Global historic pandemics caused by the FAM-1 genotype of the Irish potato pathogen *Phytophthora infestans*

Amanda C. Saville[1], Jean B. Ristaino*[1,2]

[1] Department of Entomology and Plant Pathology, [2]Emerging Plant Disease and Global Food Security Cluster 112 Derieux Place, Box 7613, North Carolina State University, Raleigh, NC USA

Abstract

The FAM-1 genotype of *Phytophthora infestans* caused late blight in the 1840s in the US and Europe and was responsible for the Irish famine. We examined 140 herbarium specimens collected between 1845 and 1991 from six continents and used 12-plex microsatellite genotyping (SSR) to identify FAM-1 and the mtDNA lineage (Herb-1/ Ia) present in historic samples. FAM-1 was detected in approximately 73% of the historic specimens and was found on 6 continents. The US-1 genotype was found in only 27% of the samples and was found later on all continents except Australia/Oceania. FAM-1 was the first genotype detected in almost all the former British colonies from which samples were available. The data from historic samples suggest the FAM-1 genotype was widespread, diverse, and spread more widely than US-1. The famine lineage spread to six continents over 140 years, and likely spread during global colonization from Europe.

Introduction

Emerging plant diseases threaten crop production and forest ecosystems worldwide[1,2]. Movement of pathogens due to increased trade of plants and plant products has exacerbated outbreaks of plant diseases. *Phytophthora infestans* (Mont.) de Bary caused the Irish potato famine of the 1840s[2], limits potato and tomato production today, and threatens global food

security worldwide[1]. While *P. infestans* can be spread aerially through asexual sporangia, long distance movement of the pathogen is mainly due to transport of infected tubers for use as seed potatoes[2].

The history of *P. infestans* consists of a series of migrations, combined with periodic displacements of one clonal lineage with another, both of which have occurred at local and global scales[2-5]. The disease first appeared in the US in 1843, near the ports of New York, Philadelphia and in surrounding states (Supplementary Table 1) [4,5]. The disease was first reported on the European continent in Belgium in 1845, after which it spread throughout Europe and then into Ireland[2,5,6]. Historic populations of the pathogen in the US and Europe have been studied using mycological herbarium specimens to better understand the pathogen's origin, identify the outbreak strain, and track its spread from the Americas to Europe[6-13]. Herbarium specimens collected in the 1840s and later from original outbreak specimens revealed that the famine lineage was a Ia mitochondrial haplotype, disputing previous theories that the US-1 (Ib haplotype) lineage caused the famine [2,6,8,12,13]. The clonal lineage that caused the Irish potato famine (FAM-1) was identified, the genome was sequenced, and it's shared ancestry with *P. andina*, a sister species from South America, was documented[6,11].

Microsatellite genotyping (SSRs) has also been used widely to study the population biology of *P. infestans*[2,10,14]. *P. infestans*-infected leaves from historic specimens collected in North America and Europe were genotyped using SSRs and migration from North America into Europe was documented[10]. The FAM-1 genotype caused the first outbreaks in the US and Europe and was eventually displaced by the newly emerging US-1 genotype around the 1930s-1950s[9,10]. US-1 continued to persist globally until the 1980s, when it was replaced by more

aggressive lineages out of Mexico and Europe, but can still be found in select populations today[3,15].

Historic migrations of the pathogen into the US and Europe have been studied, but little is known about migrations of the FAM-1 genotype to other continents after the original 19th century outbreaks[6,10,12,13]. It has been suggested that FAM-1 was less fit and went extinct[8,16]. The earliest known records of *P. infestans* in Asia indicate it was present in India between 1870 and 1880[17], in Australia, in Tasmania in 1907[4,18], and in Kenya in 1941[19]. Genotyping of samples of *P. infestans* from eastern Africa from the mid-21st century revealed the presence of the US-1 genotype[20,21]. Recent genotyping of *P. infestans* from herbarium samples from South and Central America identified the FAM-1 genotype in Colombia, Guatemala, and Costa Rica between the 1910s and 1940s, suggesting that the FAM-1 genotype was present in these regions for many years[2,10].

The goal of this study was to examine the population structure of historic *P. infestans* using a large global set of outbreak samples[22]. The primary objectives of this research were: (1) to infer the population structure of historic *P. infestans* on six continents (2) Examine the spatial biogeography and diversity of FAM-1 and US-1 genotypes over space and time; and (3) compare the impact of host diversity on genotype diversity and (4) infer putative migration pathways of the pathogen into Africa, Asia and Australia/Oceania.

Results

Population Structure

A total of 137 historic samples were genotyped with microsatellites (Table 1). The FAM-1 (n=101) and US-1 (n=36) genotypes were identified in the set of historic samples. A subset of 67 samples were genotyped for mtDNA haplotype and identified as the Herb-1 mitochondrial

haplotype, while 32 specimens, all US_1's were the Ib mitochondrial haplotype. The FAM-1 genotype was the first lineage detected in specimens from most countries sampled (Table 1). (Fig. 1)

The earliest FAM-1 genotype was found in France in 1845 (K 71) and the most recent FAM-1 was found in Malaysia in 1987 (K 174) (Fig 1, Table1). This indicates that the famine lineage circulated for more than 144 years. In contrast, the earliest US-1 genotype was identified later in the US in 1931 (BPI 186927) and the most recent sample was from India in 1991 (IMI 344673) indicating that US-1 genotype circulated for 60 years, less than half the time of FAM-1.

Greater subclonal variation was found in FAM-1 than US-1 genotypes on all continents but Africa, where US-1 was more diverse (Table 1). There were 85 multilocus genotypes (MLGs) of FAM-1 and 36 MLGs of US-1 (Table 1). The greatest diversity of FAM-1 MLGs occurred in specimens from North America, Europe and Asia. FAM-1 displayed higher diversity values across all calculated indices, as well as a higher index of association (Ia) and standardized index of association ($\bar{r}_{d)}$ than US-1.

The majority of the herbarium samples (111) were collected from potato, while only 13 were from tomato (Table 1). The remaining specimens were from petunia or wild species of *Solanum species*. There were higher numbers of MLGs from potato than tomato, and greater genetic diversity and higher indices of association/$\bar{r}_d$ were found in specimens from potato than tomato.

Several SSR loci were useful for distinguishing the FAM-1 genotype from US-1. Diagnostic loci were Pi70 (192/192 in FAM-1 and 189/192 in US-1), PiG11 (160/200 in FAM-1 and 152/156/200 in US-1), PinfSSR2 (173/173 in FAM-1 and 173/177 in US-1), and Pi4B

(209/213 in FAM-1 and 213/217 in US-1). The average number of alleles was higher in FAM-1 (5.33) than in US-1(4.5) (Supplementary Table 2).

Structure analysis was done and the optimal K value was two based on results from Structure Harvester. At K=2, the Structure analysis grouped samples into two groups based on genotype (FAM-1 or US-1). FAM-1 was found before US-1 on each continent (Fig. 2). However. At K=3, US-1 genotypes were more homogeneous, while FAM-1 genotypes displayed allelic variation between the two remaining K groups. At K=3, it was noted that FAM-1 genotypes showed more allelic diversity and shifted assignment from one K group to the other over time, beginning around 1911. Both FAM-1 and US-1 occurred in many geographic locations.

A minimum spanning network (MSN) was made with haplotypes based on the continent where the samples were collected. Two large groups within the MSN consisting of either FAM-1 or US-1 genotype were observed (Fig. 3). Within the genotype clusters there was no genetic substructuring based on continent. However, the FAM-1 had a larger number of MLGs and more branches in the MSN than the US-1 and greater subclonal variation was observed (Table 1). No exclusive clusters were observed by country, host, or continent, but higher numbers of haplotypes of FAM-1 were found in North America and Asia than elsewhere. There were fewer FAM-1 MLGs from Africa and they occurred on fewer branches of the MSN than MLGs from other continents, most likely due to the more recent introductions there. In contrast, there were more MLGs of US-1 in Africa than North America or Europe, indicating diversification of the US-1 genotype in east Africa.

A similar structure was observed from the neighbor-joining tree, with two large clades that contained either FAM-1 or US-1 genotypes (Fig. 4). When the neighbor-joining tree was

expanded to include modern global samples of *P. infestans*, most of the samples identified as FAM-1 or US-1 genotypes also formed homogeneous clusters within the larger neighbor-joining tree using SSRs (Fig. S1).

Migration

Potential migration paths of *P. infestans* into Africa and Asia were tested using specimens identified as FAM-1genotype from Europe and North America. We calculated the probabilities of scenarios that hypothesized movement of *P. infestans* based on a North American source, a European source, and either source with a constant or a varying population size, and a source based on an admixture of European and North American populations (Fig. 5).

The most likely scenario indicated a divergence first of the European genotypes of FAM-1 from a North American source, followed by a divergence of the African and or Asian FAM-1 genotypes occurred from a European source (Scenario 1, P=0.3506 and 0.3222 for Africa and Asia, respectively) (Supplementary Table 3). Confidence in the scenario choice was evaluated by using simulated datasets to calculate error percentages between the three scenarios with the highest probabilities. Estimation of type I error for scenarios using the Asian data revealed that 53.2% of simulated datasets using this scenario resulted in the highest posterior probability for Scenario 1 when compared to the two scenarios with the next highest probabilities (Scenarios 3, 2) (type I error, 0.468) (Fig. 5 and Supplementary Table 3). For scenarios using the African data, estimation of type I error indicated that 35.6% of simulated datasets resulted in the highest posterior probability for Scenario 1 when compared to the next highest probabilities (Scenarios 2, 4) (type I error, 0.644).

Discussion

We examined historic outbreaks of *P. infestans* from global historic sources to better understand the history of the spread of the pathogen after the first recorded outbreaks in Europe and the US using herbarium voucher specimens collected worldwide. Our data revealed the widespread presence of the FAM-1 genotype throughout the world and its dominance until the 1930s, when the US-1 lineage began to emerge globally. By the end of the 1950s, FAM-1 had almost completely disappeared from collections and was displaced by US-1 genotype, most likely through movement of potatoes with resistance breeding efforts[2,10]. The only post-1950s FAM-1 genotype observed was collected in a single sample from Malaysia in 1987. This unusual sample suggests that the FAM-1 genotype may have continued to persist in remote areas of the world for a longer period. The Malaysian FAM-1 genotype had variable alleles at several loci when compared to earlier FAM-1 genotypes, suggesting the accumulation of mutations over time or potential hybridization with another lineage such as US-1. It clustered closely with the US-1 genotype in a neighbor- joining tree of a larger set of samples. Further work is underway to sequence the genome of this specimen to understand the variation in more detail.

The earliest known record of the US-1 genotype in potato is from 1931 in the US[10]. Based on specimens analyzed in this study, the earliest known records of US-1 in Africa are from 1953 in Cameroon and Nigeria. For Asia, the oldest US-1 sample observed in this study was from 1952 in China. In South America, the oldest US-1 lineage was from Bolivia in 1944. US-1 was not identified from the samples we examined from Australia in this study, although further work is underway in our lab to genotype more specimens from Australia.

There is scarce information on the history of the early emergence of US-1, but records from the literature suggest an approximate period for multiple parts of the world. In 1947, while documenting the history of late blight in Tasmania, Oldaker commented on an unusual outbreak

of *P. infestans* in 1938, in which disease was more sporadic than it had been in previous years, but treatment with copper formulations proved effective in controlling the pathogen[18]. In 1951, Nattrass wrote that potatoes bred for *P. infestans* resistance were failing with the emergence of a new biotype that appeared in Tanzania (then called Tanganyika) in 1946[19]. Our data suggest the new biotype observed was likely the US-1 lineage in Africa. The proximity of these outbreaks suggests that US-1 genotype was spread during potato breeding trials, facilitated by the continued movement of tubers over long distances.

Both the FAM-1 and US-1 genotype predominantly formed two clusters that excluded all modern lineages from Europe, North America, and South America. Our previous studies of North American populations of *P. infestans* suggested a Mexican origin for many of the recent lineages of *P. infestans* circulating in the US[14]. The US-1 genotype has been displaced by other lineages emerging from either Europe or Mexico with metalaxyl resistance or the ability to overcome host resistance genes being bred into potatoes and tomatoes at the time.

Migration analyses of FAM-1 outbreak samples was analyzed using DIYABC analysis and data suggest that both the African and Asian genotypes of FAM-1 most likely emerged from a European source. Outbreaks of the disease caused by FAM-1 genotype first occurred in North America and subsequently spread to Europe[10]. This coincides with historic records and our previous studies that support the migration of *P. infestans* into Europe after outbreaks occurred in North America[2,10,22]. The top migration scenario for emergence into Africa and Asia is from a European source. We do not attempt to identify source countries within Europe but likely sources from historical records include countries such as the UK and or the NL.

Our data and examination of global herbarium sources suggest that the pathogen likely moved first on potatoes and then spread later into tomato[22.] There was greater genetic diversity

among potato than tomato genotypes of FAM-1 and more MLGs among the potato genotypes. There were also more infected potato than tomato specimens in a global search of archival collections[22]. The FAM-1 genotype was also more genetically diverse than the US-1 genotype.

The findings of our study are supported by historical reports published by researchers contemporary to the time of the initial outbreaks and provide insight into potential sources[4,17-19,23]. Potatoes were disseminated across the world by European sailors and missionaries, with varying degrees of adoption by local populations[24]. As European colonists moved into new regions, potatoes moved with them. Potatoes were actively encouraged as food for native people during colonization and touted as cheap and nutritious. Potatoes were cited by European scholars to "elevate the happiness and well-being of native peoples", and subsequently were useful for developing the labor force of the colonizing empire[25]. In India, this mentality resulted in the dissemination of potatoes to local villages by horticultural and agricultural societies, despite the crop having already been adopted as a cash crop to sell to British soldiers[25,26].

Potatoes continued to move into colonial regions long after their establishment. A colonial handbook for Kenya from 1920 states that potatoes grown from locally produced seed were not as productive, and recommended regularly importing fresh seed potatoes from Europe. This would have provided an obvious avenue for the introduction of *P. infestans*[27]. Regular imports of potatoes were observed in other parts of the world as well. A 1904 agricultural report for the Philippines compared the quality of natively grown potatoes to imported ones found in Manila markets, suggesting a potential introduction route, mostly likely via the US[28]. In India, the pathogen was reported in the area circa 1870-1880 based on reports from local agri-horticultural societies. In letters it was stated that a major late blight outbreak in the Nilgiri region around 1893 was the result of the importation of potatoes from a large nursery in

England[17]. In East Africa it was believed that the first outbreak located outside of Nairobi, Kenya, was the result of an importation of Kerr's Pink potatoes for planting from the United Kingdom, bolstered by a wet and rainy season[4]. In West Africa, however, it was thought that *P. infestans* was introduced as the result of the importation of potatoes from France. While intended for use as food, potatoes were also planted, resulting in the propagation of the pathogen[23]. *Phytophthora infestans* followed the movements by colonists of potato, leading to its introduction from the US and Europe into the African, Asian and Austalian/Oceanic continents.

With the extensive reach of the British Empire (Fig. 1), it is likely that many introductions of the pathogen were the result of movement of potatoes on British ships, with multiple introductions over time as new shipments of tubers were imported into colonies. We are currently doing whole genome analysis of globally sourced specimens to provide more information on the role of host diversity and host jumps in global pathogen spread. The FAM-1 genotype was diploid[11] and asexual, was able to colonize susceptible potato on six continents and thus caused global pandemics. Our data document that the FAM-1 genotype adapted to many environments, occurred mostly on potato, and remained aggressive for over 140 years.

Methods

Over 1280 late blight specimens collected in the 19th and 20th century are in herbaria on six continents[22]. We sampled specimens from 37 countries on six continents including North America, South America, Europe, Africa, Asia and Oceania (Fig. 1) (Supplementary Table 1). *Phytophthora infestans* was sampled from herbarium specimens collected in Africa, Asia, Europe, Oceania, North America, and South America. A total of 137 samples were genotyped with microsatellites, consisting of 18 African samples (1942 – 1973), 32 Asian samples (1901 –

1991), 2 Oceania samples (1911 – 1917), 29 European samples (1873 – 1970), 48 North American samples (1855 – 1958), and 8 South American samples (1913 – 1967) (Supplementary Table 1).

We collected SSR data for an additional 194 samples was obtained from databases, published studies, and theses[10,29,30]. These included data from modern populations from Saville et al.[10] and representatives of current common European lineages from the outbreak tracking system Euroblight.org. In addition we included a subset of data from publications on current Mexican populations[29,30].

DNA Extraction, PCR, and Genotyping

DNA was extracted from lesions present on each herbarium voucher using either a Qiagen DNEasy Plant Mini Kit (Qiagen, Valencia, CA) or a modified CTAB method using DNEasy Plant Mini Kit spin columns for cleaning and purifying DNA[12]. The presence of *P. infestans* DNA was checked using species specific primers[31]. All work with herbarium DNA was conducted in a lab in which no modern DNA of *P. infestans* is used, using separate equipment and reagents.

Mitochondrial haplotyping of samples was conducted using primers and PCR cycling conditions developed by Griffith and Shaw to detect the presence of the Ib haplotype[11,32]. For detecting the Herb-1 haplotype we utilized primers previously developed that target a single nucleotide polymorphism within the haplotype[10]. Amplicons were sequenced at the Genomic Sciences Laboratory at North Carolina State University.

Samples were genotyped using a 12-plex system of primers for the identification of *P. infestans* lineages using microsatellite loci[33]. To compensate for the low levels of DNA present in extractions, a modification of the PCR protocol was used that increased primer concentration,

sample size, and cycling times[15]. The Qiagen Type-It Microsatellite PCR kit (Qiagen Corporation, Valenica CA) was used for PCR reactions, and sample volumes were modified to run a 12.5µL reaction, consisting of 6.25µl of Type-It 2X master mix, 1.3µl of a 10X primer mix (Supplementary Table 4), 1.95µl ddH2O, and 1 – 3µl of DNA extract. Thermal cycling conditions consisted of initial denaturation at 95ºC for 5 min, followed by 33 cycles of 95ºC for 30 seconds, 58ºC for 90 seconds, and 72ºC for 30 seconds, and then a final extension period for 30 minutes at 60ºC. Fragments were analyzed on an Applied Biosystems 3730xl DNA analyzer at the Genomic Sciences Laboratory at North Carolina State University using 1-3µl of PCR product in a 10.3µL reaction mix consisting of 10µL highly deionized formamide and 0.3µL LIZ500 size standard (Applied Biosystems, Foster City, CA).. Alleles were scored in Geneious 11.1.5 (Biomatters Ltd., Auckland, NZ) using microsatellite plugin 1.4.6. Alleles were named using bin ranges from previously published work[33,34].

SSR Data Analysis

Because of the age of herbarium DNA, recovery rates of microsatellite loci from *P. infestans* are lower than they would be for DNA extracted from modern samples, resulting in increased missing data. To reduce the amount of variability due to missing data, only samples with data from at least five SSR loci were used. Data were divided into six categories based on continent: Africa (Afr), Asia (As), Europe (EU), North America (NA), Oceania (Oc), and South America (SA). The broad structure of the populations was evaluated via model-based Bayesian clustering using the program Structure v. 2.3.3 [35]. Before analysis by Structure, the data were clone corrected (clones were removed such that each population contains only one representative of each haplotype) using the R library poppr v. 2.8.1 [36] and R v. 3.5.2 [37]. Data were clone corrected using their region of collection as a population. The data were run using a 20,000

repeat burn-in and 1,000,000 MCMC repeats under a no admixture model, with each individual sample representing its own population. Independent runs of the model used *K* values from 1 to 10 with 10 replicate runs at each value of *K*. The optimal *K* was estimated using the second order rate of change (the "Evanno method") in the web tool Structure Harvester [38,39]. All runs for the optimal *K* value, as well as non-optimal *K* values, were averaged using CLUMPP v. 1.1.2 [40] using the Greedy algorithm (M=2) with the pairwise matrix similarity statistic G' (S=2). The Greedy algorithm was used with 1000 repeats of randomly selected input orders. The resulting output was visualized with the program distruct v. 1.1[41]. Poppr was also used to infer population statistics including: the number of samples (N), the number of multilocus genotypes (MLG), the number of expected MLGs at the smallest sample size of at least 10 (eMLG)[42], the Shannon-Weiner index of MLG diversity (H)[43], the Stoddart and Taylor index of MLG diversity (G)[44], Simpson index corrected for sample size by multiplying the index value by N/(N-1) (λ)[45,46], evenness[47-49], Nei's unbiased gene diversity (Hexp)[50], the index of association (Ia)[51,52], and the standardized index of association($\bar{r}_d$)[53].

Relationships between locations and haplotypes of samples were further explored using a minimum spanning network (MSN) based on Bruvo's distance using the R library adegenet v 2.1.1 [54]. In addition, a neighbor-joining (NJ) tree based on Bruvo's distance was constructed using the poppr R library and a combination (genome addition and genome loss) model. In order to utilize a complete dataset in the NJ tree for the purposes of bootstrapping, five loci with low recovery rates were removed for tree construction (PinfSSR8, PinfSSR4, Pi63, PinfSSR11, Pi4B). Any remaining samples still containing missing data were removed. The tree was bootstrapped using 1000 samplings.

297 An additional 7-plex neighbor joining tree was generated as above using the combined
298 dataset from herbarium and modern samples. Due to the bootstrapping function used
299 (bruvo.boot), no samples with missing data or null alleles could be used. Therefore, samples with
300 putative null alleles were also removed from the dataset for the neighbor-joining tree.

301 Migration of FAM-1

302 Migration routes of the FAM-1 lineage of *P. infestans* into Africa and Asia from Europe
303 and/or North America were examined using Approximate Bayesian Comparison (ABC), as
304 implemented in the program DIYABC v. 2.0.4[55]. Tested migration scenarios for both African
305 and Asian populations included direct divergence from Europe or North America, divergence
306 including a change in population size, or admixture between European and North American
307 populations. Parameter range priors were initialized with values from Saville et al.[10] and then
308 iteratively modified to better fit our data (Supplementary Table 5). A total of 5 million datasets
309 were simulated. Scenario probabilities were determined through comparison of the observed
310 dataset to simulated datasets generated by DIYABC. A logistic regression of these differences
311 was computed using ten proportions of the simulated dataset as the dependent variable and
312 corresponding differences between the observed and simulated datasets as the independent
313 variable. The value calculated using 50,000 simulated datasets was taken as the scenario's
314 overall probability. Confidence in the highest scenario was evaluated using a type I error tests, in
315 which the data were compared against 500 simulated data sets assuming the scenario with the
316 highest probability is true and the number of times the scenario in question was correctly or
317 incorrectly applied to the data was determined.

318 Acknowledgments

Appreciation is expressed to the following herbaria for providing herbarium material: Royal Botanic Gardens Kew Mycological Herbarium (K), the National Botanic Garden, Glasnevin, Dublin (DBN), the USDA National Fungus Collection, Beltsville, MD (BPI), the CABI Bioscience collection, Egham (IMI), the Farlow Herbarium Harvard University, Cambridge, MA (FH), the Museum of Evolutionary Biology, Uppsala University, Uppsala (UPS), the Cornell Plant Pathology Herbarium, Ithaca, NY (CUP), and the University of Florida Herbarium, Gainesville, FL (FLAS). Thanks to David Cooke, James Hutton Institute, for providing a set of European lineages from EuroBlight database shown in Supplemental Figure S1. Funding was provided by USDA AFRI Grant Number 5197-NCSU-USDA-3179 and USDA AFRI Grant 2011-68004-30154 and the North Carolina Agricultural Research Service.

Author Contribution

Jbr collected the samples and conceived experiments. AS conducted experiments and analysed data. JBR and AS interpreted data and co-wrote the paper. JBR and AS contributed equally as coauthors of this work

Data Availability

Raw SSR data and binning rules can be found on GitHub link to be provided once paper is accepted. See the references in the Supplementary Information for data used in the analysis.

References

1 Anderson, P. K. *et al.* Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol Evol* **19**, 535-544, doi:10.1016/j.tree.2004.07.021 (2004).

2 Ristaino, J., Cooke, D. E., Acuña, I. & Muñoz, M. in *Emerging Plant Diseases and Global Food Security* (eds J. Ristaino & A. Records) 101-133 (American Phytopathological Society Press, 2020).

3 Hu, C. H. *et al.* Recent genotypes of *Phytophthora infestans* in the eastern United States reveal clonal populations and reappearance of mefenoxam sensitivity. *Plant Disease* **96**, 1323-1330 (2012).

4 Cox, A. E. & Large, E. C. Potato blight epidemics throughout the world. *Agriculture Handbook* **174**, 1-230 (1960).

5 Bourke, P. M. A. Emergence of Potato blight, 1843-46. *Nature* **203**, 805-808(1964).

6 Martin, M. D. *et al.* Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nature Communications* **4**, 2172, doi:10.1038/ncomms3172; 10.1038/ncomms3172 (2013).

7 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*. *Molecular Biology and Evolution* **31**, 1414-1420, doi:10.1093/molbev/msu086 (2014).

8 Yoshida, K. *et al.* The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife* **2**, e00731, doi:10.7554/eLife.00731 (2013).

9 Yoshida, K. *et al.* Mining herbaria for plant pathogen genomes: back to the future. *PLOS Pathogens* **10**, e1004028, doi:10.1371/journal.ppat.1004028 (2014).

10 Saville, A. C., Martin, M. D. & Ristaino, J. B. Historic late blight outbreaks caused by a widespread dominant lineage of *Phytophthora infestans* (Mont.) de Bary. *PLOS ONE* **11**, e0168381, doi:10.1371/journal.pone.0168381 (2016).

11 Martin, M. D. *et al.* Genomic characterization of a South American *Phytophthora* hybrid mandates reassessment of the geographic origins of *Phytophthora infestans*. *Mol Biol Evol* **33**, 478-491, doi:10.1093/molbev/msv241 (2016).

12 May, K. J. & Ristaino, J. B. Identity of the mtDNA haplotype(s) of *Phytophthora infestans* in historical specimens from the Irish potato famine. *Mycological Research* **108**, 471-479 (2004).

13 Ristaino, J. B., Groves, C. T. & Parra, G. R. PCR amplification of the Irish potato famine pathogen from historic specimens. *Nature* **411**, 695-697, doi:10.1038/35079606 (2001).

14 Saville, A. & Ristaino, J. B. Genetic structure and subclonal variation of extant and recent U.S. lineages of *Phytophthora infestans*. *Phytopathology* **109**, 1614-1627, doi:10.1094/phyto-09-18-0357-r (2019).

15 Cooke, D. E. L., Cano, L. M., Raffaele, S., Bain, R. A., Cooke, L. R., Etherington, G. J., Deahl, K., 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., and Kamoun, S. Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. PLOS Pathogens **8,** e1002940. (2012).

16 Birch, P. R. J. & Cooke, D. E. L. The early days of late blight. *eLife* **2**, e00954-e00954, doi:10.7554/eLife.00954 (2013).

17 Butler, E. J. Potato diseases of India. *The Agricultural Ledger* **10**, 87-124 (1903).

18 Oldaker, C. E. W. Blight (*Phytophthora infestans*) of potatoes in Tasmania. *Tasmanian Journal of Agriculture* **18**, 137-140 (1947).

19 Nattrass, R. M. & Ryan, M. New hosts of *Phytophthora infestans* in Kenya. *Nature* **168**, 85-86, doi:10.1038/168085b0 (1951).

20 Vega-Sánchez, M. E. *et al.* Host adaptation to potato and tomato within the US−1 clonal lineage of *Phytophthora infestans* in Uganda and Kenya. *Plant Pathology* **49**, 531-539, doi:10.1046/j.1365-3059.2000.00487.x (2000).

21 Ochwo, M. K. N. *et al.* Genetic diversity of *Phytophthora infestans* (Mont.) de Bary in the Eastern and Western highlands of Uganda. *Journal of Phytopathology* **150**, 541-542, doi:10.1046/j.1439-0434.2002.00794.x (2002).

22 Ristaino, J. B. The importance of mycological and plant herbaria in tracking plant killers. *Frontiers in Ecology and Evolution* **7**, doi:10.3389/fevo.2019.00521 (2020).

23 Russell, T. A. Potato blight in West Africa. *Empire Journal of Experimental Agriculture* **22**, 19-22 (1954).

24 Laufer, B. & Wilbur, C. M. *The American plant migration. Part I: the potato. Part I: the potato*. (Field Museum of Natural History, 1938).

25 Earle, R. Food, colonialism and the quantum of happiness. *History Workshop Journal* **84**, 170-193, doi:10.1093/hwj/dbx046 (2017).

26 Arnold, D. Agriculture and 'improvement' in early colonial India: A pre-history of development. *Journal of Agrarian Change* **5**, 505-525, doi:10.1111/j.1471-0366.2005.00110.x (2005).

27 Great Britain. *A handbook of Kenya Colony (British East Africa) and the Kenya Protectorate (protectorate of Zanzibar)*. 413 (H.M. Stationery Off.; Printed by F. Hall at the University Press, 1920).

28 Philippines. Report of the Bureau of Agriculture for the year ended August 31, 1904. *Report of the Insular Bureau of Agriculture 1905*, v. (1902).

29 Shakya, S. K., Larsen, M. M., Cuenca-Condoy, M. M., Lozoya-Saldaña, H. & Grünwald, N. J. Variation in genetic diversity of *Phytophthora infestans* populations in Mexico from the center of origin outwards. *Plant Disease* **102**, 1534-1540, doi:10.1094/PDIS-11-17-1801-RE (2018).

30 Wang, J. *et al.* High levels of diversity and population structure in the potato late blight pathogen at the Mexico centre of origin. *Molecular Ecology* **26**, 1091-1107, doi:10.1111/mec.14000 (2017).

31 Ristaino, J. B., Hu, C. H. & Fitt, B. D. L. Evidence for presence of the founder Ia mtDNA haplotype of *Phytophthora infestans* in 19th century potato tubers from the Rothamsted archives. *Plant Pathology* **62**, 492-500, doi:doi:10.1111/j.1365-3059.2012.02680.x (2013).

32 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. *Applied and Environmental Microbiology* **64**, 4007-4014 (1998).

33 Li, Y., Cooke, D. E. L., Jacobsen, E. & van der Lee, T. Efficient multiplex simple sequence repeat genotyping of the oomycete plant pathogen *Phytophthora infestans*. *Journal of Microbiological Methods* **92**, 316-322 (2013).

34 Martin, F. N. *et al.* Insights into evolving global populations of Phytophthora infestans via new complementary mtDNA haplotype markers and nuclear SSRs. *PLOS ONE* **14**, e0208606, doi:10.1371/journal.pone.0208606 (2019).

35 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959 (2000).

36 Kamvar, Z. N., Tabima, J. F. & Grünwald, N. J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281, doi:10.7717/peerj.281 [doi] (2014).

37 R: A language and environment for statistical computing v. 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria, 2019).

38 Earl, D. A. & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361, doi:10.1007/s12686-011-9548-7 (2012).

39 Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611-2620, doi:10.1111/j.1365-294X.2005.02553.x (2005).

40 Jakobsson, M. & Rosenberg, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806, doi:10.1093/bioinformatics/btm233 (2007).

41 Rosenberg, N. A. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138, doi:10.1046/j.1471-8286.2003.00566.x (2004).

42 Hurlbert, S. H. The nonconcept of species diversity: A critique and alternative parameters. *Ecology* **52**, 577-586, doi:10.2307/1934145 (1971).

43 Shannon, C. E. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.* **5**, 3-55, doi:10.1145/584091.584093 (2001).

44 Stoddart, J. A. & Taylor, J. F. Genotypic diversity: estimation and prediction in samples. *Genetics* **118**, 705 (1988).

45 Simpson, E. H. Measurement of diversity. *Nature* **163**, 688, doi:10.1038/163688a0 (1949).

46 Wang, J. *et al.* High levels of diversity and population structure in the potato late blight pathogen at the Mexico centre of origin. *Molecular Ecology* **26**, 1091-1107, doi:10.1111/mec.14000 (2017).

47 Pielou, E. C. *Ecological Diversity*. (New York : Wiley, 1975., 1975).

48 Ludwig, J. A. R., J.F. *Statistical ecology : a primer on methods and computing*. (New York : Wiley, c1988., 1988).

49 Grünwald, N. J., Goodwin, S. B., Milgroom, M. G. & Fry, W. E. Analysis of genotypic diversity data for populations of microorganisms. *Phytopathology* **93**, 738-746, doi:10.1094/PHYTO.2003.93.6.738 (2003).

50 Nei, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583 (1978).

51 Brown, A. H. D., Feldman, M. W. & Nevo, E. Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* **96**, 523 (1980).

52 Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4384-4388 (1993).

53 Agapow, P.-M. & Burt, A. Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* **1**, 101-102, doi:10.1046/j.1471-8278.2000.00014.x (2001).

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

55 Cornuet, J. M., Ravigne, V. & Estoup, A. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* **11**, 401, doi:10.1186/1471-2105-11-401 (2010).

Table 1: Population statistics of populations of *Phytophthora infestans* from historic global outbreaks sorted by genotype and by region and based on microsatellite data obtained from herbarium specimens.

Population[a]		***N***[b]	**MLG**	**eMLG(SE)** [d]	**H**	**G**	**λ**	**Evenness**	**Hexp**	**Ia**[e]	$\bar{r}_d$
Africa	***Genotype***										
	FAM-1	5	5	5.00(0.00)	1.61	5.0	1	1.000	0.437	0.936	0.1759
	US-1	13	13	10(7.3e-8)	2.56	13.0	0.9999	1.000	0.529	0.417	0.0571
	Host										
	Solanum lycopersicum	3	3	3(0e0)	1.10	3.0	1	1.000	0.555	3.38	0.647
	S. tuberosum	13	13	10(7.30e-8)	2.56	13.0	0.9999	1.000	0.569	1.51	0.159
	Other (*Petunia sp.*, *S. incanum*)	2	2	2(0e0)	0.69	2.0	1	1.000	0.764	NA	NA
	Total	18	18	10(5.43e-7)	2.89	18.0	0.9995	1.000	0.568	1.80	0.191
Asia	***Genotype***										
	FAM-1	21	21	10.00(0.00)	3.05	21.0	0.9996	1.000	0.404	0.136	0.0188
	US-1	11	11	10.00(0.00)	2.40	11.0	0.9999	1.000	0.481	0.0458	0.00671
	Host										
	S. lycopersicum	4	4	4(0e0)	1.39	4.0	1	1.000	0.561	3.25	0.411

	S. tuberosum	23	23	10(5.03e-7)	3.14	23.0	1.0005	1.000	0.497	1.08	0.111
	Other (*S. laciniatum*, *S. lyratum*, *S. marginatum*, *S. melongena*, *S. xanthocarpum*)	5	5	5(0e0)	1.61	5.0	1	1.000	0.533	0.15	0.0261
	Total	32	32	10(0e0)	3.47	32.0	1.0002	1.000	0.544	1.113	0.1168
Europe	***Genotype***										
	FAM-1	26	26	10.00(1.09e-6)	3.26	26.0	1.0005	1.000	0.390	0.274	0.0315
	US-1	3	3	3(0e0)	1.10	3.0	1.0005	1.000	0.560	-1.11e-16	-8.33e-17
	Host										
	S. lycopersicum	3	3	3(0e0)	1.10	3.0	1.334	1.000	0.386	-0.50	-0.50
	S. tuberosum	24	24	10(4.38e-7)	3.18	24.0	0.9997	1.000	0.506	2.77	0.26
	Other (*S. dulcamara*, *S. nigrum*)	2	2	2(0e0)	0.69	2.0	1	1.000	0.548	NA	NA
	Total	29	29	10(1.03e-6)	3.37	29.0	1.0005	1.000	0.495	2.53	0.23
North America	***Genotype***										

	FAM-1	42	31	8.91(9.63e-1)	3.24	18.4	0.9691	0.711	0.337	-0.189	-0.0303
	US-1	6	6	6(0e0)	1.79	6.0	0.9996	1.000	0.540	1.120	0.123
	Host										
	S. lycopersicum	3	3	3(0e0)	1.10	3.0	1.0005	1.000	0.490	0.333	0.075
	S. tuberosum	43	43	10(9.46e-7e0)	3.76	43.0	0.9998	1.000	0.457	3.695	0.354
	Other (*S. nigrum*, *S. sp.*)	2	2	2(0e0)	0.69	2.0	1	1.000	0.488	NA	NA
	Total	48	48	10(2.87e-6)	3.87	48.0	0.9998	NA	0.453	3.461	0.336
Australia/ Oceania[c]	***Total***	2	2	2.00(0e0)	0.69	2.0	1	1.000	0.286	NA	NA
South America	***Genotype***										
	FAM-1	5	5	5(0e0)	1.609	5.0	1	1.000	0.480	2.289	0.2501
	US-1	3	3	3(0e0)	1.10	3	1.0005	1.000	0.436	-0.333	-.200
	Host										
	Other (*Petunia sp.*)	1	1	1(0e0)	0.0	1.0	NA	NA	0.364	NA	NA
	S. tuberosum	7	7	7(0e0)	1.95	7.0	0.9998	1.000	0.519	2.326	0.2204
	Total	8	8	8(0e0)	2.08	8.0	0.875	1.000	0.566	0.982	0.0866

Totals	***Genotype***									
	FAM-1	101	85	32.7(1.57)	4.30	53.4	0.9908	0.718	0.381	0.41	0.0702
	US-1	36	36	36(0e0)	3.58	36	0.9997	1.000	0.508	0.232	0.0258
	Host									
	S. lycopersicum	13	13	12(0e0)	2.57	13.0	0.9999	1.000	0.461	1.45	0.194
	S. tuberosum	112	99	11.7(5.4e-1)	4.51	73.8	0.9950	0.813	0.461	2.12	0.220
	Other	12	12	12(0e0)	2.48	12.0	1.0003	1.000	0.561	1.41	0.135
	Total	137	121	34(1.36e-0)[d]	4.69	82.7	0.9952	0.757	0.471	1.903	0.1952

[a] All isolates, including those with missing data (minimum=5) are included.

[b]n: number of individuals ; MLG: number of multilocus genotypes (MLG); eMLG: expected number of MLGs at smallest size of at least ten; SE: Standard error; H: Shannon-Weiner Index of MLG diversity; G: Stoddart and Taylor Index of MLG diversity; λ: corrected Simpson's Index; Hexp: Nei's 1978 gene diversity; Ia: Index of Association; $\bar{r}_d$: standardized index of association.

[c]Only two samples were collected from Oceania, and both were collected from potato and identified as FAM-1

[d]eMLG for the entire sample set was based on population counts by genotype

[e]Data for Ia and $\bar{r}_d$ were clone corrected before analysis.

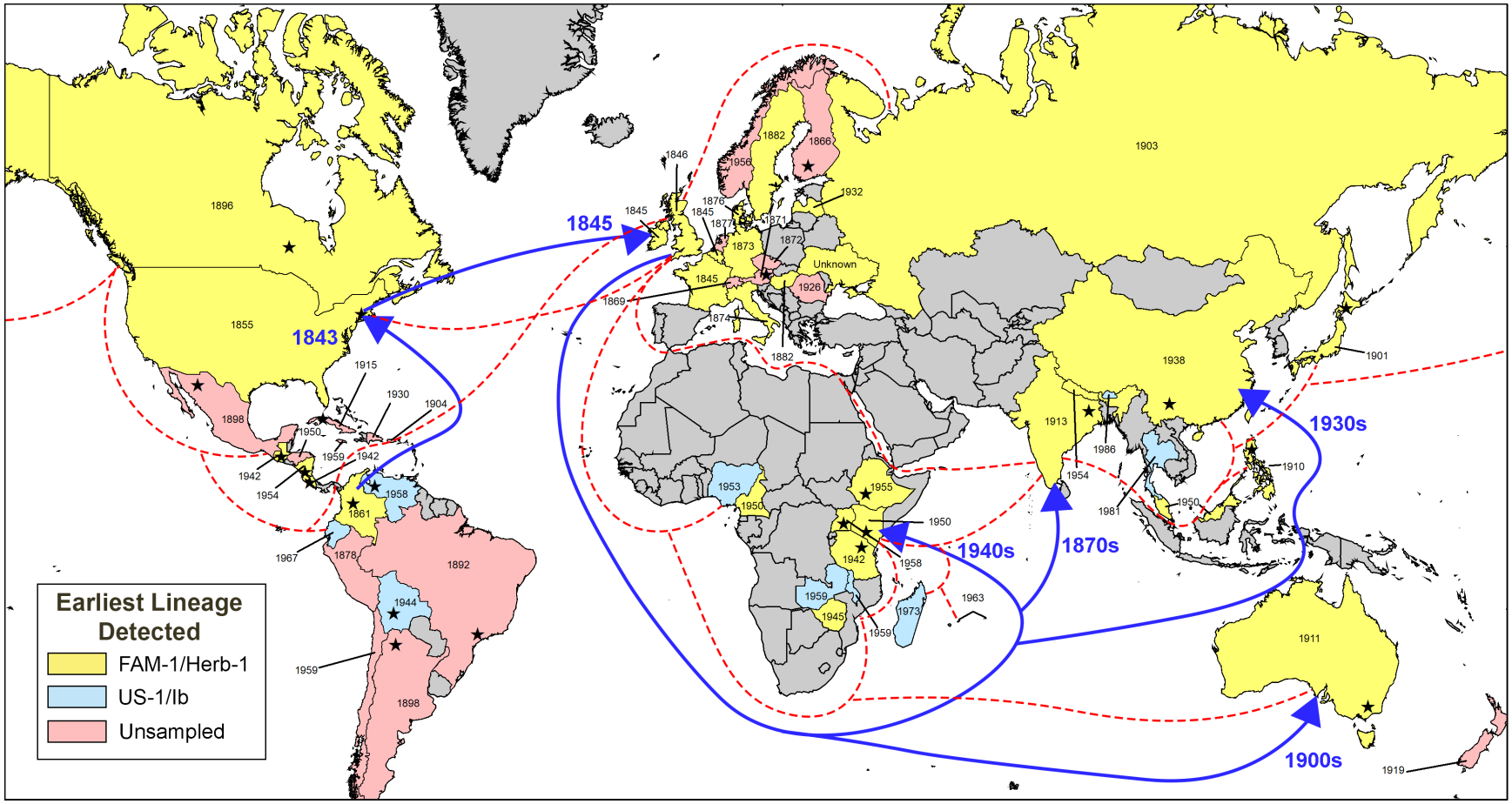

Figure 1: Global map of early outbreaks of late blight caused by *Phytophthora infestans.* Years within each country indicate the date of the earliest known specimen, while color indicates if the genotype was FAM-1, US-1 or unsampled. Dotted lines indicate representative trade routes of the British Empire circa 1932. Arrows indicate the most likely migration path taken by the FAM-1 lineage into Africa and Asia based on DIYABC analysis and trade routes. Stars within each country indicate the approximate location of the first recorded outbreak, if known.

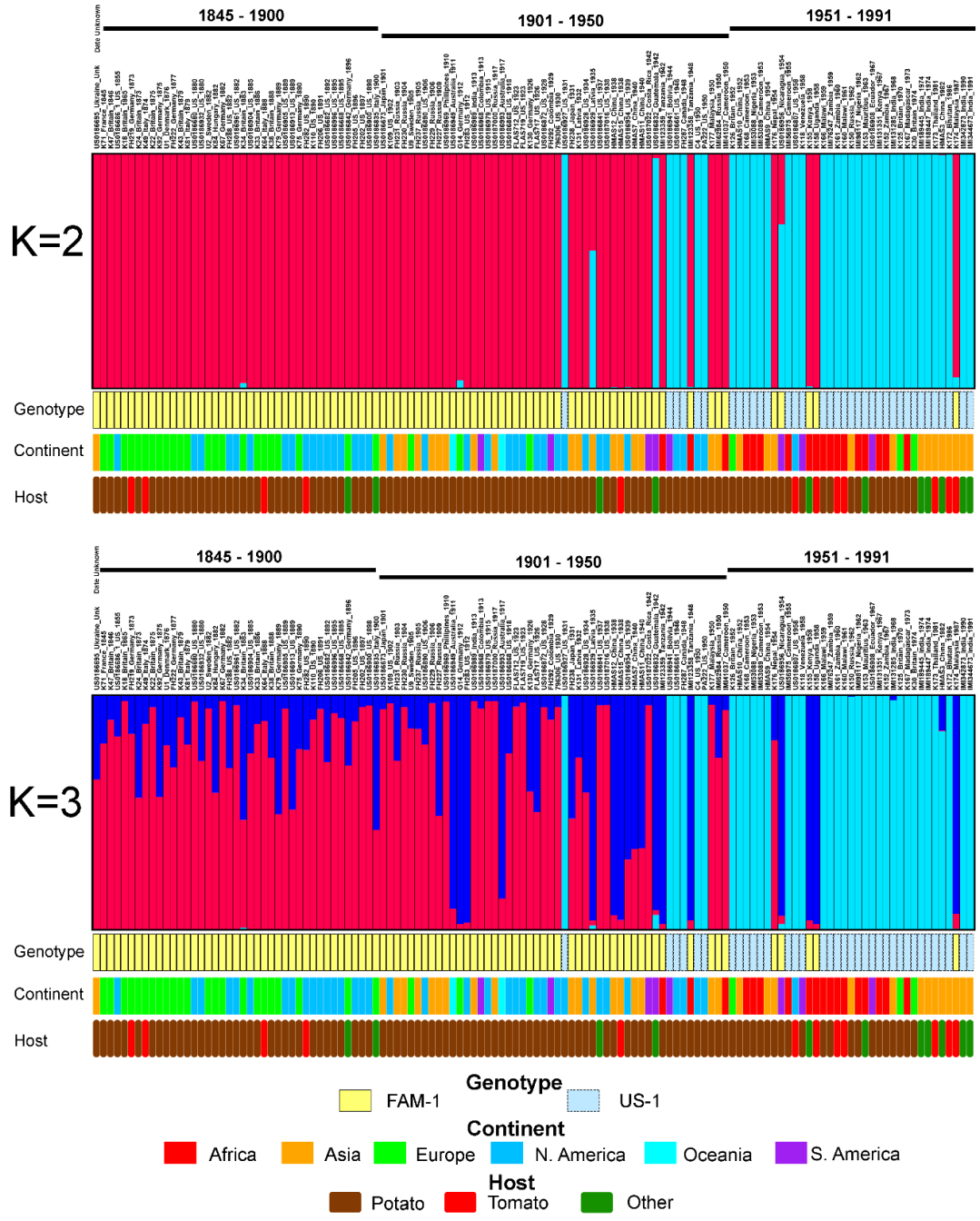

Figure 2: Structure analysis of SSR genotypes of *Phytophthora infestans* from herbarium specimens. Specimens are arranged in ascending chronological order. Samples were clone corrected based on region before analyzing. Results displayed are based under the assumption of two or three groups within the dataset (K=2, K=3). K=2 was calculated to be the most likely based on the second order rate of change.

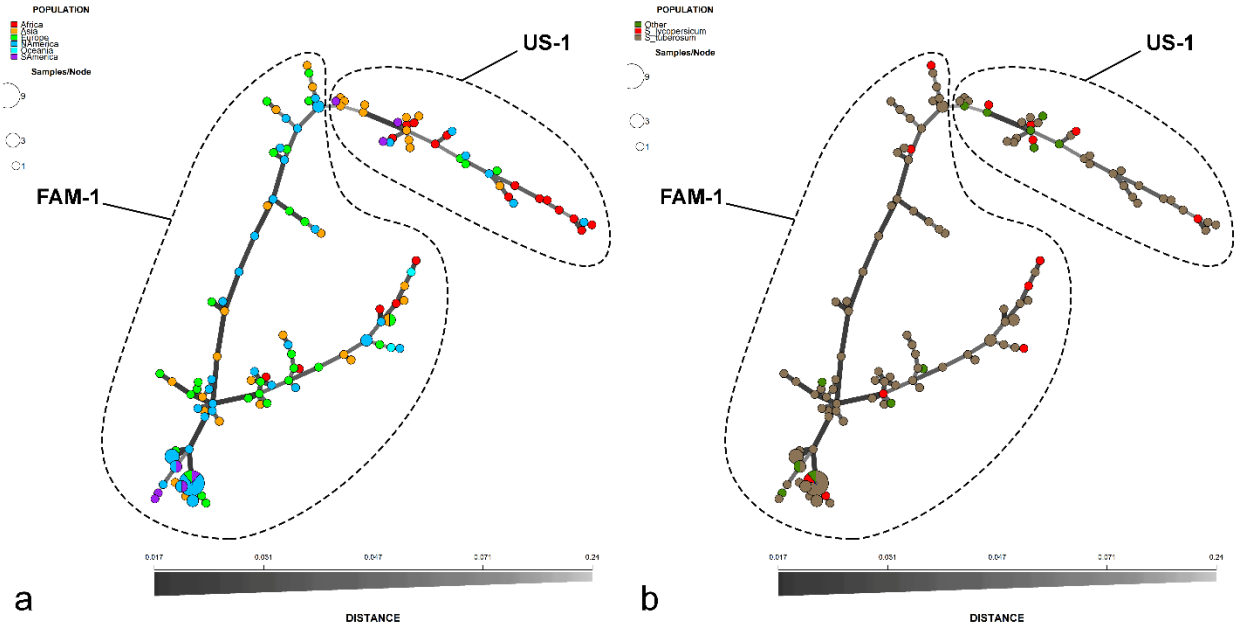

Figure 3: Minimum spanning network of SSR genotypes of *Phytophthora infestans* from herbarium specimens. Data are colored-coded based on (a) the continent where they were collected or (b) host. Genetic distance between haplotypes is indicated by the shade and thickness of the branches.

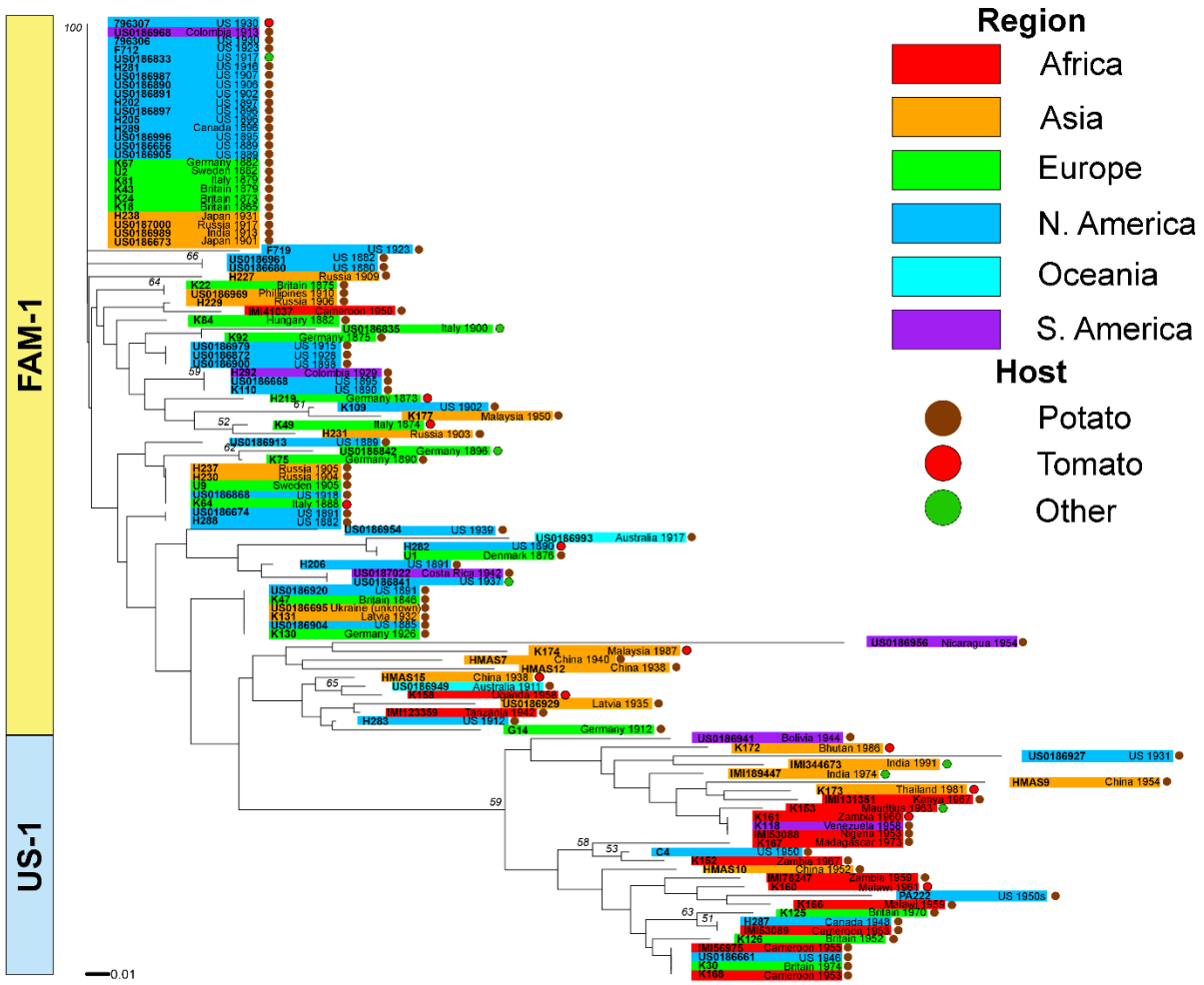

Figure 4. Neighbor joining tree of 7-plex SSR genotypes of *Phytophthora infestan* from herbarium specimens, Specimens are colored-coded based on the continent and host of sample. Bootstrapping was performed with 1000 replicates.

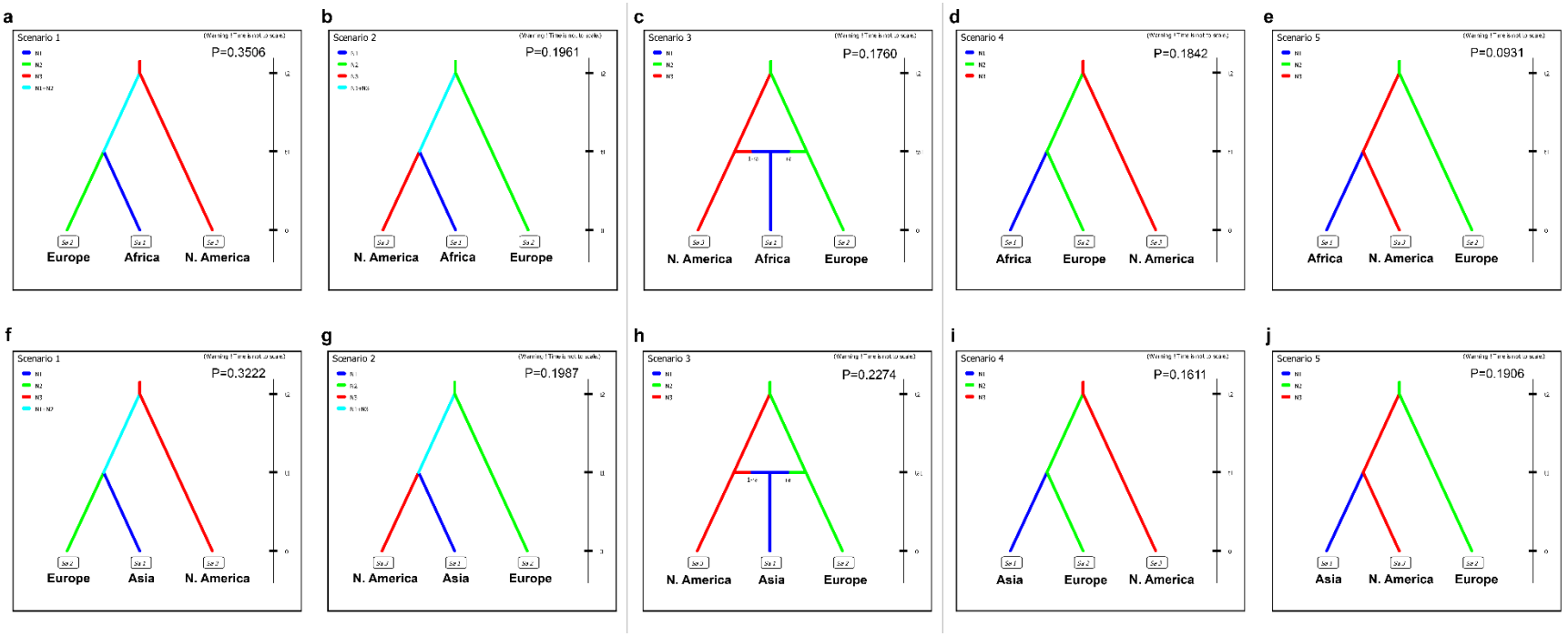

Figure 5. Posterior probabilities of migration scenarios involving populations of the FAM-1 genotype of *Phytophthora infestans* from North America and Europe to Africa(a-e) or North America and Europe to Asia (f-j). Probabilities are based on 1% of the simulated data.