NTNU
Norwegian University of
Science and Technology

Genome sequencing and phylogenomic analysis of the devastating plant pathogen Phytophthora infestans

Dhirendra Niroula

MSc in Biology
Submission date: November 2018
Supervisor: Michael D Martin, IBI
Co-supervisor: Mika Bendiksby, IBI

Norwegian University of Science and Technology
Department of Biology

Acknowledgements

This Master thesis was written at Norwegian University of Science and Technology (NTNU), University Museum. The financial support for the study was covered by Assoc. Prof. Michael Martin's Onsager fellowship start-up award. I would like to express my deep gratitude to my supervisor Michael Martin for his help, advice, support, and guidance for positively directing my thesis. Your consistent guidance in running the server and different softwares, as well as comments and suggestions in structuring my thesis, has made possible to successfully complete my thesis. I also want to express my warm regards to my co-supervisor Mika Bendiksby for her timely guidance in the crucial steps of my thesis. I would also like to thank all the group members Venessa, Malene, Andraes, Victoria and Ida for their direct and indirect help and suggestions in structuring my thesis work. I remember a special instance when Vanessa, Malene, and Andraes had given me suggestions and feedbacks in thesis presentation. Chao Song-thank you for being with me in the study room until 9 o' clock in the evening. Your help, especially, in running software was commendable. Also, thanks for answering my call even at 12:30 at the midnight and tolerating my silly questions. Thanks Erik for your support in the laboratory.

I would also like to thank formally all my friends for their incessant love and support. I am always indebted towards my friends for their presence in every thick and thins. I would like to thank my god, my parents for their unconditional love and support. All credit goes to your love and care for who I am right now. I will be indebted throughout my life for your love. A huge thanks to my fellow Cell and Molecular biology program and NTNU Museum classmates for two years of the wonderful company. Most of all, I cannot stand without thanking Prabina for all her love, support and care through every ups and downs. Thank you for believing in me and always encouraging me for whatever I do.

Abstract

The oomycete *Phytophthora infestans* is a plant pathogen responsible for late blight of potato and tomato. This rapidly evolving pathogen is of persistent concern because of the damage it has caused since the infamous Irish potato famine of 1845. Although many *P. infestans* genetic studies have been completed, the actual center of origin and famine-era strain has not been determined with certainty. We built genome libraries and sequenced 108 present-day isolates of *P. infestans* from six different locations of Mexico. We assembled 107 resequenced mitochondrial (mtDNA) genome sequences with 80 previously studied sequences from the global sources for our phylogenomic analysis. We also used 14 high-depth genomic sequences for effector gene analysis. The major objectives of this study are: to complete a phylogenomic analysis of *P. infestans* isolates from mitochondrial genome sequences (n = 187), and to characterize novel RXLR gene sequences assembled from the high-depth nuclear genome (n = 14) sequencing data. The phylogenomic analysis led us to observe that HERB-1 haplotypes are actually very common in Mexico. Toluca, which is often considered to harbor the most genetically diverse populations and to be the center of origin of the pathogen, is not found to be especially genetically diverse in the mitochondrial genome, because there we found a predominance of only the Ia mtDNA lineage. Our study also finds that the Ib mtDNA lineage is rare in recent times. In contrast to previous studies, our results do not support the claim of divergence of type I and type II lineage in South America. We name a new mtDNA lineage as lineage III, which is common in the study of isolates from Michoacán, and we suspect these isolates to belong to the 'aggressive lineage of global origin' (AGG). We discovered 10 novel RXLR genes in this study from an assembly of unmapped nuclear genome sequence reads. Cluster analysis reveals that most of these novel genes belong to a large cluster consisting of the RXLR genes that according to the *P. infestans* reference genome annotation are not classified as secreted RXLR effector gene families. This cluster is the largest, and new RXLR genes discovered belong to this cluster, indicating the cluster to be highly dynamic in nature.

Keywords: *Phytophthora infestans*, late blight, Irish potato famine, Mexico, phylogenetics, mtDNA, lineages, HERB-1, RXLR, cluster analysis

Table of Contents

Acknowledgements ..... **i**

Abstract ..... **iii**

Table of contents ..... **v**

List of Tables ..... **vii**

List of Figures ..... **ix**

1. Introduction **1**

1.1 Background information and relevant theory 1

1.1.1 Biology of *P. infestans* 2

1.1.2 Host-pathogen interaction 3

1.1.3 Origin of the pathogen and famine-era strain 4

1.1.4 Evolutionary ability of *P. infestans* 5

1.1.5 Phylogenetic analysis 6

1.2 Study Objectives 7

2. Materials and Methods **9**

2.1 Sonication and quantification of extracted DNA 9

2.2 Single-tube library preparation and amplification 11

2.3 Library pool preparation and sequencing 13

2.4 Sequence data processing and mapping 13

2.5 Variant discovery and genotyping 14

2.6 Phylogenetic analyses 14

2.7 Assembly of novel RXLR effectors 15

2.8 RXLR gene cluster analysis 16

3. Results **17**

3.1 Results from sequencing 17

3.2 Phylogenetic analysis 19

3.3 Categorization of study isolates according to mtDNA lineages 28

3.4 Characterization of novel RXLR effectors 28

4. Discussion **33**

4.1 Phylogenetic analysis 33

4.2 Characterization of novel RXLR effectors 37

4.3 Limitations of the study 38

4.4 Recommendations for future work 39

5. Conclusion **41**

References **43**

Appendix **55**

List of Tables

Table 2. 1 Table representing the region, host, year of isolation of those DNA extracts and number of isolates collected in each year 10

Table 3. 1 Table showing grouping of all the samples from different geographical locations to their respective mitochondrial DNA lineages. 26

Table 3. 2 Statistics for *de novo* assemblies of the unmapped reads. 28

Table 3. 3 Table indicating de novo assembly of unmapped reads for the detection of novel, non-reference RXLR protein. 29

Table A. 1 Provenance and necessary information regarding *Phytophthora spp*. samples used in this study.. 55

List of Figures

Fig 1. 1 (a) Black wilted, winded, necrotic lesion on potato leaf (b) Black necrotic lesion on potato stem (c) Rotted potato tuber infected by late blight pathogen (d) Tomato foliage and fruit affected by *P. infestans* 2

Fig 2. 1 General layout of the laboratory project from fragmentation of DNA to sequencing and read alignment to reference genome 9

Fig 2. 2 Representation of the geographical distribution of the study isolates in six different regions of Mexico.... 10

Fig 3. 1 Comparison of total nucleotides retained in all samples. 17

Fig 3. 2 Comparison of the nuclear genome sequencing depth achieved in all samples. 17

Fig 3. 3 Comparison of average sequencing depth of nuclear and mitochondrial genomes. 18

Fig 3. 4 Fraction of clonality for both mitochondrial and nuclear genome.... 18

Fig 3. 5 The mitochondrial DNA phylogeny of *P. infestans* isolates from Mexico and other previously studied samples using Bayesian dating analysis 21

Fig 3. 6 Section of the phylogenetic tree representing mtDNA lineages IIb, IIa and III. This section is represented by A in Fig 3.5. 22

Fig 3. 7 Section of the phylogenetic tree representing mtDNA lineages US-1/Ib and a part of lineage Ia. This section is represented by the B in Fig 3.5. 23

Fig 3. 8 Section of the phylogenetic tree representing part of mtDNA lineage Ia and part of HERB-1. This section is represented by the C in Fig 3.5. 24

Fig 3. 9 Section of the phylogenetic tree representing a part of mtDNA lineage HERB-1. This section is represented by D in Fig 3.5. 25

Fig 3. 10 The pie chart represents the categorization of mtDNA lineages in each of the six Mexican regions where our study samples come from.... 26

Fig 3. 11 Graph showing number of gene clusters.... 32

Fig A. 1 Mitochondrial DNA phylogeny of *P. infestans* isolates from Mexico and other previously studied samples using Bayesian dating analysis 63

CHAPTER 1

Introduction

1.1 Background information and relevant theory

Plant diseases caused by oomycetes are a major threat to worldwide food security [1]. Late blight of potato caused by *Phytophthora infestans*, an oomycete, has been a serious issue since 1845. Millions of people died during that period (the famine era, 1845-1849) in Ireland due to potato famine [2]. Every year millions of dollars are spent on fungicides in controlling the pathogen [3]. Regardless, it still exists as a major threat possibly due to its rapid evolutionary mechanisms [4].

There are few features that make *P. infestans* and its disease emergence of extreme interest and importance. Firstly, the devastating Irish potato famine caused by the pathogen took the lives of millions of people. This led to curiosity about the disease. Secondly, the devastating disease caused by the pathogen regularly appears in surprising new locations and with increased intensity. Thirdly, due to the rapid flux of *P. infestans* globally, the proper disease control has been a major issue [5].

Many recent studies have revealed new interesting findings as well as raised a few questions about this pathogen. Studies performing genome sequencing of this pathogen to look at the evolutionary history and the geographical distribution of different haplotypes have tried to answer various questions like the pathogen's center of origin, migration pattern of the pathogen and the famine-era strain [6–8]. However, those answers were based on only a few genome sequences from samples of the global origin (limited number of Mexican samples).

Therefore, this thesis will focus on the study of the evolutionary history of the pathogen through mitochondrial genome phylogenetic analysis, including more comprehensive sampling from Mexico (the probable location of *P. infestans*' center of origin) and all previously sequenced samples of global origin. Similarly, another focus is on the characterization of effector genes (which encode effector proteins) implicated in host-pathogen interactions [9]. For the pathogen to invade plant cell, it secretes effector proteins leading to effector triggered susceptibility [10,11]. We particularly focus on the

characterization of RXLR effectors. Hundreds of them are present in *P. infestans* genome [12,13], and we focus on the discovery and characterization of novel ones not found in the reference genome annotation.

Fig 1. 1 (a) Black wilted, winded, necrotic lesion on potato leaf (b) Black necrotic lesion on potato stem (c) Rotted potato tuber infected by late blight pathogen (d) Tomato foliage and fruit affected by *P. infestans* [14] .

1.1.1 Biology of *P. infestans*

In the tree of life, oomycetes are classified within the Stramenopiles lineage of the supergroup Chromalveolates of the eukaryotes [1,15]. Initially, *P. infestans* and other oomycetes were classified as a fungus due to their filamentous growth nature, feed upon by absorption and reproduction by spore [4,16]. However, after the development of molecular techniques oomycetes were found to be distantly related to fungi and assigned to a separate group [4]. *Phytophthora infestans* is a heterothallic organism and consists of two mating types, A1 and A2, and both sexual and asexual reproduction takes place [17,18]. The onset of late blight occurs with certain symptoms which can be observed on both tubers and foliages [19] (Fig 1.1).

1.1.2 Host-pathogen interaction

For pathogens to invade plant cells, pathogens secrete effector proteins that can manipulate host defense response leading to effector triggered susceptibility [10,20]. When a resistant protein in plant recognizes an effector, effector-triggered immunity is activated, often resulting in a hypersensitive response. The effector, in this case, acts as an avirulence factor [19]. Effectors of *P. infestans* are of two types based on their expression of effector proteins within the host: extracellular (apoplastic) or cytoplasmic [11]. Apoplastic effectors are divided into three groups: small cysteine-rich proteins, enzyme inhibitors, and the Nep-1 like (NPL) family. Most of them target plant defense-related proteins, proteases and glucanases [21–23]. Similarly, cytoplasmic effectors are divided into two classes: the RXLR protein family and the Crinkler protein family. Both are delivered inside the plant cell via a haustorium [24]. The RXLR effector is major cytoplasmic effector, and hundreds of them are present in the *P. infestans* genome [12,13]. The RXLR family protein can be categorized into two groups: IPIO and AVR3a. The IPIO group consists of three classes (I, II and III) [25,26]. Similarly, RXLR effector AVR3a has two different forms $AVR3a^{KI}$ (containing amino acid C19, K80, and I103) and $AVR3A^{EM}$ (containing amino acid S19, E80 and M103) [27].

Particular characteristics of RXLR (Arg-X-Leu-Arg, where X is any amino acid) effectors are the presence of the conserved N-terminus and highly divergent C-terminus in the amino acid sequence of the protein [12]. The conserved N-terminus end consists of a signal peptide and RXLR motif [12]. In addition to RXLR motif, dEER or EER motif is present downstream of RXLR [28]. N-terminal signal peptide is responsible for the secretion and RXLR motif is responsible for translocation inside the cytoplasm whereas, the C- terminal domain has an effector function [12,24]. Both the E's in EER is replaced by D, and R is replaced by K in some of the RXLR effector proteins [12]. A meticulous bioinformatic examination (on the basis of the Hidden Markov Model (HMM) score and MEME search result) [28] of the C-terminal end of the effector has revealed the presence of three conserved W-Y-L motifs, often present in the repeated fashion [13,29]. They are arranged as a module and repeated up to eight times [28]. The W motif has tryptophan residue at position 5 and is 25 amino acids long and is found in about 60% of the RXLR superfamily members. In about 95% of RXLR effectors, the W motif is followed by Y and/or L motif having a length of 22 and 33 amino acids, respectively [28]

1.1.3 Origin of the pathogen and famine-era strain

The late blight of potato (and tomato) appeared quite suddenly in the 19th century, but it was unclear from where the pathogen came [30,31]. There has been an increasing number of research in recent years to find out the probable location of pathogen's origin, with the debate focused on whether it is Central Mexico or South America [32]. Mexico is regarded as the center of origin of the pathogen based on three grounds [33]. First, *P. ipomoeae* and *P. mirabilis*, which are closely related to *P. infestans*, are both endemic to central Mexico and cause disease in *Ipomoea* spp. and *Mirabilis jalapa* respectively [34,35]. Second, genetically diverse populations of *P. infestans* are present in the Toluca Valley of Mexico and contain both mating types (A1 and A2) at an approximately 1:1 ratio. Third, migration of *P. infestans* from Mexico to Europe occurred in the 19th century. Prior to that, only mating type A1 was found worldwide outside Mexico [36,37], limiting these populations to asexual reproduction. Similarly, the hypothesis that South American is the center of origin of the pathogen is based on four lines of evidence. First, the presence of resistant genes in endogenous and wild *Solanum* species indicates the presence of *P. infestans* species in South American Andes for centuries (Hawkes 1990). Second, based on the rooted gene genealogies of nuclear RAS locus and mitochondrial P3 and P4 region Gloss et al (2014) reported the pathogen origin to be in the Andes. Third, Martin et al. (2014) found the type II mitochondrial DNA lineage out of two lineages to be ancestral lineage and found in South America. Fourth, the nucleotide diversity was found to be more in South American isolates than Mexican isolates [40].

Initially, it was proposed a single haplotype of *P. infestans* (Ib) was responsible for the 19th-century epidemics, and this lineage was thought to be predominant in recent times [41]. However, genetic analysis revealed that the 19th-century potato famine was caused by mitochondrial lineage Ia and the introduction of Ib lineage came later, in the 20th century [2,42]. The mtDNA lineage Ia in the 19th century was predominant not only in Toluca Valley of Mexico [4,43,44] but also found in South America [45]. South- America is thought to be the center of origin of 19th-century Ib lineage which later spread in North America and Europe. Three mtDNA SNPs were studied to draw this conclusion [2,42]. Further data on the famine-era strains were obtained by two parallel studies comparing herbarium samples collected between 1845 to 1896 with diverse modern isolates [46,47]. Yoshida et al. (2013) sequenced ancient DNA and modern genomes and found that 19th-century epidemic was

caused by HERB-1 mtDNA lineage. They further stated that nowadays the HERB-1 genotype is rare or hardly present [47]. But, in a separate study using phylogenetic analysis, Martin et al. (2014) identified HERB-1 lineage in modern populations from both South America and Mexico.

1.1.4 Evolutionary ability of *P. infestans*

Mutation, population size, mode of reproduction, migration, and selection play major roles in the evolutionary ability of pathogens [48]. One of the prime examples of the evolutionary potential of the pathogen is an appearance of an aggressive clonal lineage 13_A2 in Europe in the early 21st century [49]. This aggressive genotype has more recently become pandemic and has caused severe outbreaks in India [50] and China [51]. This aggressive clonal lineage outcompetes other aggressive lineages in the potato field and is not susceptible to plant host resistance [52]. The genome of *P. infestans* is organized in a unique way, having both gene-dense and gene-sparse regions [13]. In the gene sparse regions, the level of repeat content is high [13]. These gene-sparse regions show high copy number variation, silent or replacement (missense and nonsense) mutations [49], and are involved in pathogenesis through secreted proteins [53]. Similarly, mode of reproduction is also an important factor for the evolution of pathogens, and affects the gene distribution among the species [48]. Sexual reproduction in the pathogen enhances the survival mechanism; it also acts as a source of variation [48]. In addition, migration also plays a crucial role in the evolution of this pathogen. Migration of the A2 mating type into regions where only A1 was initially present led to sexual reproduction [54]. There are three theories that describe the migration patterns of *P. infestans* [55]. According to the first theory, migration of the pathogen took place from Mexico to the United States and then to Europe in the mid-19th century [37]. The second theory suggested the migration of the pathogen from the Andean region (near Peru) to the United States and then to Europe [56]. The third theory suggested migration of pathogen from Mexico to Peru to the United States and then to Europe [30].

Several impactful studies focus on population structure and genetic diversity of *P. infestans* in Toluca Valley of Mexico [32,44,57,58] whereas fewer target also to other locations of Mexico [59–61] and most studies use non-genomic methods. The dynamics of the *P. infestans* worldwide have been impacted by the population biology of *P. infestans* in the

central highland of Mexico [61]. Yet, it is not possible to predict the diversity of *P. infestans* by looking at the population structure of only one region (Toluca valley) of Mexico. Goodwin et al. (1992) used DNA fingerprinting and allozyme analysis to investigate the diversity of *P. infestans* between Central and Northern Mexico. A more recent study found sexual reproduction and high level of genetic diversity of *P. infestans* on potato crop in the Mexican states of Michoacán, Tlaxaca, and Toluca by using SSR genotyping and gene flow [61]. Wang et al. (2017) further suggested that *P. infestans* exists as a metapopulation in Toluca Valley and is the source of genetic diversity in other locations of Mexico.

1.1.5 Phylogenetic analysis

The mitochondrial genome is often used in the study of evolutionary relatedness among *Phytophthora* isolates [62], because it is reasonable in size, exists in multiple copies in each cell, is resistant to degradation, lacks recombination, and has high substitution rate [63]. In contrast, the nuclear genome does not have a high substitution rate, but a high recombination rate leads to genetic diversity [64]. There are various methods for inferring phylogenetic trees (including neighbour-joining, maximum parsimony, maximum likelihood and Bayesian Inference (BI)) (Felenstein 2004). The neighbour-joining method constructs a phylogenetic tree based on evolutionary distance data [66]. Maximum parsimony finds a tree with a minimum number of evolutionary changes [67]. Standard statistical techniques are used by the maximum likelihood approach for inferring probability distribution in order to assign probabilities to a particular phylogenetic tree [68]. In this method bases of all the sequences at each site are considered as independent, and then the log- likelihood of having these bases for a given topology is computed by using the same evolutionary model [69]. The Bayesian inference is a branch of statistical inference which uses prior knowledge in assessing the probability of model parameter in the presence of new data [70]. BEAST is a software for phylogenetic analysis of molecular data, which estimates rooted, time measured phylogenies based on various model parameters [70].

1.2 Study Objectives

General Objectives

1) Perform a phylogenomic analysis of mitochondrial genome sequences from all available *P. infestans* isolates to find the evolutionary relatedness among the isolates.

2) Characterize novel RXLR effectors from nuclear genome sequences of a subsample of Mexican isolates based on conserved motifs and gene cluster analysis.

Specific Objectives

1) Perform low-depth genome sequencing (Illumina Next-generation sequencing platform) of present-day isolates from Mexico.

2) Determine mitogenome sequences from these samples and previously published samples from global sources.

3) Characterize the diversity of mtDNA haplotypes in each of the six Mexican study locations.

4) Perform high-depth genome sequencing of at least two isolates from each of the six Mexican study areas.

5) Assemble novel RXLR sequences and then categorize them with a gene cluster analysis.

CHAPTER 2

Materials and Methods

The samples studied (these samples will be termed as 'study samples/isolates' hereafter) in this project were 112 DNA extracts of *P. infestans* cultures isolated from potato infected by the late blight from six different regions of Mexico (Fig 2.2). The provenance of these samples (Table A.1) is described in Wang et al. (2017) [61]. Together, 16 samples were isolated from Chapingo (9 (2015), 7 (2016)), 14 samples were isolated from Toluca in 2015, 18 samples were isolated from Michoacán in 2007, 26 samples were isolated from Juchitepec (11 (2015), 15 (2016)), 19 samples were isolated from Tlaxcala (4 (2001), 15(2015)), and 19 samples were collected from San Geronimo (16 (2015), 3 (2016)). Out of the 112 total samples provided by collaborators in the US, only 108 samples were pooled for sequencing (Isolates CHCL14, CHCL18, CHCL19, and CHG1 were not prepared for sequencing because of their extremely low DNA concentrations) (Table 2.1). Different methods were followed until the preparation of indexed DNA libraries that could be pooled and send it to the sequencing centre in Oslo. After approximately one month, the raw data was provided by the sequencing centre. Methods followed in the laboratory and the bioinformatics softwares applied in the study are described below.

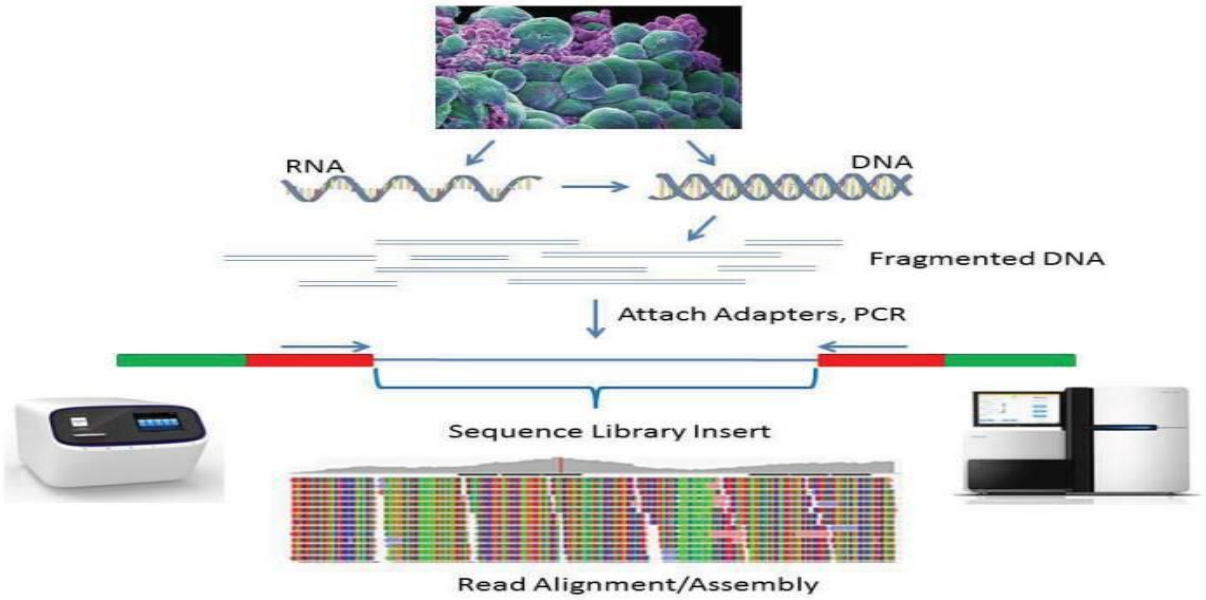

Fig 2. 1 General layout of the laboratory project from fragmentation of DNA to sequencing and read alignment to reference genome [71]

2.1. Sonication and quantification of extracted DNA

Sonication (shearing) of DNA was done to reduce the genomic DNA fragments to approximately 350 to 400 bp length. For this Bioruptor Pico instrument was used. After bringing the sample volume to 100 µl using EB buffer, 12 samples were sheared simultaneously using 7 sonication cycles at 15 sec on and 90 sec off format. After the double-strand breaking of DNA by sonication, the DNA extracts were checked if they were fragmented to the size we had expected (350-400 bp) on a 2% agarose gel stained with SYBR Safe DNA Stain. The Qubit2.0 fluorometer was used with the Qubit HS (High Sensitivity) and BR (Broad Range) assay kits to estimate the concentration of double-stranded DNA using 2 µl samples of the sheared DNA extracts.

Fig 2. 2 Representation of the geographical distribution of the study isolates in six different regions of Mexico. Different colors are used to represent different regions Tlaxcala (Red), Toluca (Pink), Chapingo (Yellow), San Geronimo (Green), Michoacán (Faint Blue) and Juchitepec (Dark Blue).

Table 2. 1 Table representing the region, host, year of isolation of those DNA extracts and number of isolates collected in each year (four samples not included for sequencing not included in the table).

Regions	Host	Collection year	Number of isolates
Chapingo	Potato	2015	9
		2016	3
Toluca	Potato	2015	14
Tlaxcala	Potato	2015	15
		2001	4
Michoacan	Potato	2007	18
Juchitepec	Potato	2015	11
		2016	15
San Geronimo	Potato	2015	16
		2016	3
Total			108

2.2 Single-tube library preparation and amplification

The BEST 2.0 single-tube protocol was followed for preparing blunt-end, shotgun Illumina libraries [72]. The reagents and buffers used in library preparation methods were enzymes (T4 Polynucleotide kinase, T4 DNA polymerase, T4 DNA ligase, Bst 2.0 Warmstart polymerase), buffers (ligase buffer, isothermal amplification buffer) and other reagents (dNTPs, molecular grade water, PEG-4000, Tween 20, NaCl solution) etc. The BEST library preparation method has three major steps namely end-repair, ligation, and fill-in process. Hence, three different master mixes; end-repair master mix, ligase master mix, and fill-in master mix were used during the process. The end-repair master mix was used to repair the blunt end and overhangs present in fragmented DNA. The ligase master mix was used to ligate the adapters to the DNA and fill-in master mix was used to fill-in the space created during the ligation. 32 µl of EBT was used to elute the purified libraries.

SPRI beads were used to purify the built libraries. A SPRI beads to DNA volume ratio of 1:1 was used to remove DNA fragment shorter than approximately 100bp from the libraries. The clean-up of the library was done both on the 16 places magnetic rack as well as 96 well magnetic plate. First, 60 µl of built library sample was mixed with 60 µl of SPRI

beads. Care was taken to slowly pipette in and out because SPRI beads solution is highly viscous. The mixture was incubated at RT for 5 minutes and kept in the magnetic rack for few minutes. The supernatant was discarded, and the beads were washed twice with freshly prepared 80% ethanol. The beads were then allowed to dry for approx. 5 minutes at RT but not over dried because over drying may lead to fragmentation of DNA. The tubes were taken out of the magnetic rack and beads were eluted with EBT buffer. Finally, tubes were again put back in the magnetic rack, allowed the beads to be attracted towards the magnet and supernatant was pipetted out as a purified library.

A quantitative PCR analysis was performed on each library to test the newly built libraries and to estimate the number of cycles required for optimal amplification in indexing PCR. The purified libraries were diluted 1:20 (1 µl of library template and 19 µl of EBT was added) for quantitative PCR analysis. Other reagents used for qPCR were dNTPs, IS7 and IS8 primers AmpliTaq Gold buffer, $MgCl_2$, AmpliTaq Gold polymerase, SYBR Green mix, and molecular biology H_2O. The qPCR conditions were set beforehand loading the samples in the qPCR. The conditions were; start for 10 min at 95°C, denature for 30 sec at 95°C, annealing for 1 min at 60°C, an extension for 45 secs at 72°C. The conditions from denaturing to the extension were repeated for 40 cycles, followed by a melt curve. The qPCR quantifies the amplifiable DNA present in each of the samples. Amplification of purified library was done to find the exact number of cycles needed for each sample in the indexing PCR.

After estimating the number of cycles required to give to each of the samples by the help of qPCR, the unamplified libraries were subjected to indexing PCR. In this process, each of the samples was indexed (barcoded) using unique forward (P5) and reverse (P7) primers. Unique barcoded primers were used so that it would be possible to demultiplex the sequences after Illumina sequencing. The PCR reactions were performed in volumes of 50 µl and contained dNTPs (final conc. 0.2 µM, F index primer (final conc. 0.2 µM), R index primer (final conc. 0.2 µM), AmpliTaq Gold Polymerase (final conc. 0.05 U/µl), PCR buffer II (final conc. 1X), $MgCl_2$ (final conc. 2.5 mM), BSA (final conc. 0.4 mg/ml), 5 µl of undiluted library template, and molecular biology water. Initial denaturation was performed at 95°C for 10 min, then denaturation at 95°C for 30 sec, the annealing temperature was set at 60°C for 1 min and the extension temperature of 72°C for 45 sec. The steps from denaturation to extension was repeated 14-22 times according to the values calculated from qPCR. The final holding

temperature for the samples was set at 4°C. The indexed libraries were purified using Qiagen QIAquick column and eluted in 32 µl EB buffer.

2.3 Library pool preparation and sequencing

An Agilent 2100 BioAnalyzer was used to estimate the distribution of DNA fragment lengths, concentration, and molarity of each purified, indexed library. Indexed libraries were pooled for sequencing in the desired proportion using the DNA molarity values estimated from the BioAnalyzer and Qubit measurements. The pooled samples were sequenced both at the NTNU University Museum and at the Norwegian National Sequencing Center in Oslo on various Illumina sequencing platforms. Data was generated in several runs: MiniSeq 150-bp SR (Single-Read), NextSeq 500 75-bp SR, HiSeq X 150-bp PE (Paired-End), and HiSeq4000 150-bp PE.

2.4 Sequence data processing and mapping

Shotgun genomic sequencing data from all previously published and publicly available *P. infestans* and *P. andina* samples were obtained from the European Nucleotide Archive via accession codes PRJEB2285 (Raffaele et al. 2010), PRJEB2282 (Raffaele et al. 2010), PRJEB2284 (Raffaele et al. 2010), PRJEB2285 (Raffaele et al. 2010), ERS226850 (Cooke et al. 2012), PRJEB4015 (Martin et al. 2013), PRJEB1717 (Yoshida et al. 2013), PRJEB1877 (Yoshida et al. 2013), PRJEB1879 (Yoshida et al. 2013), PRJNA275926 (Martin et al. 2016), PRJNA52431, and PRJEB20998. This raw sequence data, along with the sequence data generated in this study, was processed using the PALEOMIX software pipeline [73]. This pipeline simplifies the removal of library adapter contamination and performs mapping and filtering of sequence reads. Low-quality bases and residual adapter sequences were trimmed with the software AdapterRemoval v2.2 [74]. Paired reads overlapping by at least 11 bp were collapsed into a single read with a re-calibrated base quality. After trimming, reads shorter than 25 bp were discarded. Independent mapping against the draft reference nuclear and mitochondrial genome assemblies of *P. infestans* strain T30-4 (Haas et al. 2009) was performed with the backtrack algorithm of the software BWA v0.7.16 [75]. BWA's default settings were used, except that seeding was deactivated for historical samples. The

MarkDuplicates function of Picard v2.9.1 was used to remove PCR duplicates. Reads with mapping quality (MAPQ) scores <30 were discarded. Read alignments around insertion/deletion events were optimized with the RealignerTargetCreator and IndelRealigner tools included in the Genome Analysis Toolkit (GATK) v3.7 [76]. To reduce the damage pattern of ancient DNA sequenced from the historical samples, MapDamage2 [77] was used to recalibrate base qualities of aligned sequence reads. Statistics on genome and exome sequencing depth and amplified library clonality were generated by PALEOMIX.

2.5 Variant discovery and genotyping

The UnifiedGenotyper algorithm within the Genome Analysis Toolkit (GATK) v3.8 was used to discover SNPs in the resulting aligned sequence data simultaneously in all individual BAM (Binary sequence Alignment Map) file, using parameters minGQ=30 and minQ=30. The ploidy level was assumed to be 2*n* for the nuclear genome and 1*n* for the mitochondrial genome. The maximum allowed number of alternate alleles was set to 3. INDELs were included in the resulting VCF (Variant Call Format) file. A custom Python script was used to convert the resulting genotype calls into multiple sequence alignments (MSA) in FASTA format. During this process, INDELs and genotypes with a quality score (GQ) of less than 20 were masked with N, and heterozygous genotypes were converted to the corresponding IUPAC ambiguity codes. An MSA of the entire mitochondrial genome region was generated. For the nuclear genome, to reduce the time of phylogenetic inference, only the first 1.5 Mbp of the coding regions of the largest supercontig in the *P. infestans* genome assembly were included in the MSA. Because of their extremely low depth in these regions, the samples JFH89, Pi1845B, and KM177497 were excluded from further analysis.

2.6 Phylogenetic analyses

In total 187 samples were included in the phylogenetic analysis. These included 107 study samples, 80 previously published samples (including reference genome T30-4), including six unpublished samples collected in Denmark (Pinf_B_13_8_35, Pinf_D_13_3_5, Pinf_F_13_7_33, Pinf_A_13_5_10, Pinf_E_13_8_102 and Pinf_C_13_7_28). BEAST v1.8.2 (which comes with BEAST, Beauti, and LogfileTracer) was used for the phylogenomic

analysis of the mitochondrial genome. Beauti v1.8.2 was used to prepare the Beast input file with the following run parameters: HKY+G nucleotide substitution model, strict clock as rate variation among branches, and tree prior set as flexible skyride coalescent. For this model, the MCMC was run for 40M generations and sampled every 4000 generations. The first 4M generations of the run were discarded as burn-in. The posterior distribution of parameters was estimated by Markov Chain Monte Carlo (MCMC) sampling. Samples were drawn every 4000 MCMC steps over a total of 40,000,000 steps. The summary statistics from the Beast analysis were visualized using LogfileTracer. The resulting phylogenetic tree was visualized in FigTree v1.4.0. The node support (posterior probability) at the point of branching was visualized, and the scale axis option was used to visualize the evolutionary date of divergence of each isolate from the most recent common ancestor. The mtDNA lineages of the previously published reference samples and posterior probability support at the node was used as a basis for the categorization of the isolates into different mtDNA lineages.

2.7 Assembly of novel RXLR effectors

Genomes of 14 Mexican samples were sequenced high depth to study novel effector genes. These samples were: CHG3 and CH02 from Chapingo, TG10 and TF1 from Tlaxcala, JFH99, JFH109, and JFH32 from Juchitepec, Mich7006 and Mich7014 from Michoacán, SGFX and SGF78 from San Geronimo, and T11, T12 and T17 from Toluca. Sequences not mapping to the reference genome were used to discover non-reference genes (absent from the T30-4 genome annotation) encoding novel RXLR proteins in the high-depth resequenced sample genomes. For each of these samples, sequences that did not map to the T30-4 reference genome assembly were used as input to a *de novo* assembly performed with ABySS v2.0 [78] using a k-mer length of 85. The total assembly length and the N_{50} statistic were used to quantify the completed assemblies. N_{50} is the size of a contig such that 50% of the sequence assembly of a particular contig is equal to or larger than the size of N_{50}. The N_{50} helps us find the quality of our genome assembly since they are distributed into contigs of different sizes. A custom script was used to translate all assembled contig nucleotide sequence to amino acid sequence for six possible reading frames. Then these amino acids sequences were searched for all the ORFs (Open Reading Frames) with lengths of at least 70 residues [49]. SignalP v4.1 was used to detect signal peptides characteristic of secreted effector proteins. All the amino

acid sequences were then sorted based on SignalP discrimination score (D-score ≥ 0.9) and signal peptide within residues 1-30 were searched for RXLR motifs within residues 25-60. X here stands for any amino acid. In this way, different probable peptide sequences for each sample were produced. These proteins were used for server-based BLAST against the T30-4 reference genome and other recently available reference genomes available in the gene bank. This removes the non-*Phytophthora* hits with E-value greater than $1e^{-10}$ and any perfect matches to reference genome including T30-4.

2.8 RXLR gene cluster analysis

An RXLR gene cluster analysis was performed, including our newly discovered genes along with all previously known RXLR genes from the T30-4 reference genome assembly. We used software mcl [79] with an inflation value of 1.5 to cluster annotate RXLR effectors into tribes. As input for the RXLR gene clustering, we used the result of protein-protein self-BLAST analysis (cutoff E-value of 10^{-5}) of the entire T30-4 RXLR arsenal in addition to the novel RXLR amino acid sequences discovered by Martin et al. (2013), Cooke et al. (2012) and Yin et al. (2013) and against itself [80].

CHAPTER 3

Results

3.1 Results from sequencing

We were interested to look at some of the important parameters such as retained nucleotides, depth of coverage and fraction of the PCR duplicates. For the majority of the isolates, the total amount of sequencing data generated was approximately 2 Gbp (Fig 3.1). Sample JFH32 received approximately 34 Gbp. We also compared the achieved nuclear genome sequencing depth (mean number of reads overlapping every position in the reference genome) for all the samples. This value ranged between 3 and 45, with sample JFH32 being the highest (Fig 3.2).

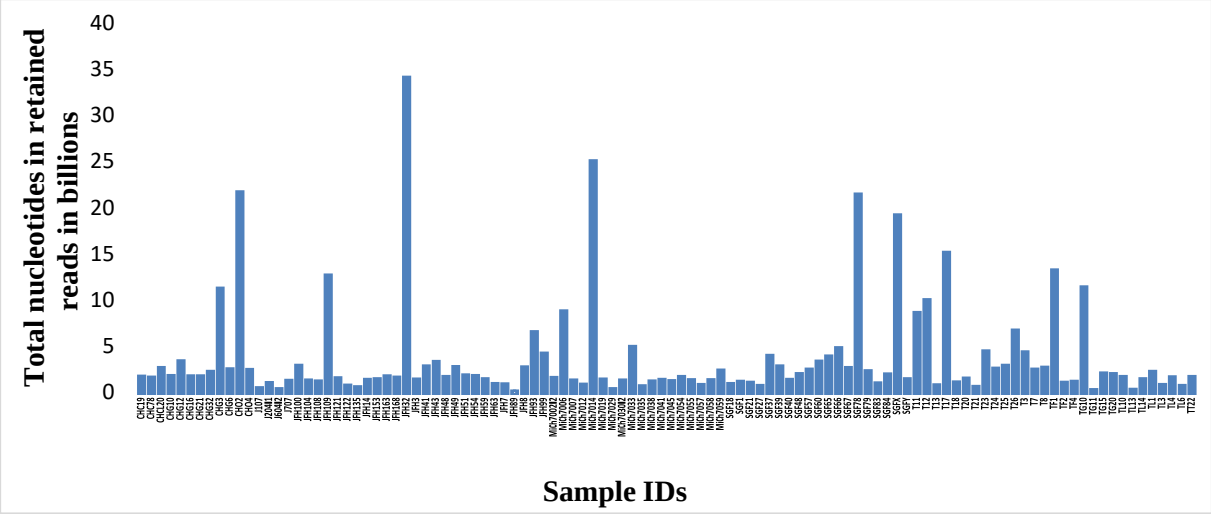

Fig 3. 1 Comparison of total nucleotides retained in all samples.

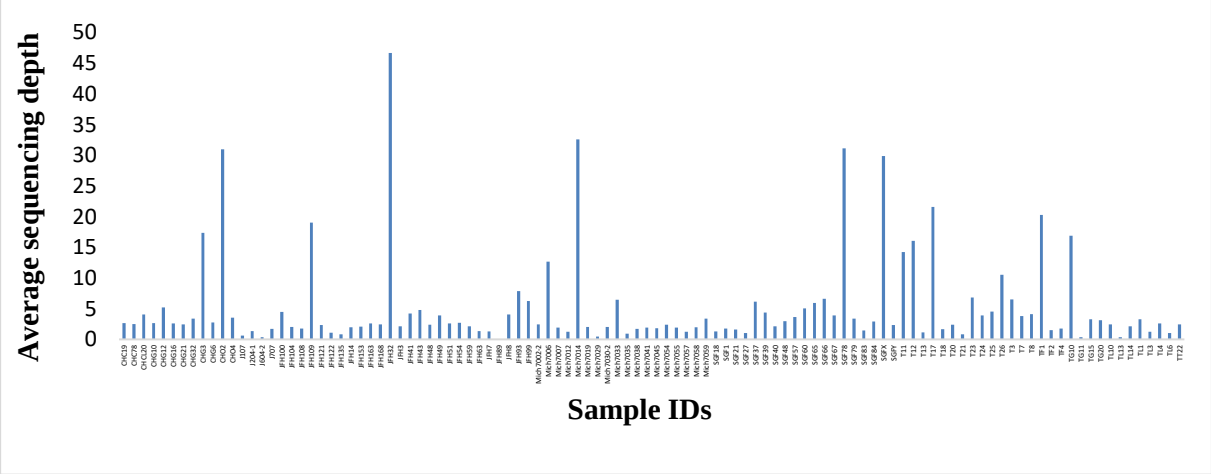

Fig 3. 2 Comparison of the nuclear genome sequencing depth achieved in all samples.

Likewise, a comparison of average sequencing depth was performed for both nuclear and mitochondrial genomes. As predicted due to its higher copy number, the average sequencing depth for the mitochondrial genome was higher than that of the nuclear genome. The average sequencing depths for nuclear and mitochondrial genomes were 5.23 (s.d. = 7.66) and 529.27 (s.d. = 667) respectively (Fig. 3.3). Similarly, the proportion of reads that were PCR duplicates for mitochondrial and nuclear genomes were 0.27 (s.d = 0.17) and 0.13 (s.d = 0.05), respectively (Fig 3.4).

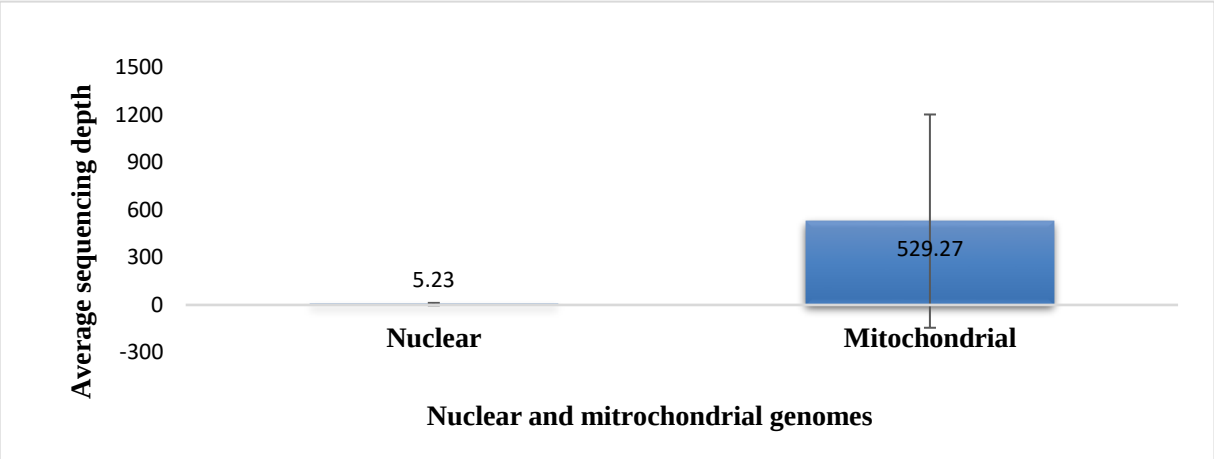

Fig 3. 3 Comparison of average sequencing depth of nuclear and mitochondrial genomes. The whiskers indicate the standard deviation from the mean.

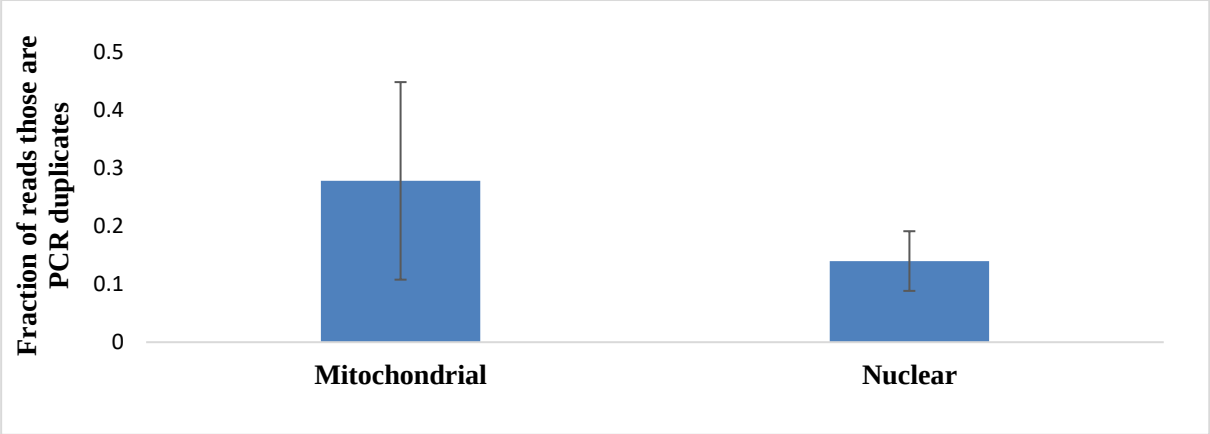

Fig 3. 4 Fraction of reads those are PCR duplicates for both mitochondrial and nuclear genome. Whiskers indicate standard deviations from the mean.

3.2 Phylogenetic analysis

The Bayesian phylogenetic tree for the mitochondrial genome (Fig 3.5) shows the categorization of isolates in six different lineages (IIa, IIb, III, US-1/1b, Ia and HERB-1). Two *Phytophthora* isolates (Pand_PaX (*P. andina*) and Pand_EC3425 (*P. mirabilis*)) were included as outgroups in this phylogenetic analysis. The root of the tree was dated to 640 (95% HPD (highest posterior density): 475-800) years ago. Similarly, the most recent common ancestor (MRCA) of lineage Ia dates 134 years ago (95% HPD: 100-150), and MRCA of HERB-1 lineage dates 234 years ago (95% HPD: 225-375). Likewise, the MRCA of lineage III dates 79 years ago (95% HPD: 50-100), MRCA of US-1/Ib dates 213 years ago (95% HPD: 180-260). In the same way, MRCA of IIa and IIb lineage dates to 143 (95% HPD: 100-200) and 256 (95% HPD: 200-320) years ago respectively.

Only one isolate (P10127) was categorized as mtDNA lineage IIb, which has 100% posterior support at the node (Fig 3.6). Other samples (PCZ098, PCO033, and PCO038) previously categorized as IIb [6] are now grouped in IIa lineage. Our study samples JFH43, SGFX, JFH48 and JFH100, Mich7035 and TL14, including two of the unpublished samples from Denmark (Pinf_C_13_7_28 and Pinf_E_13_8_102) group within the lineage IIa.

Only a few samples group within the lineage US-1/Ib (Fig 3.7). None of our study samples grouped in this category. All the samples under this lineage were also categorized as US-1/Ib in the previous studies [6,47]. Lineage Ib diverged from lineages Ia and HERB-1 with 99% posterior probability support at the node. Lineage Ia and HERB-1 were separated with posterior probability support of 99% at the node. From Fig 3.5, we can see that divergence of these two lineages occurred 254 years ago (95% HPD: 205-350). Many of our study samples grouped in the lineage Ia. However, none of the samples from Michoacán belong to this lineage. As can be seen in Table 3.1, six isolates from Juchitepec, seven isolates from San Geronimo, 12 isolates from Toluca, eight isolates from Tlaxcala, and two isolates from Chapingo group within lineage Ia. All the samples previously categorized as Ia [6,47] were also categorized as Ia in this study. In addition, the samples PIC97605, P8143, P6635, P6636, PIC97207, P3681 and P8144 which were previously ungrouped as a distinct lineage, group together with Ia lineage samples (Fig 3.8).

As mentioned previously, lineage HERB-1 diverged from all other lineages with 99% posterior probability support at the node. Clearly, the sample Pand_EC3394 (the red branch shown in the phylogenetic tree below (Fig 3.8)) was well differentiated, as its divergence from the rest of the group receives 99% posterior probability support at the node. This lineage was the most geographically diverse, with the samples from all six sampled regions of Mexico (Table 3.1). Altogether, 16 isolates from Juchitepec, two from Toluca, 10 from San-Geronimo, five from Tlaxcala, seven from Chapingo, and three from Michoacán group within lineage HERB-1 (Table 3.1). All the samples categorized as HERB-1 in the previous studies [6,7] were also categorized as HERB-1 in this study.

A unique lineage name was designated to one of the clade (just below lineage IIa) where most of the samples from Michoacán belong. The lineage, which was previously not given any lineage name [6], was designated as lineage III (Fig 3.6). Fourteen of our 18 isolates from Michoacán group within this lineage. In addition, three isolates from Chapingo, one from San-Geronimo, and five isolates from Tlaxcala group in this lineage. Thus, in total 23 study samples from four locations of Mexico group under this lineage (Table 3.1).

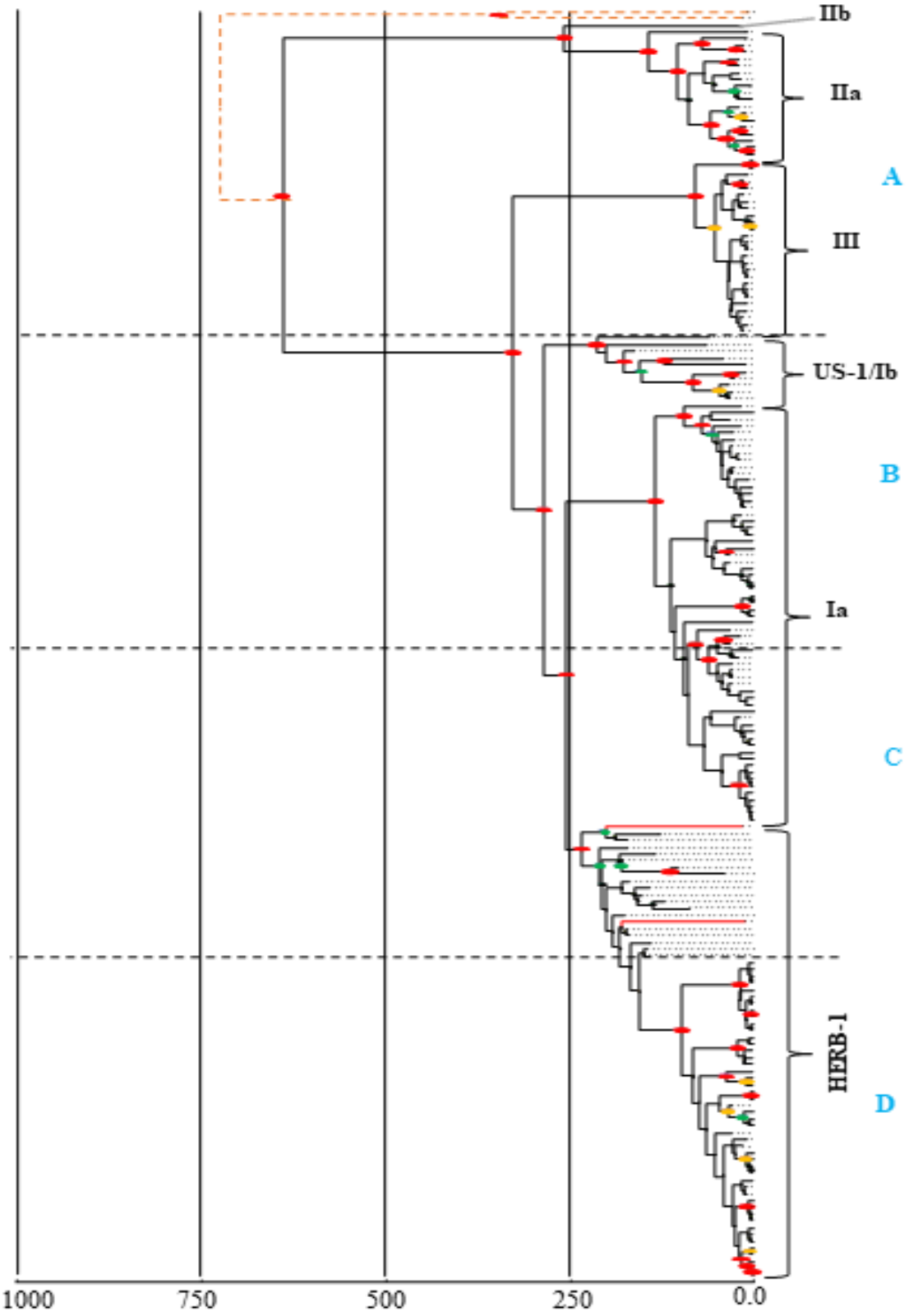
IIb
IIa
A
III
US-1/Ib
B
Ia
C
HERB-1
D
1000
750
500
250
0.0

Fig. 3. 5 The mitochondrial DNA phylogeny of *P. infestans* isolates from Mexico and other previously studied samples using Bayesian dating analysis. The major mtDNA lineages are indicated by the brackets on the right side of the tree. Red branches indicate *P. andina* isolates with *P. infestans*-like mitogenomes. The orange dotted lines indicate the position of two outgroups included in this analysis (branch length does not represent the true length). The posterior probability support is indicated by different coloured ovals at the node (90-100% represented by red, 75-90% indicated by yellow, and 50-75% indicated by green). The phylogenetic tree is partitioned into four parts (A, B, C and D) as shown on the right side for easier visualization of individual parts (see Fig 3.6, Fig 3.7, Fig 3.8, Fig 3.9). The sample ID of the isolates are removed because they are hard to visualize. The time scale of nodes and branch lengths is shown at the bottom of the tree (years before 2016).

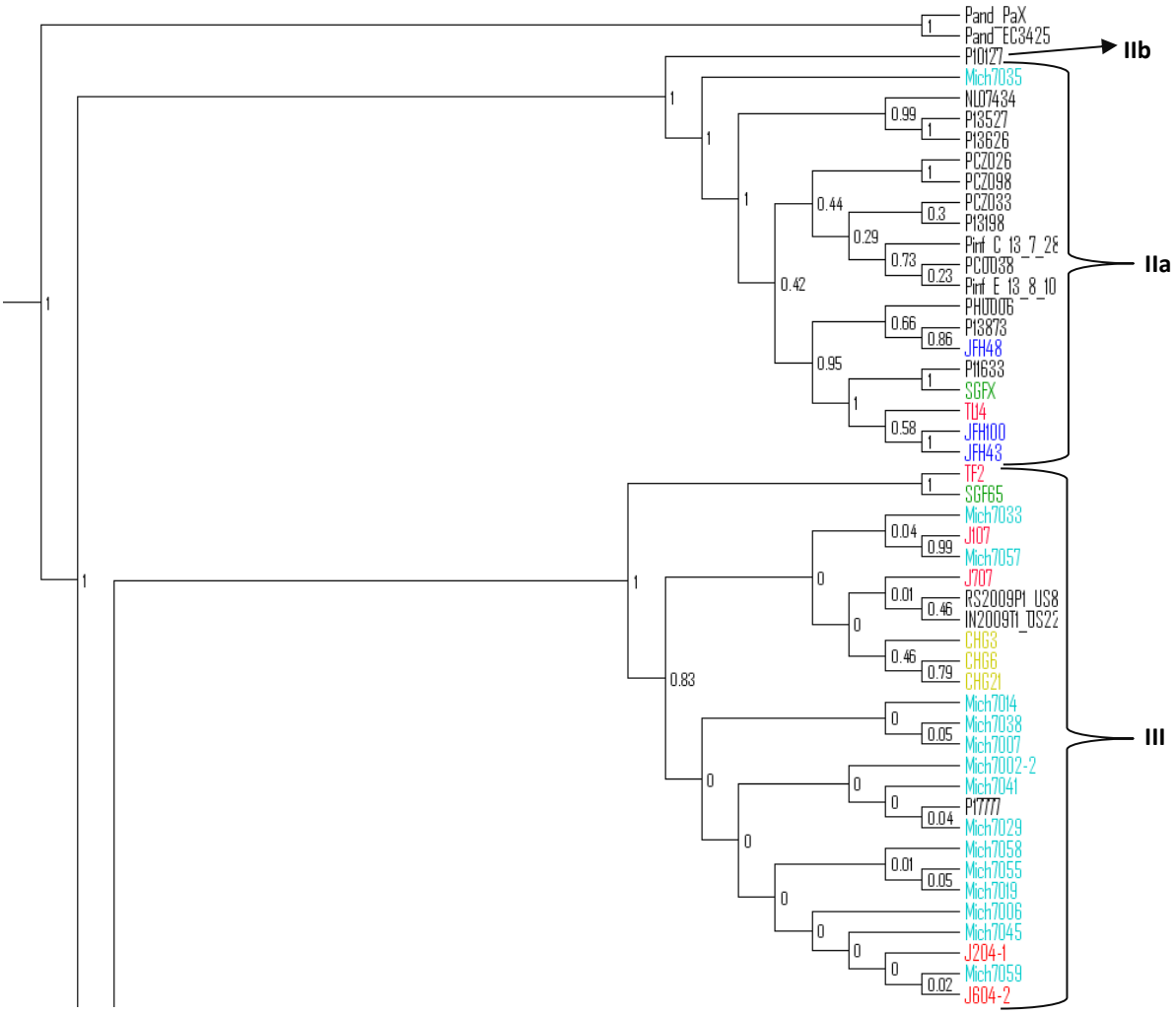

Fig. 3. 6 Section of the phylogenetic tree representing mtDNA lineages IIb, IIa and III. This section is represented by A in Fig 3.5. The scale axis is not included to make it easier to visualize. The colour of isolates corresponds to the geographical locations where they were collected (Fig 2.2).

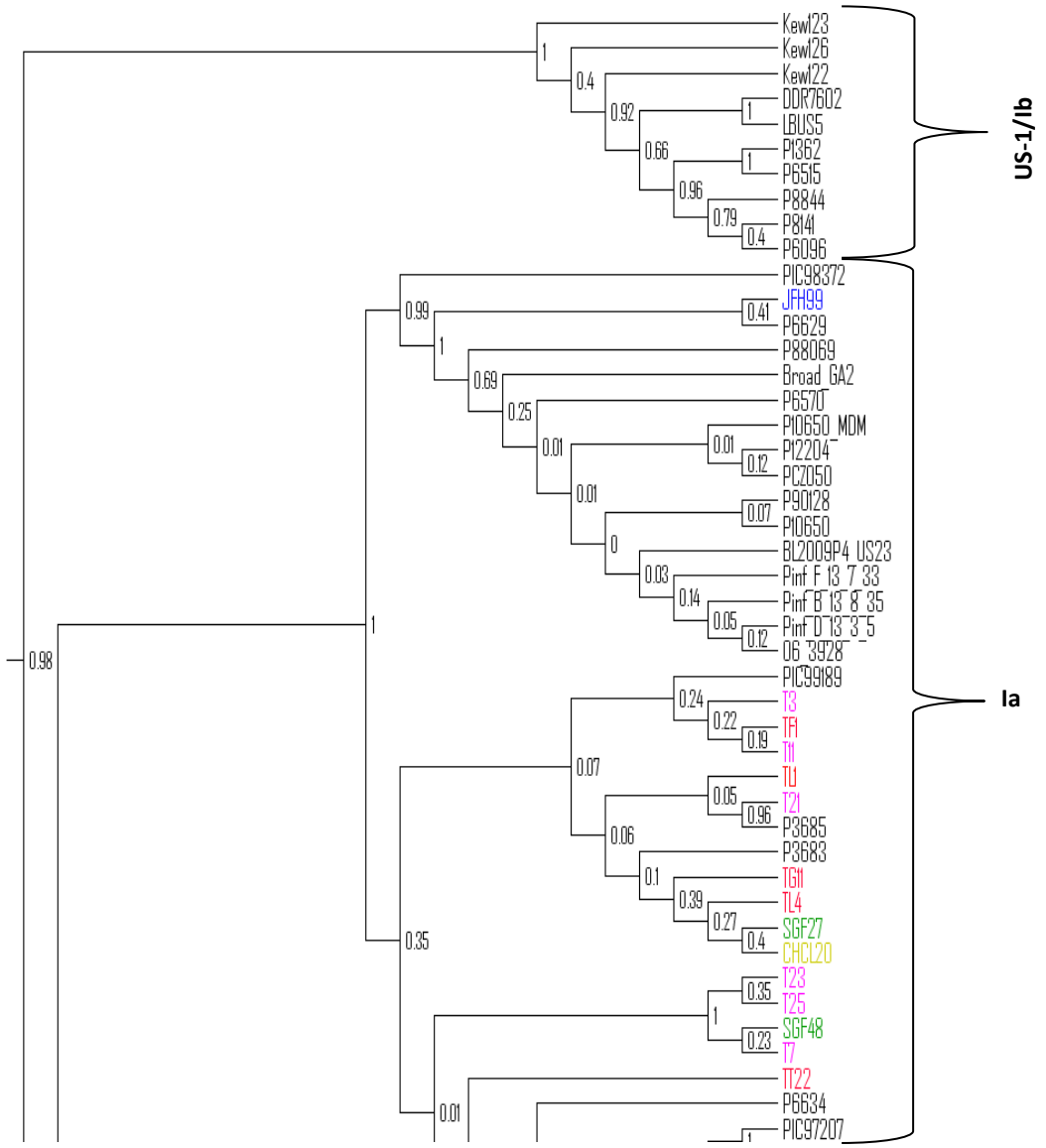

Fig. 3. 7 Section of the phylogenetic tree representing mtDNA lineages US-1/Ib and a part of lineage Ia. This section is represented by the B in Fig 3.5. The scale axis is not included to make it easier to visualize. The colour of isolates corresponds to the geographical locations where they were collected (Fig 2.2).

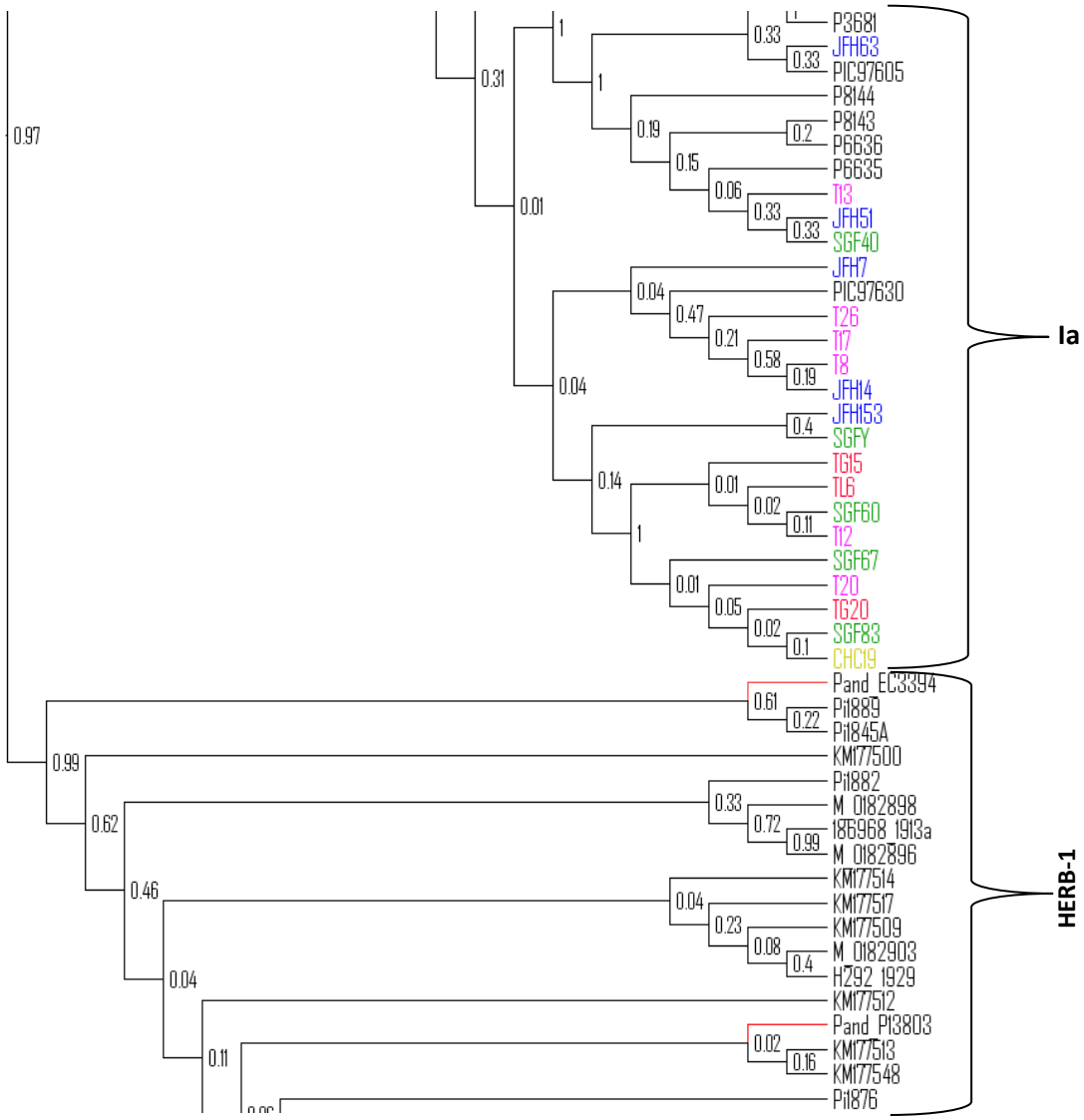

Fig 3. 8 Section of the phylogenetic tree representing part of mtDNA lineage Ia and part of HERB-1. This section is represented by the C in Fig 3.5. The scale axis is not included to make it easier to visualize. The colour of isolates corresponds to the geographical locations where they were collected (Fig 2.2).

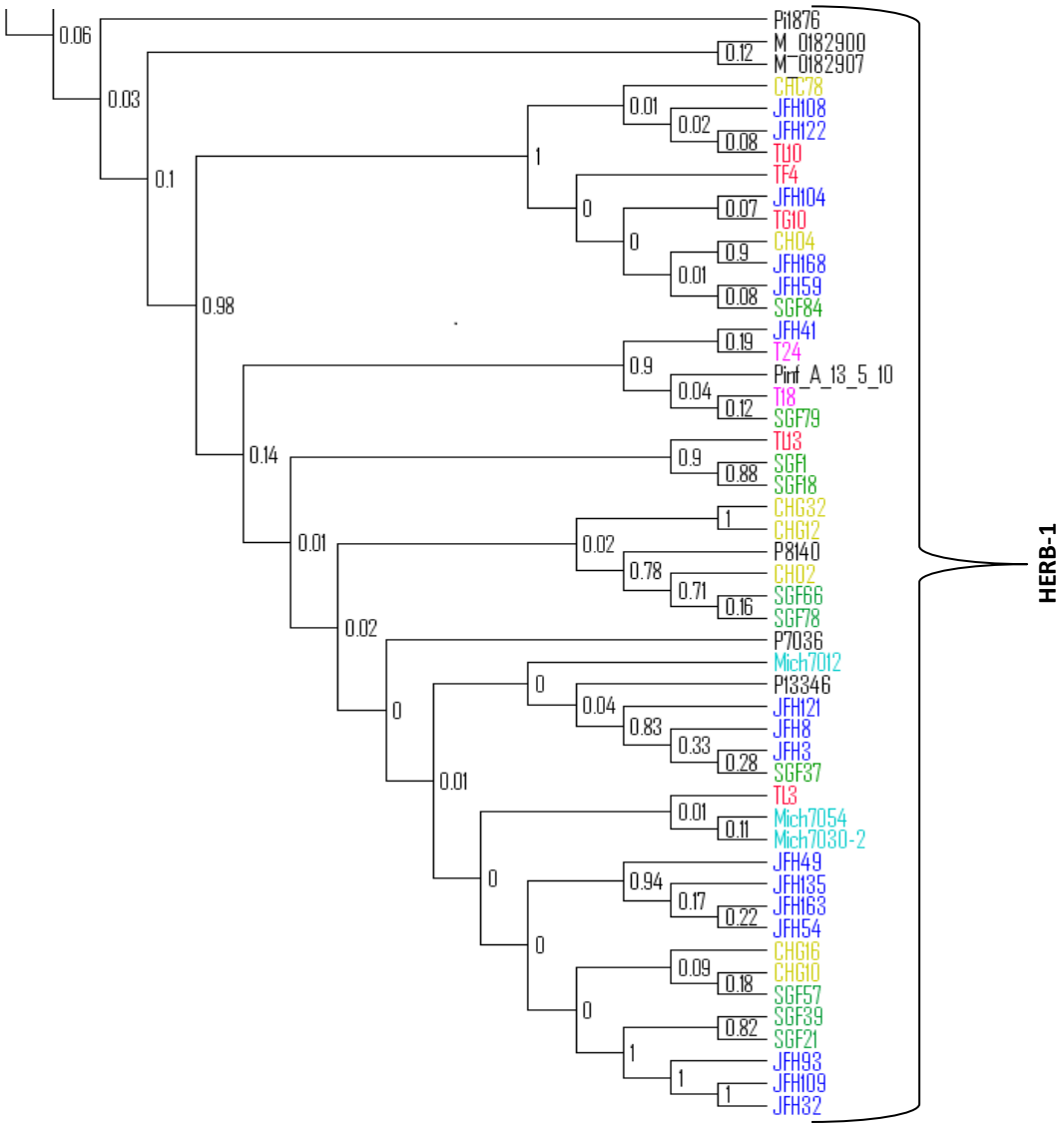

Fig 3. 9 Section of the phylogenetic tree representing a part of mtDNA lineage HERB-1. This section is represented by D in Fig 3.5. The scale axis is not included to make it easier to visualize. The colour of isolates corresponds to the geographical locations where they were collected (Fig 2.2).

3.3 Categorization of study isolates according to mtDNA lineages

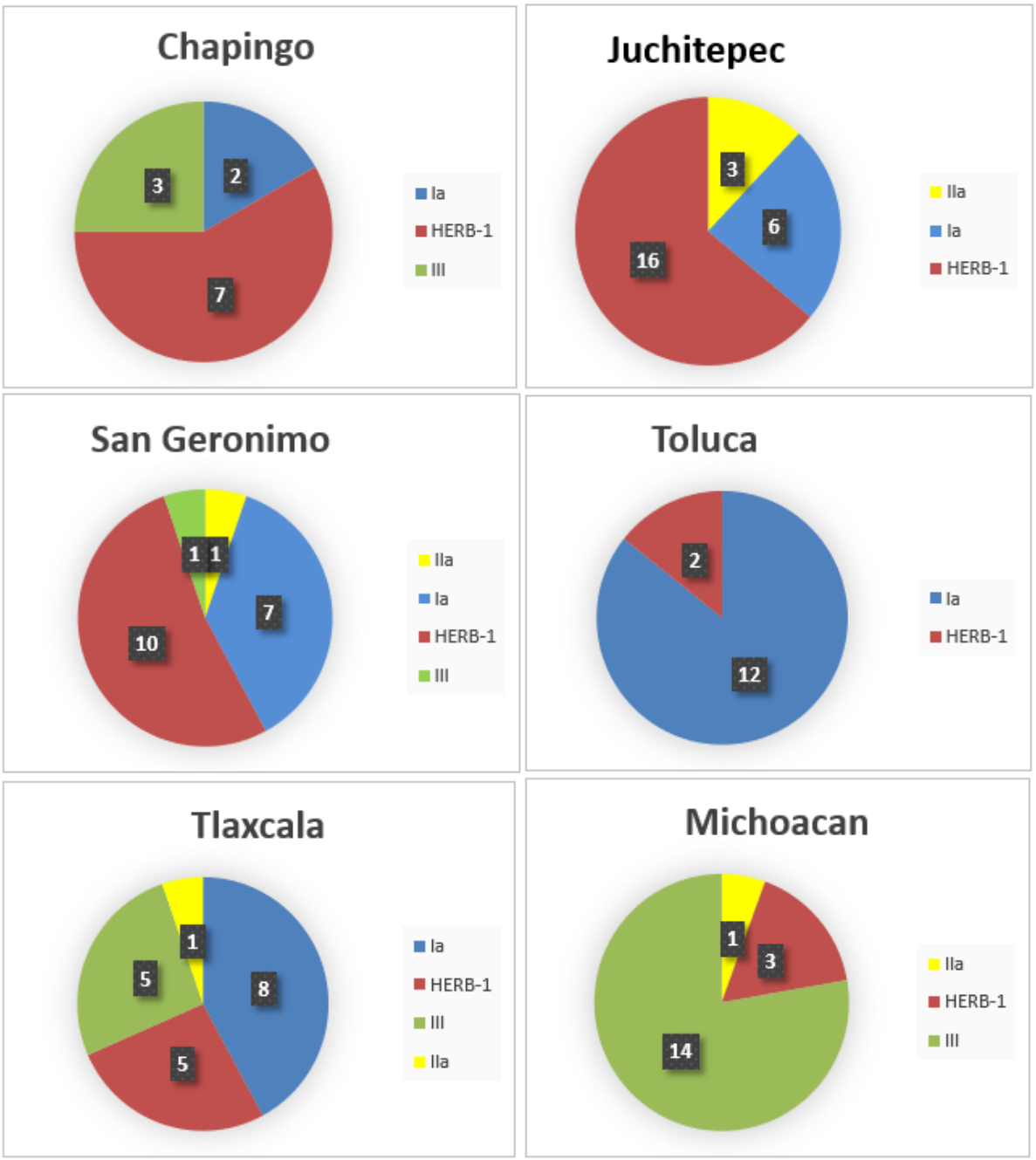

Fig 3. 10 The pie chart represents the categorization of mtDNA lineages in each of the six Mexican regions where our study samples come from. Also displayed is the number of isolates carrying each major mtDNA lineage in each location.

Table 3. 1 Table showing grouping of all the samples from different geographical locations to their respective mitochondrial DNA lineages.

Locations of sample collection	Mitochondrial DNA lineages					
	US-1/Ib	IIb	IIa	Ia	HERB-1	III
Chapingo	-	-	-	CHC19, CHCL20	CHC78, CH04, CHG32, CHG12, CHG16, CHG10, CH02	CHG3, CHG6, CHG21
Juchitepec	-	-	JFH48, JFH43, JFH100	JFH7, JFH153, JFH14, JFH99, JFH63, JFH51,	JFH122, JFH59, JFH104, JFH108, JFH168, JFH121, JFH3, JFH8, JFH93, JFH32, JFH109, JFH41, JFH49, JFH163, JFH54, JFH135	-
San-Geronimo	-	-	SGFX	SGFY, SGF67, SGF83, SGF48, SGF60, SGF27, SGF40	SGF1, SGF18, SGF84, SGF78, SGF66, SGF37, SGF21, SGF39, SGF79, SGF57	SGF65
Toluca	-	-	-	T13, T8, T26, T17, T12, T20, T7, T25, T23, T21, T3, T11	T18, T24	-
Tlaxcala	-	-	TL14	TT22, TL6, TG20, TG15, TL1, TG11, TL4, TF1	TL13, TF4, TG10, TL10, TL3	TF2, J107, J204-1, J604-2, J707
Michoacan	-	-	Mich7035	-	Mich7054, Mich7030-2, Mich7012	Mich7057, Mich7033, Mich7038, Mich7059, Mich7045, Mich7029, Mich7006, Mich7002-2, Mich7014, Mich7058, Mich7019, Mich7007, Mich7041, Mich7055

A unique pattern of sample distribution was observed in Juchitepec, Toluca, and Michoacán. In Juchitepec, 16 of 25 isolates belong to the HERB-1 mtDNA lineage (Fig 3.10). Similarly, in Toluca 12 of 14 samples belong to the Ia mtDNA lineage. In the same way, a unique pattern of sample distribution was seen in Michoacán. Out of the 18 samples taken into consideration for our study 14 of them were found to be closely related to the previously

categorized mtDNA samples without any lineage and are now placed under lineage III (see also Table 3.1). Thus, in total none of the study samples was categorized as IIb and US-1/Ib, six samples were categorized as IIa, 35 isolates were categorized as Ia, and a substantial portion of isolates (43) were categorized as HERB-1.

3.4 Characterization of novel RXLR effectors

Shotgun sequencing data from the 14 genomes sequenced to high depth was used for the identification of novel RXLR effector genes via *de novo* assembly of unmapped reads. For each of these samples, the number of input unmapped reads, assembly N_{50}, the number of assembled contigs having a length more than 200bp, length of the longest contig, and the total length of the assembled contig (bp) were calculated (Table 3.2). The N_{50} (weighted median contig size) for all samples ranged between 300-500 bp. The total length of assembled contigs for each sample was between 18-36 Mbp.

Table 3. 2 Statistics for *de novo* assemblies of the unmapped reads.

Assembly	**Number of input unmapped reads**	**Assembly N_{50} (bp)**	**Number of assembled contigs longer than 200bp**	**Largest contig length (bp)**	**Total length (Mbp)**
CHG3	446,219	404	86,894	14,524	34.5
SGF78	742,467	459	73,144	20,153	32.9
CHO2	1,226,892	466	75,173	27,419	34.1
JFH109	228,666	432	64,880	41,545	28.0
SGFX	830,196	464	78,992	39,643	36.0
T11	435,869	449	55,192	56,568	24.7
JFH32	1,589,115	480	76,649	27,189	35.7
JFH99	152,569	367	63,309	14,483	23.7
T17	128,229	383	66,230	18,466	26.0
Mich7014	152,146	407	56,415	28,698	23.2
T12	172,105	401	66,877	14,767	26.8
TG10	53,771	385	69,609	13,830	27.3
Mich7006	80,427	336	51,845	10,763	18.2
TF1	230,516	442	62,160	22,627	27.3

There were altogether 18 ORFs that satisfied the criteria of D score ≥ 0.9, signal peptide within 1-30 residues and RXLR motif within 25-60 residues. However, after the BLAST search, there were only 10 peptide sequences not having a perfect match to the reference genomes, which is strong support of these genes encoding novel RXLR proteins.

Table 3. 3 Table indicating de novo assembly of unmapped reads for the detection of novel, non-reference RXLR proteins. The RXLR and DER motifs are highlighted. Also, amino acids S, K and Q at position 19, 80 and 103 respectively of amino acid sequences are highlighted.

Novel gene ID	**Note**	**Closest match**		**Peptide Length**	**Peptide sequence**
		Gene ID	**Amino acid identity (%)**		
PiRXLR 1	Found in isolate CHG3	Pex50259_Cooke etal2012	42.52	153	MPLYLFVWLSVIALLLAA**S**SARATNSDFY**RSLR**WSREIP TTRFPQDSNALSLEN**DER**AGIPVSAIEKIKALLQMSSVSP **K**TLRHWVNKGKSKNAVFIRLKLD**Q**AGDNLLSNPQFVT WVAYADDFNAKFSEKATPLLSTLKAQYRDEVLSEILIA GKKVPSTEKLASRLQAEQLEGWVIAKLPKGEVFKLLKL NDKLETFLTNPQLSTWVEYMEKFVKKYPDEKTPMIKEL TKSYGNEMLARMLATADSYVKDCPDGGLESQWEIP
PiRXLR 2	Found in isolate CHO2	PiRXLRd_Martin etal2013	95.45	196	MRFMRLLLLLIAFAMLNIAKAAELDHTSSDCGG**RQLR**T ITTNDKEQRSLSIPSAIESGITRWRIKIWMRNKMPDHSVL EKLSLAGVTEKTLTRDPKFKIFQKVKVESWLKEKASTT KVWEKLGLHRLPITDVEKADEFSTFVKYVAALNEKAK NIQFKKWPKPFEGGSPEEMAAKSYILKKLKRNDIDRRIM LGVFM
PiRXLR 3	Found in isolate CHO2	PITG_15032T0	41.21	264	MPLYLFVWLSVIALLLAA**S**SARATNSDFY**RSLR**WSREIP TTRFPQDSNALSLEN**DER**AGIPVSAIEKIKALLQMSSVSP **K**TLRHWVNKGKSKNAVFIRLKLD**Q**AGDNLLSNPQFVT WVAYADDFNAKFSEKATPLLSTLKAQYRDEVLSEILIA GKKVPSTEKLASRLQAEQLEGWVIAKLPKGEVFKLLKL NDKLETFLTNPQLSTWVEYMEKFVKKYPDEKTPMIKEL TKSYGNEMLARMLATADSYVKDCPDGGLESQWEIP
PiRXLR 4	Found in isolate	PiRXLRd_Martin etal2013	97.12	196	MRFMRLLLLLMAFAMLNITKAAELDHTSSDCGG**RQLR** TITTNDKEQRSLSIPSAIESGITRWRIKIWMRNKMPDHSV LEKLSLAGVTEKTLTRDPKFKIFQKVKVESWLKEKAST

	JFH99				TKVWENLGMHRLPFTDVEKADEFSTFVKYVAALNEKA KNIQFKKWPKPFEGGSPEEMAAKSYILKKLKRNDIDRRI MLGVFM
PiRXLR 5	Found in isolate JFH109	PITG_16 283T0	38.64	260	MPLYLFVWLSVIALLLTA**S**SARATNSDFY**RSLR**WSREIP TTRFPQDSNALSLEN**DER**AGIPVSAIEKIKALLQMSSVSP **K**TLRHWVNKGKSKNAVFIRLKLD**Q**AGDNLLSNPQFVT WVAYADDFNAKFSEKATPLLSTLKAQYRDEVLSEILIA GKKVPSTEKLASRLQAEQLEGWVIAKLPKGEVQVNDK LETFLTNPQLSTWVEYMEKFVKKYPDEKTPMIKELTKS YGNEMLARMLATADSYVKDCPDGGLESQWEIP
PiRXLR 6	Found in isolate JFH109	PiRXLR d_Martin etal2013	95.45	196	MRFMRLLLLLIAFAMLNIAKAAELDHTSSDCGG**RQLR**T ITTNDKEQRSLSIPSAIESGITRWRIKIWMRNKMPDHSVL EKLSLAGVTEKTLTRDPKFKIFQKVKVESWLKEKASTT KVWEKLGLHRLPITDVEKADEFSTFVKYVAALNEKAK NIQFKKWPKPFEGGSPEEMAAKSYILKKLKRNDIDRRIM LGVFM
PiRXLR 7	Found in isolate SGF78	PiRXLR d_Martin etal2013	95.45	196	MRFMRLLLLLIAFAMLNIAKAAELDHTSSDCGG**RQLR**T ITTNDKEQRSLSIPSAIESGITRWRIKIWMRNKMPDHSVL EKLSLAGVTEKTLTRDPKFKIFQKVKVESWLKEKASTT KVWEKLGLHRLPITDVEKADEFSTFVKYVAALNEKAK NIQFKKWPKPFEGGSPEEMAAKSYILKKLKRNDIDRRIM LGVFM
PiRXLR 8	Found in isolate SGF78	PITG_16 283T0	38.64	260	MPLYLFVWLSVIALLLTA**S**SARATNSDFY**RSLR**WSREIP TTRFPQDSNALSLEN**DER**AGIPVSAIEKIKALLQMSSVSP **K**TLRHWVNKGKSKNAVFIRLKLD**Q**AGDNLLSNPQFVT WVAYADDFNAKFSEKATPLLSTLKAQYRDEVLSEILIA GKKVPSTEKLASRLQAEQLEGWVIAKLPKGEVQVNDK LETFLTNPQLSTWVEYMEKFVKKYPDEKTPMIKELTKS YGNEMLARMLATADSYVKDCPDGGLESQWEIP
PiRXLR 9	Found in isolate T11	PITG_16 283T0	38.64	264	MPLYLFVWLSVIALLLAA**S**SARATNSDFY**RSLR**WSREIP TTRFPQDSNALSLEN**DER**AGIPVSAIEKIKALLQMSSVSP **K**TLRHWVNKGKSKNAVFIRLKLD**Q**AGDNLLSNPQFVT WVAYADDFNAKFSEKATPLLSTLKAQYRDEVLSEILIA GKKVPSTEKLASRLQAEQLEGWVIAKLPKGEVFKLLKL NDKLETFLTNPQLSTWVEYMEKFVKKYPDEKTPMIKEL TKSYGNEMLARMLATADSYVKDCPDGGLESQWEIP
PiRXLR 10	Found in	PITG_15 032T0	41.21	264	MPLYLFVWLSVIALLLAA**S**SARATNSDFY**RSLR**WSREIP TTRFPQDSNALSLEN**DER**AGIPVSAIEKIKALLQMSSVSP

	isolate T12				**K**TLRHWVNKGKSKNAVFIRLKLD**Q**AGDNLLSNPQFVT WVAYADDFNAKFSEKATPLLSTLKAQYRDEVLSEILIA GKKVPSTEKLASRLQAEQLEGWVIAKLPKGEVFKLLKL NDKLETFLTNPQLSTWVEYMEKFVKKYPDEKTPMIKEL TKSYGNEMLARMLATADSYVKDCPDGGLESQWEIP

Novel RXLR effectors were identified in only seven (CHG3, CH02, JFH99, JFH109, SGF78, T11, and T12) of 14 high-depth nuclear genome sequences analyzed in this study (Table 3.3). A unique ID was given to each of these putative RXLR effectors. Ten genes from these seven isolates' nuclear genomes were identified as putative novel RXLRs (Table 3.3). All the novel genes with encoded proteins have an RXLR motif occurring between 25-60 residues. However, none of the novel RXLR proteins has dEER or EER motif in the N-terminus. Six of the 10 novel effector proteins (PiRXLR1, PiRXLR3, PiRXLR5, PiRXLR8, PiRXLR9, and PiRXLR10) have an RXLR motif as well as a DER motif (rather than an EER motif) in the N-terminal. The RXLR-DER motif is the N-terminal conserved motif in RXLR effector protein sequences. However, in four of our novel RXLR protein sequences (PiRXLR2, PiRXLR4, PiRXLR6, and PiRXLR7), we only found the RXLR motif, but not the dEER or EER motif. In all the novel effector proteins with N-terminal RXLR-DER motifs, a unique pattern of amino acid S (19), K (80) and Q (103) (amino acid S at position 19, K at position 80 and Q at position (103)) were identified. In addition, we characterized these novel RXLR effectors by determining the RXLR families/classes to which our novel RXLR genes belong. Altogether, 90 RXLR gene clusters were discovered in the T30-4 reference genome annotation (Fig 3.11). Of our 10 novel RXLR gene sequences, six belong to the largest RXLR gene cluster, which consists of 49 genes. The remaining four novel RXLR gene sequences belong to a smaller RXLR gene cluster consisting of 7 genes.

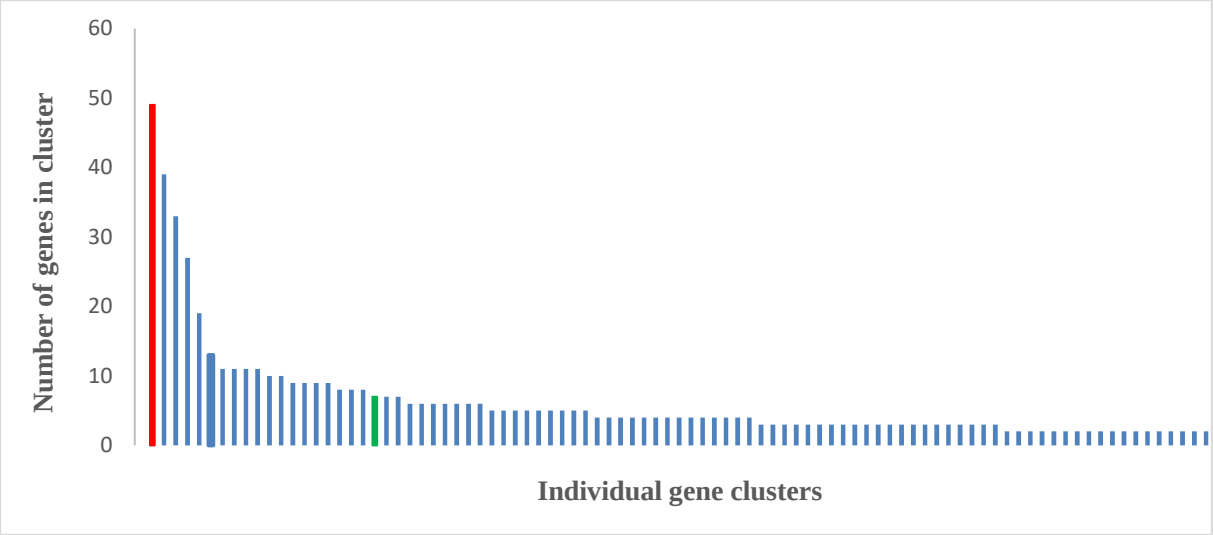

Fig 3. 11 Graph showing number of the gene clusters. The red and green bars show the gene clusters to which our novel RXLR effector proteins belong. The red cluster contains six effector genes, and the green cluster contains four effector genes.

CHAPTER 4

Discussion

The majority of genetic studies that have examined *P. infestans* are limited to only the Toluca Valley region, which is considered to be the centre of origin of this pathogen. Only very few studies encompass samples from different locations of Mexico (not just Toluca Valley), and all of them have used traditional, non-genomic methods to draw their conclusions. This study has expanded a previous phylogenomic analysis of *P. infestans* (Martin et al. 2016) from global sources to incorporate present-day *P. infestans* samples from six different locations in Mexico. In addition, this study has also discovered and characterized novel RXLR effectors (from nuclear genome sequences) from deep-sequenced samples present in six different locations of Mexico.

4.1 Phylogenetic analysis

We generated and assembled genomic data from 187 samples (107 present-day Mexican isolates and 80 previously sequenced genomes) for this study. This is by far the largest number of mitochondrial genome sequences assembled for a phylogenetic analysis of *P. infestans*. Previously, Martin et al. (2016) analyzed 71 mitochondrial sequences from global sources for their study of mtDNA phylogenetics.

Our analysis grouped our study isolates into six major mtDNA lineages (Fig 3.5). A unique mtDNA lineage name (III) was given to one of the lineages, which includes the majority of isolates from Michoacán (Fig 3.6). The clade also contains previously studied samples IN2009T1_US22, RS2009P1 and P17777. These three samples were not given any mitogenome lineage name by previous studies [6,8]. Similarly, Yoshida et al. (2013) also had not categorized the sample P17777 into any mitogenome lineage. Martin et al. (2016) categorized IN2009T1_US22, RS2009P1, and P17777) into group AGG ("aggressive isolates of global origin") according to a nuclear genome phylogeny. From this, we can speculate that the study samples under mtDNA lineage III, which includes most samples from Michoacán, are related to other, well known aggressive *P. infestans* isolates. However, to reach a firm

conclusion this needs to be confirmed with a phylogenetic analysis of the nuclear genomes in our study group.

One major discrepancy between this study and previous work is the grouping of particular samples within lineage IIb and IIa. Three samples (PCZ098, PCZ033, and PCO038) previously categorized as IIb by Martin et al. (2016) were instead categorized by us as IIa. In our phylogenetic tree (Fig 3.6), these samples are closely related to samples like PCZ026 and P13198 from the IIa lineage. For instance, PCZ033 and P1398 diverge from each other with only 30% posterior node support, which indicates they are genetically related (rather than distinct). PCZ026 and PCZ098 diverge from a common node with 100% node support. However, these two samples diverge from the rest of the IIa lineage with only 44% posterior support, indicating they should be grouped together with the samples from IIa lineage. Moreover, samples PCZ033, PCO038 and PCZ098 lie in the same clade as samples from the IIa lineage, indicating their genetic nearness.

However, in this phylogenetic analysis, we used only one set of Bayesian model parameters (HKY+G nucleotide substitution, flexible skyride coalescent model for prior distribution and a strict molecular clock) to generate the phylogenetic tree, but the choice of prior distribution can strongly affect the posterior (result). Therefore, it is important to test the sensitivity of our phylogenetic analysis to different prior distributions [70]. In a similar study, Yoshida et al. (2013) tested five different models and used evolutionary birth-death model (which better suited their data) to generate their results. Thus, drawing firm conclusions about the correct phylogenetic topology and node dating necessitates further analysis of our data with different choices of prior distributions.

Our study is consistent with several previous studies in that we find the US-1/Ib haplotype to be rare in Mexico (Fig 3.7). Nine of the 11 samples (Kew123, Kew126, Kew122, DDR7602, LBUS5, P13626, P6515, P8844, P6096, P8141) which were also previously published [6,8] and categorized under the lineage Ib are from the non-Mexican sources (P8141 and P1362 being an exception). This is consistent with previous findings that US-1/Ib haplotypes are not found widely in the extant Mexican population of *P. infestans* [43,44,57]. This finding is also consistent with a study by Garay-Serrano et al. (2007). They studied 14 samples of *P. infestans* from central Mexico and found only two of them carried the Ib haplotype. The basis of their haplotype determination was amplification and digestion of P2

and P4 regions and comparison of amplicons to reference strains of a known haplotype. On the other hand, our Bayesian phylogeny supports the fact that the Ia and HERB-1 lineages are more closely related than the HERB-1 and US-1/Ib lineages. However, a recent review article claims that US-1/Ib emerged from HERB-1 lineage [81]. Our claim is based on the fact that most common ancestor from which the Ia and HERB-1 emerge is relatively younger (254 years ago) than that of common ancestor where US-1/Ib and HERB-1 emerge (285 years ago). This claim is further supported by the earlier branching of US-1/b and other two lineages (HERB-1 and Ia) with 99% posterior support at their ancestral node. This is consistent with recent findings by Martin et al. (2014) and Martin et al. (2016).

Twelve of 14 study samples from Toluca fall within the Ia mtDNA lineage (Fig 3.10). The remaining two Tolucan samples grouped within HERB-1 mtDNA lineage. This result is consistent with previous findings that the predominant population in the central highlands of Mexico (Toluca) is mtDNA lineage Ia [44]. This conclusion was drawn from a study of mtDNA haplotypes based on restriction fragment patterns of 170 isolates collected from three locations of Toluca, which concluded that all of them belong to the Ia mtDNA lineage [44]. A similar result was obtained by Gavino and Fry (2002); they found mtDNA haplotype Ia was the dominant *P. infestans* haplotype in the samples from Toluca. Our study also demands the reassessment of the previous hypothesis that Toluca Valley contains the most diverse populations of *P. infestans* [58]. Our study clearly highlights that the other five Mexican *P. infestans* populations (rather than Toluca) are more diverse (Table 3.1/Fig 3.10), at least according to their mtDNA haplotypes. However, we only analyzed 14 samples from Toluca. A larger population genetic analysis would be required before countering previous claims that the Toluca Valley *P. infestans* population is the most diverse [41,82,83]. On the other hand, our study is consistent with previous work showing that isolates with Ia mtDNA haplotypes are observed in many locations world-wide [43,84,85]. We found high frequencies of Ia haplotypes in our five different Mexican study locations as well as world-wide (Table 3.1, see also Table A.1).

Our study further bolsters previous findings that the HERB-1 lineage is neither extinct nor rare [6,8]. This assertion is based on the observation that 40% of our Mexican samples were categorized within the HERB-1 mtDNA lineage (Table 3.1), which contrasts with the claim of Yoshida et al. (2013). In their article, they claimed that the HERB-1 lineage is extinct

or rarely present in modern times. Our study also agrees with the finding of Martin et al. (2014) that HERB-1 mitogenomes diverged before their introduction to Europe. This agreement is based on two observations. First, two 19th-century historic samples (Pi1845A and Pi1889, dating to 1845 and 1889, respectively) diverged earlier than other HERB-1 lineage samples. Second, the estimated age of most recent common ancestor (MRCA) of the HERB-1 lineage is 234 (95% HPD: 225-375) years. This age of MRCA HERB-1 lineage is older than the first appearance of late blight in the U.S (1843) [86] and Europe (1845) [87] (this age (234 years) is also older by 52 years than the 182-years estimate by Yoshida et al. (2013). Furthermore, our study is consistent with some previous studies that HERB-1 haplotype was the predominant haplotype in historical late blight epidemics [6,8,46,47]. This assertion is based on two findings in our study. First, the historical specimens collected from the time of late blight epidemics included in our study, group within HERB-1 lineage. Second, 40% of our study samples, as well as other previously studied samples grouped in this lineage, are from the diverse global sources.

The root of our phylogenetic tree (common ancestor of all *P. infestans* mitogenomes) dates to 640 years ago (95% HPD: 475-800) (Fig 3.5). This age does not correspond with the result of Yoshida et al. (2013), where the root dated to 460 years ago (95% HPD: 853-1,863). However, our result agrees better with the result of Martin et al. (2016), where the root of the phylogenetic tree from mitogenome sequences from the global sources dated to 650 years ago. Also, this result concurs with the previous result where the phylogenetic tree was rooted back to 1470 A.D (638 years ago, 95% HPD: 313-1,022) [8]. However, the 95% HPD of Yoshida et al. (2013) overlaps with our 95% HPD range. There are several factors that can lead evolutionary rates to be high or low when they are calibrated using the sample age [88,89]. Some of the factors that affect the evolutionary rate are mutation over very short time scales, selection, random genetic drift, and high levels of homoplasy [90]. An alternative explanation for this inconsistency might be that we used the flexible skyride coalescent for prior distribution and did not partition our sequence data before phylogenetic analysis. This may affect the estimation of the evolutionary age of the tree [70]. A possible reason for this is that different parts of the genome evolve at different rates and with different patterns (e.g. nucleotide substitution rates and base composition biases) [91–93]. In addition, lateral gene

transfer, gene duplication and loss, and ancestral polymorphism affect the evolutionary rate [94,95].

This mtDNA phylogenomic study provides useful insights into past patterns of migration of this important pathogen. In this study, the most divergent mtDNA lineage (type II), basal to all others, contains samples from Mexico, South America, the United States, as well as Europe (Fig 3.6, see also Table A.1), which does not support the previous findings, based on much more limited genetic data and number of samples, that lineage I and lineage II necessarily diverged in South America [40,96]. Two recent studies (Martin et al. 2014; Martin et al. 2016) also implicitly agree with the claim that lineage type I and type II diverged in South America. These studies probably reached this conclusion because none of the samples from Mexico grouped under the type II lineage. We claim that divergence of type I and II lineage did not take place in South America because our study found six Mexican isolates grouped in the lineage IIa. All lineages except IIb can be found in Mexico, and the ages of all the lineages are different, the oldest being IIb (MRCA 257 years ago) and HERB-1 (MRCA 234 years ago). Still, we cannot say with confidence where the divergence of lineage I and lineage II took place.

4.2 Characterization of novel RXLR effectors

We discovered 10 novel RXLR effector genes from high-depth sequenced nuclear genomes in this study (Table 3.3). However, validation of the putative novel RXLR effectors by PCR was not done as in the previous study [46], which is one of the limitations of this study. None of our novel effectors follows the classical definition of RXLR gene [97], as none contain a dEER or EER motif. As a principle, N-terminal end of the RXLR effector should have RXLR-EER (or RXLR-dEER motifs). Six of our 10 novel RXLR effectors sequences contain a RXLR-DER in the N-terminal motif (DER rather than EER) (Table 3.3). We followed the selection of RXLR effectors based on RXLR-DER as done by Haas et al. (2009) in their studies [13]. However, Haas et al. (2009) used a flexible method where they selected RXLR effectors even if E's in EER motif replaced by D, and R replaced by K. Furthermore, the gene cluster analysis using software mcl led us to obtain 90 RXLR gene clusters. This number is slightly larger than previous RXLR clusters analysis by (Martin et al. 2013), which identified

84 RXLR gene clusters). The likely reason for this discrepancy is that we also included in our cluster analysis new RXLR genes discovered more recently (Martin et al. 2013; Cooke et al. 2012). In our six novel RXLR effectors with RXLR-DER motifs, a unique pattern of amino acid was observed (S^{19} K^{80} Q^{103}, where superscript refers to the position of amino acids in the RXLR effector peptide sequence). This pattern is similar to the amino acid pattern of two alleles of AVR3a effector, $AVR3a^{EM}$ (S^{19} E^{80} M^{103}) and $AVR3a^{KI}$ (C^{19} K^{80} I^{103}) [23,98]. However, BLAST search of these RXLR genes with previously known RXLR gene do not match to any of the known Avr3a or Avr3a-like RXLR genes. On top of this, the cluster analysis reveals that none of these genes having Avr3a-like pattern (amino acid pattern like Avr3a genes) cluster with known Avr3a like genes. All these six RXLR effectors fall in the larger cluster having 49 RXLR genes (Fig 3.11) and all other genes in this cluster are not classified as any secreted RXLR effector family. We also speculate that this larger cluster is highly dynamic and that surveying more isolates would identify many new genes belonging to this cluster. In addition, the clustering of four putative novel RXLR effectors (without the N-terminal EER or DER motif) with the RXLR gene not having EER motif (PiRXLRd) by Martin et al. (2013) signifies that they are related. They form a small cluster of seven genes including our novel genes (Fig 3.11). Two of them are from the reference genome T30-4 (PITG_06432 and PITG_16663), and they also do not have N-terminal EER motif. PITG_16663 is registered as an Avr1 RXLR gene in the NCBI gene bank and forms a cluster with our four novel RXLR genes (without EER or DER motifs), indicating those novel proteins are encoded by Avr1-like RXLR genes. However, this needs to be verified by further RXLR gene analysis. To date, there has been no characterization of the C-terminal region (W-Y-L domain) of RXLR effectors without the use of motif finding software (e.g. HMM and/or MEME) [13,28,99]. Due to lack of available time, we could not use these methods to characterize our putative novel RXLR effectors based on C-terminal conserved motifs.

4.3 Limitations of the study

The mitochondrial genome phylogenomic analysis in BEAST extended our knowledge on the evolutionary history and migration pattern of *P. infestans* using only a single combination of parameters (HKY+G nucleotide substitution, flexible skyride coalescent prior and strict molecular clock without the partitioning of data). But, a comparison of results from a panel of

multiple parameters (e.g. multiple coalescent models) may have provided a more confident estimation of the samples' evolutionary relatedness and may have resolved the contradictory phylogenetic root node age in comparison with a previous study. Verification of the IIa/IIb discrepancy of three samples (Fig 3.6) in comparison according to previous studies might be performed by using a panel of different parameters in BEAST as well as by using a maximum-likelihood phylogenetic method. A study of the nuclear genome phylogeny would have given us more confidence in finding whether samples grouping in type III lineage belongs to the AGG nuclear genome cluster. On the other hand, we were able to discover 10 novel RXLR effector genes and categorized them manually. However, use of motifs searching software like MEME and HMM would have made categorizing the RXLR effectors more efficient.

4.4 Recommendations for future work

The results of this study present us with various possibilities for further research. The tree topology can be further verified by using maximum-likelihood phylogenetic analysis of the mitogenome sequences (as in Martin et al. (2014) and Martin et al. (2016)). Similarly, it might also be possible to estimate the node ages of the of phylogenetic tree root and MRCA of various lineages with higher precision by partitioning the sequence data (as in Martin et al. (2014) and Martin et al. (2016)). Moreover, evolutionary histories may also vary among genes because of various processes like lateral gene transfer, gene duplication and loss, and due to ancestral polymorphism [94,95]. Therefore, when dealing with large molecular data, it is important to partition the genomic data into various parts (e.g. three codon positions, tRNA, rRNA, and intragenic sites as done by Martin et al. (2014) and Martin et al. (2016)) to account for diversity of evolutionary characteristics across different regions of genome [100,101]. However, partitioning of molecular data is not always the solution for the best estimate of the phylogenetic tree. Phylogenies generated from partitioned data are subjected to sampling variation [102]. This means that estimates from partitioned data are more variable. This also means that it is hard to find the same taxonomic grouping from estimates from different partitions. Therefore, heterogeneity of molecular data should be considered when partitioning them [102]. Since the model used for priors might affect the tree topology and age, we suggest to carry out the BEAST analysis using different models and then select the best model for the

study [81]. On the other hand, the novel RXLR effector genes discovered could be verified using direct PCR amplification (as in Martin et al. (2013)). We also recommend the use of software such as MEME and HMM, which assist in the efficient categorization of the RXLR effector genes motifs (e.g. C-terminal conserved W-Y-L motifs) [103].

CHAPTER 5

Conclusion

The root of our phylogenetic tree (640 years ago) was very close to recent similar studies. Three samples formerly categorized in lineage IIb were categorized as IIa based on node support value and their clustering in a single clade with other members of the IIa lineage. In addition, we assigned a new mtDNA lineage name (III) to describe a group of samples (mostly from Michoacán) that was previously unnamed. Since three previously studied samples under this lineage were described as AGG (aggressive isolates of the global sources), we expect that the samples within this lineage will also be assigned to the AGG nuclear genome cluster. This study further supports previous findings that HERB-1 haplotypes are neither rare nor extinct. Moreover, our work supports a few recent studies finding that Ib and HERB-1 lineages are more distant than HERB-1 and Ia. In agreement with previous studies, we also found that Mexico's Toluca Valley is enriched with the mtDNA lineage Ia, but is not exceptionally genetically diverse as previously suggested. From our results, we suggest that the divergence of lineage I and II did not take place in South America as previously suggested, and we cannot say with confidence where the divergence took place. Another important finding of this thesis is that we discovered and characterized 10 novel RXLR genes, which we further characterized using amino acid motifs and cluster analysis.

References

1. Keeling, P. J.; Burger, G.; Durnford, D. G.; Lang, B. F.; Lee, R. W.; Pearlman, R. E.; Roger, A. J.; Gray, M. W. The tree of eukaryotes. *Trends Ecol. Evol.* **2005**, *20*, 670–676.
2. Ristaino, J. B.; Groves, C. T.; Parra, G. R. PCR amplification of the Irish potato famine pathogen from historic specimens. *Nature* **2001**, *411*, 695–697, doi:10.1038/35079606.
3. Turner, R. S. After the famine: Plant pathology, Phytophthora infestans, and the late blight of potatoes, 1845—1960. *Hist Stud Phys Biol Sci* **2005**, *35*, 341–370.
4. Fry, W. E.; Grünwald, N. J. Introduction to oomycetes. The Plant Health Instructor. DOI: 10.1094. *Plant Heal. Instr.* **2010**.
5. Fry, W. E.; Birch, P. R. J.; Judelson, H. S.; Grünwald, N. J.; Danies, G.; Everts, K. L.; Gevens, A. J.; Gugino, B. K.; Johnson, D. A.; Johnson, S. B.; McGrath, M. T.; Myers, K. L.; Ristaino, J. B.; Roberts, P. D.; Secor, G.; Smart, C. D. Five Reasons to Consider *Phytophthora infestans* a Reemerging Pathogen. *Phytopathology* **2015**, *105*, 966–981, doi:10.1094/PHYTO-01-15-0005-FI.
6. Martin, M. D.; Vieira, F. G.; Ho, S. Y. W.; Wales, N.; Schubert, M.; Seguin-Orlando, A.; Ristaino, J. B.; Gilbert, M. T. P. Genomic characterization of a south American phytophthora hybrid mandates reassessment of the geographic origins of phytophthora infestans. *Mol. Biol. Evol.* **2016**, *33*, 478–491, doi:10.1093/molbev/msv241.
7. Yoshida, K.; Burbano, H. A.; Krause, J.; Thines, M.; Weigel, D.; Kamoun, S. Mining herbaria for plant pathogen genomes: back to the future. *PLoS Pathog.* **2014**, *10*, e1004028.
8. Martin, M. D.; Ho, S. Y. W.; Wales, N.; Ristaino, J. B.; Gilbert, M. T. P. Persistence of the mitochondrial lineage responsible for the Irish potato famine in extant new world Phytophthora infestans. *Mol. Biol. Evol.* **2014**, *31*, 1414–1420, doi:10.1093/molbev/msu086.
9. Flor, H. H. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* **1971**, *9*, 275–296.
10. Jones, J. D. G.; Dangl, J. L. The plant immune system. *Nature* **2006**, *444*, 323.
11. Kamoun, S. A catalogue of the effector secretome of plant pathogenic oomycetes.

Annu. Rev. Phytopathol. **2006**, *44*.

12. Whisson, S. C.; Boevink, P. C.; Moleleki, L.; Avrova, A. O.; Morales, J. G.; Gilroy, E. M.; Armstrong, M. R.; Grouffaud, S.; Van West, P.; Chapman, S.; Hein, I.; Toth, I. K.; Pritchard, L.; Birch, P. R. J. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **2007**, *450*, 115, doi:10.1038/nature06203.
13. Haas, B. J.; Kamoun, S.; Zody, M. C.; Jiang, R. H. Y.; Handsaker, R. E.; Cano, L. M.; Grabherr, M.; Kodira, C. D.; Raffaele, S.; Torto-Alalibo, T. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature* **2009**, *461*, 393.
14. Fry, W. E.; Goodwin, S. B. Resurgence of the Irish potato famine fungus. *Bioscience* **1997**, *47*, 363–371.
15. McLaughlin, D. J.; Hibbett, D. S.; Lutzoni, F.; Spatafora, J. W.; Vilgalys, R. The search for the fungal tree of life. *Trends Microbiol.* **2009**, *17*, 488–497.
16. Tian, M.; Win, J.; Song, J.; van der Hoorn, R.; van der Knaap, E.; Kamoun, S. A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease. *Plant Physiol.* **2007**, *143*, 364–77, doi:10.1104/pp.106.090050.
17. Gallegly, M. E. Genetics of pathogenicity of Phytophthora infestans. *Annu. Rev. Phytopathol.* **1968**, *6*, 375–396.
18. Montarry, J.; Andrivon, D.; Glais, I.; Corbiere, R.; Mialdea, G.; Delmotte, F. Microsatellite markers reveal two admixed genetic groups and an ongoing displacement within the French population of the invasive plant pathogen Phytophthora infestans. *Mol. Ecol.* **2010**, *19*, 1965–1977.
19. Agrios, G. N. Plant Pathology. 5th eds. *Dep. Plant Pathol. Univ. Florida. United States Am.* **2005**.
20. Kamoun, S. Groovy times: filamentous pathogen effectors revealed. *Curr. Opin. Plant Biol.* **2007**, *10*, 358–365.
21. van Poppel, P. M. J. A.; Guo, J.; de Vondervoort, P. J. I. V; Jung, M. W. M.; Birch, P. R. J.; Whisson, S. C.; Govers, F. The Phytophthora infestans Avirulence Gene Avr4 Encodes an RXLR- dEER Effector. *Mol. Plant-Microbe Interact.* **2008**, *21*, 1460–1470, doi:Doi 10.1094/Mpmi-21-11-1460.
22. Oh, S.-K. S.-K.; Young, C.; Lee, M.; Oliva, R.; Bozkurt, T. O.; Cano, L. M.; Win, J.;

Bos, J. I. B. B.; Liu, H.-Y. H.-Y.; van Damme, M.; Morgan, W.; Choi, D.; Van der Vossen, E. A. G.; Vleeshouwers, V. G. A. A.; Kamoun, S. In planta expression screens of Phytophthora infestans RXLR effectors reveal diverse phenotypes, including activation of the Solanum bulbocastanum disease resistance protein Rpi-blb2. *Plant Cell* **2009**, *21*, 2928–2947, doi:10.1105/tpc.109.068247.

23. Armstrong, M. R.; Whisson, S. C.; Pritchard, L.; Bos, J. I. B.; Venter, E.; Avrova, A. O.; Rehmany, A. P.; Böhme, U.; Brooks, K.; Cherevach, I. An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7766–7771.
24. Schornack, S.; Huitema, E.; Cano, L. M.; Bozkurt, T. O.; Oliva, R.; van Damme, M.; Schwizer, S.; Raffaele, S.; Chaparro-Garcia, A.; Farrer, R. Ten things to know about oomycete effectors. *Mol. Plant Pathol.* **2009**, *10*, 795–803.
25. Halterman, D. A.; Chen, Y.; Sopee, J.; Berduo-Sandoval, J.; Sánchez-Pérez, A. Competition between Phytophthora infestans effectors leads to increased aggressiveness on plants containing broad-spectrum late blight resistance. *PLoS One* **2010**, *5*, e10536, doi:10.1371/journal.pone.0010536.
26. Champouret, N.; Bouwmeester, K.; Rietman, H.; van der Lee, T.; Maliepaard, C.; Heupink, A.; van de Vondervoort, P. J. I.; Jacobsen, E.; Visser, R. G. F.; van der Vossen, E. A. G. Phytophthora infestans isolates lacking class I ipiO variants are virulent on Rpi-blb1 potato. *Mol. plant-microbe Interact.* **2009**, *22*, 1535–1545.
27. Bos, J. I. B.; Armstrong, M. R.; Gilroy, E. M.; Boevink, P. C.; Hein, I.; Taylor, R. M.; Zhendong, T.; Engelhardt, S.; Vetukuri, R. R.; Harrower, B. Phytophthora infestans effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proc. Natl. Acad. Sci.* **2010**, *107*, 9909–9914.
28. Jiang, R. H. Y.; Tripathy, S.; Govers, F.; Tyler, B. M. RXLR effector reservoir in two Phytophthora species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proc. Natl. Acad. Sci.* **2008**, *105*, 4874–4879.
29. Eddy, S. S. R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763, doi:btb114.
30. Andrivon, D. The origin of Phytophthora infestans populations present in Europe in the 1840s: a critical review of historical and scientific evidence. *Plant Pathol.* **1996**, *45*,

1027–1035.

31. Berkeley, M. J.; Broome, C. E. IX. Notices of British hypogœous Fungi. *J. Nat. Hist.* **1846**, *18*, 73–82.
32. Goss, E. M.; Tabima, J. F.; Cooke, D. E. L.; Restrepo, S.; Fry, W. E.; Forbes, G. A.; Fieland, V. J.; Cardenas, M.; Grünwald, N. J. The Irish potato famine pathogen Phytophthora infestans originated in central Mexico rather than the Andes. *Proc. Natl. Acad. Sci.* **2014**, *111*, 8791–8796.
33. Grünwald, N. J.; Flier, W. G. The Biology of *Phytophthora infestans* at Its Center of Origin. *Annu. Rev. Phytopathol.* **2005**, *43*, 171–190, doi:10.1146/annurev.phyto.43.040204.135906.
34. Goodwin, S. B.; Legard, D. E.; Smart, C. D.; Levy, M.; Fry, W. E. Gene flow analysis of molecular markers confirms that Phytophthora mirabilis and P. infestans are separate species. *Mycologia* **1999**, 796–810.
35. Flier, W. G.; Grünwald, N. J.; Kroon, L. P. N. M.; Van Den Bosch, T. B. M.; Garay-Serrano, E.; Lozoya-Saldaña, H.; Bonants, P. J. M.; Turkensteen, L. J. Phytophthora ipomoeae sp. nov., a new homothallic species causing leaf blight on Ipomoea longipedunculata in the Toluca Valley of central Mexico. *Mycol. Res.* **2002**, *106*, 848–856.
36. Niederhauser, J. S. *Phytophthora infestans: the Mexican connection*; Cambridge University Press, Cambridge, 1991;
37. Dyer, A.; Matusralt, M.; Drenth, A.; Cohen, A.; Splelman, L. Historical and recent migrations of Phytophthora infestans: chronology, pathways, and implications. *Plant Dis.* **1993**, *77*, 653–661.
38. Hawkes, J. G. *The potato: evolution, biodiversity and genetic resources.*; Belhaven Press, **1990**; ISBN 1852930454.
39. Junchaya La Rosa de Abad, G. *Phytopthora infestans en la zona central del Perú: rango de hospedantes variabilidad y resistencia varietal*; Universidad Nacional Agraria La Molina, Lima (Peru). Programa Académico de Graduados. Especialidad en Fitopatología, **1983**;
40. Gómez-Alpizar, L.; Carbone, I.; Ristaino, J. B. An Andean origin of Phytophthora infestans inferred from mitochondrial and nuclear gene genealogies. *Proc. Natl. Acad.*

Sci. **2007**, *104*, 3306–3311.

41. Goodwin, S. B.; Cohen, B. A.; Fry, W. E. Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 11591–5, doi:10.1073/pnas.91.24.11591.
42. May, K. J.; Ristaino, J. B. Identity of the mtDNA haplotype (s) of Phytophthora infestans in historical specimens from the Irish potato famine. *Mycol. Res.* **2004**, *108*, 471–479.
43. Gavino, P. D.; Fry, W. E. Diversity in and evidence for selection on the mitochondrial genome of Phytophthora infestans. *Mycologia* **2002**, *94*, 781–793.
44. Flier, W. G.; Grunwald, N. J.; Kroon, L. P.; Sturbaum, A. K.; van den Bosch, T. B.; Garay-Serrano, E.; Lozoya-Saldana, H.; Fry, W. E.; Turkensteen, L. J. The Population Structure of Phytophthora infestans from the Toluca Valley of Central Mexico Suggests Genetic Differentiation Between Populations from Cultivated Potato and Wild Solanum spp. *Phytopathology* **2003**, *93*, 382–390, doi:10.1094/PHYTO.2003.93.4.382.
45. Perez, W. G.; Gamboa, J. S.; Falcon, Y. V; Coca, M.; Raymundo, R. M.; Nelson, R. J. Genetic structure of Peruvian populations of Phytophthora infestans. *Phytopathology* **2001**, *91*, 956–965.
46. Martin, M. D.; Cappellini, E.; Samaniego, J. A.; Zepeda, M. L.; Campos, P. F.; Seguin-Orlando, A.; Wales, N.; Orlando, L.; Ho, S. Y. W.; Dietrich, F. S.; Mieczkowski, P. A.; Heitman, J.; Willerslev, E.; Krogh, A.; Ristaino, J. B.; Gilbert, M. T. P. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nat. Commun.* **2013**, *4*, 1–7, doi:10.1038/ncomms3172.
47. Yoshida, K.; Schuenemann, V. J.; Cano, L. M.; Pais, M.; Mishra, B.; Sharma, R.; Lanz, C.; Martin, F. N.; Kamoun, S.; Krause, J.; Thines, M.; Weigel, D.; Burbano, H. A. The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *Elife* **2013**, *2013*, 1–25, doi:10.7554/eLife.00731.
48. Gavino, P. D.; Smart, C. D.; Sandrock, R. W.; Miller, J. S.; Hamm, P. B.; Lee, T. Y.; Davis, R. M.; Fry, W. E. Implications of sexual reproduction for Phytophthora infestans in the United States: generation of an aggressive lineage. *Plant Dis.* **2000**, *84*, 731–735.
49. Cooke, D. E. L.; Cano, L. M.; Raffaele, S.; Bain, R. A.; Cooke, L. R.; Etherington, G.

J.; Deahl, K. L.; Farrer, R. A.; Gilroy, E. M.; Goss, E. M.; Grünwald, N. J.; Hein, I.; MacLean, D.; McNicol, J. W.; Randall, E.; Oliva, R. F.; Pel, M. A.; Shaw, D. S.; Squires, J. N.; Taylor, M. C.; Vleeshouwers, V. G. A. A.; Birch, P. R. J.; Lees, A. K.; Kamoun, S. Genome Analyses of an Aggressive and Invasive Lineage of the Irish Potato Famine Pathogen. *PLoS Pathog.* **2012**, *8*, doi:10.1371/journal.ppat.1002940.

50. Chowdappa, P.; Kumar, N. B. J.; Madhura, S.; Kumar, M. S. P.; Myers, K. L.; Fry, W. E.; Squires, J. N.; Cooke, D. E. L. Emergence of 13_ A 2 Blue Lineage of Phytophthora infestans was Responsible for Severe Outbreaks of Late Blight on Tomato in South-West India. *J. Phytopathol.* **2013**, *161*, 49–58.
51. Li, Y.; Van der Lee, T.; Zhu, J. H.; Jin, G. H.; Lan, C. Z.; Zhu, S. X.; Zhang, R. F.; Liu, B. W.; Zhao, Z. J.; Kessel, G. Population structure of P hytophthora infestans in China–geographic clusters and presence of the EU genotype Blue_13. *Plant Pathol.* **2013**, *62*, 932–942.
52. Cooke, D. E. L.; Cano, L. M.; Raffaele, S.; Bain, R. A.; Cooke, L. R.; Etherington, G. J.; Deahl, K. L.; Farrer, R. A.; Gilroy, E. M.; Goss, E. M. Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. *PLoS Pathog.* **2012**, *8*, e1002940.
53. Raffaele, S.; Farrer, R. A.; Cano, L. M.; Studholme, D. J.; MacLean, D.; Thines, M.; Jiang, R. H. Y. Y.; Zody, M. C.; Kunjeti, S. G.; Donofrio, N. M.; Meyers, B. C.; Nusbaum, C.; Kamoun, S. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science (80-.).* **2010**, *330*, 1540–1543, doi:10.1126/science.1193070.
54. Fry, W. E.; Goodwin, S. B. Re-emergence of potato and tomato late blight in the United States. *Plant Dis.* **1997**, *81*, 1349–1357.
55. Fry, W. E.; Goodwin, S. B.; Dyer, A. T.; Matuszak, J. M.; Drenth, A.; Tooley, P. W.; Sujkowski, L. S. YJ Koh, BA Cohen, LJ Spielman, KL Deahl, DA Inglis & KP Sandlan. 1993. Historical and recent migrations of Phytophthora infestans: chronology, pathways, and implications. *Plant Dis.* **1993**, *77*, 653–661.
56. Tooley, P. W.; Therrien, C. D.; Ritch, D. L. Mating type, race composition, nuclear DNA content, and isozyme analysis of Peruvian isolates of Phytophthora infestans. *Phytopathology* **1989**, *79*, 478–481.

57. Grünwald, N. J.; Flier, W. G. The biology of Phytophthora infestans at its center of origin. *Annu. Rev. Phytopathol.* **2005**, *43*, 171–190.
58. Grünwald, N. J.; Flier, W. G.; Sturbaum, A. K.; Garay-Serrano, E.; van den Bosch, T. B. M.; Smart, C. D.; Matuszak, J. M.; Lozoya-Saldaña, H.; Turkensteen, L. J.; Fry, W. E. Population Structure of *Phytophthora infestans* in the Toluca Valley Region of Central Mexico. *Phytopathology* **2001**, *91*, 882–890, doi:10.1094/PHYTO.2001.91.9.882.
59. Goodwin, S. B. Population Genetics of Soilborne Fungal Plant Pathogens: The Population Genetics of *Phytophthora*. *Phytopathology* **1997**, *87*, 462–473, doi:10.1094/PHYTO.1997.87.4.448.
60. Shakya, S. K.; Larsen, M. M.; Cuenca-Condoy, M. M.; Lozoya-Saldaña, H.; Grünwald, N. J.; Grünwald, N. J. Variation in Genetic Diversity of *Phytophthora infestans* Populations in Mexico from the Center of Origin Outwards. *Plant Dis.* **2018**, *102*, 1534–1540, doi:10.1094/PDIS-11-17-1801-RE.
61. Wang, J.; Fernández-Pavía, S. P.; Larsen, M. M.; Garay-Serrano, E.; Gregorio-Cipriano, R.; Rodríguez-Alvarado, G.; Grünwald, N. J.; Goss, E. M. High levels of diversity and population structure in the potato late blight pathogen at the Mexico centre of origin. *Mol. Ecol.* **2017**, *26*, 1091–1107, doi:10.1111/mec.14000.
62. Lassiter, E. S.; Russ, C.; Nusbaum, C.; Zeng, Q.; Saville, A. C.; Olarte, R. A.; Carbone, I.; Hu, C.-H.; Seguin-Orlando, A.; Samaniego, J. A. Mitochondrial genome sequences reveal evolutionary relationships of the Phytophthora 1c clade species. *Curr. Genet.* **2015**, *61*, 567–577.
63. Gupta, A.; Bhardwaj, A.; Sharma, P.; Pal, Y. Mitochondrial DNA-a tool for phylogenetic and biodiversity search in equines. *J. Biodivers. Endanger. Species* **2015**, *2015*.
64. Hartl, D. L.; Clark, A. G. Principles of population genetics. 2007. *Sunderland, Massachusetts Fourth Ed. Sinauer Assoc. Google Sch.*
65. Felsenstein, J. *Inferring phylogenies*; Sinauer associates Sunderland, MA, 2004; Vol. 2;.
66. Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.

67. Farris, J. S. Methods for computing Wagner trees. *Syst. Biol.* **1970**, *19*, 83–92.
68. Chor, B.; Tuller, T. Maximum likelihood of evolutionary trees: hardness and approximation. *Bioinformatics* **2005**, *21*, i97–i106.
69. Holmes, S. Statistics for phylogenetic trees. *Theor. Popul. Biol.* **2003**, *63*, 17–32.
70. Bouckaert, R.; Heled, J.; Kühnert, D.; Vaughan, T.; Wu, C. H.; Xie, D.; Suchard, M. A.; Rambaut, A.; Drummond, A. J. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **2014**, *10*, 1–6, doi:10.1371/journal.pcbi.1003537.
71. Head, S. R.; Komori, H. K.; LaMere, S. A.; Whisenant, T.; Van Nieuwerburgh, F.; Salomon, D. R.; Ordoukhanian, P. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques* **2014**, *56*, 61.
72. Carøe, C.; Gopalakrishnan, S.; Vinner, L.; Mak, S. S. T. T.; Sinding, M. H. S.; Samaniego, J. A.; Wales, N.; Sicheritz-Pontén, T.; Gilbert, M. T. P.; Sicheritz-Pontén, T.; Gilbert, M. T. P. Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **2018**, *9*, 410–419, doi:10.1111/2041-210X.12871.
73. Schubert, M.; Jónsson, H.; Chang, D.; Der Sarkissian, C.; Ermini, L.; Ginolhac, A.; Albrechtsen, A.; Dupanloup, I.; Foucal, A.; Petersen, B. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci.* **2014**, *111*, E5661–E5669.
74. Schubert, M.; Lindgreen, S.; Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **2016**, *9*, 88.
75. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760, doi:10.1093/bioinformatics/btp324.
76. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**.
77. Jónsson, H.; Ginolhac, A.; Schubert, M.; Johnson, P. L. F. F.; Orlando, L. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **2013**, *29*, 1682–1684, doi:10.1093/bioinformatics/btt193.
78. Jackman, S. D.; Vandervalk, B. P.; Mohamadi, H.; Chu, J.; Yeo, S.; Hammond, S. A.;

Jahesh, G.; Khan, H.; Coombe, L.; Warren, R. L. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* **2017**, gr-214346.

79. Van Dongen, S. M. Graph clustering by flow simulation 2000.
80. Enright, A. J.; Van Dongen, S.; Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **2002**, *30*, 1575–1584.
81. Birch, Paul RJ and Cooke, D. EL The early days of late blight Large-scale. *Elife* **2013**, *2013*, 3–5, doi:10.7554/eLife.00731.
82. Goodwin, S. B.; Spielman, L. J.; Matuszak, J. M.; Bergeron, S. N.; Fry, W. E. Clonal diversity and genetic differentiation of Phytophthora infestans populations in northern and central Mexico. *Phytopathol.* **1992**.
83. Tooley, P. W.; Fry, W. E.; Gonzalez, M. J. V. Isozyme characterization of sexual and asexual Phytophthora infestans populations. *J. Hered.* **1985**, *76*, 431–435.
84. Carter, D. A.; Archer, S. A.; Buck, K. W.; Shaw, D. S.; Shattock, R. C. Restriction fragment length polymorphisms of mitochondrial DNA of Phytophthora infestans. *Mycol. Res.* **1990**, *94*, 1123–1128.
85. Griffith, G. W.; Shaw, D. S. Polymorphisms in Phytophthora infestans: four mitochondrial haplotypes are detected after PCR amplification of DNA from pure cultures or from host lesions. *Appl. Environ. Microbiol.* **1998**, *64*, 4007–4014.
86. Stevens, N. E. The dark ages in Plant Pathology in America: 1830—1870. *J. Washingt. Acad. Sci.* **1933**, *23*, 435–446.
87. Bourke, P. M. A. Emergence of potato blight, 1843–46. *Nature* **1964**, *203*, 805.
88. Ho, S. Y. W.; Lanfear, R.; Phillips, M. J.; Barnes, I.; Thomas, J. A.; Kolokotronis, S. O.; Shapiro, B. Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Syst. Biol.* **2011**, *60*, 366–375, doi:10.1093/sysbio/syq099.
89. Gibbs, A. J.; Fargette, D.; García-Arenal, F.; Gibbs, M. J. Time - The emerging dimension of plant virus studies. *J. Gen. Virol.* **2010**, *91*, 13–22, doi:10.1099/vir.0.015925-0.
90. Howell, N.; Smejkal, C. B.; Mackey, D. A.; Chinnery, P. F.; Turnbull, D. M.; Herrnstadt, C. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am. J. Hum.*

Genet. **2003**, *72*, 659–670.

91. Yang, Z. A space-time process model for the evolution of DNA sequences. *Genetics* **1995**, *139*, 993–1005.
92. Springer, M. S.; Amrine, H. M.; Burk, A.; Stanhope, M. J. Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with heterogeneous base composition. *Syst. Biol.* **1999**, *48*, 65–75.
93. Shapiro, B.; Rambaut, A.; Drummond, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **2005**, *23*, 7–9.
94. Maddison, W. P. Gene trees in species trees. *Syst. Biol.* **1997**, *46*, 523–536.
95. Nichols, R. Gene trees and species trees are not the same. *Trends Ecol. Evol.* **2001**, *16*, 358–364.
96. Avila-Adame, C.; Gómez-Alpizar, L.; Zismann, V.; Jones, K. M.; Buell, C. R.; Ristaino, J. B. Mitochondrial genome sequences and molecular evolution of the Irish potato famine pathogen, Phytophthora infestans. *Curr. Genet.* **2006**, *49*, 39–46, doi:10.1007/s00294-005-0016-3.
97. Win, J.; Kamoun, S. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Signal. Behav.* **2008**, *3*, 251–253, doi:10.4161/psb.3.4.5182.
98. Bos, J. I. B.; Kanneganti, T.; Young, C.; Cakir, C.; Huitema, E.; Win, J.; Armstrong, M. R.; Birch, P. R. J.; Kamoun, S. The C-terminal half of Phytophthora infestans RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in Nicotiana benthamiana. *Plant J.* **2006**, *48*, 165–176.
99. Ye, W.; Wang, Y.; Wang, Y. Bioinformatics analysis reveals abundant short alpha-helices as a common structural feature of oomycete RxLR effector proteins. *PLoS One* **2015**, *10*, 1–13, doi:10.1371/journal.pone.0135240.
100. dos Reis, M.; Inoue, J.; Hasegawa, M.; Asher, R. J.; Donoghue, P. C. J.; Yang, Z. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. London B Biol. Sci.* **2012**, *279*, 3491–

3500.

101. Meusemann, K.; von Reumont, B. M.; Simon, S.; Roeding, F.; Strauss, S.; Kück, P.; Ebersberger, I.; Walzl, M.; Pass, G.; Breuers, S. A phylogenomic approach to resolve the arthropod tree of life. *Mol. Biol. Evol.* **2010**, *27*, 2451–2464.
102. Huelsenbeck, J. P.; Bull, J. J.; Cunningham, C. W. Combining data in phylogenetic analysis. *Trends Ecol. Evol.* **1996**, *11*, 152–158.
103. Boutemy, L. S.; King, S. R. F.; Win, J.; Hughes, R. K.; Clarke, T. A.; Blumenschein, T. M. A.; Kamoun, S.; Banfield, M. J. Structures of Phytophthora RXLR effector proteins: A conserved but adaptable fold underpins functional diversity. *J. Biol. Chem.* **2011**, *286*, 35834–35842, doi:10.1074/jbc.M111.262303.

Appendix

Table A. 1 Provenance and necessary information regarding *Phytophthora* spp. samples used in this study. Pi, *Phytophthora infestans*. Pa, *Phytophthora andina*. Pp, *Phytophthora ipomoeae*. Pm *P. mirabilis*. ENA, European Nucleotide Archive. M2013, Martin et al. (2013). Y2013, Yoshida et al. (2013). R2010, Raffael et al. (2010). C2012, Cooke et al. (2012). WPC, World *Phytophthora* collection. H, Host herbarium specimen. M, Mycleium.

Sample ID	Alternate IDs	Species	Host	Location	Collection date	Tissue source	RFLP genotype and/or mtDNA lineage	Sequence reads source
Pi1845A	K91	Pi	*S. tuberosum*	Audenarde, Belgium	1845	H	HERB-1	M2013
KM177500		Pi	*S. tuberosum*	England	1845	H	HERB-1	Y2013
KM177513		Pi	*S. tuberosum*	Ireland	1846	H	HERB-1	Y2013
KM177514		Pi	*S. tuberosum*	Ireland	1847	H	HERB-1	Y2013
KM177548		Pi	*S. tuberosum*	England	1847	H	HERB-1	Y2013
KM177509		Pi	*S. tuberosum*	England	1865	H	HERB-1	Y2013
M-0182900		Pi	*S. lycopersicum*	Germany	1873	H	HERB-1	Y2013
KM177517		Pi	*S. tuberosum*	Wales	1875	H	HERB-1	Y2013
M-0182907		Pi	*S. tuberosum*	Germany	1875	H	HERB-1	Y2013
Pi1876	UPS 1; Up1	Pi	*S. tuberosum*	Skårop, Denmark	1876	H	HERB-1	M2013
M-0182896		Pi	*S. tuberosum*	Germany	1877	H	HERB-1	Y2013
M-0182906		Pi	*S. tuberosum*	Germany	1877	H	HERB-1	Y2013
Pi1882	UPS 2; Up2	Pi	*S. tuberosum*	Stockholm, Sweden	1882	H	HERB-1	M2013
Pi1889	K 79	Pi	*S. tuberosum*	Germany	1889	H	HERB-1	M2013
M-0182903		Pi	*S. tuberosum*	Canada	1896	H	HERB-1	Y2013
Kew126		Pi	*S. tuberosum*	Britain	1952	H	US-1/1b	This report
Kew122		Pi	*S. tuberosum*	Britain	1955	H	US-1/1b	This report
Kew123		Pi	*S. tuberosum*	Ireland	1955	H	US-1/1b	This report
M-0182898		Pi	*S. tuberosum*	Germany	before 1863	H	HERB-1	Y2013
KM177512		Pi	*S. tuberosum*	England	Unknown	H	HERB-1	Y2013
Pi1845B	K16	Pi	*S. tuberosum*	Great Britain	1845	H	HERB-1	M2013
DDR7602		Pi	*S. tuberosum*	Germany	1976	M	US-1/1b	Y2013
P1362		Pi	*S. tuberosum*	Mexico	1979	M	US-1/Ib	Y2013
P8844	Hohl, Pineda 151	Pi	*S. tuberosum*	Peru	1982	M	US-1/Ib	Martin et al 2016

P3681	Tooley 529	Pi	*S. tuberosum*	Mexico	1983	M	Ia	Martin et al 2016
P3683	Tooley 550	Pi	*S. stoloniferum*	Mexico	1983	M	Ia	Martin et al 2016
P3685	Tooley 533	Pi	*S. tuberosum*	Mexico	1983	M	Ia	Martin et al 2016
P6629	Tooley 511	Pi	*S. tuberosum*	Mexico	1983	M	Ia	Martin et al 2016
P6634	PD_02383; Speil543	Pi	*S. tuberosum*	Mexico	1983	M	Ia	Martin et al 2016
P6635	PD_02388; Fry 580	Pi	*S. demissum*	Central Mexico	1986	M	Ia	Martin et al 2016
P7036	Goodwin 575	Pi	*S. tuberosum*	Toluca, Mexico	1986	M	HERB-1	Martin et al 2016
P8140	PD_01946; P15103; Fry 572	Pi	*S. tuberosum*	Toluca, Mexico	1986	M	HERB-1	Martin et al 2016
P8141	Fry 616	Pi	*S. tuberosum*	Toluca, Mexico	1987	M	US-1/1b	Martin et al 2016
P8143	PD_01949; P15154; Fry 619	Pi	*S. tuberosum*	Toluca, Mexico	1987	M	Ia	Martin et al 2016
P8144	Fry 622	Pi	*S. tuberosum*	Toluca, Mexico	1987	M	Ia	Martin et al 2016
T30-4	Broad_GA2	Pi	*S. tuberosum*	Netherlands	1988	M	Ia	R2010
P6515	CIP 27 ; PD_00883; Fry 543	Pi	*S. tuberosum*	Peru	1989	M	US-1/Ib	Martin et al 2016
P6570	Davidse 89018; P570	Pi	*S. tuberosum*	Netherlands	1989	M	IIa	Martin et al 2016
90128		Pi	*S. tuberosum*	Netherlands	1990	M	Ia	R2010
PHU006		Pi	*S. tuberosum*	Peru	1996	M	*EC-1*, IIa	Martin et al 2016
P12204		Pi	*S. tuberosum*	Scotland	1996	M	Ia	Y2013
PCO038	PCO028	Pi	*S. tuberosum*	Peru	1997	M	*EC-1*, IIa	Martin et al 2016
PCZ026		Pi	*S. tuberosum*	Cusco, Peru	1997	M	*PE-6*, IIa	Martin et al 2016
PCZ033		Pi	*S. tuberosum*	Cusco, Peru	1997	M	*EC1.2*, IIa	Martin et al 2016
PCZ050		Pi	*S. tuberosum*	Peru	1997	M	*PE-3*, Ia	Martin et al 2016
PCZ098		Pi	*S. tuberosum*	Cusco, Peru	1997	M	*EC1.3*, IIa	Martin et al 2016
PIC97207		Pi	*S. tuberosum*	Mexico	1997	M	Ia	Martin et al 2016
PIC97605		Pi	*S. tuberosum*	Mexico	1997	M	Ia	Martin et al 2016
PIC97630		Pi	*S. tuberosum*	Mexico	1997	M	Ia	Martin et al 2016
P10650	MX980099/ P10650_MDM	Pi	*S. tuberosum*	Toluca, Mexico	1998	M	Ia	Martin et al 2016; Y2013
P13198	CIP3198 tuq	Pi	*S. tuquerrense*	Napo, Ecuador	1998	M	*EC-1*, IIa	Martin et al 2016
PIC99189		Pi	*S. stoloniferum*	Mexico	1999	M	Ia	R2010
P13346	CIP3346 col	Pi	*S. colombianum*	Napo, Ecuador	2001	M	HERB-1	Martin et al 2016
P13527		Pi	*S. andreanum*	Ecuador	2002	M	IIa	Y2013

P10127		Pi	*S. lycopersicum*	USA	2002	M	IIb	Y2013
P13626	CIP3626 tbr	Pi	*S. tuberosum*	Ecuador	2003	M	IIa	Martin et al 2016
LBUS5		Pi	*Petunia hybrida*	South Africa	2005	M	US-1/Ib	Y2013
P3873	CIP3873 tbr	Pi	*S. tuberosum*	Cañar, Ecuador	2005	M	IIa	Martin et al 2016
06_3928A	06_3928	Pi	*S. tuberosum*	England	2006	M	*13_A2*, Ia	C2012
NL07434		Pi	*S. tuberosum*	Netherlands	2007	M	IIa	Y2013
BL2009P4_23	PA112	Pi	*S. tuberosum*	Pennsylvania, USA	2009	M	*US-23*, Ia	M2013
IN2009T1	PA114	Pi	*S. tuberosum*	Pennsylvania, USA	2009	M	*US-22*, Ia	M2013
P17777		Pi	*S. lycopersicum*	USA	2009	M	*US-22*, Ia	Y2013
RS2009P1	PA117	Pi	*S. tuberosum*	Pennsylvania, USA	2009	M	*US-8,* Ia	M2013
P6636	PD_01096; Spelman 618	Pi	*S. tuberosum*	Toluca, Mexico	1987	M	Ia	Martin et al 2016
PIC98372		Pi	*S. demissum*	Mexico	1998	M	Ia	Martin et al 2016
P13803	CIP 3803	Pa	*S. betaceum*	Pichincha, Ecuador	2004	M	*EC-3*, HERB-1	Martin et al 2016
P6096	CIP 17; Tooley 801	Pi	*S. tuberosum*	Paurcartambo, Peru	1984	M	US-1/Ib	Y2013
Pinf_C_13_7_28		Pi	*Solanum* spp.	Denmark	2013	M	IIa	
EC3394	P_andEC3394	Pa	*S. betaceum*	Tungurahua, Ecuador	2001	M	*EC-3*, HERB-1	Martin et al 2016
EC3425	P_andEC3425	Pm	*S. brevifolium*	Ecuador	1999	M		R2010; Y2013
CHC19		Pi	*S. tuberosum*	Chapingo, Mexico	2015		Ia	study Sample
CHC78		Pi	*S. tuberosum*	Chapingo, Mexico	2016		HERB-1	study Sample
CHCL20		Pi	*S. tuberosum*	Chapingo, Mexico	2015		Ia	study Sample
CHG10		Pi	*S. tuberosum*	Chapingo, Mexico	2015		HERB-1	study Sample
CHG12		Pi	*S. tuberosum*	Chapingo, Mexico	2015		HERB-1	study Sample
CHG16		Pi	*S. tuberosum*	Chapingo, Mexico	2015		HERB-1	study Sample
CHG21		Pi	*S. tuberosum*	Chapingo, Mexico	2015		III	study Sample

CHG3		Pi	*S. tuberosum*	Chapingo, Mexico	2015		III	study Sample
CHG32		Pi	*S. tuberosum*	Chapingo, Mexico	2015		HERB-1	study Sample
CHG6		Pi	*S. tuberosum*	Chapingo, Mexico	2015		III	study Sample
CH02		Pi	*S. tuberosum*	Chapingo, Mexico	2016		HERB-1	study Sample
CH04		Pi	*S. tuberosum*	Chapingo, Mexico	2016		HERB-1	study Sample
JFH100		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		IIa	study Sample
JFH104		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		HERB-1	study Sample
JFH108		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		HERB-1	study Sample
JFH109		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		HERB-1	study Sample
JFH121		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		HERB-1	study Sample
JFH122		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		HERB-1	study Sample
JFH135		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		HERB-1	study Sample
JFH14		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		Ia	study Sample
JFH153		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		Ia	study Sample
JFH163		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		HERB-1	study Sample
JFH168		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		HERB-1	study Sample
JFH3		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		HERB-1	study Sample
JFH32		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		HERB-1	study Sample
JFH41		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		HERB-1	study Sample
JFH43		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		IIa	study Sample
JFH48		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		IIa	study Sample
JFH49		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		HERB-1	study Sample

JFH51		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		Ia	study Sample
JFH54		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		HERB-1	study Sample
JFH59		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		HERB-1	study Sample
JFH63		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		Ia	study Sample
JFH7		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		Ia	study Sample
JFH8		Pi	*S. tuberosum*	Juchitepec, Mexico	2015		HERB-1	study Sample
JFH93		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		HERB-1	study Sample
JFH99		Pi	*S. tuberosum*	Juchitepec, Mexico	2016		Ia	study Sample
SGF(X)		Pi	*S. tuberosum*	San_Geronimo	2015		IIa	study Sample
SGF(Y)		Pi	*S. tuberosum*	San_Geronimo	2015		Ia	study Sample
SGF1		Pi	*S. tuberosum*	San_Geronimo	2015		HERB-1	study Sample
SGF18		Pi	*S. tuberosum*	San_Geronimo	2015		HERB-1	study Sample
SGF21		Pi	*S. tuberosum*	San_Geronimo	2015		HERB-1	study Sample
SGF27		Pi	*S. tuberosum*	San_Geronimo	2015		Ia	study Sample
SGF37		Pi	*S. tuberosum*	San_Geronimo	2015		HERB-1	study Sample
SGF39		Pi	*S. tuberosum*	San_Geronimo	2016		HERB-1	study Sample
SGF40		Pi	*S. tuberosum*	San_Geronimo	2015		Ia	study Sample
SGF48		Pi	*S. tuberosum*	San_Geronimo	2016		Ia	study Sample
SGF57		Pi	*S. tuberosum*	San_Geronimo	2016		HERB-1	study Sample
SGF60		Pi	*S. tuberosum*	San_Geronimo	2015		Ia	study Sample
SGF65		Pi	*S. tuberosum*	San_Geronimo	2015		III	study Sample
SGF66		Pi	*S. tuberosum*	San_Geronimo	2015		HERB-1	study Sample
SGF67		Pi	*S. tuberosum*	San_Geronimo	2015		Ia	study Sample
SGF78		Pi	*S. tuberosum*	San_Geronimo	2015		HERB-1	study Sample
SGF79		Pi	*S. tuberosum*	San_Geronimo	2015		HERB-1	study Sample
SGF83		Pi	*S. tuberosum*	San_Geronimo	2015		Ia	study Sample
SGF84		Pi	*S. tuberosum*	San_Geronimo	2015		HERB-1	study Sample
T11		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T12		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample

T13		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T17		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T18		Pi	*S. tuberosum*	Toluca	2015		HERB-1	study Sample
T20		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T21		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T23		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T24		Pi	*S. tuberosum*	Toluca	2015		HERB-1	study Sample
T25		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T26		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T3		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T7		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
T8		Pi	*S. tuberosum*	Toluca	2015		Ia	study Sample
TF1		Pi	*S. tuberosum*	Tlaxcala	2015		Ia	study Sample
TF2		Pi	*S. tuberosum*	Tlaxcala	2015		III	study Sample
TF4		Pi	*S. tuberosum*	Tlaxcala	2015		HERB-1	study Sample
TG10		Pi	*S. tuberosum*	Tlaxcala	2015		HERB-1	study Sample
TG11		Pi	*S. tuberosum*	Tlaxcala	2015		Ia	study Sample
TG15		Pi	*S. tuberosum*	Tlaxcala	2015		Ia	study Sample
TG20		Pi	*S. tuberosum*	Tlaxcala	2015		Ia	study Sample
TL1		Pi	*S. tuberosum*	Tlaxcala	2015		Ia	study Sample
TL10		Pi	*S. tuberosum*	Tlaxcala	2015		HERB-1	study Sample
TL13		Pi	*S. tuberosum*	Tlaxcala	2015		HERB-1	study Sample
TL14		Pi	*S. tuberosum*	Tlaxcala	2015		IIa	study Sample
TL3		Pi	*S. tuberosum*	Tlaxcala	2015		HERB-1	study Sample
TL4		Pi	*S. tuberosum*	Tlaxcala	2015		Ia	study Sample
TL6		Pi	*S. tuberosum*	Tlaxcala	2015		Ia	study Sample
TT22		Pi	*S. tuberosum*	Tlaxcala	2015		Ia	study Sample
J107		Pi	*Solanum lycopersicum*	Tlaxcala	2001		III	Wang et al 2017
J204-1	J204	Pi	*Solanum lycopersicum*	Tlaxcala	2001		III	Wang et al 2017
J604-2	J604	Pi	*Solanum lycopersicum*	Tlaxcala	2001		III	Wang et al 2017
J707		Pi	*Solanum lycopersicum*	Tlaxcala	2001		III	Wang et al 2017

Mich7002-2		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2017
Mich7006		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2017
Mich7007		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2017
Mich7012		Pi	*S. tuberosum*	Michoacan	2007		HERB-1	Wang et al 2017
Mich7014		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2017
Mich7019		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2017
Mich7029		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2017
Mich7035		Pi	*S. tuberosum*	Michoacan	2007		IIa	Wang et al 2007
Mich7038		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2007
Mich7041		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2007
Mich7045		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2007
Mich7054		Pi	*S. tuberosum*	Michoacan	2007		HERB-1	Wang et al 2007
Mich7055		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2007
Mich7057		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2007
Mich7058		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2007
Mich7059		Pi	*S. tuberosum*	Michoacan	2007		III	Wang et al 2007
Mich7030-2		Pi	*S. tuberosum*	Michoacan	2007		HERB-1	Wang et al 2007
P13873	P3873	Pi	*S. tuberosum*	Canar, Equador	2005		IIa	
P11633			*Solanum lycopersicum*	Hungary	2005		IIa	Y 2013
P88069				Netherlands	1998		HERB-1	
P90128	90128		*Solanum spp.*	Netherlands	1990		Ia	
Pinf_B_13_8_35		Pi	*S. tuberosum*	Denmark	2013	M	US-1/Ib	
Pinf_D_13_3_5		Pi	*S. tuberosum*	Denmark	2013	M	Ia	
Pinf_F_13_7_33		Pi	*S. tuberosum*	Denmark	2013	M	US-1/Ib	
H292_1929					1929		HERB-1	
186968_1913a			*S. tuberosum*	Columbia	1913		HERB-1	
Pinf_A_13_5_10		Pi	*S. tuberosum*	Denmark	2013	M	HERB-1	
Mich7033		Pi	S. tuberosum	Mexico	2007		III	
Pand_Pax	Pax	Pa	Solanum spp.		2009	M		Martin et al. 2016
Pinf_E_13_8_102		Pi		Denmark	2013	M	IIa	

Pinf_C_13_7_28		Pi		Denmark	2013	M	IIa	

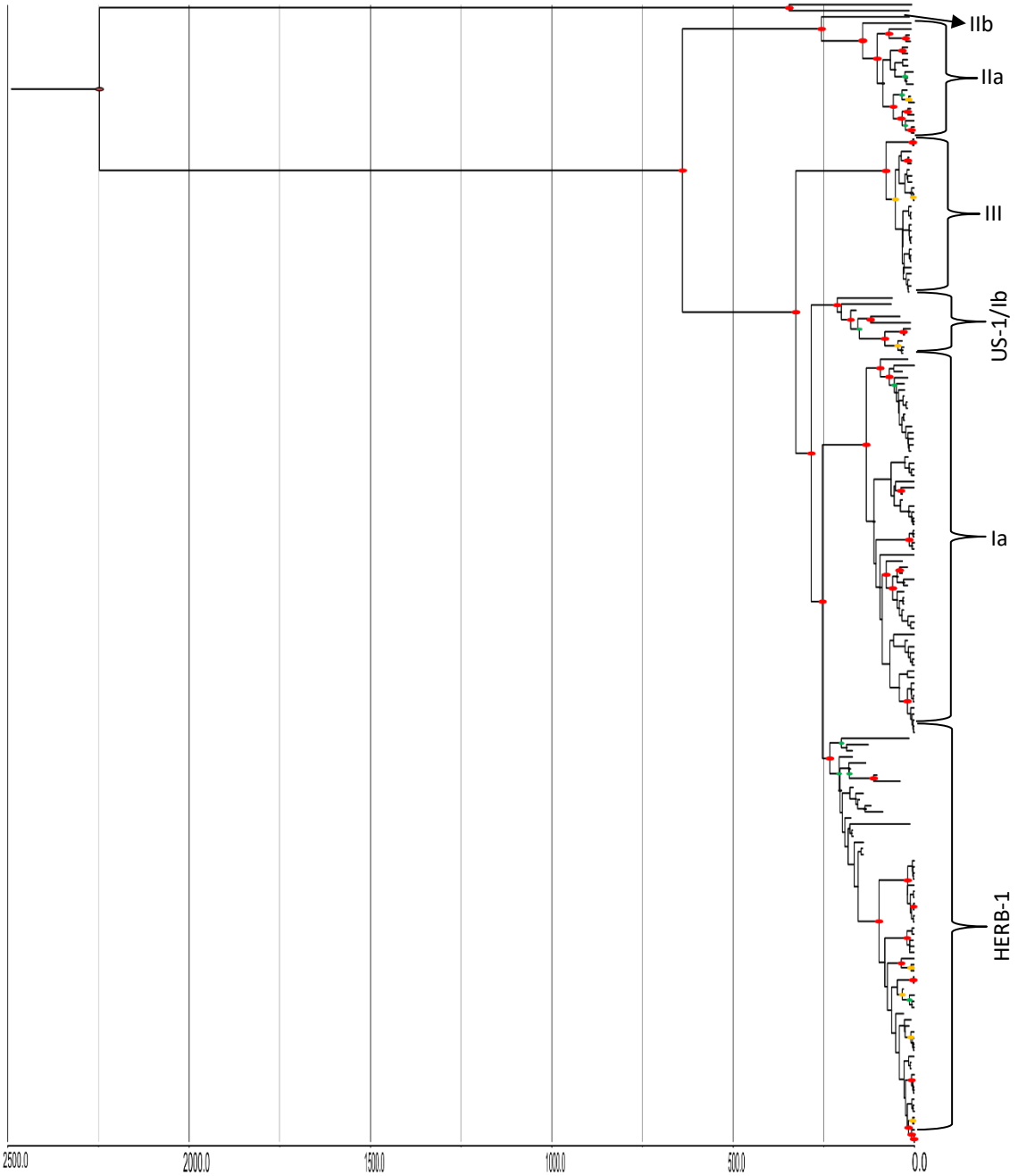

Fig A. 1 Mitochondrial DNA phylogeny of *P. infestans* isolates from Mexico and other previously studied samples using Bayesian dating analysis. The major mtDNA lineages are indicated by the brackets on the right side of the tree. The tree also includes two outgroups, it shows the real age of the tree with inclusion of the outgroups. The posterior probability support is indicated by different coloured circles at the node (90-100% represented by red, 75-90% indicated by yellow, and 50-75% indicated by green). The sample ID of the isolates are removed because they are hard to visualize. The tree has a scale to visualize the age at the common node (shown at the bottom of the tree with time in years going back from the present time).