguard household food safety worldwide, especially in the least developed and food-deficit countries, since it may reveal Moko disease epidemiology, historical population events associated with changes in virulence or antibiotic resistance, sources of bacterial infection, routes of dissemination, and geographical areas affected and prone to be.

The present study will analyze evolutionary relationships, demographic history, and phylogeography of American Rs-pII strains infecting *Musa spp.* to provide valuable information for future epidemiological research focused on the appropriate management of the Moko disease through strategies safe for both human health and the environment.

## **1.3 Objectives**

### **1.3.1 General Objective**

Describe the evolutionary and phylogeographic dynamics of the American strains of Rs-pII causing Moko disease in Musa species.

### **1.3.2 Specific Objectives**

- Determine the best nucleotide substitution model for the genomic dataset of the studied Moko disease-causing Rs-pII strains.
- Identify the optimal molecular clock model for the American Rs strains in phylotype II through the Bayes factor computed from marginal likelihoods obtained using BEAST software.
- Analyze the phylogenetic relationships among the American Rs-pII strains using BEAST.
- Assess the demographic history of the American Rs-pII strains using BEAST.
- Analyze the dissemination patterns of Rs-pII strains throughout America making a phylogeographic analysis.

# Chapter 2

## Methodology

### 2.1 Sequence data and LCBs-based genome alignment

For this project, genome sequences corresponding to *Ralstonia solanacearum* strains from the American continent that infect any species of *Musa* were gathered by filtering the genomic database available in NCBI's database (National Library of Medicine (US), National Center for Biotechnology Information, 1988) to date (September 2023) based on host and geographic location inclusion criteria. After excluding strains associated with insufficient or questionable information for the study, such as those lacking essential sampling details or including inconsistent location data, the genomic sequences of thirty-three *Ralstonia solanacearum* strains isolated between 1959 and 2015 from banana ( $n = 12$ ), plantain ( $n = 16$ ), and other *Musa* species ( $n = 5$ ) were downloaded in FASTA (.fasta) and GENBANK (.gb) formats for subsequent analysis, see Table 2.1.

The retrieved genomic sequences of Rs-pII strains were aligned using the Mauve software package (Darling et al., 2004, 2010) with default settings. The Mauve algorithm efficiently aligned the genomes to identify locally collinear blocks that are highly conserved segments (5kb long each) among the genomes of the Rs-pII strains. Subsequently, the LCBs strip Subset script concatenated these core blocks, removing redundant genome sequences. Comprising multiple genome alignments constructed by LCBs of closely 5000 bp, the final file

Table 2.1: Genomic dataset of 33 Rs-pII strains.

Strain	GenBank accession	Host	Collection date	Geographic Location
UA-1579	GCA_003860725.1	Banana	2014	Uraba, Colombia
UA-1609	GCA_003860765.1	Banana	2014	Uraba, Colombia
UA-1611	GCA_003860685.1	Banana	2014	Magdalena, Colombia
UA-1612	GCA_003860665.1	Banana	2014	Magdalena, Colombia
CCRMRs277	GCA_014210395.1	Banana	2015	Manicore, Amazonas, Brasil
CCRMRs287	GCA_014210375.1	Banana	2015	Benjamin Constant, Amazonas, Brasil
CCRMRs304	GCA_014210335.1	Banana	2015	Fonte Boa, Amazonas, Brasil
CCRMRsB7	GCA_014210345.1	Banana	2012	Anama, Amazonas, Brasil
UW179	GCA_000825805.2	Banana	1961	Colombia
BA7	GCA_029220005.1	Banana	1984	Grenada
UW98	GCA_028464675.1	Banana	1960	Costa Rica
UW158	GCA_028464865.1	Banana	1962	Peru
UA-1591	GCA_003860705.1	Plantain	2014	Quindio, Colombia
UA-1617	GCA_003860745.1	Plantain	2014	Quindio, Colombia
CIAT_078	GCA_012562465.1	Plantain	2004	Colombia
UW163	GCA_001587135.1	Plantain	1967	Nauta, Peru
UW128	GCA_023646455.1	Plantain	1962	Peru
CFBP1416	GCA_000825925.2	Plantain	1967	Costa Rica
UW156	GCA_023077575.1	Plantain	1962	Peru
CIP417.UW70	GCA_021117115.1	Plantain	1960	Colombia
UW162	GCA_023077495.1	Plantain	1963	Peru
UW175	GCA_023077475.1	Plantain	1961	Colombia
UW132	GCA_028464965.1	Plantain	1962	Peru
UW140	GCA_028464615.1	Plantain	1963	Costa Rica
UW176	GCA_028464775.1	Plantain	1961	Colombia
UW177	GCA_028464765.1	Plantain	1961	Colombia
UW190	GCA_028464665.1	Plantain	1964	Peru
UW192	GCA_028464745.1	Plantain	1964	Peru
IBSBF 2570	GCA_003590585.2	Musa sp.	2012	Sergipe, Brasil
SFC	GCA_003590625.2	Musa sp.	2012	Sergipe, Brasil
IBSBF 2571	GCA_003590605.1	Musa sp.	2010	Sergipe, Brasil
PD:1445	GCA_015912065.1	Musa sp.	1959	Panama
IBSBF1900	GCA_001373275.1	Musa sp.	2000	Brasil

sp., unspecified species.

in XMFA format was converted into FASTA format through a Perl script (Katz, 2014). After aligning all the genomes in just under 1,5 million nucleotides, the size of each of the sequences represented approximately 24% of its original size (see Table 2.2).

Table 2.2 also includes further details about the genome sequencing level retrieved from NCBI about the sequences subjected to this study.

Table 2.2: Genomic properties of the Rs-pII strains under study.

Strain	Phylotype- Sequevar	Assembly Level	Initial genome size (Mbp)	GC (%)	Genes	Protein- coding genes	Size after LCBs-bGA (bp)
UA-1579	IIB-4	Chromosome	5,08	67,0	4597	4373	1317221
UA-1609	IIB-4	Chromosome	5,07	67,0	4591	4357	1317225
UA-1611	IIA-6	Chromosome	5,20	67,0	4731	4436	1317675
UA-1612	IIA-6	Chromosome	5,00	67,0	4644	4266	1317675
CCRMRs277	IIA-24	Scaffold	5,63	67,0	4453	4264	1317752
CCRMRs287	IIB-4	Scaffold	5,42	66,5	4800	4628	1317243
CCRMRs304	IIA-24	Scaffold	5,63	67,0	4432	4242	1317757
CCRMRsB7	IIB-25	Scaffold	5,85	66,5	5098	4958	1317502
UW179	IIB-4	Scaffold	5,43	66,5	5072	4761	1317273
UA-1591	IIB-4	Chromosome	5,35	66,5	4819	4612	1317254
UA-1617	IIB-4	Chromosome	5,36	66,5	4825	4634	1317271
CIAT.078	IIB-4	Complete genome	5,39	66,5	4915	4548	1317067
UW163	IIB-4	Complete genome	5,60	66,5	5020	4840	1317242
UW128	IIB-4	Scaffold	5,49	66,5	4936	4720	1317242
CFBP1416	IIB-3	Scaffold	5,74	66,5	5448	5096	1317547
UW156	IIB-ND	Contig	5,66	66,5	5067	4882	1317235
CIP417_UW70	IIB-ND	Chromosome	5,49	66,5	4910	4735	1317270
UW162	IIB-4	Contig	5,73	66,5	5148	4956	1317230
UW175	IIB-ND	Contig	5,37	66,5	4839	4654	1317271
IBSBF 2570	IIA-53	Chromosome	5,72	67,0	4589	4402	1317641
SFC	IIA-53	Chromosome	5,71	67,0	4682	4476	1317637
IBSBF 2571	IIA-53	Complete genome	5,43	66,5	4862	4641	1317222
PD:1445	IIB-ND	Contig	5,47	66,5	5023	4793	1317579
IBSBF1900	IIA-24	Scaffold	5,81	66,5	5540	5070	1317754
BA7	IIA-6	Chromosome	5,75	66,5	5115	4871	1317678
UW98	IIB-3	Scaffold	5,65	66,5	5172	4963	1317578
UW158	IIB-4	Contig	5,46	66,5	4944	4761	1317249
UW132	IIB-4	Scaffold	5,48	66,5	4938	4761	1317243
UW140	IIB-3	Scaffold	5,59	66,5	5147	4925	1317585
UW176	IIB-4	Scaffold	5,37	66,5	4831	4655	1317268
UW177	IIB-4	Scaffold	5,37	66,5	4823	4649	1317267
UW190	IIB-4	Scaffold	5,49	66,5	4950	4776	1317242
UW192	IIB-4	Scaffold	5,47	66,5	4942	4759	1317241

ND, no data available.

## 2.2 Nucleotide substitution model estimation

Among other parameters, evolutionary analysis performed using the methods mentioned below requires a priori knowledge of the best-fit nucleotide substitution model.

The most appropriate DNA substitution model for the dataset subject to analysis was computed through IQ-TREE v1.6.12. software (Nguyen et al., 2015). The -m TEST standard algorithm implemented in the software conducted tree reconstruction and non-parametric bootstrap analysis to estimate the most fitting substitution model and related parameters for the concatenated LCBs of the Rs-pII strains.

## 2.3 Bayes factor comparison for molecular clock model selection

The aligned genomic sequences of Rs-pII strains were tested for the optimal molecular clock model selection by evaluating the fit of the data to the strict clock (Zuckerkandl and Pauling, 1965) or to the uncorrelated lognormal relaxed clock (Drummond et al., 2006) through calculating twice the natural logarithm of the Bayes factor (Jeffreys, 1957) using formula 2.1 and from the log marginal likelihoods computed using path sampling (PS) and stepping-stone sampling (SS) (Baele, Lemey, et al., 2012; Baele, Li, et al., 2012) included in BEAST v1.10.4 package (Bouckaert et al., 2014).

$$2\log BF = 2 \times (\log ML_{RC} - \log ML_{SC}) \quad (2.1)$$

Where:

- BF represents the Bayes factor for the strict clock model against the uncorrelated lognormal relaxed clock model.
- $ML_{RC}$  is the marginal likelihood for the uncorrelated log-normal relaxed clock model.
- $ML_{SC}$  is the marginal likelihood for the strict clock model.

To calculate the logarithm of the marginal likelihood associated with the uncorrelated relaxed clock model, the Bayesian phylogenetic inference was set up in BEAST software under the general time reversible substitution model, gamma plus invariant sites as site heterogeneity model, and coalescent constant population tree model (Drummond et al., 2002; Kingman, 1982). Identical settings were employed to compute the marginal likelihood logarithm related to the strict clock model.

## 2.4 Evolutionary reconstruction

The species tree and demographic skyline plot for evolutionary inference were estimated under the Bayesian framework using BEAST v2.6.7 (Bouckaert et al., 2014). BEAST was run together with the Broad-platform Evolutionary Analysis General Likelihood Evaluator (BEAGLE) v4.0.0 (Ayres et al., 2012; Suchard and Rambaut, 2009) to accelerate the highly complex phylogenetic analysis and optimize computational resources.

Based on the previous results, the Bayesian phylogenetic analysis was set up in BEAUti v2.6.7 to construct phylogenetic relationships from the aligned genomic sequence dataset under the GTR + F + I nucleotide substitution model, the number of categories for the gamma site heterogeneity was equal to 4, and tip dates were assigned to each sequence regarding known sampling dates to set up clock under strict model because analyses done to select the best clock model indicate that the strict clock fits the data well. Additionally, the tree prior was set up to reflect a constant population growth, and the substitution rate of partition was set up to Log Normal distribution with an M parameter of  $10^{-7}$ . The Markov chain Monte Carlo (MCMC) analysis was run in BEAST for 30 million generations, sampling at intervals of 2000 generations. All parameters' effective sample size (ESS) was assessed to be above 200 by uploading the .log file in Tracer v1.7.2 (Rambaut et al., 2018) to ensure the convergence of phylogenetic tree topologies and other parameters derived from the MCMC analysis (Ho, 2012; Lanfear et al., 2016). The summary tree was generated from the sampled trees with the maximum clade credibility tree after discarding 10% of the generations as burnin using the

.trees file in TreeAnnotator v2.6.7 (Drummond and Rambaut, 2007). Node heights were estimated as common ancestor heights. FigTree v1.4.4 (Rambaut, 2018) was used to annotate and display the phylogenetic tree.

Reconstructing the demographic history and population dynamics on BEAST v2.6.7 used coalescent Bayesian skyline for the tree prior and similar settings than for phylogeny inference. The Bayesian skyline plot was visualized on Tracer v1.7.2 by uploading the trees file.

## **2.5 Discrete Phylogeographic reconstruction**

The discrete phylogeographic migration of the American Rs-pII strains was assessed using the Bayesian MCMC method implemented on the software BEAST v2.7.5 (Bouckaert et al., 2019). To this end, the sampling geographical information of every Rs strain was manually added in Bioedit v7.7.1 (Hall et al., 2011). Using BEAUti from the BEAST package, a Bayesian phylogeographic inference framework was configured with a strict clock and constant-size coalescent model for tree prior. Within this framework, sampling dates were included as tip dates and sampling locations as a discrete partition. Furthermore, the proportion of invariant sites, frequency, and GTR substitution model priors were set up to the values summarized in Table 3.1, the gamma site heterogeneity comprised four categories, and the gamma shape parameter was modeled using a gamma distribution. The prior distribution used to model nucleotide frequencies and the molecular clock rate was specified as Log Normal. MCMC analysis was run in BEAST for 150 million iterations, subsampling every 2000 iterations. The MCMC convergence was evaluated in Tracer v1.7.2 (Rambaut et al., 2018). Results were satisfactory since the ESS for all parameters was over 200. The maximum clade credibility tree was generated in TreeAnnotator by summarizing the sampled trees after excluding the initial 10% of them as burn-in. The discrete trait phylogeography was analyzed through Spatial Phylogenetics Reconstruction of Evolutionary Dynamics using Data-Driven Documents (spread3) v0.9.7.1 (Bielejec et al., 2016). Spread3 converted the .trees file containing the maximum clade credibility tree, the .dat file with the geographic coordinates

associated with the sampling locations presented in Table 2.3, and the geoJSON file for the American region (<https://groups.google.com/g/d3-js/c/3iIL32UqAkk?pli=1>) to a JavaScript Object Notation (JSON) file. The software used the output JSON file in rendering, generating phylogeographic visualizations in a Keyhole Markup Language (KML) file suitable for displaying the resulting spreading network in Google Earth Pro (GoogleLLC, 2018).

Table 2.3: Sampling location information of the Rs-pII strains under study.

Strain	Geographic location	Latitude	Longitude
UA-1579	Uraba, Colombia	6,828350	-75,513070
UA-1609	Uraba, Colombia	6,828350	-75,513070
UA-1611	Magdalena, Colombia	10,411301	-74,405661
UA-1612	Magdalena, Colombia	10,411301	-74,405661
CCRMRs277	Manicore, Amazonas, Brazil	-5,813187	-61,298990
CCRMRs287	Benjamin Constant, Amazonas, Brazil	-4,376009	-70,030181
CCRMRs304	Fonte Boa, Amazonas, Brazil	-2,515165	-66,095897
CCRMRsB7	Anama, Amazonas, Brazil	-2,068587	-66,000061
UW179	Colombia	0,819321	-73,791619
UA-1591	Quindio, Colombia	4,461019	-75,667356
UA-1617	Quindio, Colombia	4,461019	-75,667356
CIAT_078	Colombia	0,819321	-73,791619
UW163	Nauta, Peru	-4,505359	-73,580987
UW128	Peru	-9,189967	-75,015151
CFBP1416	Costa Rica	9,748917	-83,753428
UW156	Peru	-9,189967	-75,015151
CIP417_UW70	Colombia	0,819321	-73,791619
UW162	Peru	-9,189967	-75,015151
UW175	Colombia	0,819321	-73,791619
IBSBF 2570	Sergipe, Brazil	-11,14577	-37,348306
SFC	Sergipe, Brazil	-11,14577	-37,348306
IBSBF 2571	Sergipe, Brazil	-11,14577	-37,348306
PD:1445	Panama	8,537981	-80,782127
IBSBF1900	Brazil	-14,235001	-51,925279
BA7	Grenada	12,116500	-61,679000
UW98	Costa Rica	9,748917	-83,753428
UW158	Peru	-9,189967	-75,015151
UW132	Peru	-9,189967	-75,015151
UW140	Costa Rica	9,748917	-83,753428
UW176	Colombia	0,819321	-73,791619
UW177	Colombia	0,819321	-73,791619
UW190	Peru	-9,189967	-75,015151
UW192	Peru	-9,189967	-75,015151

# Chapter 3

## Results

### 3.1 Phylogenomic reconstruction

The genetic makeup within bacteria differs significantly, even between closely related taxa (Lawrence and Hendrickson, 2005). Studies demonstrate that bacterial adaptive evolution is mostly shaped by horizontal gene transfer, along with other evolutionary forces such as mutation, contributing to genetic variability. Horizontal genomic exchange, which is more common in bacteria than in eukaryotic organisms, leaves a distinct signature in the genome (Arnold et al., 2022; Rodríguez-Beltrán et al., 2021).

BEAST is a potent and versatile analysis package for estimating evolutionary parameters and testing hypotheses (Drummond and Rambaut, 2007). It implements Bayesian inference, a plethora of prior distributions, the Markov Chain Monte Carlo (MCMC) sampling, stochastic models of molecular sequence evolution, and highly parametric coalescent analysis to address discordance among inferred gene trees resulting from biological phenomena, such as horizontal gene transfer, and thus to reconstruct accurate and reproducible phylogenetic trees from diverse sequence datasets (Castro-Nallar et al., 2015; Daubin et al., 2003).

Bayesian inference in BEAST for all the analyses conducted in this research required the nucleotide substitution and molecular clock models that best describe the evolutionary processes underlying the genomic data.

The model selection performed by IQ-TREE identified the general time reversible (GTR) model with gamma plus invariant sites heterogeneity (+ F + I) as the most fitting nucleotide substitution model for the genomic data of the Rs-pII strains. Additional parameters related to the substitution model are listed in table 3.1.

Table 3.1: Overview of Substitution Model Assessment.

Parameter	Value
Model for the base substitution rates	GTR
Type of base frequencies	+ F
Model of the rate heterogeneity across sites	+ I
Proportion of invariable sites	0,8964
A-C rate	1,0509
A-G rate	8,6336
A-T rate	0,5541
C-G rate	0,8000
C-T rate	6,9449
G-T rate	1,0000

The model comparison based on the Bayes factor (table 3.2) provided very strong evidence supporting the strict clock model over the uncorrelated log normal relaxed clock model as the best molecular clock model for the American Rs-pII strains.

Table 3.2: Molecular clock model comparison based on the Bayes factor estimated from the marginal likelihoods.

Clock model	Log Marginal Likelihood	
	PS	SS
Uncorrelated log normal relaxed clock	-2233472,39	-2232434,50
Strict clock	-2220501,82	-2220496,17
$2 \log BF$	-25941,14	-23876,66

The Bayesian maximum-credibility phylogenies shown in figures 3.1 and 3.2 were obtained from the aligned and concatenated dataset (chromosome, contig, scaffold, and complete genomic sequences) that includes 33 isolates of Rs-pII through trial-and-error approach in BEAST configured with GTR + F + I (general time-reversible with nucleotide frequencies and proportion of invariant sites) nucleotide substitution model and the best-fit strict clock.

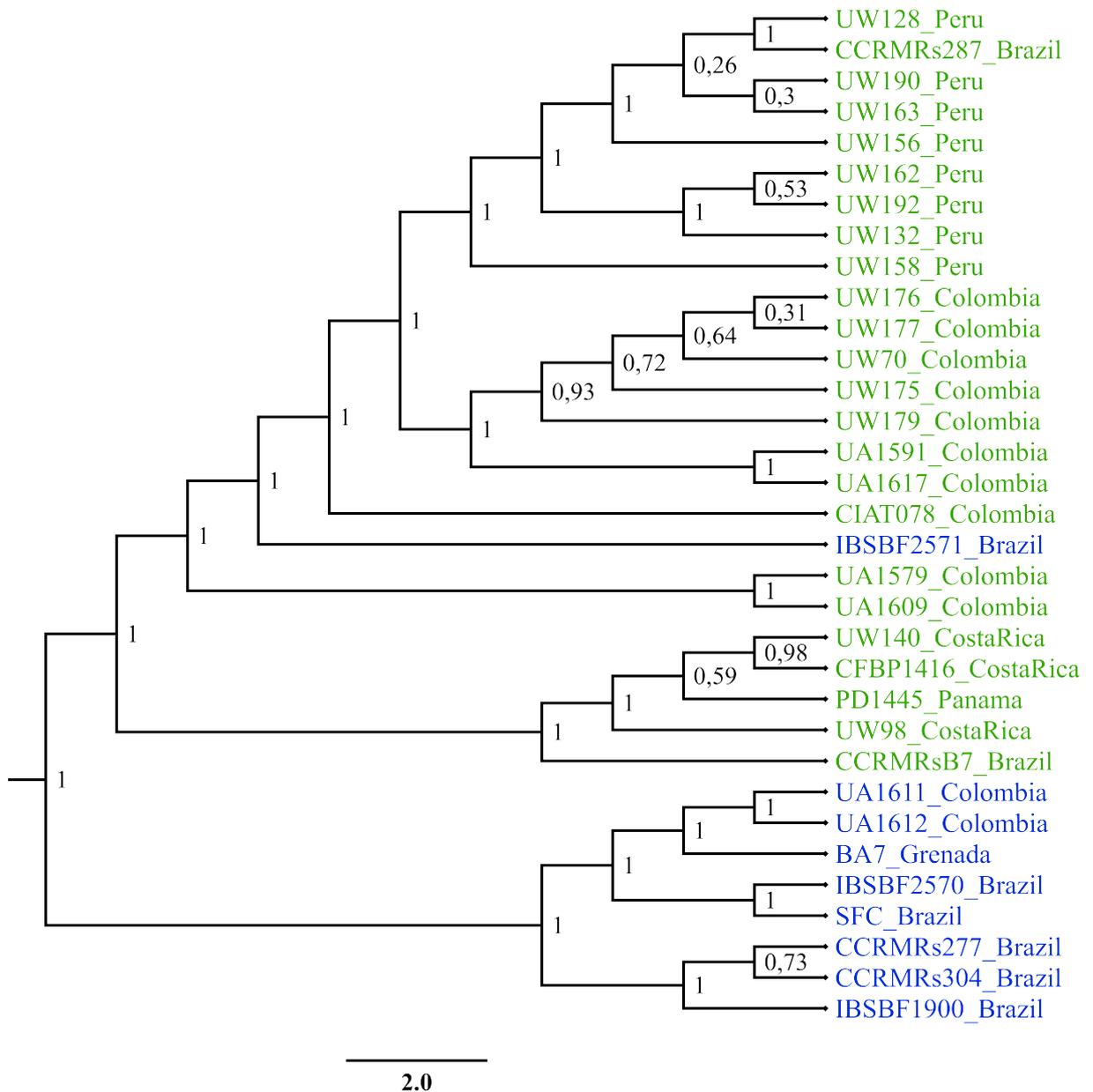


Figure 3.1: Bayesian maximum clade credibility tree for the 33 strains belonging to Rs-pII based upon concatenated nucleotide sequences of 5000 bp long LCBs. The tree is rooted at the midpoint, and branch lengths are cladogram-like. Values assigned to internal nodes indicate the posterior probability (pp) supporting the accuracy of the clade according to the model. The scale bar provides a reference for the number of nucleotide substitutions per site. The two major clusters, phylotypes IIA and IIB, are highlighted in blue and green, respectively.

However, the tree in figure 3.2 was obtained from the Bayesian framework used for the phylogeographic reconstruction, thereby incorporating both genomic and geographic information of the Rs-pII isolates. Both phylogenetic trees are summary trees derived from sampled trees verified to have all parameters with an effective sample size (ESS) exceeding 200.

When compared using beta.phylo.io (Robinson et al., 2016), the trees show a significant (Robinson-Foulds distance = 97%) agreement in their topology (Figure 3.3).

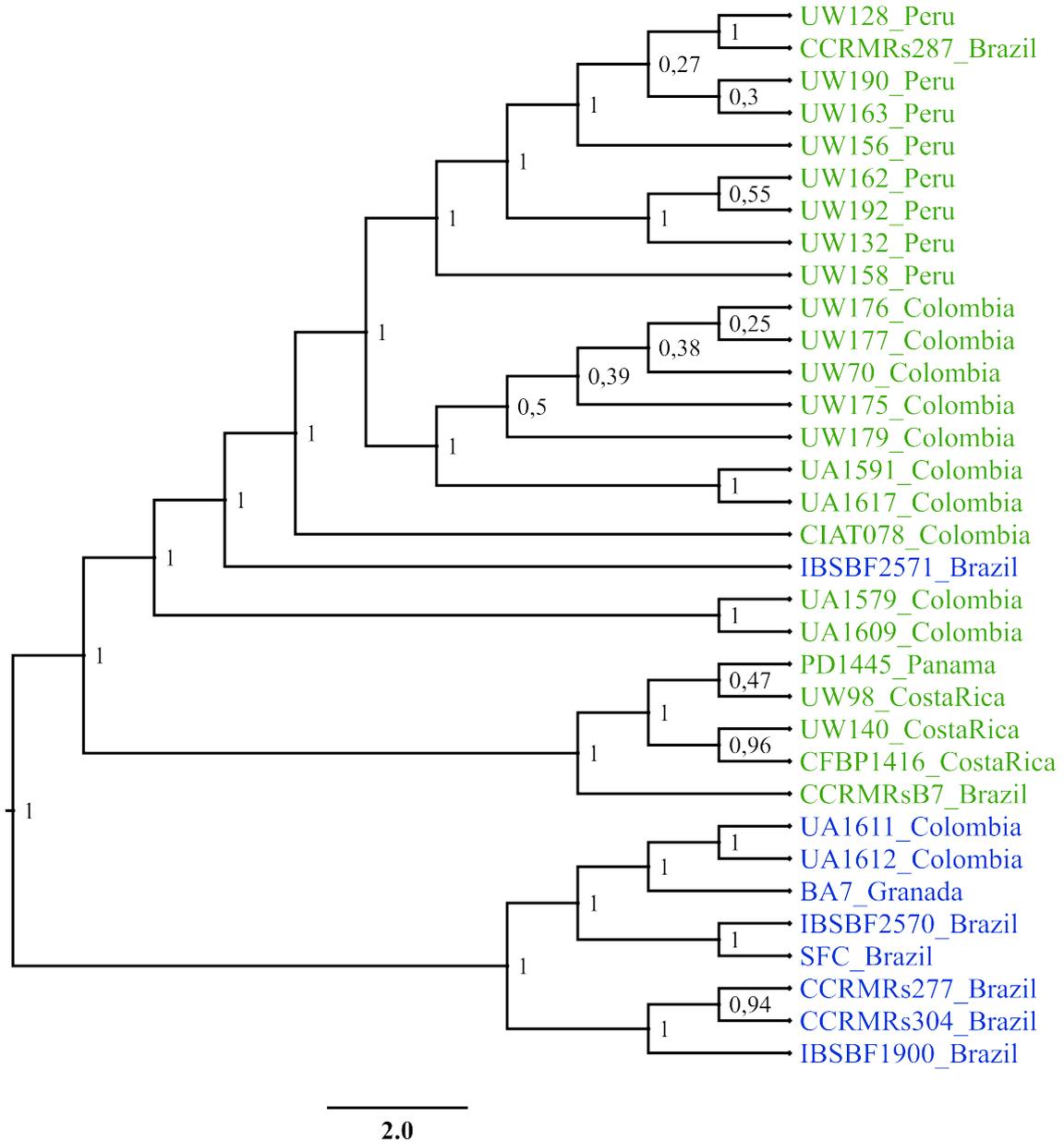


Figure 3.2: Bayesian maximum clade credibility tree inferred for the phylogeographic history of *Ralstonia solanacearum* phylotype II. The tree is rooted at the midpoint, and branch lengths are cladogram-like. Values assigned to internal nodes represent the posterior probability. The scale bar provides a reference for the number of nucleotide substitutions per site. The two major clusters, phylotypes IIA and IIB, are highlighted in blue and green, respectively.

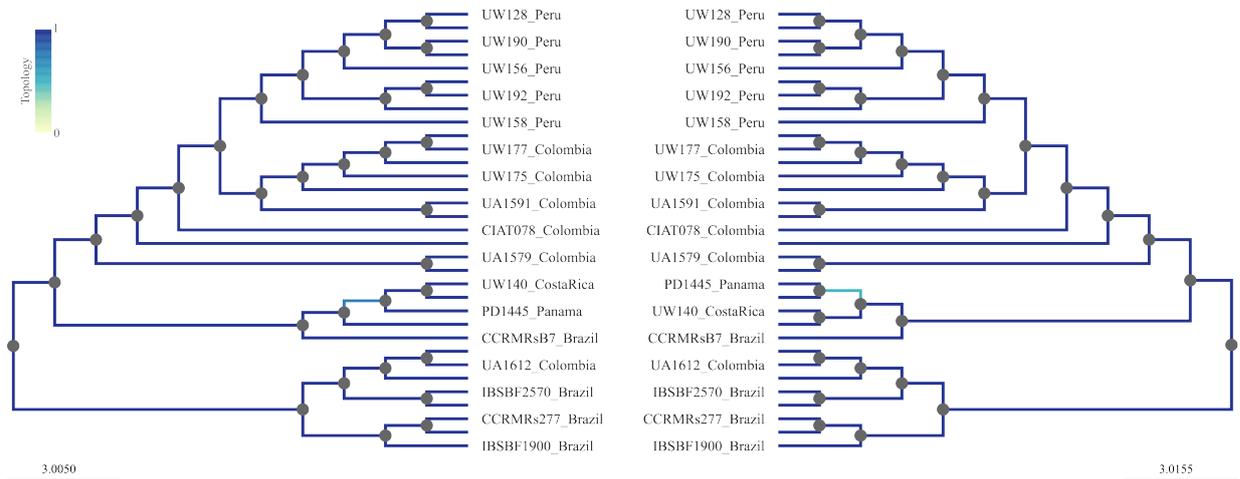


Figure 3.3: Comparison of the phylogenetic trees for the studied Rs-pII strains based on genetic data (left tree) and both genetic and geographic data (right tree). The blue-to-yellow color gradient represents the similarity of the best-matching branches between the two trees. The scale bar indicates nucleotide substitutions per site.

The phylogenomic analysis clearly reveals two statistically significant (*posterior probability* ( $pp$ )  $\approx 1$ ) monophyletic clusters within the Rs-pII population that share a most recent common ancestor (Fig. 3.1). Comprising 24% of the isolates ( $n = 8$ ), the first clade exclusively consists of strains belonging to phylotype IIA. These strains were isolated from Colombia, Grenada, and Brazil. The second clade contains the remaining isolates ( $n = 25$ ). Except for IBSBF2571 strain that was inaccurately reported by Albuquerque et al., 2014 to correspond to phylotype IIA, this clade predominantly includes strains of phylotype IIB isolated from Peru, Colombia, Costa Rica, Panama, and Brazil. In particular, the basal branches of the clade associated with phylotype IIB are occupied by Brazilian (IBSBF2571) and Colombian (CIAT078) strains. Based on the length of their branches measured in nucleotide substitutions per site, these South American isolates exhibit relatively large genetic distances compared to the strains of this clade.

Moreover, some highly significant ( $pp \approx 1$ ) subclades were strongly defined based on geographical criteria: Four South American subclades and one Central American subclade. The Peruvian subgroup primarily included phylotype IIB isolates from Peru, except for a sole strain (CCRMrs287) isolated from Brazil. The two pure Colombian subgroups exclusively consisted of phylotype IIB isolates from Colombia. The Brazilian subgroup comprised a mixed collection

of South American strains of phylotype IIA from Brazil and Colombia, except for one strain (BA7) from the Caribbean, Grenada. Likewise, the Costa Rican subgroup included mixed isolates, phylotype IIB strains predominantly from Costa Rica, and a few others from Panama, and Brazil. It is important to highlight that the Brazilian strain (CCRMrsB7) is located on the most basal branch of this subclade.

## 3.2 Demographic history

BEAST can perform Bayesian analysis to infer demographic history, which involves the changes in population size over time. To this end, it includes Bayesian skyline reconstruction that plots population size fluctuations across historical periods by simultaneously estimating parameters associated with genealogy, population history, and substitution model to overcome the phylogenetic and coalescent uncertainties present in the inferred genealogy due to the intraspecific nature of the molecular data (Ho and Shapiro, 2011).

The Bayesian skyline plot in figure 3.4 denotes the demographic history concerning the Rs-pII that infects *Musa* species. This figure reveals that the population of Rs-pII pathogenic to *Musa* species sustains a relatively constant size of roughly 1000  $N_e$  (Number of individuals or genetic units) between 1959 and 2015. Nonetheless, it is noteworthy that the width of the confidence interval is initially wider and becomes narrower over time, owing to the fact that phylogenetic uncertainty associated with the population size is higher for more remote periods because of the inherent variability in the underlying data or the availability of sufficient ancient representatives.

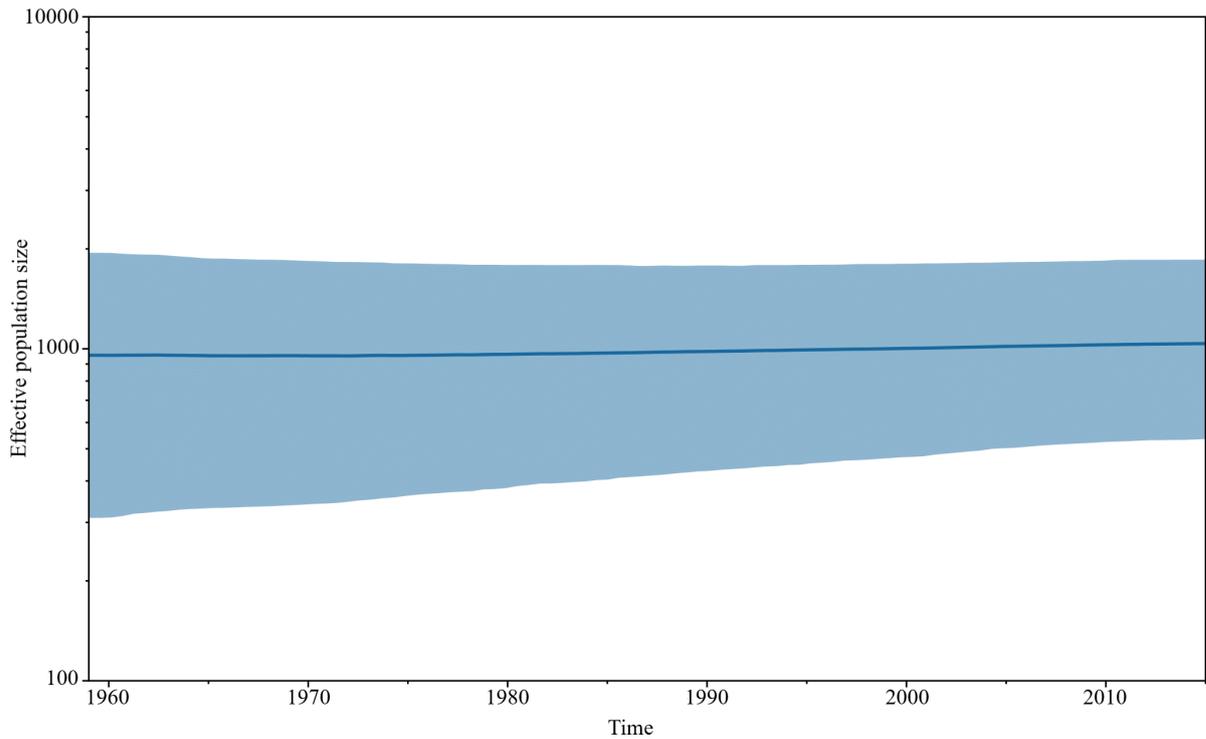


Figure 3.4: Bayesian Skyline plot inferred from the multiple alignment of 33 Rs-pII genomes. The plot illustrates effective population size ( $N_e$ ) as a function of time. Time reported in the x-axis is constrained by the sampling dates of the Rs-pII strains, spanning from 1959 to 2015. The solid blue line show the mean population size, while the blue shaded areas indicate the credibility interval determined by the 95% highest posterior probability density (HPD) interval.

### 3.3 Discrete phylogeographic reconstruction

Combining BEAST and spread3 brings out a powerful approach for pathogen phylogeographic reconstructions. BEAST integrates genetic sequence data and discrete or continuous geographic location information to model the spatio-temporal dispersal of evolutionary lineages. Subsequently, spread3 uses the resulting phylogenies with location annotation for visualizing the estimates of pathogen spread based on phylogenetic data on a geographic map (Nahata et al., 2022).

Figure 3.5 derives from the Bayesian phylogeographic approach described above. This phylogeography was co-estimated alongside the maximum-credibility genealogical tree displayed in figure 3.2 by applying a strict molecular clock and GTR + F + I nucleotide substitution model under constant-size coalescent model and using the genomic dataset along

with the sampling locations incorporated as a discrete partition.

The Rs-pII strains responsible for Moko disease in *Musa* species used for diverse studies primarily come from America (Fegan and Prior, 2006). Hypothesizing that the origin of the phytopathogen tracks to America, our phylogeographic analysis included twenty-eight isolates from South America (Colombia, Brazil, and Peru), four from Central America (Costa Rica and Panama), and one from the Caribbean (Grenada). Although outbreaks of the pathogen have been recorded in all banana-producing American countries, Rs-pII strains isolated from other regions of America, such as Mexico from North America (Obrador-Sánchez et al., 2017) or Ecuador from South America (Delgado et al., 2014), were not included in the analysis due to the unavailability of their genomic information in open-access databases.

The phylogeographic analysis identified a total of 16 intracontinental migration pathways of the Rs-pII exhibiting pathogenicity to *Musa* species throughout America (Figs 3.5 a,b,c). Figure 3.5 a shows that the diffusion of the phytopathogen began in Brazil. The pathogen initially spread within the country, from Anama toward Manicore. Concurrently, it spread northward to Magdalena, Colombia. Spreading from Magdalena to Grenada, the phytopathogen reached the Caribbean. Subsequently, the plant pathogen spread from Brazil (Manicore) to Uraba, Colombia, where transmission of the phytopathogen extended into other Colombian regions. Figure 3.5 b reveals that the dissemination of Rs-pII strains persisted mainly in South America, spreading from Colombia to Peru and Brazil. Concomitantly, dispersal pathways headed toward Central America. The propagation of the phytopathogen extended from Brazil (Anama) to Costa Rica and later from Costa Rica to Panama. Figure 3.5 c shows that the following migration paths arose in parallel from Grenada and Colombia to Sergipe in Brazil, and from Peru to Benjamin Constant, Brazil. Upon returning to Brazil, the phytopathogen continued to propagate across the Brazilian Amazon and other several regions north of Brazil, such as Fonte Boa. Meanwhile, internal spread persisted within Colombia as the pathogen disseminated to Quindio.

As a result of the dispersal patterns, Figure 3.5 d reveals that Colombia and Brazil were the most commonly associated countries as focal source and recipient locations, followed by

Peru, Costa Rica, and Grenada. Anama and Manicore from Brazil and Uraba from Colombia seemed to play a key role in seeding the spread of Rs-pII strains throughout their respective countries. Brazil (especially Anama and Manicore), Colombia (particularly Magdalena), Peru, and Grenada may contribute significantly to the Rs-pII epidemics across South America since they appear to be potential areas for the dissemination of the phytopathogen to neighboring nations within the region. Furthermore, Anama from Brazil and Magdalena from Colombia might be the hot spots of the diffusion of Rs-pII strains toward Central America and the Caribbean, respectively.

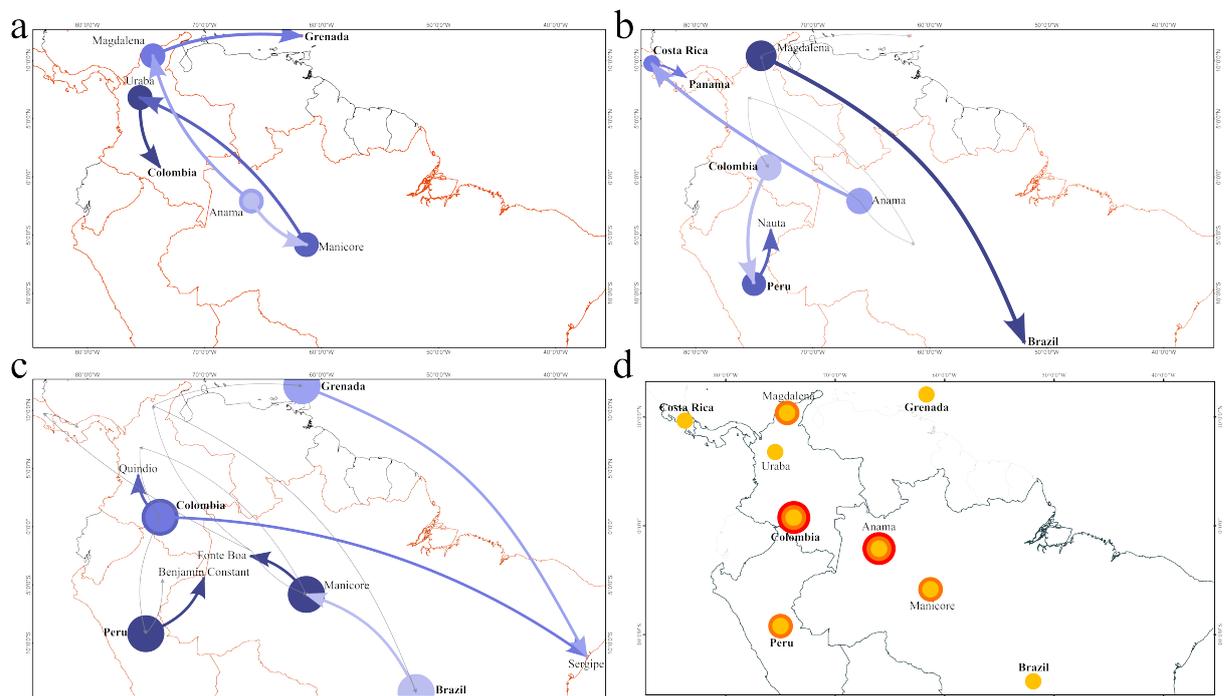


Figure 3.5: Phylogeography of the *Ralstonia solanacearum* phylotype II infecting *Musa* species across America. **a,b,c**, Spatial dynamics of Rs-pII strains spreading across America. The color intensity of the circles and arrows represents the order of the transmission patterns, and the arrows indicate the direction of the migration. The red line outlines the geographical extent of the bacterial spread covered by the phylogeographic analysis. **d**, Geographic distribution of the spreading centers of the plant bacterium pathogenic to *Musa* species. The number of circles indicates the number of dispersal events originating from each respective location.

# Chapter 4

## Discussion

In this study, the phylogenomics, demographic history, and phylogeography concerning the Moko disease-causing Rs-pII were reconstructed by integrating genomic and geographic data through advanced computational methods with the intention of gaining new insights into the evolutionary relationships, historical population dynamics, spatial distribution, geographic origin, and factors and processes driving the evolutionary trajectory of the Rs-pII population pathogenic to *Musa* species.

The phylogenetic trees independently constructed with and without geographic data shared similar branching patterns, differing in the placement of the Panamian (PD1445) and Costa Rican (UW98, UW140, CFBP1416) strains. Notably, these strains were consistently classified in both trees into the same clade, phylotype IIB, indicating that minor discrepancies in phylogenetic relationships were inconsequential. Nevertheless, despite the strong topological congruence, the genealogical tree presented in figure 3.1 exhibited more substantial support, as demonstrated by posterior probability scores nearing 1 for most nodes. Henceforth, the tree in figure 3.1 was used to define the clades and elucidate the evolutionary relationships among Rs-pII specimens.

Moreover, the topological consistency of the genealogy trees suggests that the phylogenetic inference for Rs-pII strains is solid and independent of geographic criteria. However, clustering clades comprising Moko disease-causing bacterial strains isolated from a

particular American region hints at the involvement of geographical factors in shaping Rs-pII evolutionary relationships. The congruence between the classification of RSSC phylotypes and their respective geographic sampling location proposed in previous phylogenetic research (Fegan, Prior, et al., 2005) supports this conjecture.

Furthermore, phylogenetic findings are consistent with current phylogeny of the RSSC, confirming the classification of Moko disease-causing Rs-pII into two discernible clusters denoted as IIA and IIB (Castillo and Greenberg, 2007; Etmiani et al., 2020; Pais et al., 2022; Paudel et al., 2020).

Within the clade of Rs-pIIB strains, the Colombian and Brazilian strains linked to the most basal branches may be regarded as more ancestral relative to all other Rs-pIIB isolates. Thus, South America likely represents the geographic origin of the most recent common ancestor (MRCA) of the Rs-pIIB pathogenic to *Musa* species (Clarke et al., 2015; Goss et al., 2014), which is in concordance with literature (Albuquerque et al., 2014; Blomme et al., 2017; Buddenhagen, 2007). Moreover, the relative genetic distances indicate a higher relative diversity in Brazil than in Colombia (Stukenbrock and McDonald, 2008), suggesting that the MRCA of the phylotype IIB of Moko disease-causing *Rs* bacteria, especially of the Peruvian and Colombian strains, is originally from Brazil. Similarly, given its position in the phylogenetic tree as the basal branch of the Costa Rican subgroup, Brazil possibly represents the geographic origin of the ancestor of Costa Rican Rs-pIIB strains.

Furthermore, our data indicate that both phylotypes IIA and IIB of *Ralstonia solanacearum* co-exist in some South American countries, especially Colombia and Brazil. This direct interaction of distinct Rs-pII lineages within the same geographical region might suggest sympatry, a feasible evolutionary scenario for bacteria (Cordero and Polz, 2014). Consistently, previous studies indicate that the open nature of its pangenome (Castillo, 2023) and its wide geographical dispersion (Wicker et al., 2012) support that sympatric lifestyles might have influenced the speciation of *Ralstonia solanacearum* species complex (RSSC). It is imperative that phytosanitary authorities recognize that the prevalence of the two Rs-pII phylotypes, A and B, in countries like Colombia and Brazil must be addressed through an

integrated approach that optimally incorporates targeted methods and resources to manage both Rs-pII lineages effectively.

Moreover, results also reveal the exclusive presence of Rs-pIIA strains in Grenada and Rs-pIIB in Costa Rica, Panama, and Peru, indicating allopatric speciation (Whitaker, 2006). The isolation of Rs-pII populations due to geographic barriers, such as the spatial distance and environment, may have hindered the gene flow and driven their evolutionary divergence (Castillo, 2023; Castillo and Greenberg, 2007; Chase et al., 2019). The presence of a single Rs-pII phylotype within a country may require the implementation of strategies specifically tailored to combat the phylotype in that context. Such an approach would not only enhance the overall effectiveness of control measures but also ensure their safety for the local ecosystem while optimizing resource allocation.

Previous research in comparative evolutionary genomics and reverse ecology indicates that RSSC phylotype diversification varies depending on geographic distribution, standing out that phylotype II is genetically the most diverse (Hong et al., 2012; Ramírez et al., 2020; Remenant et al., 2010; Sharma et al., 2022). Hence, our results regarding the independent evolution of *Rs* phylotypes IIA and IIB to adapt in a number of separate geographic regions across the Americas could imply that Rs-pIIB exhibits higher genetic diversity than Rs-pIIA. The phylogenetic and phylogeographic analyses further corroborate this assertion, revealing that Rs-pIIB exhibits a greater relative genetic distance over Rs-pIIA. However, it is noteworthy that strains belonging to *Rs* phylotype IIA and IIB are not evenly distributed among and within countries since Rs-pIIA is underrepresented, comprising only 30% of total strains compared to Rs-pIIB. Consequently, our findings may not accurately reflect their actual geographic distribution and genetic diversity. This disproportionate representation results from the limited availability of freely accessible research resources in the Americas, particularly in developing countries like Ecuador, as well as sampling bias in open-access genomic databases, likely stemming from geographical limitations. Further research in genetic diversity is required to support this assumption since the current literature reflects a need for consensus. Whereas certain studies indicate that Rs-pIIB exhibits higher levels of diversity (Santiago et al., 2017),

others argue that Rs-pIIA demonstrates greater diversification, attributed to its elevated rates of homologous recombination (Wicker et al., 2012). For the successful management of Moko disease, phytosanitary organizations must be aware of the genetic diversity of Rs-pII. High levels of genetic diversity or rapid diversification patterns may anticipate the emergence of new strains. Consequently, controlling this phytopathogen may require the adoption of flexible strategies that remain effective in the long term against evolving threats.

Bacterial populations are susceptible to undergoing different historical events, including genetic bottleneck, expansion, and contraction events, but the timing of such events varies depending on specific factors, such as mutation rate, selective pressure, life-history traits, geographic context, and population structure (Avitia et al., 2014; Castillo and Agathos, 2019; Gibson and Eyre-Walker, 2019; Kuo and Ochman, 2009; Woolfit and Bromham, 2003). The Bayesian skyline analysis of population dynamics in Rs-pII reveals that the demographic history of the phytopathogenic bacteria exhibits a stable behavior between 1959 and 2015. Although the widespread distribution of this phytopathogenic bacteria across all regions of America might indicate the occurrence of expansion events, our demographic findings suggest that the bacterial population has not undergone significant historical events. However, studies applying multilocus sequence analysis provide evidence that corroborate the ongoing diversification and recent population expansion of Rs-pIIA (Wicker et al., 2012). Our results are probably not compelling evidence to substantiate our presumptions since there is currently insufficient freely available genomic data from the appropriate temporal and spatial scales to accurately estimate the demographic history of the Rs-pII population. Additional molecular studies are necessary to describe better the demographic dynamics of the studied *Musa* species-infecting bacteria and determine its relationship with geographical distribution.

The phylogeographic findings reveal numerous intracontinental dispersal events of the Rs-pII population in different countries with similar environmental conditions throughout South America, Central America, and the Caribbean, including Brazil, Colombia, Peru, Costa Rica, Panama, and Grenada. Over time, this extensive network of long- and short-distance spread pathways has contributed to the long-term establishment of the Moko disease-causing bacteria

in the American continent. Research and historical outbreak records reveal that the actual geographic distribution of the phytopathogenic bacteria extends even further, indicating its presence in *Musa* species plantings located in additional regions such as Trinidad and Tobago, Grenadines, Guyana, Belize, Martinique, Jamaica, Suriname, Ecuador, Uruguay, Venezuela, Honduras, Guatemala, El Salvador, Nicaragua, and Mexico (Blomme et al., 2017; CAB International, 2014; CABI, 2021; Ramírez et al., 2020). It is noteworthy that the narrow spatial and temporal frame of the genomic data relative to the context of the evolutionary history being studied may constrain the scope of the phylogeographic inference, leading to confounding the dispersal patterns.

Results show that Brazil was the initial epicenter for the geographical spread of the Rs-pII and the source of bacterial dispersion in Colombia and Costa Rica. This finding suggests that this country may be the geographical origin of the MRCA of the pathogen, especially of the strains found in Colombia and Costa Rica, as previously proposed. A growing body of evidence from the literature adds weight to the argument that Brazil is the origin of the Rs-pII. Studies point out that the Moko disease-inducing ecotype comprising phylotype IIB strains in Martinique (Wicker et al., 2007) and French Guiana (Deberdt et al., 2014) was introduced through banana cutting originating from the Amazonian region of Brazil. However, more extensive phylogeographic research is warranted to validate these hypotheses.

Our data indicates that Moko disease-causing Rs-pII persisted predominately in South America. It is strongly plausible that this distribution pattern is primarily attributed to the tropical climatic conditions of these regions, given that the colonization ability of the RSSC phytopathogenic bacteria is temperature-dependent (Bittner et al., 2016), and the inefficiency of traditional agronomic practices for controlling the bacterial pathogen (Ahmed et al., 2022). The phytopathogen bacteria reached Brazil, Colombia, and Peru multiple times, expanding along them. This agrees with historical information describing recurring bacterial wilt epidemics affecting *Musa* species crops in those South American countries since 1950 (French et al., 1993; Mariano et al., 1998; Munar-Vivas et al., 2010).

As transmission patterns suggest, South America is implicated as the most frequent

source of the Moko disease epidemics caused by Rs-pII across the Americas. In particular, our results suggest that Anama, situated in the Brazilian Amazon, is a key hotspot for bacterial dissemination, followed by Manicore (Brazil) and Magdalena (Colombia), which is plausible since they are the leading banana exporting centers in their countries (Aguirre et al., 2015; Voora et al., 2020). Previous research indicates that the Moko disease-causing bacteria that prevailed in Brazil spread toward the Caribbean region by mechanical transmission (Deberdt et al., 2014) and throughout South America by insects or as a result of poorly intensive biological management (Blomme et al., 2017). Strategic planning of rigorous containment measures, along with monitoring and regulatory methods for agricultural product export activities in hotspots of Rs-pII dissemination, such as Anama and Magdalena, may be essential for eradicating the Moko disease propagation in high-dispersion areas and preventing the Rs-pII from spreading and establishing in new regions.

# Chapter 5

## Conclusions

The phylogenomic reconstruction classifies the Rs-pII infecting *Musa* species into phylotypes IIA and IIB. The genomic data alone have proven robust enough to reliably reconstruct the evolutionary relationships of the phytopathogen.

The geographic distribution of the Moko disease-causing Rs-pII throughout America has imprinted a distinctive signature in the genome of each of its members, shaping their evolutionary relationships. Thus, integrating geographic data may corroborate the phylogenetic findings without significantly altering the tree structure.

The phylotypes IIA and IIB of *Musa* species-infecting *Ralstonia solanacearum* are exclusively prevalent in specific countries of the Americas; however, they also coexist within other geographical regions of the same continent. Given the influence of geographical distribution over speciation, Rs-pII exhibits both allopatric and sympatric lifestyles (Castillo, 2023).

The findings underscore the complexity of population dynamics within the Rs-pII. Despite its extensive network of long- and short-distance spread pathways across America, the demographic history of Rs-pII has been described as constant between 1959 and 2015. Thus, the relationship between the geographic distribution of Rs-pII and its historical changes in population size remains ambiguous.

The phylogeographic reconstruction has determined the dispersion patterns of Rs-pII

across the Americas, highlighting its epidemiological distribution in South America.

South America is the most frequent recipient and source of the Moko disease epidemics caused by Rs-pII. Key hotspots of bacterial dissemination are distributed predominately in Brazil and Colombia. Tropical climatic conditions, export processes of agricultural products, and traditional agronomic practices ineffective for biological control of pathogens foster the spread of Rs-pII from South America, especially Brazil and Colombia, across all of America.

Phylogenomic results indicate that South America, specifically Brazil, is the geographical origin of the most common recent ancestor of Rs-pII, especially of Rs-pIIB strains from Peru, Colombia, and Costa Rica. Consistently, phylogeographic analysis tracks the geographic origin of the Rs-pII strains from Colombia and Costa Rica to Brazil.

# Chapter 6

## Recommendations and future research

This study describes the evolutionary dynamics and geographic dispersal patterns of the Moko disease-causing Rs-pII, laying the groundwork for future investigations into the mechanisms driving genetic diversity and adaptation of the phytopathogen.

Future research should replicate the phylogenomic, demographic history, and phylogeographic analyses using an expanded dataset comprising a statistically representative number of genomes of Rs-pIIA and Rs-pIIB strains isolated in recent years from a wider geographic range to validate and extend the findings of this study.

# Bibliography

- Agrocalidad. (2022). Plan de acción para el control de *Ralstonia solanacearum* raza 2. *Ministerio de Agricultura y Ganadería*, 50.
- Aguirre, S. E., Piraneque, N. V., & Rodríguez, J. (2015). Relationship between the nutritional status of banana plants and black sigatoka severity in the Magdalena region of Colombia. *Agronomía Colombiana*, 33(3), 348–355.
- Ahmed, W., Yang, J., Tan, Y., Munir, S., Liu, Q., Zhang, J., Ji, G., & Zhao, Z. (2022). *Ralstonia solanacearum*, a deadly pathogen: Revisiting the bacterial wilt biocontrol practices in tobacco and other Solanaceae. *Rhizosphere*, 21, 100479.
- Ailloud, F., Lowe, T., Cellier, G., Roche, D., Allen, C., & Prior, P. (2015). Comparative genomic analysis of *Ralstonia solanacearum* reveals candidate genes for host specificity. *BMC Genomics*, 16(1), 1–11.
- Albuquerque, G. M., Santos, L. A., Felix, K. C., Rollemberg, C. L., Silva, A. M., Souza, E. B., Cellier, G., Prior, P., & Mariano, R. L. (2014). Moko disease-causing strains of *Ralstonia solanacearum* from Brazil extend known diversity in paraphyletic phylotype II. *Phytopathology*, 104(11), 1175–1182.
- Alengebawy, A., Abdelkhalek, S. T., Qureshi, S. R., & Wang, M.-Q. (2021). Heavy metals and pesticides toxicity in agricultural soil and plants: Ecological risks and human health implications. *Toxics*, 9(3), 42.
- Alvarez, E., Pantoja, A., Ganan, L., & Ceballos, G. (2013). Estado del arte y opciones de manejo del moko y la sigatoka negra en América Latina y el Caribe. *CGIAR Repository*.

- Alvarez, J. A., Rodriguez Gaviria, P. A., & Marin Montoya, M. (2008). Detección molecular de *Ralstonia solanacearum* en agroecosistemas bananeros de Colombia. *Tropical Plant Pathology*, *33*, 197–203.
- Álvarez, B., Biosca, E. G., & López, M. M. (2010). On the life of *Ralstonia solanacearum*, a destructive bacterial plant pathogen. *Current research, technology and education topics in applied microbiology and microbial biotechnology*, *1*, 267–279.
- Ando, T. (2009). Bayesian factor analysis with fat-tailed factors and its exact marginal likelihood. *Journal of Multivariate Analysis*, *100*(8), 1717–1726.
- Aoun, N., Avalos, J. K., Cope-Arguello, M. L., Enriquez, C., Nguyen, T.-H., Olyushinets, A., Subuyuj, G. A., Tom, C., & Lowe-Power, T. M. (2023). Genome resource announcement of 3 phylotype I and 14 phylotype II *Ralstonia solanacearum* species complex isolates from South America. *PhytoFrontiers™*, *3*(4), 859–862.
- Arnold, B. J., Huang, I.-T., & Hanage, W. P. (2022). Horizontal gene transfer and adaptive evolution in bacteria. *Nature Reviews Microbiology*, *20*(4), 206–218.
- Avise, J. C. (2000). *Phylogeography: The history and formation of species*. Harvard University Press.
- Avise, J. C. (2009). Phylogeography: Retrospect and prospect. *Journal of Biogeography*, *36*(1), 3–15.
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, *18*(1), 489–522.
- Avitia, M., Escalante, A. E., Rebollar, E. A., Moreno-Letelier, A., Eguiarte, L. E., & Souza, V. (2014). Population expansions shared among coexisting bacterial lineages are revealed by genetic evidence. *PeerJ*, *2*, e696.
- Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., Huelsenbeck, J. P., Ronquist, F., Swofford, D. L., Cummings, M. P., et al. (2012).

- Beagle: An application programming interface and high-performance computing library for statistical phylogenetics. *Systematic biology*, 61(1), 170–173.
- Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., & Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular biology and evolution*, 29(9), 2157–2167.
- Baele, G., Lemey, P., & Suchard, M. A. (2016). Genealogical working distributions for bayesian model testing with phylogenetic uncertainty. *Systematic Biology*, 65(2), 250–264.
- Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A., & Lemey, P. (2012). Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Molecular biology and evolution*, 30(2), 239–243.
- Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate bayesian computation in population genetics. *Genetics*, 162(4), 2025–2035.
- Beutler, J., Holden, S., Georgoulis, S., Williams, D., Norman, D. J., & Lowe-Power, T. M. (2023). Whole genome sequencing suggests that “nonpathogenicity on banana (npb)” is the ancestral state of the *ralstonia solanacearum* iib-4 lineage. *PhytoFrontiers*<sup>TM</sup>, PHYTOFR–06.
- Bielejec, F., Baele, G., Vrancken, B., Suchard, M. A., Rambaut, A., & Lemey, P. (2016). Spread3: Interactive visualization of spatiotemporal history and trait evolutionary processes. *Molecular biology and evolution*, 33(8), 2167–2169.
- Bittner, R., Arellano, C., & Mila, A. (2016). Effect of temperature and resistance of tobacco cultivars to the progression of bacterial wilt, caused by *ralstonia solanacearum*. *Plant and Soil*, 408, 299–310.
- Blomme, G., Dita, M., Jacobsen, K. S., Pérez Vicente, L., Molina, A., Ocimati, W., Poussier, S., & Prior, P. (2017). Bacterial diseases of bananas and enset: Current state of knowledge and integrated approaches toward sustainable management. *Frontiers in plant science*, 8, 1290.

- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). Beast 2: A software platform for bayesian evolutionary analysis. *PLoS computational biology*, *10*(4), e1003537.
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., et al. (2019). Beast 2.5: An advanced software platform for bayesian evolutionary analysis. *PLoS computational biology*, *15*(4), e1006650.
- Buddenhagen, I. (2007). Blood bacterial wilt of banana: History, field biology and solution, 57–68.
- CAB International. (2014). Invasive species compendium. *ralstonia solanacearum* race 2. *Invasive Species Compendium*.  
<https://doi.org/http://www.cabi.org/isc/datasheet/44999/aqb>
- CABI. (2021). *Ralstonia solanacearum* race 2 (moko disease). *CABI compendium*.  
<https://doi.org/https://doi.org/10.1079/cabicompendium.44999>
- Castillo, J. A. (2023). Analysis of the speciation process suggests a dual lifestyle in the plant pathogen *ralstonia solanacearum* species complex. *European Journal of Plant Pathology*, *166*(2), 251–257.
- Castillo, J. A., & Agathos, S. N. (2019). A genome-wide scan for genes under balancing selection in the plant pathogen *ralstonia solanacearum*. *BMC Evolutionary Biology*, *19*, 1–16.
- Castillo, J. A., & Greenberg, J. T. (2007). Evolutionary dynamics of *ralstonia solanacearum*. *Applied and Environmental Microbiology*, *73*(4), 1225–1238.
- Castro-Nallar, E., Hasan, N. A., Cebula, T. A., Colwell, R. R., Robison, R. A., Johnson, W. E., & Crandall, K. A. (2015). Concordance and discordance of sequence survey methods for molecular epidemiology. *PeerJ*, *3*, e761.
- Chase, A. B., Arevalo, P., Brodie, E. L., Polz, M. F., Karaoz, U., & Martiny, J. B. (2019). Maintenance of sympatric and allopatric populations in free-living terrestrial bacteria. *MBio*, *10*(5), 10–1128.

- Clarke, C. R., Studholme, D. J., Hayes, B., Runde, B., Weisberg, A., Cai, R., Wroblewski, T., Daunay, M.-C., Wicker, E., Castillo, J. A., et al. (2015). Genome-enabled phylogeographic investigation of the quarantine pathogen *Ralstonia solanacearum* race 3 biovar 2 and screening for sources of resistance against its core effectors. *Phytopathology*, *105*(5), 597–607.
- Cordero, O. X., & Polz, M. F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nature Reviews Microbiology*, *12*(4), 263–273.
- Darling, Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome research*, *14*(7), 1394–1403.
- Darling, Mau, B., & Perna, N. T. (2010). Progressivemaue: Multiple genome alignment with gene gain, loss and rearrangement. *PloS one*, *5*(6), e11147.
- Daubin, V., Moran, N. A., & Ochman, H. (2003). Phylogenetics and the cohesion of bacterial genomes. *Science*, *301*(5634), 829–832.
- Deberdt, P., Guyot, J., Coranson-Beaudu, R., Launay, J., Noreskal, M., Rivière, P., Vigné, F., Laplace, D., Lebreton, L., & Wicker, E. (2014). Diversity of *Ralstonia solanacearum* in French Guiana expands knowledge of the “emerging ecotype”. *Phytopathology*, *104*(6), 586–596.
- Delgado, R., Morillo, E., Buitrón, J., Bustamante, A., Sotomayor, I., et al. (2014). First report of moko disease caused by *Ralstonia solanacearum* race 2 in plantain (*Musa aab*) in Ecuador. *New Disease Reports*, *30*(23), 2044–588.
- de Pedro-Jové, R., Puigvert, M., Sebastià, P., Macho, A. P., Monteiro, J. S., Coll, N. S., Setubal, J. C., & Valls, M. (2021). Dynamic expression of *Ralstonia solanacearum* virulence factors and metabolism-controlling genes during plant infection. *BMC genomics*, *22*, 1–18.
- Drenth, A., & Kema, G. (2021). The vulnerability of bananas to globally emerging disease threats. *Phytopathology*, *111*(12), 2146–2161.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS biology*, *4*(5), e88.

- Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., & Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, *161*(3), 1307–1320.
- Drummond, A. J., & Rambaut, A. (2007). Beast: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, *7*(1), 1–8.
- Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular biology and evolution*, *22*(5), 1185–1192.
- Emerson, B. C., Cicconardi, F., Fanciulli, P. P., & Shaw, P. J. (2011). Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1576), 2391–2402.
- Emerson, B. C., & Hewitt, G. M. (2005). Phylogeography. *Current biology*, *15*(10), R367–R371.
- Etminani, F., Yousefvand, M., & Harighi, B. (2020). Phylogenetic analysis and molecular signatures specific to the *Ralstonia solanacearum* species complex. *European Journal of Plant Pathology*, *158*, 261–279.
- Fan, Y., Wu, R., Chen, M.-H., Kuo, L., & Lewis, P. O. (2011). Choosing among partition models in bayesian phylogenetics. *Molecular biology and evolution*, *28*(1), 523–532.
- Fegan, M., & Prior, P. (2006). Diverse members of the *Ralstonia solanacearum* species complex cause bacterial wilts of banana. *Australasian Plant Pathology*, *35*, 93–101.
- Fegan, M., Prior, P., et al. (2005). How complex is the *Ralstonia solanacearum* species complex. *Bacterial wilt disease and the Ralstonia solanacearum species complex*, *1*(15), 449–461.
- Florens, J. P., Mouchart, M., & Rolin, J.-M. (2019). *Elements of bayesian statistics*. CRC Press.
- Food and Agriculture Organization of the United Nations (FAO). (2022). *Banana statistical compendium 2021*.
- French, E., Aley, P., Torres, E., & Nydegger, U. (1993). Diversity of *Pseudomonas solanacearum* in Peru and Brazil, 70–70.

- French, E., Sequeira, L., et al. (1970). Strains of *Pseudomonas solanacearum* from central and south America: A comparative study. *Phytopathology*, *60*, 506–512.
- Friel, N., & Pettitt, A. N. (2008). Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *70*(3), 589–607.
- Ghosh, J. K., Delampady, M., & Samanta, T. (2006). *An introduction to Bayesian analysis: Theory and methods* (Vol. 725). Springer.
- Gibson, B., & Eyre-Walker, A. (2019). Investigating evolutionary rate variation in bacteria. *Journal of Molecular Evolution*, *87*(9), 317–326.
- Google LLC. (2018). *Google Earth Pro*.  
[https://www.google.com/intl/es\\_ALL/earth/about/versions/#download-pro](https://www.google.com/intl/es_ALL/earth/about/versions/#download-pro)
- Goss, E. M., Tabima, J. F., Cooke, D. E., Restrepo, S., Fry, W. E., Forbes, G. A., Field, V. J., Cardenas, M., & Grünwald, N. J. (2014). The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. *Proceedings of the National Academy of Sciences*, *111*(24), 8791–8796.
- Grant, W. S. (2015). Problems and cautions with sequence mismatch analysis and Bayesian skyline plots to infer historical demography. *Journal of Heredity*, *106*(4), 333–346.
- Hall, T., Bioinformatics, I., Carlsbad, C., et al. (2011). BioEdit: An important software for molecular biology. *GERF Bull Biosci*, *2*(1), 60–61.
- Hayward, A. (1994). *Pseudomonas solanacearum*. *Pathogenesis and host specificity in plant diseases: histopathological, biochemical, genetic and molecular bases*, *1*, 139–151.
- Hikichi, Y., Yoshimochi, T., Tsujimoto, S., Shinohara, R., Nakaho, K., Kanda, A., Kiba, A., & Ohnishi, K. (2007). Global regulation of pathogenicity mechanism of *Ralstonia solanacearum*. *Plant Biotechnology*, *24*(1), 149–154.
- Ho, S. Y. (2012). Phylogenetic analysis of ancient DNA using BEAST. *Ancient DNA: Methods and Protocols*, 229–241.
- Ho, S. Y., & Shapiro, B. (2011). Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, *11*(3), 423–434.

- Holder, M. T., Lewis, P. O., Swofford, D. L., Bryant, D., Chen, M., & Kuo, L. (2014). Variable tree topology stepping-stone marginal likelihood estimation. *Bayesian phylogenetics: methods, algorithms, and applications, 1*, 95–111.
- Hong, J. C., Norman, D. J., Reed, D. L., Momol, M. T., & Jones, J. B. (2012). Diversity among *Ralstonia solanacearum* strains isolated from the southeastern United States. *Phytopathology, 102*(10), 924–936.
- Horita, M., Tsuchiya, K., Suga, Y., Yano, K., Waki, T., Kurose, D., & Furuya, N. (2014). Current classification of *Ralstonia solanacearum* and genetic diversity of the strains in Japan. *Journal of general plant pathology, 80*, 455–465.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science, 294*(5550), 2310–2314.
- Iversen, G. R. (1984). *Bayesian statistical inference*. Sage.
- Jeffreys, H. (1957). *Theory of probability*. Oxford University Press.
- Kago, E., Kinyua, Z., Maingi, J., & Okemo, P. (2017). Diversity of *Ralstonia solanacearum* strains in solanaceous crops production regions of central Kenya. *Journal of Experimental Agriculture International, 16*(1), 1–12.
- Kaplan, D., & Depaoli, S. (2013). Bayesian statistical methods. *Oxford handbook of quantitative methods*, 407–437.
- Kass, R. E. (1993). Bayes factors in practice. *Journal of the Royal Statistical Society: Series D (The Statistician), 42*(5), 551–560.
- Katz, L. (2014). *Lskscripts*. <https://github.com/lskatz/lskScripts>
- Kingman, J. F. C. (1982). The coalescent. *Stochastic processes and their applications, 13*(3), 235–248.
- Kuo, C.-H., & Ochman, H. (2009). Inferring clocks when lacking rocks: The variable rates of molecular evolution in bacteria. *Biology Direct, 4*, 1–10.
- Landry, D., González-Fuente, M., Deslandes, L., & Peeters, N. (2020). The large, diverse, and robust arsenal of *Ralstonia solanacearum* type III effectors and their in planta functions. *Molecular Plant Pathology, 21*(10), 1377–1388.

- Lanfear, R., Hua, X., & Warren, D. L. (2016). Estimating the effective sample size of tree topologies from bayesian phylogenetic analyses. *Genome biology and evolution*, 8(8), 2319–2332.
- Larget, B., & Simon, D. L. (1999). Markov chain monte carlo algorithms for the bayesian analysis of phylogenetic trees. *Molecular biology and evolution*, 16(6), 750–759.
- Lawrence, J. G., & Hendrickson, H. (2005). Genome evolution in bacteria: Order beneath chaos. *Current opinion in microbiology*, 8(5), 572–578.
- Leonard, S., Hommais, F., Nasser, W., & Reverchon, S. (2017). Plant–phytopathogen interactions: Bacterial responses to environmental and plant stimuli. *Environmental microbiology*, 19(5), 1689–1716.
- Lopez, P., & Baptiste, E. (2009). Molecular phylogeny: Reconstructing the forest. *Comptes rendus biologies*, 332(2-3), 171–182.
- Mariano, R., Silveira, N., & Michereff, S. (1998). Bacterial wilt in brazil: Current status and control methods, 386–393.
- Mather, N., Traves, S. M., & Ho, S. Y. (2020). A practical introduction to sequentially markovian coalescent methods for estimating demographic history from genomic data. *Ecology and evolution*, 10(1), 579–589.
- Munar-Vivas, O., Morales-Osorio, J. G., & Castañeda-Sánchez, D. A. (2010). Use of field-integrated information in gis-based maps to evaluate moko disease (*ralstonia solanacearum*) in banana growing farms in colombia. *Crop Protection*, 29(9), 936–941.
- Nahata, K. D., Bielejec, F., Monetta, J., Dellicour, S., Rambaut, A., Suchard, M. A., Baele, G., & Lemey, P. (2022). Spread 4: Online visualisation of pathogen phylogeographic reconstructions. *Virus evolution*, 8(2), veac088.
- Nascimento, F. F., Reis, M. d., & Yang, Z. (2017). A biologist’s guide to bayesian phylogenetic analysis. *Nature ecology & evolution*, 1(10), 1446–1454.
- National Library of Medicine (US), National Center for Biotechnology Information. (1988). *National center for biotechnology information (ncbi)*.  
<https://www.ncbi.nlm.nih.gov/>

- Navascués, M., Leblois, R., & Burgarella, C. (2017). Demographic inference through approximate-bayesian-computation skyline plots. *PeerJ*, 5, e3530.
- N'guessan, C. A., Brisse, S., Le Roux-Nio, A.-C., Poussier, S., Koné, D., & Wicker, E. (2013). Development of variable number of tandem repeats typing schemes for *Ralstonia solanacearum*, the agent of bacterial wilt, banana moko disease and potato brown rot. *Journal of microbiological methods*, 92(3), 366–374.
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). Iq-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, 32(1), 268–274.
- Obrador-Sánchez, J. A., Tzec-Simá, M., Higuera-Ciapara, I., & Canto-Canché, B. (2017). Genetic diversity of *Ralstonia solanacearum* strains from Mexico associated with moko disease. *European Journal of Plant Pathology*, 149, 817–830.
- O'Hagan, A. (2004). Bayesian statistics: Principles and benefits. *Frontis*, 31–45.
- Pais, A. K. L., Santos, L. V. S. d., Albuquerque, G. M. R., Farias, A. R. G. d., Silva Junior, W. J., Balbino, V. d. Q., Silva, A. M. F., Gama, M. A. S. d., & Souza, E. B. d. (2022). Comparative genomics and phylogenomics of the *Ralstonia solanacearum* moko ecotype and its symptomatological variants. *Genetics and Molecular Biology*, 45, e20220038.
- Pardo, J. M., López-Alvarez, D., Ceballos, G., Alvarez, E., & Cuellar, W. J. (2019). Detection of *Ralstonia solanacearum* phylotype II, race 2 causing moko disease and validation of genetic resistance observed in the hybrid plantain FHIA-21. *Tropical plant pathology*, 44, 371–379.
- Paudel, S., Dobhal, S., Alvarez, A. M., & Arif, M. (2020). Taxonomy and phylogenetic research on *Ralstonia solanacearum* species complex: A complex pathogen with extraordinary economic consequences. *Pathogens*, 9(11), 886.
- Peillex, C., & Pelletier, M. (2020). The impact and toxicity of glyphosate and glyphosate-based herbicides on health and immunity. *Journal of immunotoxicology*, 17(1), 163–174.
- Philippe, H., Delsuc, F., Brinkmann, H., & Lartillot, N. (2005). Phylogenomics. *Annu. Rev. Ecol. Evol. Syst.*, 36, 541–562.

- Ploetz, R. C. (2008). Black sigatoka and moko: Impact and spread of two destructive banana diseases in the caribbean basin.
- Rambaut, A. (2018). *Figtree*. <http://tree.bio.ed.ac.uk/software/figtree/>
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using tracer 1.7. *Systematic biology*, *67*(5), 901–904.
- Ramírez, M., Moncada, R., Villegas-Escobar, V., Jackson, R. W., & Ramírez, C. (2020). Phylogenetic and pathogenic variability of strains of *ralstonia solanacearum* causing moko disease in colombia. *Plant Pathology*, *69*(2), 360–369.
- Remenant, B., Coupat-Goutaland, B., Guidot, A., Cellier, G., Wicker, E., Allen, C., Fegan, M., Pruvost, O., Elbaz, M., Calteau, A., et al. (2010). Genomes of three tomato pathogens within the *ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC genomics*, *11*, 1–16.
- Robinson, O., Dylus, D., & Dessimoz, C. (2016). Phylo. io: Interactive viewing and comparison of large phylogenetic trees on the web. *Molecular biology and evolution*, *33*(8), 2163–2166.
- Rodríguez-Beltrán, J., DelaFuente, J., Leon-Sampedro, R., MacLean, R. C., & San Millan, A. (2021). Beyond horizontal gene transfer: The role of plasmids in bacterial evolution. *Nature Reviews Microbiology*, *19*(6), 347–359.
- Rudge, D. W. (1998). A bayesian analysis of strategies in evolutionary biology. *Perspectives on Science*, *6*(4), 341–360.
- Safni, I., Cleenwerck, I., De Vos, P., Fegan, M., Sly, L., & Kappler, U. (2014). Polyphasic taxonomic revision of the *ralstonia solanacearum* species complex: Proposal to emend the descriptions of *ralstonia solanacearum* and *ralstonia syzygii* and reclassify current *r. syzygii* strains as *ralstonia syzygii* subsp. *syzygii* subsp. nov., *r. solanacearum* phylotype iv strains as *ralstonia syzygii* subsp. *indonesiensis* subsp. nov., banana blood disease bacterium strains as *ralstonia syzygii* subsp. *celebesensis* subsp. nov. and *r. solanacearum* phylotype i and iii strains as *ralstonia pseudosolanacearum* sp. nov.

- International journal of systematic and evolutionary microbiology*, 64(Pt\_9), 3087–3103.
- Salmona, J., Heller, R., Lascoux, M., & Shafer, A. (2019). Inferring demographic history using genomic data. *Population genomics: Concepts, approaches and applications*, 511–537.
- Santiago, T., Lopes, C., Caetano-Anollés, G., & Mizubuti, E. (2017). Phyloptype and sequevar variability of *Ralstonia solanacearum* in Brazil, an ancient centre of diversity of the pathogen. *Plant Pathology*, 66(3), 383–392.
- Schmalz, X., Biurrun Manresa, J., & Zhang, L. (2023). What is a Bayes factor? *Psychological Methods*, 28(3), 705–718.
- Semple, C., & Steel, M. (2003). Phylogenetics. *Oxford University Press*, 24, 1–239.
- Sharma, P., Johnson, M. A., Mazloom, R., Allen, C., Heath, L. S., Lowe-Power, T. M., & Vinatzer, B. A. (2022). Meta-analysis of the *ralstonia solanacearum* species complex (rssc) based on comparative evolutionary genomics and reverse ecology. *Microbial Genomics*, 8(3), 000791.
- Shen, F., Yin, W., Song, S., Zhang, Z., Ye, P., Zhang, Y., Zhou, J., He, F., Li, P., & Deng, Y. (2020). *Ralstonia solanacearum* promotes pathogenicity by utilizing l-glutamic acid from host plants. *Molecular Plant Pathology*, 21(8), 1099–1110.
- Silva, S. d. O., Veras, S. d. M., Gasparotto, L., de Matos, A. P., Cordeiro, Z., & Boher, B. (2000). Evaluation of *musa* spp. for resistance to moko disease (*ralstonia solanacearum*, race 2).
- Sleator, R. D. (2011). Phylogenetics. *Archives of microbiology*, 193, 235–239.
- Stukenbrock, E. H., & McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annu. Rev. Phytopathol.*, 46, 75–100.
- Suchard, M. A., & Rambaut, A. (2009). Many-core algorithms for statistical phylogenetics. *Bioinformatics*, 25(11), 1370–1376.
- Tripathi, L., Ntui, V. O., & Tripathi, J. N. (2022). Control of bacterial diseases of banana using crispr/cas-based gene editing. *International Journal of Molecular Sciences*, 23(7), 3619.

- Vailleau, F., & Genin, S. (2023). *Ralstonia solanacearum*: An arsenal of virulence strategies and prospects for resistance. *Annual Review of Phytopathology*, *61*, 25–47.
- Valls, M., Genin, S., & Boucher, C. (2006). Integrated regulation of the type iii secretion system and other virulence determinants in *ralstonia solanacearum*. *PLoS pathogens*, *2*(8), e82.
- Voora, V., Larrea, C., & Bermudez, S. (2020). *Global market report: Bananas*. International Institute for Sustainable Development Winnipeg, MB, Canada.
- Whitaker, R. J. (2006). Allopatric origins of microbial species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1475), 1975–1984.
- Wicker, E., Grassart, L., Coranson-Beaudu, R., Mian, D., Guilbaud, C., Fegan, M., & Prior, P. (2007). *Ralstonia solanacearum* strains from martinique (french west indies) exhibiting a new pathogenic potential. *Applied and environmental microbiology*, *73*(21), 6790–6801.
- Wicker, E., Lefeuvre, P., De Cambiaire, J.-C., Lemaire, C., Poussier, S., & Prior, P. (2012). Contrasting recombination patterns and demographic histories of the plant pathogen *ralstonia solanacearum* inferred from mlsa. *The ISME journal*, *6*(5), 961–974.
- Wiley, E. O., & Lieberman, B. S. (2011). *Phylogenetics: Theory and practice of phylogenetic systematics*. John Wiley & Sons.
- Wilkinson, D. J. (2007). Bayesian methods in bioinformatics and computational systems biology. *Briefings in bioinformatics*, *8*(2), 109–116.
- Woods, A. (1984). Moko disease: Atypical symptoms induced by afluoidal variants of *pseudomonas solanacearum* in banana plants.
- Woolfit, M., & Bromham, L. (2003). Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Molecular Biology and Evolution*, *20*(9), 1545–1555.
- Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. *Nature reviews genetics*, *13*(5), 303–314.
- Zhi, X.-Y., Zhao, W., Li, W.-J., & Zhao, G.-P. (2012). Prokaryotic systematics in the genomics era. *Antonie Van Leeuwenhoek*, *101*, 21–34.

Zuckerlandl, E., & Pauling, L. (1965). Molecules as documents of evolutionary history.

*Journal of theoretical biology*, 8(2), 357–366.

Zulperi, D., & Sijam, K. (2014). First report of *Ralstonia solanacearum* race 2 biovar 1 causing moko disease of banana in Malaysia. *Plant Disease*, 98(2), 275–275.