Genomic Analyses of Dominant U.S. Clonal Lineages of *Phytophthora infestans* Reveals a Shared Common Ancestry for Clonal Lineages US11 and US18 and a Lack of Recently Shared Ancestry Among All Other U.S. Lineages

B. J. Knaus, J. F. Tabima, C. E. Davis, H. S. Judelson, and N. J. Grünwald

First and fifth authors: Horticultural Crop Research Unit, United States Department of Agriculture–Agricultural Research Service, Corvallis, OR 97330; second and fifth authors: Department of Botany and Plant Pathology, Oregon State University, Corvallis 97331; and third and fourth authors: Department of Plant Pathology, University of California, Riverside 92521.
Accepted for publication 14 June 2016.

ABSTRACT

Knaus, B. J., Tabima, J. F., Davis, C. E., Judelson, H. S., and Grünwald, N. J. 2016. Genomic analyses of dominant U.S. clonal lineages of *Phytophthora infestans* reveals a shared common ancestry for clonal lineages US11 and US18 and a lack of recently shared ancestry among all other U.S. lineages. Phytopathology 106:1393-1403.

Populations of the potato and tomato late-blight pathogen *Phytophthora infestans* are well known for emerging as novel clonal lineages. These successions of dominant clones have historically been named US1 through US24, in order of appearance, since their first characterization using molecular markers. Hypothetically, these lineages can emerge through divergence from other U.S. lineages, recombination among lineages, or as novel, independent lineages originating outside the United States. We tested for the presence of phylogenetic relationships among U.S. lineages using a population of 31 whole-genome sequences, including dominant U.S. clonal lineages as well as available samples from global populations. We analyzed ancestry of the whole mitochondrial genome and samples of nuclear loci, including supercontigs 1.1 and 1.5 as well as several previously characterized coding regions. We found support for a shared ancestry among lineages US11 and US18 from the mitochondrial genome as well as from one nuclear haplotype on each supercontig analyzed. The other nuclear haplotype from each sample assorted independently, indicating an independent ancestry. We found no support for emergence of any other of the U.S. lineages from a common ancestor shared with the other U.S. lineages. Each of the U.S. clonal lineages fit a model where populations of new clonal lineages emerge via migration from a source population that is sexual in nature and potentially located in central Mexico or elsewhere. This work provides novel insights into patterns of emergence of clonal lineages in plant pathogen genomes.

Many plant pathogens are characterized as having a clonal or partially clonal lifestyle. The advent of molecular markers has facilitated the inference of clonality in populations (Milgroom 1996). Genotyping of emerging or reemerging pathogen populations that cause epidemics provides the ability to determine whether newly genotyped populations share ancestry with those characterized previously and whether populations contain the same clonal lineages (Grünwald and Goss 2011). Such comparisons are facilitated by a molecular marker system and nomenclature that refers to the genotypes revealed by that marker system (Fry et al. 1993; Grünwald et al. 2009). When a comparison matches a characterized strain, phenotypic behavior of the newly discovered strain may be assessed. For example, when a newly isolated, clonal strain matches an existing strain that has resistance to a fungicide, it may be assumed that this new strain is likely to have resistance as well (Saville et al. 2015), although these correlations are not absolute and mutations can occur within a clone that might affect phenotype. Clonality also allows development of lineage-specific markers that can facilitate more rapid identification (Gagnon et al. 2014).

Often, the assumption is made that strains sharing the same multilocus genotype and, thus, sharing identity by state (IBS) across all loci also share identity by descent (IBD) (Carbone and Kohn 2004; Grünwald and Goss 2011). However, IBS may not reflect IBD because lineages with the same genotype might have arisen from different ancestors. For example, presence of a shared allele might have occurred via homoplasy. Conversely, when strains do not share IBS, they might still share ancestry because one or more mutations might have accumulated in loci in another line. Here, we reassess the concept of clonal lineage in the potato and tomato late-blight pathogen *Phytophthora infestans*, considered to be one of the top 10 oomycete pathogens worldwide (Kamoun et al. 2015), to determine whether IBD is the case among U.S. lineages of *P. infestans*.

The concept of clonal lineages in *P. infestans* has a long history in plant pathology. A clonal lineage, by definition, consists of individuals sharing IBD but showing some accumulated mutations within lineages as lineages diverge in space and time, typically leading to gradual clonal divergence. The restriction fragment length polymorphism probe RG57 (Goodwin et al. 1992) was the first molecular marker system employed to diagnose lineages of *P. infestans*. Subsequently, other markers, such as the allozymes *glucose-6-phosphate isomerase* and *peptidase*, as well as mating type, were employed to diagnose the U.S. lineages (Goodwin et al. 1995). As new marker systems such as mitochondrial haplotypes were developed (Griffith and Shaw 1998), they were also used to help assign individuals to lineages and identify new lineages. This has resulted in a recipe of molecular markers implemented to diagnose lineages of *P. infestans* used to date (Adler et al. 2004; Alkher et al. 2015; Cooke and Lees 2004; Drenth et al. 1993; Fry and Goodwin 1997a,b; Fry et al. 1993, 2013; Hu et al. 2012; Li et al. 2013; Petchaboon et al. 2014; Peters et al. 2014; Wangsomboondee

et al. 2002). A similar approach has been used to characterize the population behavior of other clonal *Phytophthora* pathogens such as *P. ramorum* (Kamvar et al. 2015).

The United States has witnessed the emergence of a number of lineages of *P. infestans* over the past few decades. Prior to the early 1990s, populations of *P. infestans* throughout most of the world were considered to be a single clonal lineage, US1 (Goodwin et al. 1994). During the 1990s, the United States witnessed a reemergence of late blight (Fry and Goodwin 1997a,b). Among the lineages described during this period were lineages US8 and US11, both resistant to the fungicide metalaxyl. More recently, these lineages have been largely replaced by new lineages (Fry et al. 2013; Hu et al. 2012), including US22, US23, and US24, which are susceptible to metalaxyl. The latter two lineages remain abundant today. An understanding of the evolutionary relationships among these lineages could provide an understanding of how these epidemics emerge. At this point, it is not clear whether a given clonal lineage arose from another clonal lineage by mutation, recombination, or other genetic mechanisms (Fig. 1B, C, and D) (Fry et al. 2013; Martin et al. 2014; Yoshida et al. 2013) or independently from a source population, possibly outside the United States, via migration (Fig. 1A).

An evolutionary perspective can address many of the challenges associated with the identification of clonal lineages. From an evolutionary perspective, a clonal lineage is a nonidentical population of closely related lineages sharing IBD. Here, we examine whether clones of the late blight pathogen *P. infestans* arose from a single immigrant, shared ancestor (Fig. 1B, C, or D) or are independent migrants from a source population (Fig. 1A). We sampled published whole-genome sequences and resequenced additional strains of *P. infestans* using Illumina technology to infer whether strains belonging to different clonal lineages observed in the United States share IBD by sampling different genetic loci and regions in these genomes.

MATERIALS AND METHODS

Analysis of microsatellites. Microsatellite data were obtained from http://Phytophthora-id.org/ (Grünwald et al. 2011) and, therefore, contain samples different from the mitochondrial and genomic datasets presented below. Microsatellite analysis based on protocols developed by Li et al. (2013) was used to ensure that the strains we used for whole-genome sequencing were, indeed, placed into the respective clonal lineages and could reproduce relationships shown by Fry et al. (2013). Isolates belonging to the predominant U.S. clonal lineages US8, US11, and US22 to -24 were included with a reference panel available at http://Phytophthora-id.org/ (Grünwald et al. 2011). Simple-sequence repeat data were analyzed using the neighbor-joining algorithm based on Bruvo's distance and 1,000 bootstrap replicates (Kamvar et al. 2014). The resulting tree was plotted in R (R Core Team 2015) using ape (Paradis et al. 2004).

Analysis of the mitochondrial genome. Samples from previously published sources as well as novel sequencing were used in the analysis of the mitochondrion (Table 1). Characterization of the 12 new samples is presented in Table 1. Sequence reads from several sequencing technologies were mapped to the *P. infestans* IIa

mitochondrial genome (Avila-Adame et al. 2006; Broad Institute 2014) (note that analyses used the version from the Broad Institute) using bwa mem 0.7.10 (Li 2013; Li and Durbin 2009). The aligned SAM files were converted to BAM format, sorted, had mate coordinates and related flags filled (for paired-end reads), and indexed with SAMtools (Li et al. 2009). To improve the alignment of the reads, the GATK's RealignerTargetCreator and Indel-Realigner were used (DePristo et al. 2011; McKenna et al. 2010; Van der Auwera et al. 2013). The reads from isolates P10127, P10650, P11633, P12204, P1362, P6096, and P7722 were determined by GATK to have unusually high-quality scores, indicating a legacy encoding. These samples were rerun using the GATK's '–fix_misencoded_quality_scores' option. After realignment, the MD tags were created with SAMtools calmd (Li et al. 2009). Variants were called using the GATK's HaplotypeCaller, setting the ploidy level to one. Variants were filtered for quality by selecting only variants with a cumulative coverage between 1×10^3 and 6×10^3 and a mapping quality between 59.8 and 60.2. A 2-kbp indel differentiates the type I and type II mitochondrial haplotypes (Avila-Adame et al. 2006). This indel includes Orf99 through Orf198 or positions 4,400 to 6,650 in the type IIa sequence obtained from the Broad Institute. This region was excluded from analyses using the vcfR package (Knaus and Grunwald in press) and custom scripts in the R language (R Core Team 2015). Inference of phylogeny was performed with RAxML (Stamatakis 2014) using the GTRCAT model and 1,000 bootstrap replicates. The resulting tree was plotted in R (R Core Team 2015) using ape (Paradis et al. 2004).

Analysis of nuclear loci. Two large supercontigs, supercontig_1.1 (6,928,287 bp) and supercontig_1.5 (4,768,765 bp), were analyzed similarly to that of the mitochondrial DNA, except that reads were mapped to the T30-4 reference (Haas et al. 2009). To handle alignments around indels, a second round of bam improvement was performed. After genotype calling, the genotypes were phased using beagle4 (Browning and Browning 2007). Beagle4 also imputed genotypes but, due to a lack of family structure in our dataset, we were concerned that this may not have been effective. To mitigate our concerns, imputed genotypes were rescored as missing data (using output from the GATK haplotype caller). In order to avoid the inclusion of variants with anomalous coverages (e.g., no data or excessively high coverage), a 90% confidence interval for variant coverage was made by using the 5th and 95th quantiles of variant sequence depth for each sample. This confidence interval was then used to score variants with anomalous coverage as missing data. To focus variant calling on the highest-quality portion of each supercontig, variants were only retained when they were within the coordinates of core orthologous genes. Core orthologous genes were determined as genes that were conserved in *P. infestans*, *P. sojae*, and *P. ramorum* as being orthologous and occurring as a single-copy locus (Haas et al. 2009). Because of this conservation, we assume that these loci represent a high-quality fraction of the genome. The resulting sequences were used to infer phylogeny using the likelihood method of RAxML (Stamatakis 2014). Because we used the phased haplotypes, each sample was
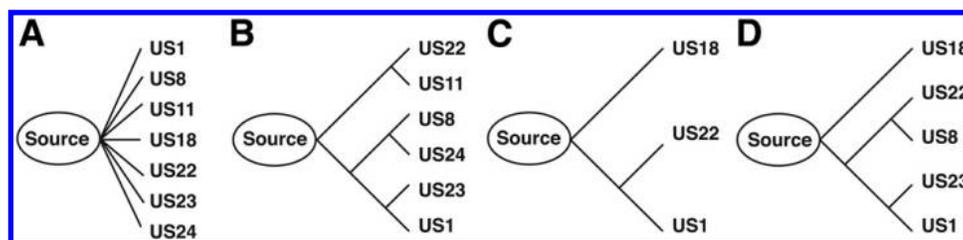

Fig. 1. Cartoon representation of several hypothetical evolutionary scenarios for U.S. lineages of *Phytophthora infestans*. Clonal lineages of *P. infestans* may have emerged as **A**, independent evolutionary lineages that do not share ancestry with any other lineage from one or more source populations. Alternatively, certain clonal lineages may have arisen from existing lineages, as presented by previous reports using **B**, microsatellite data (Fry et al. 2013) or mitochondrial genome analysis: **C**, Yoshida et al. 2013 and **D**, Martin et al. 2014.

represented twice in the analysis, once for each haplotype, based on the assumption that *P. infestans* is diploid. The resulting trees were plotted in R (R Core Team 2015) using ape (Paradis et al. 2004). In order to explore heterozygosity among the samples, the number of heterozygous positions per sample for the core orthologous genes on supercontigs 1 and 5 were calculated and violin plots were created in ggplot2 (Wickham 2009).

To allow for comparison with previous work based on single genes, we extracted several genes from our above workflow for single-gene analyses. This is intended to provide a perspective on how single-gene analyses may compare with inferences based on supercontigs. The genes β-tubulin (PITG_00156) and PITG_11126 were used for comparison with previous work (Blair et al. 2008; Gómez-Alpizar et al. 2007; Goss et al. 2011, 2014; Kroon et al. 2004; Oliva et al. 2010). We also selected two core orthologs from supercontig_1.1 and supercontig_1.5, each which were close to 1,000 bp (i.e., could theoretically be polymerase chain reaction amplified and spanned by two Sanger sequencing reactions) and had a number of segregating sites that were representative of the core orthologs. This search resulted in the selection of genes encoding NEP1 (PITG_00142) and trafficking protein particle complex subunit 6B (PITG_00953) from supercontig_1.1, and SNARE (PITG_04424) and Serin S09X (PITG_04066) from supercontig_1.5 (Table 2). Variant information for these loci was extracted from the genomic data and sequences were constructed from them (no Sanger sequencing was performed for data presented here). These genes were analyzed using a Bayesian coalescent approach to reconstruct phylogenetic relationships using BEAST 2.3.0 (Bouckaert et al. 2014). To set the priors for the analysis in BEAST, we obtained the most likely models of nucleotide substitution for each alignment using jModelTest 2.1 (Darriba et al. 2012). BEAST analyses were conducted with independent Markov chain Monte Carlo (MCMC) simulations of 10 million iterations for each locus. For each MCMC run, we sampled every 1,000 generations

and discarded nonstationary samples after a burn-in of 2.5 million iterations. Effective sample size estimates were typically greater than 200 and parameter trace plots supported good MCMC mixing at each locus. Maximum clade credibility trees were obtained using TreeAnnotator in the BEAST suite using common ancestor heights. The final tree was visualized in FigTree 1.3.1 (Rambaut and Drummond 2009).

In order to provide an analysis that is complementary to the inference of phylogeny, unsupervised clustering was performed on the variants of supercontigs 1.1 and 1.5 used for phylogeny. The software fastSTRUCTURE (Raj et al. 2014) uses plink format files as input. The VCF format files were converted to plink format using VCFtools version 0.1.12b (Danecek et al. 2011). The *.bim and *.bam format files were created using plink v1.07 (Purcell et al. 2007). The software fastSTRUCTURE was used to infer the probability of group membership for each sample, where the number of groups ranged from two to five. The fastSTRUCTURE script structure.py was used to infer probability of group membership and the script chooseK.py was used to infer the optimal number of clusters. The results were plotted in R (R Core Team 2015).

Estimation of divergence times for US1 and US22. In order to estimate the divergence times of the lineages US1 and US22 from their respective ancestors, the mean number of substitutions per site was extracted from the likelihood tree inferred for supercontig_1.1. The number of nucleotides contained in the core orthologs for supercontig_1.1 was then multiplied by the mean number of substitutions for each branch to estimate the number of substitutions which define each branch. This value represents the number of substitutions that mark the collective divergence of each lineage from their ancestor in the sample. In order to convert this number to a rate, the number of years since the Great Famine (170 years) and the time since displacement of US1 in North America (25 years) were used.

TABLE 1. Whole genomes of *Phytophthora infestans* strains obtained for the analyses of the mitochondrion, supercontig_1.1 and supercontig_1.5, and loci presented in Table 2

Strain	Genome source	Provenance	Year isolated[a]	Sequencing technology[b]
T30-4	Broad Institute	The Netherlands	1995	se36, se45, 454
90128	Raffaele et al. (2010)	The Netherlands	–	se36
PIC99189	Raffaele et al. (2010)	Mexico	NA	se36
NL07434	Yoshida et al. (2013)	The Netherlands	2007	pe76
P12204	Yoshida et al. (2013)	Scotland	1996	pe101
P13527	Yoshida et al. (2013)	Ecuador	2002	pe76
P13626	Yoshida et al. (2013)	Ecuador	2003	pe76
13_A2 (blue13)	Cooke et al. (2012)	England	2006	pe76
Pic97136	Goss et al. (2014)	Mexico	1997	pe101
Pic97146	Goss et al. (2014)	Mexico	1997	pe101
Pic97335	Goss et al. (2014)	Mexico	1997	pe101
Pic97442	Goss et al. (2014)	Mexico	1997	pe101
Pic97750	Flier et al. (2001)	Mexico	1997	pe101
Pic97785	Goss et al. (2014)	Mexico	1997	pe101
DDR7602 (US1)	Yoshida et al. (2013)	Germany	1976	pe76
LBUS5 (US1)	Yoshida et al. (2013)	South Africa	2005	pe76
P11633	Yoshida et al. (2013)	Hungary	2005	pe101
P6096 (US1)	Yoshida et al. (2013)	Peru	1984	pe101
P1362 (US1)	Yoshida et al. (2013)	Mexico	1979	pe101
P10650	Yoshida et al. (2013)	Mexico	2004	pe101
US040009 (US8)	This study	New York	–	pe51
RS2009P1_US8	Martin et al. (2013)	Pennsylvania	2009	pe100
FP-GCC (US11)	This study	New York	–	pe51
P10127 (US18)	Yoshida et al. (2013)	North Carolina	2002	pe101
1306a	This study	California	–	pe51
US100006 (US22)	This study	Kentucky	–	pe51
P17777 (US22)	Yoshida et al. (2013)	New York	2009	pe76
IN2009T1_US22	Martin et al. (2013)	Pennsylvania	2009	pe100
FL2009P4 (US23)	This study	Kentucky	–	pe51
BL2009P4_US23	Martin et al. (2013)	Pennsylvania	2009	pe100
ND822Pi (US24)	This study	North Dakota	–	pe51

[a] Symbol − = data not available and NA indicates not applicable.
[b] Sequencing technology consisted of single-end 36-bp (se36), single-end 45-bp (se45), paired-end 51-bp (pe51), paired-end 76-bp (pe76), paired-end 100-bp (pe100), or paired-end 101-bp Illumina sequence reads or Roche 454 reads (454).

RESULTS

To infer relationships among clonal lineages, we studied the genetic relationships of U.S. lineages and a global reference population using microsatellites, the mitochondrial genome, and single-nucleotide polymorphisms observed in core orthologs at the nuclear supercontigs 1.1 and 1.5 and at six individual genes.

Analysis of microsatellites. The microsatellite analysis supported clustering of *P. infestans* lineages US11, US22, and US23 with at least 90% bootstrap support and US24 with at least 70% bootstrap support (Fig. 2), confirming prior microsatellite work and the relationships shown in Figure 1B (Fry et al. 2013). A notable exception from the support of U.S. lineages is that the strains included for lineage US8, represented by five samples, did not cluster at a significant level and appeared in two clusters (Fig. 2). Although strains belonging to the U.S. lineages typically clustered together, there is no support for relationships among these lineages. Genetic variation within clonal lineages is also observed, reinforcing the concept that clonal lineages are not strictly identical and undergo clonal divergence.

Analysis of the mitochondrial genome. Inference of phylogenetic placements using the whole mitochondrial genome, excluding indels, based on observed single nucleotide polymorphisms is shown in Figure 3. Notable results from this analysis were that many U.S. lineages do not form a well-supported clade. Isolates from Germany (DDR7602), South Africa (LBUS5), Mexico (P1362), and Peru (P6096) are considered to be US1 (Kamoun et al. 1998; Yoshida et al. 2013) and formed a supported clade with an isolate from Hungary (P11633). The samples from lineages US8 and US22 formed a significantly supported clade (bootstrap support greater than 90%), which suggests a relationship among these lineages. The sample USA 1306a formed a supported clade with lineages US11 and US18. Sister to this clade was a well-supported clade of isolates from Ecuador and The Netherlands.

Analyses of the nuclear genome. *Phytophthora* core orthologous genes from supercontigs_1.1 and 1.5 were filtered for loci determined to be of typical sequence coverage. Summaries of the breadth of coverage and mean coverage are presented in Supplementary Figs. S1 and S2. Inference of phylogeny from phased haplotypes of supercontig_1.1 and supercontig_1.5 resulted in trees with 62 tips, showing two haplotypes per sample (Fig. 4; Table 1). Of the 1,043 predicted genes on supercontig_1.1 (Haas et al. 2009), 571 were core orthologs, with at least one segregating site. Of the 373 predicted genes on supercontig_1.5, 108 were core orthologs, with at least one segregating site. Core orthologs from supercontigs 1 and 5 tended to be approximately 1 to 2 kbp in length (Fig. 5A and B) but also included genes that were several kilobase pairs in length. These genes contained approximately 20 segregating sites each (Fig. 5C and D), with nucleotide diversities of approximately 0.001 (Fig. 5E and F).

Phylogenetic inference of the core orthologs of supercontig_1.1 resulted in several well-supported clades (Fig. 4A). The lineage US1 was represented by samples from Germany (DDR7602), South Africa (LBUS5), Hungary (P11633), Peru (P6096), and Mexico (P1362). Both haplotypes of this lineage formed a clade. Both haplotypes from samples of FP-GCC (US11) and P10127 (US18) and an isolate from California (1306a) formed a well-supported clade; however, the relationship among the two haplotypes within each strain was not supported. Clonal lineage US22 formed a well-supported clade including both haplotypes from all three samples. The lineages US8, US23, and US24 did not form well-supported clades and were, instead, distributed among clades with low support.

Phylogenetic inference of the core orthologs of supercontig_1.5 resulted in subtle differences from inferences based on supercontig_1.1 (Fig. 4B). Haplotypes from lineages US1 formed a clade. However, instead of forming one well-supported clade, the only well-supported relationship was among the haplotypes from Germany (DDR7602) and South Africa (LBUS5), which formed two separate clades, one for each haplotype. The remaining US1 samples clustered with low support. In contrast to the analysis of supercontig_1.1, samples from US11 and US18 show one haplotype as forming a well-supported clade while the samples from the other haplotype appeared unrelated. Finally, isolates from clonal lineage US22 formed supported clades for each haplotype, and these two clades were supported with bootstrap support greater than 70%.

Unsupervised clustering of core orthologous genes from supercontigs 1.1 and 1.5 resulted in support for the lineage US1 belonging to a group separate from the rest of the sample (Supplementary Figs. S9 and S10). Clustering based on core orthologous genes from supercontig_1.1 resulted in samples of US1 being different from all other strains for all numbers of groupings tested. Although the optimal number of groups was two, when the number of groups was specified as five, the samples from US22 formed a group. This may be due to this lineage's relatively high representation ($n = 3$) in the data set. Analysis based on supercontig_1.5 similarly supported the samples of US1 forming a group. The model complexity that maximized the marginal likelihood was two and the model components used to explain structure in the data were four, as opposed to a value of two for both derived from the analysis of supercontig_1.1. At values of K = 3 though K = 5, samples of US22 formed a group. At groupings of K = 4 and K = 5, US23 clustered, at least in part, with samples from Europe (T30-4, 90128, NL07434, P12204, and 13_A2) as well as with samples from Mexico (PIC99189 and PIC97335) and Ecuador (P13527 and P13626).

Analyses of six representative genes failed to consistently reveal a pattern of kinship among the U.S. lineages of *P. infestans*. Locus β-tubulin (PITG_00156) (Supplementary Fig. S3) resulted in a large, well-supported clade containing many of the strains. However, there was no supported topology within this large clade. Three other branches were well supported, one including one of the haplotypes for US22, but the other two were not congruent with other analyses. Analysis of a conserved hypothetical protein (PITG_11126) (Supplementary Fig. S4) resulted in two well-supported clades containing most individuals, with no support for the internal

TABLE 2. Summary statistics for genic regions selected for phylogenetic analysis[a]

Locus	Locus tag	References	Chromosome	Length	Segregating sites	Nucleotide diversity
β-Tubulin	PITG_00156	Goss et al. (2014)	Supercontig1.1	1,748	13	1.25E-04
Conserved hypothetical protein	PITG_11126	Goss et al. (2014)	Supercontig1.19	2,584	48	1.95E-05
NEP1	PITG_00142	This study	Supercontig1.1	871	14	1.24E-03
Trafficking protein particle complex subunit 6B	PITG_00953	This study	Supercontig1.1	692	13	2.20E-03
SNARE	PITG_04424	This study	Supercontig1.5	967	10	1.03E-03
Serin S09X	PITG_04066	This study	Supercontig1.5	919	14	1.47E-03

[a] Gene predictions were obtained from the annotations provided at the Broad Institute. Other commonly used genic regions found in the literature were identified as being multicopy genes and, hence, were not used in our analyses and included elongation factor 1α, enolase, heat shock protein 90, tigA gene fusion protein, and actin.

structure. One of the haplotypes of US1 formed a clade, as did one of the haplotypes of US23. Locus NEP1 (PITG_00142) (Supplementary Fig. S5) resulted in a well-supported clade for one of the haplotypes of US23 as well as one of the haplotypes of US1 but little else of relevance to the U.S. lineages. Analysis of trafficking protein particle complex subunit 6B (PITG_00953) (Supplementary Fig. S6) similarly resulted in a well-supported clade for one of the haplotypes of US1. However, it also included a haplotype from sample 90128, a finding not supported in other analyses. The locus SNARE (PITG_04424) (Supplementary Fig. S7) resulted in support for a clade consisting of one of the two haplotypes in the US1 clade found in other analyses but also included haplotypes from Ecuador

(P13527_1 and P13626_1) as well as Scotland (P12204_0), a relationship not inferred in other analyses. One of the US23 haplotypes also formed a supported clade. Much of the remainder of the sample was unresolved. Analysis of Serin S09X (PITG_04066) (Supplementary Fig. S8) resulted in a supported clade for one of the US1 haplotypes but also included samples from Great Britain (13_A2) and The Netherlands (NL07434_1), a relation not inferred in other analyses. In general, analyses of these candidate loci resulted in poorly resolved trees that lacked support for clades found in the nuclear, mitochondrial, and microsatellite analyses. Thus, these genetic loci did not provide further insights into the complex history of U.S. clonal lineages.

Comparison among analyses. As presented in detail above, our inferences of the evolutionary relationships among the U.S. lineages of *P. infestans* based on various genomic regions and loci, including microsatellites, the mitochondrial genome, and nuclear regions (including supercontigs 1.1 and 1.5), as well as six genetic loci, resulted in clades that were not consistent among different datasets. Results that were consistent among the analyses included a US1 clade, a US11 and US18 clade, and a US22 clade. None of the
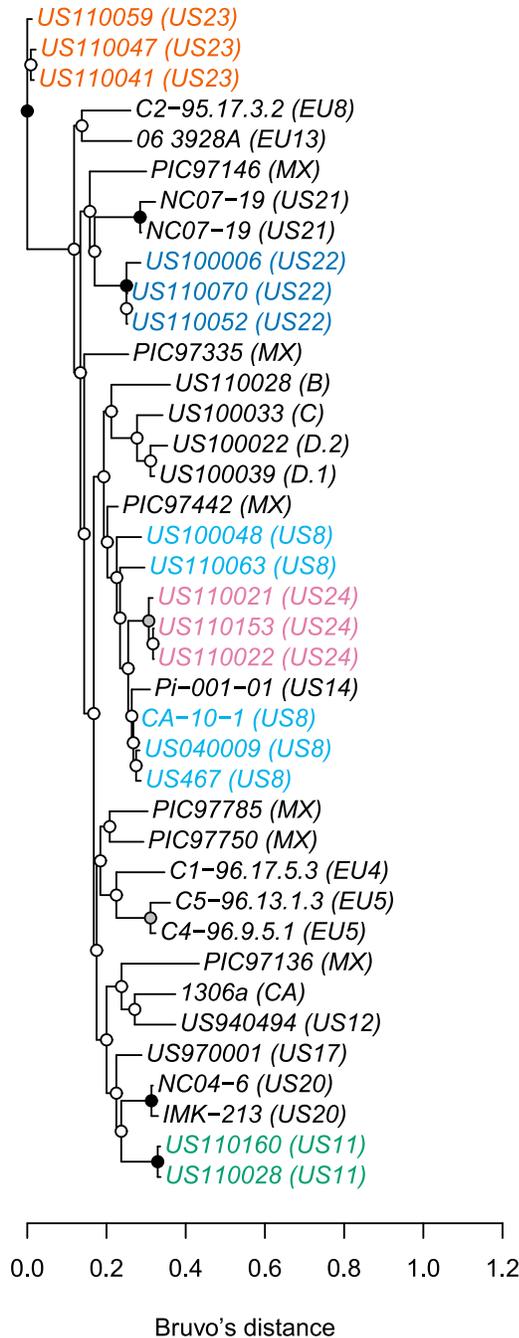

Fig. 2. Neighbor-joining tree based on 12 microsatellites and Bruvo's distance for *Phytophthora infestans*. Circles at nodes are shaded according to bootstrap proportion (based on 1,000 replicates), where black is greater than 90%, gray is 70 to 90%, and white is below 70%. Strains that are members of a U.S. clonal lineage have been shaded by lineage (colored in the online version).

Fig. 3. Maximum-likelihood tree (RAxML) with bootstrap values (based on 1,000 replicates) for the mitochondrial genome of *Phytophthora infestans* populations sampled. Circles at nodes are shaded according to bootstrap proportion, where black is greater than 90%, gray is 70 to 90%, and white is below 70%. Strains that are members of a U.S. clonal lineage have been shaded by lineage (colored in the online version).
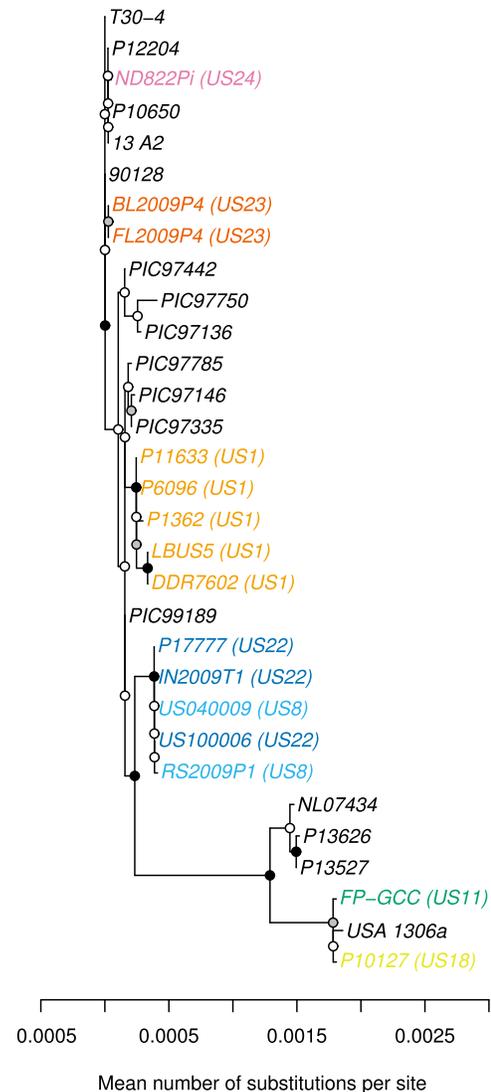
results provided strong support for scenarios inferred by previous studies (Figs. 1B, C, or D) (Fry et al. 2013; Martin et al. 2014; Yoshida et al. 2013). The clustering based on mitochondrial data and one of the two nuclear haplotypes in the US11 and US18 strains provides evidence for a shared ancestral haplotype while the other nuclear haplotype assorts independently (Figs. 3 and 4). The lineages US8, US23, and US24 did not form well-supported clades that were shared with other U.S. lineages in the data presented.

Estimation of divergence times for US1 and US22. Among the core orthologous genes analyzed on supercontig_1.1 were a total of 1,096,137 nucleotides. The average number of substitutions per site for the branch leading to the US1 clade was 1.1×10^{-3}. When multiplied by the number of nucleotides in the core orthologous genes of supercontig_1.1, it resulted in 1,147.1 substitutions that differentiate US1 from its inferred ancestor. The average number of substitutions per site for the branch leading to the US22 clade was 7.2×10^{-4}. When multiplied by the number of nucleotides in the core orthologous genes of supercontig_1.1, it resulted in 787.8

substitutions that differentiate US22 from its inferred ancestor. From these estimates of the number of substitutions, two mutation rates were calculated for each lineage based on the number of years since the Great Famine and years since the displacement of US1. Using 170 years since the Great Famine and 25 years since the displacement of US1 in North America as denominators, we get 6.16×10^{-6} and 4.18×10^{-4} substitutions per site per year that would be required for our observed divergence to be realized within these proposed time frames. Similarly, a rate of 4.23×10^{-6} and 4.18×10^{-5} substitutions per site per year would have had to occur to observe the amount of divergence in the data to explain a divergence over the 170 years since the Great Famine or the 25 years since the displacement of US1.

DISCUSSION

Are U.S. clonal lineages related? A specific objective of our work was to determine whether any of the dominant U.S. clonal
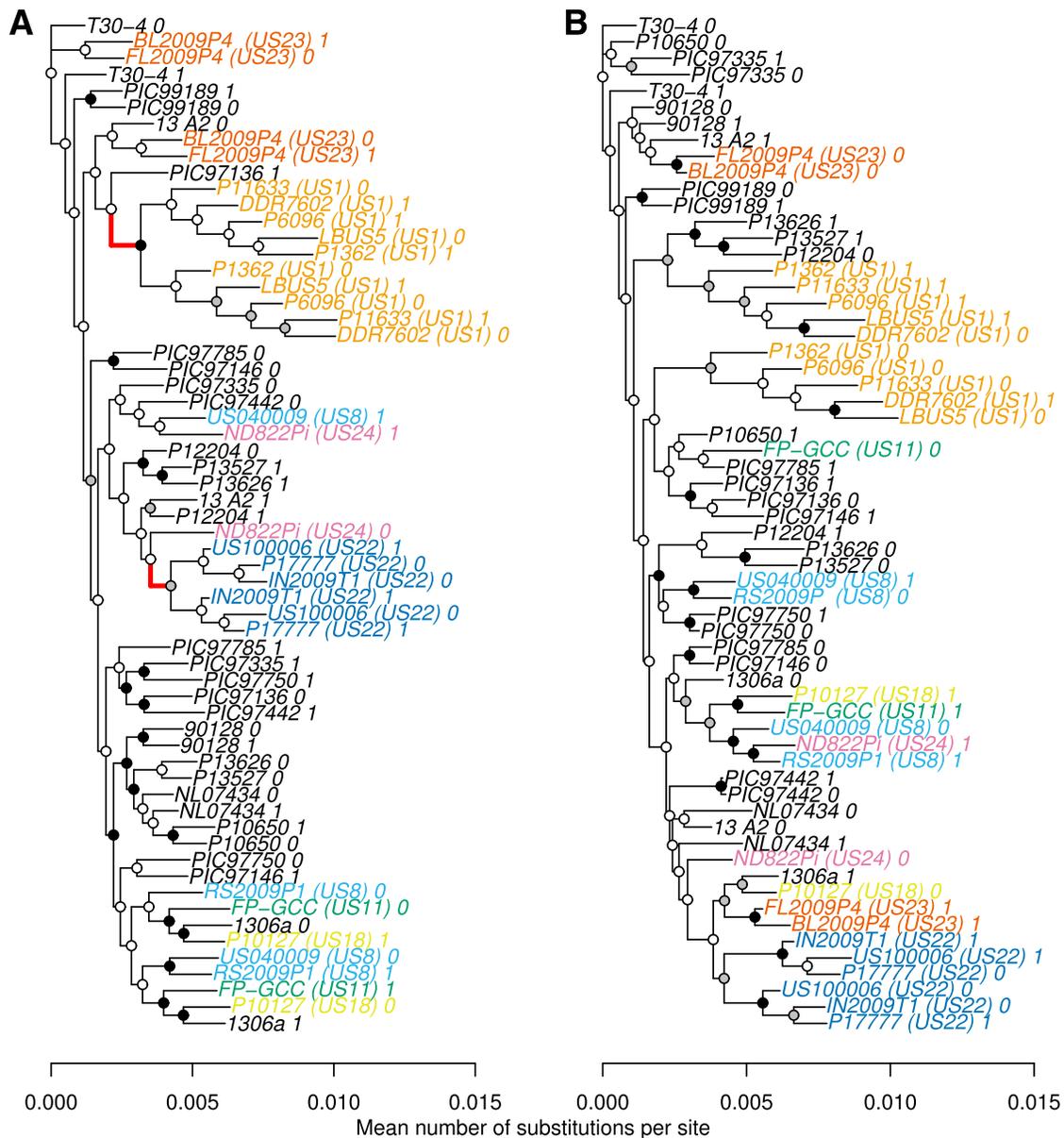

Fig. 4. Maximum-likelihood phylogeny (RAxML) with bootstrap values (based on 1,000 replicates) based on core orthologs found on **A**, supercontig_1.1 and **B**, supercontig_1.5 of *Phytophthora infestans*. Supercontig_1.1 is the longest supercontig and is relatively gene dense. Supercontig_1.5 has the highest incidence of RxLR genes and is relatively gene-sparse. Circles at nodes are shaded according to bootstrap proportion, where black is greater than 90%, gray is 70 to 90% and white is below 70%. Strains that are members of a U.S. clonal lineage have been shaded by lineage (colored in the online version). Branches used for estimates of divergence time for US1 and US22 are wider than others (colored red in the online version).

lineages of *P. infestans* might have emerged from another U.S. lineage. We found evidence supporting two distinct mechanisms. First, US11 and US18 shared one nuclear haplotype that formed a clade on both supercontig_1.1 and 1.5 but not the second nuclear haplotype that assorted independently. This pattern was supported by the clustering of the mitochondrial haplotypes as well. Thus, it appears that US11 and US18 share ancestry for one haplotype but not the other. Second, we did not find any evidence to suggest that any other U.S. clonal lineages share IBD across the various genetic regions analyzed. *Phytophthora* spp. are diploid and, thus, have two haplotypes at each nuclear locus. Thus, at any given nuclear locus, we would expect to find at least one haplotype that is inherited from a previously existing clonal lineage. Instead, we found that alleles in different regions are assorting themselves independently, as expected for a sexual population.

Lineages US11 and US18 and isolate 1306a demonstrated a close affinity to each other. The isolate referred to as 1306a was collected from California in the early 1990s (Cvitanich and Judelson 2003; Vartanian and Endo 1985) but has not been typed with the RG-57 probe to determine clonal lineage. Lineage US11 is important
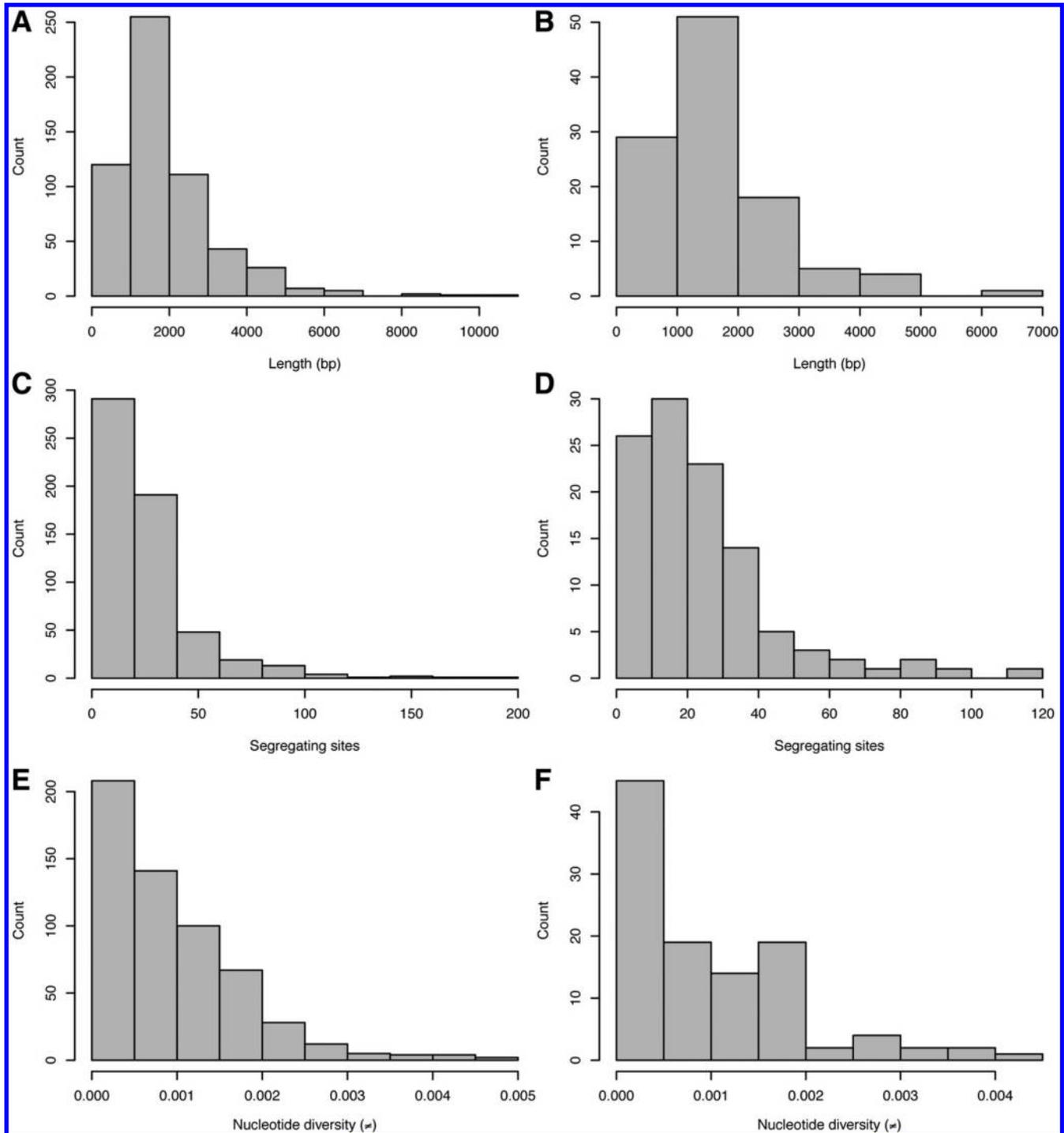

Fig. 5. Summary statistics for the 571 core orthologs from supercontig_1.1 and 108 core orthologs from supercontig_1.5 of *Phytophthora infestans* which contained at least one segregating site. Length of sequences for **A,** supercontig_1.1 (mean = 1919.7 bp) and **B,** supercontig_1.5 (mean = 1685.9). Number of segregating sites for **C,** supercontig_1.1 (mean = 26.3 sites) and **D,** supercontig_1.5 (mean = 24.4 sites). Nucleotide diversity (π) for **E,** supercontig_1.1 (mean = 0.00094) and **F,** supercontig_1.5 (mean = 0.001).

because it was one of the predominant lineages of the reemergence of *P. infestans* in the United States during the 1990s and it has been characterized as having resistance to the fungicide metalaxyl (Fry and Goodwin 1997b). The relationship among US11 and US18 provides an example of independent haplotype sorting among different genetic loci and is consistent with a scenario where (i) US11 and US18 share a common ancestor or (ii) either US11 or US18 was derived from the other clonal lineage. Either way, both US11 and US18 appear to be the product of a cross with an unknown donor ancestor for the second homologous haplotype and potential migration after recombination. Gavino et al. (2000) performed a cross between lineages US6 and US7 and, based on the RG57 banding pattern of the progeny, concluded that US11 was derived from this cross. The lineages US6 and US7 are thought to be extinct in the wild but may currently exist in culture collections. The inclusion of these isolates in future studies may provide new insights on the relations inferred here as well as those presented by Gavino et al. (2000).

Lineage US22 has been characterized as part of the most recent pandemic (Fry et al. 2013). This lineage has been described as sensitive to mefenoxam (Fry et al. 2013). Analyses of microsatellites, the mitochondrion, and supercontig_1.1 support multiple samples of US22 forming a well-supported clade. However, the clade from the tree inferred from mitochondrial data included well-supported relationships among samples of US8 intermixed with US22, a relationship not supported in the other analyses. The mitochondrion and one of the nuclear haplotypes are expected to be maternally inherited from generation to generation. This provides the expectation that inferences based on the mitochondrion should be reflected in the nuclear data; however, they are not. This is an example of a statistically significant result that is supported by one fraction of the genome but not elsewhere, and perhaps an important consideration when interpreting findings of statistical support. For example, regions of the genome that consist of a large linkage block may resolve well statistically. Regions that have more frequent recombination or less polymorphism to record their pattern of descent may not resolve well. Therefore, this mosaic nature of closely related genomes may be better represented as ancestral recombination graphs (Rasmussen et al. 2014). However, we feel that these recently presented methods may still present technical challenges in their implementation.

The lineage US1 is represented in the sample by isolates DDR7602 (reported to be US1 by Kamoun et al. [1998]) as well as isolates from South Africa (LBUS5), Peru (P6096), and Mexico (P1362) (reported as US1 by Yoshida et al. [2013]). Our analysis also included the sample P11633 from Hungary as part of this grouping. This lineage is important because it was considered historically to be the lineage that caused the Great Famine (Goodwin et al. 1994), although recent work indicates that, instead, it may have been a member of the same metapopulation (Yoshida et al. 2013). Prior to the early 1990s, US1 was thought to be globally distributed. Since the emergence of new strains in the United States and Europe, it has been displaced (Cooke et al. 2012; Fry and Goodwin 1997a,b; Fry et al. 2013). However, it remains present in regions of Africa (Njoroge et al. 2015) and Asia (Petchaboon et al. 2014). Although the US1 lineage was absent from our microsatellite data set, it formed well-supported clades in the mitochondrial tree (Fig. 3) as well as supercontig_1.1 (Fig. 5). The tree inferred from supercontig_1.5 (Fig. 5) showed clades for each haplotype of US1; however, there was no support for a close relationship among these two haplotypes. The distinctiveness of this lineage was also confirmed by unsupervised clustering (fastSTRUCTURE). The concept of this lineage as being one of the oldest U.S. lineages is consistent with it being relatively well defined, because most markers and analyses captured its presence. However, no well-supported relationships for this lineage were inferred, suggesting it to be a unique migrant from a sexual population probably located in central Mexico (Goodwin et al. 1994).

Lineages US8, US23, and US24 did not demonstrate strong relationships to other lineages. For example, in the microsatellite analysis (Fig. 2), US8 formed at least two unsupported clades. In the mitochondrial analysis, US8 showed a strong alliance with US22. However, this relationship was not supported in the analyses of supercontig_1.1 or 1.5. The distant relationship of the two haplotypes which make up US8 in this data may be an indication that the ancestors of this lineage were distantly related individuals. Heterosis that may have resulted from this cross may have contributed to this lineage's pathogenicity. The lineage US23 is similar, in that it formed a supported cluster in the microsatellite analysis. It also formed a supported clade in the mitochondrial analysis. Analyses of supercontig_1.1 and 1.5 show a distant relationship among the two haplotypes, again suggesting that the ancestors of this lineage may have been highly unrelated. Inferences from US24 are less robust due to its representation by a single sample but it shows a pattern similar to that of US8 and US23, where microsatellites show a significant cluster, the mitochondrion does not reveal significant relationships, and the supercontig fractions show a distant relation among each haplotype. A general lack of significant phylogenetic relationships among these lineages is perhaps best interpreted as being due to their relatively recent origin, where there has not been sufficient evolutionary time for the evolution of shared, derived mutations to identify these lineages. This recent origin may also be confounded by our attempt to analyze many genes if they actually lack a common pattern of descent.

Analyses of the six select genes appear in congruence with the above interpretations of the lineages US1 and US22 forming individual clades and an inferred relationship among US11 and US18. Little bootstrap support exists for the clades discussed above in any of these genes. A paucity of phylogenetically informative sites appears to have resulted in phylogenies where a small number of variants may drive support for a clade. The general lack of variants has resulted in little conflict among the characters and support for only a small number of clades. This appears congruent with the idea that these lineages are recently derived sexual progeny that may lack unique mutations to mark them at these loci, as well as the possibility that different loci may represent different patterns of descent within the group. This data are also in line with the observed shared ancestry of US11 and US18, given the lack of statistical support for clustering at the branch tips.

Ploidy in *P. infestans*. We have treated *P. infestans* as a diploid. This is important because, to our knowledge, no existing variant caller will infer ploidy. Instead, it needs to be specified as a single level of ploidy during variant calling. Our assumption of diploidy is consistent with other genomic treatments of *P. infestans* (Martin et al. 2013; Raffaele et al. 2010; Yoshida et al. 2013). Some authors have argued that some strains of *P. infestans* are not simple diploids (Li et al. in press; Yoshida et al. 2013). This fact could affect inferences presented here as well as throughout the *P. infestans* literature. In theory, if we knew which samples are of which ploidy level, we could call variants accordingly. It is possible, however, that this could be even more complicated in that only portions of any particular genome may vary in ploidy, in which case we would need to know the genomic coordinates for which fractions of which genomes are of a certain ploidy. The analyses presented here are accurate to the best of our knowledge. However, if it were to subsequently become known that some of the samples were of a different ploidy, it could dramatically affect our results, as well as those reported in the literature that have made the same assumptions. We believe that the two variants per locus we were able to call are real but that, in polyploid species, we might have missed detection of additional alleles.

Recombination and the analysis of *P. infestans* genomes. Phylogenetic inference relies on the assumption that there is a single path of descent through the generations; that is, there is a lack of recombination (Lanier and Knowles 2012). Model-based clustering methods such as structure (Pritchard et al. 2000), admixture

(Alexander et al. 2009), and fastSTRUCTURE (Raj et al. 2014) are at the opposite end of the spectrum, where they assume each locus to be independent due to recombination. We are aware that our fastSTRUCTURE analysis probably violates these assumptions when applied to individual supercontigs and, thus, it is provided with caution and shown only in the Supplementary Materials for reference. Note that the authors of the original Structure software do provide an example of its use in bacteria, where little recombination and clonality can be expected (Falush et al. 2003). The choice of the most appropriate analytical tool for the analysis of *P. infestans* is particularly challenging due to the mixed reproductive modes. At its center of origin in Mexico, sexual structures are observed and the populations are considered to be sexually reproducing (Goss et al. 2014; Grünwald et al. 2001). In the United States, no sexual structures are observed and the populations are considered to be clonal (Fry and Goodwin 1997a; Fry et al. 2013). To our knowledge, there is no analytical tool currently available that allows for the analysis of both sexual and clonal populations in the same analysis. Instead, we employed both types of analyses with the expectation that violations of the analysis have occurred in both analyses.

Are the clonal lineages recently evolved? The lineages of *P. infestans* are numbered in the order of description or observation (Forbes et al. 1998). An inherent pitfall of this system is the desire to assume some sort of evolutionary relationship among U.S. lineages (Goodwin et al. 1998). We find no evidence for the U.S. lineages to be derived from one another, with the exception of the relationship among US11 and US18. In the absence of supported relationships, we feel a more likely scenario is that the U.S. lineages are independent migrants from a standing pool of genetic diversity present at a center of origin such as Mexico (Goss et al. 2014; Grünwald and Flier 2005) or a cryptic, sexual population elsewhere.

In order to explore the age of two select lineages, we estimated two mutation rates each for the divergence of the clades containing the lineages US1 and US22. The amount of divergence we observed in the present data during the time since the Great Famine or since the displacement of US1 from North America would require mutation rates on the order of 10×10^{-6} to 10×10^{-5} substitutions per site per year. For context, the spontaneous mutation rate for *Arabidopsis thaliana* has been estimated to be 7×10^{-9} substitutions per site per generation (Ossowski et al. 2010). Similarly, mutation rates in humans are considered to be on the order of 10^{-8} to 10^{-9} substitutions per site per generation (Scally and Durbin 2012). Mutation rate estimates for these well-characterized genomes are orders of magnitude slower than our estimates, which attempt to place divergence of the U.S. lineages in the early 1990s or even the mid-19th century. Based on this information, we can conclude that there either would have to be great selective pressure to drive this mutation rate or, the scenario we prefer, these lineages are much older than 170 years old. By using only the amount of collective divergence of the lineage (that is, the time from the lineage's most recent common ancestor to its ancestor), we have provided an underestimate of each lineage's age. Each lineage's age may be better represented by including the mutational length of branches from the most recent common ancestor to the tips. This inclusion would add mutations to the numerator of our estimate without changing the denominator, resulting in an even higher estimate. If we assume that error is uniformly distributed, then it is unlikely for technical error to change two samples to the same state. The result is that most technical error will be unique to a sample and contribute to long tips in the phylogeny Therefore, our choice of focusing on the amount of collective divergence for each lineage should be considered very conservative. The divergence of mitochondrial haplotypes I and II was estimated to have occurred during the 14th (Martin et al. 2014) or 16th centuries (Yoshida et al. 2013), reinforcing the idea that these lineages predate their emergence in the United States during the late 20th and early 21st centuries. This suggests that these lineages may have been present in their center of diversity in Mexico and recently dispersed to the United States or,

alternatively, may have existed undetected in the United States for several centuries.

Our analyses of a sample available to us of the over 20 U.S. lineages identified since the 1990s using nuclear and mitochondrial sequences failed to generate a well-supported phylogeny, except for US11 and US18. Although an individual locus might have suggested phylogenetic relationships among clonal lineages, the genome-wide analysis seems to indicate independent origins of the U.S. lineages. Furthermore, these lineages appear to be much older than their description in the literature. This indicates that the U.S. lineages may have been independent and ongoing introductions of *P. infestans* from a center of diversity, such as central Mexico. This work provides an initial perspective on the genomic history of the U.S. lineages of *P. infestans*. Identification of linkage blocks, which form the unit of descent among generations, may become a critical element in future attempts to infer relationships among U.S. lineages. As we continue to explore this genomic landscape, this perspective of U.S. lineages containing the hallmark of ancestral recombination may contribute to our attempts to identify genes important for host adaptation or other traits.

LITERATURE CITED

Adler, N. E., Erselius, L. J., Chacon, M., Flier, W. G., Ordoñez, M. E., Kroon, L. P., and Forbes, G. A. 2004. Genetic diversity of *Phytophthora infestans* sensu lato in Ecuador provides new insight into the origin of this important plant pathogen. Phytopathology 94:154-162.

Alexander, D. H., Novembre, J., and Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655-1664.

Alkher, H., Islam, M. R., Wijekoon, C., Kalischuk, M., Kawchuk, L. M., Peters, R. D., Al-Mughrabi, K. I., Conn, K. L., Dobinson, K. F., Waterer, D., and Daayf, F. 2015. Characterization of *Phytophthora infestans* populations in Canada during 2012. Can. J. Plant Pathol. 37:305-314.

Avila-Adame, C., Gómez-Alpizar, L., Zismann, V., Jones, K. M., Buell, C. R., and Ristaino, J. B. 2006. Mitochondrial genome sequences and molecular evolution of the Irish potato famine pathogen, *Phytophthora infestans*. Curr. Genet. 49:39-46.

Blair, J. E., Coffey, M. D., Park, S. Y., Geiser, D. M., and Kang, S. 2008. A multi-locus phylogeny for *Phytophthora* utilizing markers derived from complete genome sequences. Fungal Genet. Biol. 45:266-277.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C., Xie, D., Suchard, M. A., Rambaut, A., and Drummond, A. J. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. PLOS Comput. Biol. 10: e1003537.

Broad Institute. 2014. *Phytophthora infestans* sequencing project. Broad Institute of Harvard and MIT. https://www.ncbi.nlm.nih.gov/bioproject/17665

Browning, S. R., and Browning, B. L. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084-1097.

Carbone, I., and Kohn, L. 2004. Inferring process from pattern in fungal population genetics. Appl. Mycol. Biotechnol. 4:29-58.

Cooke, D. E. L., Cano, L. M., Raffaele, S., Bain, R. A., Cooke, L. R., Etherington, G. J., Deahl, K. L., Farrer, R. A., Gilroy, E. M., Goss, E. M., Grünwald, N. J., Hein, I., MacLean, D., McNicol, J. W., Randall, E., Oliva, R. F., Pel, M. A., Shaw, D. S., Squires, J. N., Taylor, M. C., Vleeshouwers, V. G. A. A., Birch, P. R. J., Lees, A. K., and Kamoun, S. 2012. Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. PLoS Pathog. 8:e1002940.

Cooke, D. E. L., and Lees, A. K. 2004. Markers, old and new, for examining *Phytophthora infestans* diversity. Plant Pathol. 53:692-704.

Cvitanich, C., and Judelson, H. S. 2003. A gene expressed during sexual and asexual sporulation in *Phytophthora infestans* is a member of the Puf family of translational regulators. Eukaryot. Cell 2:465-473.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., and McVean, G. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. 2012. jModelTest 2: More models, new heuristics and parallel computing. Nat. Methods 9:772.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko. A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., and Daly, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491-498.

Drenth, A., Goodwin, S., Fry, W., and Davidse, L. 1993. Genotypic diversity of *Phytophthora infestans* in the Netherlands revealed by DNA polymorphisms. Phytopathology 83:1087-1092.

Falush, D., Wirth, T., Linz, B., Pritchard, J. K., Stephens, M., Kidd, M., Blaser, M. J., Graham, D. Y., Vacher, S., Perez-Perez, G. L., Yamaoka, Y., Mégraud, F., Otto, K., Reichard, U., Katzowitsch, E., Wang, X., Achtman, M., and Suerbaum, S. 2003. Traces of human migrations in *Helicobacter pylori* populations. Science 299:1582-1585.

Flier, W. G., Grünwald, N. J., and Turkensteen, L. J. 2001. Formation, production and viability of oospores of *Phytophthora infestans* from potato and *Solanum demissum* in the Toluca Valley, central Mexico. Mycol. Res. 105:998-1006.

Forbes, G. A., Goodwin, S. B., Drenth, A., Oyarzun, P., Ordoñez, M. E., and Fry, W. E. 1998. A global marker database for *Phytophthora infestans*. Plant Dis. 82:811-818.

Fry, W. E., and Goodwin, S. B. 1997a. Resurgence of the Irish potato famine fungus. Bioscience 47:363-371.

Fry, W. E., and Goodwin, S. B. 1997b. Re-emergence of potato and tomato late blight in the United States. Plant Dis. 81:1349-1357.

Fry, W. E., Goodwin, S. B., Dyer, A. T., Matuszak, J. M., Drenth, A., Tooley, P. W., Sujkowski, L. S., Koh, Y. J., Cohen, B. A., and Spielman, L. J. 1993. Historical and recent migrations of *Phytophthora infestans*: Chronology, pathways, and implications. Plant Dis. 77:653-661.

Fry, W. E., McGrath, M., Seaman, A., Zitter, T., McLeod, A., Danies, G., Small, I. M., Myers, K., Everts, K., Gevens, A. J., Gugino, B. K., Johnson, S. B., Judelson, H., Roberts, P., Secor, G., Seebold, K. Jr., Snover-Clift, K., Wyenandt, A., Grünwald, N. J., and Smart, C. D. 2013. The 2009 late blight pandemic in the eastern United States-causes and results. Plant Dis. 97:296-306.

Gagnon, M.-C., Bergeron, M.-J., Hamelin, R. C., Grünwald, N. J., and Bilodeau, G. J. 2014. Real-time PCR assay to distinguish *Phytophthora ramorum* lineages using the cellulose binding elicitor lectin (CBEL) locus. Can. J. Plant Pathol. 36:367-376.

Gavino, P. D., Smart, C. D., Sandrock, R. W., Miller, J. S., Hamm, P. B., Lee, T. Y., Davies, R. M., and Fry, W. E. 2000. Implications of sexual reproduction for *Phytophthora infestans* in the United States: Generation of an aggressive lineage. Plant Dis. 84:731-735.

Gómez-Alpizar, L., Carbone, I., and Ristaino, J. B. 2007. An Andean origin of *Phytophthora infestans* inferred from mitochondrial and nuclear gene genealogies. Proc. Natl. Acad. Sci. USA 104:3306-3311.

Goodwin, S. B., Cohen, B. A., and Fry, W. E. 1994. Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. Proc. Natl. Acad. Sci. USA 91:11591-11595.

Goodwin, S. B., Drenth, A., and Fry, W. E. 1992. Cloning and genetic analyses of two highly polymorphic, moderately repetitive nuclear DNAs from *Phytophthora infestans*. Curr. Genet. 22:107-115.

Goodwin, S. B., Schneider, R. E., and Fry, W. E. 1995. Use of cellulose-acetate electrophoresis for rapid identification of allozyme genotypes of *Phytophthora infestans*. Plant Dis. 79:1181-1185.

Goodwin, S. B., Smart, C. D., Sandrock, R. W., Deahl, K. L., Punja, Z. K., and Fry, W. E. 1998. Genetic change within populations of *Phytophthora infestans* in the United States and Canada during 1994 to 1996: Role of migration and recombination. Phytopathology 88:939-949.

Goss, E. M., Cardenas, M. E., Myers, K., Forbes, G. A., Fry, W. E., Restrepo, S., and Grünwald, N. J. 2011. The plant pathogen *Phytophthora andina* emerged via hybridization of an unknown *Phytophthora* species and the Irish potato famine pathogen, *P. infestans*. PLoS One 6:e24543.

Goss, E. M., Tabima, J. F., Cooke, D. E., Restrepo, S., Fry, W. E., Forbes, G. A., Fieland, V. J., Cardenas, M., and Grünwald, N. J. 2014. The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. Proc. Natl. Acad. Sci. USA 111:8791-8796.

Griffith, G. W., and Shaw, D. S. 1998. Polymorphisms in *Phytophthora infestans*: Four mitochondrial haplotypes are detected after PCR amplification of DNA from pure cultures or from host lesions. Appl. Environ. Microbiol. 64:4007-4014.

Grünwald, N. J., and Flier, W. G. 2005. The biology of *Phytophthora infestans* at its center of origin. Annu. Rev. Phytopathol. 43:171-190.

Grünwald, N. J., Flier, W. G., Sturbaum, A. K., Garay-Serrano, E., van den Bosch, T. B. M., Smart, C. D., Matuszak, J. M., Lozoya-Saldaña, H., Turkensteen, L. J., and Fry, W. E. 2001. Population structure of *Phytophthora infestans* in the Toluca valley region of central Mexico. Phytopathology 91:882-890.

Grünwald, N. J., and Goss, E. M. 2011. Evolution and population genetics of exotic and re-emerging pathogens: Traditional and novel tools and approaches. Annu. Rev. Phytopathol. 49:249-267.

Grünwald, N. J., Goss, E. M., Ivors, K., Garbelotto, M., Martin, F. N., Prospero, S., Hansen, E., Bonants, P. J. M., Hamelin, R. C., Chastagner, G., Werres, S., Rizzo, D. M., Abad, G., Beales, P., Bilodeau, G. J., Blomquist, C. L., Brasier, C., Brière, S. C., Chandelier, A., Davidson, J. M., Denman, S., Elliott, M., Frankel, S. J., Goheen, E. M., de Gruyter, H., Heungens, K., James, D., Kanaskie, A., McWilliams, M. G., Man in 't Veld, W., Moralejo, E., Osterbauer, N. K., Palm, M. E., Parke, J. L., Perez Sierra, A. M., Shamoun, S. F., Shishkoff, N., Tooley, P. W., Vettraino, A. M., Webber, J., and Widmer, T. L. 2009. Standardizing the nomenclature for clonal lineages of the sudden oak death pathogen, *Phytophthora ramorum*. Phytopathology 99:792-795.

Grünwald, N. J., Martin, F. N., Larsen, M. M., Sullivan, C. M., Press, C. M., Coffey, M. D., Hansen, E. M., and Parke, J. L. 2011. *Phytophthora*-ID.org: A sequence-based *Phytophthora* identification tool. Plant Dis. 95:337-342.

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I. B., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grünwald, N. J., Horn, K., Horner, N. R., Hu, C.-H., Huitema, E., Jeong, D.-H., Jones, A. M. E., Jones, J. D. G., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L. MacLean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J. G., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J. I., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R. J., Whisson, S. C., Judelson, H. S., and Nusbaum, C. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461:393-398.

Hu, C.-H., Perez, F. G., Donahoo, R., McLeod, A., Myers, K., Ivors, K., Secor, G., Roberts, P. D., Deahl, K. L., Fry, W. E., and Ristaino, J. B. 2012. Recent genotypes of *Phytophthora infestans* in the eastern United States reveal clonal populations and reappearance of mefenoxam sensitivity. Plant Dis. 96:1323-1330.

Kamoun, S., Furzer, O., Jones, J. D. G., Judelson, H. S., Ali, G. S., Dalio, R. J. D., Roy, S. G., Schena, L., Zambounis, A., Panabières, F., Cahill, D., Ruocco, M., Figueiredo, A., Chen, X.-R., Hulvey, J., Stam, R., Lamour, K., Gijzen, M., Tyler, B. M., Grünwald, N. J., Mukhtar, M. S., Tome, D. F. A., Tör, M., Van Den Ackerveken, G., McDowell, J., Daayf, F., Fry, W. E., Lindqvist-Kreuze, H., Meijer, H. J. G., Petre, B., Ristaino, J., Yoshida, K., Birch, P. R. J., and Govers, F. 2015. The Top 10 oomycete pathogens in molecular plant pathology. Mol. Plant Pathol. 16:413-434.

Kamoun, S., van der Lee, T., van den Berg-Velthuis, G., de Groot, K. E., and Govers, F. 1998. Loss of production of the elicitor protein INF1 in the clonal lineage US-1 of *Phytophthora infestans*. Phytopathology 88:1315-1323.

Kamvar, Z. N., Larsen, M. M., Kanaskie, A. M., Hansen, E. M., and Grünwald, N. J. 2015. Spatial and temporal analysis of populations of the sudden oak death pathogen in Oregon forests. Phytopathology 105:982-989.

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.

Knaus, B. J., and Grunwald, N. J. VcfR: An R package to manipulate and visualize VCF format data. bioRxiv. In press. doi:10.1101/041277

Knaus, B. J., Tabima, J. F., Judelson, H. S., and Grünwald, N. J. 2015. *Phytophthora infestans* US lineage diagnostic SNP data. Online publication. Ag Data Commons, United States Department of Agriculture–Agricultural Research Service National Agricultural Library. https://data.nal.usda.gov/dataset/data-genomic-analyses-dominant-us-clonal-lineages-phytophthora-infestans-reveals-recent

Kroon, L. P. N. M., Bakker, F. T., Van Den Bosch, G. B. M., Bonants, P. J. M., and Flier, W. G. 2004. Phylogenetic analysis of *Phytophthora* species based on mitochondrial and nuclear DNA sequences. Fungal Genet. Biol. 41:766-782.

Lanier, H. C., and Knowles, L. L. 2012. Is recombination a problem for species-tree analyses? Syst. Biol. 61:691-701.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Online publication. http://arxiv.org/abs/1303.3997v2

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abercasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Li, Y., Cooke, D. E. L., Jacobsen, E., and van der Lee, T. 2013. Efficient multiplex simple sequence repeat genotyping of the oomycete plant pathogen *Phytophthora infestans*. J. Microbiol. Methods 92:316-322.

Li, Y., Zhou, Q., Qian, K., van der Lee, T., and Huang, S. Successful asexual lineages of the Irish potato famine pathogen are triploid. bioRxiv. In press. doi:10.1101/024596

Martin, M. D., Cappellini, E., Samaniego, J. A., Zepeda, M. L., Campos, P. F., Seguin-Orlando, A., Wales, N., Orlando, L., Ho, S. Y. W., Dietrich, F. S., Pieczkowski, P. A., Heitman, J., Willerslev, E., Krogh, A., Ristaino, J. B., and Gilbert, M. T. P. 2013. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. Nat. Commun. 4:Article 2172. doi:10.1038/ncomms3172

Martin, M. D., Ho, S. Y., Wales, N., Ristaino, J. B., and Gilbert, M. T. P. 2014. Persistence of the mitochondrial lineage responsible for the Irish potato famine in extant New World *Phytophthora infestans*. Mol. Biol. Evol. 31: 1414-20.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297-1303.

Milgrom, M. G. 1996. Recombination and the multilocus structure of fungal populations. Annu. Rev. Phytopathol. 34:457-477.

Njoroge, A., Tusiime, G., Forbes, G., and Yuen, J. 2015. Displacement of US-1 clonal lineage by a new lineage of *Phytophthora infestans* on potato in Kenya and Uganda. Plant Pathol. 65:587-592.

Oliva, R. F., Kroon, L. P. N. M., Chacón, G., Flier, W. G., Ristaino, J. B., and Forbes, G. A. 2010. *Phytophthora andina* sp. nov., a newly identified heterothallic pathogen of solanaceous hosts in the Andean highlands. Plant Pathol. 59:613-625.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., Weigel, D., and Lynch, M. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327:92-94.

Paradis, E., Claude, J., and Strimmer, K. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289-290.

Petchaboon, K., Srisawad, N., Pongam, P., Chunwongse, C., and Chunwongse, J. 2014. Phenotypic and genotypic structure of *Phytophthora infestans* populations on tomato and potato in the North of Thailand in 2000–2002. Eur. J. Plant Pathol. 140:441-454.

Peters, R., Al-Mughrabi, K., Kalischuk, M., Dobinson, K., Conn, K., Alkher, H., Islam, M. R., Daayf, F., Lynn, J., Bizimungu, B., De Koeyer, D., Lévensque, C. A., and Kawchuk, L. M. 2014. Characterization of *Phytophthora infestans* population diversity in Canada reveals increased migration and genotype recombination. Can. J. Plant Pathol. 36:73-82.

Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., and Sham, P. C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.

Raffaele, S., Win, J., Cano, L. M., and Kamoun, S. 2010. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. BMC Genomics 11: 637.

Raj, A., Stephens, M., and Pritchard, J. K. 2014. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics 197:573-589.

Rambaut, A., and Drummond, A. 2009. FigTree v1. 3. 1. Online publication. http://tree.bio.ed.ac.uk/software/figtree/

Rasmussen, M. D., Hubisz, M. J., Gronau, I., and Siepel, A. 2014. Genomewide inference of ancestral recombination graphs. PLoS Genet. 10: e1004342.

R Core Team. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Saville, A., Graham, K., Grünwald, N. J., Myers, K., Fry, W. E., and Ristaino, J. B. 2015. Fungicide sensitivity of US genotypes of *Phytophthora infestans* (Mont.) de Bary to six oomycete-targeted compounds. Plant Dis. 99: 659-666.

Scally, A., and Durbin, R. 2012. Revising the human mutation rate: Implications for understanding human evolution. Nat. Rev. Genet. 13:745-753.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., and DePristo, M. A. 2013. From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr. Protocols Bioinf. 11:11.10.1-11.10.33.

Vartanian, V. G., and Endo, R. M. 1985. Overwintering hosts, compatibility types, and races of *Phytophthora infestans* on tomato in southern California. Plant Dis. 69:516-519.

Wangsomboondee, T., Groves, C. T., Shoemaker, P., Cubeta, M., and Ristaino, J. 2002. *Phytophthora infestans* populations from tomato and potato in North Carolina differ in genetic diversity and structure. Phytopathology 92: 1189-1195.

Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer Science & Business Media. Springer-Verlag, New York.

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., and Burbano, H. A. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. eLife 2:e00731.