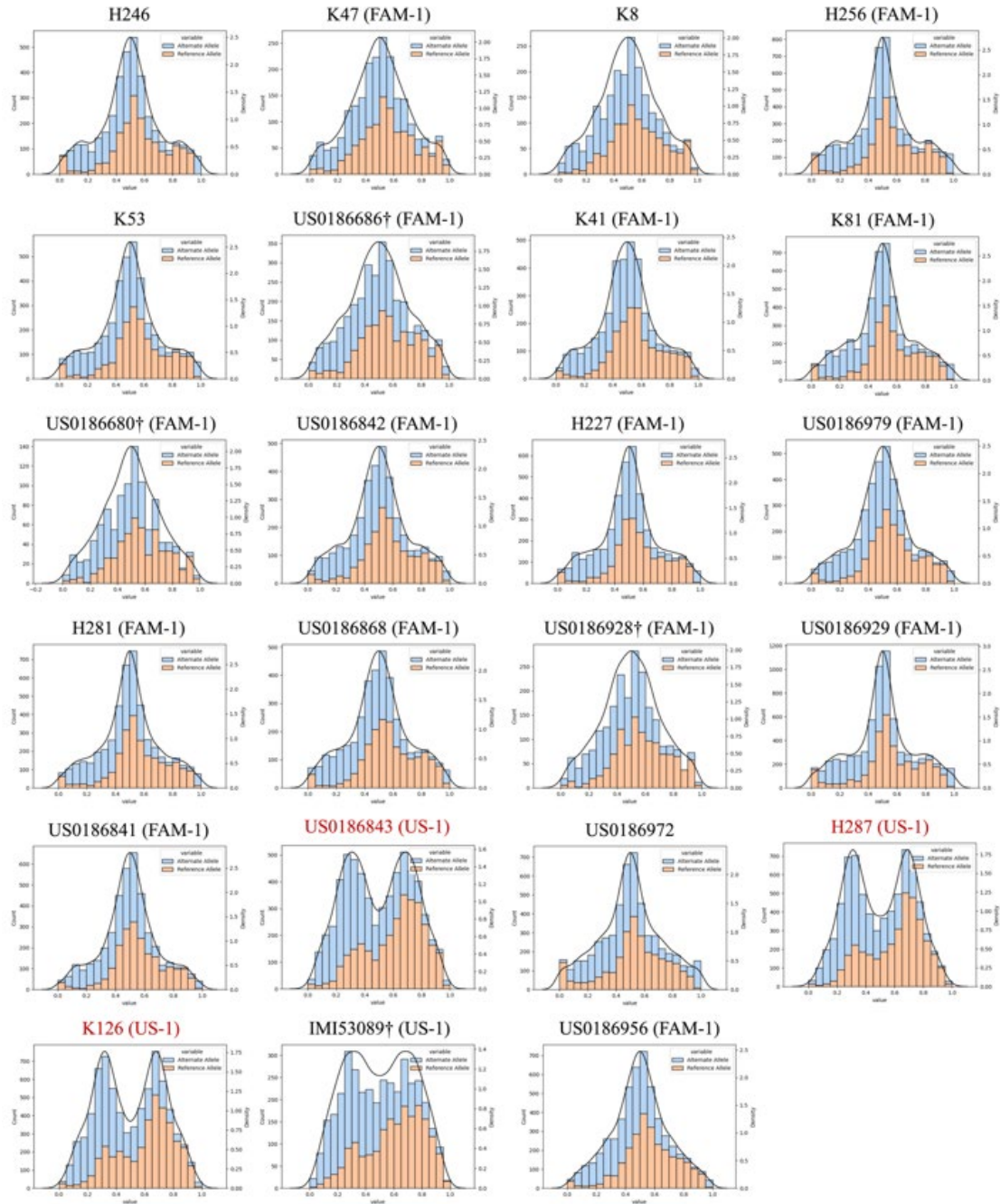


Supplementary Discussion 1. Additional *Solanum* Loci of Interest

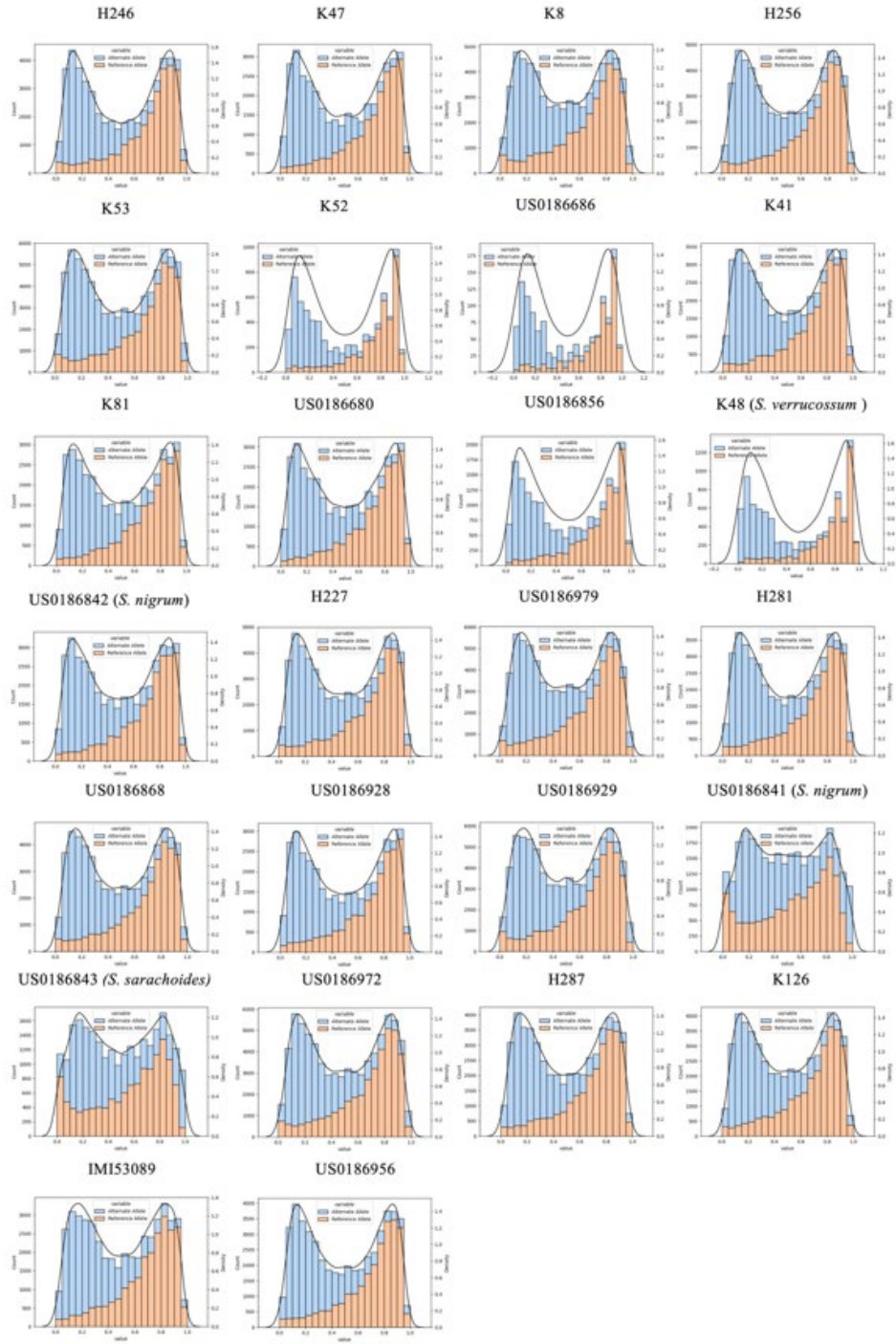
In addition to the R genes, two *Solanum* loci relevant to the evolution and domestication of *S. tuberosum* were investigated. *StCDF1*, a gene involved in day length tolerance in potato, exhibits alleles *StCDF1.1* and *StCDF1.2*, which determine tuber formation inhibition in long day conditions and tuber formation during long days, respectively. In this study, a 7 base pair insertion was observed in ten samples, including the oldest one, resulting in a premature stop codon and protein truncation similar to *StCDF1.2* (Supplementary Fig. 10). This indicated that some of the *Solanum* hosts in these historic samples were able to form tubers during long days. It is likely that these samples were considered "early maturing" varieties during the 19th century due to the earlier onset of tuber formation in long days. The diversity in this allele throughout our sampling period highlights the presence of both early and late potato varieties.

The source of the *Solanum* germplasm (either Andean or Chilean) can be partially predicted by examining mutations in the *trnV-UAC/ndhC* intergenic spacer region. This region contains a 241-base pair deletion associated with Chilean germplasm, specifically that of the "Rough Purple Chili"¹⁷. Most *Solanum* species in the herbarium samples we examined contained this deletion, indicating a Chilean origin (Supplementary Fig. 11). However, four samples, including two from wild hosts, displayed genotypes indicative of Andean origin. The original cultivated potato germplasm introduced to Europe was believed to be of Andean origin, with Chilean germplasm becoming predominant after the Famine^{17,18}. Our findings demonstrate that Chilean germplasm was already present during the Famine Period and persisted thereafter. The importation of "Rough Purple Chili" shortly after the Famine suggests the presence of other Chilean germplasm in the cultivated potato gene pool during the onset of the Famine⁶⁸.

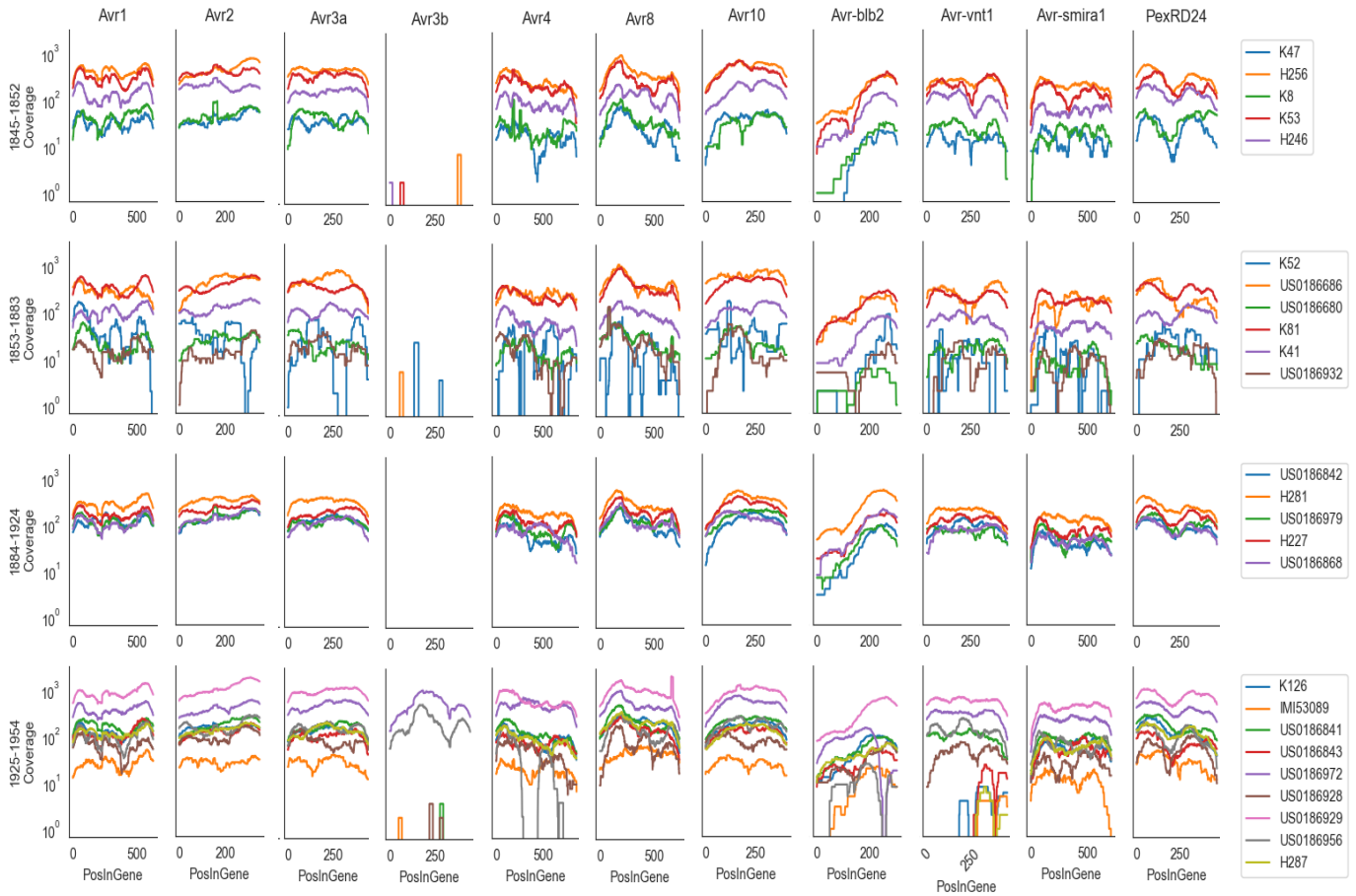
Supplementary Fig. 1. Histograms depicting the allelic fraction of both the reference and non-reference allele at biallelic SNP sites for high coverage *P. infestans* samples. In diploid organisms, biallelic SNPs should have an allelic frequency of about 50%. In triploid organisms, biallelic SNPs should have an allelic frequency of about 33% or about 66%. The samples highlighted in red are triploid. All FAM-1 genotypes were diploid while US-1 genotypes were triploid. Samples marked with a † are low confidence.



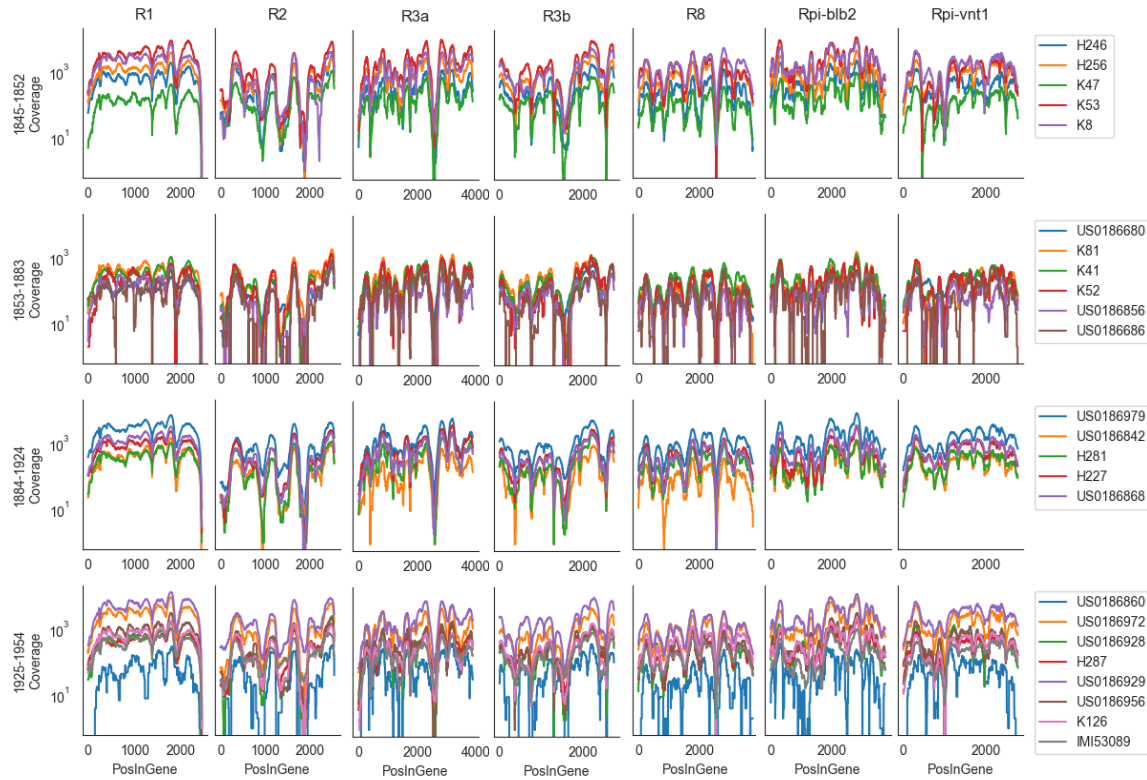
Supplementary Fig. 2. Histograms depict the allelic fraction of both the reference and non-reference allele at biallelic SNP sites for all *S. tuberosum* samples from herbarium specimens. In tetraploid organisms, biallelic SNPs should have an allelic frequency mode around 25%, 50%, and 75%. The ploidy histograms look visually like the expected distribution of tetraploids, but statistical confidence for diploidy, triploidy, and tetraploidy was low for all samples ($R^2 < 0.25$ compared to theoretical distribution graph for all plots).



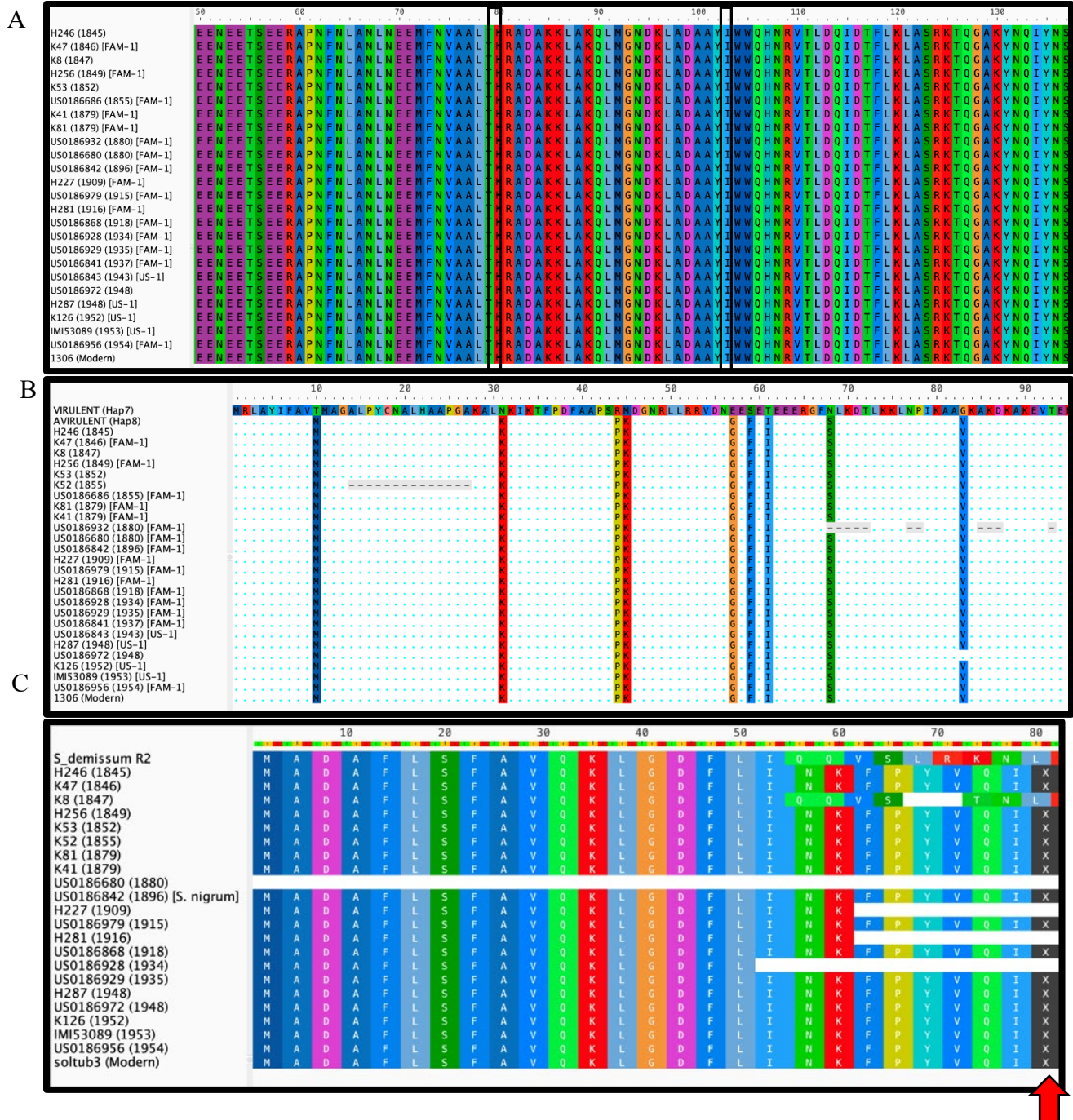
Supplementary Fig. 3. Depth of coverage at each position in 11 previously cloned *P. infestans* RXLR effectors. Only high coverage *P. infestans* samples are included in this analysis. Samples are shown by 4 time periods (1845-1854, 1853-1883, 1884-1924, 1925-1945) based on their date of collection of the herbarium sample.



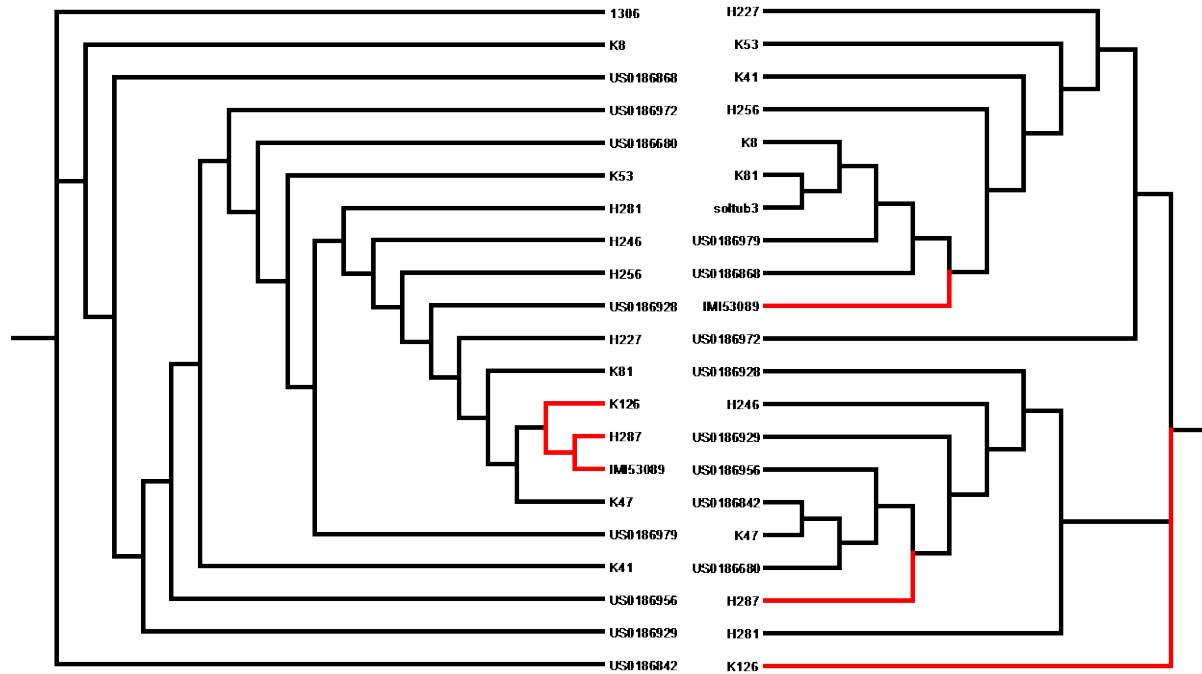
Supplementary Fig. 4 Depth of coverage at each position in 7 well-described *Solanum* spp. R genes. Only high coverage *Solanum* spp. samples are included in this analysis. Samples are shown by 4 time periods (1845-1854, 1853-1883, 1884-1924, 1925-1945) based on their date of collection of the herbarium sample.



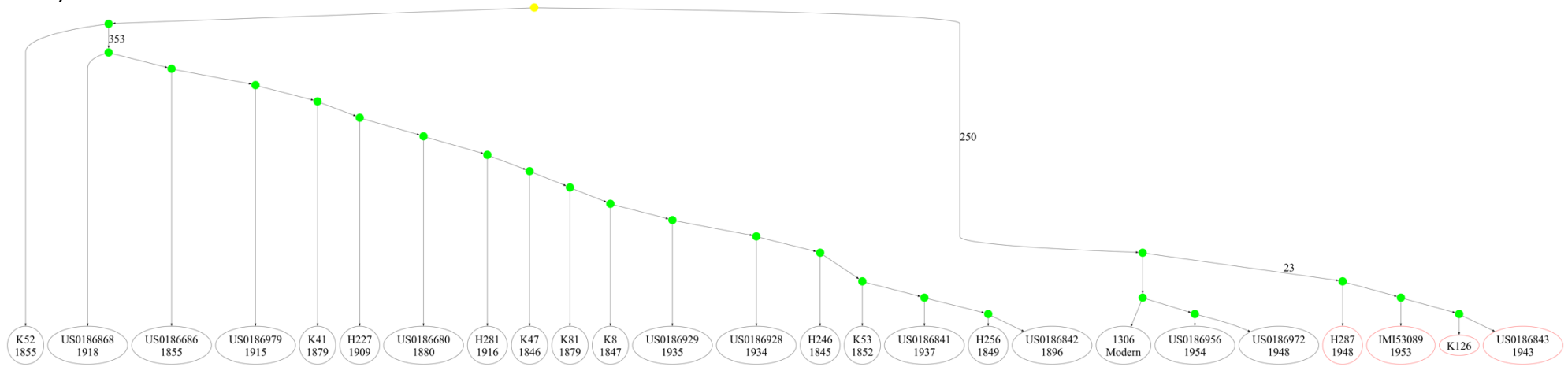
Supplementary Fig. 5. A) Excerpt of the *Avr3a* allele in all high coverage (>80%) samples. All samples are allelic variant *Avr3a*^{K80/1103} (boxed in black) which is the avirulent form of the effector and thus sensitive to the *R3a* resistance gene. *Avr3a* resistance breaking allele *Avr3a*^{E80/M103} was not found in any of the historic samples. **B)** Alignment of the entire *Avr2* gene for samples with high coverage (>80% of the gene is covered). All samples have the avirulent allele, with the exception of US0186972 which has one polymorphism which corresponds to the virulent allele. **C)** Excerpt of an amino acid alignment of the *R2* gene for samples with high coverage (>80%). Compared to the *S. demissum* *R2* allele which provides resistance to late blight, all samples have a premature stop codon (red arrow) except for K8.



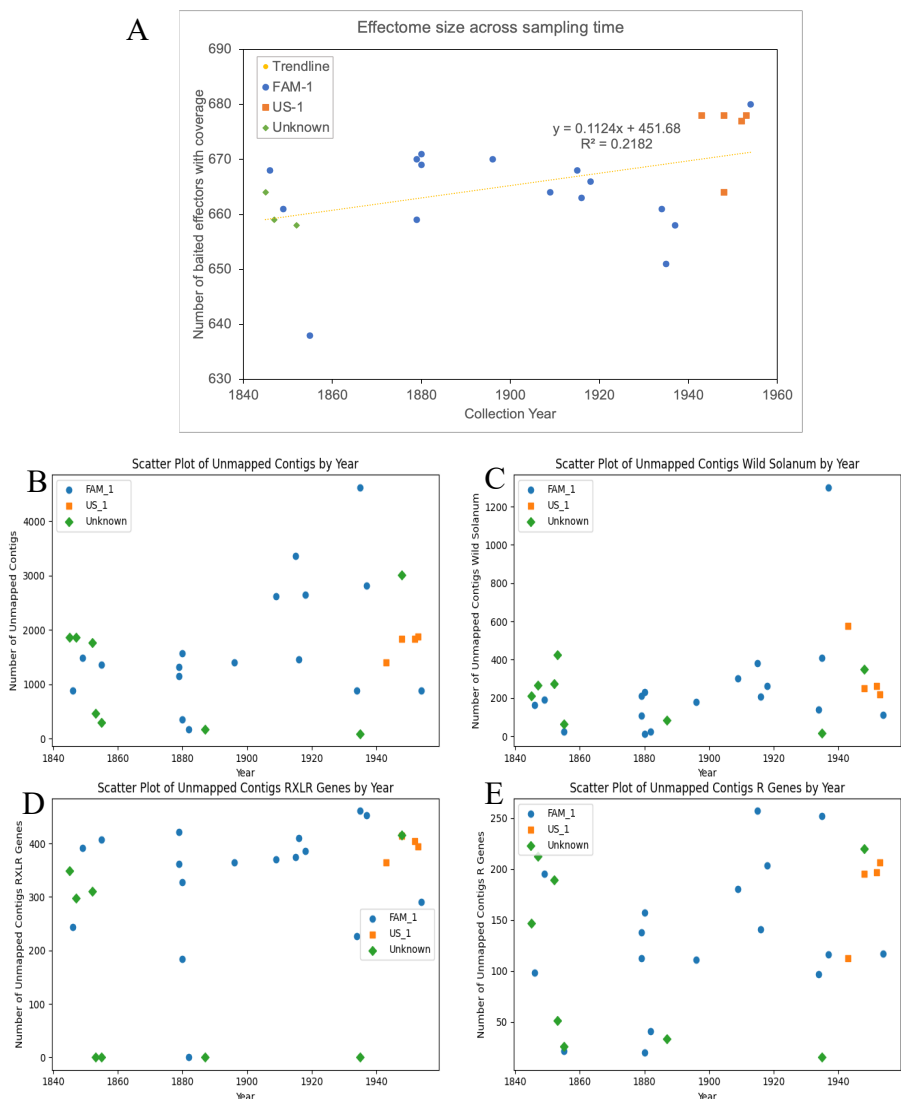
Supplementary Fig. 6. Co-phylogeny of *Avr8* (left) and *R8* (right). Samples of *Phytophthora infestans* which have US-1 genotype are highlighted in red. Samples with less than 80% of *Avr8* or *R8* covered are excluded from the phylogenies. The US-1 genotype samples are monophyletic in the *Avr8* RaxML gene tree, but they are in disparate clades in the *R8* RaxMLtree.



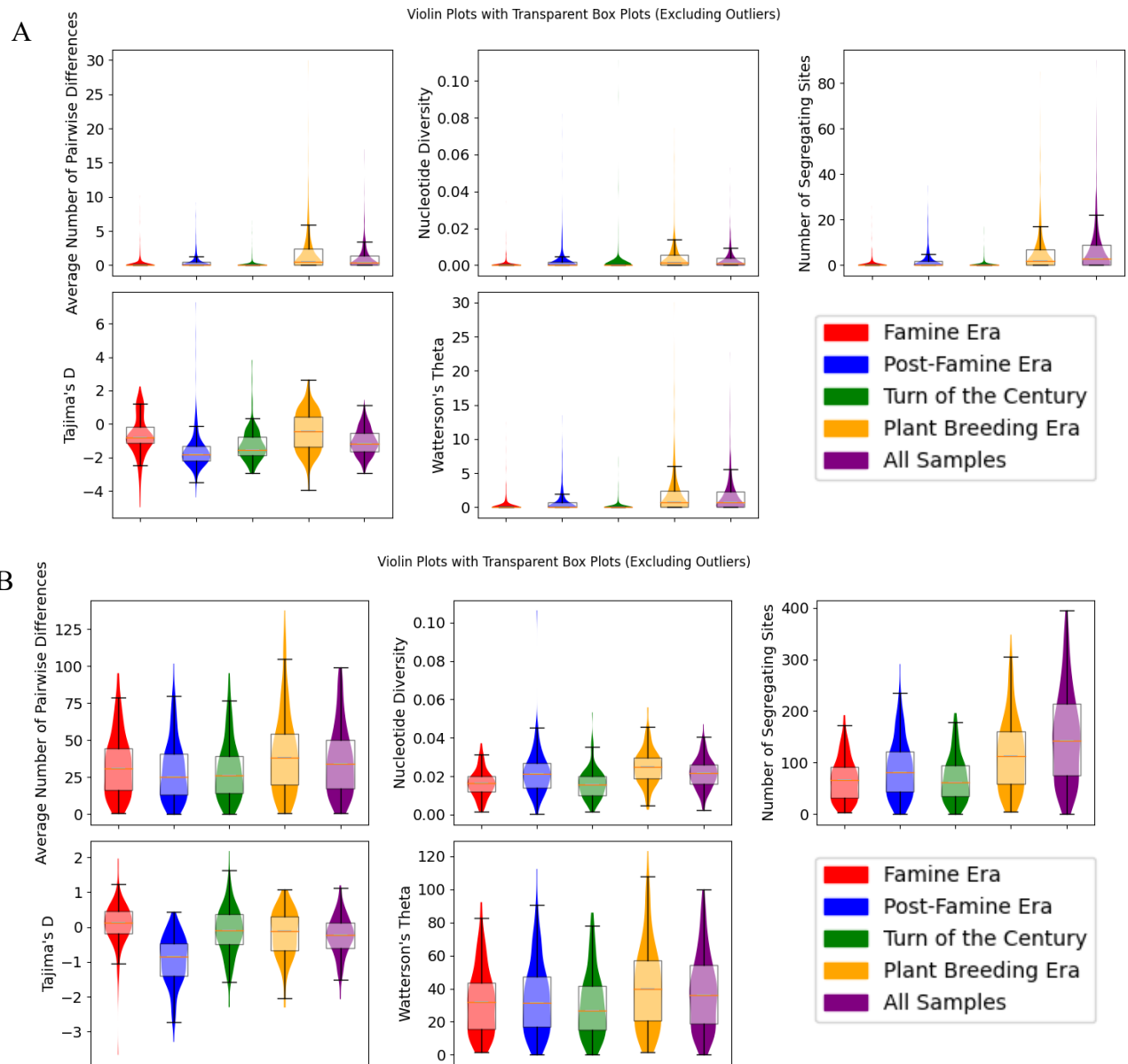
1 **Supplementary Fig. 7.** Ancestral recombination graph of *Avr10* for all samples which had high coverage (>80%) for this locus. The
2 yellow circle represents the final coalescence into the most recent common ancestor. Green circles are coalescent nodes. Leaf nodes
3 circled in red are herbarium samples identified with *Phytophthora infestans* genotype US-1. Numbers along edges represent the
4 position of mutation sites. No recombination events were observed within this gene in the sample set. All US-1 genotypes share a
5 mutation in *Avr10*. The modern reference genome *P. infestans* 1306 shares a clade with two of the more recent samples from the
6 herbarium sample dataset, US0186956 (1954 - Nicaragua) and US0186972 (1948 - Mexican).



Supplementary Fig. 8. A) Number of baited RXLR genes from the *Phytophthora infestans* 1306 reference genome with any coverage in the 25 highest coverage samples (>5X coverage depth), plotted against the year of sample collection. All alignments were adjusted to the same average coverage depth of the lowest coverage sample to control for coverage influencing effector count. **B-E)** Scatterplots showing number of unmapped contigs in various categories over the sampling period. Reads not mapping to either the *Phytophthora infestans* 1306 genome or the *Solanum tuberosum* SolTub3.0 genome were assembled into contigs. These contigs were then blasted to the NCBI nucleotide database. The number of these contigs that blasted most closely to wild *Solanum* species or other *Solanum tuberosum* cultivars, other described RXLR genes, or other described R genes were collated manually. The scatterplots show **B)** the total number of unmapped contigs, **C)** the number of unmapped contigs blasting to *Solanum*, **D)** the number of contigs blasting to RXLR genes, and **E)** the number of contigs blasting to R genes for each sample.

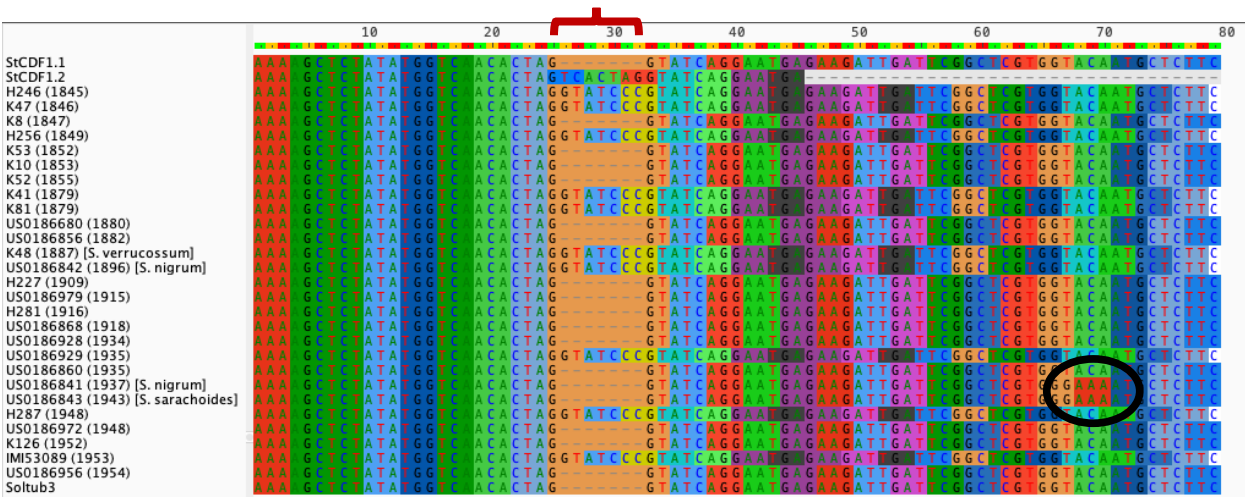


Supplementary Fig. 9. A) Violin plots of selection statistics for all RXLR genes sorted by time of sample collection. Each selection statistic was calculated separately for all RXLR genes for all samples within each era. These datapoints were then summarized into violin plots to show the density and distribution of each statistic for all RXLR genes during each period. The four eras are the Famine Era (1845-1852), Post-Famine Era (1853-1883), Turn of the Century (1884-1924), and the Plant Breeding Era (1925-1954). The final violin plots (purple) in each chart are all samples from all collection dates. **B)** Violin plots of nucleotide diversity statistics for all R genes broken out by time of sample collection. Each statistic was calculated separately for all R genes for all samples within each era. These data points were then summarized into violin plots to show the density and distribution of each statistic for all R genes during each period. The four time periods included the Famine Era (1845-1852), Post-Famine Era (1853-1883), Turn of the Century (1884-1924), and the intensive Plant Breeding Era (1925-1954). The final violin plots in each chart are all samples from all collection dates.

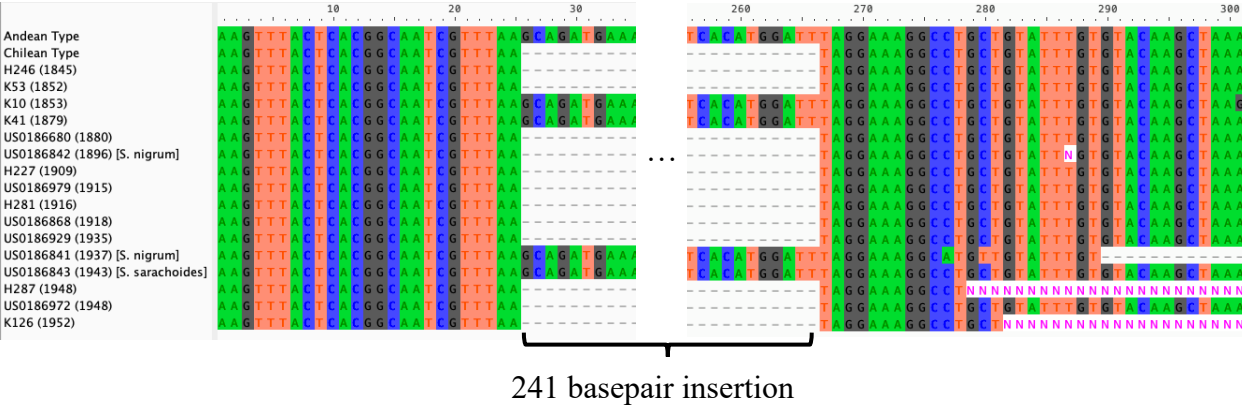


Supplementary Fig. 10. Eighty base pair regions excerpted from the *Solanum tuberosum* *StCDF1* gene for all samples which had at least 80% of the gene covered. Samples are listed chronologically by year of collection and the species is *Solanum tuberosum* unless otherwise specified. The first two rows are reference alleles of the *StCDF1* gene that influences the daylength tolerance of *S. tuberosum*. *StCDF1.1* is a gene in a pathway that limits tuber formation during long day conditions. Allele *StCDF1.2* has a 7 base pair insertion (bracketed in red) resulting in a frameshift mutation and premature stop codon, truncating the *StCDF1* protein. Allele *StCDF1.2* allows tuber formation during long day conditions. Samples US0186941 and US0186943, both from wild *Solanum* species, also have a point mutation in this region resulting in a single amino acid change (circled in black).

7 base pair insertion



Supplementary Fig. 11 Alignment of an excerpt of the *trnV-UAC/ndhC* intergenic spacer region in the *Solanum tuberosum* chloroplast. A 241 base pair deletion in this region has been used to distinguish *S. tuberosum* germplasm of Andean origin (no deletion) from that of Chilean origin (has 241 base pair deletion). Because this deletion is so large, contigs assembled from unmapped reads for each sample were also included in this alignment to ensure the insertion was not missed. The alignment includes all samples which have high coverage (>80%) for this region. All host species are *S. tuberosum* unless otherwise noted. Two samples collected in Britain, K10 and K41, have unmapped contigs showing alignment to the Andean Type allele. Two samples collected from wild species, US0186941 and US0186943, have unmapped contigs showing alignment to the Andean Type allele. The rest of the samples show the Chilean Type allele.



112 **Supplementary Table 1.** Summary of genome coverage of historic *Phytophthora infestans*
113 samples mapped to A) the *P. infestans* 1306 reference genome or B) *Solanum tuberosum*
114 SolTub3.0 genome coverage.

A) Summary of <i>P. infestans</i> 1306 reference genome genome coverage							
Sample ^a	Year	Mapped Reads ^b	% Reads Mapped to Reference Genome	Coverage Depth ^c	Reads Mapped to RXLR genes	% of reads mapped to baited effectors	Coverage Depth of RXLR genes
H246	1845	2304731	21.64%	0.51	547340	23.75%	93.63
K47	1846	470005	16.51%	0.10	130214	27.70%	23.60
K8	1847	926149	8.74%	0.10	147600	15.94%	21.71
H256	1849	2666018	31.15%	0.66	1422853	53.37%	219.37
K53	1852	3446833	17.27%	0.49	1203325	34.91%	108.54
K10	1853	428678	11.70%	0.03	41883	9.77%	1.14
K52	1855	1127342	9.89%	0.07	192777	17.10%	6.06
US0186 686	1855	3339394	26.28%	0.41	1180366	35.35%	29.48
K41	1879	926687	24.81%	0.25	355580	38.37%	69.40
K81	1879	2217574	36.96%	0.56	1240130	55.92%	182.31
US0186 680	1880	401247	12.06%	0.08	71619	17.85%	13.55
US0186 932	1880	336880	10.87%	0.05	74560	22.13%	9.32
US0186 856	1882	115924	11.25%	0.01	17437	15.04%	1.94
K48	1887	312349	9.78%	0.02	13134	4.20%	0.49
US0186 842	1896	1302663	27.99%	0.34	371627	28.53%	64.18
H227	1909	2283476	22.86%	0.59	603836	26.44%	109.86
US0186 979	1915	1508232	12.55%	0.28	387320	25.68%	61.41
H281	1916	1851772	43.32%	0.59	871206	47.05%	174.41
US0186 868	1918	1872809	14.67%	0.39	372192	19.87%	65.44
US0186 928	1934	677655	15.53%	0.10	239941	35.41%	24.26
US0186 860	1935	731730	10.51%	0.04	33761	4.61%	0.93
US0186 929	1935	6049510	21.24%	1.23	2802773	46.33%	336.39
US0186 841	1937	1311425	19.47%	0.35	510550	38.93%	79.40
US0186 843	1943	1074350	14.43%	0.21	388297	36.14%	53.36

H287	1948	987953	26.44%	0.31	382163	38.68%	80.97
US0186 972	1948	3815223	26.45%	0.94	1531080	40.13%	225.54
K126	1952	1152013	27.24%	0.37	448192	38.91%	98.09
IMI5308 9	1953	342132	12.03%	0.08	91026	26.61%	19.77
US0186 956	1954	1148379	25.78%	0.22	568188	49.48%	72.64
Average		1556177	19.64%	0.32	560033	30.15%	77.49

B) Summary of *Solanum tuberosum* SolTub3.0 genome coverage.

Sample ^a	Year	Mapped Reads ^b	% Reads Mapped to Reference Genome	Coverage Depth ^c	Reads Mapped to R genes	% of reads mapped to baited R genes	Coverage Depth of R genes
H246	1845	7873725	76.90%	0.88	1291929	16.41%	342.812
K47	1846	9569970	85.43%	0.94	1245000	13.01%	297.059
K8	1847	5904567	67.76%	0.54	1987543	33.66%	425.797
H256	1849	2921450	67.24%	0.30	700122	23.96%	185.266
K53	1852	2864288	75.43%	0.35	689799	24.08%	203.771
K10	1853	1504830	52.34%	0.21	400423	26.61%	130.85
K52	1855	2084460	56.49%	0.09	327094	15.69%	13.9856
US0186 686	1855	3585872	83.01%	0.45	835457	23.30%	252.803
K41	1879	2703771	70.75%	0.29	583483	21.58%	156.779
K81	1879	2026509	69.16%	0.22	422607	20.85%	114.058
US0186 680	1880	1148166	35.85%	0.07	194591	16.95%	23.2495
US0186 932	1880	2256807	19.75%	0.08	516092	22.87%	23.0606
US0186 856	1882	15751601	76.95%	1.06	4650042	29.52%	568.451
K48	1887	9993258	90.15%	1.00	3583104	35.86%	770.935
US0186 842	1896	2225919	36.99%	0.16	624667	28.06%	118.508
H227	1909	1754857	52.27%	0.21	316602	18.04%	90.8815
US0186 979	1915	2780101	21.85%	0.07	221643	7.97%	9.89955
H281	1916	5158220	74.97%	0.42	1143329	22.17%	247.435
US0186 868	1918	3730252	79.34%	0.38	529858	14.20%	128.479
US0186 928	1934	4103656	54.61%	0.28	759299	18.50%	145.841
US0186 860	1935	656596	63.21%	0.05	233089	35.50%	40.9613

US0186 929	1935	1163466	16.72%	0.04	144311	12.40%	7.98956
US0186 841	1937	8538437	65.95%	0.92	1380094	16.16%	349.322
US0186 843	1943	2631996	59.80%	0.19	785313	29.84%	112.648
H287	1948	16848002	58.31%	1.39	5179541	30.74%	872.36
US0186 972	1948	346132	11.17%	0.02	32853	9.49%	6.17145
K126	1952	3283564	72.42%	0.27	1198789	36.51%	220.803
IMI5308 9	1953	12084305	81.27%	1.15	3333436	27.58%	663.245
US0186 956	1954	10645373	86.60%	1.16	2927224	27.50%	658.621
Average		5039316	60.78%	0.46	1249563	22.73%	247.657

^a Samples from herbarium specimens are listed chronologically by their year of collection.

^b The number of reads obtained for each sample is listed along with the number of reads that mapped to any position in the reference genome SolTub3.0.

^c Average coverage depth across the entire genome based on the depth at each position calculated with SAMTools

Supplementary Table 2. Nucleotide diversity statistics for A) the known R genes and B) the known RXLR genes analyzed in this study.

A) Nucleotide diversity statistics for known R genes					
Locus	Average Number of Pairwise Differences ^a	Nucleotide Diversity ^b	Number of Segregating Sites ^c	Tajima's D ^d	Watterson's Theta ^e
<i>R1</i>	39.784	0.019	157	0.015	39.630
<i>R2</i>	33.998	0.018	154	-0.483	38.873
<i>R3a</i>	76.239	0.022	293	0.120	73.959
<i>R3b</i>	58.021	0.025	215	0.268	54.270
<i>R8</i>	48.216	0.017	234	-0.712	59.066
<i>Rpi-blb2</i>	69.446	0.022	267	0.118	67.396
<i>Rpi-vnt1</i>	40.538	0.018	137	0.663	34.582

B) Nucleotide diversity statistics for known RXLR genes					
Locus	Average Number of Pairwise Differences ^a	Nucleotide Diversity ^b	Number of Segregating Sites ^c	Tajima's D ^d	Watterson's Theta ^e
<i>avr1</i>	0.000	0.000	0	NA	0.000
<i>avr2</i>	0.057	0.000	1	-1.204	0.252
<i>avr3a</i>	0.524	0.002	3	-0.724	0.757
<i>avr3b</i>	0.126	0.000	0	NA	0.000
<i>avr4</i>	1.492	0.002	8	-0.793	2.019
<i>avr8</i>	3.122	0.006	19	-1.210	4.796
<i>avr10</i>	0.892	0.002	3	0.419	0.757
<i>avrblb2</i>	1.071	0.008	6	-0.835	1.515
<i>avrsmiral</i>	0.000	0.000	0	NA	0.000
<i>avrvnt1</i>	0.191	0.001	4	-2.082	1.010
<i>pexrd24</i>	4.931	0.012	22	-0.394	5.553

^a Average genetic differences between all pairs of DNA sequences in a population.

^b Average proportion of nucleotide differences between any two DNA sequences.

^c Count of positions in DNA where multiple nucleotides exist in a population.

^d Statistic that helps detect departures from neutral evolution in a population. It compares two estimators of genetic diversity: the average number of pairwise differences and the number of segregating sites. Deviations from neutrality, such as an excess of rare variants, can suggest factors like selection, population growth, or population structure.

^e Estimation of the population mutation rate per generation, considering the number of segregating sites and the sample size.