Population structure and migration of the Tobacco Blue Mold Pathogen, Peronospora tabacina into North America and Europe

Article *in* Molecular Ecology · December 2017

DOI: 10.1111/mec.14453

3 authors, including:

Jean Beagle Ristaino
North Carolina State University
189 PUBLICATIONS **4,878** CITATIONS

Received: 28 April 2016 | Revised: 13 October 2017 | Accepted: 8 November 2017

DOI: 10.1111/mec.14453

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Population structure and migration of the Tobacco Blue Mold Pathogen, *Peronospora tabacina*, into North America and Europe

Monica Blanco-Meneses[1] | Ignazio Carbone[2] | Jean B. Ristaino[1]

[1]Department of Entomology and Plant Pathology, NC State University, Raleigh, NC, USA

[2]Center for Integrated Fungal Research, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA

Correspondence
Jean B. Ristaino, Department of Entomology and Plant Pathology, NC State University, Raleigh, NC, USA.
Email: Jean_Ristaino@ncsu.edu

Present address
Monica Blanco-Meneses, Laboratorio de Técnicas Moleculares aplicadas a la Fitoprotección, CIPROC, Universidad de Costa Rica, San Jose, Costa Rica.

Funding information
Tobacco Foundation of North Carolina; Tobacco Commission of North Carolina; National Science Foundation (NSF) Genealogy of Life (GoLife), Grant/Award Number: DEB-1541418

Abstract

Tobacco blue mold, caused by *Peronospora tabacina*, is an oomycete plant pathogen that causes yearly epidemics in tobacco (*Nicotiana tabacum*) in the United States and Europe. The genetic structure of *P. tabacina* was examined to understand genetic diversity, population structure and patterns of migration. Two nuclear loci, *lgs2* and *Ypt1*, and one mitochondrial locus, *cox2*, were amplified, cloned and sequenced from fifty-four isolates of *P. tabacina* from the United States, Central America–Caribbean–Mexico (CCAM), Europe and the Middle East (EULE). Cloned sequences from the three genes showed high genetic variability across all populations. Nucleotide diversity and the population mean mutation parameter per site (Watterson's theta) were higher in EULE and CCAM and lower in U.S. populations. Neutrality tests were significant and the equilibrium model of neutral evolution was rejected, indicating an excess of recent mutations or rare alleles. Hudson's S_{nn} tests were performed to examine population subdivision and gene flow among populations. An isolation-with-migration analysis (IM) supported the hypothesis of long-distance migration of *P. tabacina* from the Caribbean region, Florida and Texas into other states in the United States. Within the European populations, the model documented migration from North Central Europe into western Europe and Lebanon, and migration from western Europe into Lebanon. The migration patterns observed support historical observations about the first disease introductions and movement in Europe. The models developed are applicable to other aerial dispersed emerging pathogens and document that high-evolutionary-risk plant pathogens can move over long distances to cause disease due to their large effective population size, population expansion and dispersal.

KEYWORDS
genetic structure, *Peronospora tabacina*, population genetics, tobacco

1 | INTRODUCTION

Peronospora tabacina Adam (syn. *P. hyoscyami* de Bary), causal agent of blue mold, is an oomycete pathogen and obligate parasite of tobacco (*Nicotiana* spp.) (Johnson, 1989; Lucas, 1980). *P. tabacina* can reproduce both asexually and sexually. Sporangia can be aerially transported by wind currents over long distances and can infect leaf surfaces (Davis & Monahan, 1991; Todd, 1981). Under cool and humid conditions, sporangiophores emerge through stomata and produce sporangia which are then released and reinfect other plants. The polycyclic pathogen can complete an infection cycle in 7–10 days, and thus, an outbreak can spread rapidly. The sexual oospores have been

rarely found, and thus, the asexual cycle and re-introduction of the pathogen each year are needed for an outbreak to occur (Heist, Nesmith, & Schardl, 2002).

Blue mold of tobacco (*Nicotiana tabacum* L.) was first reported in Australia in 1890 (Tyron, 1890) and is believed to be endemic to Australia, one of the countries of origin of the genus *Nicotiana* (Lucas, 1980). Evidence that blue mold was also endemic to the Americas, another centre of origin of *Nicotiana* spp., is also recorded. In the United States, blue mold was first observed on cultivated tobacco in 1921, followed by reports in Canada, Brazil and Argentina (1939), Chile (1953), Cuba (1957) and Mexico (1964). The disease was found in England on tobacco transplants imported from Australia in 1957, and rapidly spread into the Netherlands, Belgium and Germany by 1959 and into France, Italy, Yugoslavia, Poland, Spain, Morocco, Syria and Lebanon by the 1960s (Johnson, 1989; Wolf, Dixon, McLean, & Darkis, 1933). The pathogen is also a problem in South America (Johnson, 1989; Lea, 1999) and southwestern and southeastern Europe, the Middle East and North Africa (Delon & Schiltz, 1989).

Other important emerging plant pathogens that spread via aerially dispersed spores include soybean rust, cucurbit and grape downy mildews. Understanding the migration patterns of aerially dispersed plant pathogens can enable better means to manage diseases (Anderson et al., 2004). Several pathways of migration of *P. tabacina* may be operating in the United States and Europe. First, inoculum produced on tobacco in the Caribbean basin could be disseminated unidirectionally northwards to spring and summer-grown tobacco in the more temperate zones arriving anew each year from suspected source regions in the Caribbean or Central America (Davis & Main, 1986; Nesmith, 1984, 1990). Similarly in the Mediterranean region and northern Europe, putative sources of inoculum are from North Africa or the Middle East (Ledez, 1988; Main & Spurr, 1990).

Wild plant hosts can harbour plant pathogens that then jump host to domesticated crops. Inoculum of *P. tabacina* may survive on the wild fiddle leaf tobacco (*Nicotiana repanda* Willd.), an annual herb native to central and southern Texas and adjacent regions of Mexico (Reuveni, Nesmith, Siegel, & Keeny, 1988; Wolf, 1947). It has also been postulated using atmospheric trajectory analysis that Texas could be the source for blue mold spores in some years in North Carolina (Davis, Main, & Nesmith, 1990).

Movement of plant pathogens in planting material is a common means by which pathogens emerge in new geographic regions. A third possible source of inoculum could include survival of the pathogen as oospores on roots, stems or host debris in the field (Caiazzo, Tarantino, Porrone, & Lahoz, 2006; Heist et al., 2002; LaMondia & Aylor, 2001; Lucas, 1980; Spurr & Todd, 1982). The pathogen can also survive and spread in infected tobacco seedling beds or in transplants in greenhouses. As the symptoms and signs of the disease are not visible immediately after infection and systemic infections can occur, infected seedlings could also be an important inoculum source involved in spread of the pathogen between tobacco-growing regions.

Biosurveillance of plant pathogens can enable growers to make more targeted disease management decisions. The North American Blue Mold Warning System was initially established after the blue mold epidemics in 1979 and reported blue mold outbreaks in the United States until 2011 (Davis & Main, 1986; Ledez, 1988; Nesmith, 1984, 1990). The warning system was modelled after the European Blue Mold Warning Service developed by the Centre for Cooperation in Scientific Research Relative to Tobacco (CORESTA) for the Euro-Mediterranean zone in 1961 (Coresta, 2009; Ledez, 1988). Atmospheric satellite weather data are used to examine, integrate and predict the movement of the pathogen in space and time. However, validation of the migration patterns predicted by the forecasting system using genetic data from actual outbreaks has not been performed previously.

We examined the genetic structure of populations of *P. tabacina* to test two hypotheses that: (i) migration of the pathogen occurs from the Caribbean region into the United States by long-distance aerial dispersal and (ii) that migration of the pathogen occurs from the Mediterranean region into Europe by long-distance aerial dispersal. Specific objectives of this study were to analyse the population structure of *P. tabacina* from the United States, Caribbean, Mexico and Europe and the Mediterranean region using multilocus sequencing of one mitochondrial and two nuclear genes in order to (i) estimate genetic variability in pathogen populations and perform neutrality tests to assess departures from population-size constancy, (ii) estimate the magnitude and direction of migration and its contribution to the pathogen genetic structure and (iii) examine the population structure and evolutionary relationships of the haplotypes of *P. tabacina* between regions. Specifically, we predict higher genetic diversity at locations with longer established and/or larger effective population sizes, such as overwintering sites.

2 | MATERIALS AND METHODS

2.1 | Sample collection and procedures

Blue mold occurs annually in the eastern United States and Caribbean area on tobacco crops. We collect *Peronospora tabacina* annually and maintain in cryoculture more than 260 isolates in the Department of Entomology and Plant Pathology at North Carolina State University. Fifty-four isolates of *P. tabacina* sampled from different geographic locations from the United States, Central America, the Caribbean, Europe and the Middle East were selected (Table S1). Total genomic DNA was extracted using the standard cetyltrimethylammonium bromide (CTAB) protocol (Murray & Thompson, 1980). Each DNA extraction was checked for the presence of *P. tabacina* DNA using the specific PCR primer PTAB (Ristaino, Johnson, Blanco-Meneses, & Liu, 2007) and the real-time diagnostic primer PtabBM TaqMan (Blanco-Meneses & Ristaino, 2011). Initially, different primer pairs from nuclear, ribosomal and mitochondrial genes were tested and sequenced using 15 isolates of *P. tabacina* and one sequence of each gene were submitted to GenBank (Table S2). Three genes were chosen for subsequent analysis.

Polymorphisms were found in the intergenic spacer *Igs2* region of the ribosomal DNA (rDNA), the Ras-related protein (*Ypt1*) gene and the mitochondrial cytochrome c oxidase subunit 2 (*cox2*) gene.

These regions were chosen for subsequent amplification, cloning and sequencing. The presence of heterozygous sites was confirmed by cloning using the TA cloning kit PGEM-T easy vector (Promega, Madison, Wisconsin). Ten white colonies were selected, and the QIAprep Spin Miniprep Kit (QIAGEN, Valencia, California) was used for the purification of the plasmid DNA. DNA was stored at −20°C for further use and long-term storage. Six of ten clones of each isolate ($n = 54$) were randomly selected for sequencing. A total of 324 clones were analysed for each of the three gene regions. Each clone was sequenced using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit and analysed on an ABI PRISM 377 automated sequencer (Applied Biosystem, Foster City, California) at the North Carolina State University Genomic Sciences Laboratory.

2.2 | Data analysis

Multiple DNA sequences were aligned and edited manually using either BioEdit Sequence Alignment Editor Version 7.0.5.3 (Hall, 1999) or Clustal X (Thompson, Higgins, & Gibson, 1994). The total population was divided into four geographic subpopulations after analysis: (i) United States (isolates from U.S. pooled); (ii) CCAM (isolates from Mexico, Guatemala, Dominican Republic and Nicaragua); (iii) EULE (isolates from Bulgaria, France, Germany, Hungary, Lebanon and Poland) and (iv) United States vs. CCAM (clones from isolates from each U.S. state vs. clones from each isolate from CCAM countries). If overwintering occurs outside the United States and/or Europe, then directional gene flow and greater haplotype diversity at source sites should be observed.

Each gene or region was analysed separately for each subpopulation. Cloned sequences were formatted using the Phylogeny Inference Package (PHYLIP) format (Felsenstein, 1989). Initially, sequences were collapsed into haplotypes removing indels and excluding infinite-site violations using SNAP MAP (Aylor, Price, & Carbone, 2006; Price & Carbone, 2005) from SNAP WORKBENCH version 2.0 (Aylor et al., 2006; Monacell & Carbone, 2014; Price & Carbone, 2005). Incompatible sites were removed from the sequences using CladeEx (Bowden, Price, & Carbone, 2008) in the SNAP WORKBENCH platform. Then, the sequences were collapsed with phenotypes into haplotypes removing insertion and deletions (indels) and excluding infinite-site violations using the SNAP MAP (Aylor et al., 2006). Base substitutions were categorized as phylogenetically informative or uninformative and transitions vs. transversions and each of the polymorphic sites received a specific site number on the consensus sequence.

2.3 | Nucleotide diversity, neutrality tests and population subdivision

DNASP version 4.50.3 (Rozas, Sanchez-delBarrio, Messeguer, & Rozas, 2003) and the Neutrality Test Program (Li & Fu, 2002) were used to estimate nucleotide diversity and perform neutrality tests. Analysis was performed on the total population, separate regional subpopulations (U.S., CCAM and EULE subpopulations) and the individual states in the U.S. subpopulation. For each analysis, the population mutation parameter per nucleotide site θ using Watterson's (1975) θ_w, based on the number of segregating sites, s, and the average pairwise nucleotide diversity, π (Tajima, 1983), were estimated. Neutrality test statistics were calculated with Tajima's D (Tajima, 1989), Fu and Li's D^* and F^* (Fu & Li, 1993), and Fu's Fs (Fu, 1997) and tested using 5,000 simulations to accept or reject the hypothesis that mutations in these regions follow the neutral model of molecular evolution (Kimura, 1980). These neutrality tests assume a constant population size, no recombination and no migration and that each isolate is taken from a single randomly mating population and random sampling, that is, the population evolves according to the Wright–Fisher model. Departures of neutrality may be caused by evolutionary forces such as population subdivision, population shrinkage, overdominance selection, population growth, genetic hitchhiking and background selection (Fu, 1997). Carbone, Liu, Hillman, and Milgroom (2004) and Carbone, Jakobek, Ramirez-Prado, and Horn (2007) examined population growth and background selection by comparing Tajima's D, Fu and Li's D^* and F^* and Fu's Fs tests. If only Tajima's D, Fu and Li's D^* and F^* are significant and of Fu's Fs is nonsignificant, background selection is assumed (Fu, 1997). If only Fu's Fs is significant, there is evidence for population growth or genetic hitchhiking. Nonsignificant tests can be also informative; when Tajima's D and Fu's Fs tend to be uniformly negative, there is an excess of rare variants, and indication of genetic hitchhiking/selective sweep or population growth, and tend to be uniformly positive when there is an excess of intermediate-frequency alleles, indicating a model under population subdivision or balancing selection. The neutrality tests provide an inference on population processes and enable detection of deviations in population-size constancy that may result in departures of neutrality (Carbone et al., 2007).

Nonparametric tests for population subdivision were performed using Hudson's test, the nearest neighbour statistics S_{nn} (Hudson, Boos, & Kaplan, 1992) on the SNAP Workbench platform. This statistic uses the infinite-site model to examine how often the nearest sequence neighbours are from the same populations. The S_{nn} statistic allows pooling of different populations into genetically subdivided groups of regions. To calculate a significant value for the S_{nn} statistic, a permutation test of 1,000 pseudoreplicates was performed. When $S_{nn} = 1$, the two or more populations are highly genetically differentiated, whereas an S_{nn} near 0.5 indicates that the two or more populations are part of one panmictic population.

Hudson's tests were performed for populations from individual states in the United States and for populations from countries in the CCAM regions. The three gene regions were analysed separately, and then, populations were pooled when nonsignificant values occurred. The S_{nn} test was repeated a second and third time until it was impossible to pool the populations further. A second analysis was performed for the EULE subpopulation, and the same procedure for pooling was performed.

2.4 | Migration analysis

Once the putative boundaries of the subpopulations were determined, the observed patterns of genetic divergence were further

explored to determine whether there was evidence for constant migration among populations and shared ancestral polymorphism.

The migration with division (MDIV) programme was used to test for equilibrium migration vs. shared ancestral polymorphisms between two subdivided populations (Nielsen & Wakeley, 2001). This approach uses either an infinite-site or finite-site model without recombination and implements both likelihood and Bayesian methods using a Markov chain Monte Carlo coalescent simulation for jointly estimating the population mean mutation rate (θ), divergence time (T), migration rate (M) and time since the most recent common ancestor (TMRCA) between two subdivided populations. This approach assumes that populations are descended from one panmictic population that may (or may not) have subsequently exchanged migrants. We ran the MDIV analysis after first collapsing the data into haplotypes for the largest nonrecombining partition to avoid assumption violations.

If MDIV showed evidence of equilibrium migration, the isolation-with-migration (IM) programme (Hey & Nielsen, 2004; Nielsen & Wakeley, 2001) within the SNAP Workbench platform was used to estimate population divergence time, population mean mutation rate and direction of migration. Both MDIV and IM assume neutrality and no recombination. Under the null hypothesis of isolation, an expected migration near 0 would be expected. To run the IM programme, the following parameters were used: one million steps for the burn-in duration, 12 and 24 hr were used between outputs and a value of 20 was used for the maximum migration rate and the maximum time of population splitting. Multiple chains (total of six different runs for every two populations compared) were run using different starting values for the parameters. If all chains give similar results, it was assumed as good evidence of convergence for reliable parameter estimation.

2.5 | Phylogenetic and structure analysis

Maximum-likelihood (ML) phylogenetic analysis of the nuclear (*lgs2*, *ypt1*) and mitochondrial locus (*cox2*) sequences was performed using RAXML version 8, which is accessible through the CIPRES RESTful application (CRA) programmer interface (Miller et al., 2015). Phylogenies were displayed in circle tree format using T-BAS v2.0 (I. Carbone, unpublished), an extension of the T-BAS v1.0 toolkit (Carbone et al., 2017). Statistical support for tree branches was based on 1000 bootstrap replicates. Structure analysis and the optimal number of *K* clusters were performed using STRUCTURE version 2.3.1 (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000). To achieve good mixing and assess convergence of log-likelihood values, MCMC sampling used 20,000 iterations after a burn-in period of 20,000, and three simulations for *K* ranging from 1 to 6 were performed using the "full search" method in CLUMPP v1.1.2 (Jakobsson & Rosenberg, 2007). Two methods, LnP(D) and delta *K* (Evanno, Regnault, & Goudet, 2005) implemented in Structure Harvester v0.6.93 (Dent & von Holdt, 2012), were used to estimate the optimal number of *K* clusters. The results were visualized in T-BAS v2.0 as outer rings of a multilocus ML phylogeny. Principal component analysis (PCA) and the number of significant axes of variation (eigenvectors) were determined using the Tracy–Widom statistic (Tracy & Widom, 1994) with the number of *K* clusters estimated using the Gap Statistic (Tibshirani, Walther, & Hastie, 2001). Locus statistics and population statistics were performed at each geographic level using the R library poppr (Kamvar, Tabina, & Grünwald, 2014). Poppr was utilized to generate minimum-spanning networks (MSN) for each Multi Locus Genotype (MLG) at each spatial level.

2.6 | Validation of the North American blue mold warning

Isolates from the year 2006 were used to validate the trajectories generated by the forecast system for that year. The 2006 subpopulation was selected because we had the highest number of isolates collected from different geographical locations in the United States and CCAM regions in that year: North Carolina (NC) (n = 1), Kentucky (KY) (n = 2), Texas (TX) (n = 1), Dominican Republic (n = 3) and Nicaragua (n = 2) (Table S1). Population subdivision, gene flow and coalescent analysis were performed on the 2006 subpopulations as described above.

3 | RESULTS

3.1 | DNA sequence variability

A total of 954 clones from 54 blue mold samples were analysed using the *lgs2* region, *cox*2 gene and *Ypt1* region. A total of 435 nucleotides were sequenced for the *lgs*2 region, 493 nucleotides were sequenced for the *Ypt*1 gene and 631 nucleotides for *cox*2 gene. The numbers of segregating sites were 116 nucleotides for the *lgs*2 region (26.36%), 126 nucleotides for the *Ypt*1 region (25.40%) and 107 nucleotides for the *cox*2 gene (16.98%) (Table 1).

Summary statistics describing the nucleotide diversity in the pooled and individual populations for both nuclear and mitochondrial loci are presented in Table 1. Nucleotide diversity (π) estimates for the total population were 8.11×10^{-3} for the *lgs*2 gene, 10.28×10^{-3} for the *Ypt*1 region and 7.66×10^{-3} for the *cox*2 gene. The mean mutation parameters per site or Waterson's θ_w for the total population were similar and were 0.157, 0.159 and 0.157 for the *lgs2*, *Ypt*1 and *cox*2 genes.

Nucleotide diversity at the *lgs*2 region and the *Ypt*1 region were higher for the EULE (53.04×10^{-3} and 71.09×10^{-3}) and CCAM (42.4×10^{-3} and 64.17×10^{-3}) subpopulations than for the U.S. subpopulations (20.48×10^{-3} and 15.68×10^{-3}) (Table 1). However, nucleotide diversity for the mitochondrial *cox*2 gene was highest for the CCAM (31.63×10^{-3}) subpopulation. Similarly, θ_w was higher for the EULE and CCAM populations than for the U.S. populations for both nuclear regions. Waterson's θ_w for the *cox*2 gene was higher for CCAM than U.S. and EULE subpopulations.

Populations of *P. tabacina* from North Carolina burley and flue-cured tobacco were compared, and nucleotide diversity was higher in the NC-Burley for the *Ypt*1 region (103.88×10^3) but was higher for the NC-Flue-cured population for the *lgs*2 region (91.87×10^3) and the *cox*2 gene (1.09×10^{-3}). Little difference in θ_w was

TABLE 1 Population statistics, diversity estimates and neutrality tests based on variation in nuclear and mitochondrial regions of *Peronospora tabacina*

	Sample summaries					Parameter estimates		Test of neutrality statistic			
Locus/Population	*l*	*n*	*s*	*h*	*k*	π (*SE*) $\times 10^{-3}$	$\theta_w \times 10^{-3}$	Tajima's *D* statistic	Fu and Li's *D** statistic	Fu and Li's *F** statistic	Fu's *Fs* statistic
Nuclear (*Igs2* region)											
U.S.	435	180	68	51	1.393	20.48 (1.64 $\times 10^{-3}$)	173.39	−2.628 ***	−10.307 ***	−7.884 ***	−20.056 ***
CCAM	435	78	54	36	2.29	42.41 (3.97 $\times 10^{-3}$)	202.94	−2.503 ***	−8.323 ***	−6.749 ***	−21.279 ***
EULE	435	66	27	25	1.432	53.04 (5.82 $\times 10^{-3}$)	210.12	−2.273 ***	−5.550 ***	−4.872 ***	−21.239 ***
TOTAL	435	324	116	93	0.941	8.11 (7.00 $\times 10^{-4}$)	157.32	−2.788 ***	−12.642 ***	−8.925 ***	ns
NC-BURLEY	435	36	25	17	1.951	78.03 (1.27 $\times 10^{-2}$)	241.15	−2.214 **	−4.869 ***	−4.350 ***	−12.529 ***
NC-FLUE-CURED	432	36	16	11	1.47	91.87 (1.73 $\times 10^{-2}$)	241.15	−1.922 **	−4.340 ***	−3.881 ***	−5.392 *
(*Ypt* 1 region)											
U.S.	495	174	89	58	1.395	15.68 (1.74 $\times 10^{-3}$)	174.42	−2.760 ***	−9.191 ***	−7.165 ***	−20.234 ***
CCAM	493	66	34	28	2.182	64.17 (5.97 $\times 10^{-3}$)	210.12	−2.160 *	−5.385 ***	−4.681 ***	−20.192 ***
EULE	494	66	27	24	1.919	71.09 (7.21 $\times 10^{-3}$)	210.12	−2.012 *	−5.550 ***	−4.757 ***	−20.446 ***
TOTAL	493	306	126	93	1.295	10.28 (8.10 $\times 10^{-4}$)	158.75	−2.769***	−11.709 ***	−8.350 ***	−20.051 ***
NC-BURLEY	490	36	27	18	2.805	103.88 (1.37 $\times 10^{-2}$)	241.15	−1.877 **	−4.156 **	−3.704 **	−10.397 **
NC-FLUE-CURED	489	36	27	18	2.481	91.89 (1.52 $\times 10^{-2}$)	241.15	−2.041 **	−5.017 ***	−4.381 ***	−10.028 **
Mitochondrial (*cox* 2 gene)											
U.S.	626	180	65	46	0.744	1.25 (1.80 $\times 10^{-4}$)	18.88	−2.776 ***	−11.691 ***	−8.854 ***	−21.098 ***
CCAM	624	78	40	29	1.473	31.63 (6.82 $\times 10^{-3}$)	202.94	−2.534 ***	−4.453 ***	−4.205 ***	−21.368 ***
EULE	622	66	23	21	0.697	1.14 (2.20 $\times 10^{-4}$)	7.92	−2.554 ***	−6.943 ***	−5.993 ***	−20.799 ***
TOTAL	631	324	107	82	0.82	7.66 (1.02 $\times 10^{-3}$)	157.32	−2.787 ***	−12.295 ***	−8.775 ***	ns
NC-BURLEY	604	36	9	8	0.552	0.91 (2.9 $\times 10^{-4}$)	3.59	−2.099 **	−3.726 **	−3.520 ***	−6.436 **
NC-FLUE-CURED	609	36	12	11	0.667	1.09 (3.0 $\times 10^{-4}$)	4.75	−2.286 ***	−4.786 ***	−4.358 ***	−10.943 ***

U.S., United States; CCAM, Central America, Carribean and Mexican; EULE, Europe and Lebonon; *l*, # of nucleotide sites; *n*, sample size (number of clones); *s*, segregating nucleotide sites; *h*, haplotypes; *k*, average number of pairwise nucleotide differences; π, average number of base differences per site; SE, standard error; θ_w, population mean mutation rate per site or Watterson's theta estimator; ns, not significant; *.01 < *p* < .05; **.001 < *p* < .01; ***$p < .001$.

observed for NC-Burley and NC-Flue-cured subpopulations for any gene analysed (Table 1).

Nucleotide diversity (π) was also estimated for the U.S. populations separately. As the population sampled from North Carolina was larger than the populations from other states, we examined the effect of population size on the nucleotide diversity estimates (Table S3). Populations from North Carolina ($n = 72$) had lower nucleotide diversity estimates for the nuclear gene regions than populations from many other states. Waterson's θ_w for the North Carolina subpopulation was intermediate in value for the nuclear gene regions analysed but highest for the *cox*2 gene when compared with isolates from the other tobacco states (Table S3). Interestingly, the highest mitochondrial diversity occurred in Texas populations, a putative overwintering source in the United States (Table S3).

3.2 | Tests of neutrality

Four neutrality tests were performed to determine whether the data departed from an equilibrium model of neutral evolution. Highly significant and negative values were observed for all the neutrality tests for the three subpopulations indicating that the neutral model of evolution could be rejected (Table 1). For the total population, Fu's *Fs* was not significant for the *Igs2* and *cox2* genes indicating background selection. The negative values for all of the tests indicate an excess of recent mutations in the populations and population expansion or growth.

3.3 | Population genetic structure

There were a high number of low-frequency haplotypes in the populations. There were 93 haplotypes identified for the nuclear genes *Igs2* and *Ypt1*, and 82 haplotypes identified for the *cox*2 gene (Tables S4–S6).

For the *Igs2* region, one high-frequency haplotype (H90) dominated the populations. Ninety-three different haplotypes were identified of which 15 were frequent haplotypes (present in 2 or more clones) and 79 were less frequent haplotypes (presented in only 1 clone). The total *Igs2* haplotypes were differentiated by 116 single nucleotide polymorphism (SNPs), 104 transitions and 12 transversion substitutions (Table S4). Eight transition substitutions were

informative for the data analysis and were located at sites 9, 32, 59, 62, 71, 83, 108 and 112, and only one transversion substitution was informative and located at the 69 site of the consensus for the *Igs2* region (Table S4).

For the *Igs2* region, the predominant haplotype H90 was present in 209 of the 324 clones (64.51%). Greatest numbers of haplotypes (n = 24) were found in the North Carolina (22.58%) and Kentucky (n = 15, 11.83%) subpopulations; the other tobacco-producing states had fewer haplotypes (Table S4). Within the CCAM population, 14 of the haplotypes were present in Dominican Republic (15.05%) and 12 in Mexico (12.90%) (Table S7). Within EULE, 11 of the haplotypes (11.83%) were present in France while fewer than six haplotypes (1.08% and 6.45%) were found elsewhere in the EULE countries (Tables S7 and S8). NC had 15 unique haplotypes (18.99%), followed by 11 in KY (13.92%), 10 in Mexico (12.66%), nine in the Dominican Republic (11.39%) and eight in France (10.13%). The number of unique haplotypes for the other regions was lower than 4 (Tables S7 and S8).

For the *Ypt1* region, a total of 93 different haplotypes were found, 17 were frequent and 76 less frequent haplotypes (Table S5). Similarly, one high-frequency haplotype was predominant (H-1). The total haplotypes were differentiated by 126 SNPs, of which 107 were transitions and 19 were transversion substitutions. Eight transition substitutions were informative for the data analysis and were located on the 9, 25, 26, 57, 58, 72, 98 and 101, and only one transversion substitution was informative and located at the 96 site of the consensus for the *Ypt1* region (Tables S5 and S8).

The predominant haplotype named H1 was present in 176 of the 324 clones (54.32%) of the *Ypt1* region and was found in all of the subpopulations studied (Table S7). Within the U.S. population, 30 of the 93 haplotypes were present in NC (32.26%) and six were found in TX, PA, CT and GA (6.45% each). The other tobacco-producing states had between two and four (2.15%–4.30%) haplotypes (Table S7). Within the CCAM population, 16 of the haplotypes were present in Dominican Republic (17.20%) and six of the haplotypes were in Mexico (6.45%). Within EULE, six of the haplotypes were present in France, Bulgaria and Poland, respectively (6.45% each) (Tables S7 and S8). The highest number of unique haplotypes were found in NC (24 of 76, 31.58%), followed by Dominican Republic (13 of 76, 17.10%). The number of unique haplotypes for the other regions was low (0–5; 0%–6.58%) (Table S7).

The *cox2* gene sequence was the less variable when compared with the other two gene regions (Tables S7 and S8). A total of 82 haplotypes were found for this gene, eight were frequent and 74 less frequent haplotypes. Similarly, one high-frequency haplotype was predominant (H-1). The total haplotypes were characterized by 107 SNPs, of which 87 were transition and 20 were transversion substitutions. Six transition substitutions were informative for the data analysis and were located on the 16, 31, 33, 39, 46 and 107, and two transversion substitutions were informative and located at the 1 and 5 sites of the consensus for the *cox2* gene (Table S6).

For the *cox2* gene, the predominant haplotype, H1, was present in 235 of the 324 clones (75.53%) and it was present in every geographical location analysed (Table S6). Within the U.S. population, 16 of the 82 haplotypes were present in NC (19.51%), seven in PA and TX (8.54%) and six in KY (7.32%). The other tobacco-producing states had fewer haplotypes (6.10% and 12.19%) (Tables S6 and S7). Within the CCAM population, 15 of the haplotypes were present in Dominican Republic (17.20%), eight in Nicaragua (9.76%) and five in Mexico (6.10%). Within EULE, nine of the haplotypes were present in France (10.98%) and four in Germany (4.88%) (Table S7). The highest number of unique haplotypes for the *cox2* gene was found in North Carolina (15 of 74, 20.27%), followed by Dominican Republic (10 of 74, 13.51%) and France (7 of 74 haplotypes, 9.46%). The number of unique haplotypes for the other regions was low (between 0 and 6, 0%–8.10%) (Table S8).

3.4 | Population subdivision

Hudson's tests S_{nn} were performed to examine population structure and subdivision within and among populations (Hudson et al., 1992). Sequences from the United States and the CCAM subpopulations were analysed together, and sequences from Europe were analysed separately. Populations from Mexico, Guatemala, Nicaragua and the Dominican Republic in the CCAM region were not genetically subdivided from each other for the two nuclear and one mitochondrial region examined, and the sequences were pooled together for the CCAM region (Table 2, Figure 1). The Hudson's test was performed a second time, and populations from NC and KY, PA, MD and CT were not genetically subdivided, so they were pooled together (Table S10, Figure 1). The S_{nn} test was performed a third time, and the populations were classified into the North United States (NC, KY, GA, VA, PA, MD and CT) and South United States/CCAM (TX, FL and the CCAM) (Table S9, Figure 1). The S_{nn} test was performed a fourth time, and the North and South populations were genetically subdivided for the Ypt1/4 region and the *cox*2 gene but not for the *Igs*2 region (Table S11).

The EULE population was analysed separately using the Hudson's S_{nn} test. Initially, the population was pooled into four different groups: northern Europe (Poland and Germany), Central Europe (Bulgaria and Hungary), western Europe (France) and Lebanon (Table 3, Figure 1). The test was performed again and northern and Central Europe were not subdivided, so the data were pooled (Table S12, Figure 1). When the test was performed a third time, northern/Central Europe was genetically subdivided from western Europe for the *Igs*2 and the Ypt1/5 regions but not for the *cox*2 gene. Both populations were not genetically subdivided from Lebanon for all three gene regions (Tables S12 and S13).

3.5 | Phylogenetic and structure analysis

There were two large populations in the nuclear loci (*Ypt*1 and *Igs2*) (Figure 2). In the nuclear phylogeny, population 1 (predominantly blue) based on structure inference using the Evanno method at K = 2 (third outer ring) and overlapping predominantly with PCA cluster 2 (innermost ring) was represented by isolates in all regions

TABLE 2 The nearest neighbour statistic (S_{nn}) evaluated in the United States and CCAM subpopulations of *Peronospora tabacina* pathogen. The total population was used for the test

	Texas	Florida	Georgia	N. Carolina	Kentucky	Virginia	Maryland	Pennsylvania	Connecticut	Mexico	Guatemala	Nicaragua
U.S.–CCAM—Igs2												
Texas												
Florida	ns											
Georgia	0.015 *	ns										
North Carolina	ns	ns	ns									
Kentucky	0.021 *	ns	ns	ns								
Virginia	0.000 ***	ns	0.001 **	0.000 ***	0.001 **							
Maryland	ns	ns	ns	ns	ns	0.000 ***						
Pennsylvania	0.001 **	ns	ns	ns	ns	0.000 ***	ns					
Connecticut	ns	ns	ns	ns	ns	0.000 ***	ns	ns				
Mexico	ns	ns	ns	ns	ns	0.000 ***	ns	ns	ns			
Guatemala	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		
Nicaragua	ns	ns	ns	ns	ns	0.000 ***	ns	ns	ns	ns	ns	
Domican Rep.	ns	ns	ns	ns	ns	0.000 ***	ns	ns	ns	ns	ns	ns
U.S.–CCAM—Ypt 1												
Texas												
Florida	ns											
Georgia	ns	ns										
North Carolina	0.013 *	ns	ns									
Kentucky	0.004 **	ns	0.014 *	0.000 ***								
Virginia	ns	ns	ns	ns	0.015 *							
Maryland	ns	ns	0.007 **	0.026 *	0.042 *	ns						
Pennsylvania	0.004 **	ns	0.001 **	0.000 ***	0.002 **	0.004 ** I	0.021 *					
Connecticut	ns	ns	0.037 *	ns	0.038 *	ns	ns	0.007 **				
Mexico	ns	ns	ns	ns	0.001 **	ns	0.006 **	0.000 ***	0.028 *			
Guatemala	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		
Nicaragua	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	
Domican Rep.	ns	ns	ns	0.000 ***	ns	ns	ns	ns	ns	ns	ns	ns
U.S.–CCAM—cox 2												
Texas												
Florida	ns											
Georgia	ns	ns										
North Carolina	ns	ns	ns									

(Continues)

TABLE 2 (Continued)

	Texas	Florida	Georgia	N. Carolina	Kentucky	Virginia	Maryland	Pennsylvania	Connecticut	Mexico	Guatemala	Nicaragua
Kentucky	ns	ns	ns	ns								
Virginia	0.027 *	ns	0.000 ***	0.002 **	0.000 ***							
Maryland	ns	ns	ns	ns	ns	ns						
Pennsylvania	ns	ns	ns	0.000 ***	0.000 ***	ns	ns					
Connecticut	ns	ns	ns	ns	ns	0.020 *	ns	ns				
Mexico	0.001 **	ns	ns	ns	ns	0.000 ***	0.008 **	0.000 ***	ns			
Guatemala	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		
Nicaragua	ns	ns	ns	ns	ns	0.002 **	ns	ns	ns	ns	ns	
Domincan Rep.	ns	ns	ns	0.013 *	ns	0.032 *	ns	0.010 *	ns	ns	ns	ns

ns, not significant; *.01 < p < .05; **.001 < p < .01; ***p < .001.

except Nicaragua, Guatemala, Lebanon and Germany (second outer ring). Population 2 (orange histogram bars), which aligns predominantly with cluster 0 from PCA, had representatives across all regions except Bulgaria and Hungary (Figure 2). The fourth ring in the nuclear tree shows the structure LnP inference of $K = 4$. The nuclear structure analysis supports the strong population stratification between the South (e.g., Nicaragua, Guatemala) and North (e.g., NC, KY, VA) in the Americas and the high gene flow between North Europe (e.g., Germany) and Middle East (Lebanon). For *cox2*, the phylogeny was poorly resolved and is consistent with one large population corresponding to PCA cluster 4 (Figure S1). For the *cox2* mitochondrial locus, the Evanno estimate was $K = 4$ (third ring) and $K = 6$ (outermost ring) based on LnP.

A frequent predominant haplotype that was shared among United States, CCAM and EULE populations was observed in the minimum-spanning networks. There were many rare emergent haplotypes derived from the dominant haplotypes in each region for each locus examined (Figures S3–S5 and Tables S4–S6).

3.6 | Migration analysis

The MDIV programme estimates the population mean mutation rate (θ), divergence time (T), migration rate (M) and time since the most recent common ancestor (TMRCA) between two subdivided populations (Figure 1 and Figure S1). The direction of migration for the populations was examined using the isolation-with-migration programme (Hey & Nielsen, 2004). When the U.S./CCAM populations were compared, migration from the South United States/CCAM to the North United States was supported by data from the two nuclear gene regions and the mitochondrial gene. The migration rates from one population to another were very low. These results were supported by the two nuclear region, *Igs*2 (North $m_1 = 0.67$, South $m_2 = 1.08$) and *Ypt*1 (North $m_1 = 0.26$, South $m_2 = 0.27$), and the mitochondrial region (North $m_1 = 0.90$, South $m_2 = 1.02$) (Figure 1).

The three subdivided geographic regions were analysed in the EULE population: NCE, WE and LE. Migration rates were higher between populations in these regions than between populations in the United States. When direction of migration was tested between NCE and WE, migration from the NCE population ($m_1 = 7.29$) to the WE population ($m_2 = 6.94$) was identified. When direction of migration was tested between NCE and LE, migration was identified from NCE ($m_1 = 10.89$) to LE ($m_2 = 7.16$), and finally, when WE and LE populations were analysed, migration from WE ($m_1 = 9.09$) into LE ($m_2 = 6.66$) was identified. These results indicated that there is migration occurring from the Northern part of Europe (NCE) into France (WE) and Lebanon, and from France into Lebanon as well (Figure 1).

4 | DISCUSSION

We analysed the genetic structure of populations of *Peronospora tabacina* from a global collection of isolates from 11 countries in

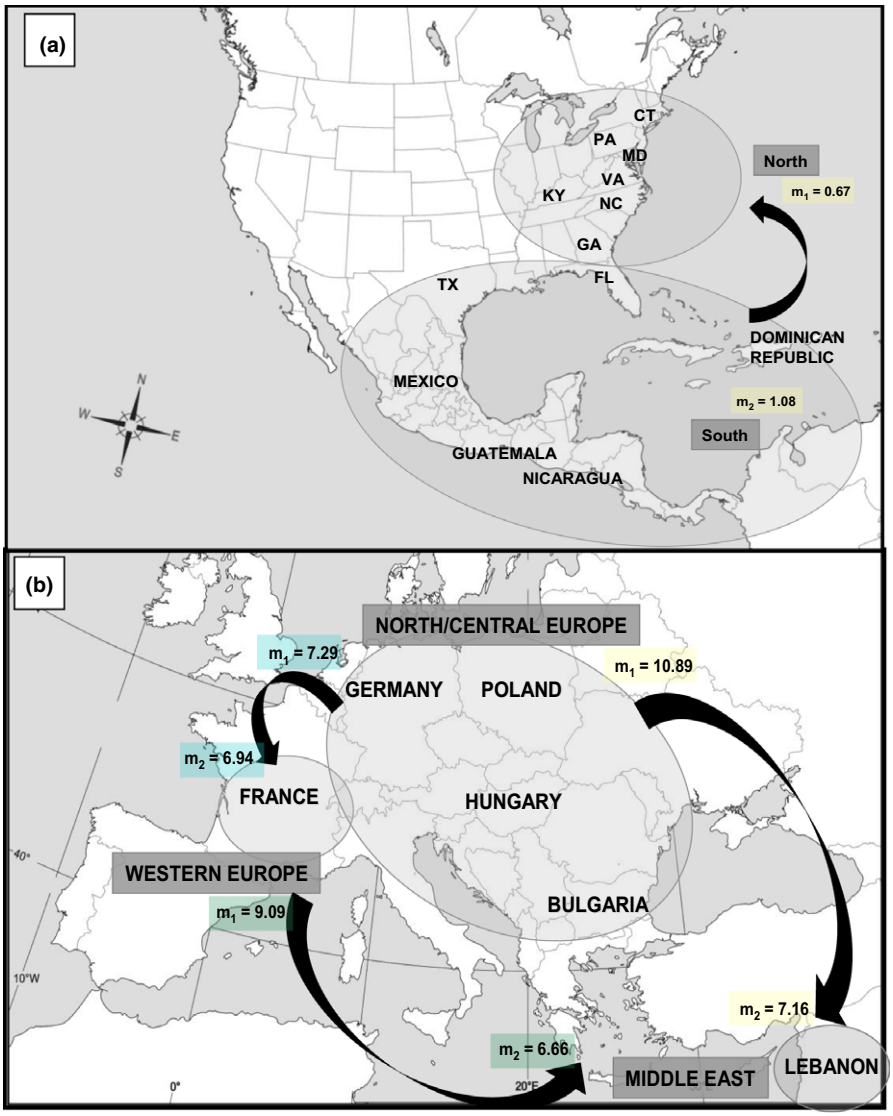

FIGURE 1 Map of migration scenarios. Isolation with migration (IM) programme was used to evaluate the (a). United States/CCAM and (b). EULE subpopulation of *Peronospora tabacina* pathogen. Rates of migration are shown in the figures for the *Igs*2 regions, but the direction of migration was supported by the two nuclear regions and the mitochondrial gene

three geographic regions including the United States, Central America, the Caribbean and Mexico (CCAM) and Europe and Lebanon (EULE). Nucleotide sequences from nuclear, ribosomal and mitochondrial genes showed high genetic variability across all populations and the dominance of a single high-frequency haplotype and many less frequent rare haplotypes.

Several previous studies have examined the genetic structure of populations of *P. tabacina*. Trigiano et al., 2012 found 10 polymorphic microsatellite loci using 44 isolates of the pathogen and demonstrated some genetic differentiation between North and South American isolates and isolates of European and Middle Eastern origin. Sukno, Taylor, & Farman, 2002 studied 35 isolates from nine tobacco states in the United States and 10 haplotypes were differentiated by RFLPs. Seven haplotypes were found in Kentucky while other United States regions had fewer haplotypes. Edreva, Delon, & Coussirat, 1998 used three isoenzymes and found little variation in a collection of isolates of *P. tabacina* from France and Bulgaria. Variation in allozyme alleles was only observed when isolates collected from *N. tabacum* were passed to *N. repanda*.

High levels of genetic variation were found among populations of *P. tabacina* when DNA from allelic regions of several loci was cloned and sequenced in our work. Other downy mildews such as cucurbit downy mildew also exhibit high levels of genetic diversity in the United States in Florida and North Carolina (Choi, Thines, Han, & Shin, 2011; Ojiambo, Gent, Quesada-Ocampo, Hausbeck, & Holmes, 2015). Several possible mechanisms could explain the high genetic variation observed in the blue mold pathogen. The most common form of dispersal of this pathogen is aerial, long-distance dispersal from putative overwintering sites in Texas and the Caribbean. Long-distance dispersal is an important survival strategy for many fungal pathogens including rusts, powdery mildews and downy mildews, enabling them to colonize new territory rapidly or to migrate between summer and winter habitats (Brown & Hovmoller, 2002; Ojiambo et al., 2015). The production of huge numbers of spores in each polycyclic event is essential for reproduction and survival because the pathogen is completely dependent on living host tissue. *P. tabacina* causes a polycyclic disease and aerial dispersal over long distances, and expansive population growth occurs after

TABLE 3 The nearest neighbour statistic (S_{nn}) evaluated on the EULE subpopulation of *Peronospora tabacina* pathogen. All geographic subpopulations used for the test

	Hungary	Germany	Poland	Bulgaria	Lebanon
EULE- Igs 2					
Hungary					
Germany	0.000 ***				
Poland	0.000 ***	ns			
Bulgaria	ns	ns	0.000 ***		
Lebanon	ns	0.045 *	0.000 ***	ns	
France	0.002 **	0.043 *	0.016 *	0.054 *	0.024 *
EULE—Ypt 1					
Hungary					
Germany	0.003 **				
Poland	0.001 **	ns			
Bulgaria	0.027 *	0.017 *	0.001 **		
Lebanon	0.001 **	0.002 **	0.000 ***	ns	
France	0.012 *	ns	0.049 *	0.001 **	0.000 ***
EULE—cox 2					
Hungary					
Germany	ns				
Poland	ns	ns			
Bulgaria	ns	ns	ns		
Lebanon	ns	ns	0.053 *	ns	
France	ns	ns	ns	ns	ns

ns, not significant; *.01 < p < .05; **.001 < p < .01; ***p < .001.

infection. *P. tabacina* could be considered a high-evolutionary-risk plant pathogen due to its large effective population size and mechanism of dispersal (McDonald & Linde, 2002). Population size affects the probability that mutants will be present and can influence diversity of genes.

Population expansion, recombination, migration and mutation have influenced the population structure of *P. tabacina*. Neutrality tests were significant for all populations and the equilibrium model of neutral evolution was rejected, indicating an excess of recent mutations or rare alleles. Tests of neutrality including Tajima's *D*, Fu and Li's D^*, F^* and Fu's *Fs* showed significant and negative values for the subpopulations. The significant values indicated a departure from neutrality, assuming that the population does not have a constant population size and recombination and migration could be occurring. There was evidence for migration in the populations we examined. In addition, the negative values indicate an excess of recent mutations in the population. It is interesting to note that the Fu's *Fs* statistic test was not significant for the analysis of the total population for the *Igs*2 region and *cox*2 gene. This could indicate a process of background selection. In cases where there is strong negative selection on a locus combined with a high mutation rate, the frequent purging of deleterious variants may result in the occasional removal of linked variation, producing a decrease in the level of variation surrounding the locus under selection. The result is the accidental purging of nondeleterious alleles due to such spatial proximity to deleterious alleles (Fu, 1997). A significant departure from neutrality could also indicate population expansion. The data from the minimum-spanning networks indicated the presence of a dominant haplotype and many more rare haplotypes, a signature of population expansion.

4.1 | Migrations from Central America and Caribbean to North America

Our data indicate that gene flow is occurring between Central American and Caribbean regions and migration is occurring to the United States. CCAM could serve as a reservoir for inoculum as the region has the appropriate weather for the development of the pathogen and tobacco is grown there year-round. Tobacco grown in Florida and Texas could also serve as reservoirs for pathogen inoculum. The data presented here confirm previous data from forecasting models and support the hypothesis of long-distance dispersal of this pathogen from the Caribbean region, Florida and Texas into the tobacco-growing states further north (Aylor & Taylor, 1983; Davis & Main, 1986). For many years, the Tobacco Blue Mold Forecasting System at N.C. State University has predicted the trajectory of tobacco blue mold epidemics in tobacco-growing regions in the United States, based on weather data and reports of source populations in the Caribbean (Main, Keever, Holmes, & Davis, 2001). Our data provide the first genetic evidence to support the long-distance migration theory of *P. tabacina*. We used Hudson's statistics to examine population subdivision and gene flow within CCAM. Subpopulations within this region were not subdivided, and gene flow was documented. Interestingly, putative overwintering subpopulations in Texas and Florida were not subdivided from subpopulations in CCAM and were pooled into a South population. The other U.S. subpopulations were genetically subdivided from CCAM subpopulations and divided into two populations: north and south. An isolation-with-migration analysis (IM) indicated migration from south to north supporting the hypothesis of long-distance migration of *P. tabacina* from the Caribbean regions, Florida and Texas into other states further north. PCA also indicated the south and north divide. These data also support previous reports of disease occurrence first in the CCAM region (including Cuba) and later as the green tobacco belt is planted further north (Davis & Main, 1986; Nesmith, 1984). We were unable to collect and genotype isolates from Cuba due to a U.S. imposed trade embargo with Cuba that prohibited exchange of plant samples. Thus, Cuban samples were not included among isolates from the CCAM region.

Our data demonstrated greater genetic diversity in the *Cox2* mitochondrial gene in populations from Texas, a putative overwintering site of the pathogen. These results agree with previous reports that inoculum of *P. tabacina* is known to survive on wild *Nicotiana* repanda in Texas (Reuveni et al., 1988) and overwinter and form oospores in Florida (McGrath & Miller, 1958). Both these sources of inoculum may contribute to epidemics in regions further north. Our

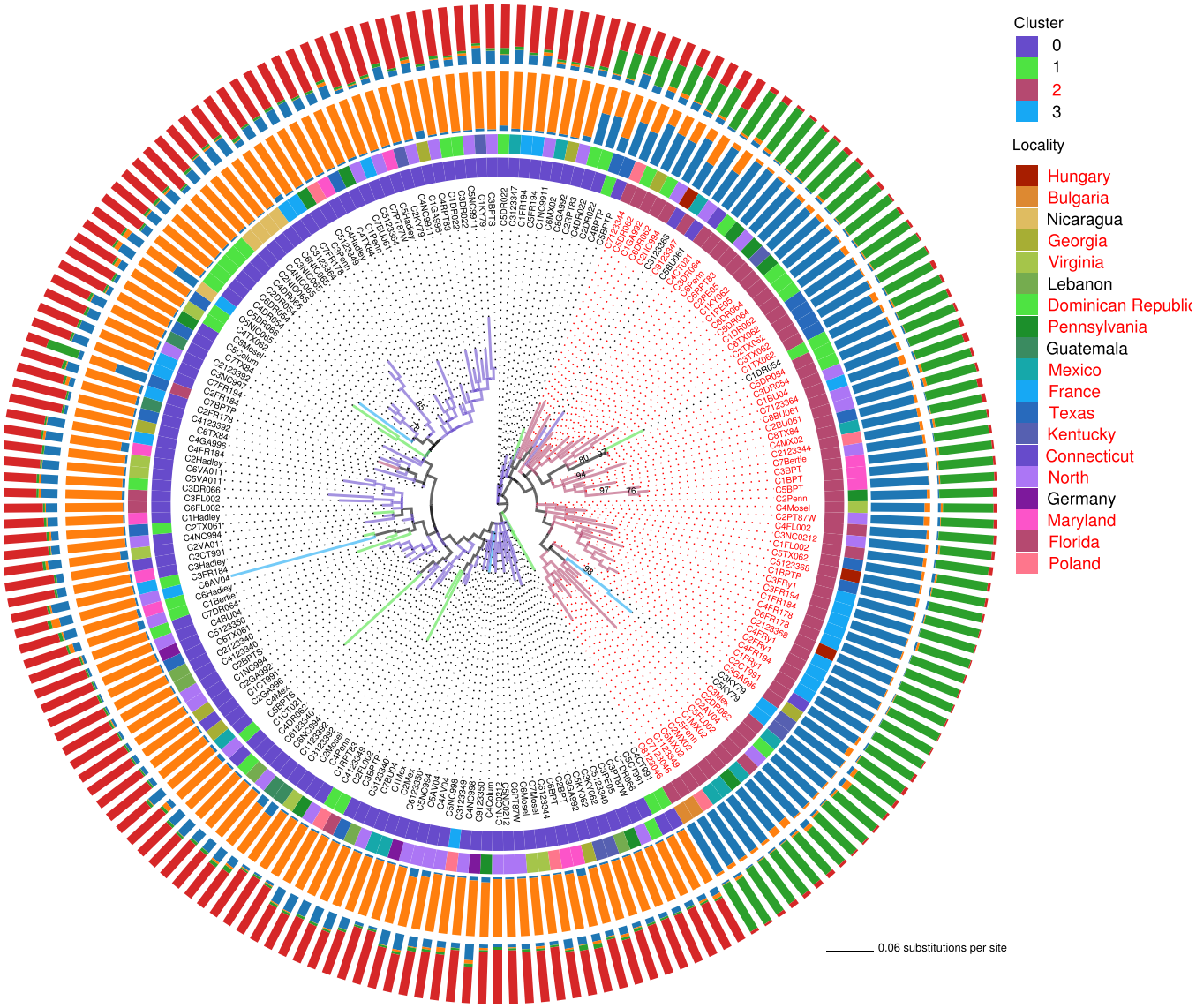

FIGURE 2 Maximum-likelihood (ML) phylogenetic analysis of the nuclear loci (*Ypt1* and *lgs2*) sequences from *Peronospora tabacina* was performed using RAxML version 8 (centre), principal component analysis (PCA) (innermost ring), geographic locality shown in the second ring, structure analysis using the Evanno method at $K = 2$ (third ring) and LnP inference at $K = 4$ (outermost ring). For each individual, the colours of histogram bars correspond to the relative proportion of each cluster. The tree branches are coloured according to PCA cluster, and bootstrap values >70 are labelled along branches

population statistics indicate that mitochondrial gene mutation rate and diversity were higher in Texas then elsewhere supporting Texas as a putative overwintering source of the pathogen.

Nucleotide diversity (π) and Watterson's θw were also higher for the CCAM subpopulation for all the gene regions examined. Subpopulations in the Dominican Republic for the CCAM and North Carolina from the United States had the highest population mean mutation rate (Watterson's estimate), the highest total number of haplotypes and the highest number of unique haplotypes among all populations examined. The CCAM region is clearly a centre of diversity for *P. tabacina* and provides the source of inoculum for blue mold epidemics into the United States.

High levels of genetic diversity have been reported in other oomycete pathogens such as *Peronospora effusa* (Choi et al., 2011), *Peronospora farinosa* (Danielsen & Lubeck, 2010) and *Phytophthora* spp. (Chamnanpunt, Shan, & Tyler, 2001; Judelson, 2007), and this diversity has been attributed to factors such as genetic instability, transposable elements, gene conversion, mitotic recombination and/or dispensable chromosomes (Haas et al., 2009; Kamoun, 2003). The *P. tabacina* genome has been recently sequenced, and the genome size and organization suggest that like other oomycetes, the genome has a large number of retrotransposons (Derevnina et al., 2015). SNP analysis revealed that 32% of the genome was heterozygous and that outcrossing has occurred. As nucleotide diversity is the average number of pairwise nucleotide differences between sequences in a sample and the Watterson's θ_w is based on the number of polymorphic sites in a sample, π depends on both the number of polymorphic sites and their

frequency, whereas θ_w is independent of frequencies (Nei & Li, 1979; Watterson, 1975). A population that has a high number of haplotypes could inflate the estimate for the θ_w, as was observed in our data.

4.2 | European migrations

High nucleotide diversity and similar mutation rate estimates were observed for CCAM and EULE populations when compared with the U.S. subpopulations for the nuclear gene regions examined. Nucleotide diversity of the mitochondrial gene for the EULE subpopulations was lower than CCAM estimates. Hudson's statistics separated the EULE subpopulation into three genetically subdivided populations including North Central Europe, western Europe and Lebanon. The isolation and migration model clearly documented gene flow between the regions, and the direction of migration was determined. When migration analysis was performed, migration rates were high and direction of migrations was established from North Central Europe to France and from North Central Europe to Lebanon. Directional migration from France to Lebanon was also observed.

The first introductions of this pathogen into Europe occurred in the UK in 1958 and continental Europe in 1959 from infected tobacco from Australia, the putative centre origin of the pathogen. In continental Europe, migration occurred first to the Netherlands and Germany and later to the east to France and then Middle East and North Africa. Our model of migration supports the historic disease reports (LaMondia & Aylor, 2001; Ledez, 1988). The oldest isolates studied in our work were from Germany and Lebanon in 1963 which is close to the time of the original introductions into Europe. Transport of inoculum by winds was believed to be involved in movement of the pathogen in these first European epidemics. Further analysis using more European samples collected more recently from other countries could provide better detail on the current population structure within European populations of *P. tabacina* and the direction of migration.

4.3 | Mechanisms of genetic variation in coenocytic oomycetes

Peronospora tabacina is coenocytic and lacks crosswalls; therefore, it is a multinucleate organism. *P. tabacina*'s sporangiospores have from 4 to 35 nuclei per sporangium (Trigiano & Spurr, 1987; Trigiano, Van Dyke, Spurr, & Main, 1985). The high amount of spores produced in each yearly epidemic plus the multikaryon characteristic of *P. tabacina* increases the possibility of high amounts of variation in subsequent populations.

Peronospora tabacina reproduces predominately by asexual means. The mechanism by which nuclei migrate in the coenocytic hyphae or the level of variation among nuclei is unknown. It is possible that nuclei may separate in mycelium and subsequently recombine to form divergent lineages. It is possible that in one asexual event, many or all of the nuclei pass to the subsequent generation. It is also possible that developing spores receive one nucleus that then divides to form all the other nuclei. This phenomenon has been observed in mycorrhizal fungi (Kuhn, Hijri, & Sanders, 2001; Sanders, 2002) that also show high levels of genetic variation. In some mycorrhizal species (*Glomus* spp.), it is believed that recombination can occur between different nuclei that coexist in one single spore (Bever, Kang, Kaonongbua, & Wang, 2008; Pawlowska & Taylor, 2004; Schardl & Craven, 2003). Intraspore genetic diversity has been reported in mycorrhizal spores. Sanders, Alt, Groppe, Boller, and Wiemken (1995) first showed that mycorrhizal spores contain variable ribosomal DNA (rDNA) sequences. This means that the heterokaryotic organisms have single spores containing more than one genome that coexist together (Sanders, 2002). For *P. tabacina*, the presence of genetic variability was observed in different clones from one single isolate. Further analysis of nuclei separately to determine whether intraspore genetic diversity occurs in *P. tabacina* is needed.

Another explanation for the genetic variation observed could involve nonmeiotic recombination of genetic material between cohabitating nuclei within heterokaryons (i.e., parasexuality) (Leslie, 1993). Chromosomal exchange is due to mitotic crossing over and takes place during mitotic divisions. The resultant nucleus, because of crossing over and random loss of chromosomes, is genetically different from the parents (Kamoun, 2003).

It has been reported that the genome of *P. tabacina* may be comprised of repetitive DNA (Derevnina et al., 2015; Sukno et al., 2002) as has been reported in other oomycetes including *Bremia lactuca* (Francis, Hulbert, & Michelmore, 1990) and *Phytophthora infestans* (Haas et al., 2009). The genome of *P. infestans* is greatly expanded with repetitive DNA. Rearrangements in the genome are associated with transposon activity which results in high levels of genomic variation (Haas et al., 2009). Retrotransposons have been reported in both *P. tabacina* and in a related *Hyaloperonospora* sp. (Coates & Beynon, 2010; Derevnina et al., 2015).

Another possible explanation for the high variability in *P. tabacina* is the presence of sexual recombination. The presence of oospores in this pathogen has been reported in vitro in plants after inoculation and incubation under laboratory conditions. In tobacco fields, the presence of oospores was reported in the United States at very low levels (Heist et al., 2002). Sexual reproduction is known to occur in populations from overwintering sites in Florida, Texas, and has been reported in Canada (LaMondia & Aylor, 2001; McGrath & Miller, 1958; Reuveni et al., 1988). The presence of sexual recombination in the overwintering sites and dispersal of diverse clonal lineages further north are the possibilities as a large number of rare alleles were found in the populations examined (McGrath & Miller, 1958; Reuveni et al., 1988; Spurr & Todd, 1982). The presence of incompatible sites would support conflicting phylogenies and provide evidence of recombination and genetic exchange. However, further work will be necessary to confirm this hypothesis. Sexual recombination has been reported in many other oomycetes including *Bremia lactuca* on wild Lettuce species (Lebeda & Blok, 1990), *Phytophthora infestans*, *P. capsici*,

P. sojae and *Plasmopara viticola* (Judelson, 2007). Indeed, evidence of outcrossing has been reported from analysis of the recently sequenced *P. tabacina* genome (*Derevnina* et al., 2015).

The nuclear regions and mitochondrial gene regions identified and sequenced in this study have provided valuable information on the genetic structure of populations of *P. tabacina*. Further studies using multiple single sequence repeats (SSRs) on a larger set of samples to compare U.S. populations to those in other nearby tobacco-growing regions are underway. It would be interesting to include Cuba in further sampling to determine whether the pathogen is introduced to the United States each year from there as has been suggested in the forecasting trajectories. The recent opening of diplomatic relations with Cuba may enable such studies to proceed. The data presented here have validated previous historic hypotheses about the long-distance migration of this pathogen into the United States and Europe and provide valuable tools for others to use in studies with downy mildews. This analysis has provided a basis for further studies to better understand the genetic processes that allow this pathogen to survive and colonize new areas. As the pathogen does not recognize international borders, understanding the population structure of *P. tabacina* will lead to improved strategies for global management of the disease.

ACKNOWLEDGEMENTS

We thank Tobacco Foundation of North Carolina and the Tobacco Commission of North Carolina for funding the first author's graduate assistantship to conduct the work. The development of T-BAS v2.0 is supported by the National Science Foundation (NSF) Genealogy of Life (GoLife) programme to IC (DEB-1541418). We are grateful to Carole Saravitz and the Phytotron Facility and the Genome Research Laboratory at North Carolina State University for space and sequencing. Tobacco producers and scientists in the United States, Central America, the Caribbean and Europe assisted with the collection of blue mold samples. We thank emeritus faculty Charlie Main, Harvey Spurr and Paul Shoemaker for providing historic collections and interesting discussions on the pathogen. We thank Kelly Ivors for serving on Monica Blanco's thesis committee. We thank Amanda Saville for conducting the minimum-spanning network analysis of the data and submitting the DNA sequences to GenBank.

AUTHOR CONTRIBUTION

M.B.M. conducted the research, analysed data and wrote paper. I.C. analysed data. J.B.R. conceived research, analysed data and wrote paper.

DATA ACCESSIBILITY

DNA sequences for the genes amplified were submitted to GenBank and accession numbers are shown in parentheses: intergenic spacer region 2 partial sequence (MF981863); Ras-related protein (ypt1) gene partial cds (MF981864); cytochrome oxidase subunit II (COX2) gene, partial cds; mitochondrial gene for mitochondrial product (MF981865); Beta-tubulin gene, partial cds (MF981866 and MF981867); large subunit ribosomal RNA gene, partial sequence (MF981868); 28S ribosomal RNA gene, partial sequence (MF981869); tRNA-Arg gene, partial sequence; NADH dehydrogenase subunit 1 (ND1) gene, partial cds (MF981870); NADH dehydrogenase subunit 4L (nad4L) gene, complete cds; and NADH dehydrogenase subunit 1 (nad1) gene, partial cds, mitochondrial genes encoding mitochondrial proteins, accession number (MG545997); ATP synthase F1 subunit 9 (atp9) and NADH dehydrogenase subunit 9 (nad9) genes, partial sequence; mitochondrial, accession number (M6759486); tRNA-Met gene, partial sequence; tRNA-Pro (trnP(ugg)) and tRNA-Met (trnM (cau)) genes, complete sequence; ribosomal protein L14 (rpl14) and ribosomal protein L5 (rpl5) genes, complete cds; and tRNA-Gly (trnG (gcc)) gene, partial sequence; mitochondrial (MF981871, MG545998); tRNA-Tyr (trnY) and small subunit ribosomal RNA (rns) genes, partial sequence; mitochondrial, partial cds (MF981872); cytochrome c oxidase subunit 2 (cox2) gene, partial cds; orf32 gene, complete cds; and cytochrome c oxidase subunit 1 (cox1) gene, partial cds; mitochondrial (MF981873).

ORCID

Jean B. Ristaino http://orcid.org/0000-0002-9458-0514

REFERENCES

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution*, *19*(10), 535–544. https://doi.org/10.1016/j.tree.2004.07.021

Aylor, D. L., Price, E. W., & Carbone, I. (2006). SNAP: Combine and Map modules for multilocus population genetic analysis. *Bioinformatics*, *22*, 1399–1401. https://doi.org/10.1093/bioinformatics/btl136

Aylor, D., & Taylor, G. (1983). Escape of *Peronospora tabacina* spores from a field of diseased tobacco plants. *Phytopathology*, *73*(4), 525–529. https://doi.org/10.1094/Phyto-73-525

Bever, J. D., Kang, H. J., Kaonongbua, W., & Wang, M. (2008). Genomic Organization and mechanisms of inheritance in arbuscular mycorrhizal fungi: Contrasting the evidence and implications of current theories. In A. Varma & B. Hock (Eds.), *Mycorrhiza* (pp. 135–148). Springer Berlin Heidelberg: New York, NY, USA. https://doi.org/10.1007/978-3-540-78826-3

Blanco-Meneses, M., & Ristaino, J. B. (2011). Detection and quantification of *Peronospora tabacina* using a real-time polymerase chain reaction assay. *Plant Disease*, *95*, 673–682. https://doi.org/10.1094/PDIS-05-10-0333

Bowden, L. C., Price, E. W., & Carbone, I. (2008). SNAP Clade and Matrix, Version 2. Distributed over the Internet. Retrieved from http://www.cals.ncsu.edu/plantpath/faculty/carbone/home.html.

Brown, J. K. M., & Hovmoller, M. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, *297*, 537–541. https://doi.org/10.1126/science.1072678

Caiazzo, R., Tarantino, P., Porrone, F., & Lahoz, E. (2006). Detection and early diagnosis of *Peronospora tabacina* Adam in tobacco plant with systemic infection. *Journal of Phytopathology*, *154*, 432–435. https://doi.org/10.1111/j.1439-0434.2006.01123.x

Carbone, I., Jakobek, J. L., Ramirez-Prado, J. H., & Horn, B. W. (2007). Recombination, balancing selection and adaptive evolution in the aflatoxin gene cluster of *Aspergillus parasiticus*. *Molecular Ecology*, *16*, 4401–4417. https://doi.org/10.1111/j.1365-294X.2007.03464.x

Carbone, I., Liu, Y., Hillman, B. I., & Milgroom, M. G. (2004). Recombination and migration of *Cryphonectria hypovirus 1* as inferred from gene genealogies and the coalescent. *Genetics*, *166*, 1611–1629. https://doi.org/10.1534/genetics.166.4.1611

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Kauff, F., U'Ren, J. M., May, G., & Lutzoni, F. (2017). T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads, and metadata visualization: An example with the Pezizomycotina tree of life. *Bioinformatics*, *33*, 1160–1168.

Chamnanpunt, J., Shan, W. X., & Tyler, B. M. (2001). High frequency mitotic gene conversion in genetic hybrids of the oomycete *Phytophthora sojae*. *Proceedings of the National Academy of Sciences*, *98*, 14530–14535. https://doi.org/10.1073/pnas.251464498

Choi, Y. J., Thines, M., Han, J. G., & Shin, H. D. (2011). Mitochondrial phylogeny reveals intraspecific variation in *Peronospora effusa*, the spinach downy mildew pathogen. *Journal of Microbiology*, *49*, 1039–1043. https://doi.org/10.1007/s12275-011-1069-2

Coates, M. E., & Beynon, J. M. (2010). *Hyaloperonospora arabidopsidis* as a pathogen model. *Annual Review of Phytopathology*, *48*, 329–345. https://doi.org/10.1146/annurev-phyto-080508-094422

CORESTA (2009). Euro-Mediterranean Blue Mould Information Service. Retrieved from: http://www.coresta.org/.

Danielsen, S., & Lubeck, M. (2010). Universally primed-PCR indicates geographical variation of *Peronospora farinose* ex. *Chenopodium quinoa*. *Journal of Basic Microbiology*, *50*, 104–109. https://doi.org/10.1002/jobm.200900019

Davis, J. M., & Main, C. E. (1986). Applying atmospheric trajectory analysis to problems in epidemiology. *Plant Disease*, *70*, 490–497. https://doi.org/10.1094/PD-70-490

Davis, J. M., Main, C. E., & Nesmith, W. C. (1990). Aerobiological aspects of the Kentucky blue mold epidemic of 1958. In W. E. McKeen (Ed.), *Blue mold of tobacco* (pp. 55–71). St. Paul, MN: American Phytopathological Society Press.

Davis, J. M., & Monahan, J. F. (1991). Climatology of air parcel trajectories related to the atmospheric transport of *Peronospora tabacina*. *Plant Disease*, *75*, 706–711. https://doi.org/10.1094/PD-75-0706

Delon, R., & Schiltz, P. (1989). Spread and control of blue mold in Europe, North Africa and the Middle East. In W. E. McKeen (Ed.), *Blue mold of tobacco* (pp. 19–42). St. Paul, MN: American Phytopathological Society Press.

Dent, E. A., & von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361.

Derevnina, L., Chin-Wo-Reyes, S., Martin, F., Wood, K., Froenicke, L., & Spring Michelmore, O. R. (2015). Genome sequence and architecture of the tobacco downy mildew pathogen *Peronospora tabacina*. *Molecular Plant-Microbe Interactions*, *28*, 1198–1215. https://doi.org/10.1094/MPMI-05-15-0112-R

Edreva, A., Delon, R., & Coussirat, J. C. (1998). Variability of *Peronospora tabacina* A. – and isoenzyme study. *Beiträge zur Tabakforschung International*, *18*(1), 3–13.

Evanno, G., Regnault, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure. A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*, 1567–1587.

Felsenstein, J. (1989). PHYLIP-Phylogeny Inference Package (version 3.2). *Cladistics*, *5*, 164–166.

Francis, D. M., Hulbert, S. H., & Michelmore, R. W. (1990). Genome size and complexity of the obligate fungal pathogen, *Bremia lactucae*. *Experimental Mycology*, *14*, 229–309.

Fu, Y. X. (1997). Statistical test of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*, 915–925.

Fu, Y. X., & Li, W. H. (1993). Statistical test of neutrality of mutations. *Genetics*, *133*, 693–709.

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. J., Handsaker, R. E., Cano, L. M., . . . Nusbaum, C. (2009). Genome sequence and comparative analysis of the Irish potato famine pathogen, *Phytophthora infestans*. *Nature*, *461*, 393–398.

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Heist, E. P., Nesmith, W. C., & Schardl, C. L. (2002). Interactions of *Peronospora tabacina* with roots of *Nicotiana* spp. in gnotobiotic association. *Phytopathology*, *92*, 400–405. https://doi.org/10.1094/PHYTO.2002.92.4.400

Hey, J., & Nielsen, R. (2004). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *PNAS*, *104*(8), 2785–2790.

Hudson, R. R., Boos, D. D., & Kaplan, N. L. (1992). A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, *9*, 138–151.

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

Johnson, G. I. (1989). *Peronospora hyoscyami* de Bary: Taxonomic history, strains, and host range. In W. E. McKeen (Ed.), *Blue mold of tobacco* (pp. 55–71). St. Paul, MN: American Phytopathological Society Press.

Judelson, H. S. (2007). Sexual reproduction in plant pathogenic oomycetes: Biology and impact on disease. In J. Heitman, et al., (Eds.), *Sex in fungi: Molecular determination and evolutionary implications* (pp. 445–458). Washington, D.C.: ASM Press.

Kamoun, S. (2003). Molecular genetics of pathogenic oomycetes. *Eukaryotic Cell*, *2*, 191–199. https://doi.org/10.1128/EC.2.2.191-199.2003

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

Kimura, M. (1980). A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*, 111–120. https://doi.org/10.1007/BF01731581

Kuhn, G., Hijri, M., & Sanders, I. R. (2001). Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature*, *414*, 745–748. https://doi.org/10.1038/414745a

LaMondia, J. A., & Aylor, D. E. (2001). Epidemiology and management of a periodically introduced pathogen. *Biological Invasions*, *3*, 273–282. https://doi.org/10.1023/A:1015273512111

Lea, H. W. (1999). Resistance of tobacco to pandemic blue mold (*Peronospora hyoscyami* de Bary syn *P. tabacina* Adam): A historical overview. *Australian Journal of Experimental Agriculture*, *39*, 115–118. https://doi.org/10.1071/EA97163

Lebeda, A., & Blok, I. (1990). Sexual compatibility types of *Bremia lactucae* isolates originating from *Lactuca serriola*. *Netherlands Journal of Plant Pathology*, *96*, 51–54. https://doi.org/10.1007/BF01976607

Ledez, P. (1988). The CORESTA tobacco blue mold warning service for the Euro-Mediterranean zone. In C. E. Main & H. W. Spurr (Eds.), *Blue mold disease of tobacco, symposium proceedings* (pp. 79–91). Charlotte, NC: Delmar Printing.

Leslie, J. F. (1993). Fungal vegetative compatibility. *Annual review of Phytopathology*, *31*, 127–150. https://doi.org/10.1146/annurev.py.31.090193.001015

Li, H., & Fu, Y. (2002). Neutrality test program. Release 1.1. Retrieved from http://hgc.sph.uth.tmc.edu/neutrality_test

Lucas, G. B. (1980). The war against blue mold. *Science*, *210*, 147–153. https://doi.org/10.1126/science.210.4466.147

Main, C. E., Keever, T., Holmes, G., & Davis, J. (2001). Forecasting long-range transport of downy mildew spores and plant disease epidemics. APS Net Feature, April-May 2001, Retrieved from https://www.apsnet.org/publications/apsnetfeatures/Pages/DownyMildew.aspx

Main, C. E., & Spurr, H. W. (1990). Blue mold – the problem and the phenomenon. In C. E. Main & H. W. Spurr (Eds.), *Blue mold disease of tobacco, symposium proceedings* (pp. ix–x). Charlotte, NC: Delmar Printing.

McDonald, B. A., & Linde, C. (2002). Pathogen Population Genetics, evolutionary potential, and durable resistance. *Annual review of Phytopathology*, *40*, 349–379. https://doi.org/10.1146/annurev.phyto.40.120501.101443

McGrath, H., & Miller, P. R. (1958). Blue mold of tobacco. *Plant Disease Report Supplement*, *250*, 1–35.

Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., … O'Leary, M. A. (2015). A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evolutionary Bioinformatics*, *11*, 43–48.

Monacell, J. T., & Carbone, I. (2014). Mobyle SNAP Workbench: A web-based analysis portal for population genetics and evolutionary genomics. *Bioinformatics*, *30*, 1488–1490. https://doi.org/10.1093/bioinformatics/btu055

Murray, H. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight DNA. *Nucleic Acids Research*, *8*, 4321–4325. https://doi.org/10.1093/nar/8.19.4321

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nesmith, W. C. (1984). The North American blue mold warning system. *Plant Disease*, *68*(11), 933–936. https://doi.org/10.1094/PD-68-933

Nesmith, W. C. (1990). The North American Blue Mold Warning System – an update. In C. E. Main & H. W. Spurr (Eds.), *Blue mold disease of tobacco, symposium proceedings* (pp. 73–78). Charlotte, NC: Delmar Printing.

Nielsen, R., & Wakeley, J. (2001). Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics*, *158*, 885–896.

Ojiambo, P. S., Gent, D. H., Quesada-Ocampo, L. M., Hausbeck, M. K., & Holmes, G. J. (2015). Epidemiology and population biology of *Pseudoperonospora cubensis*: A model system for management of downy mildews. *Annual review of Phytopathology*, *53*, 223–246. https://doi.org/10.1146/annurev-phyto-080614-120048

Pawlowska, T., & Taylor, J. (2004). Organization of genetic variation in individual of arbuscular mycorrhizal fungi. *Nature*, *427*, 733–737. https://doi.org/10.1038/nature02290

Price, E. W., & Carbone, I. (2005). SNAP: Workbench management tool for evolutionary population genetic analysis. *Bioinformatics*, *21*, 402–404. https://doi.org/10.1093/bioinformatics/bti003

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

Reuveni, M., Nesmith, W. C., Siegel, M. R., & Keeny, T. M. (1988). Virulence of an endemic isolate of *Peronospora tabacina* from Texas to *Nicotiana tabacum* and *N. repanda*. *Plant Disease*, *72*, 1024–1027. https://doi.org/10.1094/PD-72-1024

Ristaino, J. B., Johnson, A., Blanco-Meneses, M., & Liu, B. (2007). Identification of the tobacco blue mold pathogen, *Peronospora tabacina*, by polymerase chain reaction. *Plant Disease*, *91*, 685–691. https://doi.org/10.1094/PDIS-91-6-0685

Rozas, J., Sanchez-delBarrio, J. C., Messeguer, X., & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, *19*, 2496–2497. https://doi.org/10.1093/bioinformatics/btg359

Sanders, I. R. (2002). Ecology and evolution of multigenomic arbuscular mycorrhizal fungi. *The American Naturalist*, *160*, S128–S141. https://doi.org/10.1086/342085

Sanders, I. R., Alt, M., Groppe, K., Boller, T., & Wiemken, A. (1995). Identification of ribosomal DNA polymorphism among and within spores of the Glomales: Application to studies on the genetic diversity of arbuscular mycorrhizal fungal communities. *New Phytologist*, *130*, 419–427. https://doi.org/10.1111/j.1469-8137.1995.tb01836.x

Schardl, C. L., & Craven, K. D. (2003). Interspecific hybridization in plant-associated fungi and oomycetes: A review. *Molecular Ecology*, *12*, 2861–2873. https://doi.org/10.1046/j.1365-294X.2003.01965.x

Spurr, H. W., & Todd, F. A. (1982). Oospores in blue mold diseased North Carolina Burley and Flue-cured tobacco. *Tobacco Science*, *26*, 44–46.

Sukno, S. A., Taylor, A. M., & Farman, M. L. (2002). Genetic uniformity among isolates of *Peronospora tabacina*, the tobacco blue mold pathogen. *Phytopathology*, *92*, 1236–1244. https://doi.org/10.1094/PHYTO.2002.92.11.1236

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*, 437–460.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*, 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Royal Statistical Society*, *63*, 411–423. https://doi.org/10.1111/1467-9868.00293

Todd, F. A. (1981). The blue mold story. p. 9–25. In: *Report: 29th Tobacco Workers Conference*. Lexington, KY.

Tracy, C., & Widom, H. (1994). Level-spacing distribution and the Airy kernel. *Communications in Mathematical Physics*, *159*, 151–174. https://doi.org/10.1007/BF02100489

Trigiano, R. N., & Spurr, H. W. (1987). The development of the multinucleate condition of *Peronospora tabacina* sporangia. *Mycologia*, *79*(3), 353–357. https://doi.org/10.2307/3807455

Trigiano, R. N., Van Dyke, C. G., Spurr, H. W., & Main, C. E. (1985). Ultrastructure of sporangiophore and sporangium ontogeny of *Peronospora tabacina*. *Tobacco Science*, *29*, 116–121.

Trigiano, R. N., Wadl, P. A., Dean, D., Johnson, D., Scheffler, B. E., Runge, F., … Spring, O. (2012). Twelve polymorphic microsatellite loci for *Peronospora tabacina*, the cause of tobacco blue mold. *Mycologia*, *104*, 633–640. https://doi.org/10.3852/11-288

Tyron, H. (1890). Tobacco disease, p. 29. In: *The Queenslander*, Queensland Newspapers, Brisbane.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*, 256–276. https://doi.org/10.1016/0040-5809(75)90020-9

Wolf, F. A. (1947). Tobacco downy mildew, endemic to Texas and Mexico. *Phytopathology*, *37*, 721–729.

Wolf, F. A., Dixon, L. F., McLean, R., & Darkis, F. R. (1933). Downy mildew of tobacco. *Plant Disease Report*, *17*, 37–39.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Blanco-Meneses M, Carbone I, Ristaino JB. Population structure and migration of the Tobacco Blue Mold Pathogen, *Peronospora tabacina*, into North America and Europe. *Mol Ecol*. 2018;00:1–15. https://doi.org/10.1111/mec.14453