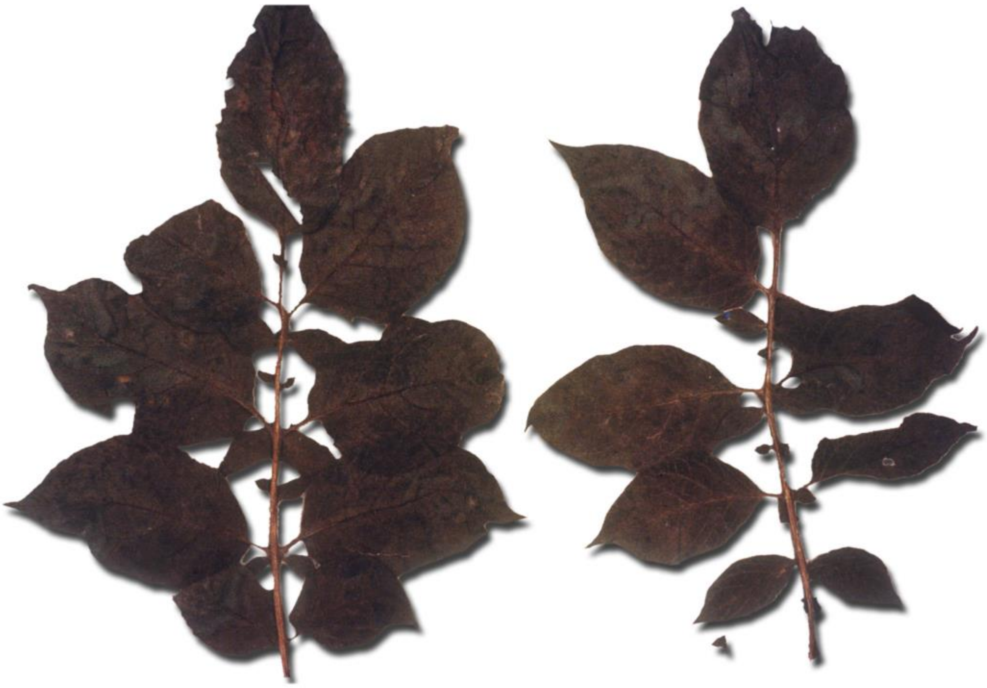
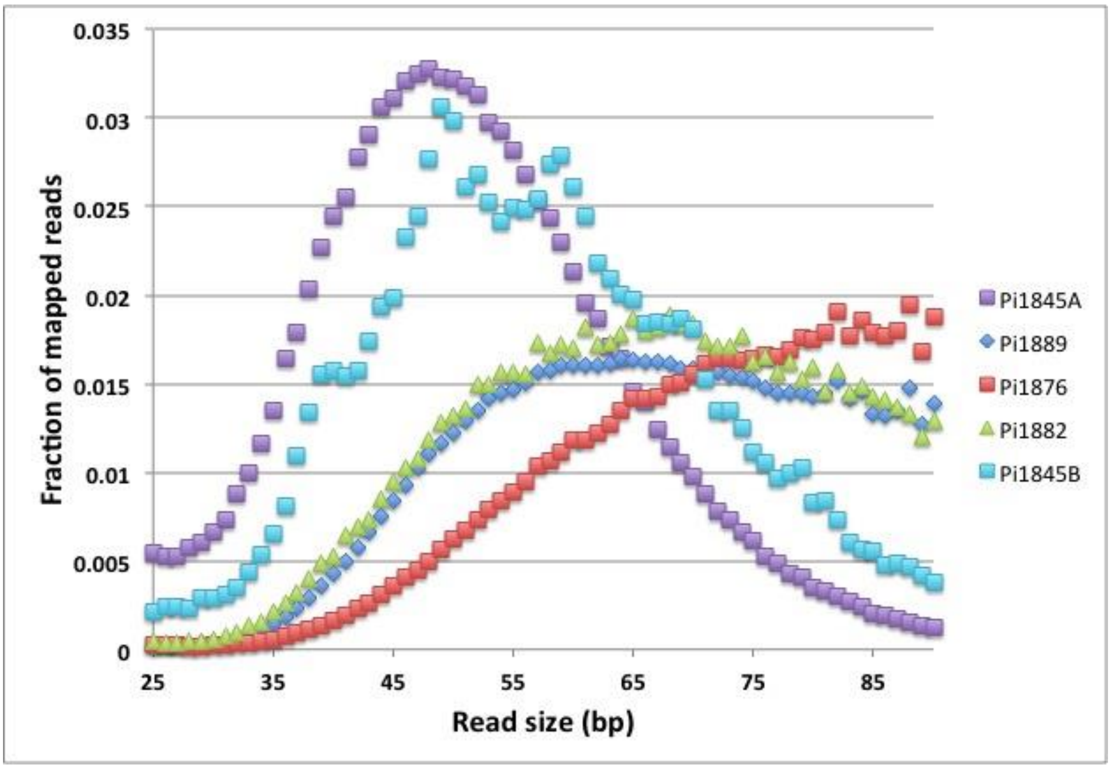


N.° 1198. Botrytis fallax DESMAZ.

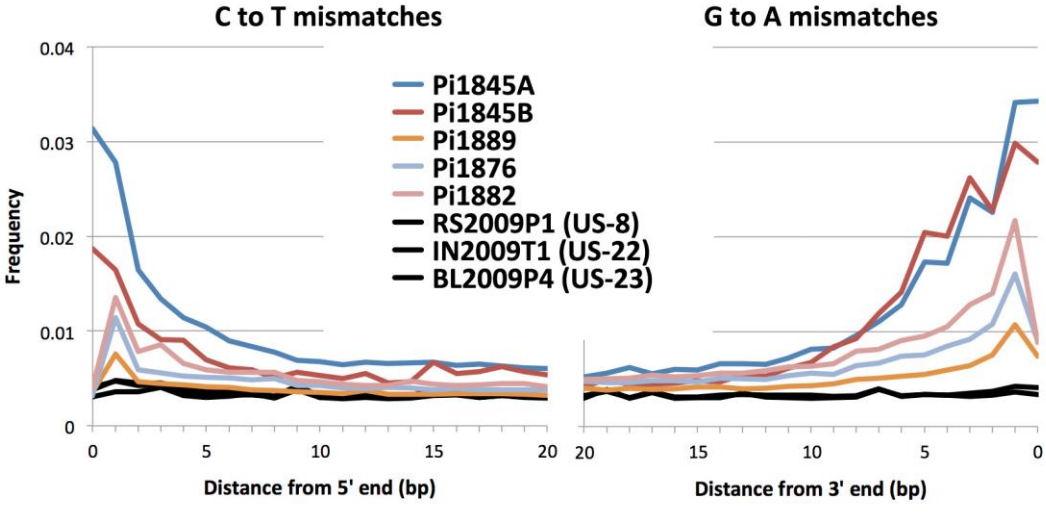
DESMAZ. *Pl. crypt. de France, n.°* 1492. — Kx. *Rech. Fl. crypt. des Fland.*, 3^{e} *cent.*, *pag.* 45, *n.°* 83. — Botrytis infestans *Mont.* — Botrytis vastatrix *Lib.* — Pritchardia solani *Muhlenb.* — Choléra de la Pomme de terre.

Sur les feuilles languissantes du *Solanum tuberosum*, aux environs d'Audenarde. (M.^{r} TOSQUINET.)

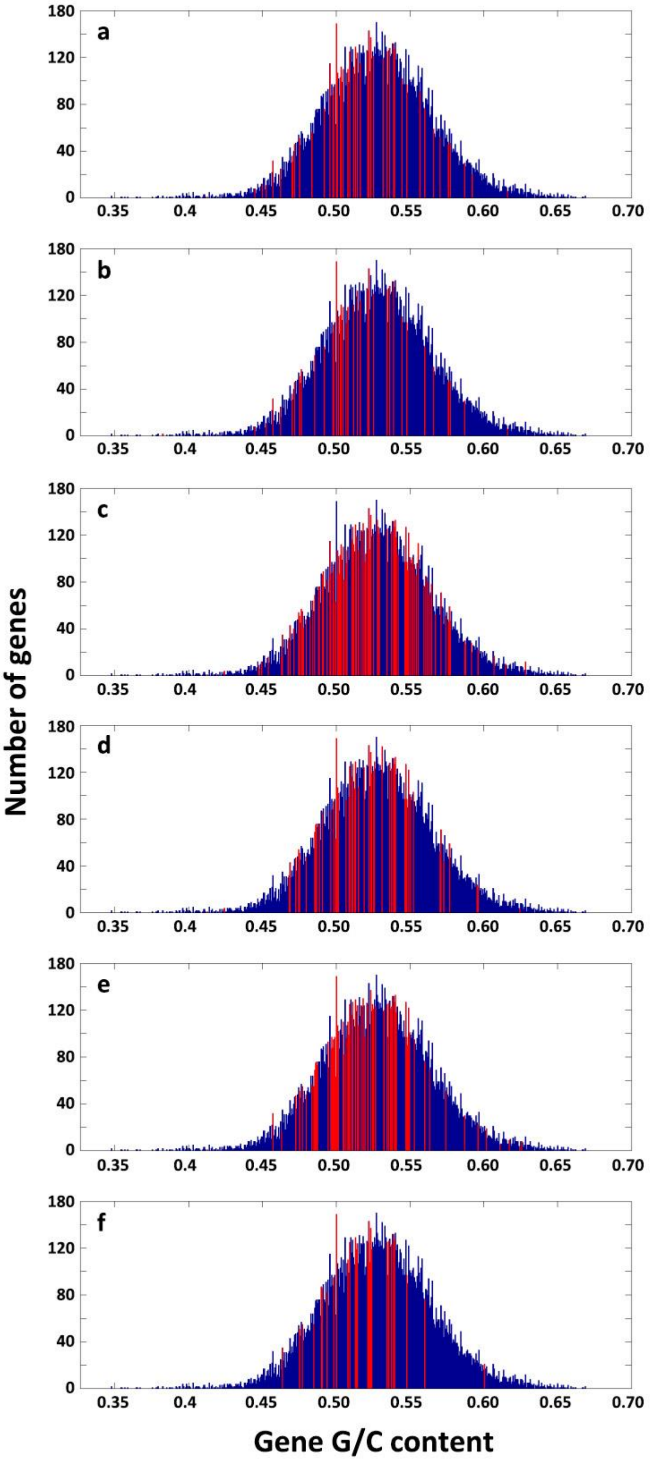
Supplementary Figure S1. Stylized photo of specimen Pi1845A, collected from Belgium in 1845. The specimen is poorly preserved, but blight-induced lesions were discernible on the leaves. The specimen's original label is reproduced below.



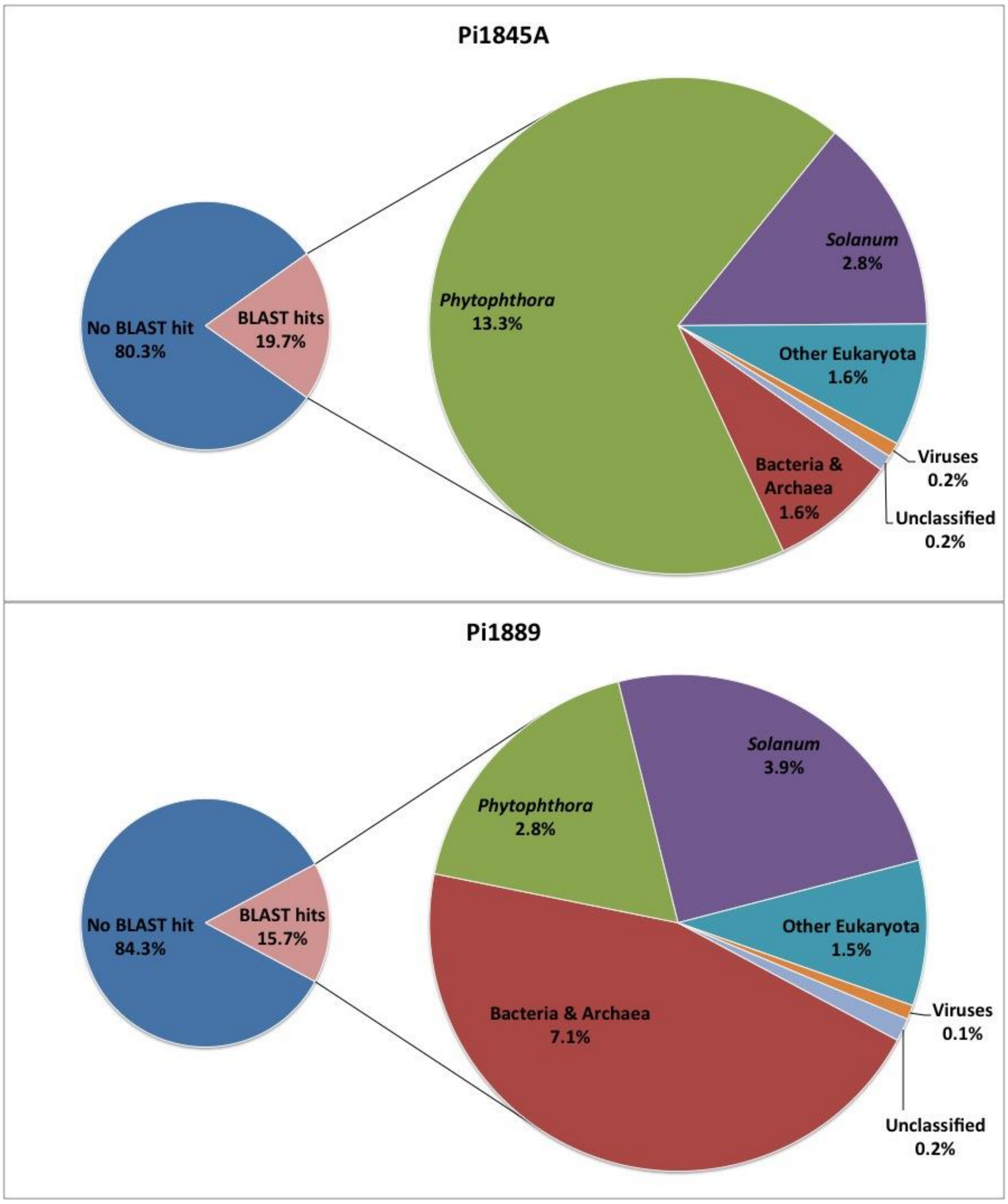
Supplementary Figure S2. Length distribution of reads mapped to *P. infestans* reference genome for each historical sample library. Distributions were drawn from a random sample of trimmed, mapped Single Read sequences from each library after removing PCR duplicates.



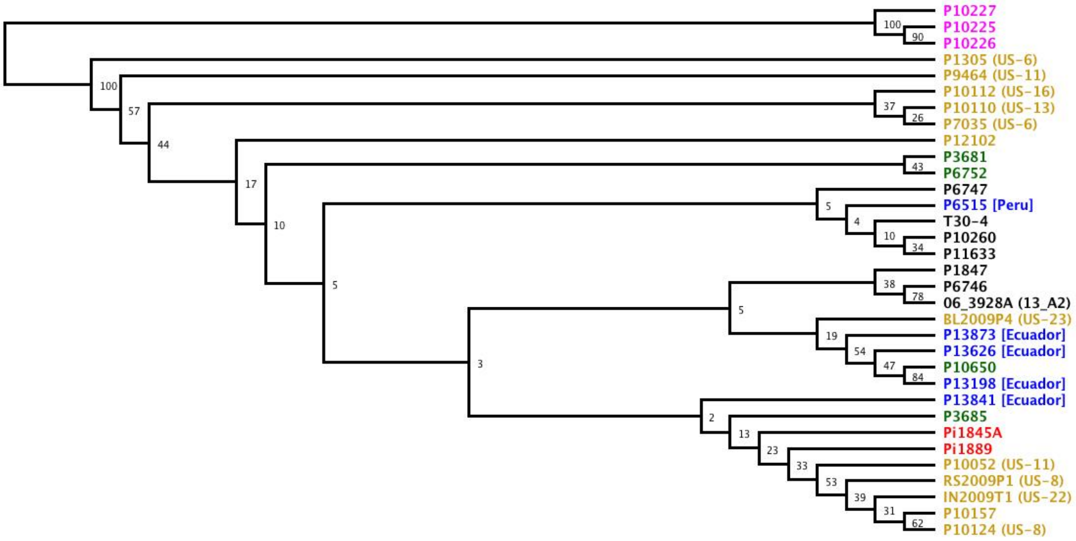
Supplementary Figure S3. Postmortem DNA damage patterns estimated from mismatch rates.



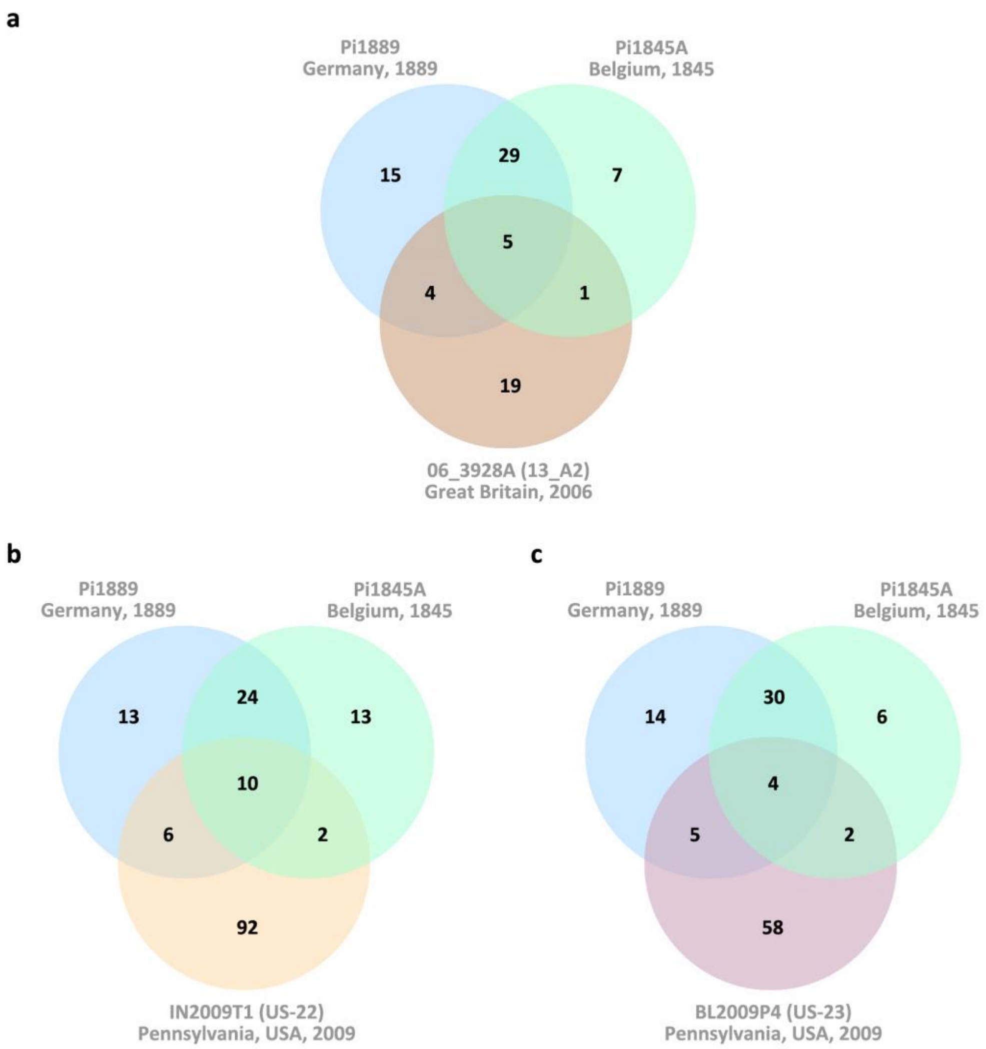
Supplementary Figure S4. G/C content of genes absent from resequenced genomes. Bar width defines a range of G/C values, while bar height represents the number of genes from the T30-4 reference genome in that bin. Genes absent from a resequenced genome fall within bins defined by red bars. 128 reference genes were not included in the analysis because their sequences in the reference genome contained undetermined bases. An additional 15 reference genes are not plotted because they are extreme G/C content outliers. **a**, Pi1845A; **b**, Pi1889; **c**, RS2009P1 (*US-8*); **d**, IN2009T1 (*US-22*); **e**, BL2009P4 (*US-23*); **f**, 06_3928A (*13_A2*).



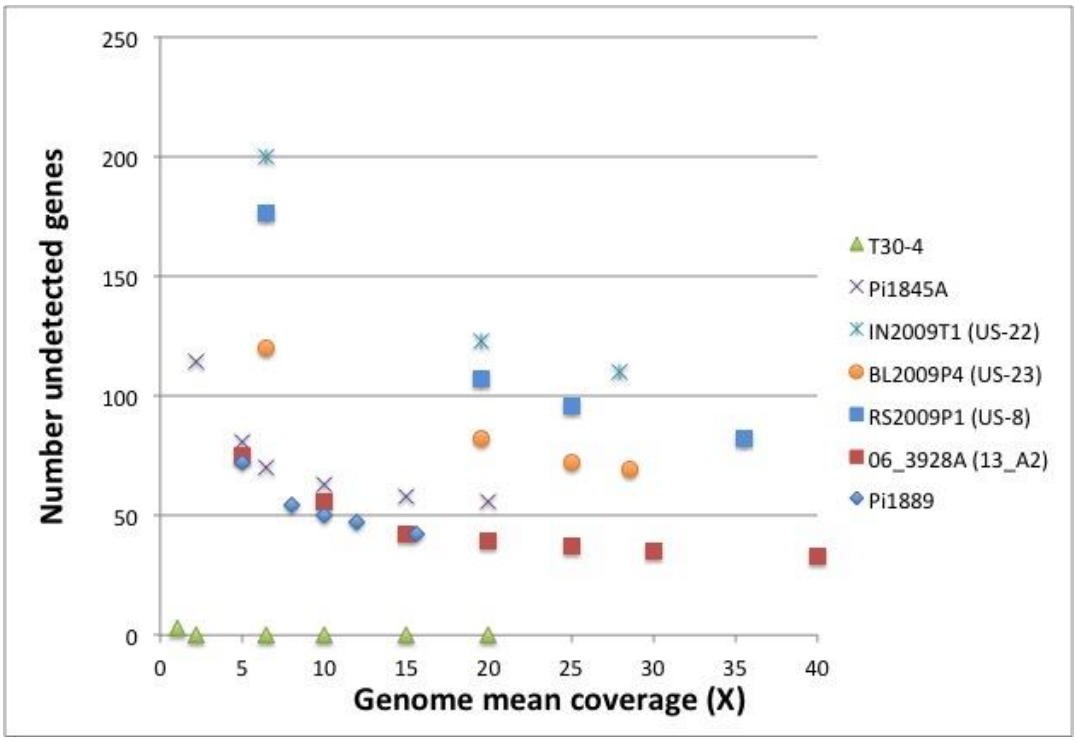
Supplementary Figure S5. Metagenomic composition of historical sample library reads not mapped to the T30-4 reference genome.



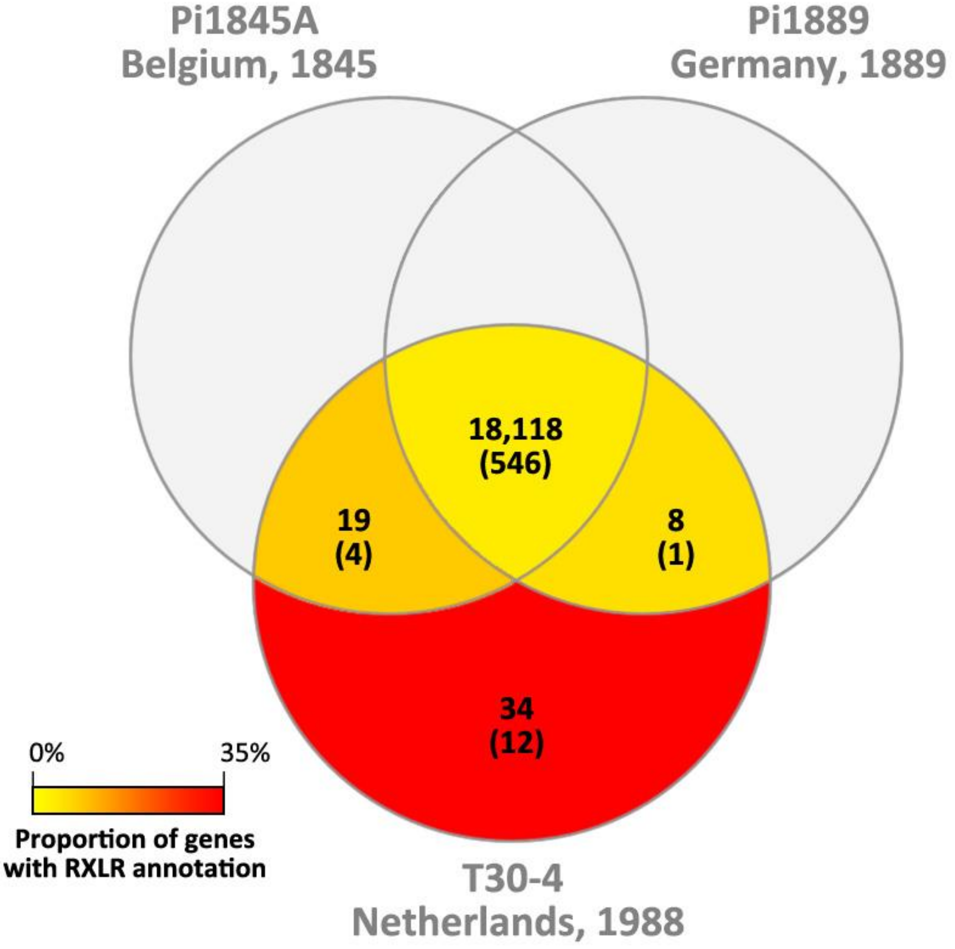
Supplementary Figure S6. Neighbor-joining phylogenetic tree of multi-locus nuclear gene sequences from diverse global isolates of *P. infestans*. Diploid sequence data for ten loci were used to build the tree: 160S ribosomal protein L10, Actin-related protein 2/3 complex, Beta-tubulin, conserved hypothetical protein PUA, Homogentisate 1,2-dioxygenase, TRP1, Phosphatidylinositol-4-phosphate-5-kinase, Pelota, Rab1 family GTPase PiYPT1 and Rab1 family GTPase PiYPT1 intron #1. The taxa labeled pink are outgroup *P. ipomoeae* isolates. Labels of the remaining *P. infestans* taxa are color-coded according to the time and geographic region from which they were collected: gold, USA; green, Mexico; blue, South America; black, modern Europe; red, historical Europe. Where available, multi-locus genotypes for isolates are provided in parentheses[51]. Specific collection location provided in brackets for the South American isolates. Bootstrap support values are indicated at the nodes.



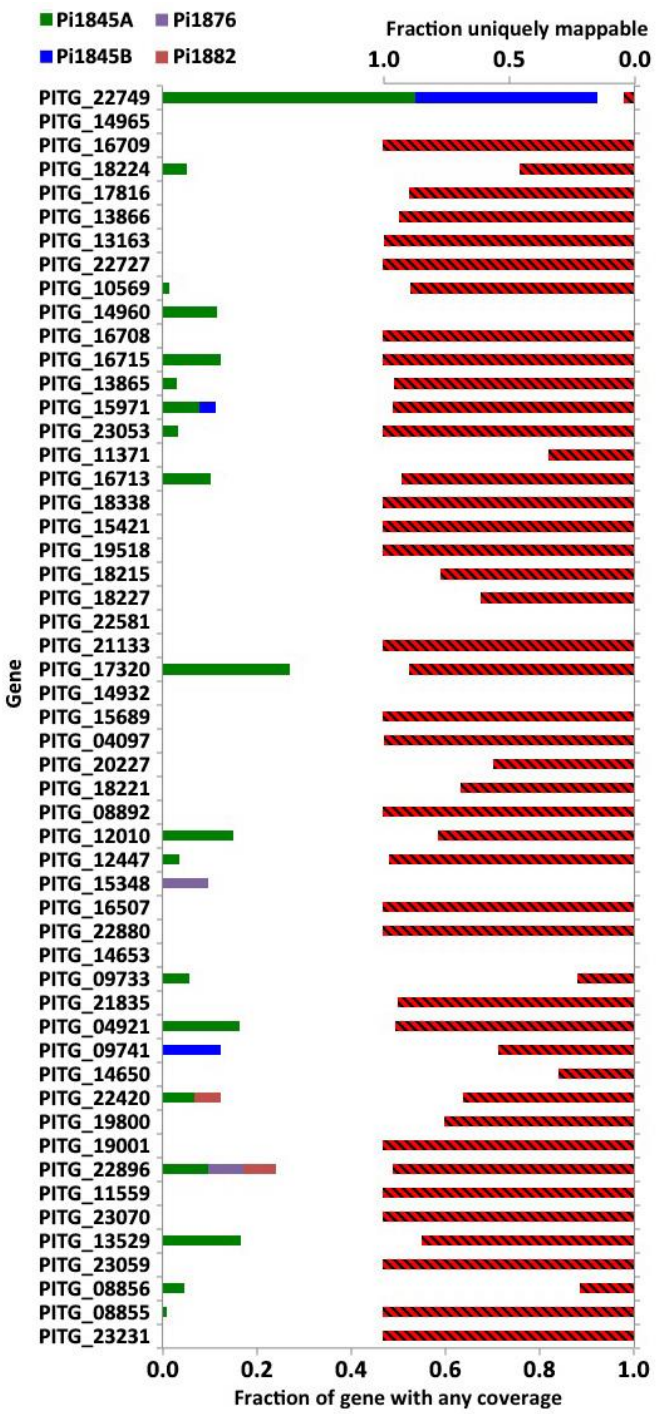
Supplementary Figure S7. Relationships of shared absence of T30-4 reference genes among samples from three different time periods. Circles represent the collection of genes absent from each sample, where overlapping regions are labeled with the number of shared absent genes for those samples whose circles overlap. Historical samples Pi1845A and Pi1889 are compared to modern samples **a**, 06_3928A (*13_A2*) **b**, IN2009T1 (*US-22*) and **c**, BL2009P4 (*US-23*).



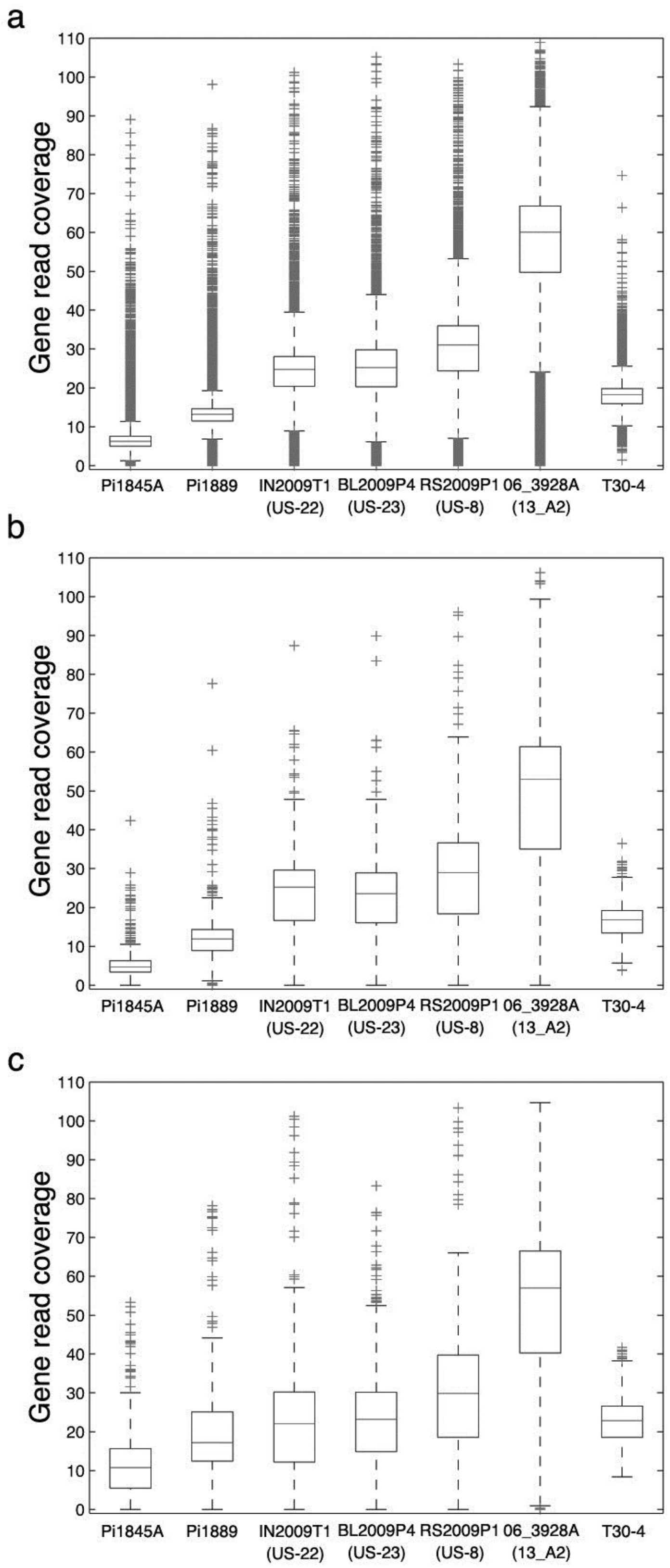
Supplementary Figure S8. Number of undetected genes versus genome average depth of coverage.



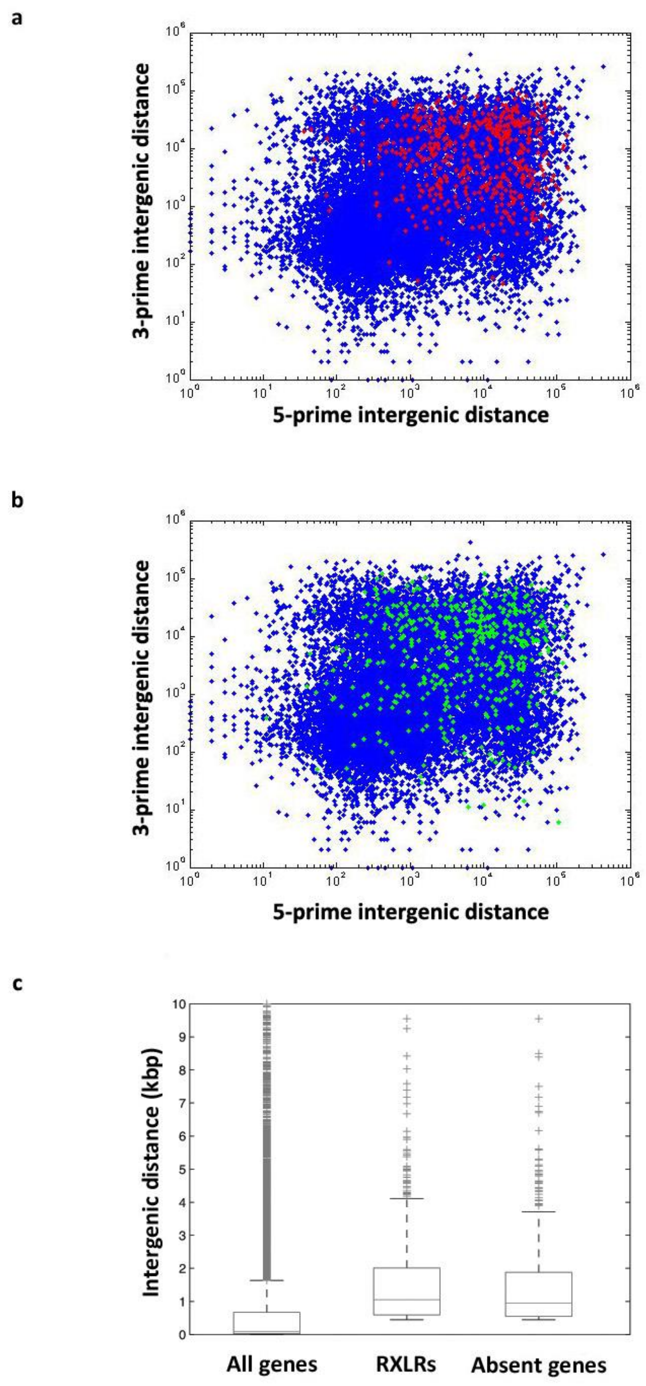
Supplementary Figure S9. Presence of reference genes at three time periods in Europe. Numbers inside overlapping portions of each circle indicate the total number of genes in that category determined to be present by mapping analysis, with the number of RXLRs in that category in parentheses. The colour of the region indicates the proportion of those genes with an RXLR annotation. Grey regions represent hypothetical genes for which data are not available because they are not present in the T30-4 genome assembly. 35 reference genes were not detected in either historical sample Pi1889 or Pi1845A.



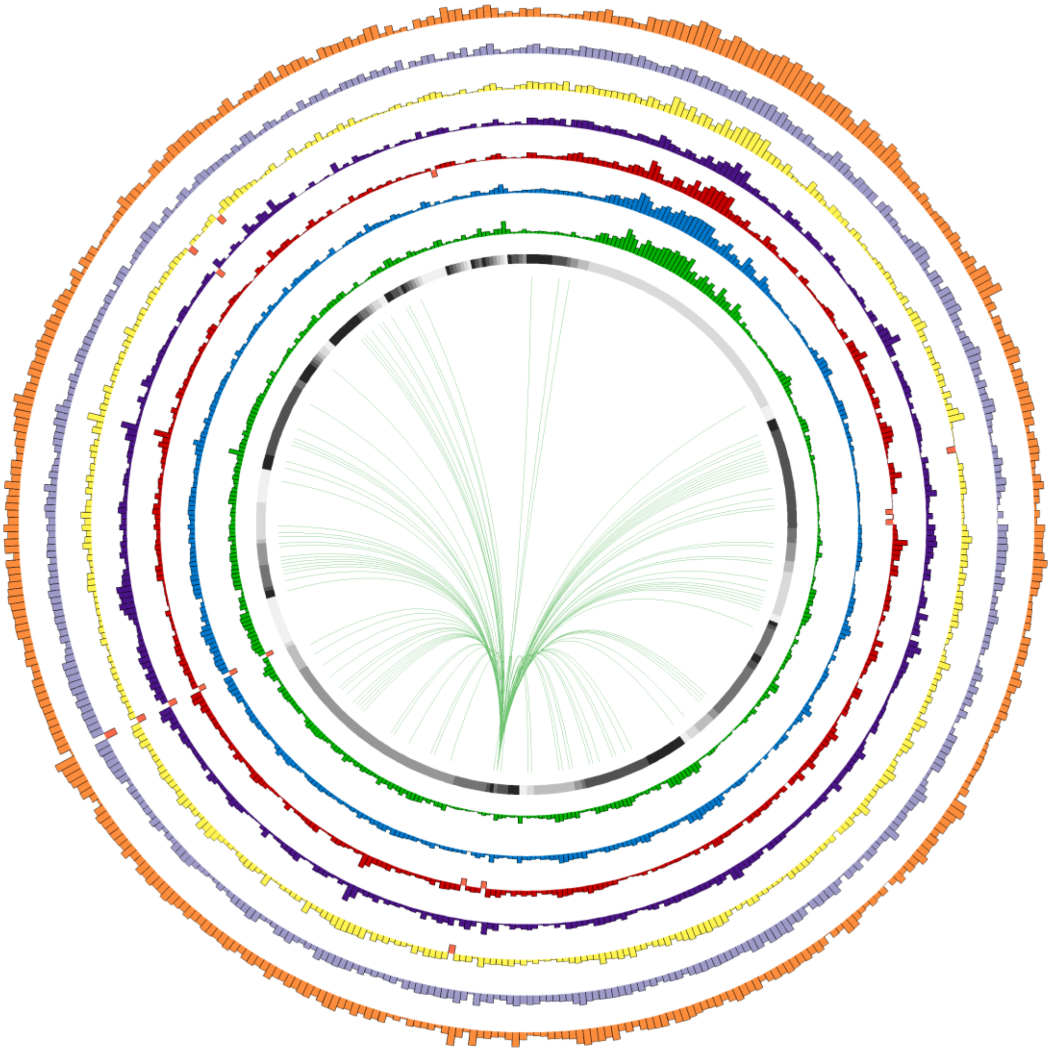
Supplementary Figure S10. Coverage of genes absent in historical sample Pi1889. The leftmost stacked bars represent the fraction of total gene length is covered by at least one sequence read in an older historical sample. The rightmost hashed red bars represent the fraction of total gene length that is uniquely mappable. Because none of these genes is fully covered in the older samples, it suggests that the detection of these genes samples older than Pi1889 could be an experimental artifact related to reads that map to multiple locations in the genome. However, many of these differentially absent genes are almost entirely uniquely mappable.



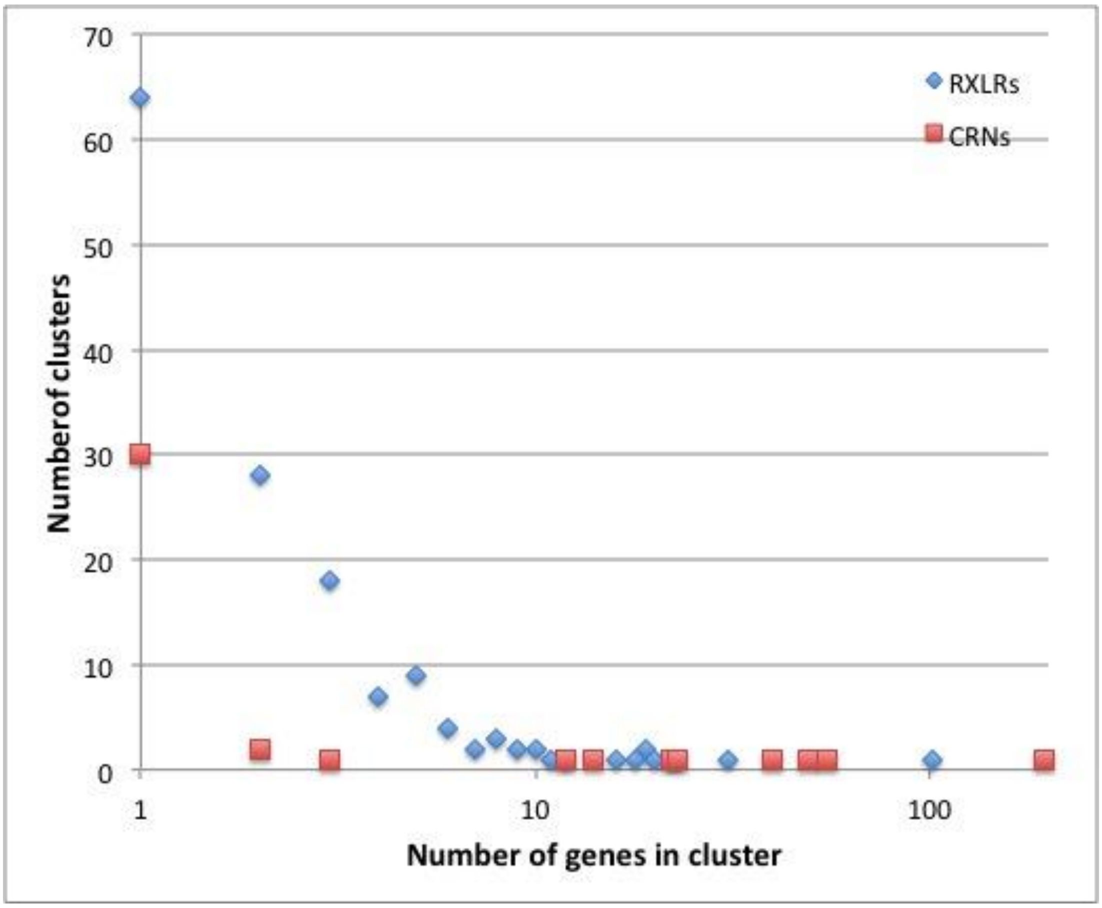
Supplementary Figure S11. Per-gene read depth-of-coverage for resequenced samples. Box and whisker plots show median (horizontal line), first and third quartile (top and bottom of box) and 1.5 times the interquartile distance (whiskers) for **a**, the entire genome; **b**, only RXLR effectors; **c**, only CRN effectors.



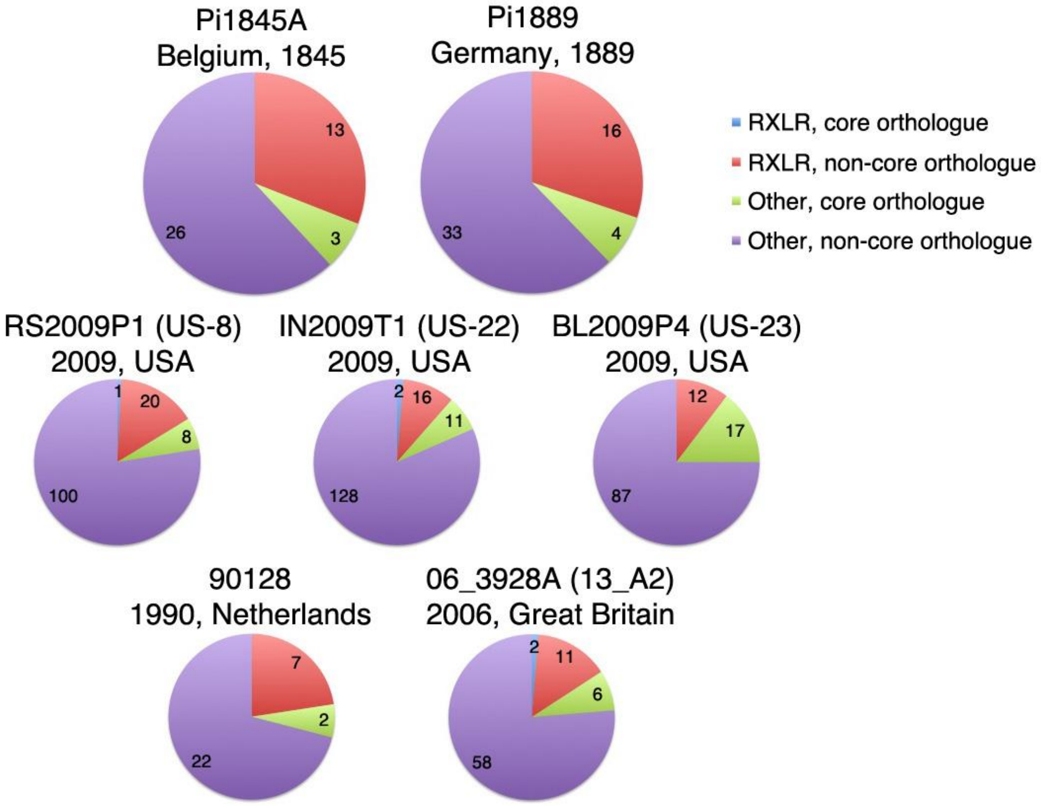
Supplementary Figure S12. Intergenic distances in different gene classes. **a,** Distance to the nearest 5' and 3' neighbor for all reference genes (blue; *n*=18,179) and all reference RXLR effectors (red; *n*=523). **b,** Distance to the nearest 5' and 3' neighbor for all reference genes (blue) reference genes absent in at least one of samples Pi1845A, Pi1889, RS2009P1 (*US-8*), IN2009T1 (*US-22*), BL2009P4 (*US-23*) or 06_3928A (*13_A2*; *n*=473), green. **c**, Box and whisker plots show median (horizontal line), first and third quartile (top and bottom of box) and 1.5 times the interquartile distance (whiskers).



Supplementary Figure S13. Visualization of sequencing coverage distribution across all reference CRN effectors and relatedness among those differentially absent. Bar height represent the mean-normalized coverage of 452 reference CRN effector genes in the resequenced genome of a particular sample. Genes are arranged according to their 5' to 3' physical position on the T30-4 reference genome assembly supercontigs. The black section at the top of the inner ring represents the seven CRNs contained on supercontig 1. CRNs in supercontigs 2, 3, … 4847 follow in clockwise fashion, each represented by a different shade of gray. Historical samples Pi1845A (green) and Pi1889 (blue) are plotted along the two inner rings. The resequenced reference strain T30-4 (orange) is plotted in the outer ring, the remaining inner rings represent modern isolates IN2009T1 (*US-22*; red), BL2009P4 (*US-23*; purple), RS2009P1 (*US-8*; yellow) and 06_3928A (*13_A2*; light purple). Orange bars extending below the axis indicate genes undetected in a particular sample (Supplementary Table 4). The central links connect a gene absent in at least one resequenced genome to all other members of its tribe, with a unique color for each tribe (Supplementary Figure S14).



Supplementary Figure S14. Markov CLuster Algorithm (MCL) gene cluster size and frequency compared between two effector classes.



Supplementary Figure S15. Proportion of conserved strain T30-4 reference genes absent from resequenced *P. infestans* genomes. Core orthologous genes are also present in diverse *Phytophthora* species, as determined by Haas et al (2009)[26]. The number of absent genes in each category is indicated on each piece of the pie.

Supplementary Table S1. Accessions and reads included in the analyses.

Sample name	Species	Accession or isolate ID	Date	Comment	Collection location	Host	Sampling reference	Sequencing reference
Pi1845A	*P. infestans*	K91	1845	mtDNA haplotype Ia	Audernade, Belgium	*S. tuberosum*	9	This report
Pi1845B	*P. infestans*	K16	1845	mtDNA haplotype Ia	Great Britain	*S. tuberosum*	9	This report
Pi1876	*P. infestans*	UPS 1	1876	mtDNA haplotype Ia	Skårop, Denmark	*S. tuberosum*	9	This report
Pi1882	*P. infestans*	UPS 2	1882	mtDNA haplotype Ia	Stockholm, Sweden	*S. tuberosum*	9	This report
Pi1889	*P. infestans*	K 79	1889	mtDNA haplotype Ia	Germany	*S. tuberosum*	9	This report
RS2009P1 (*US-8*)	*P. infestans*	PA117	2009	A2 mating type; US-8 genotype	Pennsylvania, USA	*S. tuberosum*	13	This report
IN2009T1 (*US-22*)	*P. infestans*	PA114	2009	A2 mating type; US-22 genotype	Pennsylvania, USA	*S. tuberosum*	13	This report
BL2009P4 (*US-23*)	*P. infestans*	PA112	2009	A1 mating type; US-23 genotype	Pennsylvania, USA	*S. tuberosum*	13	This report
90128	*P. infestans*	90128	1990	A2 mating type	Netherlands	*S. tuberosum*	-	53
T30-4	*P. infestans*	T30-4	1980/ 1988	Lab-derived F1 progeny of 2 isolates	Netherlands	*S. tuberosum*	68	53
06_3928A (*13_A2*)	*P. infestans*	06_3928A	2006	A2 mating type; genotype 13_A2	Great Britain	*S. tuberosum*	12	12
P. mirabilis	*P. mirabilis*	PIC99114	-	-	Mexico	*Mirabilis jalapa*	-	53

Supplementary Table S2. Genome coverage statistics. *, after PCR duplicate removal. SR, Single Read. PE, Paired End.

	Number of sequence reads (millions)					
		Sequencing platform				
Sample name	***Total***	*Illumina GAII*	*Illumina HiSeq2000*	**Mapped to T30-4 genome sequence**	**Per-base mean depth of coverage***	**Sequencing reference**
Pi1845A	1,386.0	-	969.0 SR 417.0 PE	183.8 (13%)	16X	This report
Pi1845B	383.9	-	76.6 SR 307.2 PE	50.6 (13%)	3X	This report
Pi1876	552.4	-	182.1 SR 370.2 PE	44.9 (8%)	7X	This report
Pi1882	577.7	-	203.0 SR 374.7 PE	34.7 (6%)	3X	This report
Pi1889	797.6	27.6 SR	196.0 SR 429.5 PE	132.5 (17%)	22X	This report
RS2009P1 (*US-8*)	70.5	-	70.5 PE	65.3 (93%)	36X	This report
IN2009T1 (*US-22*)	80.4	-	80.4 PE	72.8 (91%)	28X	This report
BL2009P4 (*US-23*)	100.8	-	100.8 PE	78.7 (78%)	29X	This report
90128	13.5	13.5 SR	-	9.6 (76%)	2X	53
06_3928A (*13_A2*)	221.6	221.6 SR	-	215.7 (97%)	62X	12
T30-4	258.6	258.6 SR	-	239.2 (93%)	20X	53
P. mirabilis	172.3	172.3 SR	-		11X	53

Supplementary Table S3. Genes absent in either sample Pi1845A or Pi1889 and their status in other samples. *P*, present. *A*, absent. All p-values were significant ($p < 0.001$). A list of genes induced or repressed more than 2-fold during T30-4 infection of *S. tuberosum* were retrieved from Haas et al (2009)[26]. *GCC, G/C content of gene. PUM, Percentage of gene length that is uniquely mappable.*

						Sample ID													
Gene ID	**Description**	**Gene length (bp)**	**GCC (%)**	**PUM (%)**	**T30-4 super-contig**	**Pi1845B**	**Pi1845A**	**Pi1876**	**Pi1882**	**Pi1889**	**IN2009T1 *(US-22)***	**BL2009P4 *(US-23)***	**RS2009P1 *(US-8)***	**06_2938A *(13_A2)***	**90128**	**T30-4**	**Absent in Pi1889, present in older historical sample**	**Absent in Pi1889, present in Pi1845A**	**Induced or repressed during infection**
PITG_04921	hypothetical protein	456	59.2	95.0	6	A	P	A	A	A	P	P	P	P	P	P	**X**	**X**	
PITG_22749	conserved hypothetical protein	1542	53.9	4.3	6	P	P	A	A	A	P	P	P	P	P	P	**X**	**X**	
PITG_09733	hypothetical protein	1546	50.9	22.4	15	A	P	A	A	A	P	P	A	P	P	P	**X**	**X**	
PITG_22896	secreted RxLR effector peptide, putative	585	48.4	96.1	15	A	P	P	P	A	P	P	A	P	A	P	**X**	**X**	
PITG_08855	ATP-binding Cassette (ABC) Superfamily	2869	52.4	100.0	16	A	P	A	A	A	A	A	A	A	A	P	**X**	**X**	
PITG_08856	conserved hypothetical protein	639	51.3	21.4	16	A	P	A	A	A	A	A	A	A	A	P	**X**	**X**	
PITG_10569	hypothetical protein	341	54.0	89.1	18	A	P	A	A	A	A	P	A	A	P	P	**X**	**X**	
PITG_12010	secreted RxLR effector peptide, putative	252	54.0	77.8	21	A	P	A	A	A	A	P	P	P	P	P	**X**	**X**	
PITG_12447	conserved hypothetical protein	1451	52.2	97.5	26	A	P	A	A	A	P	P	P	A	P	P	**X**	**X**	**X**
PITG_13529	secreted RxLR effector peptide, putative	489	50.3	84.3	29	A	P	A	A	A	P	P	P	P	A	P	**X**	**X**	
PITG_13865	cytochrome P450, putative	1545	50.3	95.3	30	A	P	A	A	A	P	P	P	P	P	P	**X**	**X**	
PITG_14960	secreted RxLR effector peptide, putative	498	49.6	0.0	33	A	P	A	A	A	P	P	P	P	P	P	**X**	**X**	
PITG_23053	NPP1-like protein, pseudogene	757	50.1	100.0	37	A	P	A	A	A	A	A	A	P	P	P	**X**	**X**	
PITG_15971	hypothetical protein	898	55.7	95.9	39	P	P	A	A	A	P	P	P	P	P	P	**X**	**X**	
PITG_16713	RXLR effector family protein, putative	460	45.0	92.6	51	A	P	A	A	A	P	P	P	P	P	P	**X**	**X**	
PITG_16715	hypothetical protein	498	51.6	100.0	51	A	P	A	A	A	P	P	P	P	P	P	**X**	**X**	
PITG_17320	hypothetical protein	328	57.0	89.6	53	A	P	A	A	A	P	A	P	P	P	P	**X**	**X**	
PITG_18224	conserved hypothetical protein	461	53.1	45.3	66	A	P	A	A	A	P	P	P	P	P	P	**X**	**X**	
PITG_22420	conserved hypothetical protein	528	41.3	68.2	2991	A	P	A	P	A	A	A	A	P	P	P	**X**	**X**	**X**
PITG_02837	conserved hypothetical protein	399	50.1	46.1	3	A	A	P	A	P	P	P	P	P	P	P			
PITG_07583	conserved hypothetical protein	640	53.9	66.4	11	A	A	A	A	P	A	A	P	A	P	P			
PITG_09739	secreted RxLR effector peptide, putative	381	50.7	93.7	15	A	A	A	A	P	P	P	A	P	A	P			
PITG_09030	dynein light chain, putative	333	52.6	10.5	16	A	A	P	A	P	A	P	A	P	A	P			
PITG_14334	conserved hypothetical protein	303	38.3	1.0	34	A	A	P	P	P	P	P	P	P	P	P			

PITG_19936	NPP1-like protein	750	52.5	100.0	140	A	A	A	A	P	P	P	P	P	P	P			
PITG_20511	conserved hypothetical protein	422	47.6	1.4	146	A	A	A	P	P	P	P	P	P	P	P			
PITG_22489	hypothetical protein	847	48.5	79.5	3538	A	A	A	A	P	P	A	P	P	P	P			
PITG_04097	secreted RxLR effector peptide, putative	324	50.3	99.4	5	A	A	A	A	A	P	P	P	P	P	P			**X**
PITG_22727	secreted RxLR effector peptide, putative	213	50.3	100.0	5	A	A	A	A	A	P	P	P	P	P	P			**X**
PITG_22880	secreted RxLR effector peptide, putative	606	47.2	100.0	14	A	A	A	A	A	P	P	P	P	P	P			
PITG_09741	secreted RxLR effector peptide, putative	414	50.5	53.9	15	P	A	A	A	A	P	P	A	P	P	P	**X**		
PITG_08892	conserved hypothetical protein	975	57.7	100.0	16	A	A	A	A	A	A	P	P	P	P	P			
PITG_11559	hypothetical protein	521	50.5	100.0	22	A	A	A	A	A	P	P	A	P	A	P			
PITG_11371	conserved hypothetical protein	744	44.5	33.9	23	A	A	A	A	A	P	P	P	P	P	P			
PITG_13163	P-type ATPase (P-ATPase) superfamily	621	46.2	99.7	25	A	A	A	A	A	P	P	P	P	P	P			
PITG_13866	conserved hypothetical protein	303	50.8	93.4	30	A	A	A	A	A	P	P	P	P	P	P			
PITG_14932	secreted RxLR effector peptide, putative	270	50.4	0.0	33	A	A	A	A	A	P	P	P	P	P	P			
PITG_14965	secreted RxLR effector peptide, putative	270	50.0	0.0	33	A	A	A	A	A	P	P	P	P	P	P			
PITG_15348	Mitochondrial Carrier (MC) Family	1043	49.2	0.0	35	A	A	P	A	A	P	P	P	P	P	P	**X**		
PITG_14650	conserved hypothetical protein	929	-	30.0	36	A	A	A	A	A	A	P	A	P	P	P			
PITG_14653	conserved hypothetical protein	612	52.1	0.0	36	A	A	A	A	A	A	P	A	P	P	P			
PITG_15421	hypothetical protein	448	47.5	100.0	37	A	A	A	A	A	A	A	A	P	P	P			
PITG_15689	hypothetical protein	465	57.6	100.0	40	A	A	A	A	A	P	P	P	P	P	P			
PITG_23059	cys-rich secreted peptide, putative	219	53.4	100.0	40	A	A	A	A	A	A	P	A	A	A	P			**X**
PITG_16507	glycoside hydrolase, putative	1749	52.2	100.0	44	A	A	A	A	A	A	A	P	P	P	P			
PITG_23070	PcF and SCR74-like cys-rich secreted peptide, putative	213	58.7	100.0	45	A	A	A	A	A	A	P	A	P	A	P			
PITG_16708	secreted RxLR effector peptide, putative	498	47.0	100.0	51	A	A	A	A	A	P	P	P	P	P	P			
PITG_16709	conserved hypothetical protein	414	61.6	100.0	51	A	A	A	A	A	P	P	P	P	P	P			
PITG_17816	Crinkler (CRN) family protein, pseudogene	1701	49.9	89.6	56	A	A	A	A	A	A	A	A	A	P	P			
PITG_18215	secreted RxLR effector peptide, putative	447	51.5	77.0	66	A	A	A	A	A	P	P	P	A	P	P			**X**
PITG_18221	secreted RxLR effector peptide, putative	318	56.0	69.2	66	A	A	A	A	A	P	P	A	A	P	P			**X**
PITG_18227	conserved hypothetical protein	1032	-	61.0	66	A	A	A	A	A	P	P	P	P	P	P			

PITG_18338	glycoside hydrolase, putative	1095	53.6	100.0	67	A	A	A	A	A	P	P	P	P	P	P			
PITG_19001	PcF and SCR74-like cys-rich secreted peptide, putative	312	52.6	100.0	74	A	A	A	A	A	P	P	P	P	A	P			
PITG_19518	secreted RxLR effector peptide, putative	804	49.8	100.0	95	A	A	A	A	A	A	P	A	P	P	P			
PITG_20227	hypothetical protein	1420	45.7	56.1	126	A	A	A	A	A	P	P	A	P	P	P			
PITG_19800	secreted RxLR effector peptide, putative	621	45.2	75.5	130	A	A	A	A	A	P	P	P	P	P	P			
PITG_21133	hypothetical protein	905	56.6	100.0	260	A	A	A	A	A	P	P	P	P	P	P			
PITG_21835	hypothetical protein	755	51.7	93.8	539	A	A	A	A	A	P	P	P	P	P	P			
PITG_23231	secreted RxLR effector peptide, putative	219	54.8	100.0	2210	A	A	A	A	A	A	A	A	A	A	P			
PITG_22581	conserved hypothetical protein	393	54.5	0.0	4154	A	A	A	A	A	P	P	P	P	P	P			
						224	*42*	*136*	*145*	*53*	*110*	*69*	*82*	*29*	*31*	*0*	**Total number of undetected genes**		
						16X	*3X*	*7X*	*3X*	*22X*	*28X*	*29X*	*36X*	*62X*	*2X*	*20X*	**Genome mean depth of coverage**		

Supplementary Table S4. Statistics of the *de novo* assemblies of unmapped reads.

Assembly	**Number of input unmapped reads**	**Assembly N_{50} (bp)**	**Number of assembled contigs longer than 200bp**	**Largest contig length (bp)**	**Total length (Mbp)**
Pi1889, block 1	150,000,000	323	435,946	31,703	134.95
Pi1889, block 2	150,000,000	311	519,765	48,275	162.97
Pi1889, block 3	179,186,264	312	371,872	18,111	110.92
Pi1845A, block 1	150,000,000	297	8,856	2,490	2.60
Pi1845A, block 2	172,895,913	625	4,922	2,945	1.53
RS2009P1 (*US-8*)	22,145,260	473	11,505	13,291	5.29
IN2009T1 (*US-22*)	7,614,824	223	11,302	11,820	5.21
BL2009P4 (*US-23*)	5,216,036	3,058	9,850	10,817	4.63

Supplementary Table S5. Novel, non-reference RXLR protein sequences mined from *de novo* assembly of unmapped reads. RXLR-EER motif residues are indicated in bold.

Novel gene ID	**Note**	**Source contig coverage**	**Closest match in T30-4 genome**		**Peptide length**	**Peptide sequence**
			Gene ID	**Amino acid identity (%)**		
PiRXLRa	*Found in RS2009P1 (US-8) and IN2009T1 (US-22) assemblies.*	10.42X (US-8) 15.76X (US-22)	PITG_ 22990	82	82	MYLIYVLLTVAALLRSCCASVTAGYSVAGLAQGGSDK**RILR**DNQRDNVDRAMDVAAD**EER**AGQGLVNRILGRNDLKMEKLSR
PiRXLRb	*Found in RS2009P1 (US-8) assembly.*	4.35X	*N/A*	*N/A*	131	MRTSITLLVLVVALLARHIAAATEASPR**RLLR**RHDTVENTAERDSTNT**EER**GITTFVKDFLKKRQIKQLVTKNKTNKQIFDKGITPNEVWFALNIPKLQSKISLDQLGKHPKLIAWRNYDAYVARTMRGNI
PiRXLRc	*Found in RS2009P1 (US-8) assembly.*	8.54X	PITG_ 22798	66	187	MRRCYILIAIAVVLSGIASVVADSSQDKLMAVEGDQTTGTVN**RFLR**RDDELSAENT**EER**IVAGDIPLSARMINNIYKVEKRIVDPKLADELLEKPGLKTLKTHLDAALPYSERAKVFERWHADGVDPSSITKALKVHPAIAKKYNTVSTMYDLYVKSAAIKRLTELKRKSDNDLADAVRLKRQRINE
PiRXLRd	*Found in IN2009T1 (US-22) assembly.*	10.44X	PITG_ 16663	42	104	MRFMRLLLLLMAFAMLNITKAAELDHTSSDCGG**RQLR**TITTNDKEQRSLSIPSAIESGITRWRIKIWMRNKMPDHSVLENLALAGVTGKTLTRDPKFKIFQKVK
PiRXLRe	*Found in BL2009P4 (US-23) assembly.*	5.53X	PITG_ 22990	81	82	MYLIYVLLTVAALLRSCCASVTAGYSVAGLAQGGSDK**RLLR**DNQRDNVDRAMDVAAD**EER**AGQGLVNRILGRNDLKMEKLSR

Supplementary Table S6. Results of PCR experiments to confirm the presence of novel assembled RXLR genes in modern isolates of *P. infestans*. Loci PiRXLRa and PiRXLRe were amplified using identical primers.

				Expect PCR band?			PCR results		
Locus	**Expected size (bp)**	**Forward primer**	**Reverse primer**	**RS2009P1 (US-8)**	**IN2009T1 (US-22)**	**BL2009P4 (US-23)**	**RS2009P1 (US-8)**	**IN2009T1 (US-22)**	**BL2009P4 (US-23)**
PiRXLRa	294	PiRXLRa_F	PiRXLRa_R	+	+	+	+	+	+
PiRXLRb	478	PiRXLRb_F	PiRXLRb_R	+	-	-	+	-	-
PiRXLRc	900	PiRXLRc_F	PiRXLRc_R	+	-	-	+	-	+
PiRXLRd	333	PiRXLRd_F	PiRXLRd_R	-	+	-	-	+	-
PiRXLRe	294	PiRXLRa_F	PiRXLRa_R	+	+	+	+	+	+

Supplementary Table S7. Variant amino acid sequences of putative avirulence genes. The sites of non-synonymous mutation relative to the T30-4 reference genome sequence are standard amino acid residue one-letter codes and the 1-based codon position from the start codon. Premature stop codons are indicated by *. The alphabetic allele notation corresponds to that of Table 1 in the main text.

Locus ID	*PITG_14371*
Annotation	*Avr3a*
Allele A	C19 K80 I103
Allele B	S19 E80 M103
Locus ID	*PITG_14368*
Annotation	*Avr3a-like*
Allele A	G109
Allele B	R109
Locus ID	*PITG_14374*
Annotation	*Avr3a-like*
Allele A	I27 F56 *70
Allele B	V27 S56 E70
Locus ID	*PITG_21388*
Annotation	*Avrblb1*
Allele A	T11 G30 D32 A68 N78 ?117 ?122 P129 A134 G135 N143
Allele B	N11 V30 Y32 S68 K78 S117 R122 L129 A134 S135 K143
Allele C	N11 V30 Y32 S68 K78 L117 I122 L129 G134 S135 N143
Locus ID	*PITG_06077*
Annotation	*Avr2 family*
Allele A	S13 Q104 L119
Allele B	A13 L104 *119
Allele C	A13 Q104 *119
Locus ID	*PITG_05121*
Annotation	*Avr2 family*
Allele A	E56 D58 R81 I90 R113
Allele B	E56 D58 R81 V90 G113
Allele C	K56 ?58 E81 V90 G113
Allele D	K56 ?58 E81 V90 R113
Allele E	K56 ?58 E81 V90 R113

Supplementary Methods

Estimation of a phylogenomic tree

Phylogenetic inference was conservatively restricted to positions in the *P. infestans* reference genome that are uniquely mappable (mappability score = 1) as determined by the mappability tool in GEnome Multitool (GEM) library[43] when using a with a *k*-mer size of 65bp (the approximate overall mean mapped read size for the historical samples) and a maximum number of 3 mismatches (the maximum number of mismatches allowed when mapping a read of this size in BWA). These uniquely mappable regions of the genome amounted to 66.72Mbp (35.12% of the total contig length of the T30-4 reference genome assembly). Reads aligned to the reference genome with a MAPQ score less than 25 were not included in phylogenetic analyses. Insertion/deletion events were ignored.

A multiple sequence alignment (MSA) of the 11 taxa with mean genome coverage greater than or equal to 3X in the phylogenetic analysis was populated with polymorphic bases called from VCF files with GATK's UnifiedGenotyper tool[44] using a minimum variant confidence score of Q30 and minimum coverage of 1 read. To mitigate any bias that may have been present due to different values of average genomic depth of coverage between taxa, heterozygous positions were randomly called as one of the two alleles. One artifact of this technique is that it increases the total branch length to all the tips of the phylogenetic tree. Base positions not meeting quality thresholds were coded as N (missing data) for that taxon. The final alignment consisted of 127,213 SNP positions. RAxML version 7.3.5[45] was used with the GTRGAMMA substitution model (General Time Reversible model with the Γ model of rate heterogeneity[46] and four discrete rate categories). A consensus tree was created from 100 bootstrap replicates (Figure 1). *P. mirabilis* isolate PIC99114 was manually assigned to be the root of the phylogenetic tree.

The tree topology, which supports the historical strains as a monophyletic group distinct from all modern isolates, was robust to different methods for tree generation:

a. To investigate the possibility that sequencing errors related to postmortem DNA damage might affect the tree topology, the MSA (with transitions masked) was converted to a binary alignment where Y (C or T) = 0 and R (G or A) = 1. This alignment should be considered extremely conservative, where GC>AT substitution errors, related to damage, are reduced. The alignment was reduced to 34,569 SNP positions. RAxML was used with the BINGAMMA substitution model. In the consensus tree after 100 bootstrap replicates, the topology remained unchanged except that strain T30-4 was a poorly supported (30%) sister taxon to the group containing all other modern isolates. All other nodes remained supported at 98% or above.

b. The tree was also generated with PHYML version 3.0[47] using the best-fitting model of nucleotide substitution (TVM) as determined by jModelTest version 2.0.2[48,49] with the Akaike Information Criterion[50]. In the consensus tree after 100 bootstrap replicates, the topology was unchanged, with all nodes receiving 100% bootstrap support.

c. An alignment of only positions containing transversion substitutions from supercontigs 1.1-1.4, and totaling 22,049 SNPs, yielded the same topology. Positions containing transitions were completely masked from the alignment in order to test the influence of postmortem DNA damage. After 100 bootstrap replicates, all nodes were highly supported.

d. An alignment of supercontigs 1.1-1.4 where all positions containing autapomorphies (character states unique to only one taxon) were completely masked, totaling 32,520 SNP positions, yielded the same topology. After 100 bootstrap replicates, all nodes were highly supported.

Estimated position of historical isolates in global phylogeny

To more fully provide context for understanding the phylogeny presented in Figure 1, we sought to place our historical samples in a more complete phylogenetic tree including *P. infestans* samples from diverse global sources. For this we obtained GENBANK sequence accessions from the most complete intra-species, multi-locus nuclear gene phylogeny yet published for *P. infestans*, which used 15 nuclear loci[51]. We mined diploid sequences of these loci from our VCF files using custom scripts and chose the 10 loci (60S ribosomal protein L10, Actin-related protein 2/3 complex, Beta-tubulin, conserved hypothetical protein PUA, Homogentisate 1,2-dioxygenase, TRP1, Phosphatidylinositol-4-phosphate-5-kinase, Pelota, Rab1 family GTPase PiYPT1, Rab1 family GTPase PiYPT1 intron #1) that were well covered in samples Pi1845A, Pi1889, RS2009P1 (*US-8*), IN2009T1 (*US-22*), BL2009P4 (*US-23*) and T30-4. Coverage of polymorphic sites within these loci was too poor in samples Pi1876, Pi1882 and Pi1845B to be included in the alignment. These sequences were manually aligned to corresponding sequences from 24 North American, South American and European isolates of *P. infestans* from the previously published dataset, and three additional outgroup isolates of *P. ipomoeae*. We concatenated these 10 alignments to create a super-alignment 7,084bp in length and consisting of 34 taxa, which was used to construct a simple neighbor-joining tree using Jukes and Cantor genetic distances[52] (Supplementary Figure S6). The tree was manually rooted with the three isolates of *P. ipomoeae*. Although phylogenetic inference is limited with the dataset, it does support that Pi1845A and Pi1889 are closely related, as are BL2009P4 (*US-23*) and 06_3928A (*13_A2*).

Identification of absent genes

To identify reference genes absent from the genomes of each sample, we used a conservative method published previously[12,53]. After removal of PCR duplicates, all mapped reads were used in the analysis, including those that mapped to multiple genomic locations. A gene was considered absent for a sample if zero reads mapped to any location along the entire length of the gene (Supplementary Figure S7, Supplementary Table S3). PCR confirmation of absent genes in historical sample libraries was not possible due to their very short insert sizes. To examine how the number of genes falsely identified as absent scales with the genome-wide mean depth of coverage in each sample, we used the software Picard version 1.66's DownsampleSam function to probabilistically remove reads mapped to the reference genome to achieve a predefined genome-wide mean read depth. The number of genes with no read coverage decreases as genome average depth of coverage increases (Supplementary Figure S8). With sufficient depth of coverage, the number of absent genes for each sample should saturate at a characteristic value, the actual number of absent reference genes. Although none of the samples have reached saturation at their maximum depth of coverage, they all appear to be approaching a plateau at which few additional genes would be identified as absence false positives. This is particularly true for sample Pi1889. However, cautiously we propose that readers view the genes identified in this analysis as “candidates” for absence until further sequencing on new suitable samples can be performed (Supplementary Figure S10).

The number of uniquely mappable bases (identified as above) was calculated for each gene absent in samples Pi1845A and Pi1889. The median gene absent in sample Pi1889 was 521bp in length and 95% uniquely mappable. The median gene absent in sample Pi1845A was 482bp in length and 94% uniquely mappable.

P-values for each absent gene from each sample were calculated as the probability p of observing no reads mapping to the gene. Assuming independence and uniformity, the number of reads mapping n is a Poisson-distributed random variable. The probability that a read maps to a gene G of length g in a genome of length L = 190Mbp is $(g + 2r)/L$, where we conservatively include a flanking region of maximum read length r on either side of the gene. The expected number of reads matching is $\lambda = pN$, where N for each sample is the total number of reads after removing PCR duplicates. The probability P_G of observing exactly zero covered bases of gene G with a sample size of N total reads is then given by the Poisson distribution

$$P_G(n = 0;\ \lambda) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda}.$$

For calculation of the p-value, we multiply this by the number of genes in the genome (the Bonferroni correction for multiple testing), which is also a conservative correction. Our p-value threshold to support a gene identified as absent in a particular sample was $p < 0.0001$.

This statistical approximation does not account for bias introduced by the G/C content of individual genes, which is known to affect read depth of coverage in regions of extreme G/C-bias[54,55]. However, an analysis of G/C content of all reference genes showed that the genes absent from the Pi1889 genome were not biased toward extreme values of G/C content (Supplementary Figure S4; Supplementary Table S3). Among Pi1889’s absent genes, 34 were within one standard deviation (0.0376) of the genome-wide mean G/C content (0.5281), while seven of Pi1889’s absent genes were not within two standard deviations of the mean. Finally, 25 of Pi1845A’s absent genes were within one standard deviation of the genome-wide mean G/C content, while six were not within two standard deviations of the mean.

Confirmation of gene absence by de novo assembly of unmapped reads

We used a *de novo* assembly approach to provide more confidence that the absent genes identified were indeed absent from the DNA libraries we generated from historical samples. We used the software SAMtools version 0.1.18[56] to extract unmapped reads from BAM files after sequentially mapping reads. The reads for each sequenced sample/isolate were assembled using the software Velvet version 1.2.07[57] with a *k*-mer length of 53bp and a minimum contig length of 200bp[12]. In order to minimize memory usage while still utilizing all unmapped reads for each sample, sequence reads were assembled in subsets of 150-175M reads. Statistics of these assemblies are reported in Supplementary Table S4. Unmapped reads from sample Pi1889 produced much larger total assembly lengths, presumably because they contained a larger fraction of bacterial and potato sequences (Supplementary Figure S5).

We used the software BLAT version 34x12[58] to align assembled contig sequences of unmapped reads from samples Pi1845A, Pi1889, RS2009P1 (*US-8*), IN2009T1 (*US-22*), BL2009P4 (*US-23*) to gene sequences in the T30-4 reference genome annotation using a search cutoff of at least 90% sequence identity. This search yielded no alignments to genes identified as absent in those samples, confirming that reads for these absent genes could not be retrieved from the unmapped reads.

Assembly of novel RXLR effector protein sequences

The *de novo* assemblies of unmapped reads were mined to discover potential non-reference genes encoding novel RXLR proteins in the resequenced sample genomes. We used custom scripts to translate every assembled contig nucleotide sequence to amino acid sequence for the six possible reading frames, then searched these amino acid sequences for all open reading frames (ORFs) at least 70 residues in length, following Cooke *et al*. (2012)[12]. These sequences were analyzed with the software SignalP version 4.1[59], which identifies amino acid sequences with a signal peptide that is characteristic of secreted effector proteins. All amino acid sequences with a SignalP discrimination score (D-score) greater than or equal to 0.9 and a signal peptide within residues 1-30 were searched for the RXLR motif within residues 25-60, where X stands for any amino acid. These criteria produced a varied number of candidate peptide sequences for each sample. The peptide sequences were used in a web-based protein BLAST search against the NCBI non-redundant protein sequence database, removing any non-*Phytophthora* hits with E-value greater than $1e^{-10}$ or any perfect matches to strain T30-4 protein sequences. Finally, the assembled contigs containing the remaining putative ORFs were used in a nucleotide BLAST search against the NCBI non-redundant protein sequence database, removing any non-*Phytophthora* hits with E-value greater than $1e^{-10}$. The remaining protein sequences containing the RXLR motif were considered putative novel genes to be confirmed by PCR assay (Supplementary Table S5).

Five novel RXLR gene sequences were discovered in the modern samples RS2009P1 (*US-8*), IN2009T1 (*US-22*) and BL2009P4 (*US-23*). All but one (PiRXLRd) of these novel amino acid sequences contained an RXLR-EER motif[60]. Two of the loci (PiRXLRa and PiRXLRe) differed by only a single non-synonymous nucleotide substitution in the RXLR motif and were approximately 80% identical to strain T30-4 reference gene PITG_22990. The novel gene PiRXLRc was mined from the US-8 assembly and shared remarkable similarity to the gene *Pex644* previously found[12] in the *P. infestans* isolate 06_3928Ab (*13_A2*), differing by the deletion of a single Z residue at the C-terminus.

No novel RXLR genes could be identified in the historical samples, despite that their assembly N_{50} values were within the range of those of the modern samples (Supplementary Table S4). It may be that the assembly process for these samples was hindered by both the shorter inserts (historical median 67bp, modern median 278bp) and the presence of non-*P. infestans* metagenomic contamination in the historical libraries[61,62] (Supplementary Figure S5).

PCR validation of novel RXLR genes

The presence of putative assembled RXLR genes identified in the previous section was confirmed for RS2009P1 (*US-8*), IN2009T1 (*US-22*), BL2009P4 (*US-23*) samples through PCR amplification assays using genomic DNA. Primer sequences were designed for the five putative genes using the software Primer3[63] within Geneious Pro version 5.5. The primer sequences used were: PiRXLRa_F: ACT GCC AGT TCA CAA TTC GCC A, PiRXLRa_R: AGG TCA TTC CTC CCA AGA ATT CGG T, PiRXLRb_F: CCT CCG GAC TGG GGC CTT CT, PiRXLRb_R: CGT GCT CGT GGT GGC TCT CC, PiRXLRc_F: TCG CAA CAT CCG GCC CGA AA, PiRXLRc_R: GCA TCT CCG GGG ACC CTT AGC, PiRXLRd_R: CGT CAG AGT CTT TCC TGT CAC TCC A, PiRXLRd_F: TGA CGC TTG CAT GCC ACA GA. The PiRXLRa primers were also used to amplify PiRXLRe from BL2009P4 (*US-23*) as it likely targets the same locus. Presence and absence of the expected fragment sizes were confirmed by PCR and gel electrophoresis for all loci except PiRXLRc, which yielded a band in isolate BL2009P4 (*US-23*; Supplementary Table S6), although was detected in its full length only from the assembly of isolate RS2009P1 (*US-8*). At over 560bp, the length of gene PiRXLRc is the longest of the five novel loci and is well beyond the N_{50} value of the BL2009P4 (*US-23*) assembly (473bp). All PCR products were verified by Sanger sequencing to have the predicted sequence from the assembly analysis.

Analysis of gene relatedness amongst absent RXLR and CRN effectors

The software mcl[64,65] was used (with an inflation value of 1.5) to cluster annotated RXLR effectors into tribes using as input the results of a protein BLAST alignment (cutoff E-value of 0.00001) of the entire set of T30-4 RXLR gene amino acid sequences against itself[66]. The analysis was performed independently for annotated CRN effectors. Genes with no protein coding sequence in the T30-4 genome annotation were ignored. These analyses produced 84 clusters of at least two RXLR genes and 11 clusters of at least two CRN genes (Supplementary Figure 14).

Per-sample mean depth-of-coverage of the RXLR (Figure 2) and CRN (Supplementary Figure 13) genes and the relationships between those genes differentially absent (Supplementary Figure S9, Supplementary Figure S15) amongst the high-coverage resequenced genomes were represented using the software Circos version 0.63[67]. Comparative distributions of the mean depth-of-coverage for every gene in the genome and different effector classes are shown in Supplementary Figure S11.

Analysis of intergenic lengths of absent genes

To calculate the distance between every gene and its nearest neighbor, absolute genomic position for genes in the T30-4 genome annotation was mined from the genome annotation using custom scripts (Supplementary Figure S12a). 5'-distance to the nearest 5' neighbor gene was calculated as the difference between the position of the 5' end of the gene and the 5' nearest neighbor’s 3’ end. 3’-distance to the nearest 3’ neighbor gene was calculated as the difference between the position of the 3’ end of the gene and the 3’ nearest neighbor’s 5’ end. Genes closest to either edge of an assembled contig were not considered in the analysis. To simplify plotting intergenic distance data on a logarithmic scale, the distance between opposite-strand genes that overlapped was deemed to be 1bp. Genes absent from at least one of samples Pi1845A, Pi1889, RS2009P1 (*US-8*), IN2009T1 (*US-22*), BL2009P4 (*US-23*) or 06_3928A (*13_A2*) had a median intergenic distance of 7,494bp, on the order of that of RXLR effectors (9,488bp) and notably different to the median intergenic distance of all reference genes in the genome (884bp; Supplementary Figure S12b).

Non-synonymous substitution in avirulence gene sequences

For each reference gene with a putative avirulence annotation, an MSA of available genome sequences (Supplementary Table S1) was populated with diploid bases called from VCF files with the software GATK’s UnifiedGenotyper algorithm using a minimum variant confidence score of Q20 and minimum depth of one read. For each sample, gene coding region sequences were translated to an amino acid sequence using the software Geneious Pro version 5.5. Amino acid sequences of putative avirulence loci described in the main text and summarized in Table 1 are shown in Supplementary Table S7.

Supplementary References

43. Derrien, T. *et al*. Fast computation and applications of genome mappability. *PLoS ONE* **7**: e30377. doi:10.1371/journal.pone.0030377 (2012).

44. DePristo, M. *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491-498 (2011).

45. Stamatakis, S. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).

46. Yang, Z. Among-site rate variation and its impact on phylogenetic analyses. *Trends. Ecol. Evol.* **11**, 367-372 (1996).

47. Guindon, S. *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**, 307-321 (2010).

48. Guindon, S. & Gascuel, O. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* **52**, 696-704 (2003).

49. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772 (2012).

50. Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716-723 (1974).

51. Blair, J. E., Coffey, M. D. & Martin, F. N. Species tree estimation for the late blight pathogen, Phytophthora infestans, and close relatives. PLoS ONE **7**, e37003. doi:10.1371/journal.pone.0037003 (2012).

52. Jukes, T. H. & Cantor, C. R. Evolution of protein molecules. In: *Mammalian Protein Metabolism*, Munro, H. N. (ed), pp. 21-132 (Academic Press, New York, 1969).

53. Raffaele, S. *et al*. Genome evolution following host jumps in the Irish potato famine lineage. *Science* **330**, 1540-1543 (2010).

54. Bentley, D. R. *et al*. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59 (2008).

55. Yoon, S., Xuan, Z., Makarov, V., Ye, K. & Sebat, J. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Research* **19**, 1586-1592 (2009).

56. Li, H. *et al*. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).

57. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-829 (2008).

58. Kent, W. J. BLAT: the BLAST-like alignment tool. *Genome Research* **12**, 656-664 (2002).

59. Petersen, T. N. *et al*. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786 (2011).

60. Jiang, R. H. Y. *et al*. RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proc. Natl. Acad. Sci. USA* **105**, 4874-4879 (2008).

61. Mende, D. R., *et al*. Assessment of metagenomic assembly using simulated next generation sequencing data. *PLoS ONE* **7**, e31386 (2012).

62. Kunin, V. *et al*. A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews*, **72**: 557-578 (2008).

63. Rozen, S. & Skaletsky, H. J. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S. & Misener, S. (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, pp. 365-386 (Humana Press, Totowa, New Jersey, 2000).

64. van Dongen, S. *Graph clustering by flow simulation*. Ph.D. thesis, University of Utrecht (May, 2000).

65. van Dongen, S. *A cluster algorithm for graphs*. Technical Report INS-R0010, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam (May, 2000).

66. Enright, A. J., van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**, 1575-1584 (2002).

67. Krzywinski, M. *et al*. Circos: an information aesthetic for comparative genomics. *Genome Research* **19**, 1639-1645 (2009).

68. Drenth, A., Janssen, E. M. & Govers, F. Formation and survival of oospores of Phytophthora infestans under natural conditions. *Plant Pathology* **44**, 86-94 (1995).