Theoretical and Applied Genetics (2026) 139:92
https://doi.org/10.1007/s00122-026-05179-9

ORIGINAL ARTICLE

Phased potato genome assembly and association genetics enable delineation of the H1 resistance locus against potato cyst nematodes

Yuk Woon Cheung[1,2] · Lynn H. Brown[1,2] · Thomas M. Adams[1] · Brian Harrower[1] · Amanpreet Kaur[1,3] · Gaynor McKenzie[1] · Jamie Orr[1] · James Price[1] · Vikrant Singh[1] · Moray Smith[2] · Micha Bayer[1] · Ingo Hein[1,2]

Received: 26 September 2025 / Accepted: 4 February 2026 / Published online: 12 March 2026

Abstract

The complexity of potato genetics, characterised by tetrasomic inheritance, has contributed to slower genetic gain in potato compared to other major crops. Disease resistance genes, often found in large clusters of highly similar paralogs and alleles, further complicate genetic studies. The *H1* resistance locus, introgressed into potato cultivars from *Solanum tuberosum* ssp. *andigena*, has been successfully used for over 60 years to control *Globodera rostochiensis* in Europe. Although previous genetic studies mapped this resistance to chromosome 5, the complete structure of the locus remained elusive. To reduce genomic complexity, we generated a dihaploid of the cultivar 'Athlete', DH4_Athlete, carrying the *H1* resistance locus, and produced a phased haplotype representation of the *H1* interval using Oxford Nanopore sequencing. Combined with RenSeq-based association genetics, this approach allowed us to reconstruct the entire *H1* locus, including recombination points at both the 5′ and 3′ ends of the interval.

Introduction

Potato (*Solanum tuberosum*) is the world's most important non-cereal food crop, consumed by over a billion people and serving as a dietary staple in numerous countries (Birch et al. 2012; Wijesinha-Bettoni and Mouillé 2019). In 2023, global production reached approximately 383 million tonnes, underlining its significance in global food and economic security (FAO 2023). However, the sustainability of potato production is increasingly challenged by biotic stresses, particularly pests and pathogens, which collectively account for an estimated 21% yield loss (Savary et al. 2019). Among these, potato cyst nematodes (PCN) pose one of the most important threats, contributing to around 9% of total yield reductions worldwide whilst threatening the potato supply chain as contaminated land is unsuitable for seed potato production (Gartner et al. 2021).

PCN in Europe commonly refers to two major species, *Globodera rostochiensis* and *G. pallida*, soil-borne nematodes with complex, obligate parasitic life cycles. A major challenge in PCN management is their ability to remain dormant in soil, enabled by a tightly regulated hatching process. After fertilisation, the cuticle of the gravid female hardens, forming a cyst that encases often hundreds of eggs. This cyst detaches from the host roots and persists in the soil for over 20 years, with eggs hatching only in response to specific chemical signals released by the roots of a suitable host plant (Wainer and Dinh 2021; Price et al. 2021). This remarkable longevity makes PCN difficult to eradicate and has led to its designation as a quarantine pest in over 100 countries, severely restricting international seed potato trade and causing significant economic losses (Bairwa et al. 2024).

Resistance breeding against *G. rostochiensis* has yielded notable successes with the discovery of the *H1* resistance, which confers near-complete resistance to the Ro1 and Ro4

Communicated by Jeffrey Endelman.

Yuk Woon Cheung, Lynn H. Brown have authors contributed equally to this study.

✉ Micha Bayer
Micha.Bayer@hutton.ac.uk

✉ Ingo Hein
Ingo.Hein@hutton.ac.uk

[1] The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK

[2] The University of Dundee, Nethergate, Dundee DD1 4HN, UK

[3] International Potato Centre (CIP), Avenida La Molina 1895, La Molina, Apartado 1558, Lima, Peru

pathotypes that are prevalent in many potato growing countries outside South America (Evans and Brodie 1980). The wild subspecies *Solanum tuberosum* ssp. *andigena* CPC 1673 was one of the earliest potato lines to present the resistance phenotype to potato cyst nematodes, observed while screening 1200 wild accession (Ellenby 1952). It was collected by Dr. Cárdenas in La Paz, Bolivia, and received in the UK in December 1944. Notably, the material arrived as insect-damaged tubers, which were ultimately discarded by 1949. The line survived only as true seed, which was later grown and used in a sibling cross to produce TBRADG 3520 in 1962 (G. McKenzie, personal communication). Lack of PCN resistance in cultivated varieties, led to the use of wild species in breeding programmes, which has deployed resistance genes such as H1 in varieties around the world for over 60 years (Spychalla and De Jong 2024).

Although the *H1* gene has not yet been cloned, it has been mapped to the distal end of chromosome 5, a region rich in NLR (nucleotide binding leucine rich-repeat) resistance genes. Assembly of the H1 introgression segment is challenging due to its high density of resistance gene analogues and associated structural complexity. Once assembled, however, resistant and susceptible haplotypes are clearly distinguishable. In contrast, genetic delimitation of the locus remains constrained by low recombination rates. Previous studies, including those by Bakker et al. (2004) and Finkers-Tomczak et al. (2011), defined a region within the resistant haplotype SH83-92–488. However, this region did not reveal the causal gene responsible for resistance to *G. rostochiensis*.

NLR proteins are a major class of intracellular immune receptors in plants, typically comprising a central NB-ARC (nucleotide-binding adaptor shared by APAF-1, certain *R* gene products, and CED-4) domain, a C-terminal LRR (leucine-rich repeat) domain, and an N-terminal CC (coiled-coil) or TIR (toll/interleukin-1 receptor-type) domain. To target and analyse these genes, RenSeq (Resistance gene enrichment Sequencing) was developed to specifically sequence NLR-encoding genes from complex plant genomes (Jupe et al. 2013). This approach has enabled comprehensive annotation and evolutionary studies of the NLRome, the complete set of NLR genes in a genome.

RenSeq has since evolved into a suite of methods, including SMRT-RenSeq (PacBio-based long-read sequencing) (Witek et al. 2016), AgRenSeq (*k*-mer-based association genetics) (Arora et al. 2019), and dRenSeq (diagnostic detection of known NLRs) (Van Weymers et al. 2016; Armstrong et al. 2019). These were recently unified into the SMRT-AgRenSeq-d pipeline, which has proven effective in resolving complex loci, such as the *Gpa5* locus in tetraploid potato, where 9 candidate NLR genes linked to the Gpa5 resistance were identified (Wang et al. 2023; Adams et al. 2023).

In this study, we initially used SMRT-AgRenSeq-d to identify candidate NLR genes associated with *H1*-mediated resistance in the cultivar 'Buster'. To develop a complete and accurate representation of the *H1* locus, we adapted AgRenSeq to produce phased Oxford Nanopore haplotypes from the dihaploid clone DH4_Athlete, derived from the resistant cultivar 'Athlete' using IVP48. We further demonstrate that the resolved haplotype alone represents a significant advancement for genome-based association genetics, enabling high-resolution analysis of trait associations. This combination of phased haplotypes and association genetics provides a detailed representation of the *H1* haplotype, including recombination events at both the 5′ and 3′ ends of the locus.

Results

SMRT-RenSeq

The potato cultivar 'Buster', released in 2020, is well known for its robust resistance to potato cyst nematode (PCN), particularly *G. rostochiensis* pathotypes Ro1 and Ro4, and was used as a reference to represent the NLRome of a resistant cultivar. Using the SMRT-RenSeq workflow from the nfHISS pipeline (Wang et al. 2023; Adams et al. 2023), contigs were assembled from PacBio HiFi RenSeq reads of 'Buster'. The assembly yielded 6292 contigs with an N50 of 11,379 bp. Of these, 2017 contigs contained a total of 2870 NLR-like sequences, of which 1,537 were annotated as complete NLRs, with the remainder made up of pseudogenous and/or incomplete NLRs. A dRenSeq analysis confirmed the presence of several previously cloned resistance genes in 'Buster', including *Rpi-R3a*, *Rpi-R3b*, *Rx*, and *Gpa2*. Each of these genes were fully and accurately represented on individual contigs in the PacBio HiFi assembly (Table 1).

Development of an association panel

To investigate the genetic basis of the H1 resistance and to identify candidate resistance genes at the *H1* locus, we developed an association panel comprising 126 potato cultivars. This panel included accessions with contrasting phenotypes: 58 highly resistant and 68 highly susceptible cultivars. The resistance and susceptibility phenotypes of 126 cultivars to *G. rostochiensis* pathotypes Ro1 and Ro4 were evaluated, with *H1*-linked marker information available for a subset. Phenotypic and marker data were compiled from publicly available databases and commercial sources. A detailed summary of the selected cultivars and their resistance profiles is provided in Table S1. Notably, the panel includes TBRADG 3520 (CPC 3520), a resistant

Table 1 Representation of four benchmark genes following HiFi-based assembly

Buster Reference NLR representation									
Reference Gene	Representative Sequence	Identity (%)	Gene length	Mismatch	Gap	Reference gene start	Reference gene end	Coverage of gene (%)	Coverage of HSP (%)
Rpi_R3a	tig00000259	100	3849	0	0	1	3849	100	100
Rpi_R3b	tig00000358	100	3852	0	0	1	3852	100	100
Gpa2	tig00002324	100	2976	0	0	1	2976	100	100
Rx	tig00001251	100	3045	0	0	1	3045	100	100

sibling clone of CPC 1673, the earliest reported source of *H1*-mediated resistance.

SMRT-AgRenSeq-d

The SMRT-AgRenSeq-d analysis, applied to the association panel described above (Table S1), identified candidate NLR contigs associated with *H1*-mediated resistance in the cultivar 'Buster'. The AgRenSeq component of the analysis is based on a *k*-mer-based approach to identify candidate genes using a linear regression model (Arora et al. 2019). The association mapping, visualised in Fig. 1A, revealed 89 contigs which contained 131 NLRs with varying positive association scores. These associated contigs, marked in red, were identified using an inclusive threshold of half the maximum association score for first pass candidate identification.

To refine candidate selection, dRenSeq was used to determine the presence or absence of NLR contigs across the resistant and susceptible cultivars, enabling the calculation of F1 scores (Fig. 1B; Table S2). The F1 score is particularly

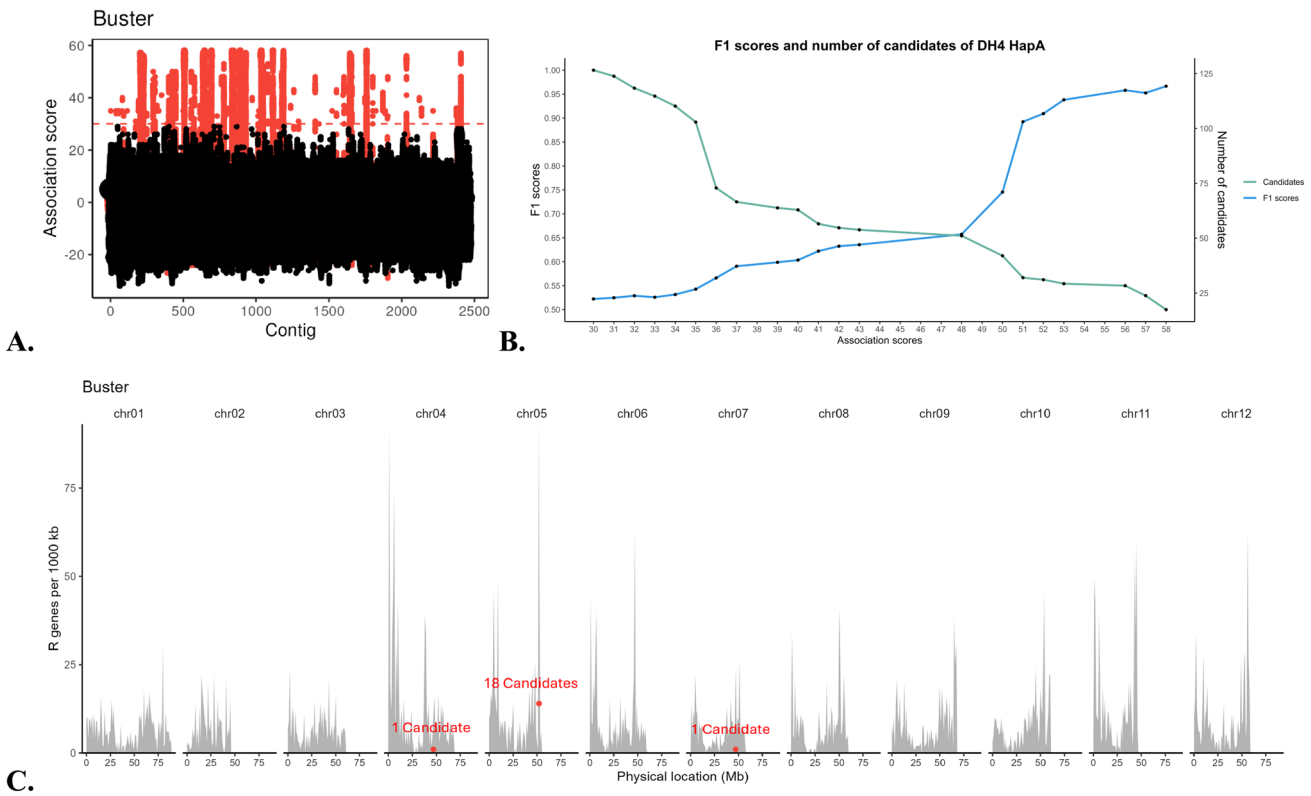
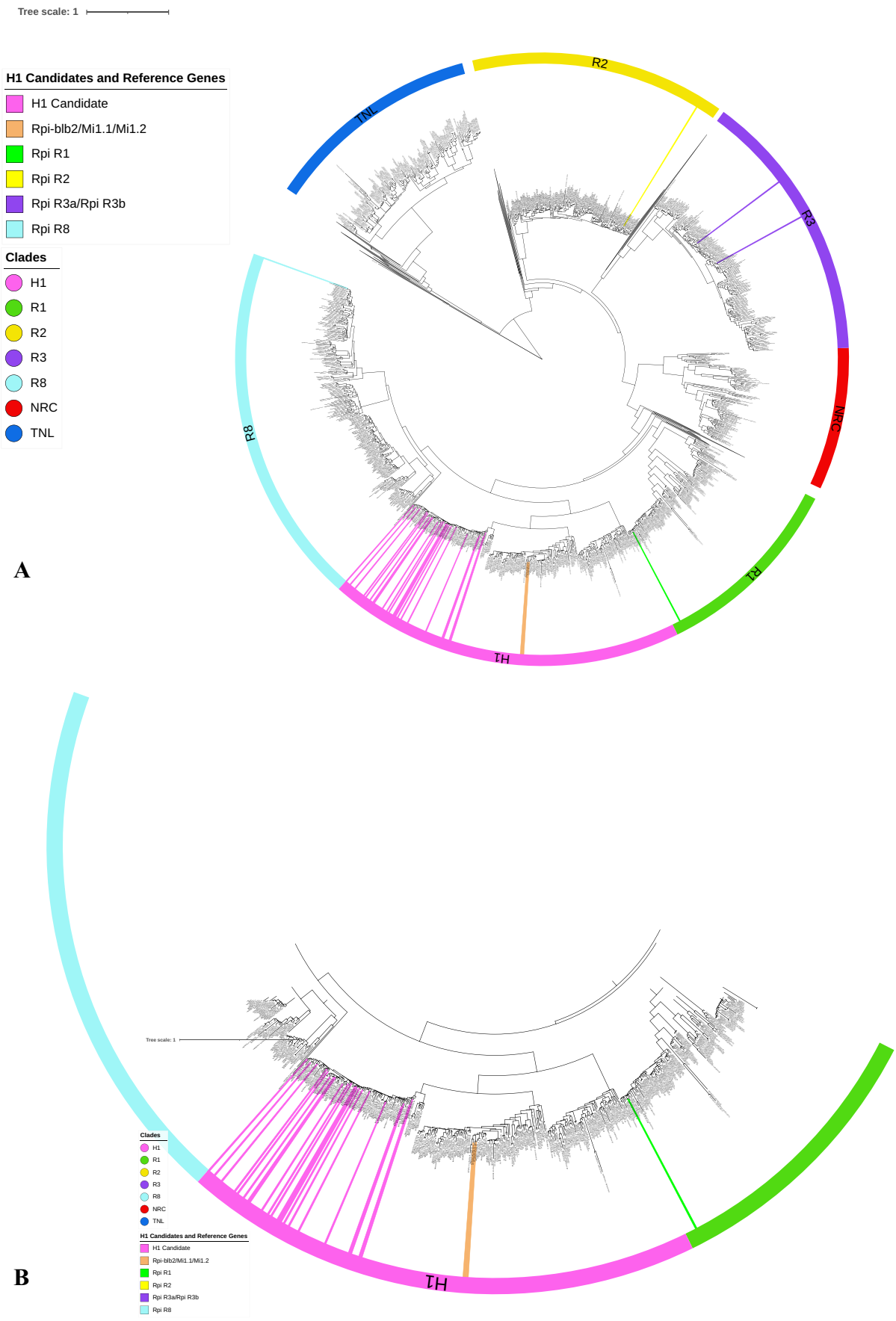


Fig. 1 SMRT-AgRenSeq-d analysis of H1. **A** Identification of NLR containing contigs in the reference cultivar 'Buster' with association to H1. Columns on the x-axis represent NLR containing contigs, with dots indicating mapped k-mers and their AgRenSeq-derived association score to H1 resistance. The dashed line highlights the chosen threshold (50% of the maximum = 29) for inclusive candidate selection. **B** F1 scores and number of candidate NLRs for each association score group. **C** The predicted positions of H1 candidates with F1 score > 0.95 are shown (red dots) in relation to all NLRs from reference Buster mapped onto DM v6.1 over all 12 chromosomes


Tree scale: 1
H1 Candidates and Reference Genes
H1 Candidate
Rpi-blb2/Mi1.1/Mi1.2
Rpi R1
Rpi R2
Rpi R3a/Rpi R3b
Rpi R8
Clades
H1
R1
R2
R3
R8
NRC
TNL
A
B

◂**Fig. 2** A Phylogenetic tree derived from an alignment of NB domains of 'Buster' NLRs following assembly of SMRT-RenSeq reads. Expanded are the clades containing previously identified functional NLRs such as Rpi-R1 *(highlighted in green),* Rpi-R2 *(Highlighted yellow),* Rpi-R3a *and* Rpi-R3b *(highlighted in purple) alongside NRCs from tomato (*Solanum lycopersicum*),* Nicotiana benthamiana *and* Beta vulgaris*. Collapsed clades consist of NLR clades which are not investigated in this study. Highlighted in pink are the* H1 *candidates with F1 score > 0.95 (No NB domain was identified in the candidate on tig00002155, hence it was excluded from this analysis). The NB containing proteins* APAF1 *from* Homo sapiens *and* CED4 *from* Caenorhabditis elegans *act as outgroups. The tree scale indicates substitutions/site. B Zoomed-in view of candidate H1 clade. Highlighted in pink are H1 candidates with F1 score > 0.95. Highlighted in orange are reference genes* Mi-1.1 *and* Mi-1.2 *from* Solanum peruvianum*, and* Rpi-blb2 *from* Solanum bulbocastanum

useful for evaluating candidate genes based on dRenSeq coverage, as it quantifies how reliably and accurately a gene's presence or absence distinguishes resistant from susceptible varieties. Using a presence/absence matrix across varieties, the F1 score highlights genes that are consistently associated with resistance, supporting confident candidate gene identification. F1 score, where absolute linkage is denoted by 1 and any variation—such as that caused by recombination—reduces the F1 score below 1. Using this approach, we identified a subset of 20 NLR contigs surpassing an F1 score of 0.95, of which five candidates had an F1 score of 1. Mapping these high-confidence candidates onto the reference genome DM v6.1, which lacks the H1 resistance, (Fig. 1C) shows their predominant localisation to the distal end of chromosome 5, aligning with the previously described genomic position of the *H1* resistance locus. However, two candidates map onto chromosomes 4 and 7 of DM v6.1, respectively.

To further investigate the relationships among the highly associated and genetically linked candidate NLRs, a phylogenetic tree was constructed using the NB domains from SMRT-RenSeq-derived NLRs in 'Buster' (Fig. 2). The resulting tree revealed that all high-confidence candidates clustered tightly within a single clade, suggesting a common evolutionary origin and potentially, functional relatedness. This clade also includes known resistance genes such as *Mi-1.1* and *Mi-1.2*, two wild tomato (*Solanum peruvianum*) genes that confer resistance to root-knot nematodes and several phloem-feeding insects (Rossi et al. 1998), as well as *Rpi-blb2*, an NLR from *Solanum bulbocastanum* that provides resistance to *Phytophthora infestans* (van der Vossen et al. 2005). These findings are consistent with a previous study of the *H1* locus by Finkers-Tomczak et al. (2011), which identified 55 resistance gene homologs (RGH) showing approximately 52% amino sequence identity with *Rpi-blb2*, a member of the CC-NB-LRR class of resistance proteins (Finkers-Tomczak et al. 2011). One exception was the candidate on tig0000215, which lacked an identifiable NB domain and was therefore excluded from the analysis.

DM lacks *H1*-mediated resistance and therefore its genome can function only as a proxy reference for genomic studies at this locus. To address this limitation, a physical representation of the H1 locus was generated using long-read Oxford Nanopore Technologies (ONT) sequencing of a dihaploid clone derived from the cultivar 'Athlete' (DH4_Athlete) using IVP48. The resulting haplotype-resolved genome assembly captured all 20 previously identified H1 candidate NLR candidates with F1 scores > 0.95, as well as four additional candidates with F1 scores < 0.95 (Table S3) and was subsequently used as a reference for association mapping.

ONT sequencing of a dihaploid derived from 'Athlete'

ONT sequencing of DH4_Athlete yielded 13.9 million reads, with an N50 of 23.5 kbp and 77.8% of reads above Q20, providing approximately 84-fold coverage based on the reference potato genome DM v6.1 size of 840 Mb (The Potato Genome Sequencing Consortium. 2011). *K*-mer spectrum analysis confirmed the genome to be diploid (Figure S1). A partially phased assembly representing both haplotypes was produced. The assembly-derived Haplotype A consisted of 1,892 contigs (849 Mb total, N50 = 57.3 Mb), while Haplotype B comprised 100 contigs (779 Mb, N50 = 31.3 Mb). Both haplotypes showed high completeness (BUSCO > 96%). Both haplotypes exhibited a high degree of gene synteny and contiguity, with multiple chromosomes being resolved as telomere-to-telomere (Table 2, see also Figs. S2 and S3).

Identification of NLRs in DH4_Athlete

Genes were predicted de novo with Helixer (Holst et al. 2023), from which NLRs were identified with Resistify (Martin et al. 2022; Smith et al. 2025). This yielded a total of 1356 NLRs across both haplotypes, with 710 NLRs assigned to Haplotype A and 646 assigned to Haplotype B (Table 2).

ONT-AgRenSeq-d

The NLRs extracted from the partially phased haplotype assembly of DH4_Athlete was utilised for an association study using the same panel of varieties with contrasting phenotypes (Table S1). This analysis demonstrated that strong association peaks existed in both Haplotype A (Fig. 3A) and Haplotype B (Fig. 3B). A dRenSeq analysis was performed in addition of the F1 score calculations. This revealed candidates with an F1 score of 1 to be found only in Haplotype A and not in Haplotype B (Fig. 3C; Tables S4 and S5).

Further, analysis of *k*-mer counts across candidates from both haplotypes revealed that the strongest association is

Table 2 Assembly statistics for Haplotype A and Haplotype B of the partially phased DH4_Athlete dihaploid genome, showing metrics for Haplotype A and Haplotype B based on ONT sequencing data

Assembly statistics for Haplotype A and Haplotype B of the partially phased DH4_Athlete dihaploid		
	Haplotype A	Haplotype B
Contig N50	57.3Mbp	31.3Mbp
Number of contigs	1892	100
Total assembly size	849Mbp	779Mbp
BUSCO (Single; Duplicated)	96.1% (90.7%; 5.4%)	96.9% (93.3%; 3.6%)
Number of NLRs	710	646
GCI score	36.7	24.5
GCI contig number	45	46

found in the partially phased haplotype A, rather than haplotype B (Figure S4). This suggests that NLRs in haplotype B share some limited conserved regions with *H1* candidates. Supporting this, a BLAST comparison confirmed that candidate athlete_chr05_B_001411 from haplotype B shares 98% identity covering 82% of the candidate athlete_chr05_A_001451 from haplotype A, accounting almost exclusively for the elevated *k*-mer count and association signal (Table S6).

Using the presence and absence of candidates that are highlighted in the cultivars 'Strachan', 'Vitabella' and 'Alouette', the conserved *H1* introgression segment associated with resistance can be delimited to approximately 270 kilobases in the H1-containing haplotype A (Fig. 4). Differences in gene length and candidate presence between the 'Buster' SMRT-RenSeq contigs and DH4_Athlete predictions reflect methodological differences between motif-based NLR detection and full gene model prediction.

Comparison of the corresponding locus in DM v6.1 as well as haplotype B which lacks H1 candidate genes suggests significant locus expansion and rearrangement of NLRs in haplotype A (Fig. 5). Consistent with this, phylogenetic analysis of candidate genes within Haplotype A provides evidence of duplication followed by diversification, a process typically associated with locus expansion (Fig. 2).

Discussion

As shown in this study, ONT-based sequencing offers key advantages for NLR analysis, notably its ability to generate long reads that span complex genomic regions, including full-length NLR genes, structural variants, and multiple NLRs within gene clusters (Fig. 5). This facilitates accurate physical mapping of resistance loci and reveals gene clusters that are often fragmented or collapsed in PacBio HiFi and short-read assemblies. ONT is particularly effective for large, repetitive genomes like tetraploid potato, or, as shown here, a dihaploid of 'Athlete' (DH4_Athlete) with reduced genomic complexity. However, assembling highly heterozygous polyploid genomes using ONT remains challenging, as phasing can be incomplete and closely related alleles may be collapsed. By integrating ONT sequencing of a dihaploid with AgRenSeq-d, several of these limitations are mitigated, including reliance on the susceptible reference genome DMv6.1 for assigning candidate chromosomal positions and incomplete representation of the genomic context in which resistance gene reside. This combined approach enables *k*-mer-based association to contigs, supporting haplotype-specific analysis and more precise allele resolution.

H1 candidates identified using 'Buster' in the PacBio-based SMRT-AgRenSeq-d pipeline showed strong concordance with those from the ONT-based DH4_Athlete association study, as confirmed by BLAST (Table S7). Minor differences mainly arose from how coding sequences (CDS) were defined. The ONT-based pipeline used Helixer for de novo gene prediction and full gene model annotation, followed by NLR detection with Resistify. In contrast, the SMRT-AgRenSeq workflow used NLR-Annotator, which identifies NLR motifs but not full gene structures, instead inferring CDS boundaries based on the inclusion of 1 kb flanking regions, sometimes leading to truncated predictions. For example, a 'Buster'-based candidate on tig00002155 lacked an NB domain and was therefore absent from ONT predictions, as Resistify only reports NB-ARC-containing genes. By contrast, ONT-based annotations generally predicted longer gene models. One such ONT gene, DH4_Athlete_Chr05_A_001462, was 16 kbp longer than its Buster counterpart, resulting in a lower F1 score (< 0.95), most likely due to bait coverage bias, consistent with reduced enrichment efficiency in more diverse gene regions. Despite these discrepancies, high-confidence candidates ($F1 > 0.95$) remained tightly clustered, and those showing absolute linkage ($F1 = 1$) were confined to a ~ 270 kbp interval. Lower-scoring candidates ($F1 < 1$) within the same region therefore reflect structural and coverage-related variation rather than reduced biological relevance.

Importantly, despite differences in gene structure predictions, all high-confidence candidate genes and their corresponding contigs identified by SMRT-AgRenSeq-d in 'Buster' were independently mapped to Haplotype A of DH4_Athlete (Fig. 4). Notably, two high-confidence candidates initially mapped to chromosomes 4 (tig00002170_nlr_1) and 7 (tig00001229_nlr_1) of the DM v6.1 reference (Fig. 1C) were reliably located within the H1 introgression segment on chromosome 5 in DH4_Athlete Haplotype A, represented by athlete_chr05_A_001138 and athlete_chr05_A_001460 (Fig. 4; Table S7).

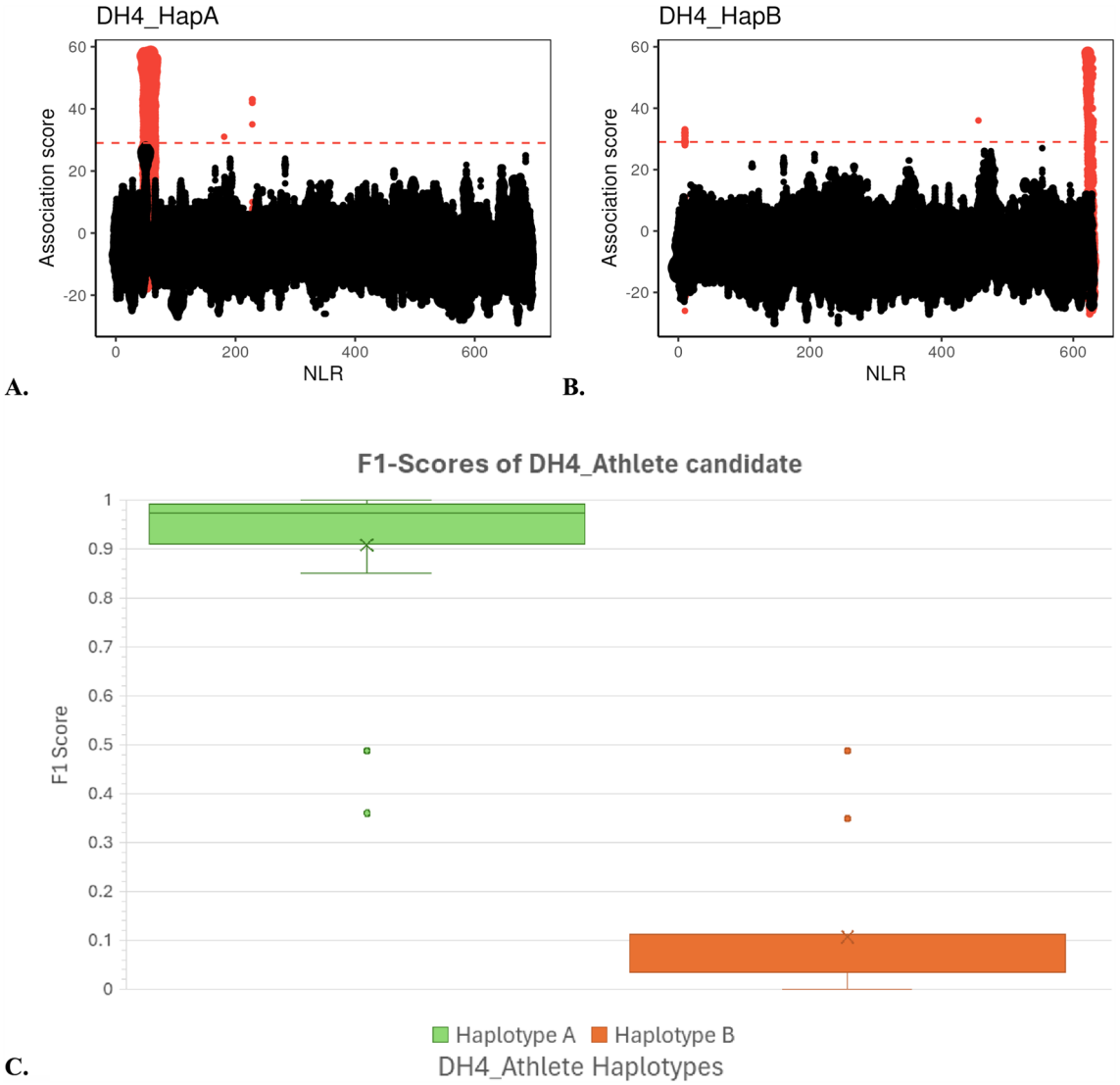


Fig. 3 **A** Association plot showing NLRs extracted from DH4_Athlete haplotype A with mapped k-mers and their respective association scores for H1. The x-axis represents individual NLRs, while dots on the y-axis indicate k-mer associations. The dashed red line marks the chosen association threshold (score = 29, e.g. 50% of maximum). **B** Equivalent association plot for DH4_Athlete haplotype B. **C** Boxplot of F1 scores for the 20 and 10 H1 candidate NLRs identified in the association studies with reference DH4_Athlete haplotype A and B, respectively

The dRenSeq analysis confirmed that the introgression segment from the resistant CPC 1632 sibling clone, the source of the original *H1* introgression segment, remains highly conserved. Finkers-Tomczak et al. (2011) identified 17 resistance gene homologs (RGHs) spanning ~ 314 kb within the *H1* introgression segment of the diploid clone SH83-92–488 (SH), including genes SH1 through SH10. In our analysis, we detected SH1–SH7 and SH10, indicating strong conservation in the northern region of the interval, previously linked to *H1* resistance in SH. However, expanded analysis across additional varieties revealed recombination events within SH2, SH3, SH4, SH5, and SH10. While SH2, SH3, SH4, SH5, and SH10 are absent in H1 resistant cultivar 'Vitabella', all five are present in resistant cultivar 'Alouette'. However, 'Alouette' lacks the first four *H1* candidate genes spanning approximately 78kbp at the proximal end of the introgression segment. Conversely, the susceptible cultivar 'Strachan' retains SH2, SH3, and SH4, but lacks SH5 and SH10suggesting recombination has decoupled resistance from these genes. We also observed variation in SH8 and SH9. For example, these genes are present in 'Vitabella', but absent in TBRADG 3520, 'Buster', and DH4_Athlete (confirmed by dRenSeq) (Table S8), helping to pinpoint a recombination breakpoint that separates the functional H1 resistance region of the introgression segment (Fig. 4).

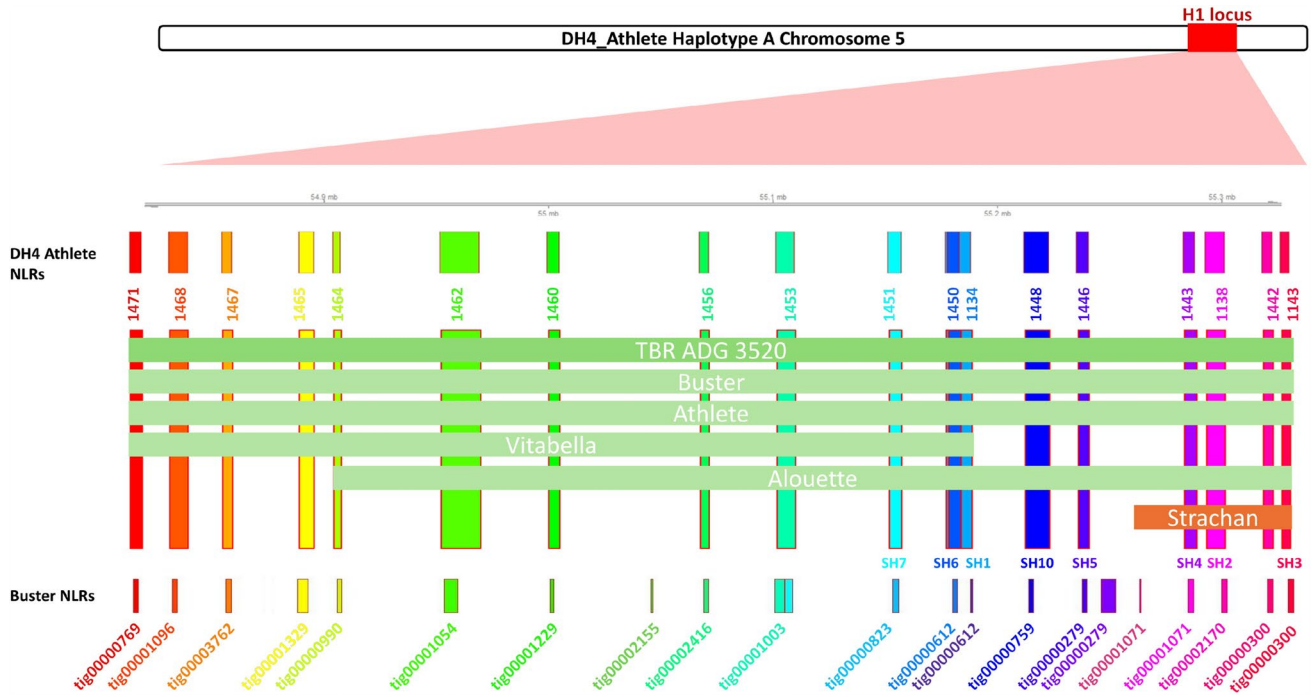


Fig. 4 *Representation of predicted candidate genes in DH4_Athlete Haplotype A. The upper panel shows the DH4_Athlete haplotype A of chromosome 5, with the H1 locus highlighted in red at the distal end (~ 54 Mb). Below, the predicted candidate genes from the association study using DH4 Athlete as a reference within this region are represented. The lower panel illustrates the conservation of the H1 introgression segment across various resistant (green bars) and susceptible (orange bar) cultivars. Candidates that are identical to* SH *genes previously reported by* Finkers-Tomczak et al. (2011) *are also indicated. Below are the candidate genes from the cv. 'Buster' SMRT-AgRenSeq-d assembly that mapped within this introgression segment*

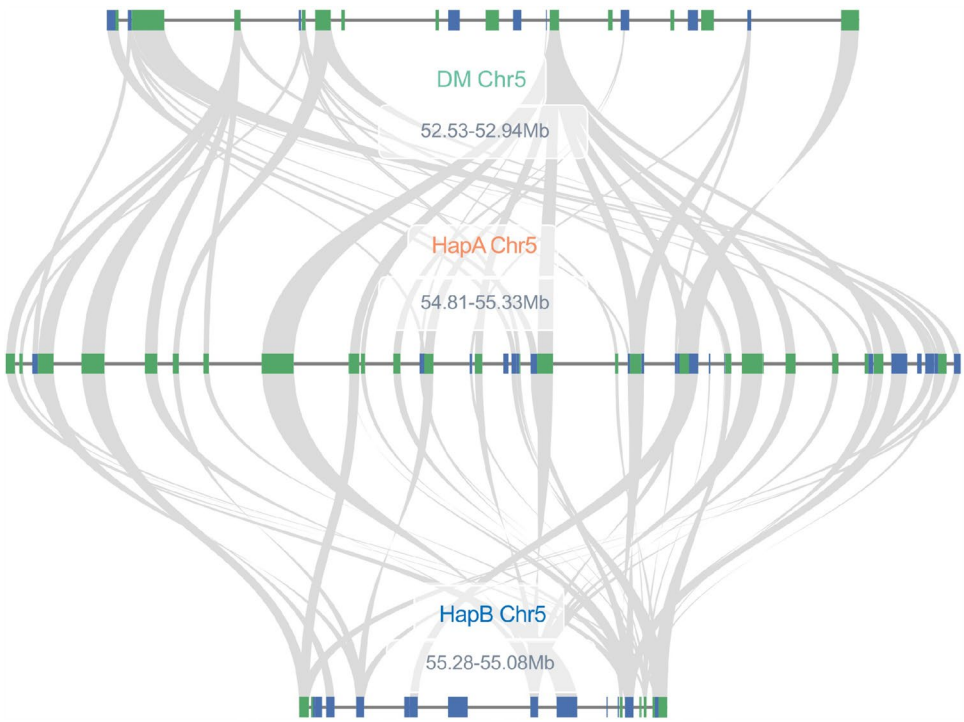


Fig. 5 Gene synteny at the H1 locus, delimited by outermost genes of the haplotype A cluster. Rectangles represent gene models predicted by Helixer. Gene orientation is indicated by colour, with blue representing genes on the forward strand and green representing genes on the reverse strand. The top track shows the corresponding region on DM v6.1 chromosome 5 containing genes with high sequence similarity to the H1 candidate genes in haplotype A represented by the middle track. The bottom track represents haplotype B. Grey connecting lines indicate syntenic relationships between genes across assemblies

In this study, ONT sequencing allowed us to fully resolve the H1 locus, including recombination boundaries and introgressed gene clusters. Importantly, the phased assembly of DH4_Athlete revealed that all high-confidence candidates from association genetics were located exclusively in haplotype A, highlighting the value of haplotype-specific information for distinguishing causal genes from non-functional homologs (Figs. 4 and 5). These findings represent the first genetically validated physical map of the H1 resistance locus in potato.

In summary, these results highlight the value of integrating long-read sequencing, partially phased genome assemblies, and association genetics to dissect complex resistance loci. This framework enables more effective candidate gene identification and supports informed, precise breeding strategies, not only for PCN resistance, but also for a broad range of traits governed by complex genomic architecture in polyploid crops.

Materials and methods

Plant materials

All potato accessions were grown in the glasshouse. In some cases, lyophilised DNA was received from collaborators. DH4_Athlete is a dihaploid plant of the tetraploid potato cultivar 'Athlete', generated by crossing with *Solanum phureja* IVP line 48, a dihaploid inducer (Hermsen and Verdenius 1973).

Extraction of DNA

As described in Wang et al. (2023), all potato accessions' DNA was extracted using the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. DNA from the cultivar 'Buster' for HiFi RenSeq sequencing was extracted according to manufacturer's instructions for the Wizard ® HMW DNA extraction Kit (Promega, Madison, WI, USA). High-molecular-weight DNA was extracted from fresh leaf tissue of dihaploid plants of the tetraploid potato cultivar 'Athlete' DH4_Athlete after three days of dark treatment, following the manufacturer's guidelines for the NucleoBond HMW DNA kit (Macherey–Nagel, Germany).

RenSeq

The RenSeq bait library version 5 was used as described in Armstrong et al. (2019). Enrichment sequencing was performed by Arbor Biosciences (Ann Arbor, MI, USA) using the Novaseq 6000 platform for Illumina reads and the PacBio Sequel II platform for HiFi reads.

Selection of association panel

The identification of *H1* used a panel of 126 cultivars which consisted of highly resistant and highly susceptible accessions to *Globodera rostochiensis* pathotype Ro1 and Ro4, where resistance was determined by consulting various databases and sources. Within the association panel were 58 resistant and 68 susceptible varieties. Phenotypes of these varieties were collected from various sources including: the European cultivated potato database, the AHDB potato database, Europlant, Solanum International Inc., Potato Pro, Bavaria Saat, The Potato Company, Greenvale, Solana United Kingdom, and from a previous study by Schultz et al. (2012) which has marker presence data available. Varieties identified as highly resistant were assigned a score of 1 while, varieties identified as highly susceptible were assigned a score of -1, summarised in Table S1.

The SMRT-AgRenSeq-d reference Buster was selected for its broad-spectrum resistance to both species of potato cyst nematode (*Globodera pallida* and *G. rostochiensis*). This was achieved through marker-assisted selection (MAS). The breeding programme combined multiple resistance loci, including GpaIV, Gpa5, Gpa6, Grp1, and H1, to pyramid resistance against diverse PCN pathotypes. This MAS-based approach enabled the deliberate stacking of complementary resistance genes, resulting in durable and broad-spectrum PCN resistance in Buster (Griffin et al. 2022).

The ONT-AgRenSeq-d reference DH4_Athlete was selected because expansion of the dRenSeq panel to include DH4_Athlete enabled the detection of all candidate *H1* resistance genes identified in the SMRT-AgRenSeq-d study. In addition, SH resistance genes identified in a previous study by Finkers-Tomczak et al. 2011 were also detected in DH4_Athlete, providing strong evidence that the *H1* introgression segment is present in this genetic background. Together, these observations confirm the suitability of DH4_Athlete for phased ONT haplotype reconstruction and detailed analysis of the *H1* locus (Table S8).

Computational analyses were performed on the James Hutton Crop Diversity HPC (High Performance Compute cluster), described by Percival-Alwyn et al. (2025).

SMRT-AgRenSeq-d

The nfHISS workflow version v1.1.0 was used to perform SMRT-AgRenSeq-d which includes SMRT-RenSeq assembly, AgRenSeq, and dRenSeq as described in Adams et al. (2023) and validated in Wang et al. (2023).

ONT-AgRenSeq-d

DH4_Athlete is a dihaploid plant of the tetraploid potato cultivar 'Athlete', generated by crossing with IVP (in vitro

pollinator) line 48. ONT sequencing libraries were prepared using the Ligation Sequencing gDNA – Native Barcoding Kit 24 V14 kit, according to manufacturer's guidelines. Libraries were sequenced on a combination of minION and PromethION R10.4.1 flow cells. After 24 h, flow cells were washed and reloaded with the same library, following the guidelines of the Flow Cell Wash Kit (EXP-WSH004; Oxford Nanopore Technologies).

Reads were basecalled using Dorado v0.9.6 (Kuśmirek 2023) with the dna_r10.4.1_e8.2_400bps_sup v4.1.0 and v5.1.0 models, depending on sampling rate. Genome ploidy and heterozygosity was determined with genomescope2 v2.0.1 (Ranallo-Benavidez et al. 2020), with 21-mers counted from ONT reads with kmc v3.2.4 (Kokot et al. 2017). Partially phased contigs were produced using hifiasm v0.25.0 (Cheng et al. 2021) with the –ont model and independently scaffolded to the doubled monoploid potato DM 1–3 516 R44 (v6.1) genome. Unscaffolded contigs were removed, and the resulting chromosomal scaffolds were merged into a single file. Genome completeness was evaluated with BUSCO v5.8.3 (Manni et al. 2021), using miniprot (Li 2023) with the solanales_odb12 lineage. Haplotype labels A and B are assembly-derived identifiers and are assigned arbitrarily. The contiguity of each haplotype assembly was assessed with GCI v1.0 (Chen et al. 2024), aligning ONT reads with minimap2 v2.30 (Li 2018) and winnowmap v2.03 (Jain et al. 2022) with 15-mers counted by meryl v1.4.1 (Rhie et al. 2020).

Genes were predicted ab initio with Helixer v0.3.4 (Holst et al. 2023) using the model land_plant_v0.3_a_0080, and transposable elements with HiTE v3.3.2 (Hu et al. 2024). NLRs were identified and classified from the gene annotations with Resistify v1.2.0 (Martin et al. 2022; Smith et al. 2025) with CoCoNat (Madeo et al. 2023) enabled to increase CC sensitivity. The resulting NLR predictions were then passed through the AgRenSeq workflow.

F1 Score

Candidate genes extracted from our highly associated contigs were further refined through generating F1 scores as described by Wang et al. (2023) and Chinchor (1992).

To assess the reliability of candidate gene identification, we calculated the F1 score, which combines precision (the proportion of predicted positives that are true positives) and recall (the proportion of true positives correctly identified) into a single metric. An F1 score of 1 indicates perfect classification, meaning that a gene's presence or absence accurately distinguishes all resistant and susceptible varieties.

We apply this to the dRenSeq table which provides candidate coverage across the panel of varieties. This approach allowed us to identify candidate genes that consistently associate with resistance, providing a robust and quantitative measure of their predictive value.

Phylogenetics

The phylogenetics workflow was as described by Wang et al. (2023). Briefly, NB domains were identified using InterProScan version 5.54–87.0 (Jones et al. 2014). The NB sequences were aligned with Clustal Omega version 1.2.4 (Sievers et al. 2011). Phylogenetic trees were constructed in R version 4.1.2 using the ape version 5.8–1 (Paradis and Schliep 2019) and phangorn version 2.12.1 (Schliep et al. 2008) packages, applying the JTT + G substitution model, identified as the most appropriate model through BIC score, and 1000 bootstrap replicates after alignment quality control. Clades were assigned using a 0.05 percentile threshold with PhyloPart version 2.1 (Prosperi et al. 2011), and the tree was visualised using IToL version 6.6 (Letunic and Bork 2021).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-026-05179-9.

Acknowledgements This work was supported by the Rural & Environment Science & Analytical Services (RESAS) Division of the Scottish Government through project JHI-B1-1, the Biotechnology and Biological Sciences Research Council (BBSRC) through award BB/X009068/1 and the Innovate UK Project 10058434: TRIP sustainable Agriculture. YWC was supported through the BBSRC PhD training program CTP-SAI award number BB/Y513684/1. LB was supported through the East of Scotland Bioscience Doctoral Training Partnership (EASTBIO DTP), funded by the BBSRC award BB/T00875X/1. JP, TA, MB, and IH were supported through the Scottish Plant Health Centre project 'PCN Action Scotland'. The authors acknowledge Research Computing at the James Hutton Institute for providing computational resources and technical support for the 'UK's Crop Diversity Bioinformatics HPC' (BBSRC grants BB/S019669/1 and BB/X019683/1), use of which has contributed to the results reported within this paper.

Author contributions YWC, LB, TMA, AK, and MS conducted the computational analyses. BH, GM, JO and VS provided material used in this study. YWC, TMA, MB, IH, JP and MS contributed towards writing of the manuscript. IH and MB concieved this work and IH secured funding. All authors have read and endorsed the manuscript.

Funding Rural and Environment Science and Analytical Services Division, JHI-B1-1, Biotechnology and Biological Sciences Research Council, BB/X009068/1, Ingo Hein, BB/Y513684/1, Ingo Hein,BB/T00875X/1, Ingo Hein, Innovate UK, 10058434: TRIP sustainable Agriculture, Ingo Hein

Data availability Raw reads are available at the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) under accession numbers PRJEB89657 and PRJEB56823. Assemblies are available under the accession numbers PRJEB90064 and PRJEB90070.

Code availability Scripts for ONT assembly of DH4_Athlete are available at Zenodo under URL https://doi.org/10.5281/zenodo.16418965.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Adams TM, Smith M, Wang Y et al (2023) HISS: snakemake-based workflows for performing SMRT-RenSeq assembly, AgRenSeq and dRenSeq for the discovery of novel plant disease resistance genes. BMC Bioinformatics 24:204. https://doi.org/10.1186/s12859-023-05335-8

Armstrong MR, Vossen J, Lim TY et al (2019) Tracking disease resistance deployment in potato breeding by enrichment sequencing. Plant Biotechnol J 17:540–549. https://doi.org/10.1111/pbi.12997

Arora S, Steuernagel B, Gaurav K et al (2019) Resistance gene cloning from a wild crop relative by sequence capture and association genetics. Nat Biotechnol 37:139–143. https://doi.org/10.1038/s41587-018-0007-9

Bairwa A, Dalamu DB et al (2024) Potato cyst nematode: resistance, management, and quarantine perspectives across the globe. approaches for potato crop improvement and stress management. Springer Nature Singapore, Singapore, pp 233–247

Bakker E, Achenbach U, Bakker J et al (2004) A high-resolution map of the H1 locus harbouring resistance to the potato cyst nematode *Globodera rostochiensis*. Theor Appl Genet 109:146–152. https://doi.org/10.1007/s00122-004-1606-z

Birch PRJ, Bryan G, Fenton B et al (2012) Crops that feed the world 8: potato: are the trends of increased global production sustainable? Food Secur 4:477–508. https://doi.org/10.1007/s12571-012-0220-1

Chen Q, Yang C, Zhang G, Wu D (2024) GCI: a continuity inspector for complete genome assembly. Bioinformatics 40(11):btae633. https://doi.org/10.1093/bioinformatics/btae633

Cheng H, Concepcion GT, Feng X et al (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods 18:170–175. https://doi.org/10.1038/s41592-020-01056-5

Chinchor N (1992) MUC-4 evaluation metrics. In: Proceedings of the 4th conference on Message understanding - MUC4 '92. Association for Computational Linguistics, Morristown, NJ, USA, pp 22

Ellenby C (1952) Resistance to the potato root eelworm, *Heterodera rostochiensis* Wollenweber. Nature 170:1016–1016. https://doi.org/10.1038/1701016a0

Evans K, Brodie BB (1980) The origin and distribution of the golden nematode and its potential in the U.S.A. Am Potato J 57:79–89. https://doi.org/10.1007/BF02853888

Finkers-Tomczak A, Bakker E, de Boer J et al (2011) Comparative sequence analysis of the potato cyst nematode resistance locus H1 reveals a major lack of co-linearity between three haplotypes in potato (*Solanum tuberosum* ssp.). Theor Appl Genet 122:595–608. https://doi.org/10.1007/s00122-010-1472-9

Food and Agriculture Organisation of the United Nations LR (2023) FAOSTAT. http://www.fao.org/nr/land/databasesinformation-systems/en/

Gartner U, Hein I, Brown LH et al (2021) Resisting potato cyst nematodes with resistance. Front Plant Sci. https://doi.org/10.3389/fpls.2021.661194

Griffin D, Bourke L, Mullins E et al (2022) Potatoes in Ireland: sixty years of potato research and development, market evolution and perspectives on future challenges. Irish J Agric Food Res 61(1):184–200. https://doi.org/10.15212/ijafr-2020-0144

Hermsen JGTh, Verdenius J (1973) Selection from *Solanum tuberosum* group phureja of genotypes combining high-frequency haploid induction with homozygosity for embryo-spot. Euphytica 22:244–259. https://doi.org/10.1007/BF00022632

Holst F, Bolger A, Günther C, et al (2023) Helixer–*de novo* Prediction of Primary Eukaryotic Gene Models Combining Deep Learning and a Hidden Markov Model

Hu K, Ni P, Xu M et al (2024) HiTE: a fast and accurate dynamic boundary adjustment approach for full-length transposable element detection and annotation. Nat Commun 15:5573. https://doi.org/10.1038/s41467-024-49912-8

Jain C, Rhie A, Hansen NF et al (2022) Long-read mapping to repetitive reference sequences using Winnowmap2. Nat Methods 19:705–710. https://doi.org/10.1038/s41592-022-01457-8

Jones P, Binns D, Chang H-Y et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Jupe F, Witek K, Verweij W et al (2013) Resistance gene enrichment sequencing (RenSeq) 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. https://doi.org/10.1111/tpj.12307

Kokot M, Długosz M, Deorowicz S (2017) KMC 3: counting and manipulating k -mer statistics. Bioinformatics 33:2759–2761. https://doi.org/10.1093/bioinformatics/btx304

Kuśmirek W (2023) Estimated nucleotide reconstruction quality symbols of basecalling tools for Oxford Nanopore sequencing. Sensors 23:6787. https://doi.org/10.3390/s23156787

Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li H (2023) Protein-to-genome alignment with miniprot. Bioinformatics 39(1):btad014. https://doi.org/10.1093/bioinformatics/btad014

Madeo G, Savojardo C, Manfredi M et al (2023) CoCoNat: a novel method based on deep learning for coiled-coil prediction. Bioinformatics. https://doi.org/10.1093/bioinformatics/btad495

Manni M, Berkeley MR, Seppey M et al (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol 38:4647–4654. https://doi.org/10.1093/molbev/msab199

Martin EC, Spiridon L, Goverse A, Petrescu A-J (2022) NLRexpress—a bundle of machine learning motif predictors—reveals motif stability underlying plant nod-like receptors diversity. Front Plant Sci 13:975888. https://doi.org/10.3389/fpls.2022.975888

Paradis E, Schliep K (2019) Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528. https://doi.org/10.1093/bioinformatics/bty633

Percival-Alwyn L, Barnes I, Clark MD et al (2025) UKCropDiversity-HPC: a collaborative high-performance computing resource approach for sustainable agriculture and biodiversity conservation.

Plants People Planet 7:969–977. https://doi.org/10.1002/ppp3.10607

Price JA, Coyne D, Blok VC, Jones JT (2021) Potato cyst nematodes *Globodera rostochiensis* and *G. pallida*. Mol Plant Pathol 22:495–507. https://doi.org/10.1111/mpp.13047

Prosperi MCF, Ciccozzi M, Fanti I et al (2011) A novel methodology for large-scale phylogeny partition. Nat Commun 2:321. https://doi.org/10.1038/ncomms1325

Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 11:1432. https://doi.org/10.1038/s41467-020-14998-3

Rhie A, Walenz BP, Koren S, Phillippy AM (2020) Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol 21:245. https://doi.org/10.1186/s13059-020-02134-9

Rossi M, Goggin FL, Milligan SB et al (1998) The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. Proc Natl Acad Sci USA 95:9750–9754. https://doi.org/10.1073/pnas.95.17.9750

Savary S, Willocquet L, Pethybridge SJ et al (2019) The global burden of pathogens and pests on major food crops. Nat Ecol Evol 3:430–439. https://doi.org/10.1038/s41559-018-0793-y

Schliep K, Paradis E, de Oliveira Martins L, et al (2008) phangorn: phylogenetic reconstruction and analysis. CRAN: Contributed Packages

Schultz L, Cogan NOI, McLean K et al (2012) Evaluation and implementation of a potential diagnostic molecular marker for *H1*-conferred potato cyst nematode resistance in potato (*Solanum tuberosum* L.). Plant Breed 131:315–321. https://doi.org/10.1111/j.1439-0523.2012.01949.x

Sievers F, Wilm A, Dineen D et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75

Smith M, Jones JT, Hein I (2025) Resistify: a novel NLR classifier that reveals helitron-associated NLR expansion in Solanaceae. Bioinform Biol Insights. https://doi.org/10.1177/11779322241308944

Spychalla P, De Jong WS (2024) Breeding for potato cyst nematode resistance in *Solanum tuberosum*. Crop Sci 64:1167–1182. https://doi.org/10.1002/csc2.21244

The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195. https://doi.org/10.1038/nature10158

van der Vossen EAG, Gros J, Sikkema A et al (2005) 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. https://doi.org/10.1111/j.1365-313X.2005.02527.x

Wainer J, Dinh Q (2021) Taxonomy, morphological and molecular identification of the potato cyst nematodes, *Globodera pallida* and *G. rostochiensis*. Plants 10:184. https://doi.org/10.3390/plants10010184

Wang Y, Brown LH, Adams TM et al (2023) SMRT–AgRenSeq-d in potato (*Solanum tuberosum*) as a method to identify candidates for the nematode resistance Gpa5. Hortic Res. https://doi.org/10.1093/hr/uhad211

Van Weymers PSM, Baker K, Chen X et al (2016) Utilizing "Omic" technologies to identify and prioritize novel sources of resistance to the oomycete pathogen phytophthora infestans in potato germplasm collections. Front Plant Sci. https://doi.org/10.3389/fpls.2016.00672

Wijesinha-Bettoni R, Mouillé B (2019) The contribution of potatoes to global food security, nutrition and healthy diets. Am J Potato Res 96:139–149. https://doi.org/10.1007/s12230-018-09697-1

Witek K, Witek K, Jupe F et al (2016) SMRT renseq protocol. Protoc Exch. https://doi.org/10.1038/protex.2016.027