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

Citation	Goss, E. M., Tabima, J. F., Cooke, D. E. L., Restrepo, S., Fry, W. E., Forbes, G. A., ... & Grünwald, N. J. (2014). The Irish potato famine pathogen Phytophthora infestans originated in central Mexico rather than the Andes. Proceedings of the National Academy of Sciences, 111(24). doi:10.1073/pnas.1401884111
DOI	10.1073/pnas.1401884111
Publisher	National Academy of Sciences
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

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

Edited by Detlef Weigel, Max Planck Institute for Developmental Biology, Tübingen, Germany, and approved May 6, 2014 (received for review January 30, 2014)

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.

Author contributions: E.M.G., D.E.L.C., S.R., G.A.F., and N.J.G. designed research; E.M.G., J.F.T., D.E.L.C., S.R., W.E.F., G.A.F., V.J.F., M.C., and N.J.G. performed research; D.E.L.C., S.R., W.E.F., G.A.F., and N.J.G. contributed new reagents/analytic tools; E.M.G., J.F.T., D.E.L.C., V.J.F., M.C., and N.J.G. analyzed data; and E.M.G., J.F.T., D.E.L.C., S.R., W.E.F., G.A.F., M.C., and N.J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KF979339–KF980878).

[1]To whom correspondence should be addressed. E-mail: grunwaln@science.oregonstate.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1401884111/-/DCSupplemental.

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 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.

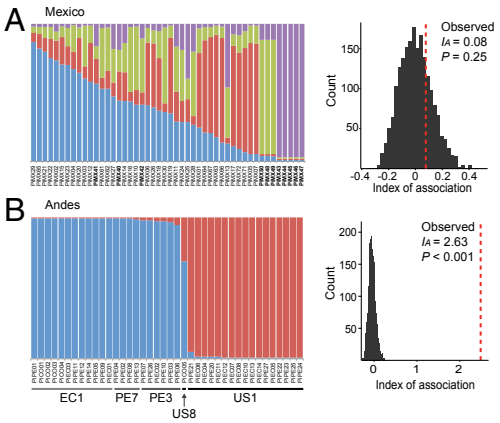


Fig. 1. Population structure of Mexico and Andes samples of *P. infestans* inferred using *structure* (32). (*A*) The Mexico sample shows admixed individuals assigned to $K = 4$ clusters. Isolates collected from patches of *S. demissum* are in bold type. Based on the index of association, I_A, there 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 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

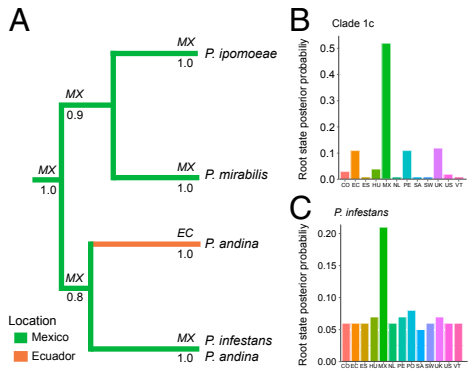


Fig. 2. Root state probabilities by country inferred using BEAST (33). (*A*) Summarized maximum clade credibility phylogeny of *Phytophthora* clade 1c species. Colors of branches indicate the most probable geographic origin of each lineage. Posterior probabilities of major branches supporting the tree 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

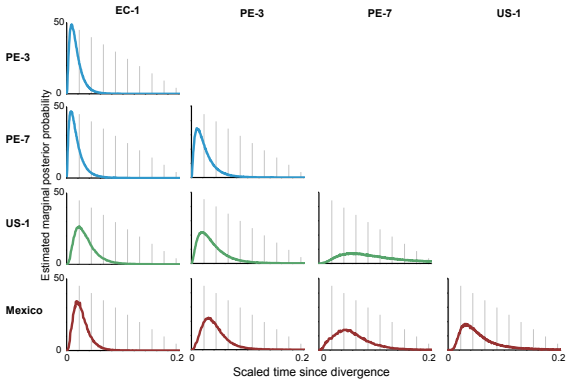


Fig. 3. Estimated marginal posterior probabilities of relative divergence times in pairwise comparisons among Andes clonal lineages (EC-1, PE-3, PE-7, and US-1) and the Mexico population estimated using IMa (64). Smaller means and modes of the scaled time since divergence indicate more recent divergence of lineages. Lineages EC-1, PE-3, and PE-7 show more recent divergence from one another than from US-1 and Mexico.

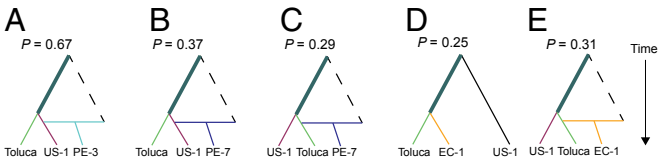


Fig. 4. Scenarios for the evolution of *P. infestans* in the Andes tested using ABC in DIYABC (35). Shown are the scenarios with the highest posterior probabilities out of 15 scenarios testing the relationships of PE-3, PE-7, and EC-1 lineages, in turn, to the US-1 lineage and Toluca Valley population (*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).

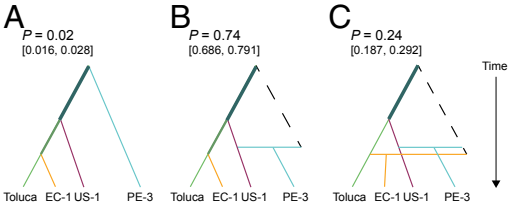


Fig. 5. Final three scenarios used to examine the relationships between the Toluca Valley population and EC-1, PE-3, and US-1 lineages in the Andes by ABC. Tree schematics are drawn as in Fig. 4. The three scenarios are simple divergence of populations such that PE-3 is an ancestral lineage (*A*); emergence 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 nonstationary 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.

ACKNOWLEDGMENTS. We are grateful to the many colleagues who provided isolates for this study. We thank Karan Fairchild, Kevin Myers, Caroline Press, and Naomi Williams for maintenance of the cultures and general technical support. This research is supported in part by US Department of Agriculture (USDA) Agricultural Research Service Grant 5358-22000-039-00D, USDA National Institute of Food and Agriculture Grant 2011-68004-30154 (to N.J.G.), and the Scottish Government (D.E.L.C.).

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 *Ipomoeae 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.

Supporting Information

Supporting Methods

Genetree analysis. We revisited the coalescent analysis of Gómez-Alpizar *et al*. (1) using their data and ours to investigate the sensitivity of root inference to estimates of migration rates. Our concern was that there is little power to estimate migration rates with only a single locus, yet these rates were used to infer the rooting of structured coalescent trees in the program Genetree (2). Furthermore, Gómez-Alpizar *et al*. considered migration to and from South America, which confounds migration from Mexico and other global populations. *P. infestans* is transported transcontinentally and intercontinentally via infected seed tubers, and there is a vibrant commercial seed tuber trade from Northwest Europe with known pathways of migration from Europe to South America.

We revised the Genetree analysis of the P3 and P4 mtDNA regions using sequences from Gómez-Alpizar *et al.* (1) for isolates from Mexico and the Andes but removing the haplotypes representing the *P. andina* isolates (haplotype 1c). The topology and rooting of the coalescent tree for the *P. infestans* haplotypes remained the same. We inferred the maximum likelihood value of theta and a symmetric migration rate for the revised data set. We then used 1×10^7 simulations to examine the inferred subpopulation of the most common recent ancestor (MRCA) of the sample while varying theta and symmetric migration rates using three sets of runs, each set using a different starting seed.

For the RAS nuclear gene, we repeated the Gómez-Alpizar *et al*. (1) analysis while varying migration rates. We used their data set and value of theta. We conducted two runs of 1×10^7 simulations for each set of migration rates. We varied migration rates from low and symmetric to the values used by Gómez-Alpizar *et al*. We also examined the coalescent history of the RAS locus using our sequences from Mexico and the Andes. The tree was rooted using ancestral states obtained from *P. ipomoeae*, *P. mirabilis*, and *P. phaseoli*. We used four runs of 1×10^6 or 1×10^7 simulations and three symmetric migration rates using two different values of theta (1.5 and 2.0) and four starting seeds.

Migration scenarios from SSR genotypes**.** The program Migrate version 3.3 was used to examine recent migration between the Andes and Mexico, using SSR genotypes. Three migration models were compared (3): bidirectional migration, migration from Mexico to the Andes only, and migration from the Andes to Mexico only. The runs used Brownian motion approximation for the stepwise mutation model, initial parameter values from F_{ST}, and mutation rates per locus estimated from the data. Bayesian inference across 10 replicates used slice sampling, uniform prior distributions from 0 to 1000 for both theta and M parameters, and four chains (temperatures: 1,000,000, 3.0, 1.5, 1.0) in which the cold chain was sampled at 10,000

steps at 200 step increments after a burnin of 500,000 steps. Models were evaluated using Bezier approximation of log marginal maximum likelihoods.

Supporting Results

Dependence of Genetree results on migration rates**.** When we removed *P. andina* from the Gómez-Alpizar *et al*. mitochondrial data set and included only the Mexican and South American isolates, the root location was uncertain such that there was equal probability of rooting in Mexico or South America when migration rates were symmetric (Table S7A). Asymmetric migration rates increased the probability of the root in the population that was the source of migration to around 0.95, irrespective of which population was assigned the higher emigration rate.

For the RAS locus, the location of the coalescent root was dependent on migration rates as well (Table S7B). When we used the Gómez-Alpizar *et al*. data under low symmetric migration rates the rooting was uncertain, and asymmetric migration rates affected the inferred population of origin. The migration rates used by Gómez-Alpizar *et al.* produced a high probability for a South American root, which replicates their result. The opposite condition produced a similarly high probability for a non-South American root. Therefore, both mitochondrial and RAS analyses were highly dependent on choice of migration parameters.

We used our data set to examine the root location of RAS for Mexican and Andean isolates only. Using RAS sequences from *P. andina*, *P. mirabilis*, *P. ipomoeae*, and *P. phaseoli*, we were able to unambiguously assign ancestral states to each segregating site and these ancestral states were used to root the *P. infestans* RAS tree. The coalescent history of the gene was simulated on the rooted topology to infer the location of the root in Mexico or the Andes. Moderate and inconclusive probabilities for root location were observed under nearly all conditions examined, including three symmetric migration rates and two values of theta (Table S8).

Key to the Gómez-Alpizar *et al*. analysis was their finding of a higher migration rate from South America to non-South American populations (i.e. Mexico, USA, and Ireland) than vice-versa using the nuclear RAS locus. Given this result, we used our SSR data to test among models of migration between the Andes and Mexico using the program Migrate (3). Multiple loci are expected to provide more robust estimates of migration rates for the nuclear genome than can be obtained from a single locus. This analysis revealed a higher migration rate from Mexico to the Andes than from the Andes to Mexico (M = 11.7 vs. 0.3, respectively). Bayesian comparison of migration models strongly supported unidirectional migration from Mexico to the Andes (Bezier approximation of the log marginal likelihood = -11,055) compared to the opposite scenario of migration from the Andes to Mexico (-12,243) or bidirectional migration (-52,357). This result is consistent with the known directions of trade in commercial seed tubers.

References Cited

1. 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 U S A* 104(9):3306-3311.
2. Griffiths RC & Tavaré S (1994) Ancestral inference in population genetics. *Statistical science* 9(3):307-319.
3. Beerli P & Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* 185(1):313-U463.
4. Nei M (1987) *Molecular Evolutionary Genetics* (Columbia University Press, New York) p 512
5. Watterson GA (1975) On the number of segregating sites in genetic models without recombination. *Theoretical Population Biology* 7:256-276.
6. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585-595.
7. Fu Y-X & Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics* 133(3):693-709.

Supporting Tables

Table S1. Isolates used in the study.

ID	Original name	Year sampled	Country	State	Host	Source[1]	Lineage
P. infestans							
PiCO01	1011		Colombia	Cundinamarca	*S. tuberosum*	Restrepo	EC-1
PiCO02	1063		Colombia	Cundinamarca	*S. tuberosum*	Restrepo	EC-1
PiCO03	1064		Colombia	Cundinamarca	*S. tuberosum*	Restrepo	EC-1
PiCO04	1068		Colombia	Cundinamarca	*S. tuberosum*	Restrepo	EC-1
PiCO05	4084		Colombia	Cundinamarca	*Physalis peruviana*	Restrepo	US-8
PiEC01	EC_3843	2004	Ecuador		potato	Forbes	EC-1
PiEC02	EC_3527	2002	Ecuador		*S. andreanum*	Forbes	EC-1
PiEC03	EC_3626	2003	Ecuador		potato	Forbes	EC-1
PiEC06	EC_3841	2004	Ecuador		*S. habrochaites*	Forbes	US-1
PiEC07	EC_3921	2006	Ecuador		*S. jugandifolium*	Forbes	US-1
PiEC08	EC_3774	2004	Ecuador		*S. ochanthum*	Forbes	US-1
PiEC10	EC_3378	2001	Ecuador		*S. lycopersicum*	Forbes	US-1
PiEC11	EC_3381	2001	Ecuador		*S. lycopersicum*	Forbes	US-1
PiEC12	EC_3150	1997	Ecuador		*S. muricatum*	Forbes	US-1
PiEC13	EC_3520	2002	Ecuador		*S. muricatum*	Forbes	US-1
PiEC14	EC_3809	2004	Ecuador		*S. caripense*	Forbes	US-1
PiES01	2005_10	2005	Estonia		potato	Fry	
PiES02	2005_17	2005	Estonia		potato	Fry	
PiES03	2005_19	2004	Estonia		potato	Fry	
PiES04	2004_4	2004	Estonia		potato	Fry	
PiES05	2004_16	2004	Estonia		potato	Fry	
PiHU02	Josze_S32		Hungary		potato	Bakonyi	

PiMX03	MX980211	1998	Mexico		potato	Fry
PiMX04	MX980230	1998	Mexico		potato	Fry
PiMX05	MX980317	1998	Mexico		potato	Fry
PiMX06	MX980352	1998	Mexico		potato	Fry
PiMX07	MX980400	1998	Mexico		potato	Fry
PiMX10	PIC97008	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX11	PIC97066	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX12	PIC97106	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX13	PIC97111	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX14	PIC97130	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX15	PIC97136	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX16	PIC97146	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX17	PIC97149	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX18	PIC97153	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX19	PIC97159	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX20	PIC97187	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX21	PIC97310	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX22	PIC97318	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX23	PIC97335	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX24	PIC97340	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX25	PIC97389	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX26	PIC97392	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX27	PIC97423	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX28	PIC97432	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX29	PIC97438	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX30	PIC97442	1997	Mexico	Toluca	potato	Flier/Grünwald/PRI
PiMX40	PIC97716	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI
PiMX41	PIC97724	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI
PiMX42	PIC97727	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI
PiMX43	PIC97744	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI
PiMX44	PIC97748	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI

PiMX45	PIC97749	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI	
PiMX46	PIC97750	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI	
PiMX47	PIC97751	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI	
PiMX48	PIC97785	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI	
PiMX49	PIC97791	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI	
PiMX50	PIC97793	1997	Mexico	Toluca	*S. demissum*	Flier/Grünwald/PRI	
PiMX61	Tlax 701	2007	Mexico	Tlaxcala	potato	Fernández Pavia	
PiMX62	Tlax 715	2007	Mexico	Tlaxcala	potato	Fernández Pavia	
PiMX63	Tlax 722	2007	Mexico	Tlaxcala	potato	Fernández Pavia	
PiMX64	Tlax 728	2007	Mexico	Tlaxcala	potato	Fernández Pavia	
PiMX65	Tlax 740	2007	Mexico	Tlaxcala	potato	Fernández Pavia	
PiMX66	Tlax 748	2007	Mexico	Tlaxcala	potato	Fernández Pavia	
PiMX67	Tlax 756	2007	Mexico	Tlaxcala	potato	Fernández Pavia	
PiMX71	T48	2003	Mexico	Tlaxcala	potato	Fernández Pavia	
PiMX72	T68	2003	Mexico	Tlaxcala	potato	Fernández Pavia	
PiNL01	NL_01096	2001	Netherlands		potato	Kessel	
PiNL02	NL_96259	1996	Netherlands		potato	Kessel	
PiPE01	BTLM 004	1997	Peru	Lima	NA	Forbes	PE-7
PiPE02	PHU 076	2003	Peru	Huánuco	potato	Forbes	EC-1
PiPE03	PHU 079	2003	Peru	Huánuco	potato	Forbes	EC-1
PiPE04	PPI 015	2000	Peru	Piura	*S. huancabambense*	Forbes	EC-1
PiPE05	PTS 031	1998	Peru	Cajamarca	*S. caripense*	Forbes	EC-1
PiPE06	1696	1995	Peru	Arequipa	potato	Forbes	PE-3
PiPE07	PCA 004	1999	Peru	Cajamarca	potato	Forbes	PE-3
PiPE08	PCZ 024	1997	Peru	Cuzco	potato	Forbes	EC-1
PiPE09	PCZ 080	1997	Peru	Cuzco	potato	Forbes	EC-1
PiPE10	PSR 001	2005	Peru	Junín	NA	Forbes	EC-1
PiPE11	PCA 020	1999	Peru	Cajamarca	NA	Forbes	PE-7
PiPE12	PLI 003	1999	Peru	Lima	NA	Forbes	PE-7
PiPE13	PLI 036	2000	Peru	Lima	*S. wittmackii*	Forbes	PE-7

PiPE14	POX 100	2003	Peru	Pasco	potato	Forbes	PE-7
PiPE20	PCA 025	1999	Peru	Cajamarca	NA	Forbes	US-1
PiPE21	PPI 009	2000	Peru	Piura	*S. caripense*	Forbes	US-1
PiPE22	PPI 013	2000	Peru	Piura	*S. caripense*	Forbes	US-1
PiPE23	PPI 014	2000	Peru	Piura	*S. caripense*	Forbes	US-1
PiPE24	PPI 023	2000	Peru	Piura	*S. caripense*	Forbes	US-1
PiPE25	PPI 028	2000	Peru	Piura	*S. caripense*	Forbes	US-1
PiPE26	PPU 048	1997	Peru	Puno	potato	Forbes	PE-3
PiPE27	PPU 097	1997	Peru	Puno	potato	Forbes	US-1
PiPE28	PCA 006	1999	Peru	Cajamarca	potato	Forbes	PE-3
PiPE29	PCA 010	1999	Peru	Cajamarca	potato	Forbes	PE-3
PiPO01	MP_618	2005	Poland		potato	Lebecka	
PiPO02	MP_622	2005	Poland		potato	Lebecka	
PiSA01	SA960008	1996	South Africa		potato	Fry	US-1
PiSW01	SE_03058	2003	Sweden		potato	Andersson	
PiSW02	SE_03087	2003	Sweden		potato	Andersson	
PiUK01	2006_3984C	2006	UK		potato	Cooke	EU_1_A1
PiUK02	2006_4012F	2006	UK		potato	Cooke	EU_3_A2
PiUK03	2006_3928A	2006	UK		potato	Cooke	EU_13_A2
PiUK04	2006_4132B	2006	UK		potato	Cooke	EU_13_A2
PiUK05	2006_3888A	2006	UK		potato	Cooke	EU_2_A1
PiUK06	2007_5866B	2007	UK		potato	Cooke	EU_5_A1
PiUK07	2006_4388D	2006	UK		potato	Cooke	EU_17_A2
PiUK08	2006_4100A	2006	UK		potato	Cooke	EU_6_A1
PiUK09	2006_4440C	2006	UK		potato	Cooke	EU_10_A2
PiUK10	2006_4232E	2006	UK		potato	Cooke	EU_8_A1
PiUS08	US040009	2004	USA		potato	Fry	US-8
PiUS11	US050007	2005	USA		tomato	Fry	US-11
PiUS12	US940494	1994	USA		tomato	Fry	US-12
PiUS17	US970001	1997	USA		tomato	Fry	US-17
PiVT01	Vn02-076	2002	Vietnam		tomato	Le	US-1

PiVT02	Vn02-106	2002	Vietnam	tomato	Le	US-1
PiVT03	Vn03-416	2003	Vietnam	potato	Le	US-1
PiVT04	Vn03-590	2003	Vietnam	potato	Le	US-1
P. mirabilis						
PmMX01	CBS 136.86		Mexico	NA	CBS	
PmMX02	CBS 678.85		Mexico	NA	CBS	
PmMX03	DF 409		Mexico	NA	Fernández Pavia	
PmMX04	G 11-3	1998	Mexico	NA	Flier/Grünwald/PRI	
PmMX05	00M 410	2000	Mexico	NA	Fernández Pavia	
PmMX06	P 3001		Mexico	NA	Flier/Grünwald/PRI	
PmMX07	P 3006		Mexico	NA	Flier/Grünwald/PRI	
PmMX08	P mirabilis DF 07	2007	Mexico	NA	Fernández Pavia	
PmMX09	P. mirabilis Mich	2003	Mexico	NA	Fernández Pavia	
PmMX10	WF014/PIC99114	1999	Mexico	NA	Flier/Grünwald/PRI	
PmMX11	WF035/PIC99135	1999	Mexico	NA	Flier/Grünwald/PRI	
P. ipomoeae						
PoMX01	00Ip5	2000	Mexico	NA	Fernández Pavia	
PoMX02	Ipom1-2	1999	Mexico	NA	Flier/Grünwald/PRI	
PoMX03	Ipom2-1	1999	Mexico	NA	Flier/Grünwald/PRI	
PoMX04	Ipom2-4	1999	Mexico	NA	Flier/Grünwald/PRI	
PoMX05	Ipom3-3	1999	Mexico	NA	Flier/Grünwald/PRI	
PoMX06	Ipom6	1999	Mexico	NA	Flier/Grünwald/PRI	
PoMX07	P. ipomoeae	1999	Mexico	NA	Fernández Pavia	
P. andina						
PaEC01	EC_3189	1998	Ecuador	*Anarrichomenum*	Forbes	
PaEC02	EC_3399	2001	Ecuador	*Anarrichomenum*	Cooke	
PaEC03	EC_3818	2004	Ecuador	*Anarrichomenum*	Cooke	
PaEC04	EC_3821	2004	Ecuador	*Anarrichomenum*	Forbes	
PaEC05	EC_3780	2004	Ecuador	*S. hispidum*	Forbes	
PaEC06	EC_3655	2003	Ecuador	*S. hispidum*	Forbes	

PaEC07	EC_3163	1998	Ecuador		*Anarrichomenum*	Forbes
PaEC08	EC_3510	2002	Ecuador		*S. betaceum*	Forbes
PaEC09	EC_3540	2002	Ecuador		*Anarrichomenum*	Forbes
PaEC10	EC_3561	2002	Ecuador		*S. quitoense*	Forbes
PaEC11	EC_3563	2002	Ecuador		*S. quitoense*	Forbes
PaEC12	EC_3678	2003	Ecuador		*Anarrichomenum*	Forbes
PaEC13	EC_3836	2004	Ecuador		*S. betaceum*	Forbes
PaEC14	EC_3860	2005	Ecuador		*Torva*	Forbes
PaEC15	EC_3864	2005	Ecuador		*Torva*	Forbes
PaEC16	EC_3865	2005	Ecuador		*S. jugandifolium*	Forbes
PaEC17	EC_3936	2006	Ecuador		*S. ochanthum*	Forbes
PaPE01	POX 102	2003	Peru		*S. betaceum*	Forbes
PaPE02	POX 103	2003	Peru		*S. betaceum*	Forbes
P. phaseoli						
PpUS01	CBS 556.88					CBS
PpUS02	P10150		USA	Delaware		Coffey

[1] Isolates were contributed by the authors and the following colleagues: Sampled by W. G. Flier and N. J. Grünwald and curated by Geert Kessel, Plant Research International (PRI), Netherlands; Geert Kessel, Plant Research International (PRI), Netherlands; Silvia Fernández Pavia, Universidad Michoacana de San Nicolás de Hidalgo, Mexico; Michael Coffey, University of California Riverside, USA; Vihn Hong Le and Arne Hermansen, Norwegian Institute for Agricultural and Environmental Research, Norway; Björn Andersson, Swedish University of Agricultural Sciences, Sweden; Renata Lebecka Młochow Research Centre, Poland; Jozsef Bakonyi, Academy of Agricultural Sciences, Hungary.

Table S2. Molecular clock and mutation rates obtained in this study for each locus for (A) Clade 1c and (B) *P. infestans* datasets. See Table S3 for indices of nucleotide variation at each locus.

A. Clade 1c

Locus	**Clock**	**Clock Rate**	**UCLD stdev**
β-tubulin	Strict	0.824	0.04
RAS	Log Normal	1.839	2.2
Trp1	Strict	1.73	0.03
PITG_11126	Strict	4.09	0.02

B. *P. infestans*

Locus	**Clock**	**Clock Rate**	**UCLD stdev**
β-tubulin	Strict	0.54	0.02
RAS	Log Normal	2.25	1.94
Trp1	Strict	1.73	0.012
PITG_11126	Strict	3.03	0.01

Table S3. Nucleotide variation by locus, species, and sampling location. Statistics given are number of individuals (N_{ind}), number of sequences (N_{seq}), length of alignment excluding gaps (L), segregating sites (S), number of heterozygous sites (Het), number of haplotypes (Hap), average pairwise nucleotide diversity (π), Watterson's theta (θ_w), Tajima's D, and Fu and Li's D* and F* (4-7).

Locus	Species/Popn	N_{ind}	N_{seq}	L	S	Het	Hap	π	θ_W	Tajima's D	Fu & Li's D*	Fu & Li's F*
PITG_11126												
	P. infestans	119	242	764	13	13[#]	10	0.00286	2.144	0.04	0.79	0.61
	Europe	22	46	764	7	7[#]	5	0.00294	1.593	1.10	1.25	1.41
	Mexico	48	96	764	9	9[#]	6	0.00154	1.752	-0.82	-0.17	-0.47
	S. America	40	82	764	11	11[#]	8	0.00394	2.210	0.97	0.79	1.01
	United States	4	8	779	2	2	2	0.00138	0.771	1.45	1.11	1.30
	Africa/Asia	5	10	779	3	3	2	0.00214	1.060	2.06*	1.15	1.53
	P. andina	19	38	748	16	16[#]	3	0.01040	3.808	3.39***	1.58**	2.56**
	P. ipomoeae	7	14	787	2	1	3	0.00052	0.629	-0.96	-0.45	-0.66
	P. mirabilis	11	22	778	11	11	3	0.00593	4.610	1.83	1.44*	1.80**
	P. phaseoli	2	4	779	0	0	1	0	0	NA	NA	NA
β-tubulin												
	P. infestans	118	238	1590	9	9[#]	8	0.00138	1.488	1.06	-0.43	0.13
	Europe	22	44	1590	6	6[#]	3	0.00061	1.379	-0.79	1.18	0.67
	Mexico	48	96	1590	7	7[#]	3	0.00185	1.363	2.75**	1.20	2.03**
	S. America	39	80	1590	9	9[#]	7	0.00113	1.817	-0.03	-0.84	-0.67
	United States	4	8	1590	5	5[#]	3	0.00124	1.928	0.08	0.75	0.65
	Africa/Asia	5	10	1590	0	0	1	0	0	NA	NA	NA
	P. andina	19	38	1590	23	23	2	0.00743	5.474	3.92***	1.70**	2.89**
	P. ipomoeae	7	14	1590	2	2	2	0.00018	0.629	-1.48	-1.83	-1.97
	P. mirabilis	11	22	1590	16	15[#]	6	0.00469	4.389	2.54**	1.54**	2.14**
	P. phaseoli	2	4	1590	0	0	1	0	0	NA	NA	NA
Trp1												
	P. infestans	117	234	813	9	7	9	0.00183	1.492	-0.004	0.42	0.32
	Europe	22	44	813	4	3	4	0.00185	0.920	1.48	-0.06	0.47
	Mexico	48	96	813	6	4	5	0.00144	1.168	-0.20	1.12	0.81
	S. America	38	76	813	6	6	6	0.00211	1.224	0.95	0.22	0.54
	United States	4	8	813	4	4	4	0.00180	1.543	-0.22	-0.18	-0.21
	Africa/Asia	5	10	813	3	3	2	0.00205	1.060	2.06*	1.15	1.53
	P. andina	19	38	813	10	10	3	0.00293	2.380	3.48***	1.40*	2.42**
	P. ipomoeae	7	14	812	0	0	1	0	0	NA	NA	NA
	P. mirabilis	11	22	813	1	1	2	0.00011	0.274	-1.16	-1.57	-1.68
	P. phaseoli	2	4	813	0	0	1	0	0	NA	NA	NA
RAS												
	P. infestans	117	234	762	15	15[#]	13	0.00364	2.487	0.290	0.311	0.364
	Europe	21	42	766	10	10	6	0.00569	2.324	2.585*	1.40	2.096**
	Mexico	47	94	766	11	11[#]	8	0.00164	2.150	-1.09	-2.51*	-2.39*
	S. America	40	80	762	10	10	5	0.00377	2.019	1.12	1.38	1.53
	United States	4	8	766	11	11	4	0.00643	4.242	0.808	1.52**	1.50
	Africa/Asia	5	10	766	1	1	2	0.00205	1.060	2.06*	1.15	1.53
	P. andina	17	34	765	23	23[#]	4	0.01433	5.625	3.28***	1.39	2.37**
	P. ipomoeae	7	14	768	3	0	2	0.00103	0.943	-0.49	1.07	0.76
	P. mirabilis	10	20	766	6	6[#]	3	0.00091	1.973	-2.13*	-3.08**	-3.25**
	P. phaseoli	1	2	764	0	0	1	0	0	NA	NA	NA

[#]Also heterozygous for one or more indel.
* $P < 0.05$
** $P < 0.02$
*** $P < 0.001$

Table S4. Results from testing relationships among Toluca population, US1 and Andean lineages as illustrated in Figure S10.

Andean lineage	Scenario	Admixed population	Probability	95% CI
PE3	1	-	0.0040	[0.0028,0.0052]
	2	-	0.0071	[0.0058,0.0084]
	3	-	0.0066	[0.0055,0.0077]
	4	PE3	0.0014	[0.0010,0.0017]
	5	Toluca	0.0062	[0.0051,0.0073]
	6	US1	0.0009	[0.0008,0.0011]
	7	**PE3**	**0.6709**	**[0.6478,0.6939]**
	8	PE3	0.1094	[0.0994,0.1194]
	9	US1	0.0023	[0.0011,0.0034]
	10	US1	0.0052	[0.0034,0.0070]
	11	Toluca	0.0020	[0.0009,0.0032]
	12	Toluca	0.1302	[0.1093,0.1511]
	13	PE3	0.0474	[0.0410,0.0537]
	14	US1	0.0011	[0.0004,0.0017]
	15	Toluca	0.0053	[0.0034,0.0073]
PE7	1	-	0.0089	[0.0055,0.0124]
	2	-	0.0327	[0.0265,0.0389]
	3	-	0.0141	[0.0107,0.0176]
	4	PE7	0.0038	[0.0025,0.0051]
	5	Toluca	0.0093	[0.0073,0.0112]
	6	US1	0.0191	[0.0148,0.0234]
	7	**PE7**	**0.3656**	**[0.3393,0.3919]**
	8	**PE7**	**0.2893**	**[0.2656,0.3131]**
	9	US1	0.0016	[0.0003,0.0030]
	10	US1	0.0248	[0.0145,0.0350]
	11	Toluca	0.0121	[0.0032,0.0211]
	12	Toluca	0.0880	[0.0673,0.1087]
	13	PE7	0.1242	[0.1069,0.1416]
	14	US1	0.0012	[0.0003,0.0022]
	15	Toluca	0.0050	[0.0021,0.0079]
EC1	**1**	**-**	**0.2524**	**[0.2335,0.2713]**
	2	-	0.0293	[0.0260,0.0326]
	3	-	0.0575	[0.0519,0.0631]
	4	EC1	0.0908	[0.0830,0.0987]
	5	Toluca	0.0843	[0.0773,0.0913]
	6	US1	0.0216	[0.0192,0.0239]
	7	EC1	0.0320	[0.0276,0.0364]
	8	**EC1**	**0.3060**	**[0.2811,0.3309]**
	9	US1	0.0071	[0.0053,0.0088]
	10	US1	0.0234	[0.0188,0.0280]
	11	Toluca	0.0177	[0.0144,0.0210]
	12	Toluca	0.0140	[0.0116,0.0163]
	13	EC1	0.0088	[0.0069,0.0106]
	14	US1	0.0517	[0.0420,0.0614]
	15	Toluca	0.0035	[0.0026,0.0043]

Table S5. Nucleotide substitution model for each locus as inferred from jModelTest and ModelGenerator. The appropriate most similar model available in BEAST was used for each analysis.

A. Clade 1c

Locus	**jModelTest**	**ModelGenerator**	**Model Used**
β-tubulin	HKY	HKY	HKY
RAS	JC	JC	JC
Trp1	K80	HKY	HKY
PITG_11126	HKY	HKY	HKY

B. *P. infestans*

Locus	**jModelTest**	**ModelGenerator**	**Model Used**
β-tubulin	HKY	HKY	HKY
RAS	F81	JC	JC
Trp1	HKY	HKY	HKY
PITG_11126	HKY	HKY	HKY

Table S6. Settings used for DIYABC analyses. All runs used the same summary statistics: within population statistics were number of segregating sites, mean of pairwise differences, variance of pairwise differences, and number of private segregating sites; between population statistics were number of segregating sites, mean of pairwise differences, and F_{ST}. Posterior probabilities of scenarios were calculated using the closest 1% of simulated datasets using logistic regression.

A. Prior distributions for scenarios in Figure S9.

Parameter		**Shape**	**Min.–Max.**	**Increment**
Population size				
	EC-1	Uniform	1–200000	1
	PE-3	Uniform	1–200000	1
	PE-7	Uniform	1–200000	1
	Ancestor of two lineages	Uniform	1–500000	1
	Ancestor of three lineages	Uniform	1–1000000	1
Time since divergence[1]				
	t1: two lineages	Uniform	1–300000	1
	t2: all three lineages	Uniform	1–500000	1
Nucleotide sequence evolution[2]				
	Mean mutation rate	Uniform	1.00×10^{-10} – 1.00×10^{-8}	
	Gamma distribution	Uniform	1.00×10^{-11} – 1.00×10^{-7}	2.00

[1] An additional condition was put on the prior distributions such that t2>t1.

[2] The Kimura 2-parameter nucleotide substitution model was used with invariant sites=90 and gamma rate variation parameter α=0.050.

B. Prior distributions for scenarios in Figure S10.

Parameter		**Shape**	**Min.–Max.**	**Increment**
Population size				
	Toluca	Uniform	1–1000000	1
	EC-1	Uniform	1–300000	1
	PE-3	Uniform	1–200000	1
	PE-7	Uniform	1–200000	1
	US-1	Uniform	1–500000	1
	Unsampled population	Uniform	1–2000000	1
	Ancestor of two lineages	Uniform	1–2000000	1
Time since divergence[3]				
	t1: two populations (scenarios 1-3)	Uniform	1–200000	1
	t2: two or three populations	Uniform	1–500000	1
	t3: sampled & unsampled populations (scenarios 7-15)	Uniform	1–10000000	1
Admixture events				
	Proportion from parent population	Uniform	0.001–0.999	0.001
	ta1: timing since admixture event (scenarios 4-12)	Uniform	1–300000	1
	ta2: timing since admixture event (scenarios 13-15)	Uniform	1–1000000	1

Nucleotide sequence evolution[4]				
	Mean mutation rate	Uniform	1.00×10^{-10} – 1.00×10^{-8}	
	Gamma distribution	Uniform	1.00×10^{-11} – 1.00×10^{-7}	2.00

[3] Additional conditions were put on the prior distributions to force the timing of events into the tree structure shown in Figure S9, including placing admixture events before or after population splits and ordering of the population splits such that t2>t1, t2>ta1, t3>ta1, ta2>t2, and t3>ta2.

[4] The Kimura 2-parameter nucleotide substitution model was used with invariant sites=90 and gamma rate variation parameter α=0.050.

C. Prior distributions for scenarios in Figure 5.

Parameter		**Shape**	**Min.–Max.**	**Increment**
Population size				
	Toluca	Uniform	1–1000000	1
	EC-1	Uniform	1–400000	1
	PE-3	Uniform	1–200000	1
	US-1	Uniform	1–400000	1
	Unsampled population	Uniform	1–1000000	1
	Ancestor to Toluca & EC-1	Uniform	10–1000000	1
	Ancestor to Toluca & US-1	Uniform	10–1000000	1
Time since divergence[5]				
	t1: Toluca–EC-1	Uniform	1–200000	1
	t2: Toluca–US-1	Uniform	1–600000	1
	t3: All populations	Uniform	1–1000000	1
Admixture events				
	Proportion from parent population	Uniform	0.001–0.999	0.001
	Timing since admixture event	Uniform	1–200000	1
Nucleotide sequence evolution[6]				
	Mean mutation rate	Uniform	1.00×10^{-10} – 1.00×10^{-8}	
	Gamma distribution	Uniform	1.00×10^{-11} – 1.00×10^{-7}	2.00

[5] Additional conditions were put on the prior distributions to force the timing of events into the tree structure shown in Figure 5, including placing admixture events before or after population splits and ordering of the population splits such that t3>t1, t3>t2.

[6] The Kimura 2-parameter nucleotide substitution model was used with invariant sites=90 and gamma rate variation parameter α=0.050.

Table S7. Inferred root subpopulation using three starting seeds for (A) the mitochondrial P3 and P4 regions using only isolates from Mexico and the Andes, and (B) the nuclear RAS locus, using the data from Gómez-Alpizar *et al.* (1).

A. P3 and P4

		Seed 1		**Seed 2**		**Seed 3**	
Theta	**Migration**	**Root location**	**Prob.**	**Root location**	**Prob.**	**Root location**	**Prob.**
1.8	0.3, 0.3	Mexico	0.505	Mexico	0.500	Andes	0.501
	1.0, 1.0	Andes	0.501	Andes	0.505	Mexico	0.508
	5.0, 5.0	Mexico	0.559	Mexico	0.556	Mexico	0.510
	10.0, 10.0	Mexico	0.587	Mexico	0.577	Andes	0.669
2.0	1.0, 1.0	Andes	0.504	Andes	0.501	Mexico	0.503
	5.0, 5.0	Andes	0.513	Andes	0.531	Andes	0.562
	10.0, 10.0	Andes	0.689	Mexico	0.727	Andes	0.549

B. RAS

		Seed 1		**Seed 2**		**Seed 3**	
Theta	**Migration**	**Root location**	**Prob.**	**Root location**	**Prob.**	**Root location**	**Prob.**
2.0	0.5, 0.5	SA	0.515	NSA	0.537	SA	0.535
	5.0, 5.0	SA	0.514	SA	0.582	SA	0.847
	10.0, 10.0	NSA	0.889	NSA	0.875	SA	0.686
	5.0, 10.0	SA	0.916	SA	0.949	SA	0.985
	10.0, 5.0	NSA	0.679	NSA	0.902	SA	0.935
	4.3, 14.6 *	SA	0.965	SA	0.992	SA	0.970
	14.6, 4.3	NSA	0.939	NSA	0.996	NSA	0.979

* Migration rates used by Gómez-Alpizar *et al*. (1)

Table S8. Root population inferred using RAS sequences from Mexico and Andean samples from this study. Migration rates were symmetric for all runs.

Simulations		Seed 1 1×10^6		Seed 2 1×10^6		Seed 3 1×10^7		Seed 4 1×10^7	
θ	**M**	**Root**	**Prob.**	**Root**	**Prob.**	**Root**	**Prob.**	**Root**	**Prob.**
1.5	0.5	Mexico	0.541	Mexico	0.584	Andes	0.510	Mexico	0.521
	5.0	Mexico	0.638	Mexico	0.703	Mexico	0.564	Mexico	0.583
	10.0	Andes	0.723	Mexico	0.975	Mexico	0.612	Andes	0.794
2.0	0.5	Andes	0.676	Andes	0.638	Mexico	0.513	Mexico	0.514
	5.0	Mexico	0.883	Andes	0.614	Mexico	0.590	Andes	0.794
	10.0	Mexico	0.714	Andes	0.688	Mexico	0.663	Mexico	0.939

Supporting Figures

Figure S1. Structure plots for K from 2 to 10 for the Andes (with and without admixture) and Mexico samples of *P. infestans*.

Figure S2. Delta K values obtained from Structure Harvester for (A) Mexico and (B) Andes (with admixture) samples of *P. infestans*. Note the highest delta K value corresponds to K=4 for Mexico and K=2 for the Andes.

Figure S3. Root state posterior probabilities for each location for (A) Clade 1c and (B) *P. infestans.* Independent analysis of each locus produced high probabilities for the root of both Clade 1c and *P. infestans* in Mexico (green bar).

Figure S4. Maximum clade credibility genealogy for the **β-tubulin** locus. Colors of branches indicate the most probable geographic origin of each lineage.

Figure S5. Maximum clade credibility genealogy for the **PITG_11126** locus. Colors of branches indicate the most probable geographic origin of each lineage.

Figure S6. Maximum clade credibility genealogy for the **RAS** locus. Colors of branches indicate the most probable geographic origin of each lineage.

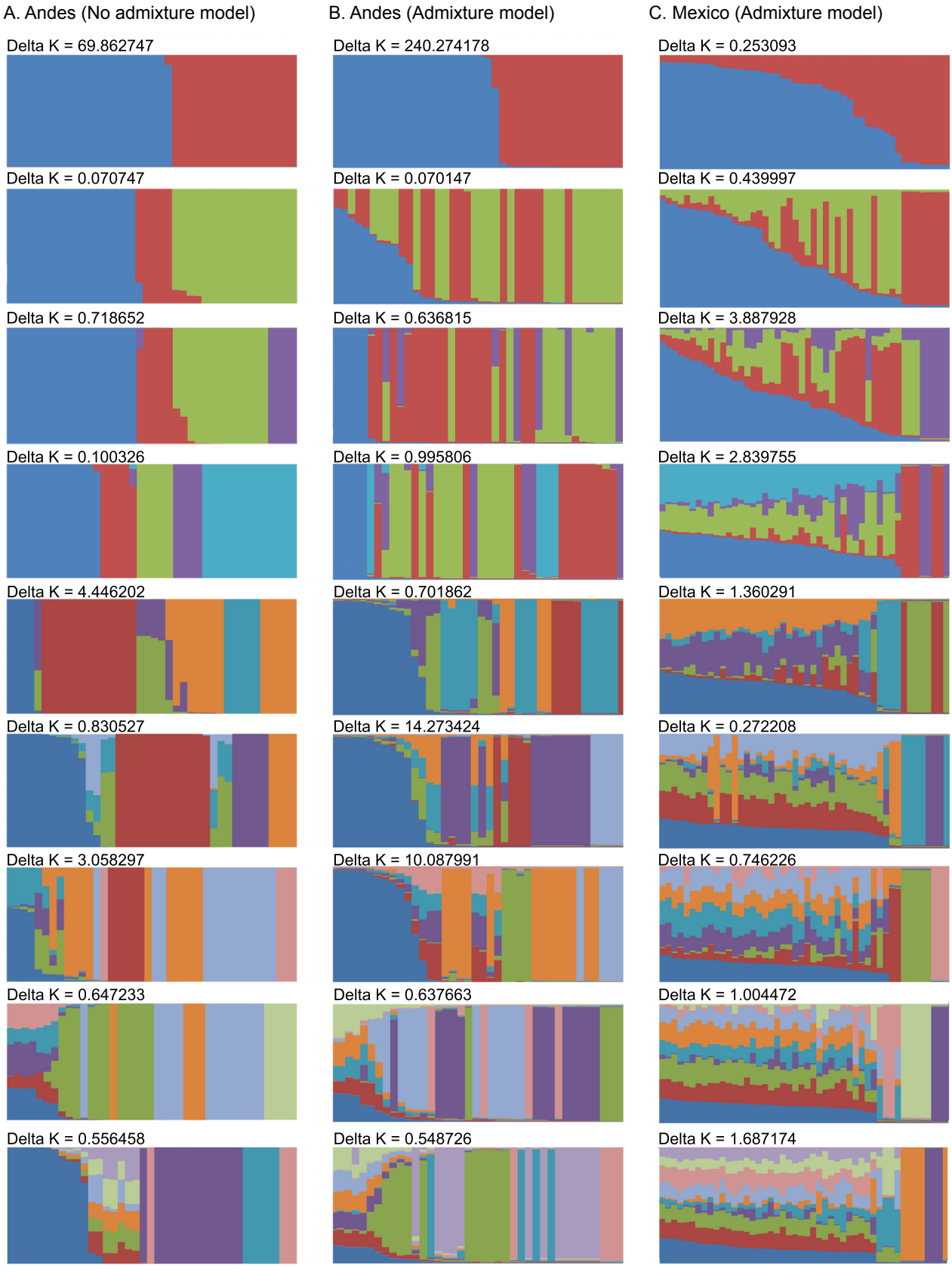
Figure S7. Maximum clade credibility genealogy for the **Trp1** locus. Colors of branches indicate the most probable geographic origin of each lineage.

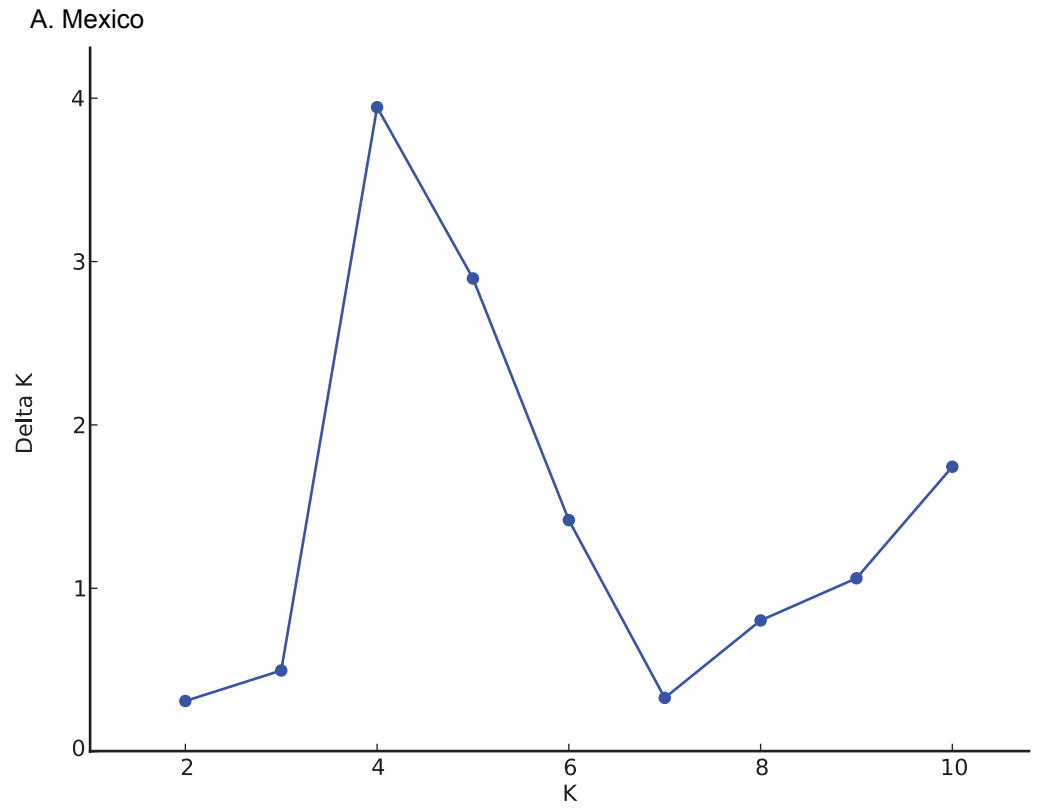
Figure S8. Maximum clade credibility phylogeny of *Phytophthora* Clade 1c from all four loci. Colors of branches indicate the most probable geographic origin of each lineage and taxon colors represent species (purple: *P. phaseoli*; blue: *P. ipomoeae*; green: *P. mirabilis*; red: *P. andina*; black: *P. infestans*).

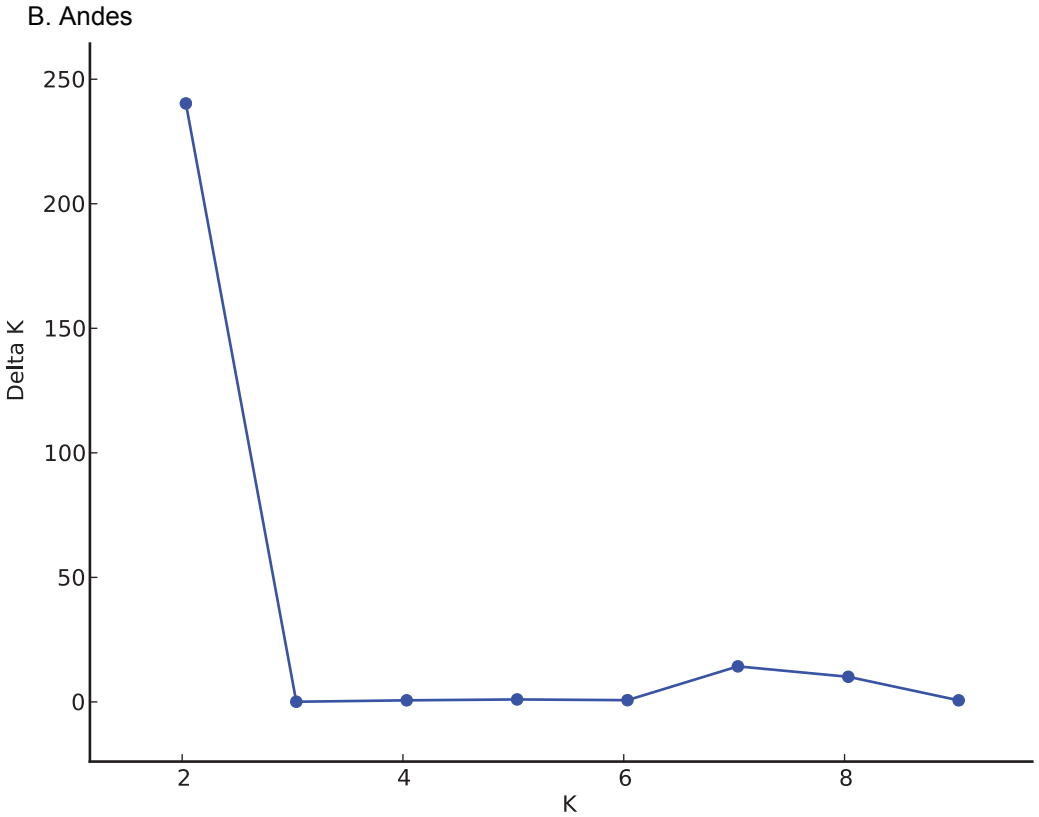
Figure S9. Scenarios used to test evolutionary relationships among Andes lineages EC-1, PE-3, and PE-7 using DIYABC. Present day populations are at the base of the tree schematic. Ancestral relationships among these populations are represented by lines intersecting in the past, with the vertex of the schematic representing the most recent common ancestor of all samples. Horizontal lines indicate admixture events between the ancestral populations connected by the horizontal line. Change in line thickness indicates a potential change in population size, such that all branches in scenarios D, E, and F had independently estimated population sizes. The scenarios in which PE-3 and PE-7 diverged more recently from each other than from EC-1 produced high posterior probabilities.

Figure S10. Scenarios used to test evolutionary relationships of EC-1, PE-3, and PE-7 lineages to US-1 in the Andes and to the Toluca, Mexico population. The 15 scenarios were compared against each other, using three independent sets of scenarios with EC-1, PE-3, and PE-7, in turn, substituted for population 'X'. 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. Andes (No admixture model)
B. Andes (Admixture model)
C. Mexico (Admixture model)
Delta K = 69.862747
Delta K = 240.274178
Delta K = 0.253093
Delta K = 0.070747
Delta K = 0.070147
Delta K = 0.439997
Delta K = 0.718652
Delta K = 0.636815
Delta K = 3.887928
Delta K = 0.100326
Delta K = 0.995806
Delta K = 2.839755
Delta K = 4.446202
Delta K = 0.701862
Delta K = 1.360291
Delta K = 0.830527
Delta K = 14.273424
Delta K = 0.272208
Delta K = 3.058297
Delta K = 10.087991
Delta K = 0.746226
Delta K = 0.647233
Delta K = 0.637663
Delta K = 1.004472
Delta K = 0.556458
Delta K = 0.548726
Delta K = 1.687174


A. Mexico
Delta K
0
1
2
3
4
2
4
6
8
10
K


B. Andes
Delta K
0
50
100
150
200
250
2
4
6
8
K

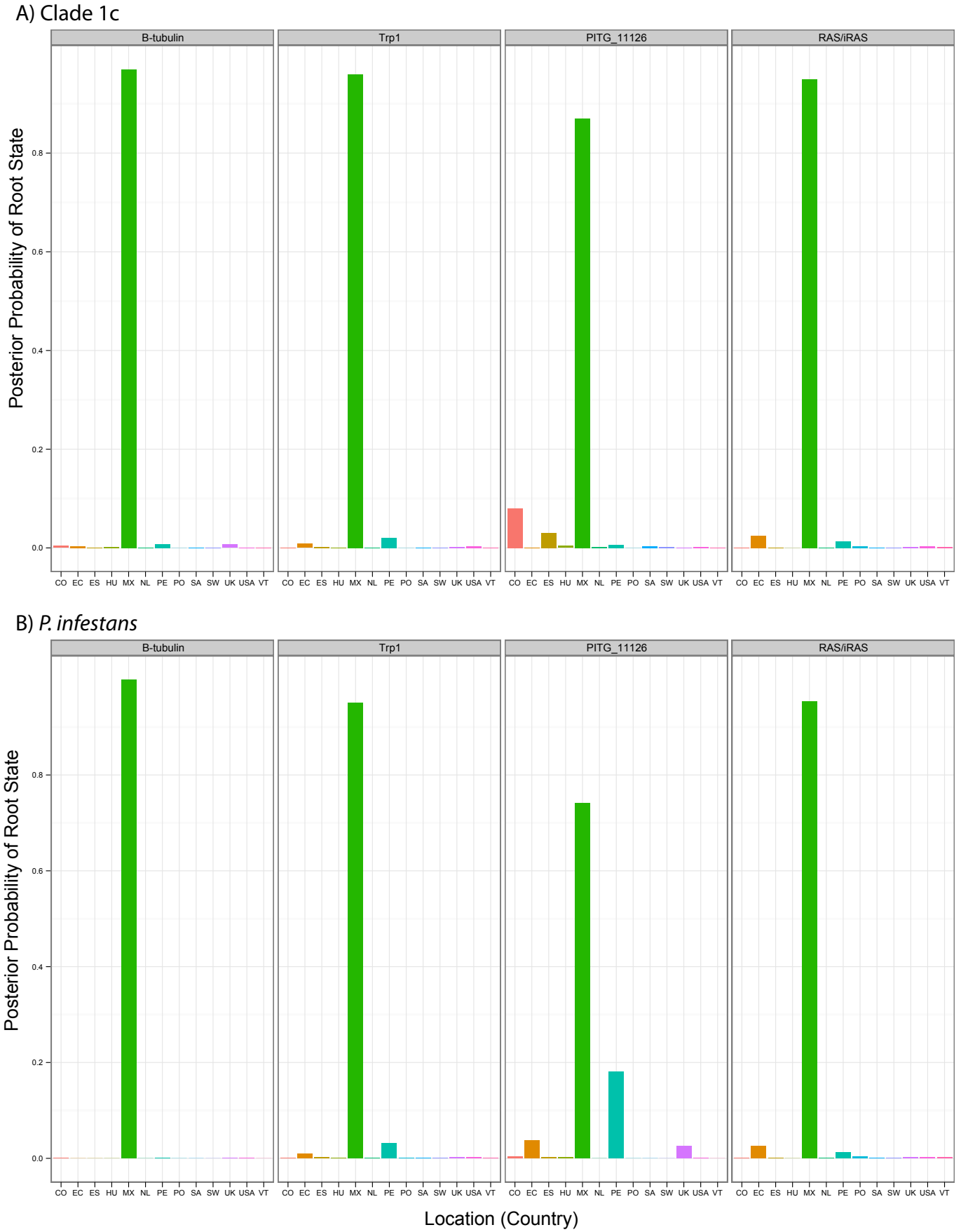


Figure S3. Root state posterior probability values for each location after 300 millions MCMC generations by locus for (A) Clade 1c species and (B) *P. infestans*. The posterior probability for a Mexico root (green bar) is higher than for any of the other countries for each locus independently and both datasets.

Figure S4. Maximum clade credibility genealogy for the **β-tubulin** locus. Colors of branches indicate the most probable geographic origin of each lineage.

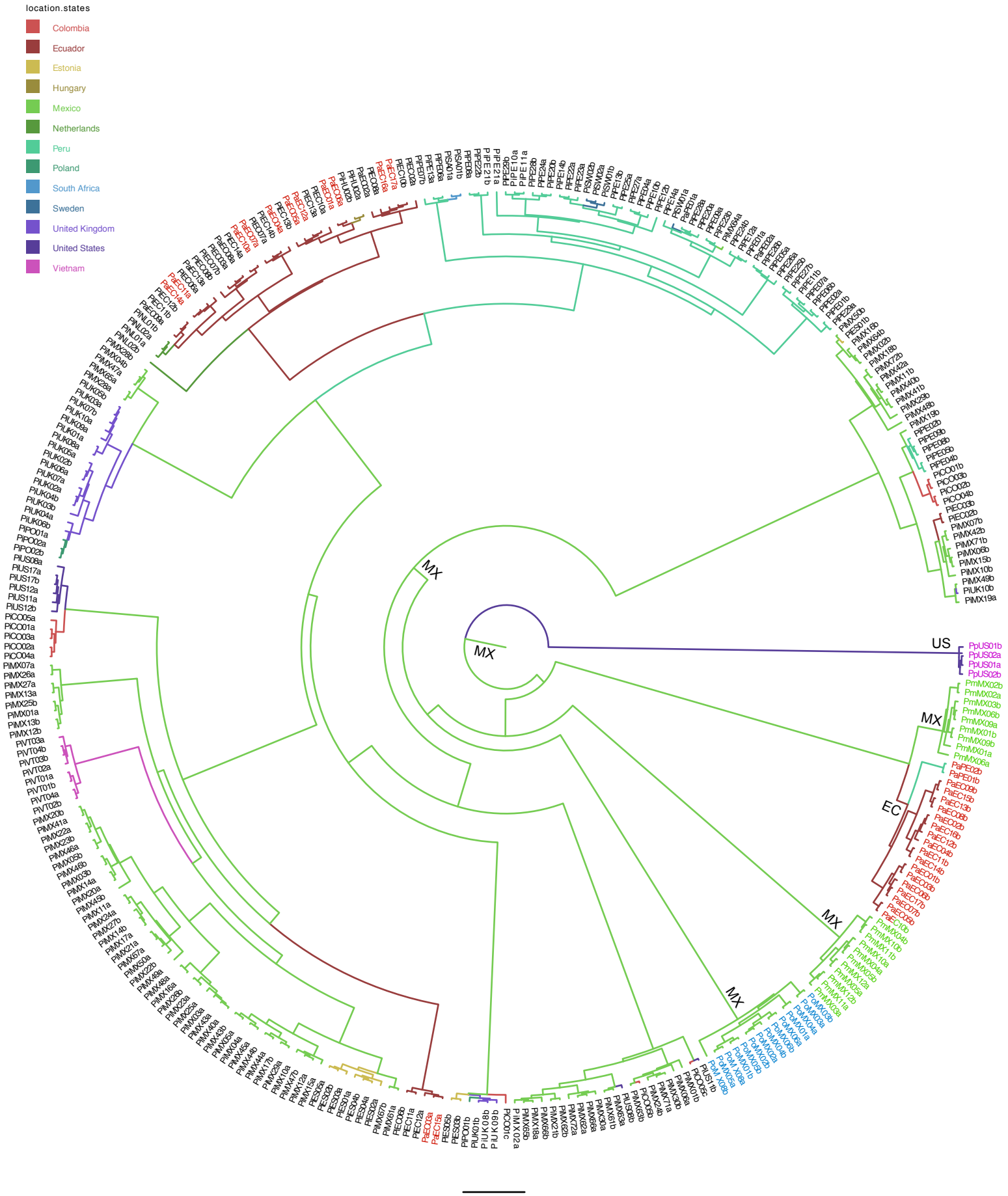


Figure S5. Maximum clade credibility genealogy for the **PITG_11126** locus. Colors of branches indicate the most probable geographic origin of each lineage.

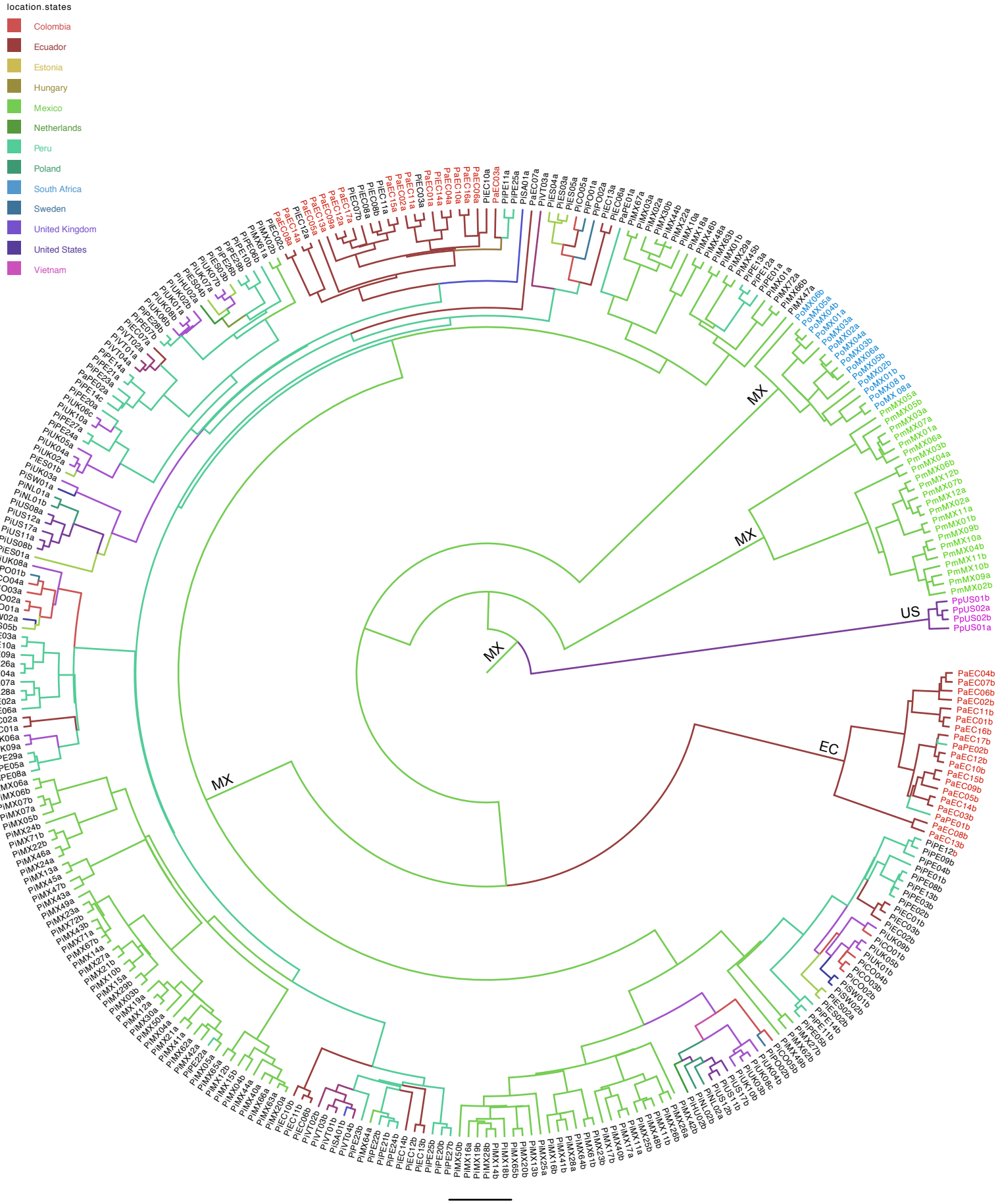


Figure S6. Maximum clade credibility genealogy for the **RAS** locus. Colors of branches indicate the most probable geographic origin of each lineage.

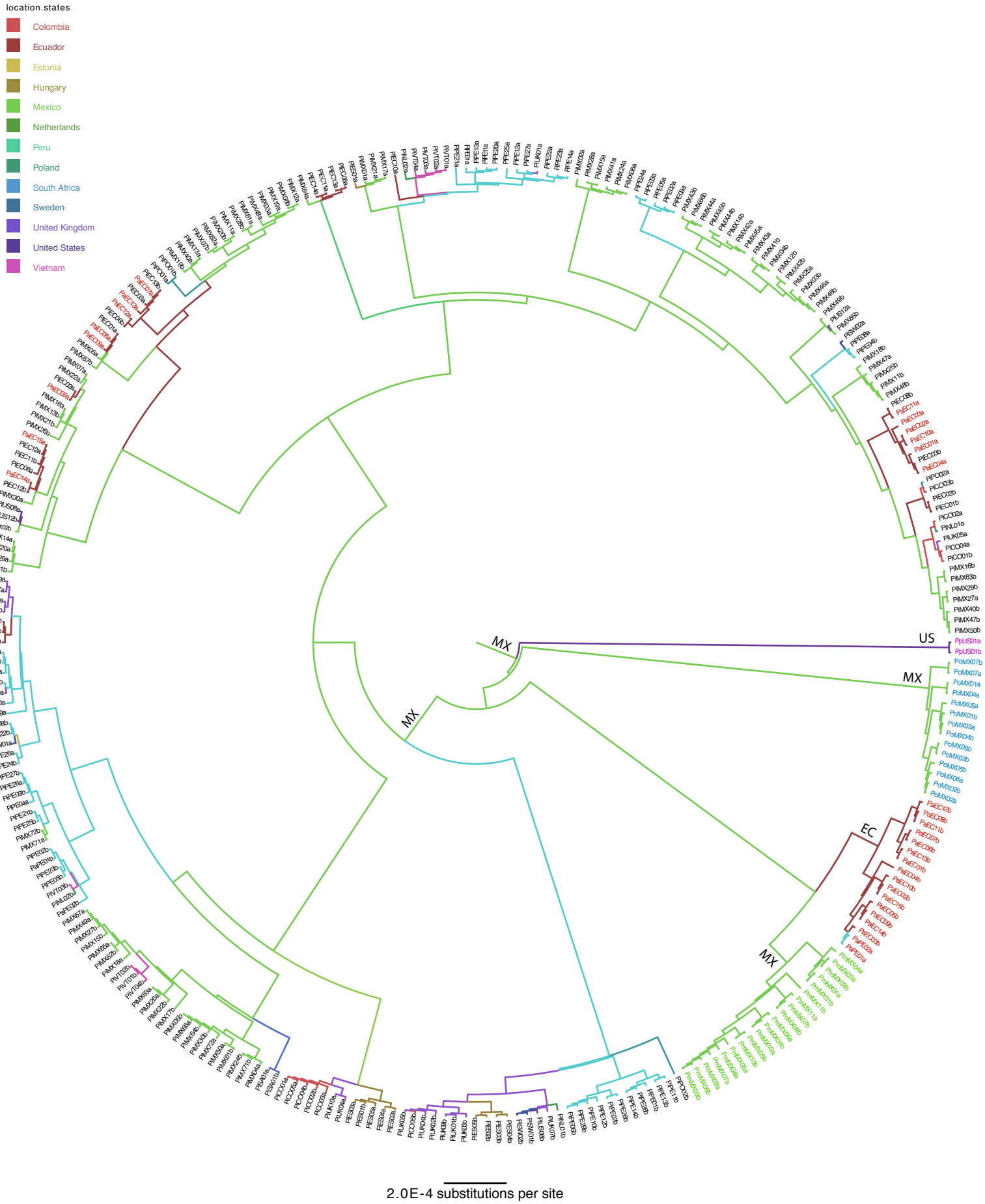
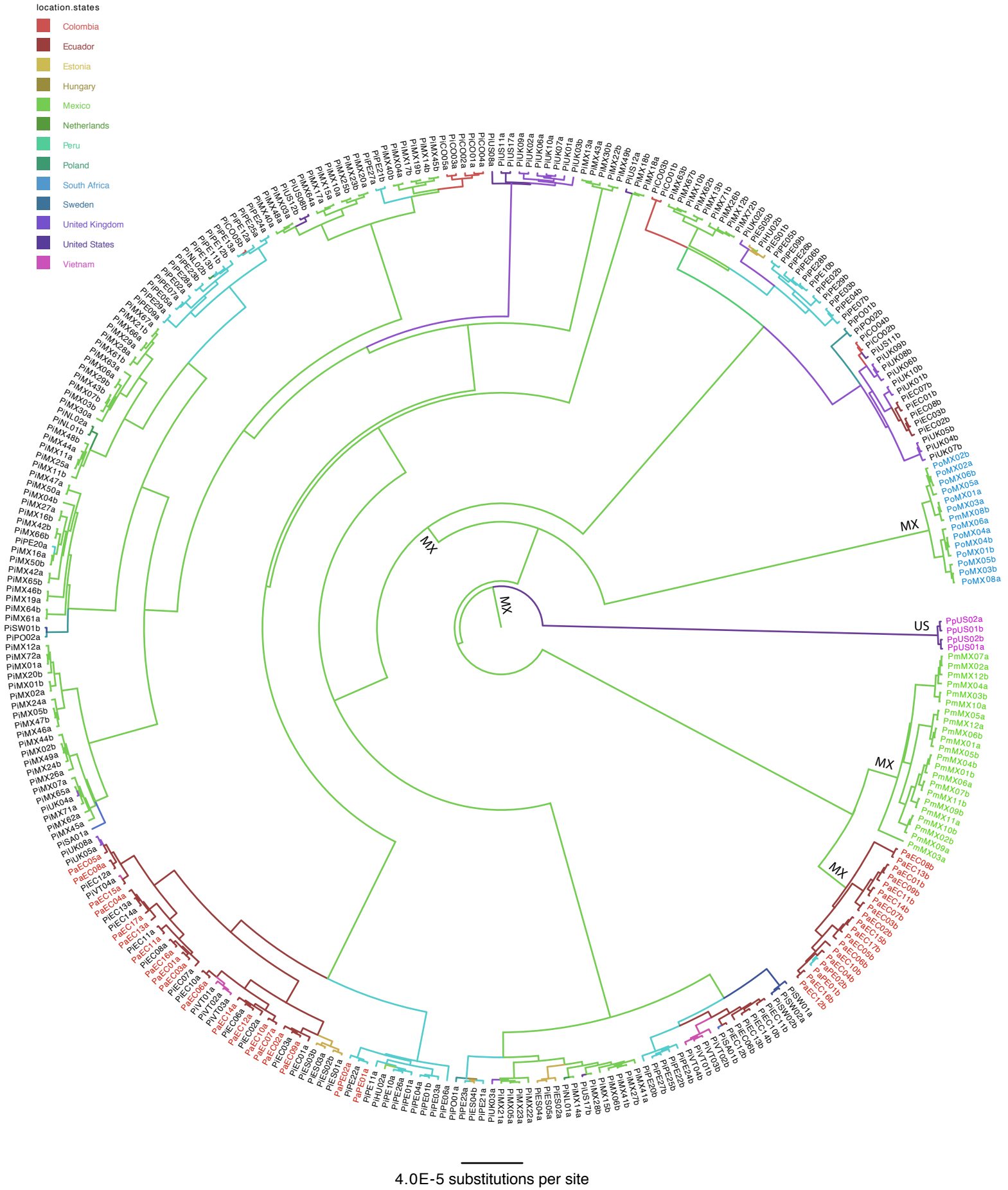
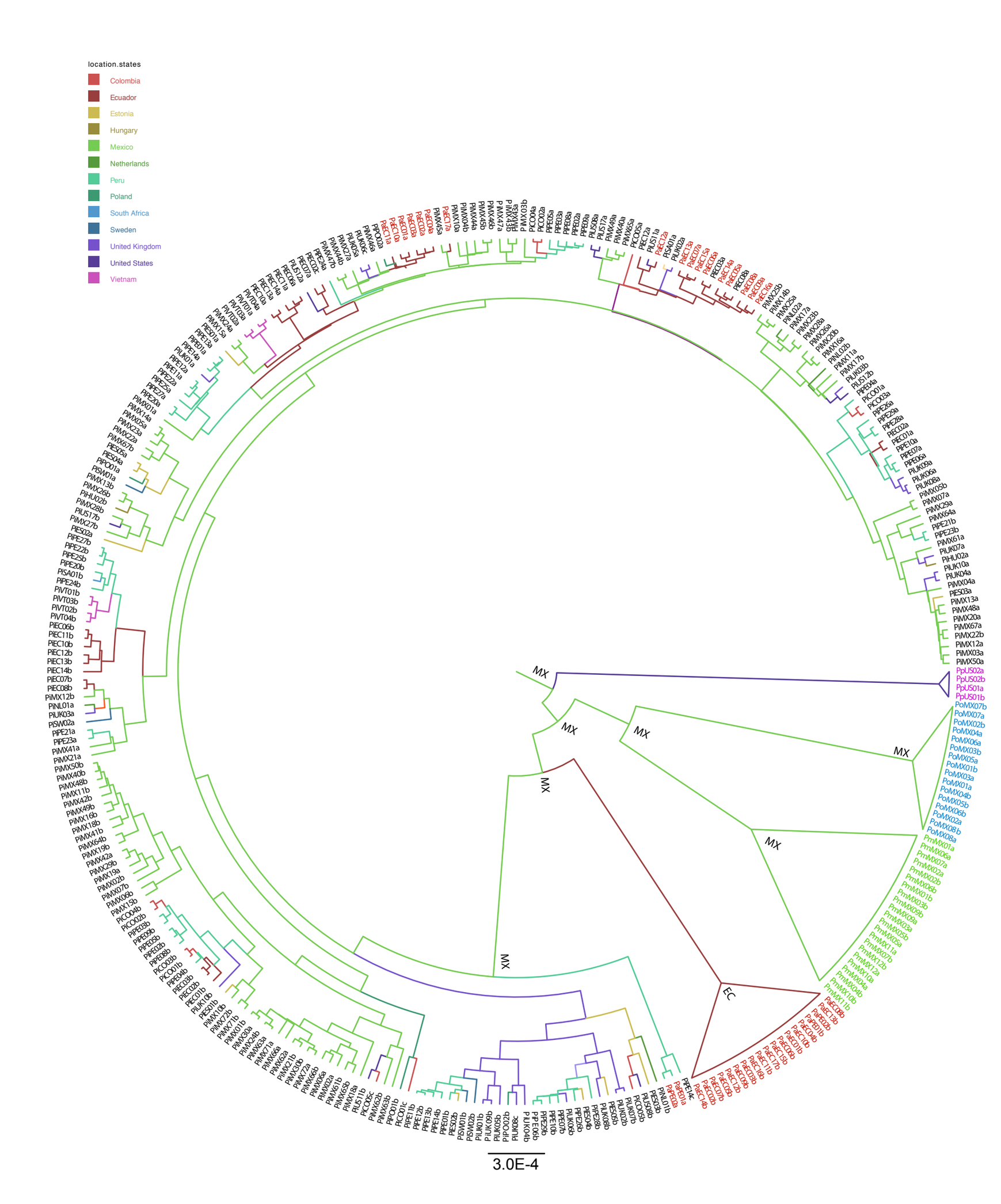
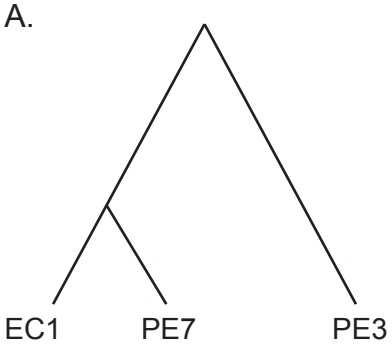


Figure S7. Maximum clade credibility genealogy for the **Trp1** locus. Colors of branches indicate the most probable geographic origin of each lineage.



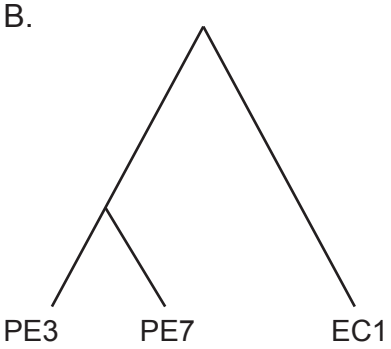

location.states
Colombia
Ecuador
Estonia
Hungary
Mexico
Netherlands
Peru
Poland
South Africa
Sweden
United Kingdom
United States
Vietnam
MX
MX
MX
MX
MX
MX
EC
3.0E-4

A.



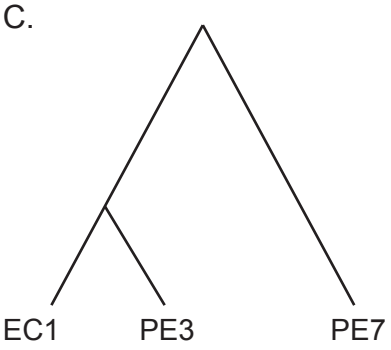
0.9 [0.7, 1.1]

B.



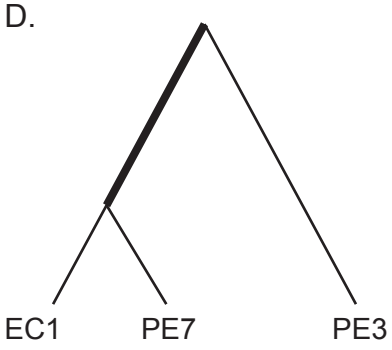
21.9 [19.7, 24.0]

C.



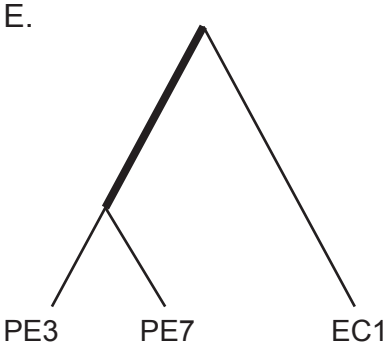
0.9 [0.7, 1.2]

D.



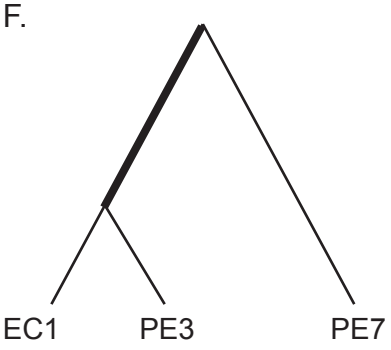
1.4 [1.1, 1.7]

E.



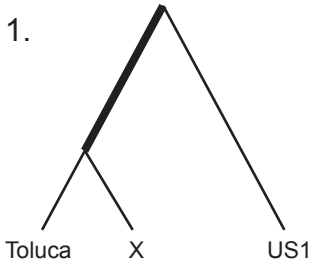
73.5 [71.1, 75.8]

F.

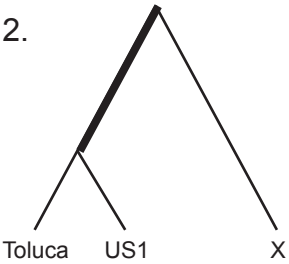


1.4 [1.1, 1.7]

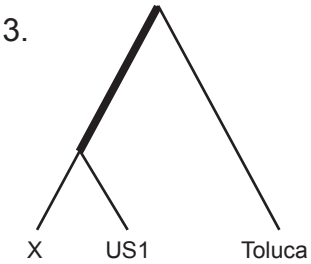
1.



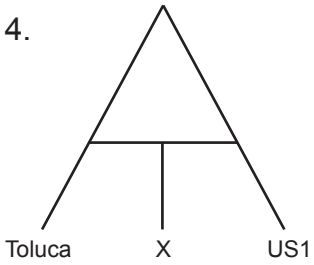
2.



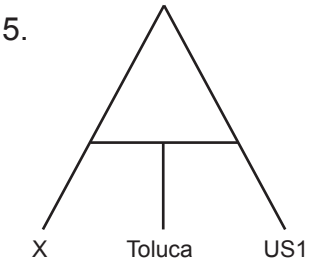
3.



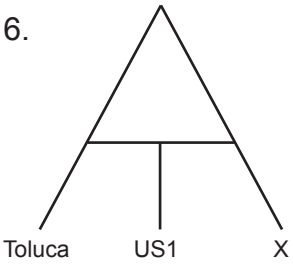
4.



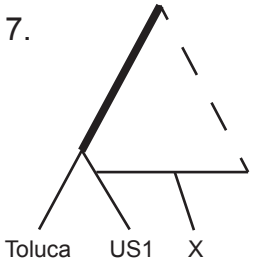
5.



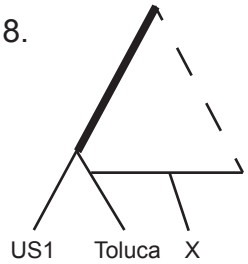
6.



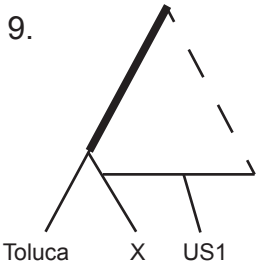
7.



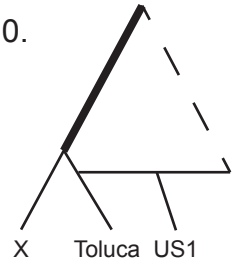
8.



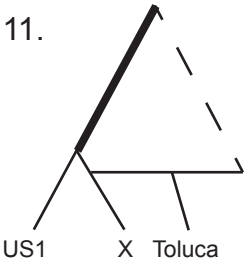
9.



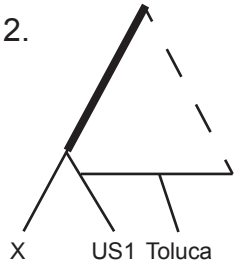
10.



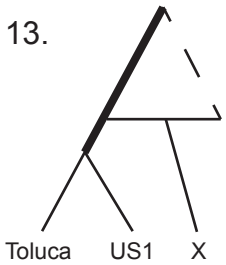
11.



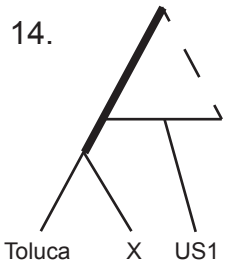
12.



13.



14.



15.