CrossMark
click for updates

The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes

Erica M. Goss[a], Javier F. Tabima[b], David E. L. Cooke[c], Silvia Restrepo[d], William E. Fry[e], Gregory A. Forbes[f], Valerie J. Fieland[b], Martha Cardenas[d], and Niklaus J. Grünwald[g,h,1]

[a]Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611; [b]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331; [c]The James Hutton Institute, Invergowrie, Dundee DD2 5DA, Scotland; [d]Department of Biological Sciences, University of the Andes, 110321 Bogota, Colombia; [e]Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY 14853; [f]CIP China Center for Asia and the Pacific, International Potato Center, Beijing 100081, China; [g]Horticultural Crops Research Laboratory, US Department of Agriculture Agricultural Research Service, Corvallis, OR 97330; and [h]Department of Botany and Plant Pathology and Center for Genome Biology and Biocomputing, Oregon State University, Corvallis, OR 97331

Phytophthora infestans is a destructive plant pathogen best known for causing the disease that triggered the Irish potato famine and remains the most costly potato pathogen to manage worldwide. Identification of *P. infestan*'s elusive center of origin is critical to understanding the mechanisms of repeated global emergence of this pathogen. There are two competing theories, placing the origin in either South America or in central Mexico, both of which are centers of diversity of *Solanum* host plants. To test these competing hypotheses, we conducted detailed phylogeographic and approximate Bayesian computation analyses, which are suitable approaches to unraveling complex demographic histories. Our analyses used microsatellite markers and sequences of four nuclear genes sampled from populations in the Andes, Mexico, and elsewhere. To infer the ancestral state, we included the closest known relatives *Phytophthora phaseoli*, *Phytophthora mirabilis*, and *Phytophthora ipomoeae*, as well as the interspecific hybrid *Phytophthora andina*. We did not find support for an Andean origin of *P. infestans*; rather, the sequence data suggest a Mexican origin. Our findings support the hypothesis that populations found in the Andes are descendants of the Mexican populations and reconcile previous findings of ancestral variation in the Andes. Although centers of origin are well documented as centers of evolution and diversity for numerous crop plants, the number of plant pathogens with a known geographic origin are limited. This work has important implications for our understanding of the coevolution of hosts and pathogens, as well as the harnessing of plant disease resistance to manage late blight.

biological invasion | coalescent analysis | oomycete | population genetics | stramenopile

The potato pathogen *Phytophthora infestans*, the causal agent of potato late blight, is the plant pathogen that has most greatly impacted humanity to date. This pathogen is best known for its causal involvement in the Irish potato famine after introduction of the HERB-1 strain to Ireland from the Americas in the 19th century (1). To this day, potato late blight remains a major threat to food security and carries a global cost conservatively estimated at more than $6 billion per year (2). In the 1980s, a single asexual lineage named US-1, possibly derived from the same metapopulation as HERB-1 (1), dominated global populations, whereas a genetically diverse and sexual population of *P. infestans* in central Mexico led to formulation of the hypothesis identifying Mexico as this pathogen's center of origin (3, 4). A competing hypothesis argues that the center of origin of the potato, the South American Andes, is the center of origin of *P. infestans* (5). This hypothesis recently gained prominence after an analysis demonstrated ancestral variation in Andean lineages of *P. infestans* (5). Other evidence supporting this hypothesis includes infection of native *Solanum* hosts and an Andean

distribution for *Phytophthora andina*, a phylogenetic relative of *P. infestans* (6).

Evidence supporting a Mexican center of origin is substantial, but inconclusive (4). Two close relatives of *P. infestans*, *Phytophthora ipomoeae* and *Phytophthora mirabilis*, are endemic to central Mexico (7, 8). *P. ipomoeae* and *P. mirabilis* cause disease on two endemic plant host groups, *Ipomoea* spp. and *Mirabilis jalapa*, respectively. Populations of *P. infestans* in the Toluca Valley, southwest of Mexico City, are genetically diverse, are in Hardy–Weinberg equilibrium, and contain mating types A1 and A2 in the expected 1:1 ratio for sexual populations (9, 10). Before a migration event from Mexico to Europe in the 1970s (11, 12), only A1 mating types of *P. infestans* were found worldwide outside of central Mexico, limiting other populations to asexual reproduction (13). Tuber-bearing native *Solanum* species occur throughout the Toluca Valley (14). Of the R genes that have been used to confer resistance to strains of *P. infestans* in potato, the majority described to date originated from *Solanum demissum* or *Solanum edinense* in the Toluca Valley, with some discovered in South America (15).

Significance

The potato late blight pathogen was introduced to Europe in the 1840s and caused the devastating loss of a staple crop, resulting in the Irish potato famine and subsequent diaspora. Research on this disease has engendered much debate, which in recent years has focused on whether the geographic origin of the pathogen is South America or central Mexico. Different lines of evidence support each hypothesis. We sequenced four nuclear genes in representative samples from Mexico and the South American Andes. An Andean origin of *P. infestans* does not receive support from detailed analyses of Andean and Mexican populations. This is one of a few examples of a pathogen with a known origin that is secondary to its current major host.

AGRICULTURAL SCIENCES

Support for the alternate hypothesis that *P. infestans* originated in the Andes is based on a coalescent analysis conducted by Gómez-Alpizar et al. (5). This analysis used the nuclear RAS locus and the mitochondrial P3 and P4 regions to infer rooted gene genealogies that showed ancestral lineages rooted in the Andes. Furthermore, the Mexico sample harbored less nucleotide diversity than the Andean population. *P. andina* was identified as the ancestral lineage for the mitochondrial genealogy; however, *P. mirabilis* and *P. ipomoeae* were not included in that study. *P. andina* has since been shown to be a hybrid species derived from *P. infestans* and a *Phytophthora* sp. unknown to science (16). Surprisingly, populations of *P. infestans* and *P. andina* are clonal in South America and are not in Hardy–Weinberg equilibrium (6, 17–19). Thus, the question of whether *P. infestans* originated in the Andes or central Mexico remained unresolved.

Powerful approaches for determining the demographic and evolutionary history of organisms are now available (20). Many of these approaches rely on the power of coalescent theory for inferring the genealogical history of a species based on a representative population sample (21–23). Bayesian phylogeography uses geographic information in light of phylogenetic uncertainty to provide model-based inference of geographic locations of ancestral strains (24). The isolation with migration (IM) model and associated software uses likelihood-based inference to infer divergence time between evolutionary lineages (25). Approximate Bayesian computation (ABC) makes use of coalescent simulations and likelihood-free inference to contrast complex demographic scenarios. Each of these methods has proven useful in reconstructing the demography of pests and pathogens (24, 26–29).

The objective of the present study was to reconcile the two competing hypotheses on the origin of *P. infestans* using Bayesian phylogenetics and ABC. We sampled key populations of *P. infestans* from central Mexico and the Andes and expanded on the analysis of Gómez-Alpizar et al. (5) by sequencing additional nuclear loci to assess support for the center of origin across multiple loci. To determine ancestral state, we added sequences from the sister taxa *P. andina*, *P. mirabilis*, *P. ipomoeae*, and *Phytophthora phaseoli*, all of which belong to *Phytophthora* clade 1c (30, 31). Finally, we aimed to reconcile the biology of *P. infestans* in Mexico with the findings of Gómez-Alpizar et al. (5) of ancestral variation in the Andes.

Results

Population Structure and Mode of Reproduction. Our sample of *P. infestans* included 40 isolates from Colombia, Ecuador, and Peru and 48 isolates from Toluca Valley and Tlaxcala State in central Mexico (*SI Appendix*, Table S1). We found 30 multilocus simple sequence repeat (SSR) genotypes in the Andes sample and 43 in the Mexico sample. The Mexico sample had slightly higher mean allelic richness per locus compared with the Andes sample (6.7 vs. 5.2) after correction for sample size by rarefaction. The mean number of private alleles per locus for the Mexico sample was more than twice that for the Andes sample (2.1 vs. 0.8). In the Mexico sample, clonality was detected in isolates sampled from single patches of the indigenous host *S. demissum*. The index of association, I_A, calculated for clone-corrected data for the Mexico isolates accepted the hypothesis of sexual reproduction ($P = 0.25$). In contrast, the hypothesis of no linkage among markers was rejected for the Andes sample ($P < 0.001$), supporting a clonal mode of reproduction.

We inferred population structure based on SSR genotypes separately for the Mexico and Andes samples of *P. infestans*, based on a priori knowledge of their sexual and clonal reproduction, respectively, using *structure* (32). The Mexico sample consisted of admixed subpopulations with at least $K = 4$ underlying groups (Fig. 1*A* and *SI Appendix*, Figs. S1 and S2), whereas the Andes sample consisted of two distinct clusters with very little admixture (Fig. 1*B* and *SI Appendix*, Figs. S1 and S2).

Phylogeographic Root of *P. infestans*. We used Bayesian multilocus phylogeographic analysis to infer the geographic location of the
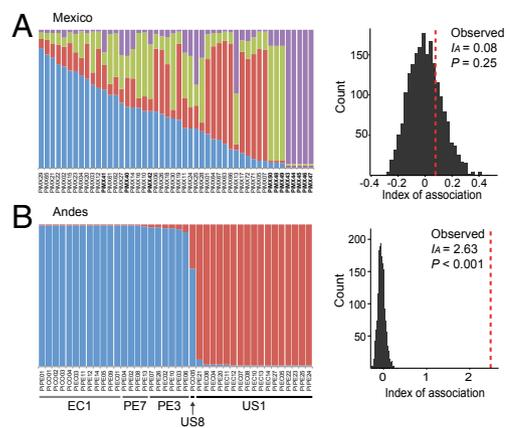

Fig. 1. Population structure of Mexico and Andes samples of *P. infestans* inferred using *structure* (32). (*A*) The Mexico sample shows admixed individuals assigned to $K = 4$ clusters. Isolates collected from patches of *S. demissum* are in bold type. Based on the index of association, I_A, there is no evidence of linkage disequilibrium among loci ($P = 0.25$), consistent with a sexually recombining population. (*B*) The Andes sample clusters into $K = 2$ distinct clades with little or no admixture. The hypothesis of no linkage among markers is rejected ($P < 0.001$), indicating a clonal population.

root ("root state") of *P. infestans* and the *Phytophthora* clade 1c species using BEAST (24, 33). For this analysis, we included a representative global sample including isolates from the now-diverse populations in Europe (*SI Appendix*, Table S1). Comparison of different molecular clocks for each sequenced nuclear locus showed that three of four loci fit a strict clock, in which the standard deviation of the uncorrelated lognormal relaxed molecular clock indicated no variation in rates among branches. In contrast, the RAS locus (intron Ras and Ras fragments) showed high variation in rates among branches and required a relaxed lognormal molecular clock (*SI Appendix*, Table S2). The PITG_11126 locus had the highest rate of evolution, whereas β-tubulin (b-tub) had the lowest substitution rate.

Root state reconstruction produced the highest posterior probabilities for Mexico as the root state of both the *P. infestans* and clade 1c datasets (Fig. 2). For each locus independently, posterior probabilities of a Mexico root were >0.8 for clade 1c (*SI Appendix*, Fig. S3). In nearly all of the *P. infestans* clades, there was an inferred ancestral connection to Mexico (*SI Appendix*, Figs. S4–S8). All of the species in Clade 1c were monophyletic with high support, except for the hybrid species *P. andina* as previously demonstrated (16, 34).

Evolution of *P. infestans* in the Andes. We found that three out of four loci had either greater nucleotide diversity or Watterson's theta for the Andes sample compared with the Mexico sample (*SI Appendix*, Table S3). To better understand the evolution of the Andes lineages and the relationship between *P. infestans* in Mexico and the Andes, we estimated pairwise times since divergence using combined SSR genotypes and nuclear sequence data using the IMa program (25). Divergence times were estimated between each of the clonal lineages in the Andes sample (US-1, EC-1, PE-3, and PE-7) and one another and the Mexico sample (Fig. 3). EC-1, PE-3, and PE-7 produced recent pairwise divergence times. EC-1 showed the most recent divergence from US-1 and the Mexico population. The time since divergence of PE-3 and US-1 was less than that of PE-3 and Mexico. The time since divergence between PE-7 and Mexico was greater than for EC-1 or PE-3 from Mexico. The marginal posterior probability distribution for the divergence time between PE-7 and US-1 was flat, indicating uncertainty in the history of PE-7.

We further explored scenarios for the evolution of the Andes lineages by testing a series of alternative models using ABC as implemented in the DIYABC program (35). We first examined
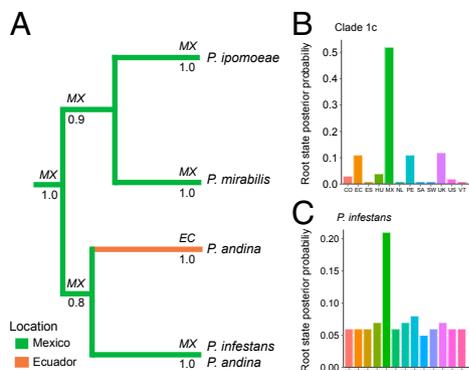
Goss et al.

Fig. 2. Root state probabilities by country inferred using BEAST (33). (*A*) Summarized maximum clade credibility phylogeny of *Phytophthora* clade 1c species. Colors of branches indicate the most probable geographic origin of each lineage. Posterior probabilities of major branches supporting the tree topology are shown below each branch. Root state posterior probabilities indicate that Mexico is the most probable origin of clade 1c (*B*) and *P. infestans* (*C*).

the relationships among the lineages EC-1, PE-3, and PE-7. Model comparison indicated that the PE-3 and PE-7 lineages have more recent shared ancestry compared with the more widespread EC-1 lineage (*SI Appendix*, Fig. S9). We next tested the relationships of each of these lineages to the Toluca population in Mexico and the US-1 lineage in the Andes (*SI Appendix*, Fig. S10). Here we used only isolates from the Toluca Valley, because we know that this population is panmictic (4). Preliminary analyses showed support for the PE-3 and PE-7 lineages being both closely related and distantly related to the other lineages.

To better understand this pattern, we included scenarios containing all possible pairwise admixture events among populations as well as an admixture with an unsampled population (36) (*SI Appendix*, Fig. S10). There was relatively strong support for PE-3 as an admixed lineage. The scenario in which PE-3 is derived from an admixture event between US-1 and an unsampled population had a posterior probability of 0.67 (Fig. 4*A* and *SI Appendix*, Table S4). All other scenarios had posterior probabilities <0.14, most <0.01. The equivalent scenario for PE-7 also had the highest posterior probability of the tested models at 0.37, but there was also support for PE-7 emerging from an admixture event between the Toluca population and the unsampled population (Fig. 4 *B* and *C* and *SI Appendix*, Table S4). The relationship of EC-1 to US-1 and the Toluca population was uncertain as well, with nearly equal support for two scenarios (Fig. 4 *D* and *E* and *SI Appendix*, Table S4). In one scenario, the EC-1 lineage recently diverged from the Toluca population (*P* = 0.25), and in the other, EC-1 was an admixture of the Toluca population and an unsampled population (*P* = 0.31). Our Andes samples were highly clonal; thus, we interpret these results as indicating that the PE-3, PE-7, and perhaps EC-1 lineages formed after a sexual event between two distinct lineages or populations.

We used the most highly supported scenarios to test more complex scenarios including all four Andean lineages and Toluca. Testing all possible admixture events proved to be too complex with support split among scenarios with various combinations of admixture events. Furthermore, the PE-7 results from both IMa and DIYABC analyses suggest that this lineage has a complex ancestry. Therefore, the final scenarios tested admixture in the evolution of EC-1 and PE-3, and excluded PE-7 (Fig. 5). We found that scenario B, in which EC-1 split from the Toluca population and PE-3 originated from an admixture event, had the highest posterior probability of 0.74 (Fig. 4). Notably, there was minimal support for scenario C, with simple ancestral divergence of PE-3 (Fig. 5).

We estimated type I and type II errors for scenario B and found that 65% of the datasets simulated under scenario B produced the highest posterior probability for scenario B (type I error, 0.352). Pseudo-observed datasets generated under scenarios A and C were wrongly assigned to scenario B at frequencies of 0.210 and 0.096, respectively (type II error). Posterior distributions for parameters were wide, indicating limited confidence in the parameter estimates.

Discussion

We found multilocus support for a Mexican origin of *P. infestans*. Bayesian phylogeographic analysis rooted both *P. infestans* and *Phytophthora* clade 1c in Mexico for each of four nuclear loci. Our results are consistent with the population biology of *P. infestans* in central Mexico and, taken together, point to Mexico as the origin of this pathogen. These results are supported by the previously noted pathogen and host characteristics. Specifically, Toluca populations are sexual, whereas South American populations of *P. infestans* are clonal (6, 9, 10, 17–19). Both mating types of *P. infestans* are known to have been present since at least the 1960s in central Mexico (37, 38).

The genealogical connections between continents that we observed in our phylogeographic analysis are consistent with the movement of *P. infestans* among widespread potato growing regions. Migration estimates using microsatellite variation also support our sequence analysis by showing migration from Mexico to the Andes, but not from the Andes to Mexico (*SI Appendix*). The commercial potato seed trade can explain much of the current global population structure of *P. infestans*; however, the early movements of *P. infestans* have not been fully reconstructed (1), and examination of the processes underlying the emergence of new, highly successful strains is ongoing (3, 39). The diverse population in central Mexico may be the ultimate source for the appearance of new strains worldwide (3, 40), although seed potatoes from Europe are behind recent migrations of virulent strains (41, 42). Detailed genetic reconstruction of global migrations of *P. infestans* may be feasible using population genomic data.

We explored the evolution of the Andean lineages and the relationship between *P. infestans* in Mexico and the Andes by testing a series of alternative models using the ABC technique. Our intention was to determine the timing of divergence of the Andean lineages and their relationship to the Toluca population in Mexico. Surprisingly, we found evidence for diversification of the Andes population as a result of admixture or hybridization. Because of our limited power to discriminate between models, we view this analysis as hypothesis-generating. Nevertheless, even moderate support for hybridization generating novel
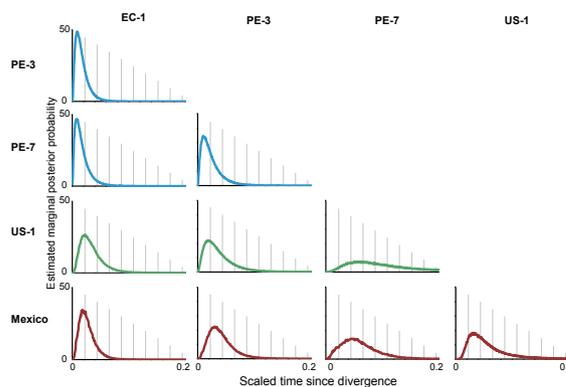
Fig. 3. Estimated marginal posterior probabilities of relative divergence times in pairwise comparisons among Andes clonal lineages (EC-1, PE-3, PE-7, and US-1) and the Mexico population estimated using IMa (64). Smaller means and modes of the scaled time since divergence indicate more recent divergence of lineages. Lineages EC-1, PE-3, and PE-7 show more recent divergence from one another than from US-1 and Mexico.
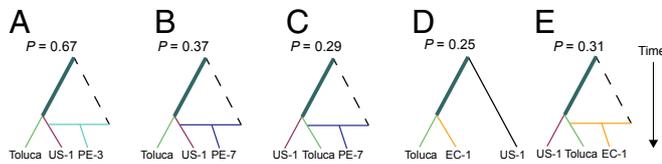
Fig. 4. Scenarios for the evolution of *P. infestans* in the Andes tested using ABC in DIYABC (35). Shown are the scenarios with the highest posterior probabilities out of 15 scenarios testing the relationships of PE-3, PE-7, and EC-1 lineages, in turn, to the US-1 lineage and Toluca Valley population (*SI Appendix*, Fig. S10). Present-day populations are at the base of the tree schematic. Ancestral relationships among these populations are represented by lines intersecting in the past, with the vertex of the schematic representing the most recent common ancestor of all samples. Horizontal lines indicate admixture events between the ancestral populations connected by the horizontal line. Potential changes in population size over time are indicated by changes in line thickness, but line thickness is not proportional to population size. The dashed line represents an unsampled population that has contributed to the genetic variation observed in sampled populations. (*A*) For PE-3, the most probable scenario includes an admixture of US-1 and an unsampled population, leading to the PE-3 lineage. (*B* and *C*) For PE-7, support is split between two scenarios containing an admixture event with an unsampled population. (*D* and *E*) For EC-1, the scenarios with the greatest support showed a simple divergence from the Toluca population or an admixture between Toluca and an unsampled population. Posterior probabilities for all tested scenarios and their 95% confidence intervals are given in *SI Appendix*, Table S4.

lineages of *P. infestans* is compelling, given that a new pathogen of *Solanum* hosts, *P. andina*, has been generated via hybridization (16), and that infrequent hybridization among lineages of *P. infestans* is suspected to be responsible for novel genotypes elsewhere (43).

Considering the diversity of *Solanum* hosts in the Andes, there is great potential for clonal diversification of *P. infestans* as available niches are colonized and rare events contribute to generation or recombination of variation. There is no evidence of genetic variation among *P. infestans* isolates from potatoes in Peru as recently as the mid-1980s (44). If there were diversity in *P. infestans* or clade 1c in the Andes, it must have been limited to other unsampled hosts. Based on our analyses, we hypothesize that the Andean diversity in *P. infestans* has been driven by global migration together with hybridization among populations established via independent migration events.

Given our results, why did Gómez-Alpizar et al. (5) infer an Andean origin? First, their mitochondrial coalescent gene tree included the Andean endemic *P. andina*, but not the Mexican sister species, and thus species selection rooted the tree in the Andes. At the time that the work was conducted, *P. andina* was not recognized as a hybrid species with two haplotypes derived from two distinct parental species (16). When we removed *P. andina* from their dataset, the location of the root was ambiguous (*SI Appendix*); therefore, the mitochondrial loci used in the analysis were not phylogeographically informative. The root of the *P. infestans* mitochondrial genome was recently dated to 460 y ago (95% highest-probability density, 300–643), around the time of the Spanish conquest of the Americas (1). Thus, the two major mitochondrial haplotypes may be the product of movement of the pathogen by humans, resulting in the formation of a new population of *P. infestans* and the evolution of diverged mtDNA haplotypes before global expansion of the pathogen some 200 y later.

Second, the coalescent root of the single nuclear locus used by Gómez-Alpizar et al. is dependent on migration rate (*SI Appendix*), which was estimated using the same locus. The RAS gene in *P. infestans* has two diverged haplotypes in the first intron. This creates two diverged clades of haplotypes, one of which is not found in the Mexican sample. This might have biased the migration rate estimates for this gene and affected the outcome of the coalescent analysis. Our analysis of this locus required a relaxed molecular clock and took an exceptionally long time to converge. Finally,

explicit treatment of geography in our BEAST analysis did not require us to make assumptions about population structure, but rather incorporated the divergent origins of our isolates into the analysis.

Resolving the origin of *P. infestans* is important to our understanding of the emergence and reemergence of damaging pathogens. *P. infestans* is one of a limited number of agricultural plant pathogens with a well-characterized center of origin (45). An expectation of long-term coevolution between a crop and its host-specific pathogen can mislead one into thinking that the pathogen originates from the crop's center of origin; however, there are other well-documented and suspected instances of host-jumping by crop pathogens (45–47). *P. infestans* appears to be an example of a pathogen originating from wild relatives in a secondary center of host diversity. Identification of the center of origin also advances our understanding of pathogen evolution outside of this region, that is, how the pathogen has changed after introduction to new environments.

Global populations of *P. infestans* have undergone rapid changes owing to both migration and evolution after migration (39, 48). Knowledge of pathogen diversity and evolution both within and outside of the center of origin is critical to crop breeding efforts (49). The long-term success of efforts to breed late blight-resistant potatoes will require breeders to account for geographic and evolutionary sources of novel variation in *P. infestans* and its hosts.

Materials and Methods

Sampling Individuals. A sampling of the known global diversity in *P. infestans* was obtained from several collaborators (*SI Appendix*, Table S1). We focused on assembling two large samples from Mexico (n = 48) and the South American Andes (n = 40) representing the known diversity from each of these regions, and included a more moderate global sample of *P. infestans* for context (*SI Appendix*, Table S1).

SSR Typing. Isolates were genotyped using 11 SSR markers as described previously (39, 50). Given that the individuals varied in ploidy, analysis was restricted to approaches that can accommodate nondiploid genetics (51).

Population Structure (SSR). Allelic richness was calculated using the ADZE program (52). Population composition was inferred using *structure* 2.3 (32) by testing the number of population clusters (K) between 1 and 20 using the admixture model for the Mexico and Andes samples (53) and the no-admixture model for the Andes sample. The no-admixture model may be more appropriate for the Andean population (i.e., Ecuador, Colombia, and Peru) given a priori knowledge of its clonality (17). Analysis was performed separately for the two different populations. A total of 10 independent runs each of 1,000,000 iterations with a burn-in period of 20,000 Markov chain Monte Carlo (MCMC) iterations were conducted. The results from *structure* were postprocessed using *Structure Harvester* (54). The ΔK method was used to evaluate the rate of change in the log probability of data between successive K values, to infer the number of clusters (55).
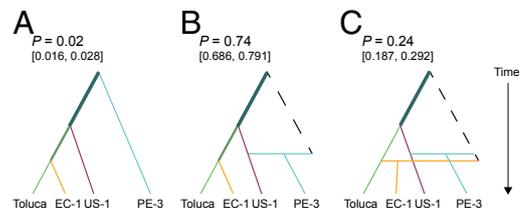

Fig. 5. Final three scenarios used to examine the relationships between the Toluca Valley population and EC-1, PE-3, and US-1 lineages in the Andes by ABC. Tree schematics are drawn as in Fig. 4. The three scenarios are simple divergence of populations such that PE-3 is an ancestral lineage (*A*); emergence of the PE-3 lineage after an admixture of US-1 and an unsampled population (*B*); and scenario B plus emergence of the EC-1 by an admixture between the Toluca population and an unsampled population (*C*). Scenario B was the most likely of the three, with a posterior probability of 0.74. The 95% confidence intervals for the posterior probabilities are in brackets.

The mode of reproduction was assessed by evaluating observed linkage among loci against expected distributions from permutation using the index of association, I_A, as proposed by Brown et al. (56) and applied in *multilocus* (57). I_A was calculated using the *poppr* package (58) in R (59) separately for the Mexico and Andes samples and evaluated with 2,000 permutations using clone-corrected data.

DNA Amplification and Sequencing. Nuclear loci were amplified and directly sequenced as described by Goss et al. (16). A larger fragment of the b-tub gene was sequenced than that sequenced by Gómez-Alpizar et al. (5). Haplotype phases of nuclear gene sequences with multiple heterozygous sites were inferred using PHASE software (51). Inferred haplotypes were confirmed by cloning PCR products for a subset of genotypes and sequencing inserts. In several instances, three distinct alleles were recovered from a given individual; these were the only times when the PHASE-inferred haplotypes were not validated. In some instances, isolates with three alleles at one locus were cloned at another locus, and only two alleles were found; thus, it was not assumed that these isolates had three distinct alleles at other loci.

For the phylogenetic and coalescent-based analyses, we removed indels and recombinant alleles from the datasets. We used the four-gamete test to detect signals of recombination, and found recombination in the b-tub, PITG_11126, and RAS alignments. For PITG_11126, the last 65 nucleotides were excluded owing to recombination in this region relative to the remainder of the fragment. For b-tub, the signal of recombination was removed when an indel was deleted from the alignment and isolate PiEC01 was excluded. For the RAS locus, PiEC07, PiUS11, and PiUS17 were excluded.

The sequences generated have been deposited in GenBank (accession nos. KF979339–KF980878).

Demographic and Phylogeographic History. We adopted a Bayesian coalescent approach to investigate divergence and phylogeographic history using BEAST 1.7.4 (33, 60). BEAST implements a method for sampling all trees that have a reasonable probability given the data. The analysis is based on haplotypes, and two haplotypes in a given individual may or may not share a most common recent ancestor. Furthermore, genealogies obtained from BEAST necessarily estimate how old the common ancestor of these haplotypes might be for each haplotype (even when haplotype sequences are identical), resulting in slight branch length variation among identical haplotypes. This is an expected outcome of coalescent analysis in which identical sequences at a given locus are expected to show divergence at the genome level. Note that although branch length variation is observed among identical haplotypes, there is no support for nodes at this level until different haplotype sequences coalesce to their common recent ancestors, and this branch length variation should not be interpreted.

To initialize BEAST, we obtained the most likely models of nucleotide substitution for each alignment using jModelTest 2.1 (61) and ModelGenerator 0.57 (62), and selected the consensus model from these programs (*SI Appendix*, Table S5). To ensure adequate sampling of parameters, we conducted the BEAST analyses with independent MCMC simulations of 200 million iterations for each locus both for *P. infestans* only and with the other clade 1c species. Analyses were conducted at least twice to ensure repeatability. For each MCMC run, we sampled every 10,000 generations and discarded non-stationary samples after 25% burn-in. Effective sample size estimates were typically greater than 200, and parameter trace plots supported MCMC convergence and good mixing in each independent run. Multilocus analyses were also run for *P. infestans* only and also for all *Phytophthora* clade 1c species. For each dataset, models of nucleotide substitution, constant coalescent priors (for the *P. infestans* dataset) or Yule speciation models (for the clade 1c dataset) were used as specified priors to improve the calculation of clock rates, geographic ancestral states, and phylogenetic relationships.

To infer the most likely geographic origin for each clade, we used the discrete phylogeographic approach as implemented by Lemey et al. (24). This method uses the geographic locations of the samples to reconstruct the ancestral states of tree nodes, including estimation of the root state posterior probability. Maximum clade credibility (MCC) trees were obtained using TreeAnnotator and visualized in FigTree 1.3.1. The bar plots for the posterior probability of the root state were created with the ggplot2 package (63) using R statistical computing and graphic language (59). Maximum clade credibility trees were obtained for each locus as well as for multigene analyses.

Divergence Times Under Isolation with Migration. IMa version 2.0 was used to estimate divergence times for each pair of Andean lineages and between Andean lineages and the Mexican population (64). The data used were the 4 nuclear loci and 8 of the 11 SSR loci. Two SSR loci that did not conform to the stepwise mutation model (D13 and G11) and one SSR locus that was nearly

monomorphic (Pi33) were excluded. The infinite sites model was used for all nuclear loci. Initial maxima for uniform prior distributions of the parameters were as follows: population size (q), 15.0; divergence time (t), 0.5; migration rate (m), 1.0. Runs including the Mexican sample were also performed using larger priors for q (20.0) and t (1.0), and similar results were obtained. The option of running the burn-in and recording periods for indefinite durations was chosen to ensure stabilization of likelihoods before the end of the burn-in and adequate sampling of the posterior distribution. Metropolis coupling was implemented using the geometric increment model and 80 chains with geometrical increment parameters of 0.97 and 0.3.

ABC. We evaluated alternative scenarios for the evolution of the Andean *P. infestans* lineages in a systematic stepwise manner using ABC with DIYABC version 1.0.4.45 (35). ABC has been used to estimate parameters for complex evolutionary models for which likelihoods are difficult or practically impossible to compute (65, 66). The DIYABC program simulates coalescent genealogies under user-specified evolutionary models using parameters drawn from prior distributions and compares statistics summarizing various aspects of the simulated data with those of the observed data. The similarity of summary statistics between observed and simulated datasets is used to calculate posterior probabilities of competing evolutionary models and posterior distributions of parameters. In effect, models that generate datasets with summary statistics close to the observed data have higher probability. We used this method to evaluate posterior probabilities for alternate scenarios representing different possible evolutionary relationships among *P. infestans* populations in Mexico and the Andes. The use of ABC to evaluate alternate scenarios has been questioned owing to potential problems with using summary statistics, but DIYABC incorporates tests that address many of these concerns (67, 68). In particular, confidence in model choice can be evaluated empirically by calculating type I and II errors using pseudo-observed datasets.

To narrow the number of evolutionary scenarios tested against one another, we used four sets of scenarios to inform a final fifth set of scenarios. We grouped the Andes isolates by clonal lineage to tease apart the evolutionary history of each lineage. Isolates were assigned to a clonal lineage based on a previously determined RG-57 fingerprint (3) and confirmed with SSR genotypes grouped using *structure* and *k*-means clustering (69). We first simulated the three different possible relationships among the Andean lineages EC-1, PE-3, and PE-7 (*SI Appendix*, Fig. S9) using the settings in *SI Appendix*, Table S6A. We next examined 15 alternative relationships for the US-1 lineage, the Toluca population, and each of the Andean lineages in turn (*SI Appendix*, Fig. S10 and Table S6B). Although the US-1 lineage has had a global distribution, for these analyses we used only US-1 isolates collected in the Andes, because our focus was on the genetic variation and evolution history of the Andes population. Based on the results of preliminary analyses, these scenarios included admixture events between lineages as well as admixture with an unsampled population.

Finally, we used the scenarios with the highest posterior probabilities to construct three final scenarios (Fig. 4 and *SI Appendix*, Table S6C). This final set tested admixture events generating the EC-1 and PE-3 lineages relative to a scenario with no admixture. We removed the PE-7 lineage from this final analysis, because there were too many plausible admixture events behind the emergence of this lineage.

We tested prior distribution settings with small numbers of simulations by running a principal components analysis on the summary statistics and comparing the overlap of the simulated and observed data. Posterior probabilities and 95% confidence intervals were estimated by polychotomic weighted logistic regression on 1% of the simulations, as implemented in DIYABC. Model checking was conducted for the final three scenarios by estimating type I and type II errors. In short, 500 pseudo-observed datasets (pods) were generated according to each evolutionary scenario using parameters drawn from the same prior distributions as used for the simulated genealogies. The posterior probability of each scenario was calculated for each pod. Type I error was estimated as the proportion of pods for which the correct scenario did not receive the highest posterior probability. Type II error was calculated as the proportion of pods erroneously assigned to a given scenario.

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

2. Haverkort AJ, et al. (2008) Societal costs of late blight in potato and prospects of durable resistance through cisgenic modification. *Potato Res* 51:1871–4528.

3. Goodwin SB, Cohen BA, Fry WE (1994) Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. *Proc Natl Acad Sci USA* 91(24):11591–11595.

4. Grünwald NJ, Flier WG (2005) The biology of *Phytophthora infestans* at its center of origin. *Annu Rev Phytopathol* 43:171–190.

5. Gómez-Alpizar L, Carbone I, Ristaino JB (2007) An Andean origin of *Phytophthora infestans* inferred from mitochondrial and nuclear gene genealogies. *Proc Natl Acad Sci USA* 104(9):3306–3311.

6. Oliva RF, et al. (2010) *Phytophthora andina* sp nov., a newly identified heterothallic pathogen of solanaceous hosts in the Andean highlands. *Plant Pathol* 59(4):613–625.

7. Flier WG, et al. (2002) *Phytophthora ipomoeae*, a new homothallic species causing late blight on *Ipomoea longipedunculata* in the Toluca Valley of central Mexico. *Mycol Res* 106(7):848–856.

8. Goodwin SB, Legard DE, Smart CD, Levy M, Fry WE (1999) Gene flow analysis of molecular markers confirms that *Phytophthora mirabilis* and *P. infestans* are separate species. *Mycologia* 91(5):796–810.

9. Grünwald NJ, et al. (2001) Population structure of *Phytophthora infestans* in the Toluca Valley region of Central Mexico. *Phytopathology* 91(9):882–890.

10. Flier WG, et al. (2003) The population structure of *Phytophthora infestans* from the Toluca Valley in central Mexico suggests genetic differentiation between populations from cultivated potato and wild *Solanum* species. *Phytopathology* 93(4):382–390.

11. Fry WE, et al. (1993) Historical and recent migrations of *Phytophthora infestans*: Chronology, pathways, and implications. *Plant Dis* 77(7):653–661.

12. Niederhauser JS (1991) *Phytophthora infestans*: The Mexican connection. *Phytophthora*, eds Lucas JA, Shattock RC, Shaw DS, Cooke LR (Cambridge Univ Press, Cambridge, UK), pp 25–45.

13. Fry WE, et al. (1992) Population genetics and intercontinental migrations of *Phytophthora infestans*. *Annu Rev Phytopathol* 30:107–130.

14. Hijmans RJ, Spooner DM (2001) Geographic distribution of wild potato species. *Am J Bot* 88(11):2101–2112.

15. Vleeshouwers VGAA, et al. (2011) Understanding and exploiting late blight resistance in the age of effectors. *Annu Rev Phytopathol* 49(1):507–531.

16. Goss EM, et al. (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(9):e24543.

17. Adler NE, et al. (2004) Genetic diversity of *Phytophthora infestans* sensu lato in Ecuador provides new insight into the origin of this important plant pathogen. *Phytopathology* 94(2):154–162.

18. Chacon MG, et al. (2006) Genetic structure of the population of *Phytophthora infestans* attacking *Solanum ochranthum* in the highlands of Ecuador. *Eur J Plant Pathol* 115(2):235–245.

19. Forbes GA, et al. (1997) Population genetic structure of *Phytophthora infestans* in Ecuador. *Phytopathology* 87(4):375–380.

20. Grünwald NJ, Goss EM (2011) Evolution and population genetics of exotic and re-emerging pathogens: Novel tools and approaches. *Annu Rev Phytopathol* 49:249–267.

21. Tavare S (1984) Lines-of-descent and genealogical processes, and their applications in population-genetics models. *Adv Appl Prob* 16(1):471–491.

22. Kingman JFC (1982) The coalescent. *Stochastic Process Appl* 13:235–248.

23. Wakeley J (2009) *Coalescent Theory: An Introduction* (Roberts & Company, Greenwood Village, CO).

24. Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *PLOS Comput Biol* 5(9):e1000520.

25. Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167(2):747–760.

26. Miller N, et al. (2005) Multiple transatlantic introductions of the western corn rootworm. *Science* 310(5750):992.

27. Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA (2007) Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Mol Biol Evol* 24(2):398–411.

28. Lombaert E, et al. (2010) Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS ONE* 5(3):e9743.

29. Dutech C, et al. (2012) The chestnut blight fungus world tour: Successive introduction events from diverse origins in an invasive plant fungal pathogen. *Mol Ecol* 21(16):3931–3946.

30. Cooke DEL, Drenth A, Duncan JM, Wagels G, Brasier CM (2000) A molecular phylogeny of *Phytophthora* and related oomycetes. *Fungal Genet Biol* 30(1):17–32.

31. Blair JE, Coffey MD, Park S-Y, Geiser DM, Kang S (2008) A multi-locus phylogeny for *Phytophthora* utilizing markers derived from complete genome sequences. *Fungal Genet Biol* 45(3):266–277.

32. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.

33. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969–1973.

34. Blair JE, Coffey MD, Martin FN (2012) Species tree estimation for the late blight pathogen, *Phytophthora infestans*, and close relatives. *PLoS ONE* 7(5):e37003.

35. Cornuet JM, Ravigné V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* 11:401.

36. Guillemaud T, Beaumont MA, Ciosi M, Cornuet J-M, Estoup A (2010) Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity (Edinb)* 104(1):88–99.

37. Gallegly ME, Galindo J (1958) Mating types and oospores of *Phytophthora infestans* in nature in Mexico. *Phytopathology* 48:274–277.

38. Flier WG, Grünwald NJ, Fry WE, Turkensteen LJ (2001) Formation, production and viability of oospores of *Phytophthora infestans* isolates from potato and *Solanum demissum* in the Toluca Valley, central Mexico. *Mycol Res* 105(8):998–1006.

39. Cooke DEL, et al. (2012) Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. *PLoS Pathog* 8(10):e1002940.

40. Spielman LJ, et al. (1991) A second world-wide migration and population displacement of *Phytophthora infestans*? *Plant Pathol* 40(3):422–430.

41. Li Y, et al. (2012) Population structure of *Phytophthora infestans* in China: Geographic clusters and presence of the EU genotype Blue_13. *Plant Pathol* 62:932–942.

42. Chowdappa P, et al. (2012) Emergence of 13_A2 blue lineage of *Phytophthora infestans* was responsible for severe outbreaks of late blight on tomato in south-west India. *J Phytopathol* 161:49–58.

43. Goodwin SB, et al. (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(9):939–949.

44. Tooley PW, Therrien CD, Ritch DL (1989) Mating type, race composition, nuclear DNA content, and isozyme analysis of Peruvian isolates of *Phytophthora infestans*. *Phytopathology* 79:478–481.

45. Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annu Rev Phytopathol* 46:75–100.

46. Zaffarano PL, McDonald BA, Linde CC (2008) Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution* 62(6):1418–1436.

47. Couch BC, et al. (2005) Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics* 170(2):613–630.

48. Goodwin SB, Sujkowski LS, Fry WE (1995) Rapid evolution of pathogenicity within clonal lineages of the potato late blight disease fungus. *Phytopathology* 85:669–676.

49. McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annu Rev Phytopathol* 40:349–379.

50. Lees AK, et al. (2006) Novel microsatellite markers for the analysis of *Phytophthora infestans* populations. *Plant Pathol* 55:311–319.

51. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68(4):978–989.

52. Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: Allelic Diversity Analyzer, version 1.0. http://rosenberglab.bioinformatics.med.umich.edu/adze.html.

53. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587.

54. Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359–361.

55. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14(8):2611–2620.

56. Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* 96(2):523–536.

57. Agapow P-M, Burt A (2001) Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* 1(1-2):101–102.

58. Kamvar ZN, Tabima JF, Grünwald NJ (2014) *Poppr*: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281.

59. R Development Core Team (2013) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

60. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.

61. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: More models, new heuristics and parallel computing. *Nat Methods* 9(8):772.

62. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29.

63. Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York), p 212.

64. Hey J (2010) Isolation with migration models for more than two populations. *Mol Biol Evol* 27(4):905–920.

65. Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annu Rev Ecol Evol Syst* 41(1):379–406.

66. Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol* 25(7):410–418.

67. Robert CP, Cornuet JM, Marin JM, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. *Proc Natl Acad Sci USA* 108(37):15112–15117.

68. Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Mol Ecol* 19(13):2609–2625.

69. Meirmans P, Van Tienderen P (2004) Genotype and Genodive: Two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes* 4:792–794.

Goss et al.