nature communications

Article

https://doi.org/10.1038/s41467-024-50749-4

Evolution of *Phytophthora infestans* on its potato host since the Irish potato famine

Received: 20 December 2023

Accepted: 18 July 2024

Published online: 05 August 2024

Check for updates

Allison Coomber[1,2], Amanda Saville[1] & Jean Beagle Ristaino[1,3]

Phytophthora infestans is a major oomycete plant pathogen, responsible for potato late blight, which led to the Irish Potato Famine from 1845–1852. Since then, potatoes resistant to this disease have been bred and deployed worldwide. Their resistance (R) genes recognize pathogen effectors responsible for virulence and then induce a plant response stopping disease progression. However, most deployed R genes are quickly overcome by the pathogen. We use targeted sequencing of effector and R genes on herbarium specimens to examine the joint evolution in both *P. infestans* and potato from 1845–1954. Currently relevant effectors are historically present in *P. infestans*, but with alternative alleles compared to modern reference genomes. The historic FAM-1 lineage has the virulent *Avr1* allele and the ability to break the *R1* resistance gene before breeders deployed it in potato. The FAM-1 lineage is diploid, but later, triploid US-1 lineages appear. We show that pathogen virulence genes and host resistance genes have undergone significant changes since the Famine, from both natural and artificial selection.

In 1843, a devastating plant disease struck US potatoes and two years later, in 1845, the pathogen spread to Europe, causing a destructive potato disease[1]. This epidemic (1845–1852) had profound consequences in Ireland, leading to the Irish Famine which resulted in the death of about 1 million people and the emigration of another 1 million people, as potatoes were a staple food[2]. The Irish Potato Famine left lasting effects as Ireland's population never fully recovered to pre-Famine levels and millions lost land that had been farmed for generations[2]. The responsible plant pathogen was identified in 1846 by M. J. Berkeley and subsequently renamed *Phytophthora infestans* in 1876 by Anton de Bary[2–4]. The pathogen is an Oomycete and can reproduce by clonal and sexual methods[5]. The clonal lineage of *P. infestans* that caused the Famine was named FAM-1 and was subsequently displaced by the US-1 lineage in the 1930s–1950s[6].

Phytophthora infestans continues to threaten potato and tomato production worldwide, necessitating the use of expensive fungicides for management[5,6]. Over the past 180 years, extensive efforts have focused on developing resistant cultivars of *Solanum tuberosum* to counter the disease[7]. Despite discovering several resistance (R) genes in wild Solanaceous relatives, achieving durable resistance to the pathogen has remained challenging[8,9]. Like other plant pathogens, *P. infestans* employs effector proteins to facilitate colonization of host plants[10,11]. Many of these effectors contain a specific amino acid motif known as the RXLR motif (Arginine – Any Amino Acid – Leucine – Arginine)[12]. In response, host plants have evolved R proteins that recognize these RXLR effector proteins, triggering an immune response that halts disease progression[10,11]. Effectors recognized by R genes are referred to as "avirulence factors," and this gene-for-gene response system accounts for the specific resistance of certain host genotypes against distinct pathogen strains[13].

All varieties planted in Ireland at the time of the Famine were highly susceptible to disease[2]. Shortly after the onset of the Irish Famine, efforts were made to select disease-resistant potatoes or import new varieties with resistance[14]. Although early efforts did not yield significant progress, breeding endeavors led to more resistant varieties, such as the Champion variety that survived the late blight outbreak in 1879[14–16]. Notable introductions included the "Rough Purple Chili" variety from Chile, which carries a distinctive genetic deletion useful for identifying Chilean sources of *Solanum* cultivars[17,18]. However, the pathogen continued to overcome deployed host resistance to the point where little effective resistance was observed by 1929[16].

[1]Department of Entomology and Plant Pathology, NC State University, Raleigh, NC, USA. [2]Functional Genomics Program, NC State University, Raleigh, NC, USA. [3]Emerging Plant Disease and Global Food Security Cluster, NC State University, Raleigh, NC, USA. e-mail: jean_ristaino@ncsu.edu

Initial explorations of close-wild relatives of potato from South America and Mexico by John Lindley and others were susceptible to late blight[15].Subsequently, identification of late blight resistant accessions in *S. demissum* from Mexico facilitated the breeding of cultivated potatoes in the early part of the 20th century[16,19,20].

Although 11 R genes from *S. demissum* were effective against various races of *P. infestans*, their introduction into domestic potato varieties proved challenging and resistance was short lived[21,22]. *Phytophthora infestans* was able to overcome the introduction of R genes from wild hosts in most instances[20,21]. Most R genes provide protection against specific races of *P. infestans*, and some races of the pathogen exhibit virulence on all 11 originally described R genes[23]. Despite recent identification of more than 20 additional R genes in the genomic era of the 21st century, achieving durable resistance to *P. infestans* has remained elusive[21,24]. Recently, the stacking of multiple R genes into transgenic potatoes from several wild species of potato has shown promise for late blight resistance in field trials conducted in Africa[25]. Genetic control, including R genes, is essential in managing *P. infestans*[26]. While genome sequencing of *P. infestans* has predicted over 500 putative effectors, only twelve avirulence effectors have been cloned and characterized (Table 1)[26–29]. Many effectors exhibit signs of rapid evolutionary adaptation, contributing to resistance breakdown in the potato late blight pathosystem[8,27].

Mycological herbarium specimens containing both host (potato) and pathogen (*P. infestans*) genetic material have been preserved since the time of the Irish Famine[1,30–32]. Sequencing of *P. infestans* from these specimens has demonstrated the presence of many currently known effectors, and the number of RXLR effectors appears to have increased over time[33,34]. However, information regarding the evolution of RXLR effectors of *P. infestans* in response to host R gene deployment over time has not been examined. Understanding the evolutionary pathways of specific known effectors could provide valuable insights into understanding the durability of deployed effectors and their associated R genes.

Both R gene identification in *S. tuberosum* and RXLR gene identification in *P. infestans* have been achieved through targeted enrichment sequencing[35–38]. This sequencing technique allows the number of reads of interest to be maximized by focusing on specific genes or genomic regions[36,38]. Targeted enrichment sequencing employs bait sequences, which consist of 60–80 base pair oligonucleotides closely aligned with the loci of interest (Fig. 1). These baits enrich the sequencing library for the genomic fragments of interest, which are then sequenced, resulting in a set of reads containing a disproportionate number of the targeted loci[36,38].

The goals of this study were fourfold: 1) to utilize targeted enrichment sequencing to simultaneously sequence both host R genes and pathogen effectors in *P. infestans* infected *Solanum* species from herbarium specimens from 1845–1954; 2) to analyze and document changes in avirulence factors of the pathogen in response to the deployment of resistance genes; 3) to investigate the temporal changes in the abundance of R genes and RXLR genes since the time of the Irish Famine; 4) to assess the evolution and diversity within R genes and RXLR effector genes over the sampling period.

Results

Genome sequencing and enrichment of targeted genomic regions

Baited enrichment sequencing was used to sequence DNA from 29 historic *P. infestans* infected *Solanum* specimens collected between 1845 and 1954 from herbaria (Table 2). The trimmed sequence reads were aligned to the whole genomes of both the *Solanum* host and *P. infestans*. Over the whole genomes, approximately 61% of reads aligned to the *Solanum tuberosum* SolTub3.0 genome, while 20% of reads aligned to the *P. infestans* 1306 genome, indicating a balanced distribution between the two genomes despite the predominance of host tissue in most samples (Supplementary Table 1B). Assembly and local alignment of unmapped reads revealed homology to other *Solanum* spp. or *Phytophthora* genomes, along with the presence of minor amounts of sequences from other microorganisms.

Across the entire *P. infestans* 1306 reference genome 521,699 SNPs were detected. The ratio of reference to alternate alleles at biallelic SNP sites was utilized to estimate the ploidy of *P. infestans* in each sample. Sixteen historic samples were classified as diploid and three (US0186943, H287, K126) were classified as triploid (Supplementary Fig. 1). All specimens genotyped as FAM-1 were diploid while US-1 genotypes were triploid. The ploidy assignment for the remaining ten samples was more challenging, with three appearing diploid with low confidence, one showing characteristics between triploid and diploid (IMI53089, US-1), and six remaining unresolved due to the limited number of high-quality biallelic SNPs.

Similarly, across the entire *S. tuberosum* SolTub3.0 reference genome, a total of 9,767,138 SNPs were detected. Although histograms

Table 1 | Summary of well-described R-gene and effector pairs analyzed in this study that are cloned or not cloned and the reference where they are reported

R gene[a]	*Solanum* spp.[b]	Status	Reference	Effector[c]	Status	Reference
R1	*S. demissum*	Cloned	75	*Avr1*	Cloned	76
R2	*S. demissum*	Cloned	44	*Avr2*	Cloned	77
R3a	*S. demissum*	Cloned	78	*Avr3a*	Cloned	79
R3b	*S. demissum*	Cloned	80	*Avr3b*	Cloned	81
R4	*S. demissum*	Not cloned	82	*Avr4*	Cloned	83
R8	*S. demissum*	Cloned	84	*Avr8*	Cloned	81,84
R10	*S. demissum*	Not cloned	85	*Avr10*	Cloned	36
Rpi-blb2	*S. bulbocastanum*	Cloned	86	*Avr-blb2*	Cloned	87
Rpi-smira1	*S. tuberosum cultivar 'Sárpo Mira'*	Not cloned	81	*Avr-smira1*	Cloned	81
Rpi-vnt1.1	*S. venturii*	Cloned	88	*Avr-vnt1*	Cloned	76,89
N/A			90	*PexRD24*	Cloned	90

[a]Eleven *S. demissum* resistance (R) genes designated R1-R11 are distinguished in a potato differential set by Black and Mastenbroek R1, R3, and R10, and to a lesser extent R2 and R4, have widely been used for introgression in European breeding programs[11,19]. Note R gene names are italicized.

[b]*Solanum* spp. host Latin names are italicized.

[c]Note that *Avr-blb1* was not successfully baited with the enrichment sequencing and it is not included in the data analysis. Effector gene names are italicized

https://doi.org/10.1038/s41467-024-50749-4

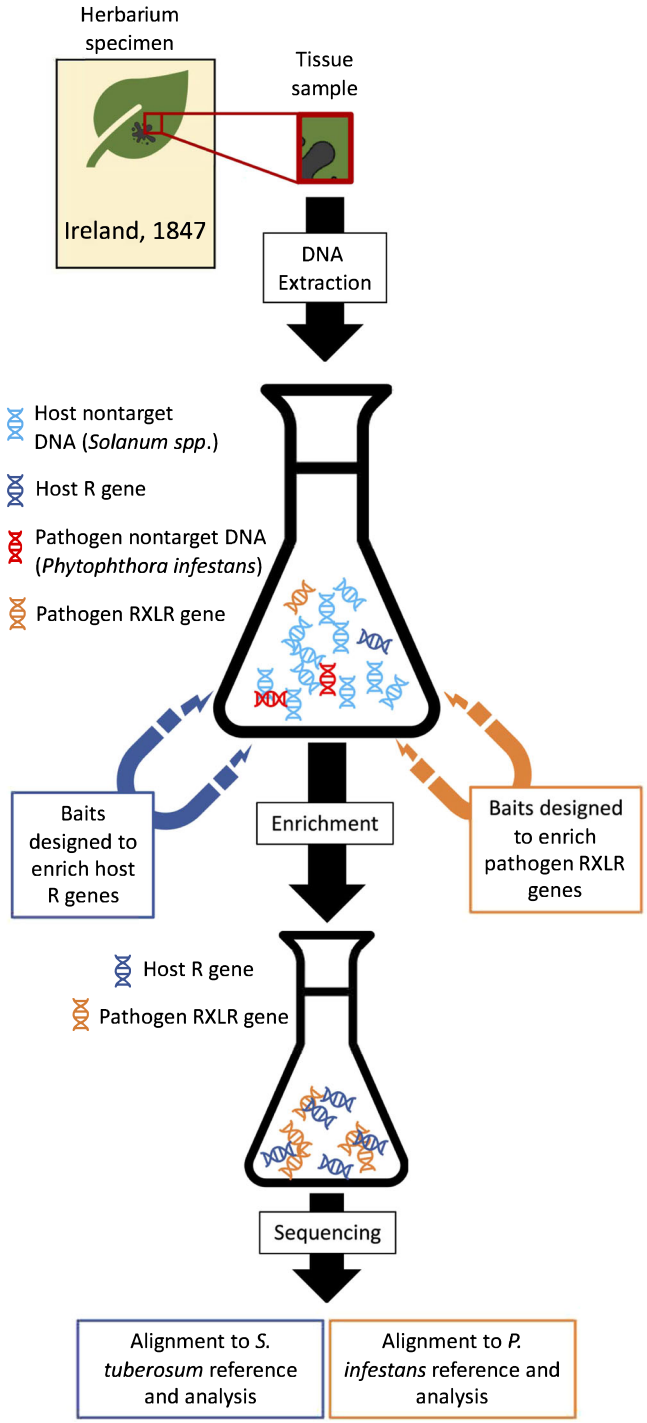

Fig. 1 | Graphic outline of sample processing. Each sample included in this study was a dried *Solanum* specimen with both host and pathogen DNA. Extracted DNA was enriched for both host R genes and pathogen RXLR genes before sequencing.

of allelic fractions at biallelic SNP sites suggested tetraploid characteristics as expected of cultivated potato, the similarity was not statistically significant (Supplementary Fig. 2).

The analysis of enrichment sequencing data revealed that, on average, 30.15% of sequenced bait reads mapping to the *P. infestans* 1306 reference genome corresponded to the targeted bait regions, resulting in an average coverage depth of 77X across the RXLR genome (Supplementary Table 1A). Within the targeted RXLR genome sequenced a total of 5459 single nucleotide polymorphisms (SNPs), 423 insertions, and 262 deletions were observed, amounting to 6144 variants across 388,942 base pairs. The ratio of nonsynonymous to synonymous variants in the RXLR genome was 1.74. For *Solanum* spp., 22.73% of reads mapping to the SolTub3.0 genome aligned to the targeted R gene bait regions, with an average coverage depth of 248X (Supplementary Table 1B). Within the targeted R genome sequenced a total of 33,454 SNPs, 796 insertions, and 658 deletions were observed, resulting in 34,908 variants across 301,133 base pairs. The ratio of nonsynonymous to synonymous variants in the R genome was 1.90. In addition to the R genes, alleles at the *StCDF1* gene and the plastid *trnV-UAC/ndhC* spacer indicate that a few historic potato samples were able to form tubers during long days while most samples have a Chilean origin (Supplemental Discussion 1).

Effectors and their associated R genes

Effectors and R genes were compared among four time periods (1845–1852, 1853–1883, 1884–1924 and 1925–1954). Note that *Avr-blb1* was not successfully baited so its presence could not be tracked in our data set of *P. infestans*. Among the 11 other known cloned effectors (Table 1), all except *Avr3b* were detected in most of the historic samples at all time periods (Supplementary Fig. 3). Interestingly, *Avr3b* was first found in herbarium samples collected in 1948 (US0186972) and 1954 (US0186956). Notably, sample US0186956 collected in 1954 from Nicaragua was a FAM-1 genotype, while sample US0186972 collected from Mexico did not match FAM-1 or US-1 and was likely a recombinant since both clonal and sexual reproduction were reported in Mexico at that time (Table 2)[34]. Our findings suggest a Central American/Mexican origin of *Avr3b*.

While coverage of the seven well-described R genes showed consistency across samples, the depth of coverage varied compared to the RXLR genes (Supplementary Fig. 4). This variability was particularly pronounced in the last period of sampling (1925–1954).

A lack of selection on the effector *Avr1* by *R1* is indicated by the fact that all samples, regardless of variation in the host *R1* gene within the same samples, exhibited identical sequences of *Avr1* (Fig. 2). All the samples of both FAM-1 and US-1 lineages had the AL virulent resistant breaking allele which is recognized by *R1* (Fig. 2A).

Investigation of allelic variation in *Avr3a* reveals the *Avr3a*^{KI} allele in all samples (Supplementary Fig. 5A). Therefore, the *P. infestans* identified in these herbarium samples would be sensitive to *R3a* mediated resistance and thus not able to overcome that host R gene. Alignment of the entire *Avr2* gene showed all samples had the avirulent allele, with the exception of the 1948 Mexican sample US0186972 which had one polymorphism which corresponds to the virulent allele (Supplementary Fig. 5B), and thus able to overcome *R2*. Alignment of historic *Solanum R2* genes compared to the *S. demissum R2* allele which provides resistance to late blight, showed all samples had a premature stop codon except for K8 (Supplementary Fig. 5C).

Phylogenetic analysis was performed for the regions containing the described effector gene *Avr8* and R gene *R8* and the phylogenies was compared (Supplementary Fig. 6). The *Avr8* alleles were differentiated based on mutations, and US-1 genotypes formed a distinct cluster. In contrast, alleles of *R8* did not cluster together according to *P. infestans* genotype.

An ancestral recombination graph of *Avr10* demonstrated that samples identified as the US-1 genotype formed a distinct cluster, and more recent samples (US0186956 and US0186972) shared a coalescent ancestor with the modern reference genome (1306) (Supplementary Fig. 7).

RXLR Genome Expansion

The RXLR genome of historic *P. infestans* displayed overall stability until the emergence of the US-1 lineage (in our sample set in1948), at which point genome expansion of numerous additional effectors was observed (Fig. 3). For FAM-1 lineages, an average of 664 covered effectors were found, compared to an average of 675 covered effectors

Table 2 | Chronological list of historical herbarium specimens used in this study, including their herbarium source, host, collector, date of collection, country, and the SSR genotype of *P. infestans* (when available)

Sample Name	Herbarium[a]	Host[b]	Collector	Date	Country (City)	SSr Genotype[c]
H246 FH-822447	FH	*S. tuberosum*	M. Desmazieres	1845	France	
K47 K-M178122	K	*S. tuberosum*	M.J. Berkeley	1846	Britain	FAM_1
K8 K-M177514	K	*S. tuberosum*	D. Moore	1847	Ireland	
H256 FH-822286	FH	*S. tuberosum*	M. J. Berkeley	1849	Ireland	FAM_1
K53 K-M 185583a	K	*S. tuberosum*	M. J. Berkeley	1852	Germany	
K10 K-M177506	K	*S. tuberosum*	M. J. Berkeley	1853	Britain	
US0186686	BPI	*S. tuberosum*	J. B. Ellis	1855	New York	FAM-1
K52 K-M 185583b	K	*S. tuberosum*	M. J. Berkeley	1855	Germany	
K41 K-M177512	K	*S. tuberosum*	M. J. Berkeley	1879	Britain	FAM-1
K81 K-M1851619	K	*S. tuberosum*	C. Spegazzini	1879	Italy	FAM-1
US0186680	BPI	*S. tuberosum*	W. Trelease	1880	WI, Madison	FAM-1
US0186932	BPI	*S. tuberosum*	F. L. Harvey	1880	ME, Orono	FAM-1
US0186856	BPI	*S. tuberosum*	G. Linhart	1882	Hungary	FAM-1
K48 K-M185632	K	*S. verrucossum*	P. Sydow	1887	Germany	
US0186842	BPI	*S. nigrum*	P. Sydow	1896	Germany, Berlin	FAM-1
H227 FH-822278	FH	*S. tuberosum*	F. Bucholtz	1909	USSR	FAM-1
US0186979	BPI	*S. tuberosum*	G.R. Lyman	1915	PA, Bath	FAM-1
H281 FH-822354	FH	*S. tuberosum*	R. Thaxter	1916	ME, Kittery Pt	FAM-1
US0186868	BPI	*S. tuberosum*	G. P. Clinton	1918	CT, Milford	FAM-1
US0186928	BPI	*S. tuberosum*	G.F. Gravatt	1934	AL, Wrangel	FAM-1
US0186860	BPI	*S. tuberosum*	A. D. McDonnell	1935	CT, Preston	
US0186929	BPI	*S. tuberosum*	K. Starcs	1935	Latvia	FAM-1
US0186841	BPI	*S. nigrum*	R. Sprague	1937	OR, Astoria	FAM-1
US0186843	BPI	*S. sarachoides*	M.W. Gardner	1943	CA, San Mateo Co.	US-1
US0186972	BPI	*S. tuberosum*	J. S. Niederhauser	1948	Mexico, Chihuahua	ND[c]
H287 FH-822365	FH	*S. tuberosum*	B. O. Saville	1948	Canada	US-1
K126 K-M185339	K	*S. tuberosum*	J.H.H.	1952	Britain	US-1
IMI53089	IMI	*S. tuberosum*	T. A. Russell	1953	Fr Cameroon	US-1
US0186956	BPI	*S. tuberosum*	S.C. Litzenberger	1954	Nicaragua, Santa Maria	FAM-1

[a]29 specimens sampled from the collections housed at: K, the Royal Botanic Gardens Kew Mycological Herbarium; BPI: the USDA National Fungus Collection, Beltsville, MD; IMI: CABI Bioscience, Egham UK; and FH, The Farlow Herbarium, Harvard University, Cambridge, MA. For FH and K, the sample number includes our internal specimen number (top) and herbarium specimen number.
[b]*Solanum* spp host Latin names are italicized.
[b]SSR genotype were done using 12 plex microsatellites as defined from Saville et al, 2021[40]. Empty cells indicate specimens were not genotyped previously by SSR genotyping. [c]ND signifies that the sample did not match any known genotype and is a recombinant.

for US-1 lineages. The coverage in the *Solanum* host R genome appeared to be more variable over time, with less apparent trends across different loci, a reflection of the different breeding efforts for *S. tuberosum* with time (Fig. 4). We assessed RXLR genome expansion by counting the number of RXLRs with coverage, normalized for overall coverage depth, and data indicated an expansion in the size of the RXLR genome over the period in which the herbarium specimens were collected (Supplementary Fig. 8A, $R^2 = 0.2182$).

Furthermore, the analysis revealed additional RXLR and R genes that did not align with the reference genome but were retrieved from the targeted baiting process. These unmapped reads were assembled into contigs and compared to the NCBI nucleotide database, revealing alignments to wild *Solanum* species and R/RXLR loci (Supplementary Fig. 8B–E). This suggests the presence of R and RXLR genes in historic samples that were not annotated in the modern reference genomes, with an increasing trend of gene alignments to wild *Solanum* species over time, coinciding with the introduction of wild *Solanum* species in breeding programs.

RXLR and R gene diversity

The well-described R genes exhibited varying numbers of segregating sites (137 to 293), and Tajima's D values ranged from −0.712 (*R8*) to 0.663 (*Rpi-vnt1*) (Supplementary Table 2). As for the well-described RXLR genes, the number of segregating sites ranged from 0 to 22, and Tajima's D values were negative or null for all loci except *Avr10* (0.419). Several RXLR genes displayed little diversity, despite some samples predating the putative introduction of R genes. Other RXLR genes showed negative Tajima's D values, indicating the potential influence of natural selection.

Nucleotide diversity statistics were also calculated for sample cohorts within each sampling period: Famine Period (1845–1852), Post Famine Period (1853–1883), Turn of the Century (1884–1924), and Plant Breeding Period (1925–1954), as well as overall (Supplementary Fig. 9). For RXLR genes, numbers of pairwise differences, nucleotide diversity and number of segregating sites were higher in the Plant Breeding Period than earlier time periods. Tajima's D was consistently negative for the RXLR genome throughout the entire sampling period (all four eras), indicating an abundance of rare polymorphisms in the dataset. This could potentially be due to an expansion in population size after a bottleneck, or positive selection on the RXLR genome. Tajima's D for the R genome was lowest during the Post Famine Period, indicating negative selection pressures during that time (Supplementary Fig. 9B).

https://doi.org/10.1038/s41467-024-50749-4

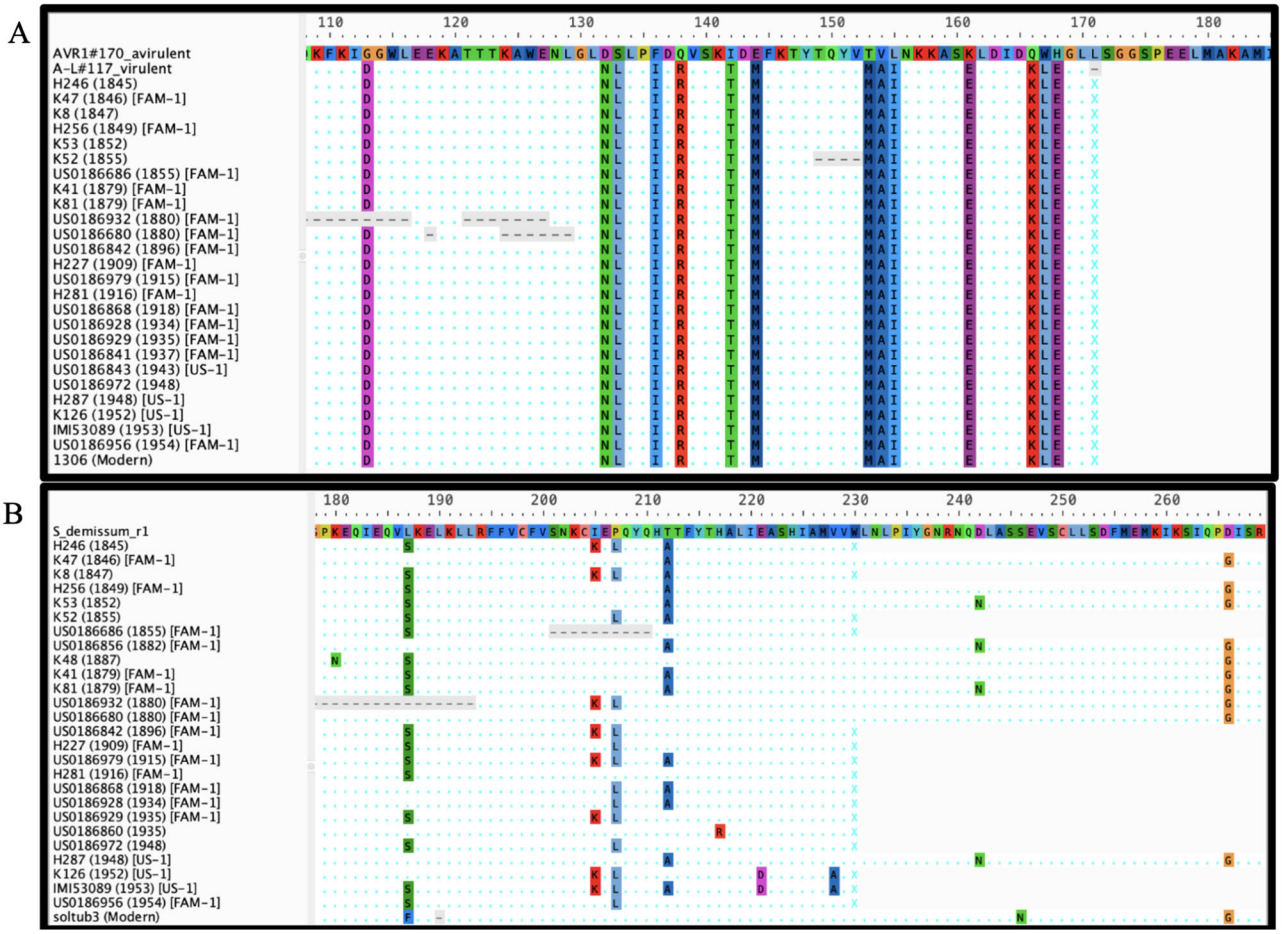

Fig. 2 | Amino acid alignment of *Avr1* and *R1*. A Amino acid alignment of a 100 amino acid excerpt of *Avr1* from P. infestans. The first line is the *Avr1* avirulent allele which is recognized by *R1*. The second line is the AL virulent, resistance-breaking allele which escapes detection by *R1*. The subsequent amino acid sequences are from high coverage samples analyzed in this study. All historic samples show a premature stop codon (indicated by x) leading to the virulent allele and are thus able to overcome *R1*. **B** Alignment of a 100 amino acid region excerpted from late blight resistance gene *R1* cloned from *Solanum demissum* and *R1* homologs in our samples and the reference genome. The functional *S. demissum* allele is shown on the top line, followed chronologically by the samples analyzed in this study. The final line is the *R1* homolog in the reference SolTub3.0 genome. Several samples have premature stop codons (indicated by x) signifying a truncation of the *R1* protein as compared to *S. demissum*.

Discussion

Simultaneous targeted sequencing of *Solanum* and *Phytophthora infestans*

Enrichment sequencing has been utilized for both *Phytophthora infestans* effectors and *Solanum* spp. R genes in previous studies[35–38]. However, this study combines enrichments of both the pathogen and host simultaneously in historic late blight specimens, allowing for the combined analysis of gene families in both species that evolved in response to each other.

One benefit of enrichment sequencing is that, by using baits to enrich the sequencing library for loci of interest, more sequencing resources can be allocated to the regions of the genome of interest. For example, RXLR genes constitute less than 1% of the *Phytophthora infestans* genome, yet, in this study, an average of 30.15% of all reads mapped to the reference genome of *P. infestans* to annotated RXLR genes (Supplementary Table 1)[36]. Ancient DNA samples are also challenging to work with due to their degraded and impure nature[39]. Ancient DNA is often highly fragmented, and in addition to the host and pathogen tissue in the sampled herbarium specimens, contaminants such as other microorganisms and bacteria are present[39]. Enrichment sequencing using small target baits improved our efficacy in obtaining the genes of interest. The samples used in this study were found to contain DNA from other microorganisms, but this DNA was present in lesser amounts compared to the targeted host and pathogen DNA and filtered out. In some samples, host leaf tissue is more prevalent than pathogen mycelial tissue. In a traditional sequencing experiment, this would result in many more reads aligning to the *S. tuberosum* reference rather than the *P. infestans* reference. However, the dual enrichment strategy employed here allowed us to capture targeted regions of both genomes successfully, even when host and pathogen were present in unequal amounts.

An increase in *P. infestans* ploidy since the Irish potato famine

Based on the allelic fractions at biallelic SNPs, the *P. infestans* genomes in the herbarium specimens in this study were a mixture of diploids and triploids. Older lineages belonging to the FAM-1 genotype were all diploid, while more recent mid-20th century US-1 lineages were triploid, consistent with previous published findings from more time-limited datasets[34,40,41]. For some lineages, determining ploidy was not straightforward, partly due to a lack of high-quality biallelic SNPs for certain samples. However, sample IMI53089 appeared to be between the triploid and diploid predicted states, which could be a result of aneuploidy (Supplementary Fig. 1). Recent work on chromosome-level assemblies of *P. infestans* indicates that aneuploids are very common[29]. More extensive sequencing coverage across all chromosomes could provide enough SNPs to determine ploidy at the chromosome level instead of the genome level. Additionally, the ploidy of *Solanum* spp. samples

https://doi.org/10.1038/s41467-024-50749-4

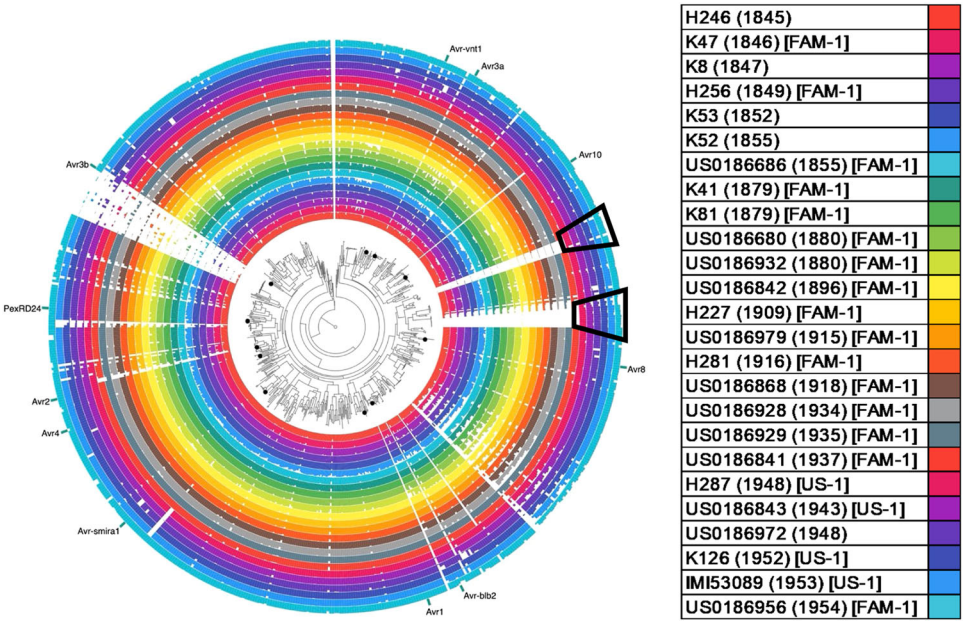

Fig. 3 | Coverage of all baited RXLR loci in the *Phytophthora infestans* 1306 reference genome. The center is a phylogeny showing the relationships between all of the baited RXLR loci. Cloned effectors are marked with a black circle on the phylogeny and a flag at the edge of the figure. The 25 samples with an average RXLR coverage depth greater than 5X are represented in concentric rings. The samples are presented chronologically, with the oldest herbarium sample (H246) in the innermost ring and the most recent specimen (US0186956) in the outermost ring. The major regions of effectome expansion during the spread of the US-1 lineage are boxed in black.

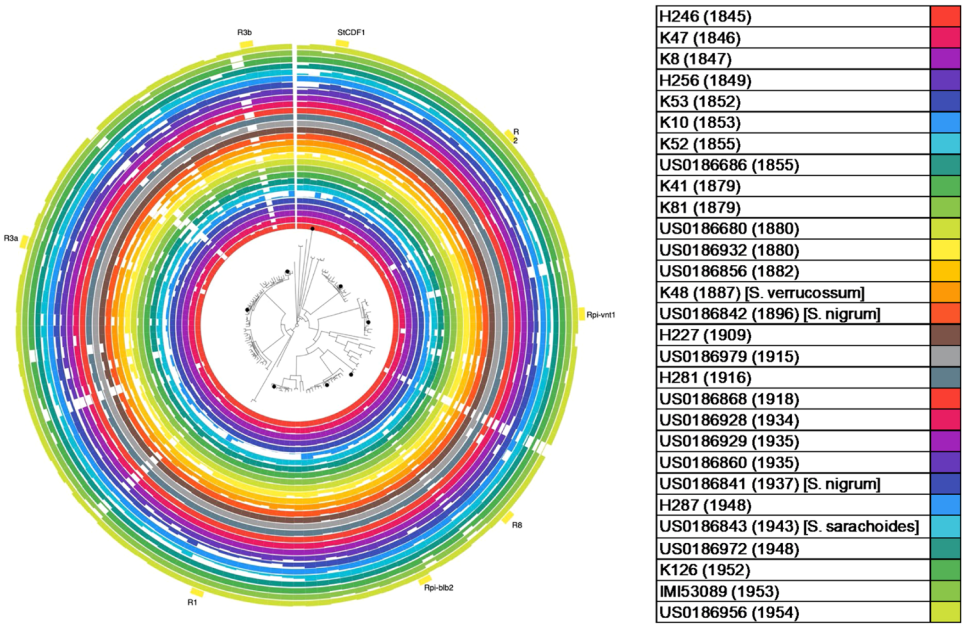

Fig. 4 | Coverage of all baited R loci in the *Solanum tuberosum* SolTub3.0 reference genome. The center is a phylogeny showing the relationships between all of the baited R loci. Cloned R genes are marked with a black circle on the phylogeny and a flag at the edge of the figure. Additionally, *StCDF1*, a locus related to potato day length toleration, is included and flagged. All 29 samples are represented in concentric rings. The samples are presented chronologically, with the oldest herbarium sample (H246) in the innermost ring and the most recent herbarium sample (US0186956) in the outermost ring.

could not be confidently assigned. Perhaps sequencing a broader number of these historic host genomes might alleviate this challenge.

Historic effectors and their associated R genes show significant variation

Consistent with previous studies, we found many RXLR genes to be historically present in all samples[33,34]. Homologs of several R genes were also identified. Out of 11 well-described RXLR genes, only *Avr3b* was absent from some samples (Supplementary Fig. 3). Samples from the final time (1925–1954) displayed more variation in effector presence or absence than other time periods, with partial absence of *Avr-blb2* in some samples (Supplementary Fig. 3). Seven well-described R genes were at least partially present in all samples. However, although there is coverage of these genes, implying the presence of some of the

gene structures in historic samples, their functionality may differ due to allelic variation. For example, all *Avr3a* alleles in this study contained the KI allele, indicating the avirulent form of the effector and sensitivity to *R3a* resistance (Supplementary Fig. 5)[42]. By comparing the *Avr2* region in all our samples to known virulent and avirulent alleles of *Avr2*, it is evident that all historic *P. infestans* in the samples tested were sensitive to *R2* resistance, with the exception of one more recent sample from 1948, US0186972, which appears to have a combination of both alleles[43]. This may represent an intermediate state between resistance breaking and sensitivity for this gene. It is notable that John Niederhauser, a World Food Prize recipient, collected this sample when he worked in Mexico in 1948. Intensive germplasm screening for development of late blight-resistant potato cultivars was underway there for many years, fueling evolution in the *P. infestans* effector genome. Similarly, some R genes exhibited incomplete or partial coverage and were predicted nonfunctional. For example, the resistance gene *R2* exhibited 154 segregating sites in samples we tested. The only sample that did not have a premature stop codon compared to the *S. demissum R2* allele was sample K8, collected in Ireland in 1847[44]. From this, we can infer that sample K8 may have had a functional *R2* protein, while all other samples were missing a significant part of the protein after translation, impacting functionality[45].

The most intriguing RXLR-R gene pair observed in this study was *Avr1* and *R1*, a resistance gene from *S. demissum* (Fig. 2)[19,46–48]. None of the *Solanum* spp. sampled in this experiment had a complete copy of the *R1* gene from *S. demissum*, with SNPs and deletions occurring throughout the gene (Fig. 2B). However, all *P. infestans* samples exhibited the resistance-breaking, virulent allele of *Avr1* (Fig. 2A) This implies that although *R1* resistance had not yet been deployed in cultivated potatoes during the Famine Period, the pathogen already possessed the ability to overcome the first resistance gene breeders would deploy later in the 20 th century. This could be attributed to the pathogen's prior exposure to this resistance gene in wild hosts. Although none of the samples had the complete *R1* resistance gene, they were also not clonally identical. We found high nucleotide diversity in the *Solanum tuberosum* loci we sequenced in this experiment, indicating that the *Solanum* germplasm grown during and shortly after the Famine was more diverse than previously reported[2,14,16]. The widespread prevalence of the virulent *Avr1* allele in our sample set demonstrates that even if *R1* resistance had been introduced earlier into cultivated potatoes, the pathogen would have been able to overcome such resistance quickly, at least in the case of *R1*. In fact, John Lindley reported in his species description paper of *S. demissum* in 1848 that *S. demissum* collected in Mexico at the time was susceptible to disease[15].

Natural and artificial selection in the RXLR and R genomes

Phylogenies of the well-described RXLR and R genes exhibit correlations with other characteristics. For example, a phylogeny of *Avr8* shows that *P. infestans* genotyped as US-1 grouped together, indicating the evolution of a shared allele of *Avr8* in the US-1 genotypes. In contrast, a phylogeny inferred from *R8* sequences did not exhibit the same pattern (Supplementary Fig. 6). This suggests that the evolved *Avr8* allele unique to US-1 was successful on different Solanum host *R8* alleles.

Ancestral recombination graphs (ARGs) of RXLR genes also provide evidence for the evolution of the RXLR genome. For example, the ARG of *Avr10* shows no recombination events, with all nodes in the tree representing coalescent events back to a common ancestor (Supplementary Fig. 7). However, the US-1 lineages clustered together in this ARG, indicating that the differences between *Avr10* in US-1 and FAM-1 were significant enough to cause divergence between these lineages. Many single SNPs were observed in *Avr10* in the FAM-1 lineage over time and no recombination was observed between FAM-1 and US-1. Notably, the more recent US-1 lineages also grouped with the reference genome.

Interestingly, there is low variation in the presence and absence of R gene content between samples (Fig. 4). However, changes in the presence and absence of RXLR genes become apparent in more recent samples, particularly those collected from 1937 and onwards (Fig. 3). This coincides with the time suggested by plant breeder Glendinning, Salaman and others of the emergence of more structured potato breeding programs in the United States and Europe, aimed at combating late blight and other potato diseases[14,19,20,22,49]. It is possible that *P. infestans* RXLR genomes were evolving to overcome newly introduced R genes during this period, which would explain the changes in RXLR gene coverage. The more recent US-1 genotype also emerged during this time; but both the US-1 and FAM-1 lineages have undergone shifts in RXLR genome content in the mid-twentieth century[33,50]. It is interesting that potato breeder Redcliffe Salaman noted in his breeding trials: "…in the autumn of 1932 our hopes were considerably dashed when certain of the immune seedlings, then growing in the open, showed signs of being attacked by Blight. In 1936 the attack was more serious, but it occurred a month later than that affecting the field crops in the neighborhood. It was thought probable that our 'immune' stocks were succumbing to some new form or biotype of *P. infestans*"[51]. Salaman was observing infection by the US-1 genotype, that appeared in the field during that time.

When comparing the number of covered effectors across the sample set while controlling for coverage depth, it becomes evident that the size of the RXLR genome has increased over time, consistent with previous findings (Supplementary Fig. 8)[33]. This expansion could be attributed to the spread of additional *P. infestans* lineages, as well as the ongoing evolution of the RXLR genome in response to plant breeding efforts. Analysis of unmapped contigs from each sample also reveals the presence of additional R and RXLR genes not found in their respective reference genomes. Further characterization and description of these additional R and RXLR loci could provide insights into the historic pan-genome of both R and RXLR genes.

Following the Irish Potato Famine, there was significant interest in identifying potato varieties resistant to late blight, either through selection or importation of new germlasm[3,20,22,49]. These early efforts were succeeded by the establishment of the US Potato Breeding Program in 1929, along with other initiatives aimed at incorporating late blight resistance from wild relatives into cultivated potato[20,22,49]. The artificial selection conducted by potato breeders to enhance the potato R genome is evident in these historic genomes, which exhibit strong signs of R gene selection (dN/dS > 1). Consequently, natural selection appears to have driven a high rate of nonsynonymous to synonymous substitutions in the *P. infestans* RXLR genome during this same period. This suggests that reports of *P. infestans* rapidly overcoming introduced resistance genes may be attributed, at least in part, to selection among effector genes. Homologs of several early-discovered R genes (*R1*, *R2*, and *R3b*) in our samples demonstrate high nucleotide diversity and many segregating sites, highlighting the allelic variety among the Solanum genome in our samples (Supplementary Table 2).

In this study, we have shown that both the RXLR genome and R genome have undergone significant changes since the Irish Famine, both from artificial and natural selection. Changes in the R genome have been reflected in the RXLR genome, such as the polymorphisms observed in *Avr2*. We could not do gene function experiments with historic genomes. Moving forward, it will be crucial to characterize the functionality and identify the causal mutations for avirulence or virulence in more modern-day *P. infestans* effector genes to better understand the historic RXLR genome. Given the limited characterization of only a few RXLR and R genes, it remains challenging to interpret the significance of historic mutations. Additionally, our sample set is constrained to what was collected historically and spans a

wide range of locations and collection years, limiting our understanding of changes in localized host and pathogen populations. However, exploring these samples has provided a snapshot into the host pathogens arms race that has characterized late blight resistance breeding for the last century and a half after the Irish Famine.

Methods

Sample selection and DNA extraction

Twenty-nine historic herbarium specimens were selected for enrichment sequencing (Table 2). Each specimen consisted of dried leaf tissue of a *Solanum* host species (mostly *S. tuberosum*) that was infected with *Phytophthora infestans*. Specimens were collected and deposited in various herbaria from 1845–1954 and were chosen for this study based on their original collection date and host.

DNA extraction was performed on each sample using a QIAGEN DNeasy Plant Mini Kit following the manufacturer's instructions (Qiagen, Valencia, CA). All DNA extractions were conducted in a dedicated laboratory solely used for ancient *P. infestans* DNA work, with no prior use for modern *P. infestans* samples. Storage of ancient samples and laboratory supplies for working with ancient samples were kept separate from modern *P. infestans* samples in an ancient DNA lab. Every precaution was taken to prevent cross-contamination among the ancient DNA samples, and there was no physical contact between specimens.

After completion of the DNA extractions, DNA quantity and quality were assessed using a QUBIT machine with the High Sensitivity Double-Stranded DNA Kit following the manufacturer's instructions (Invitrogen, Waltham, MA). In cases where the DNA concentration was too low for sequencing, extractions were repeated. A target of 1 μg of DNA was set as the baseline for all samples. For each historic herbarium specimen, at least two DNA extractions were combined to increase the DNA quantity.

Enrichment sequencing

Two bait libraries were designed based on the genes of interest (Daicel Arbor Biosciences). For the pathogen, *Phytophthora infestans*, a list of known RXLR genes was compiled from the T30-4 reference genome as well as a handful of other published RXLR genes[28]. From this list of 568 known RXLR reference genes, 8105 bait sequences, each approximately 80 nucleotides in length, were designed for RXLR target capture (https://github.com/allisoncoomber/targeted_sequencing_baits). For the host, *Solanum* spp., a list of 117 R genes for late blight resistance (and one for day length tolerance, StCDF1) was compiled from the SolTub3.0 genome and other published resources involving potato wild relatives. Using these 117 input R genes, 8647 baits, each 80 nucleotides in length, were generated (https://github.com/allisoncoomber/targeted_sequencing_baits). Both sets of baits were designed to match the target input genes (listed RXLRs and R genes) as well as any highly similar loci. The bait design focused on the regions of interest, excluding repeat-rich regions. Additionally, a Blast search was conducted against the relevant genomes to ensure that there were not a high number of off-target alignment sites.

All 29 samples were enriched with both bait libraries (Fig. 1). This allowed for the concentration of both *Phytophthora infestans* RXLR genes and *Solanum* spp. R genes in the sequencing library for each sample. Samples were sequenced on a NovaSeq with 150 bp paired end reads by Arbor Biosciences myReads (Ann Arbor, MI).

Read trimming and alignment

Raw reads received in FASTQ format were trimmed using TrimGalore, a wrapper around Cutadapt and FastQC[52–54]. TrimGalore was run in –paired mode, ensuring that both reads in a pair met all requirements. Adapters were removed along with reads less than 20 base pairs in length. A short 20 basepair minimum length was chosen because of the highly fragmented nature of the ancient DNA in these samples. A quality cutoff of 20 Phred score was used in trimming with the "–nextseq" flag within Cutadapt to prevent the G overcalling that can occur with the NovaSeq platform. FastQC results for the trimmed reads were collated with MultiQC and manually reviewed[55].

The trimmed reads were then aligned to both reference genomes. For the pathogen, *Phytophthora infestans* assembly Pinf1306_UCR_1 (PRJNA868814) was used as the reference genome[29]. For the host, *Solanum tuberosum* assembly SolTub_3.0 (PRJNA63145) was used as the reference genome[56]. The BWA-MEM algorithm from the Burrow-Wheelers Aligner was employed for aligning the reads to each respective genome[57]. From this point onward, the alignments were divided into two cohorts: one containing the pathogen alignments and the other containing the host alignments. Within each cohort, all 29 samples were kept separate. Subsequent steps were conducted in parallel for each cohort.

Variant calling

To appropriately group reads for downstream variant calling, read groups were added to each alignment using Picard AddReadGroups[58]. The read groups specified the sample, experiment, and library for each alignment. All alignments were then sorted using SAMTools[59]. Picard MarkDuplicates was then used to identify any PCR duplicates from library preparation[58]. Any duplicates that were identified as derived from amplification and not biological duplicates were ignored during variant calling.

The program mapDamage was utilized to rescale the quality scores of all alignments based on post-mortem DNA damage patterns[60]. This limits the DNA damage associated with ancient samples from influencing the alignments by reducing the quality score of the alignment where post-mortem damage is expected.

Haplotypes were called using GATK HaplotypeCaller in ERC mode to produce GVCF (Genomic Variant Call Format) files containing genotype likelihoods[61]. All GCVF haplotype call files were combined using GATK GenomicsDBI to build an indexed database. Joint genotyping for each cohort was then performed with GATK GenotypeGVCFs, resulting in one combined VCF file of genotypes for each cohort (pathogen and host-aligned reads, respectively).

Each resultant VCF file was split into separate files for INDELs and SNPs using GATK SelectVariants. INDEL and SNP variants for both cohorts were individually filtered using GATK VariantFiltration to retain only high-quality variants. Filtered INDEL and SNP VCF files were then combined, removing, or collapsing any overlapping INDELs. Finally, variants were annotated using custom-built databases for each reference with SNPEff[62].

Coverage analysis

To assess the success of the bait libraries in retrieving targeted DNA from these degraded samples, the SAMTools suite was used to evaluate coverage. SAMTools view, flagstat, depth, and bedcov were executed for each rescaled, deduplicated alignment file to determine the number of reads from the sequencing libraries that aligned to each reference genome. BED files listing the regions of the genomes containing the target genes were utilized in combination with SAMTools to determine the number of reads aligning to the baited loci.

The BEDTools suite was employed to generate files listing the coverage depth at each site from the *P. infestans* alignments for regions of all samples corresponding to well-described RXLR genes[63]. The same process was repeated with the *S. tuberosum* alignments, but this time focusing on R genes instead of RXLR genes. The depths at each position within these important genes were plotted using a custom Python script.

Additional BED files were generated corresponding to all known R genes in *S. tuberosum* and all known RXLR genes in *P. infestans* using

genome annotations and conducting BLAST searches for additional homologs. BEDTools was used to generate coverage depth files for all known R genes and RXLR genes, respectively, for each sample. A custom Python script was written to summarize these files and generate circular coverage diagrams based on the percent coverage of each R gene or RXLR gene.

Assembly and characterization of unmapped reads

We sought to investigate if there were RXLR genes or R genes which were present in these historic samples but not the modern reference genome. We expected reads from such genes would be enriched and sequenced by our experimental design, but not mapped to the reference genome. Therefore, reads that did not align to either the host or the pathogen reference genome were collected for each sample. These reads were assembled into contigs using SPAdes de novo assembler[64]. A Blast search was performed to compare these contigs to the NCBI nucleotide database using local Blast[65]. The Blast results were manually filtered to identify the closest match to each contig. Matches were summarized at the genus level based on descriptions. If the genus was *Phytophthora* or *Solanum*, the match was further examined to determine if the region of interest was a potential RXLR or R gene based on the description. For all contigs where the genus was a non-target organism, a taxonomic pie chart was created in Microsoft Excel. The number of unmapped contigs in each sample with specific characteristics (all unmapped, matching wild *Solanum*, matching RXLR genes, and matching R genes) were summarized into plots using a custom Python script.

Ploidy analysis

The *Phytophthora infestans* data from each sample was analyzed to determine the ploidy of the pathogen isolate infecting the leaf in each sample. GATK VariantsToTable was used to extract a table of all high-quality SNPs. A custom Python script was written to filter out high coverage, high quality, biallelic SNPs. Histograms of allelic fractions for each polymorphism at all high coverage biallelic SNP sites were then plotted, with fitted lines.

The program nQuire was also employed to estimate the ploidy of each *P. infestans* sample based on the frequency distribution at biallelic sites[66]. The nQuire models and ploidy estimates were compared with the histograms generated by the custom Python script to determine ploidy.

The same process was repeated for the *Solanum* spp. data for each sample to evaluate host ploidy.

Selection analysis

Alternate FASTAs were generated for all samples for all loci in the R genome, RXLR genome, and *Solanum* spp. gene *StCDF1*. The annotated, filtered variants previously called were used by BCFTools consensus to apply any changes to the reference genome for each sample. Low coverage masks were generated for each sample to replace any poorly covered regions with "N."

A total of 709 RXLR genes and 137 R genes (including *StCDF1*) were included in selection analysis. DendroPY was used to calculate average number of pairwise differences, nucleotide diversity, number of segregating sites, Tajima's D, and Watterson's Theta for each locus in both datasets[67]. A custom Python script was developed to separate the samples by collection date and plot the resulting selection statistics. SNPEff was also run on reduced VCF files containing only variants in the RXLR genome and R genome datasets for further analysis. Selection statistics for the well described R genes and RXLR genes were calculated and summarized across all time periods.

The selection tests used included: average number of pairwise differences, nucleotide diversity, Tajima's D, number of segregating sites, and Watterson's theta. Average number of pairwise differences calculates the average number of differences in nucleotide pairs between sequences in a population sample. It provides insight into the genetic diversity within a population, where a higher number indicates greater diversity. Nucleotide diversity is also a measure of genetic variation, measuring the average number of nucleotide differences per site between two DNA sequences randomly chosen from a population. Tajima's D is a statistical test used to detect departures from neutrality in DNA sequence data. Positive Tajima's D values may indicate balancing selection or population subdivision, while negative values may suggest purifying selection or population expansion. Number of segregating sites measures the number of variable sites in a population sample, where different alleles are present at a particular nucleotide position. Watterson's Theta estimates the population mutation rate based on the number of segregating sites. This provides an estimate of the long-term effective population size and mutation rate in a population.

Additionally, alternate FASTA files for specific loci of interest, such as well-described RXLR genes, R genes, and *StCDF1*, were visualized. These sequences were aligned against modern copies of these genes to investigate potential polymorphisms using Muscle[68]. Alignments were visualized in AliView[69].

RXLR genome size

To estimate the size of the RXLR genome for each sample, the differences in coverage were controlled. Samples with less than 10X coverage for the RXLR genome were excluded from the analysis. The remaining samples were down sampled to the same level of coverage as the lowest coverage sample (13.55X). The number of covered loci in the RXLR genome in each down-sampled sample was summed and plotted.

Phylogenetics

The alternate FASTA files were aligned and converted into PHYLIP format for each of the well-described RXLR effectors and R genes (Table 1). RAxML with 1000 bootstrap replicates was used to infer maximum likelihood trees from each of these alignments[70]. The resulting phylogenetic trees were visualized with Mesquite[71]. The alternate FASTA files of all genes in the R genome and RXLR genome were also generated, aligned, and concatenated into partitioned PHYLIP alignments. RAxML with 1000 bootstrap replicates was used to infer phylogenies from these alignments as well[70]. The phylogenetic trees were visualized and colored with Mesquite[71].

To compare the topologies of the well-described RXLR and R gene pairs, Visual TreeCmp was used to generate statistics comparing the two trees in each case[72].

Ancestral recombination graphs

From the generated alternate FASTA files, the homolog of RXLR gene *Avr10* was extracted from all high coverage samples as well as the reference genome. This collection of *Avr10* homologs was used as input for the program kwarg to generate ancestral recombination graphs of this locus[73]. The kwarg analysis used 1000 replicates and the minimum recombinant tree was visualized using Graphviz[74].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequence datasets generated in this study has been deposited in the NCBI database under the Sequence Read Archive project PRJNA1035512 https://www.ncbi.nlm.nih.gov/sra. The GitHub link contains sequences of the baits used. https://github.com/allisoncoomber/targeted_sequencing_baits. A source file of data is included here.

Code availability

The software used to analyze the data is cited in "Methods" section and reference lists and no custom code was written.

References

1. Ristaino, J. B. Tracking historic migrations of the Irish potato famine pathogen, *Phytophthora infestans*. *Microbes Infect.* **4**, 1369–1377 (2002).
2. Bourke, A. *'The Visitation Of God?' the Potato And The Great Irish Famine*. (Lilliput Press Ltd, 1993).
3. De Bary, A. Researches into the nature of the potato fungus, *Phytophthora infestans*. *J. Bot. Paris* **14**, 105–126 (1876).
4. Berkeley, M. J. Observations, botanical and physiological, on the potato murrain. *J. Hort. Soc.* **1**, 9–34 (1846).
5. Fry, W. E. et al. Five reasons to consider phytophthora *infestans* a reemerging pathogen. *Phytopathology* **105**, 966–981 (2015).
6. Ristaino, J. B., Cooke, D. E., Acuña, I. & Muñoz, M. *The Threat of Late Blight to Global Food Security. Emerging Plant Diseases and Global Food Security* (eds, Ristaino, J. and Records, A.) 101–132 (American Phytopathological Society Press, 2020).
7. Nowicki, M., Foolad, M. R., Nowakowska, M. & Kozik, E. U. Potato and tomato late blight caused by *Phytophthora infestans*: an overview of pathology and resistance breeding. *Plant Dis.* **96**, 4–17 (2012).
8. Leesutthiphonchai, W., Vu, A. L., Ah-Fong, A. M. V. & Judelson, H. S. How does *Phytophthora infestans* evade control efforts? modern insight into the late blight disease. *Phytopathology* **108**, 916–924 (2018).
9. Paluchowska, P., Śliwka, J. & Yin, Z. Late blight resistance genes in potato breeding. *Planta* **255**, 127 (2022).
10. Andersen, E. J., Ali, S., Byamukama, E., Yen, Y. & Nepal, M. P. Disease resistance mechanisms in plants. *Genes* **9**, 339 (2018).
11. Vleeshouwers, V. G. et al. Understanding and exploiting late blight resistance in the age of effectors. *Annu. Rev. Phytopathol.* **49**, 507–531 (2011).
12. Wawra, S. et al. The RxLR motif of the host targeting effector AVR3a of *Phytophthora infestans* is cleaved before secretion. *Plant Cell* **29**, 1184–1195 (2017).
13. Flor, H. Inheritance of pathogenicity in *Melampsora lini*. *Phytopathology* **32**, 653–669 (1942).
14. Glendinning, D. R. Potato introductions and breeding up to the early 20th century. *N. Phytologist* **94**, 479–505 (1983).
15. Lindley, J. Notes on the wild potato. *J. R. Hort. Soc.* **3**, 65–72 (1848).
16. Salaman, R. N. *Potato Varieties*. (Cambridge University Press, 1926).
17. Powell, W., Baird, E., Duncan, N. & Waugh, R. Chloroplast DNA variability in old and recently introduced potato cultivars. *Ann. Appl. Biol.* **123**, 403–410 (1993).
18. Hosaka, K. Distribution of the 241 bp deletion of chloroplast DNA in wild potato species. *Am. J. Pot. Res.* **79**, 119–123 (2002).
19. Black, W. XVII—inheritance of resistance to blight (*Phytophthora infestans*) in potatoes: inter-relationships of genes and strains. *Proc. R. Soc. Edinb., Sect. B: Bio. Sci.* **64**, 312–352 (1951).
20. Reddick, D. Problems in breeding for disease resistance. *Chron. Botanica* **6**, 73–77 (1940).
21. Fry, W. *Phytophthora infestans*: the plant (and R gene) destroyer. *Mol. Plant Pathol.* **9**, 385–402 (2008).
22. Reddick, D. & Mills, W. Building up virulence in *Phytophthora infestans*. *Am. Pot. J.* **15**, 29–34 (1938).
23. Stewart, H., Bradshaw, J. & Pande, B. The effect of the presence of R-genes for resistance to late blight (*Phytophthora infestans*) of potato (*Solanum tuberosum*) on the underlying level of field resistance. *Plant Pathol.* **52**, 193–198 (2003).
24. Stefańczyk, E., Sobkowiak, S., Brylińska, M. & Śliwka, J. Expression of the potato late blight resistance gene Rpi-phu1 and *Phytophthora infestans* effectors in the compatible and incompatible interactions in potato. *Phytopathology* **107**, 740–748 (2017).
25. Ghislain, M. et al. Stacking three late blight resistance genes from wild species directly into African highland potato varieties confers complete field resistance to local blight races. *Plant Biotech. J.* **17**, 1119–1129 (2019).
26. Varshney, R. K. & Tuberosa, R. Translational genomics in crop breeding for biotic stress resistance: an introduction. *Transl. Genomics Crop Breed.: |Biot. Stress* **1**, 1–9 (2013).
27. Win, J. et al. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* **19**, 2349–2369 (2007).
28. Haas, B. J. et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**, 393–398 (2009).
29. Matson, M. E., Liang, Q., Lonardi, S. & Judelson, H. S. Karyotype variation, spontaneous genome rearrangements affecting chemical insensitivity, and expression level polymorphisms in the plant pathogen *Phytophthora infestans* revealed using its first chromosome-scale assembly. *PLoS Pathog.* **18**, e1010869 (2022).
30. May, K. J. & Ristaino, J. B. Identity of the mtDNA haplotype (s) of *Phytophthora infestans* in historical specimens from the Irish potato famine. *Mycol. Res.* **108**, 471–479 (2004).
31. Ristaino, J. B. The importance of mycological and plant herbaria in tracking plant killers. *Front. Ecol. Evol.* **7**, 521 (2020).
32. Ristaino, J. B., Groves, C. T. & Parra, G. R. PCR amplification of the Irish potato famine pathogen from historic specimens. *Nature* **411**, 695–697 (2001).
33. Martin, M. D. et al. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nat. Commun.* **4**, 2172 (2013).
34. Yoshida, K. et al. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *elife* **2**, e00731 (2013).
35. Witek, K. et al. Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nat. Biotechnol.* **34**, 656–660 (2016).
36. Thilliez, G. J. et al. Pathogen enrichment sequencing (PenSeq) enables population genomic studies in oomycetes. *N. Phytologist* **221**, 1634–1648 (2019).
37. Lin, X. et al. Identification of Avramr1 from *Phytophthora infestans* using long read and cDNA pathogen-enrichment sequencing (PenSeq). *Mol. Plant Pathol.* **21**, 1502–1512 (2020).
38. Jupe, F. et al. Resistance gene enrichment sequencing (Ren Seq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* **76**, 530–544 (2013).
39. Orlando, L. et al. Ancient DNA analysis. *Nat. Rev. Meth. Prim.* **1**, 14 (2021).
40. Saville, A. C. & Ristaino, J. B. Global historic pandemics caused by the FAM-1 genotype of *Phytophthora infestans* on six continents. *Scien. Rep.* **11**, 12335 (2021).
41. Knaus, B. J., Tabima, J. F., Shakya, S. K., Judelson, H. S. & Grünwald, N. J. Genome-wide increased copy number is associated with emergence of dominant clones of the irish potato famine pathogen *Phytophthora infestans*. *mBio.* **11**, e00326 (2020).
42. Bos, J. I., Chaparro-Garcia, A., Quesada-Ocampo, L. M., Gardener, B. B. M. & Kamoun, S. Distinct amino acids of the *Phytophthora infestans* effector AVR3a condition activation of R3a hypersensitivity and suppression of cell death. *Mol. Plant-Microbe Inter.* **22**, 269–281 (2009).
43. Yang, L.-N. et al. The *Phytophthora infestans* AVR2 effector escapes R2 recognition through effector disordering. *Mol. Plant-Microbe Inter.* **33**, 921–931 (2020).
44. Lokossou, A. A. et al. Exploiting knowledge of R/Avr genes to rapidly clone a new LZ-NBS-LRR family of late blight resistance genes from potato linkage group IV. *Mol. Plant-Microbe Inter.* **22**, 630–641 (2009).

45. Black, W., Mastenbroek, C., Mills, W. & Peterson, L. C. A proposal for an international nomenclature of races of *Phytophthora infestans* and of genes controlling immunity in *Solanum demissum* derivatives. *Euphytica* **2**, 173–179 (1953).
46. Malcolmson, J. F. & Black, W. New R genes in Solanum demissum Lindl. and their complementary races of *Phytophthora infestans* (Mont.) de Bary. *Euphytica* **15**, 199–203 (1966).
47. Du, Y. et al. RXLR effector diversity in *Phytophthora infestans* isolates determines recognition by potato resistance proteins; the case study AVR1 and R1. *Stud. Mycol.* **89**, 85–93 (2018).
48. Reddick, D. Frost-tolerant and blight-resistant Potatoes. *Phytopathology* **20**, 987–991 (1930).
49. Martin, M. D. et al. Genomic characterization of a South American Phytophthora hybrid mandates reassessment of the geographic origins of *Phytophthora infestans*. *Mol. Biol. Evol.* **33**, 478–491 (2016).
50. Salaman, R. *Potato Varieties, Past Present and Future. The History and Social Influence of the Potato.* (Cambridge Univ. Press, 1949)
51. Ames, M. & Spooner, D. M. DNA from herbarium specimens settles a controversy about origins of the European potato. *Am. J. Bot.* **95**, 252–257 (2008).
52. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
53. Krueger, F. Trim galore. A wrapper tool around cutadapt and fastqc to consistently apply quality and adapter trimming to *FastQ files* https://github.com/FelixKrueger/TrimGalore (2015).
54. Andrews, S. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).
55. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
56. Xu, X. et al. Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195 (2011).
57. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997* (2013).
58. Broad Institute. Picard toolkit. Broad Institute, GitHub repository. https://broadinstitute.github.io/picard/ (2019).
59. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
60. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
61. Van der Auwera, G. A. & O'Connor, B. D. *Genomics In The Cloud: Using Docker, GATK, and WDL in Terra*. (O'Reilly Media, 2020).
62. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *fly* **6**, 80–92 (2012).
63. Quinlan, A. R. BEDTools: the swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinforma.* **47**, 11.12. 11–11.12. 34 (2014).
64. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes De novo assembler. *Curr. Protoc. Bioinforma.* **70**, e102 (2020).
65. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).
66. Weiß, C. L., Pais, M., Cano, L. M., Kamoun, S. & Burbano, H. A. nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinforma.* **19**, 1–8 (2018).
67. Sukumaran, J. & Holder, M. T. DendroPy: a python library for phylogenetic computing. *Bioinformatics* **26**, 1569–1571 (2010).
68. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
69. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
70. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
71. Maddison, W. & Maddison, D. Mesquite 2. a modular system for evolutionary analysis. Version 2.1. https://api.semanticscholar.org/CorpusID:116015395 (2007).
72. Goluch, T., Bogdanowicz, D. & Giaro, K. Visual treeCmp: comprehensive comparison of phylogenetic trees on the web. *Meth. Ecol. Evol.* **11**, 494–499 (2020).
73. Ignatieva, A., Lyngsø, R. B., Jenkins, P. A. & Hein, J. KwARG: parsimonious reconstruction of ancestral recombination graphs with recurrent mutation. *Bioinformatics* **37**, 3277–3284 (2021).
74. Ellson, J., Gansner, E. R., Koutsofios, E., North, S. C. & Woodhull, G. Graphviz and dynagraph—static and dynamic graph drawing tools. *Graph drawing software*, 127–148 (2004).
75. Ballvora, A. et al. The R1 gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. *Plant J.* **30**, 361–371 (2002).
76. Vleeshouwers, V. G. et al. Effector genomics accelerates discovery and functional profiling of potato disease resistance and *Phytophthora infestans* avirulence genes. *PLoS one* **3**, e2875 (2008).
77. Gilroy, E. M. et al. Presence/absence, differential expression and sequence polymorphisms between PiAVR2 and PiAVR2-like in *Phytophthora infestans* determine virulence on R2 plants. *N. Phytologist* **191**, 763–776 (2011).
78. Huang, S. et al. Comparative genomics enabled the isolation of the R3a late blight resistance gene in potato. *Plant J.* **42**, 251–261 (2005).
79. Armstrong, M. R. et al. An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. *Proc. Nat. Acad. Sci. USA* **102**, 7766–7771 (2005).
80. Li, G. et al. Cloning and characterization of R3b; members of the R3 superfamily of late blight resistance genes show sequence and functional divergence. *Mol. Plant-Microbe Inter.* **24**, 1132–1142 (2011).
81. Rietman, H. et al. Qualitative and quantitative late blight resistance in the potato cultivar Sarpo Mira is determined by the perception of five distinct RXLR effectors. *Mol. Plant-Microbe Inter.* **25**, 910–919 (2012).
82. van Poppel, P. M., Huigen, D. J. & Govers, F. Differential recognition of *Phytophthora infestans* races in potato R4 breeding lines. *Phytopathology* **99**, 1150–1155 (2009).
83. Van Poppel, P. M. et al. The *Phytophthora infestans* avirulence gene Avr4 encodes an RXLR-dEER effector. *Mol. Plant-Microbe Inter.* **21**, 1460–1470 (2008).
84. Vossen, J. H. et al. The *Solanum demissum* R8 late blight resistance gene is an Sw-5 homologue that has been deployed worldwide in late blight resistant varieties. *Theor. Appl. Genet.* **129**, 1785–1796 (2016).
85. Bradshaw, J. E., Bryan, G. J., Lees, A. K., McLean, K. & Solomon-Blackburn, R. M. Mapping the R10 and R11 genes for resistance to late blight (*Phytophthora infestans*) present in the potato (*Solanum tuberosum*) R-gene differentials of Black. *Theor. Appl. Genet.* **112**, 744–751 (2006).
86. van der Vossen, E. A. et al. The Rpi-blb2 gene from *Solanum bulbocastanum* is an Mi-1 gene homolog conferring broad-spectrum late blight resistance in potato. *Plant J.* **44**, 208–222 (2005).
87. Oh, S.-K. et al. In planta expression screens of *Phytophthora infestans* RXLR effectors reveal diverse phenotypes, including activation of the *Solanum bulbocastanum* disease resistance protein Rpi-blb2. *Plant Cell* **21**, 2928–2947 (2009).

88. Foster, S. J. et al. Rpi-vnt1. 1, a Tm-22 homolog from *Solanum venturii*, confers resistance to potato late blight. *Mol. Plant-microbe Inter.* **22**, 589–600 (2009).
89. Pel, M. A. *Mapping, isolation and characterization of genes responsible for late blight resistance in potato*. (Wageningen University and Research, 2010).
90. Ellson, J., Gansner, E. R., Koutsofios, E., North, S. C. & Woodhull, G. Graphviz and Dynagraph — Static and Dynamic Graph Drawing Tools. In *Graph Drawing Software. Mathematics and Visualization* (eds Jünger, M. & Mutzel, P.) 127–148 (Springer, 2004).

Acknowledgements

Funding for salaries for AC provided by NSF AgBioFews Training Grant Number, 2018-1966, JR Co-PI and salaries for AS, sequencing and supplies funded by Grip4PSI Grant Number 5572998, JR Co-PI, respectively. Additional research funds for AC work from a Seed Grant from the Triangle Center for Evolutionary Medicine and Grip4PSI Grant Number 5572998 JR Co-PI. Appreciation is expressed to Dr. Ross Whetten, Dr. Justin Whitehill, Dr. David Rasmussen, Dr. Jeff Thorne and Dr. Ignazio Carbone from NC State University, Ingo Hein, James Hutton Institute and Dr. Michael Martin, Norwegian Institute of Science and Technology for advice on the computational methods and workflows used to analyze data in this work.

Author contributions

A.C. and J.R. conceived the idea for the work. A.C., A.S., and J.R. edited and authored the paper. A.C. and A.S. performed the experiments. A.C. developed computational workflows and analyzed the data. J.R. is a senior author, and A.C. is the first author of the paper. Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-50749-4.

Correspondence and requests for materials should be addressed to Jean Beagle Ristaino.

Peer review information *Nature Communications* thanks Guillaume Besnard, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.