eLIFE
elifesciences.org

Figures and figure supplements

The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine

Kentaro Yoshida, et al.

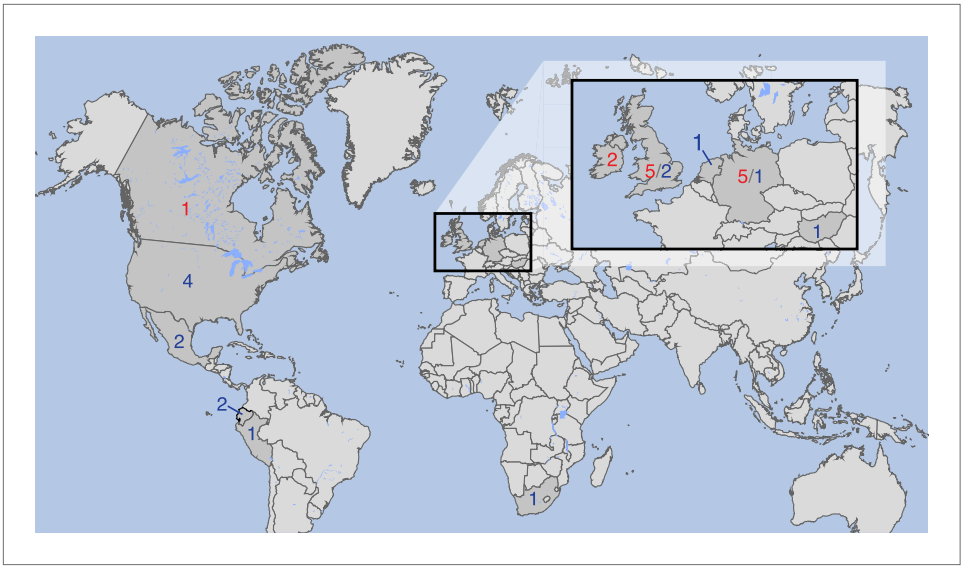


Figure 1. Countries of origin of samples used in whole-genome, mtDNA genome or both analyses. Red indicates number of historic and blue of modern samples. More information on the samples is given in ***Tables 1 and 2***.
DOI: 10.7554/eLife.00731.004

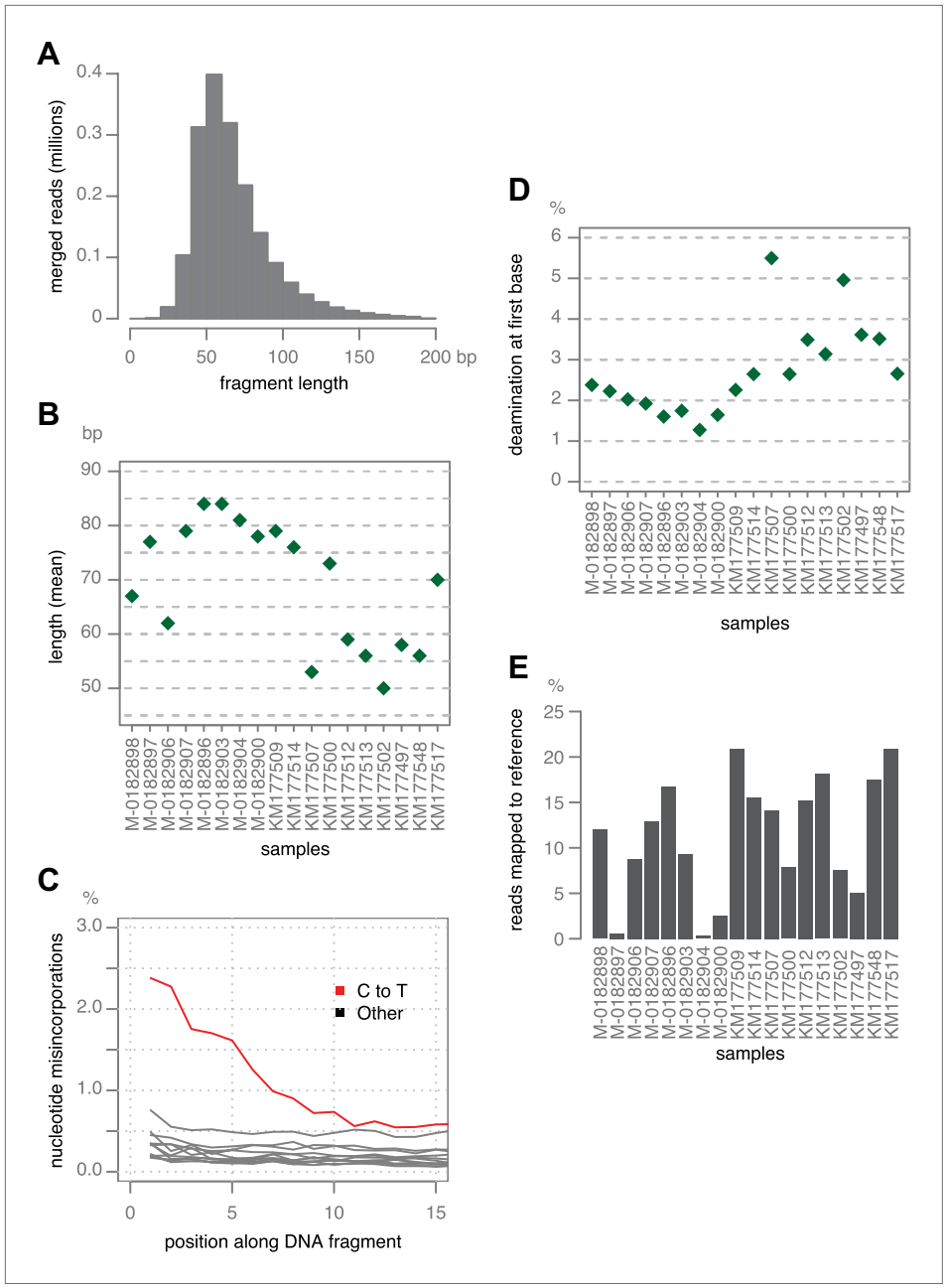


Figure 2. Ancient DNA-like characteristic of historic samples. (**A**) Lengths of merged reads from historic sample M-0182898. (**B**) Mean lengths of merged reads from historic samples. (**C**) Nucleotide mis-incorporation in reads from the historic sample M-0182898. (**D**) Deamination at first 5′ end base in historic samples. (**E**) Percentage of merged reads that mapped to the *P. infestans* reference genome.

DOI: 10.7554/eLife.00731.006

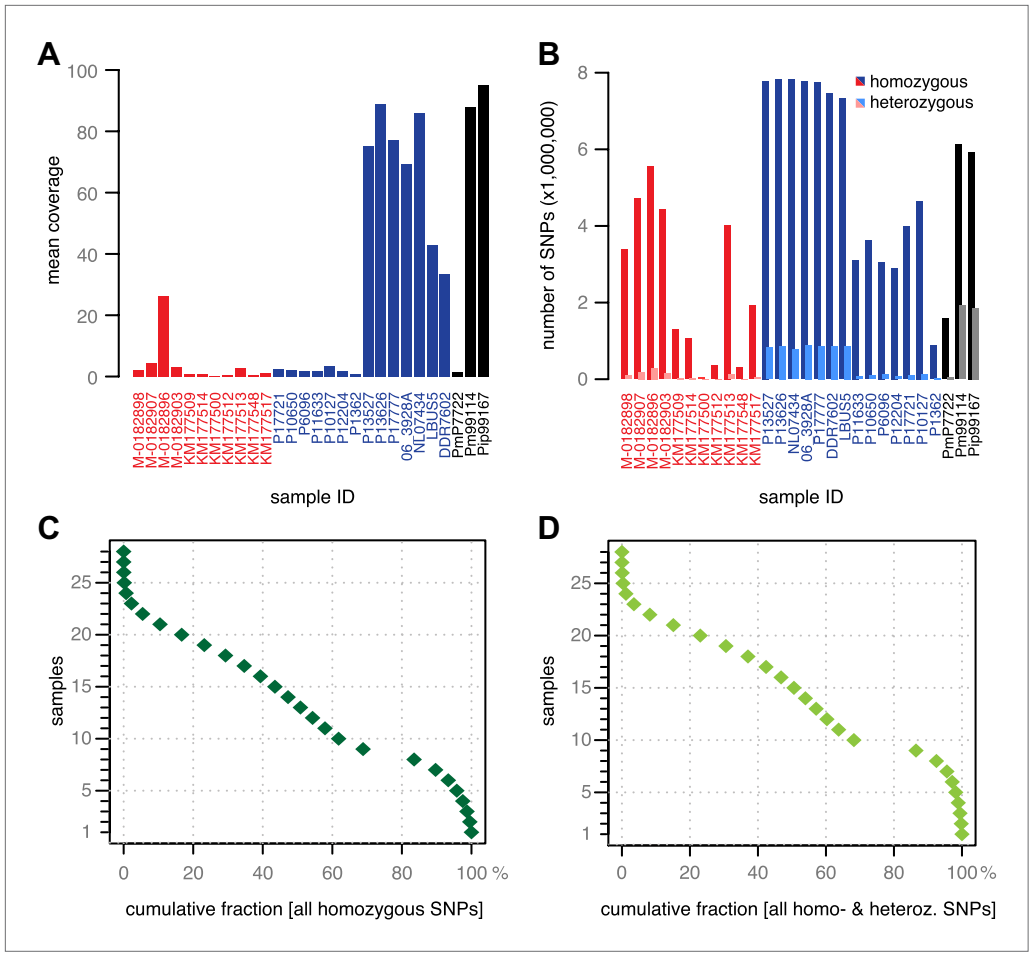
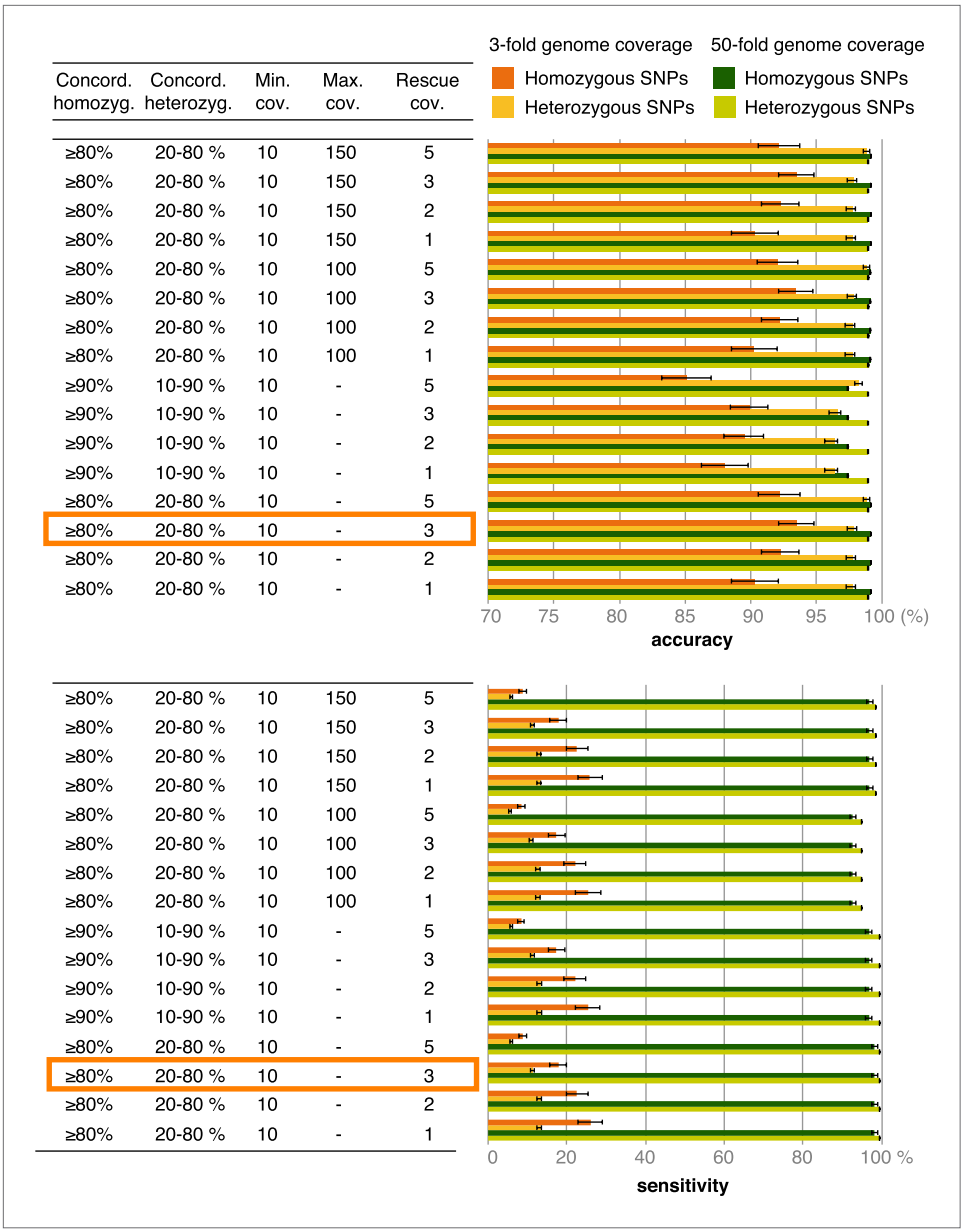


Figure 3. Coverage and SNP statistics. (**A**) Mean nuclear genome coverage from historic (red) and modern (blue) samples. (**B**) Homo- and heterozygous SNPs in each sample. (**C**) Inverse cumulative coverage for all homozygous SNPs across all samples. (**D**) Same as (**C**) for homo- and heterozygous SNPs.

DOI: 10.7554/eLife.00731.007



Concord. homozyg.	Concord. heterozyg.	Min. cov.	Max. cov.	Rescue cov.
≥80%	20-80 %	10	150	5
≥80%	20-80 %	10	150	3
≥80%	20-80 %	10	150	2
≥80%	20-80 %	10	150	1
≥80%	20-80 %	10	100	5
≥80%	20-80 %	10	100	3
≥80%	20-80 %	10	100	2
≥80%	20-80 %	10	100	1
≥90%	10-90 %	10	-	5
≥90%	10-90 %	10	-	3
≥90%	10-90 %	10	-	2
≥90%	10-90 %	10	-	1
≥80%	20-80 %	10	-	5
≥80%	20-80 %	10	-	3
≥80%	20-80 %	10	-	2
≥80%	20-80 %	10	-	1

Concord. homozyg.	Concord. heterozyg.	Min. cov.	Max. cov.	Rescue cov.
≥80%	20-80 %	10	150	5
≥80%	20-80 %	10	150	3
≥80%	20-80 %	10	150	2
≥80%	20-80 %	10	150	1
≥80%	20-80 %	10	100	5
≥80%	20-80 %	10	100	3
≥80%	20-80 %	10	100	2
≥80%	20-80 %	10	100	1
≥90%	10-90 %	10	-	5
≥90%	10-90 %	10	-	3
≥90%	10-90 %	10	-	2
≥90%	10-90 %	10	-	1
≥80%	20-80 %	10	-	5
≥80%	20-80 %	10	-	3
≥80%	20-80 %	10	-	2
≥80%	20-80 %	10	-	1

Figure 3—figure supplement 1. Accuracy and sensitivity of SNP calling at different cutoffs for SNP concordance based on 3- and 50-fold coverage of simulated data. Rescue cov.—minimum coverage required to accept SNP calls in low-coverage genomes based on these SNPs having been found in high-coverage genomes. The cutoffs enclosed in orange rectangles were used for the final analysis.

DOI: 10.7554/eLife.00731.008

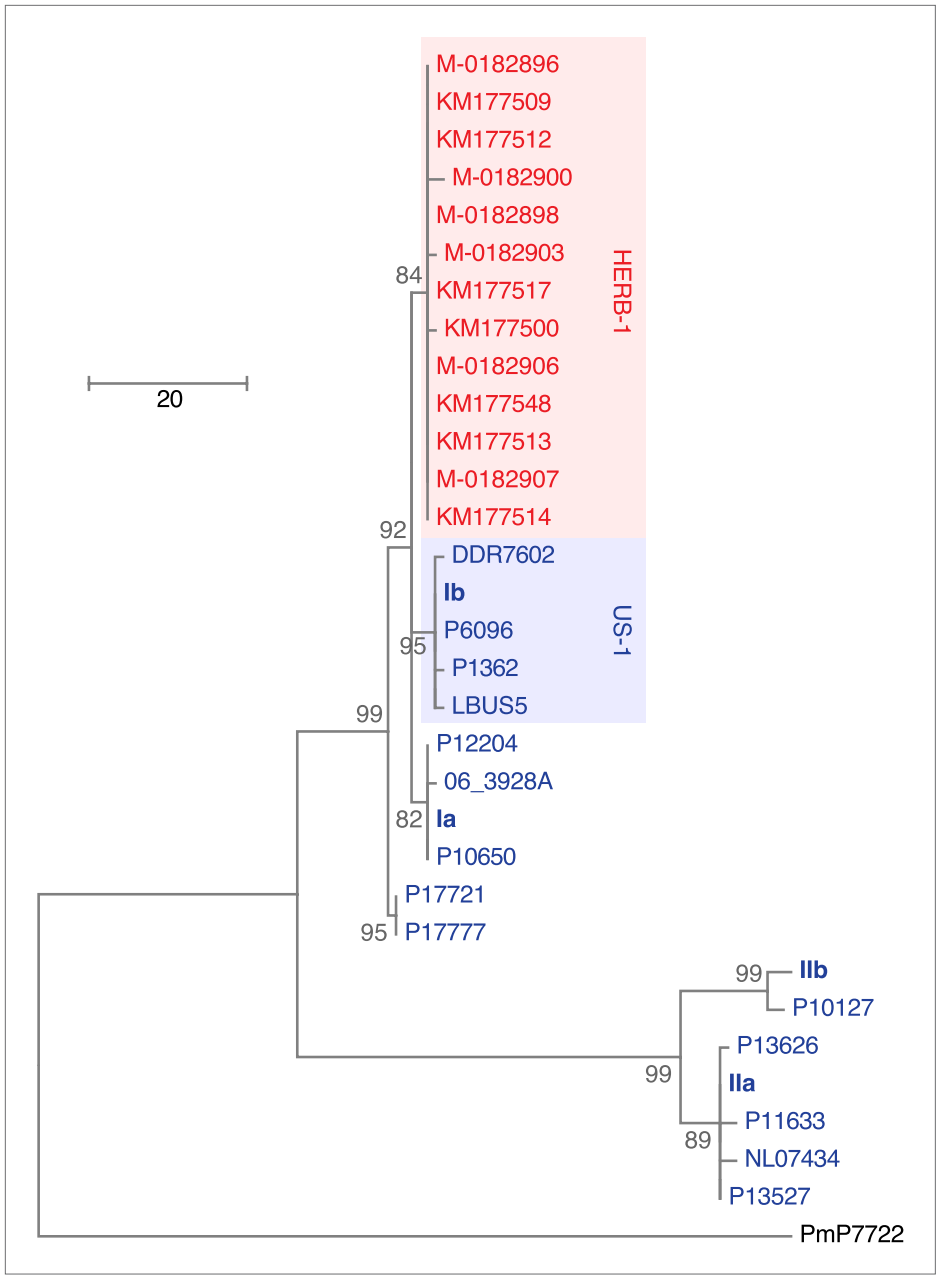


Figure 4. Maximum-parsimony phylogenetic tree of complete mtDNA genomes. Sites with less than 90% information were not considered, leaving 24,560 sites in the final dataset. Numbers at branches indicate bootstrap support (100 replicates), and scale indicates changes.

DOI: 10.7554/eLife.00731.009

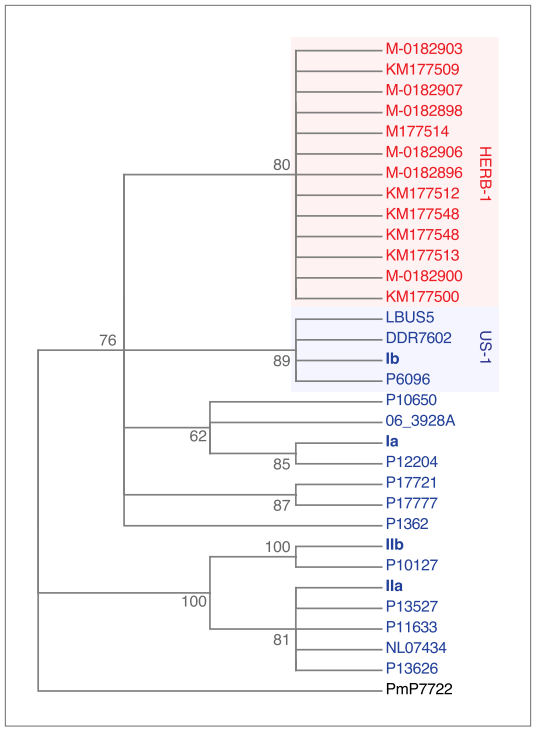


Figure 4—figure supplement 1. Maximum-likelihood phylogenetic tree of complete mtDNA genomes. Sites with less than 90% information were not considered, leaving 24,560 sites in the final dataset. Numbers at branches indicate bootstrap support (100 replicates).

DOI: 10.7554/eLife.00731.010

Figure 4—figure supplement 2. mtDNA sequences around diagnostic *Msp1* restriction site (grey) for reference haplotype modern strains (blue) and historic strains (red). The *Msp1* (CCGG) restriction site is only present in the Ib haplotype; all other strains have a C-to-T substitution (CTGG).

DOI: 10.7554/eLife.00731.011

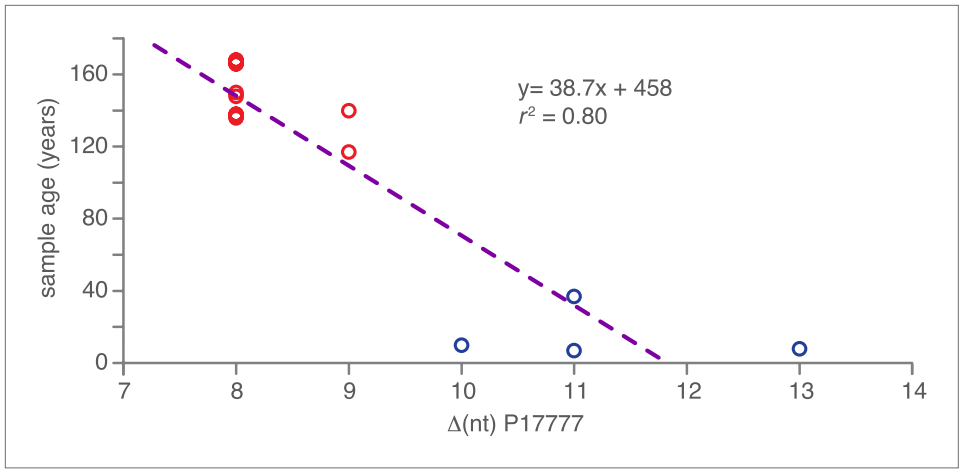


Figure 5. Correlation between nucleotide distance of mtDNA genomes of HERB-1/haplotype Ia/haplotype Ib clade to the outgroup P17777 and sample age in calendar years before present.

DOI: 10.7554/eLife.00731.012

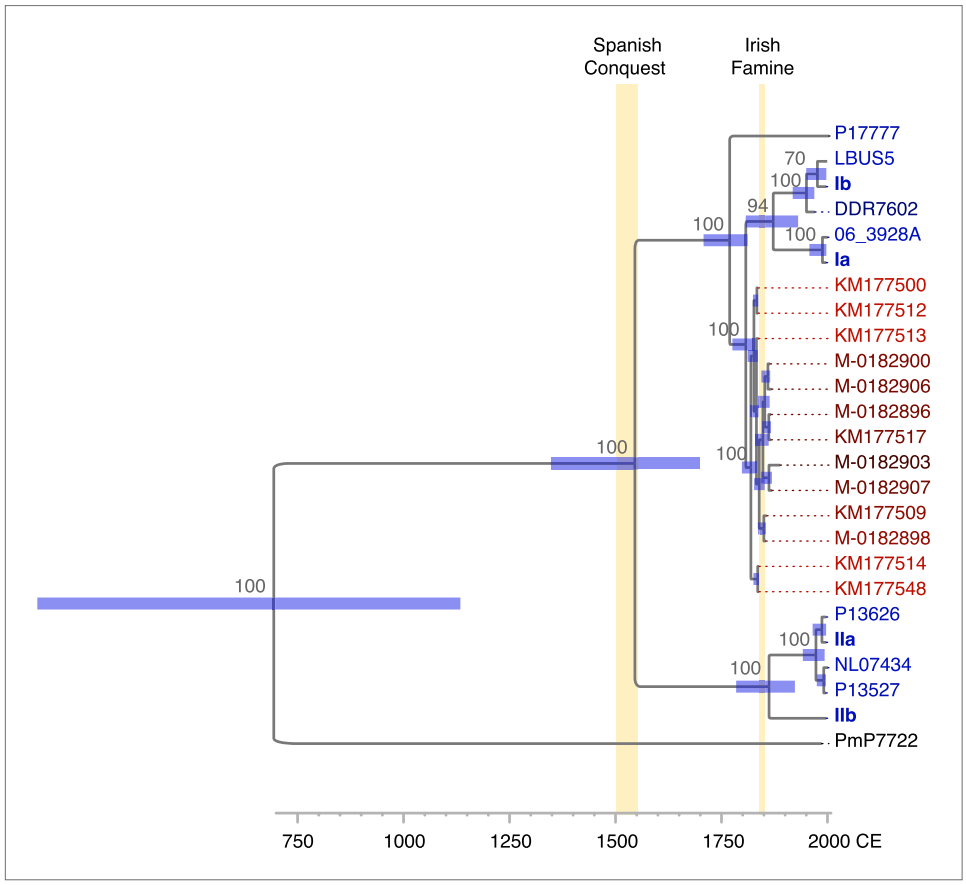


Figure 6. Divergence estimates of mtDNA genomes. Bayesian consensus tree from 147,000 inferred trees. Posterior probability support above 50% is shown next to each node. Blue horizontal bars represent the 95% HPD interval for the node height. Light yellow bars indicate major historical events discussed in the text. See ***Figure 5*** and ***Table 3*** for detailed estimates at the four main nodes in *P. infestans*.

DOI: 10.7554/eLife.00731.013

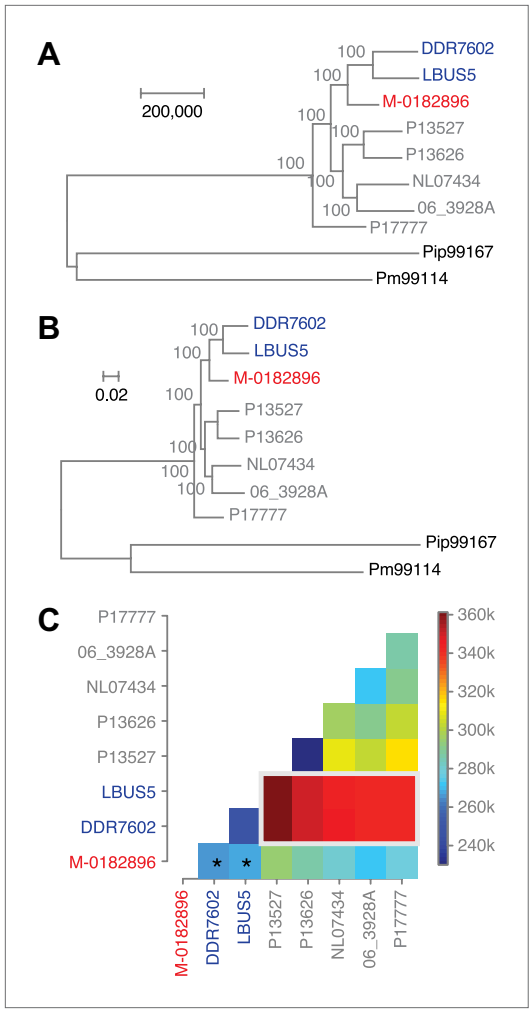


Figure 7. Phylogenetic trees of high-coverage nuclear genomes using both homozygous and heterozygous SNPs. (**A**) Maximum-parsimony tree, considering only sites with at least 95% information, leaving 4,498,351 sites in the final dataset. Numbers at branches indicate bootstrap support (100 replicates), and scale indicates genetic distance. (**B**) Maximum-likelihood tree. (**C**) Heat map of genetic differentiation (color scale indicates SNP differences). US-1 strains DDR7062 and LBUS5 have the genomes sequences closest to M-0182896 (asterisks). The two US-1 isolates in turn are outliers compared to all other modern strains (highlighted by a gray box).

DOI: 10.7554/eLife.00731.015

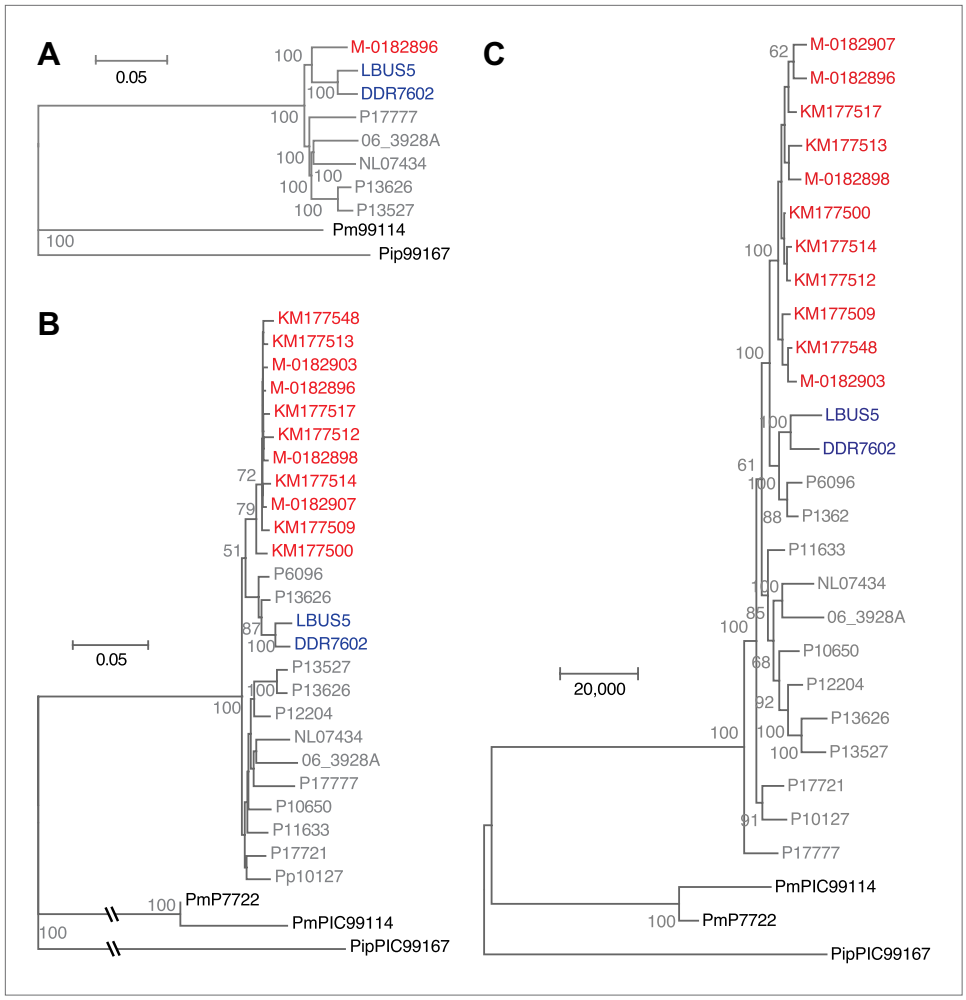


Figure 7—figure supplement 1. Phylogenetic trees of high- and low-coverage nuclear genomes. (**A**) Neighbor-joining tree of high-coverage genomes using 4,595,012 homo- and heterozygous SNPs. Numbers at branches indicate bootstrap support (100 replicates), and scale indicates genetic distance. (**B**) Neighbor-joining tree of high- and low-coverage genomes using 2,101,039 homozygous and heterozygous SNPs. Numbers at branches indicate bootstrap support above 50, from 100 replicates. Scale indicates genetic distance. (**C**) Maximum parsimony tree of high- and low-coverage genomes using 315,394 SNPs homozygous and heterozygous SNPs (using only sites with at least 80% information).

DOI: 10.7554/eLife.00731.016

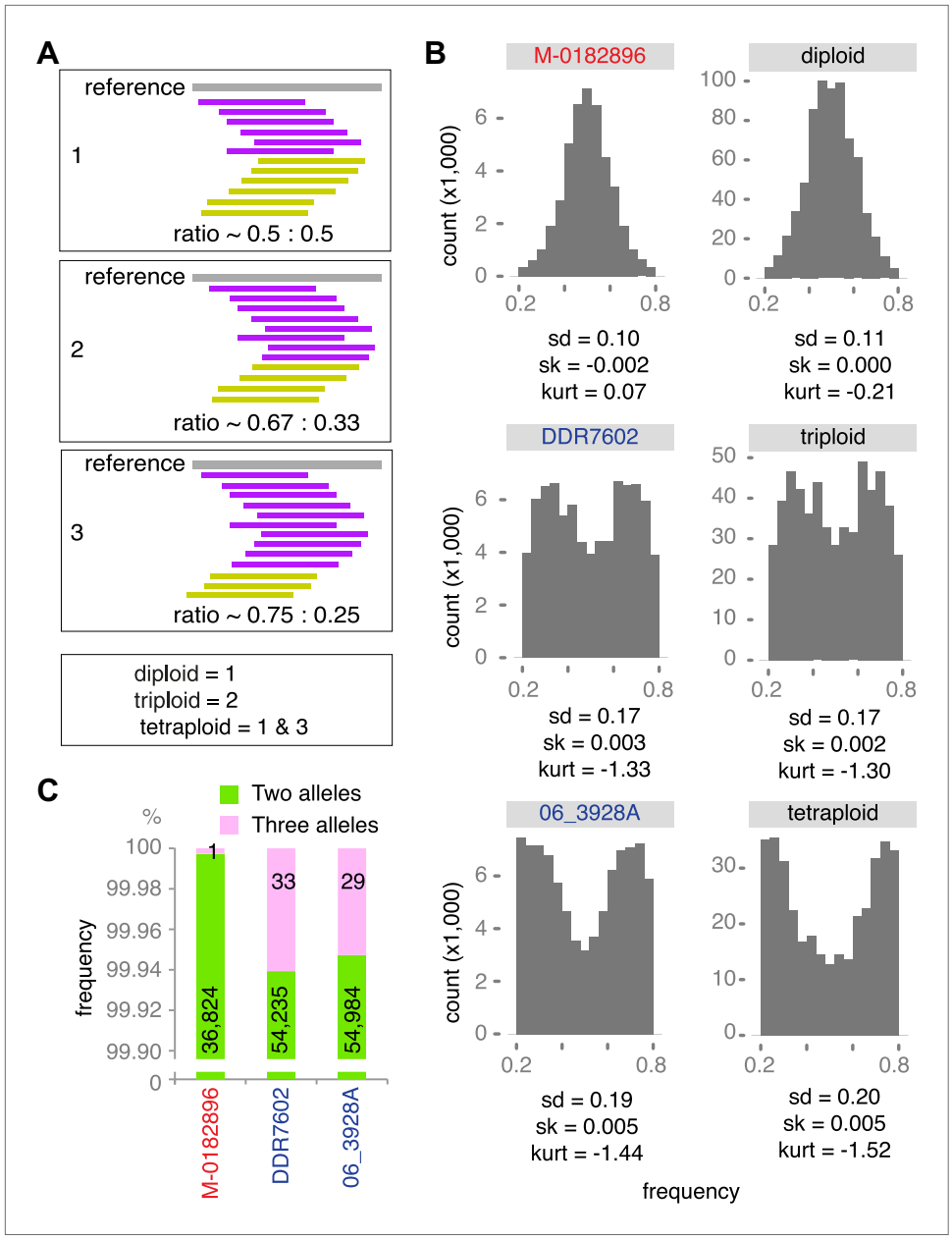


Figure 8. Ploidy analysis. (**A**) Diagram of expected read frequencies of reads at biallelic SNPs for diploid, triploid and tetraploid genomes. (**B**) Reference read frequency at biallelic SNPs in gene dense regions (GDRs) for the historic sample M-0182896, two modern samples, and simulated diploid, triploid and tetraploid genomes. The simulated tetraploid genome is assumed to have 20% of pattern 1 and 80% of pattern 3 shown in (**A**). The shape and kurtosis of the observed distributions are similar to the corresponding simulated ones. (**C**) Polymorphic positions with more than one allele in the GDR.

DOI: 10.7554/eLife.00731.017

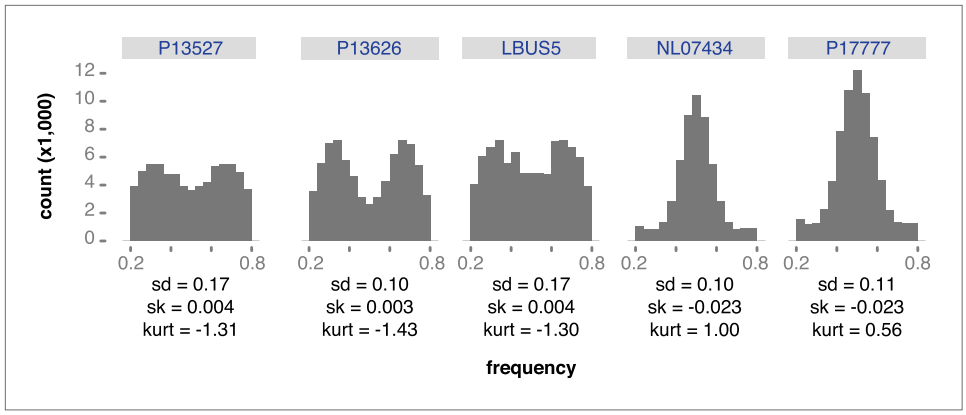


Figure 8—figure supplement 1. Reference read frequency at biallelic SNPs in gene dense regions (GDRs) for five modern high-coverage samples.

DOI: 10.7554/eLife.00731.018

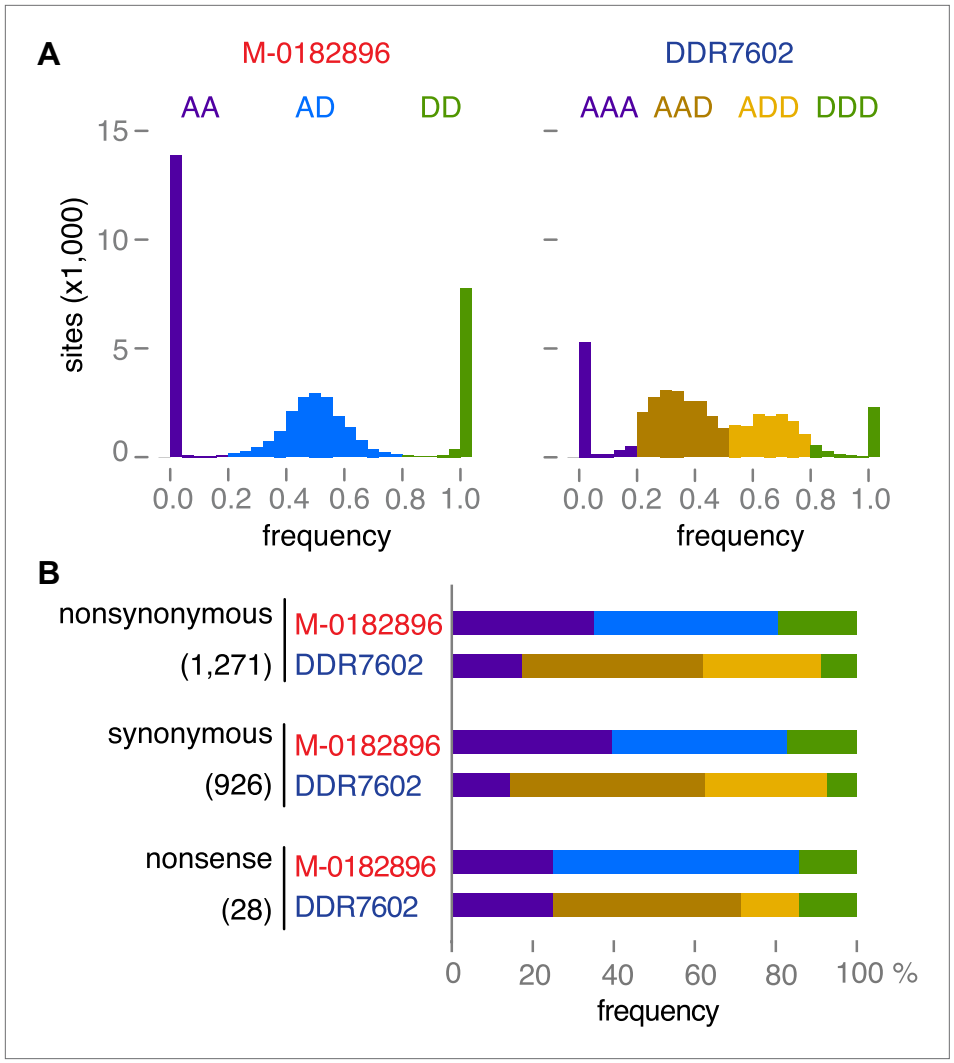


Figure 9. Read allele frequencies of historic genome M-0182896 and US-1 isolate DDR7602. Alleles were classified as ancestral or derived using outgroup species *P. mirabilis* and *P. ipomoeae*. There were 40,532 segregating sites. (**A**) Distributions of derived alleles at sites segregating between M-0182896 and DDR7602. (**B**) Annotation of the different site classes.

DOI: 10.7554/eLife.00731.019

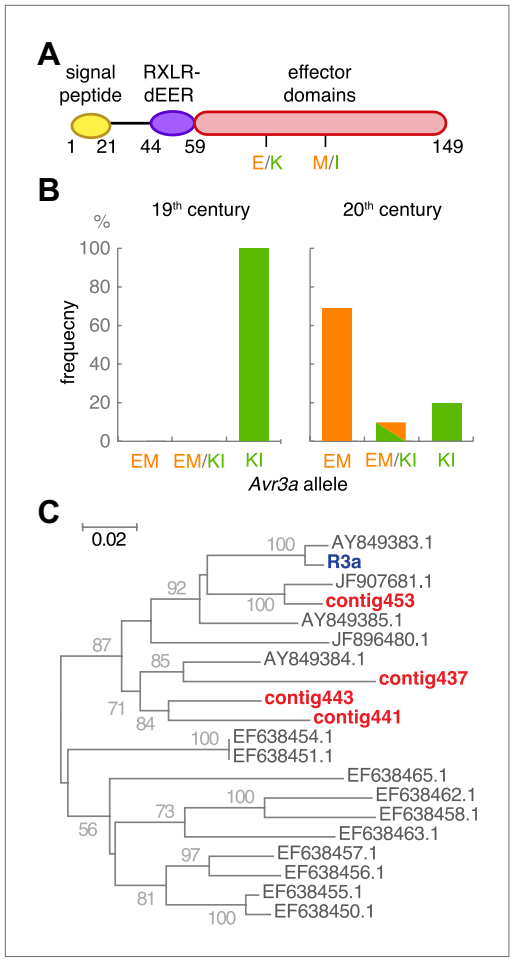


Figure 10. The effector gene *Avr3a* and its cognate resistance gene *R3a*. (**A**) Diagram of AVR3A effector protein. (**B**) Frequency of *Avr3a* alleles in historic and modern *P. infestans* strains. (**C**) Neighbor-joining tree of *R3a* homologs from potato, based on 0.67 kb partial nucleotide sequences of *S. tuberosum R3a* (blue, accession number AY849382.1) and homologs (dark grey) in GenBank, and de novo assembled contigs from M-0182896 (red). Numbers at branches indicate bootstrap support with 500 replicates. Scale indicates changes.

DOI: 10.7554/eLife.00731.023

kmer	N50 (bp)	Longest Contig (bp)	RXLR proteins with TBLASTN hit	AVR1	AVR2	AVR3a^{EM}	AVR4
41	160	2,686	25	48	81	37	66
51	169	4,616	61	39	100	46	78
61	191	8,487	89	100	100	84	87
63	219	9,428	90	77	100	84	87
65	241	9,430	92	100	99	84	87
67	248	9,509	90	100	99	97	79
69	257	15,874	88	100	99	97	79
71	259	14,583	76	100	91	97	45
81	312	36,586	76	100	99	97	45
91	362	73,532	48	75	79	90	44
101	393	68,999	19	75	79	90	7
111	429	68,999	15	75	65	69	7
121	473	25,791	6	7	8	51	7

Matched aa / AVR length (%)

0 50 100

Figure 10—figure supplement 1. Summary of de novo assembly of RXLR effector genes. TBLASTN query was performed with 549 RXLR proteins as a query and contigs as a database. When the High-scoring Segment Pair (HSP) and matched amino acids both covered ≥99% of the query length, we recorded a hit. Results with the optimal *k*-mer size are highlighted.

DOI: 10.7554/eLife.00731.024

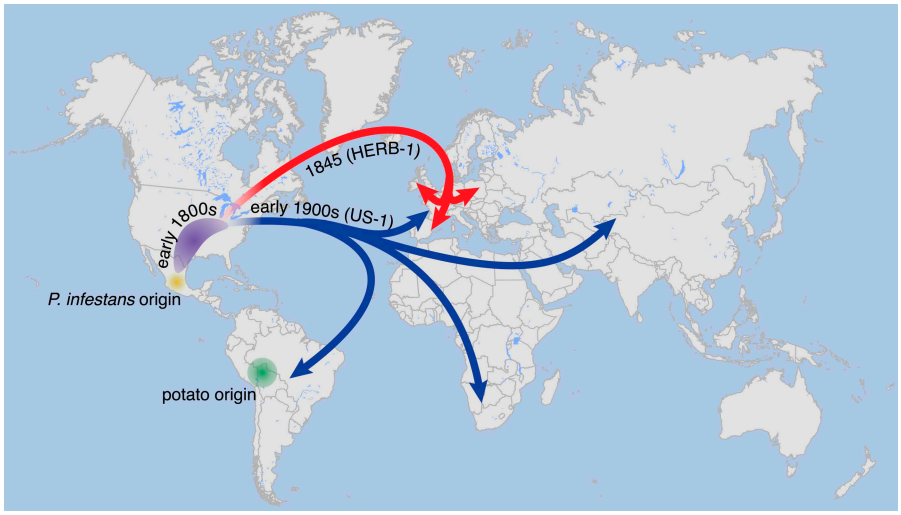


Figure 11. Suggested paths of migration and diversification of *P. infestans* lineages HERB-1 and US-1. The location of the metapopulation that gave rise to HERB-1 and US-1 remains uncertain; here it is proposed to have been in North America.

DOI: 10.7554/eLife.00731.025