G3, 2026, 16(6), jkag089
https://doi.org/10.1093/g3journal/jkag089
Advance Access Publication Date: 3 April 2026
Plant Genetics and Genomics

An improved *Solanum verrucosum* genome provides insight into potato centromeres and epigenetic regulation

Moray Smith ,[1,2,3] Amanpreet Kaur,[1,3,4] Vikrant Singh,[1] John T. Jones,[1,2] Ingo Hein [1,3,*]

[1]Cell and Molecular Sciences Department, The James Hutton Institute, Invergowrie, Dundee DD2 5DA, United Kingdom
[2]School of Biology, University of St Andrews, St Andrews KY16 9ST, United Kingdom
[3]Division of Plant Sciences, University of Dundee, Dundee DD1 4HQ, United Kingdom
[4]International Potato Center, Avenida La Molina 1895, La Molina Apartado 1558, Lima 12, Peru

*Corresponding author: Email: ingo.hein@hutton.ac.uk

Solanum verrucosum is a wild diploid potato species notable for carrying the uncharacterized *Rpi-ver1* gene conferring resistance to *Phytophthora infestans*, possessing unusually structured centromeres, and exhibiting the rare trait of self-compatibility, the genetic basis of which remains unknown. Here, we present a high-quality genome assembly and annotation of *S. verrucosum* accession CPC54. A comprehensive inventory of resistance genes is provided, and candidate genes for *Rpi-ver1* are identified through remapping approaches. The sequence and epigenetic landscape of centromeres are resolved, revealing a mosaic of repeatless and repetitive domains. These centromeres consist of tandem arrays of kilobase-sized repeats, some of which show signatures of transposable element origin. Specific subfamilies of CRM and Tekay elements appear to have adapted to centromeric regions. The basis of self-compatibility in *S. verrucosum* is investigated with respect to the *S-RNase* gene, which is found to be present and intact but expressed at very low levels. Notably, upstream insertions of transposable elements are present that may interfere with its expression.

Keywords: *Solanum verrucosum*; genome assembly; centromere; methylation; self-compatibility; disease resistance

Introduction

Solanum verrucosum is a wild diploid potato species native to Mexico. It has been widely studied as it contains valuable sources of disease resistance and can act as a bridge species, enabling crosses between species with different endosperm balance number values. In addition, it is self-compatible, unlike most tuber-bearing species of *Solanum*, and has unusual centromeres that are in a mixed state of being repeatless or formed from kilobase-sized repeats (Eijlander et al. 2000; Gong et al. 2012).

Late blight is a serious disease caused by the oomycete pathogen *Phytophthora infestans*, which has impacted potato cultivation and breeding for over 170 years (Ivanov et al. 2021). A novel, broad-spectrum dominant late blight resistance gene, *Rpi-ver1*, has been identified on chromosome 9 of this species (Liu and Halterman 2006; Chen et al. 2018). Resistance to plant viruses and tomato psyllid has also been reported in *S. verrucosum*, while 1 accession also showed resistance to a broad range of *G. pallida* populations (Carrasco et al. 2000; Castelli et al. 2005; Cooper and Bamberg 2016).

Beyond disease resistance, the unique genomic architecture of *S. verrucosum*, particularly its centromeres, presents an area of research interest. In *Arabidopsis*, centromeres are composed of megabase satellite arrays with 178 bp subunits, which are organized into higher-order repeats, a result of unequal crossover and recombination (Naish et al. 2021). Underpinning this complexity is the cyclic expansion and contraction of ATHILA LTR transposable elements in the centromere, driving evolution and speciation within *Arabidopsis* (Wlodzimierz et al. 2023a, 2023b). In contrast to *Arabidopsis*, the centromeres of potato and *S. verrucosum* are known to be partially repeatless and, where repetitive, are formed of kilobase-sized repeat subunits (Gong et al. 2012; Zhang et al. 2014). These large repeat monomers share sequence similarity with LTR retrotransposons, indicating their role in centromere evolution in *Solanum* (Gong et al. 2012). Another feature of *S. verrucosum* that makes resolving its genome of interest is its self-compatibility and lack of interspecific reproductive barriers (Behling and Douches 2023). This has previously been attributed to the lack of a functional *S-RNase*, although the genetic basis of this remains unknown (Eijlander et al. 2000).

These properties have led to *S. verrucosum* being a focus for multiple genomic studies—in 2019, an assembly of the dihaploid clone 54 was produced through a combination of short-read, long-read, and scaffolding approaches to produce an assembly with a contig N50 of 858 kbp (Paajanen et al. 2019). In 2022, an assembly of the monohaploid *S. verrucosum* clone 11H23 was produced with a contig N50 of 21 Mbp through PacBio HiFi sequencing and Hi-C scaffolding (Hosaka et al. 2022).

Here, we report an improved assembly of the dihaploid *S. verrucosum* accession 54, alongside a comprehensive gene, transposable element, and DNA methylation annotation. Through this assembly, we resolve the complete sequence of the *Rpi-ver1* locus, which contains a single pseudogenized NLR. We also investigate the sequence and epigenetic status of the *S. verrucosum* centromeres, which are revealed to be in a mix of repeatless and repetitive states—the repetitive centromeres are megabase-scaled and formed of large tandem array subunits. Finally, we investigate

Received on 30 October 2025; accepted on 17 February 2026
© The Author(s) 2026. Published by Oxford University Press on behalf of The Genetics Society of America.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

the status of the *S-RNase* gene in *S. verrucosum*, which is revealed to be densely methylated and likely silenced through transposable element activity in its promoter region.

Methods

DNA extraction, sequencing, and assembly

High-molecular-weight (HMW) DNA was extracted from 72-h dark-treated leaf tissue with the Macherey-Nagel NucleoBond HMW DNA kit following the manufacturer's specifications.

PacBio sequencing was conducted by the Norwegian Sequencing Centre using the SMRTbell ExpressTemplate Prep Kit 3.0 and sequencing on 1 8M SMRT cell on the Sequel II platform using the Sequel II Binding kit 2.2 and Sequencing chemistry v2.0. HiFi reads were generated using the CCS pipeline (SMRT Tools v11.0.0.146107).

For Oxford Nanopore Sequencing (ONT) sequencing, HMW DNA was size-selected using a BluePippin 0.75% High Pass cassette with a > 15 kbp size selection cutoff and libraries prepared with the SQK-LSK109 Ligation Sequencing kit following manufacturer's specifications. ONT libraries were sequenced on R9.4.1 minION flow cells using 50 to 75 fmol library DNA as input. ONT sequences were basecalled with guppy v4.2.2 using the model dna_r9.4.1_450bps_sup_plant.cfg.

For Hi-C sequencing, libraries were prepared according to the Dovetail® Omni-C® kit protocol. Two independent Hi-C libraries were prepared and sequenced using a NextSeq P2 kit with paired-end 150 bp reads.

To assess the homozygosity of the genome, 21-mers were counted, and a histogram was generated from HiFi reads with kmc v3.2.4 using the options -k21 -t10 -m64 -ci1 -cs10000 and kmc_tools transform histogram -cx10000 (Kokot et al. 2017). The histogram was used as input for genomescope2 v2.0.1 using the default options (Ranallo-Benavidez et al. 2020).

A primary genome assembly was produced from the HiFi reads with hifiasm v0.24.0 using the option –primary (Cheng et al. 2021). A secondary assembly was produced from the ONT reads with flye v2.9.5 using the options genome size = 750 m -nanopore (Kolmogorov et al. 2019).

Mitochondrial and chloroplast assemblies were produced from HiFi reads with oatk v1.0 using the option -c 100 and OatkDB v202030921 (Zhou et al. 2025). The chloroplast assembly was annotated with GeSeq (Tillich et al. 2017).

To assess genome contiguity and assembly quality, GCI v1.0 was run using the HiFi and ONT reads, which were simultaneously mapped with minimap2 2.28 and winnowmap v2.03 with the default settings (Li 2018; Jain et al. 2020; Chen et al. 2024a). BUSCO v6.0.0 was applied to both the genome and protein predictions using the Solanaleles OrthoDB.v12 dataset (Tegenfeldt et al. 2025).

Hi-C scaffolding

Hi-C reads were aligned to the assembly with BWA v0.7.18 using the options –5SP -T0 and filtered, sorted, deduplicated, and split with pairtools v1.10 using the options –min-mapq 40 –walks-policy 5unique –max-inter-align-gap 3 (Li 2013; Open2C et al. 2024). Alignments were used to scaffold the genome with yahs v1.2a.2 using the option –no-contig-ec (Zhou et al. 2023). The scaffolded genome was prepared for manual curation in the Juicer GUI with the yahs juicer pre module and juicertools (Durand et al. 2016). Minor corrections for chromosome fusions were made, and a final assembly was generated.

To label chromosomes, the final assembly was aligned to the DM1-3 516 R44 (v6.1) genome using mashmap v3.0.6 with default settings (Pham et al. 2020; Kille et al. 2023). The alignment was visualized with the D-GENIES web tool, and chromosome identities were inferred (Cabanettes and Klopp 2018).

RNA-seq analysis

Tissue culture plants of *S. verrucosum* were maintained on MS20 medium and kept in a growth room at a light intensity of 110 µmol $m^{-2}s^{-1}$, a temperature of 18 ± 2 °C, and a photoperiod of 16/8 h light/dark. Healthy 3-wk-old plantlets with fully expanded leaves were selected. In vitro shoots with roots were gently removed from the media and dipped for 1 min in a zoospore suspension of *P. infestans* isolate W9928C adjusted to 4 × 106 spores/mL. Dip-inoculated microshoots were blotted on sterile paper towels and then planted in fresh MS20 medium in vented containers (Generon, UK). The infected plants were kept in darkness for 16 h and then incubated under the growth conditions mentioned above. The disease severity was recorded by counting the number of leaves showing disease symptoms in 24-h intervals. The leaf samples from 3 independent replicates were collected after 0- and 24-h post infection and immediately immersed in liquid nitrogen before storage at −70 °C for further processing.

From each replicate, leaf samples were crushed to a fine powder, and 400 mg of ground sample was resuspended in 2 mL of TRIzol reagent and vortexed after the addition of 10 µL β-mercaptoethanol. The slurry was left to stand at room temperature for 5 min before centrifugation (10,000 *g*, 10 min at 4 °C). To the supernatant, 0.2 mL chloroform was added (per 1 mL) and incubated at room temperature for 5 min before centrifugation (10,000 *g*, 10 min at 4 °C). The aqueous layer was transferred. 0.5 mL isopropanol was added and transferred to a QIA RNAeasy spin column for washing in RPE buffer twice. RNA was eluted in 50 µL RNase-free water, and the integrity was checked with a Bioanalyzer 2100 (Agilent).

For RNA-seq, samples were checked for a RIN value >8 and processed at the James Hutton Institute's Genomics facility for generating RNA sequencing libraries using the standard Illumina mRNA Prep kit RNA unique dual UD Indices as recommended, with 100 ng total RNA per sample. Libraries were checked on a Qubit fluorimeter and Bioanalyzer 2100 prior to pooling equimolar amounts before sequencing. Sequencing was conducted on a NextSeq 2000 at a loading concentration of 750pM using a P3 200 kit, generating paired-end 100 bp reads.

RNA-seq mapping and read count estimation were carried out using nf-core/rnaseq v3.12.0, using the default settings. The salmon quantified read counts were imported into R with tximport v3.19, and all subsequent analysis was conducted with deseq2 v3.19. Infection, tissue-specific, and temperature response assays were analyzed in separate RNA-seq experiments. Differential expression analysis was conducted with DESeq() using the default settings, followed by lfcShrink() with the apeglm shrinkage estimator. Genes were labelled as differentially expressed if $p_{adj} < 0.01$ and $|\log_2(FC)| > 1$.

Genome annotation

Transposable element annotations were generated with Earl Grey v4.2.4 using the default settings (Baril et al. 2024). To classify LTR elements into clades, libraries were provided to TEsorter v1.4.6 using the rexdb-plant database (Zhang et al. 2022).

Tandem repeats were identified with TRASH with the options –win 10000 –m 9000 (Wlodzimierz et al. 2023a). Repeats were classified by their homology to the Earl Grey TE library—a FASTA of repeats was extracted from the TRASH.bed file, a nonredundant library was created with seqkit rmdup, and this was used as a

query against the TE library (Shen et al. 2016). Subject hits with the highest bit score were used to classify repeats according to their LTR clade, and manual validation was carried out to ensure no misclassification occurred. Repeat classifications were merged with the original TRASH.bed file to produce an informative.bed file for use in subsequent analysis.

Gene models were predicted using BRAKER3 v3.0.7 using RNA-seq and the Viridiplantae OrthoDB v.11 database (Kuznetsov et al. 2023; Gabriel et al. 2024). All RNA-seq data were aligned to the masked genome produced by Earl Grey with STAR v2.7.10 (Dobin et al. 2013). To identify NLRs that were not predicted by BRAKER3, Helixer v0.3.0 was applied to the unmasked genome with the model land_plant_v0.3_a_0080.h5 (Holst et al. 2025). Annotations that were identified as NLRs by Resistify and did not overlap with existing BRAKER3 annotations were merged into the final annotation with AGAT.

Resistify v1.1.5 was run with subcommand nlr –retain –coconat to identify complete and partial NLRs with improved CC annotation via CoCoNat, and subcommand prr to identify RLK/RLPs, which uses TMbed to identify transmembrane domains (Bernhofer and Rost 2022; Martin et al. 2022; Madeo et al. 2023; Smith et al. 2025). A phylogenetic tree of NLRs was built from a multiple sequence alignment of NB-ARC domains produced by MAFFT v7.525 with option –auto (Katoh et al. 2002).

Nanopore DNA methylation analysis was carried out using deepsignal-plant v0.1.6 (Ni et al. 2021). Basecalled fast5 files were re-squiggled with tombo v1.5.1 and extracted with deepsignal_plant extract (Stoiber et al. 2017).

Rpi-ver1 mapping

To identify the *Rpi-ver1* locus, previously generated KASP markers were mapped to the genome with BLASTn and filtered for hits with 100% identity and query length (Chen et al. 2018). The *Rpi-ver1* locus was determined to be the locus delimited by the high-confidence KASP markers. To verify that the locus fully represented *Rpi-ver1*, bulk segregant RenSeq reads from the original study were mapped to the *S. verrucosum* genome and filtered with the expected homozygosity for the parent datasets and heterozygosity for the F1 progeny bulks.

Centromere analysis

Existing CENH3 ChIP reads for *S. verrucosum* (SRR18548893) were aligned to the genome with bowtie2 v2.5.3 (Langmead and Salzberg 2012). Centromere coordinates were determined manually by inspecting each chromosome for CENH3 read peaks. To identify transposable elements with a bias for presence in the centromere, Fisher's exact tests were conducted for each family identified by Earl Grey, using inside/outside centromere counts with a *P*-value cutoff of 0.001. Only intact transposable element annotations were counted, and only families that could be confidently classified by TEsorter were investigated further. A phylogenetic tree of reverse transcriptase domains was built using iqtree -bb 1000 -nt AUTO from a domain alignment produced by TEsorter. The repeat similarity heatmap between *S. verrucosum* and *S. tuberosum* centromeres was generated with StainedGlass v0.6.

Characterization of the S-locus

To identify the homolog of *S-RNase* in *S. verrucosum* 54, known *S-RNase* sequences from *S. neorickii* and *S. chilense* (BAC00940.1 and BAC00934.1) were queried against *S. verrucosum* with BLASTp v2.9.0. The *S-RNase* homolog was determined by its higher percentage identity relative to non-S-locus RNases and its expected location on chromosome 1. The gene model predicted by BRAKER3 was cross-validated against the Helixer annotation and found to be identical.

The conserved sequence was used to design *S-RNase*-specific primers (fwd: CTGGCCTCAACTCAGATACGA, rev: ATTCCATGATT TCTAAGAGTTCCC) for gene quantification. RNA was extracted from flowers of *S. verrucosum* 54 using the previously described methodology, and cDNA was synthesized using the SuperScript IV first strand cDNA synthesis kit (Invitrogen) following the manufacturer's guidelines. The housekeeping gene exocyst complex component (*Sec3*) was amplified using primers (fwd: GCTTG CACACGCCATATCAAT, rev: TGGATTTTACCACCTTCCGCA) for gene quantification. Relative gene expression was calculated using the formula $\Delta Ct = Ct_{SRNase} - Ct_{Sec3}$.

Data analysis

Where not otherwise specified, analysis was conducted using R v4.5.0—scripts and packages were controlled by renv and are available at https://www.github.com/swiftseal/sver54_assembly. The bioinformatics and computational analyses were performed on Crop Diversity HPC, described by Percival-Alwyn et al. (2024).

Results

A comprehensive genome assembly of *S. verrucosum*

To generate a high-quality assembly of *S. verrucosum*, a combination of long-read sequencing methods was selected. A total of 39.5 Gbp of HiFi reads with an N50 of 14.8 kbp and an average quality of 28.7 were generated, producing a total assembly size of 773.8 Mbp across 2,129 contigs with a contig N50 of 46.3 Mbp. Simultaneously, a total of 39.7 Gbp of ONT reads were generated with an N50 of 14.3 kbp and an average quality of 13.63, producing a total assembly size of 686.4 Mbp across 1651 contigs with an N50 of 5.3 Mbp. As the HiFi assembly produced an assembly with a 10-fold higher contiguity, this was taken forward for scaffolding with Hi-C sequencing. A total of 502 million read pairs were generated, of which 293 million were mapped, and 21 million were marked as non-duplicate and valid contact events. Scaffolding from these produced a 696 Mbp chromosome-scale assembly composed of 12 scaffolds, 19 contigs, and a contig N50 of 46.3 Mbp (Fig. 1a).

Many of the assembled chromosomes could be described as "telomere-to-telomere"; however, this claim can be over-inflated by assembly artifacts and misassemblies (Chen et al. 2024a). To further assess the quality of the final genome assembly, the Genome Continuity Inspector (GCI) was used to evaluate contiguity and identify potential misassemblies. The complete assembly had a GCI score of 60.8, with 6/12 chromosomes being represented by a single contig, while the remaining were limited to 2/3 (Table 1). Only a single locus on chromosome 9 was noted for a significant drop in read across both HiFi and ONT reads, indicating a potential misassembly (Supplementary Fig. 1).

As expected, the genome is diploid and largely homozygous (Fig. 1c). A high BUSCO completeness (C: 99.7% [S: 99.2%, D: 0.6%], F: 0.1%, M: 0.2%, n: 7934, E: 1.1%) also demonstrated a high-quality assembly. Overall, this assembly represents a marked improvement over the monohaploid *S. verrucosum* assembly, which is composed of 12 scaffolds, 665 contigs, and a contig N50 of 20.9 Mbp, and the previous assembly of *S. verrucosum* clone 54 with a contig N50 of 0.86 Mbp (Paajanen et al. 2019; et al. 2022).

For improved gene annotation, RNA-seq was conducted across a variety of tissue levels (root, shoot, and leaf), biotic (infected with *P. infestans*), and abiotic stress conditions (heat and cold) (Table 2).

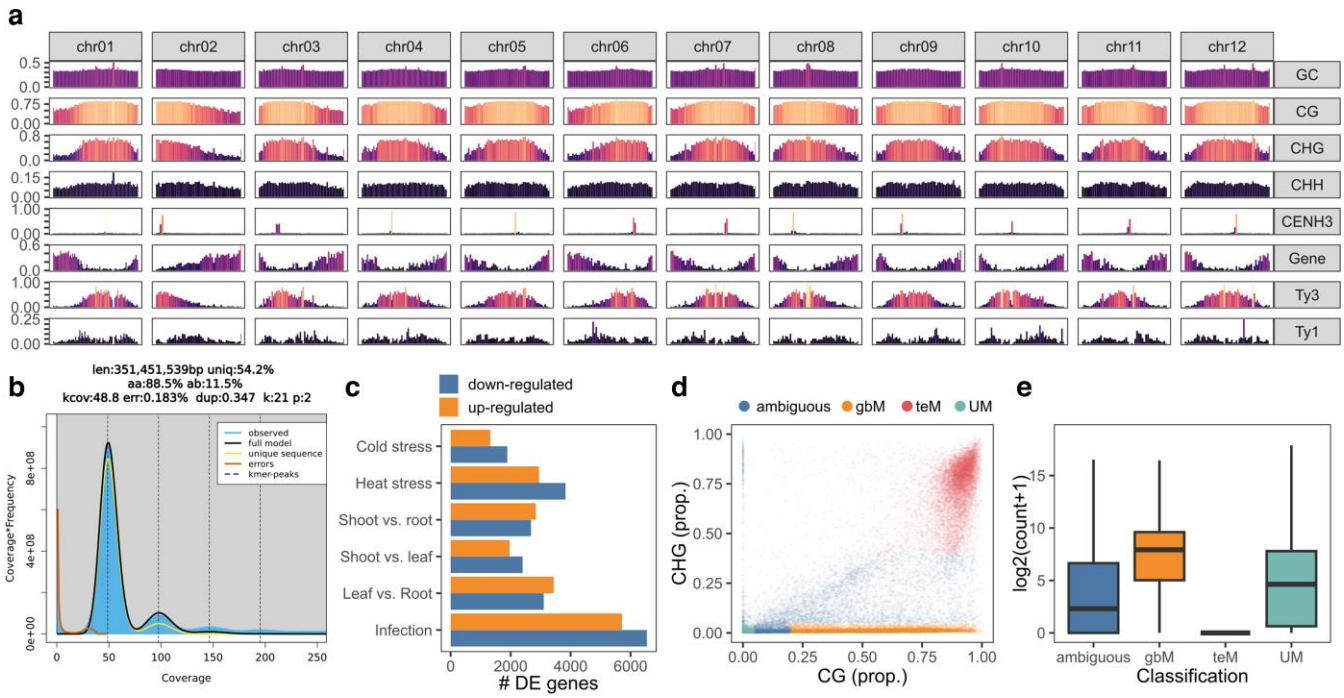

Fig. 1. a) The density of GC%, CG, CHG, and CHH methylation, CENH3 read mapping, genes, and Ty1 and Ty3 LTRs across the genome assembly in 1 Mbp windows. Coloured according to proportion. b) The GenomeScope2 profile of 21-mers extracted from HiFi reads Coverage is defined as k-mer coverage per homolog. c) The number of differentially expressed genes identified across RNA-seq conditions. d) The distribution of genes by their CG and CHG exon methylation, colored by their classification. e) The leaf RNA expression of different methylation classifications.

Table 1. The primary results of genome continuity inspector (GCI) using both HiFi and ONT reads.

Chromosome	Theoretical maximum N50	Curated N50	Curated contigs number	GCI score
chr01	84.4 Mbp	59.8 Mbp	2	48.7
chr02	46.3 Mbp	46.3 Mbp	1	100
chr03	60.3 Mbp	60.3 Mbp	1	100
chr04	66.6 Mbp	43.5 Mbp	2	45.7
chr05	51.1 Mbp	51.1 Mbp	1	100
chr06	54.6 Mbp	54.6 Mbp	1	100
chr07	53.3 Mbp	53.3 Mbp	1	100
chr08	56.8 Mbp	56.8 Mbp	1	100
chr09	59.7 Mbp	36.7 Mbp	3	34.6
chr10	57.8 Mbp	34.9 Mbp	3	34.1
chr11	46.9 Mbp	25.8 Mbp	2	39.9
chr12	56.7 Mbp	32.5 Mbp	2	41.2
Genome	56.8 Mbp	46.3 Mbp	20	60.8

Genes were annotated with a combination of BRAKER3 and Helixer, identifying 38,710 genes with a mean gene length of 3043 bp, 1.2 transcripts per gene, 4.3 exons per transcript, and a multi-to-single exon ratio of 2.75. The BUSCO completeness of the annotation was also high (C: 98.6% [S: 82.8%, D: 15.8%], F: 0.4%, M: 1.0%, n: 7934). RNA-seq analysis 24 h post infection of *P. infestans* isolate W9928C revealed higher gene down-regulation as compared to up-regulation. Similarly, during cold and heat stress, more genes were downregulated. Differential expression of genes was also observed to be organ-specific (Fig. 1c).

Two circular mitochondrial genomes were assembled, which were 417.9 and 49.3 kbp in size, respectively. Multiple mitochondrial subgenomes have been observed previously in potato (Cho et al. 2022) A single circular chloroplast genome of 155.5 kbp

Table 2. RNAseq libraries prepared for *Solanum verrucosum*.

Condition	Rep	Total reads	Duplicates (%)	Uniquely mapping (%)
Infection 0hpi	1	53,925,964	36	90
Infection 0hpi	2	62,270,652	38	89
Infection 0hpi	3	68,241,131	35	90
Infection 24hpi	1	59,017,589	41	86
Infection 24hpi	2	56,916,433	39	88
Infection 24hpi	3	65,742,064	40	89
Leaf	1	55,457,446	35	86
Leaf	2	55,866,776	36	88
Leaf	3	56,785,668	37	90
Root	1	54,483,222	35	90
Root	2	61,491,956	38	88
Root	3	61,258,248	35	92
Shoot	1	53,692,166	34	90
Shoot	2	60,784,503	36	89
Shoot	3	55,940,978	35	92
25 °C	1	61,537,559	39	87
25 °C	2	58,842,187	37	87
25 °C	3	55,652,255	35	92
35 °C	1	56,643,303	35	86
35 °C	2	59,834,457	32	88
35 °C	3	62,788,765	35	89
4 °C	1	58,432,736	38	88
4 °C	2	61,348,502	39	88
4 °C	3	57,388,920	36	90

For infection conditions, in vitro roots and shoots were treated with *P. infestans* isolate W9928C, followed by RNA extraction at the specified time point from whole plantlet tissue. For temperature conditions, RNA was extracted from leaf tissue.

was also assembled (Supplementary Fig. 2). This was in line with a recent 155.5 kbp assembly of the *S. verrucosum* chloroplast, which was contrastingly lacking in a ycf1 annotation at the IRB-SSC boundary (Zhang et al. 2024).

DNA methylation is an important characteristic of genes and transposable function, and DNA methylation was determined from the available ONT reads. The average level of methylation in the genome was 70%, 46.2%, and 9.4% in the CG, CHG, and CHH contexts, respectively—in line with previous estimates for *S. lycopersicum* and *S. melogena* (Cui et al. 2021; Lu et al. 2021). A clear trend of heightened methylation towards the centromeres was evident for all chromosomes, across all methylation contexts (Fig. 1a).

In plants, genes can be classified according to the levels of CG and CHG methylation in their exon body (Zeng et al. 2023). Accordingly, they can be classified as unmethylated (UM) if unmethylated in either context, or as having gene body methylation (gbM) if methylated in the CG context only, or as having TE-like methylation (teM) if methylated in both the CG and CHG contexts. In maize, it has been previously observed that gbM are generally highly expressed across all tissues, that UM genes are more frequently tissue-specific, and that teM genes are transcriptionally silent (Zeng et al. 2023). To determine if a similar effect is observed in *Solanum*, the same classification criteria were applied to the *S. verrucosum* genome.

Accordingly, 5,940 genes were classified as UM, 17,151 as gbM, 8,076 as teM, and 7,957 as ambiguous (Fig. 1d). Although the maize genome has a similar number of genes (~39,000), more genes were classified as being gbM (17,151 vs 8,134) and teM (8,076 vs 3,402) in *S. verrucosum*. In line with the previous investigation, genes classified as teM had substantially lower expression than those classified as gbM (Fig. 1d). No evidence of a tissue-specific expression bias was found for UM genes, likely because far fewer tissue-specific expression datasets were available for this study.

To create a high-quality transposable element annotation, the performance of TE annotation tools EDTA and Earl Grey was compared (Supplementary Fig. 3). Earl Grey classified 67.6% of the genome as being repetitive, whereas EDTA classified 56.7%. As expected, the largest proportion of repeats was classified as being LTR-derived, with both pipelines identifying similar proportions of 39.7 and 36.6%. Earl Grey identified significantly fewer LTR families ($n = 632$) than EDTA ($n = 1975$), which were also of greater mean length. A similar trend was also seen for DNA, Helitron, and SINE elements, but not LINEs, for which Earl Grey identified a larger number of families.

To further assess the completeness of the TE libraries assembled by both tools, TEsorter was applied to classify TEs by their domains. The fraction of LTRs that failed to be classified by TEsorter was 63.1 and 9.8% for EDTA and Earl Grey, respectively. TEsorter further classified 38.6% ($n = 252$) of LTRs as being complete in the Earl Grey library but only 15.5% ($n = 306$) in the EDTA library.

The smaller number of TE families, greater mean length of LTR families, and higher rate of successful LTR classification by TEsorter were taken as evidence of Earl Grey outperforming EDTA at producing a high-quality TE library. Therefore, all subsequent analysis was undertaken using the Earl Grey library.

In the genetic history of *S. verrucosum*, 2 potential bursts of LTR insertion are noted, indicated by the 2 peaks in the repeat landscape (Supplementary Fig. 4). The most recent burst also appears to have been associated with DNA type transposable elements. Interestingly, upon closer examination of the LTR history, it appears that the recent activity was associated with Tekay, Athila, and CRM Ty3 LTRs, whereas the Ty1 Clades TAR, Ikeros, and Bianca contributed to a peak of activity approximately between the 2 transposable element bursts.

Together, the curation of a high-quality assembly, DNA methylation dataset, and gene and transposable element annotations provides a valuable resource for investigations into the *S. verrucosum* genome. Below, studies into the genomic basis of disease resistance, centromere organization, and self-compatibility are described.

Taking inventory of resistance genes

Nucleotide-binding leucine-rich repeat genes (NLRs) frequently function as the sensors of pathogen-derived signals or their function, and as such are a vital component of the plant's capacity to elicit an immune response to disease. In *S. verrucosum*, a total of 490 NLRs were identified (Fig. 2a). Of these, 238 were classified as canonical coiled-coil NLRs (CNLs), 44 as toll-interleukin receptor NLRs (TNLs), and 2 as RPW8 NLRs (RNLs). Of the TNLs, 28 had C-JID domains, and 41 of the CNLs had MADA motifs (Fig. 2b). The majority of LRR domains were ~350 aa in length, although several NLRs with LRRs longer than 1,000 aa were noted (Fig. 2c).

In the process of annotating the genome, it was noted that the gene predictor Helixer identified substantially more NLRs than BRAKER3 alone. In total, 163 NLRs were identified solely by Helixer, indicating that a large fraction of NLRs evade BRAKER3 annotation. Helixer-identified NLRs had significantly lower expression ($\beta = -1.97 \log_2(\text{count} + 1)$, $SE = 0.33$, $P = 7.55e^{-09}$), suggesting that they could function as valid NLRs, and that their evasion of BRAKER3 is possibly expression-led (Fig. 2d).

Indeed, of the identified NLRs, 23.1% exhibited low to no expression across all RNA-seq conditions explored in this study (Fig. 2e). When NLRs were clustered based on their expression patterns across RNA-seq conditions, root tissue emerged as a clear outlier for a subset of NLRs, consistent with previous findings (von Dahlen et al. 2023). However, previously, it has been reported that a subset of NLRs—including *Hero* and *NRC6*—have root-specific expression in *Solanum* (Lüdke et al. 2025). While 2 *NRC6* homologs exist in *S. verrucosum* and root-specific NLR clusters were visible, no NRCs exhibited root-specific expression and instead spanned 3 separate clusters of varying NLR expression. Whether this discrepancy is dependent on tissue sampling, developmental stage, or *Solanum* species remains to be clarified. It should also be noted that in the previous study, only 1 copy of *NRC6* was identified. The second copy identified in this study was identified solely through Helixer, demonstrating the importance of capturing a complete NLR inventory prior to downstream analysis.

In addition to the NLRs, a total of 705 RLKs and 517 RLPs were identified. Of these, extracellular LRR domains were the most common (35%, $n = 423$), followed by G-LecRLK (11%, $n = 131$) and WAK (6%, $n = 71$) domains. An additional 431 RLK/RLPs had extracellular domains that did not fit into the classification schema outlined in the RLKdb (Yin et al. 2024).

Previous mapping experiments in *S. verrucosum* isolated the novel disease resistance gene *Rpi-ver1* to a 4.3 Mbp locus on chromosome 9 in the DM reference genome. Re-analysis of existing mapping data—in particular of the markers DMG400017237 and DMG400017146, which demark this locus—reduced this to a locus 1.4 Mbp in size, from 54.6 to 56.0 Mbp. Within the locus are 2 adjacent, NLR-like genes that appear to be a canonical CNL with a frameshift mutation in the NB-ARC domain. In addition, there is an Rx-CC Jacalin-like lectin domain protein and a cysteine-rich kinase RLP, both of which have been previously implicated in

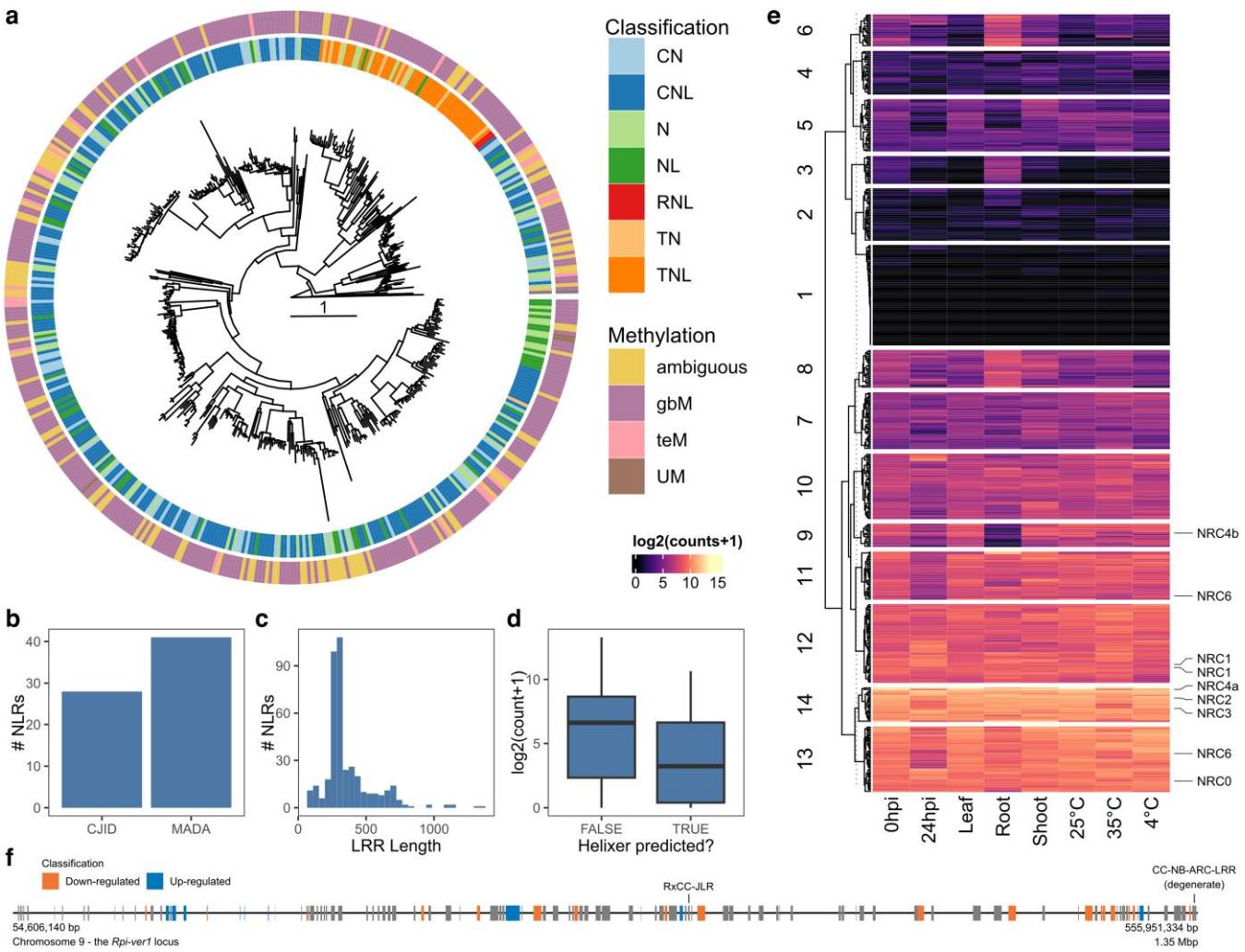

Fig. 2. a) Phylogenetic tree of NLR NB-ARC domains identified in this study, rooted to CED-4 from C. elegans. b) Number of NLRs with C-JID and MADA motifs identified. c) Histogram of LRR lengths. d) Leaf expression of NLRs identified by BRAKER3 or by Helixer. e) Clustered heatmap of NLR expression across conditions explored in this study. The position of NRCs has been highlighted. f) Representative gene models across the *Rpi-ver1* locus. Significantly differentially expressed genes under infection conditions have been highlighted. The position of putative resistance genes has been marked.

disease resistance in plants (Esch and Schaffrath 2017; Kamel et al. 2023).

The sequence and epigenetic status of centromeres

To identify putative centromeric regions, previously generated CENH3 ChIP reads were realigned to the *S. verrucosum* genome (Zhang et al. 2014). In total, 98% of the reads aligned to the primary chromosome assembly, indicating that the majority of centromere sequence had been successfully assembled, and improvement on a previous assembly where 17.2% of reads mapped to unanchored contigs (Hosaka et al. 2022). Clear peaks of CENH3 ChIP read mapping were present on each chromosome (Fig. 1a) and accordingly, centromere regions were identified (Fig. 3a).

Previous studies have shown that the centromeres of potato and *S. verrucosum* are either repeatless or formed of tandem repeat arrays with unusually large kilobase-sized subunits. As conventional transposable element annotators often fail to reliably detect tandem repeats, TRASH was applied to identify tandem repeat arrays with subunits less than 10,000 kbp. This analysis revealed a diverse range of large repeat subunits across the genome, including within the centromeric regions (Table 3). As expected, only a subset of centromeres were repetitive, specifically cen04, cen07, cen08, and cen10, while the remainder appeared repeatless. Interestingly, cen01, cen08, and cen11 showed a mixture of repeatless and repetitive regions enriched for CENH3 binding. Although repeat subunit sizes varied among centromeres, all were within the kilobase range. The majority of centromeres consisted of a uniform tandem array comprised of a single repeat subunit, except cen04, which contained multiple distinct arrays composed of varying-length repeats. In line with previous observations, centromeric repeats were enriched for CENH3 mapping relative to noncentromeric repeats (Fig. 3b). Interestingly, repeats for each centromere were organized into clearly distinct clusters when arranged by size and CENH3 mapping depth (Fig. 3c).

Manual inspection of the centromeric regions also indicated the presence of intact transposable element insertions, including insertions that disrupted the repetitive arrays. To identify transposable element families with a disposition for centromere presence, their bias towards presence in the centromere was calculated. From this, it appears that the CRM LTRs and a subgroup of Tekay LTRs have specialized to invade the *S. verrucosum* centromeres (Fig. 3d).

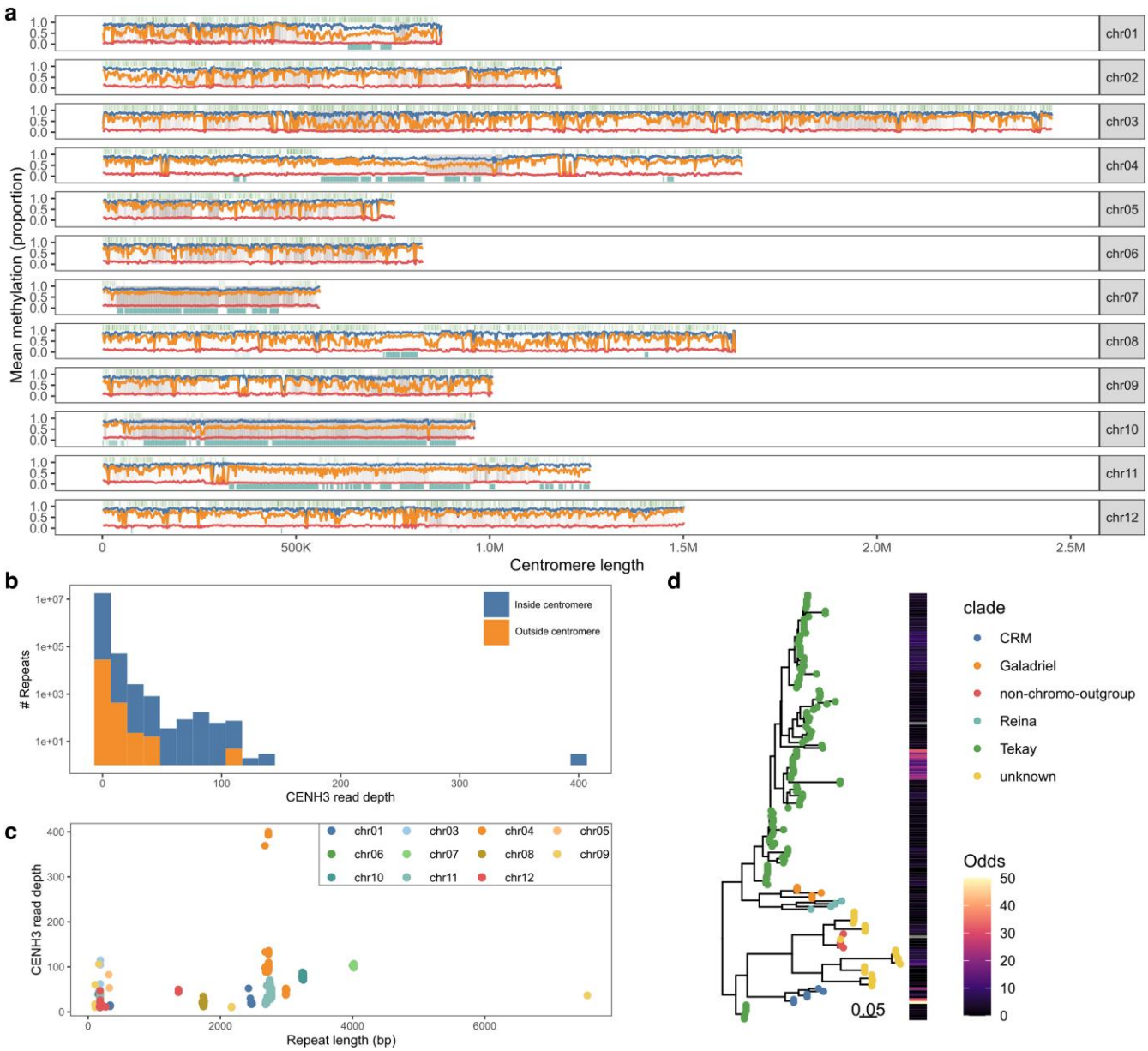

Fig. 3. a) The density of LTRs (top bar, green), CENH3 read mapping (middle background, gray), TRASH-derived repeats (bottom bar, teal), and CG, CHG, and CHG methylation (blue, orange, and red lines, respectively) across *S. verrucosum* centromeres. b) Histogram of repeats identified across the genome according to mean CENH3 read depth and on presence/absence in the centromere. c) Distribution of centromeric repeats by length and average CENH3 read depth, colored by centromere identity. d) Subset of phylogenetic tree of LTRs identified in this study built from the RT domain. LTRs colored according to their clade identified by TEsorter. The odds ratio of centromere bias from Fisher's exact test are highlighted.

Centromeres are dynamic structures, and previous studies have indicated that the centromeres of *S. verrucosum* and *S. tuberosum* are divergent (Zhang et al. 2014). As an example, a comparison of the sequence of cen7 revealed little sequence similarity between *S. verrucosum* and *S. tuberosum* (Supplementary Fig. 5). In line with the observation that *S. verrucosum* centromeric arrays are mostly uniform for each chromosome, no higher-order repeat topology was observed in either centromere.

In the centromeres of *Arabidopsis* and pepper, methylation in the CHG context is relatively reduced, potentially due to depletion of H3K9me2, which maintains methylation homeostasis (Naish et al. 2021; Zhang et al. 2025). A similar trend was not observed for *S. verrucosum*, where no clear correlation between centromeric repeats is observed, beyond the general trend of increased methylation towards the centromeric locus (Fig. 3a).

The repeatless centromeres of *S. verrucosum* are similar in structure to 2 recent assemblies of the *C. annuum* and *C. rhomboideum* genomes (Chen et al. 2024b). In both cases, the centromeres are formed of a rich landscape of Ty3 LTRs, with members of the Athila, Tekay, and CRM clades being present. In *Capsicum*, CRM LTRs are particularly dominant in the repeatless centromeres. One explanation for this enrichment is that CRM clades have specialized chromodomain and CR motifs which enable centromere targeting through an unknown mechanism (Neumann et al. 2011). The absence of these structural elements in families such as the centromere-biased Tekay LTRs observed in *S. verrucosum* indicates that other mechanisms exist that drive centromere

Table 3. The size, repeat status, and description of the *S. verrucosum* centromeres identified in this study.

Centromere	Size	Repetitive	Repeat size	Description
Cen01	0.9 Mbp	Partially	2.6 kbp	Repeatless region derived from Tekay, CRM, and Athila LTRs. Repeat array is less CENH3 enriched—repeats are CRM LTR-derived and contain RH, INT, and CHDCR domains
Cen02	1.2 Mbp	No	-	Tekay, CRM, and Athila LTR-derived
Cen03	2.5 Mbp	No	-	Tekay, CRM, and Athila LTR-derived
Cen04	1.7 Mbp	Yes	2.7/4.1 kbp	Highly repetitive—formed from multiple distinct arrays. Subunits are LTR-derived but with no identifiable LTR domains. Evidence of intact LTR insertions inside arrays
Cen05	0.8 Mbp	No	-	Tekay, CRM, and Athila LTR-derived
Cen06	0.8 Mbp	No	-	Tekay, CRM, and Athila LTR-derived
Cen07	0.6 Mbp	Yes	4.0 kbp	Highly repetitive—formed of single array. Subunits are CRM-derived and contain PROT domains. Multiple CRM insertions have disrupted the array
Cen08	1.6 Mbp	Partially	1.7 kbp	Repeatless region derived from Tekay, CRM, and Athila LTRs. Repeat array is less CENH3 enriched—repeats are LTR-derived but with no identifiable domains
Cen09	1.0 Mbp	No	-	Tekay, CRM, and Athila LTR-derived
Cen10	1.0 Mbp	Yes	3.1 kbp	Highly repetitive—formed of single array. Subunits are Tekay-derived and contain GAG domain
Cen11	1.3 Mbp	Partially	2.7 kbp	Repeatless region derived from Tekay, CRM, and Athila LTRs. Repeats are LTR-derived with no identifiable domains
Cen12	1.6 Mbp	No	-	Tekay, CRM, and Athila LTR-derived

Size is approximated from regions with dense CENH3 mapping.

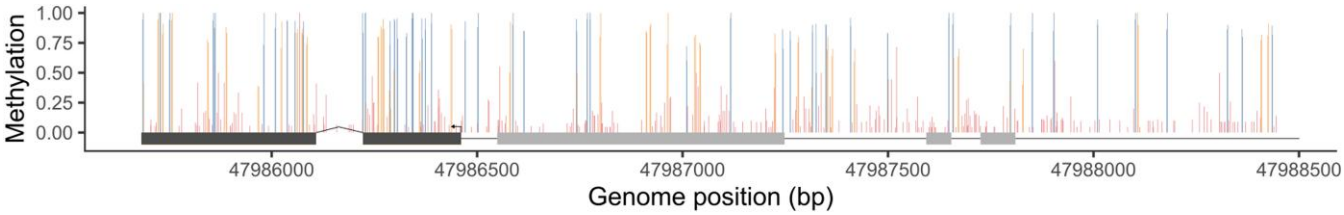

Fig. 4. The epigenetic state of *S-RNase* in *S. verrucosum*. The *S-RNase* exon sequence is highlighted in dark gray and identified upstream TIR sequences in light gray. The proportion of reads with methylated cytosines in the CG (blue), CHG (orange), and CHH (red) contexts is also visualized.

insertion or retention. Recently, a mechanism that drives centromere-bias of LTRs has been elucidated for the ATHILA family in *Arabidopsis*, whereby centromeres become "addicted" to ATHILA insertions due to their silencing transcription whilst simultaneously providing small RNAs that restore normal centromere function (Shimada et al. 2024).

Given the dynamic nature of centromeres, it should be noted that the CENH3 ChIP reads used in this study were from a previous study of a different *S. verrucosum* line—as a result, CENH3 mapping peaks presented in this post hoc analysis are unlikely to be fully representative of the true state of CENH3 mapping in this genome (Zhang et al. 2014). To fully capture the structure and activity of repeats and transposable elements in *S. verrucosum* centromeres, future investigations would benefit from producing paired ChIP sequencing and long-read centromere assemblies of multiple individuals. Significant advances in ONT sequencing chemistry and assembly software will improve the quality of assembled centromeres while simultaneously providing methylation data (Cheng et al. 2026).

The status of *S-RNase*

The self-compatibility of *S. verrucosum* has been attributed to a lack of a functional *S-RNase* in the genome, either through mutation, a lack of expression, or complete absence. A search for the *S. neorickii S-RNase* (BAC00940.1) revealed a single high-scoring hit cpc54_3362 (Identity: 48.9% E-value: $9.78e^{-68}$) on chromosome 1, the expected location of the S-locus.

The sequence of the *S. verrucosum S-RNase* gene was examined for any mutations that could result in non-functionality. No evidence of any truncations or significant insertions that might lead to a loss of function was observed. Previously identified mutations that might have led to reduced or a loss of *S-RNase* function, including a lost N-glycosylation site and histidine active sites, are also not present (Broz et al. 2021), suggesting that the *S-RNase* of *S. verrucosum* is likely to be functional.

No expression of the gene cpc54_3362 was identified in any of the RNA expression analyses. However, since the RNA-seq analysis did not specifically include flower tissue, we conducted a qRT-PCR analysis to assess *S-RNase* expression in this tissue. In flowers of *S. verrucosum* 54, *S-RNase* expression was detected with a Ct value of 32.6, relative to the housekeeping gene *Sec3* (Ct 26.6), corresponding to a relative expression level of 0.0156. Therefore, although *S-RNase* was expressed, this very low expression level may contribute to the self-compatibility observed in this accession.

The *S-RNase* gene is in a pericentromeric region that is exceptionally dense in LTRs and other transposable elements. Analysis of the 2 kb upstream region also revealed a high density of TEs (Fig. 4). However, annotations produced by Earl Grey and EDTA differed. EDTA identified 2 TIR fragments belonging to the Mariner and CACTA families within a repetitive region, whereas Earl Grey did not classify these elements fully. Instead, Earl Grey identified 2 large LINE/L1 fragments and an intermediate MULE-MuDR element in the same region.

Transposable element activity at the *S-RNase* locus is common to angiosperms, likely driving their diversification through relocating or shuffling the gene structure (Liu et al. 2025). In citrus, MITE transposable element insertions have directly led to self-compatibility transitions and are indirectly associated with tomato self-compatibility alleles (Broz et al. 2021; Hu et al. 2024).

Likely as a result of transposable element activity, the *S-RNase* gene and its surrounding locus in *S. verrucosum* are densely methylated in the CG and CHG contexts but not in the CHH context (Fig. 4). This methylation pattern is likely to affect gene expression. In almond, DNA methylation of the *S-RNase* promoter region has been closely linked to the transition to self-compatibility (Fernández i Martí et al. 2014). In *Nicotiana alata*, the promoter region of *S-RNase* accumulates dense CHH methylation specifically in pistil tissue, where the gene is highly expressed (Salcedo-Sánchez et al. 2023).

Interestingly, a recent study of a self-compatible citrus mutant reported no change in average DNA methylation across the *S-RNase* locus in style tissue. However, a distinct CHH methylation island was identified upstream of the locus, present exclusively in the self-compatible mutant (Hu et al. 2021). These findings together suggest that both the presence and context of DNA methylation in the *S-RNase* promoter region may play key roles in regulating expression and influencing self-compatibility in *S. verrucosum*.

Unexplored in this study are the potential roles of *S-RNase*-independent mechanisms of self-compatibility. The presentation of a high-quality *S. verrucosum* genome will enable future mapping studies to identify potential candidate loci.

Discussion

Here, a high-quality assembly of the *S. verrucosum* genome is presented alongside gene, transposable element, and DNA methylation annotations.

The full complement of NLRs and RLK/RLPs is resolved for *S. verrucosum*, highlighting that a significant proportion of expressed NLRs evade identification by standard gene annotation approaches such as BRAKER3. Re-analysis of existing *Rpi-ver1* mapping data reduced the size of the resistance locus to a locus on chromosome 9 between 54.6 and 56.0 Mbp and revealed its full sequence identity. Key candidates of the *Rpi-ver1* resistance gene include a pseudogenized NLR, a Rx-CC Jacalin-like lectin domain protein, and a cysteine-rich kinase RLP. Resolving the *Rpi-ver1* locus enables the design of markers for future fine-mapping experiments to identify the underlying gene.

The complete sequence of the *S. verrucosum* centromeres is also presented, revealing a complex of repeatless and repetitive sequences that appears to be actively shaped by transposable elements. Given the relative ease with which the *S. verrucosum* genome is assembled, we anticipate that future studies on the trans-generational dynamics of its centromeres will be valuable in elucidating their plasticity. An attractive hypothesis is that the repeatless centromeres may function as transitory centromeric structures until a repetitive structure can be seeded and expanded. DNA methylation is not a predictive mark of *S. verrucosum* centromeres—the identification of other centromere-specific marks, particularly chromatin modifications, will be beneficial in identifying epigenetic features that define the repetitive and repeatless centromeres.

Finally, the presence of an intact *S-RNase* gene in *S. verrucosum* was confirmed. The gene shows no structural abnormalities that would suggest non-functionality, but it is lowly expressed and densely methylated, likely due to transposable element activity in its promoter region.

Data availability

All analysis code is available at https://github.com/swiftseal/sver54_assembly. Genome and annotation data is available at https://doi.org/10.5281/zenodo.17107539 and has been deposited at ENA under accession GCA_978172885.1. Read data is available under PRJEB97628 ERP180128.

Supplemental material available at G3 online.

Acknowledgments

The authors acknowledge the Research/Scientific Computing teams at The James Hutton Institute and NIAB for providing computational resources and technical support for the "UK's Crop Diversity Bioinformatics HPC" (BBSRC grant BB/S019669/1), use of which has contributed to the results reported within this paper.

The PacBio sequencing service was provided by the Norwegian Sequencing Centre (https://www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities.

Funding

This work was supported by the Rural & Environment Science & Analytical Services (RESAS) Division of the Scottish Government through the project JHI-B1-1, the Biotechnology and Biological Sciences Research Council (BBSRC) through awards BB/S015663/1 and BB/X009068/1. MS was funded by the BBSRC award BB/T00875X/1.

Conflicts of interest

None declared.

Author contributions

MS conducted all data analysis presented in this manuscript. AK conducted all RNA-seq tissue and RNA preparation. VS prepared and analyzed qPCR expression data. MS, JJ, and IH contributed to the planning of the study. All authors substantively revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Literature cited

Baril T, Galbraith J, Hayward A. 2024. Earl grey: a fully automated user-friendly transposable element annotation and analysis pipeline. Mol Biol Evol. 41:msae068. https://doi.org/10.1093/molbev/msae068.

Behling WL, Douches DS. 2023. The effect of self-compatibility factors on interspecific compatibility in *Solanum* section petota. Plants. 12:1709. https://doi.org/10.3390/plants12081709.

Bernhofer M, Rost B. 2022. TMbed: transmembrane proteins predicted through language model embeddings. BMC Bioinformatics. 23:326. https://doi.org/10.1186/s12859-022-04873-x.

Broz AK *et al.* 2021. S-RNase alleles associated with self-compatibility in the tomato clade: structure, origins, and expression plasticity. Front Genet. 12:780793. https://doi.org/10.3389/fgene.2021.780793.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 6:e4958. https://doi.org/10.7717/peerj.4958.

Carrasco A *et al.* 2000. Transfer of PLRV resistance from *Solanum verrucosum* schlechdt to potato (*S. tuberosum* L.) by protoplast electrofusion. Potato Res. 43:31–42. https://doi.org/10.1007/BF02358511.

Castelli L *et al.* 2005. Life stage responses observed amongst fifteen wild *Solanum* species resistant to Globodera pallida. Nematology. 7:701–711. https://doi.org/10.1163/156854105775142955.

Chen X *et al.* 2018. Identification and rapid mapping of a gene conferring broad-spectrum late blight resistance in the diploid potato species *Solanum verrucosum* through DNA capture technologies. Theor Appl Genet. 131:1287–1297. https://doi.org/10.1007/s00122-018-3078-6.

Chen Q *et al.* 2024a. GCI: a continuity inspector for complete genome assembly. Bioinformatics. 40:btae633. https://doi.org/10.1093/bioinformatics/btae633.

Chen W *et al.* 2024b. Two telomere-to-telomere gapless genomes reveal insights into Capsicum evolution and capsaicinoid biosynthesis. Nat Commun. 15:4295. https://doi.org/10.1038/s41467-024-48643-0.

Cheng H *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.

Cheng H *et al.* 2026. Efficient near-telomere-to-telomere assembly of nanopore simplex reads. Nature. https://doi.org/10.1038/s41586-026-10105-6.

Cho K-S *et al.* 2022. Mitochondrial genome recombination in somatic hybrids of *Solanum commersonii* and *S. tuberosum*. Sci Rep. 12:8659. https://doi.org/10.1038/s41598-022-12661-z.

Cooper WR, Bamberg JB. 2016. Variation in susceptibility to potato psyllid, *Bactericera cockerelli* (Hemiptera: Triozidae), among *Solanum verrucosum* germplasm accessions. Am J Potato Res. 93:386–391. https://doi.org/10.1007/s12230-016-9512-x.

Cui W *et al.* 2021. A reduction in cadmium accumulation and sulphur containing compounds resulting from grafting in eggplants (*Solanum melogena*) is associated with DNA methylation. Plant Soil. 468:183–196. https://doi.org/10.1007/s11104-021-05122-5.

Dobin A *et al.* 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

Durand NC *et al.* 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3:95–98. https://doi.org/10.1016/j.cels.2016.07.002.

Eijlander R *et al.* 2000. Occurrence of self-compatibility, self-incompatibility and unilateral incompatibility after crossing diploid S. *tuberosum* (SI) with S. *verrucosum* (SC): I. Expression and inheritance of self-compatibility. Euphytica. 115:127–139. https://doi.org/10.1023/A:1003902907599.

Esch L, Schaffrath U. 2017. An update on jacalin-like lectins and their role in plant defense. Int J Mol Sci. 18:1592. https://doi.org/10.3390/ijms18071592.

Fernández i Martí A, Gradziel TM, Socias i Company R. 2014. Methylation of the S f locus in almond is associated with S-RNase loss of function. Plant Mol Biol. 86:681–689. https://doi.org/10.1007/s11103-014-0258-x.

Gabriel L *et al.* 2024. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. 34:769–777. https://doi.org/10.1101/gr.278090.123.

Gong Z *et al.* 2012. Repeatless and repeat-based centromeres in potato: implications for centromere evolution. Plant Cell. 24:3559–3574. https://doi.org/10.1105/tpc.112.100511.

Holst F *et al.* 2025. Helixer: ab initio prediction of primary eukaryotic gene models combining deep learning and a hidden Markov model. Nat Methods. 23:732–739. https://doi.org/10.1038/s41592-025-02939-1.

Hosaka AJ, Sanetomo R, Hosaka K. 2022. A de novo genome assembly of *Solanum verrucosum* Schlechtendal, a Mexican diploid species geographically isolated from other diploid A-genome species of potato relatives. G3 (Bethesda). 12:jkac166. https://doi.org/10.1093/g3journal/jkac166.

Hu J *et al.* 2021. Downregulated expression of S2-RNase attenuates self-incompatibility in "Guiyou No. 1" pummelo. Hortic Res. 8:199. https://doi.org/10.1038/s41438-021-00634-8.

Hu J *et al.* 2024. A rare inter-haplotypic recombination at the S-locus contributed to the loss of self-incompatibility in trifoliate orange. Plant Commun. 5:100940. https://doi.org/10.1016/j.xplc.2024.100940.

Ivanov AA, Ukladov EO, Golubeva TS. 2021. *Phytophthora infestans*: an overview of methods and attempts to combat late blight. J Fungi. 7:1071. https://doi.org/10.3390/jof7121071.

Jain C *et al.* 2020. Weighted minimizer sampling improves long read mapping. Bioinformatics. 36:i111–i118. https://doi.org/10.1093/bioinformatics/btaa435.

Kamel AM *et al.* 2023. The expression of Triticum aestivum cysteine-rich receptor-like protein kinase genes during leaf rust fungal infection. Plants. 12:2932. https://doi.org/10.3390/plants12162932.

Katoh K *et al.* 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066. https://doi.org/10.1093/nar/gkf436.

Kille B *et al.* 2023. Minmers are a generalization of minimizers that enable unbiased local Jaccard estimation. Bioinform 39. https://doi.org/10.1093/bioinformatics/btad512.

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.

Kolmogorov M *et al.* 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.

Kuznetsov D *et al.* 2023. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. 51:D445–D451. https://doi.org/10.1093/nar/gkac998.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. Nat Methods. 9:357–359. https://doi.org/10.1038/nmeth.1923.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [preprint]. arXiv. https://doi.org/10.48550/arXiv.1303.3997.

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

Liu Y *et al.* 2025. S-RNase evolution in self-incompatibility: phylogenomic insights into synteny with class I T2 RNase genes. Plant Physiol. 197:kiaf072. https://doi.org/10.1093/plphys/kiaf072.

Liu Z, Halterman D. 2006. Identification and characterization of RB-orthologous genes from the late blight resistant wild potato species *Solanum verrucosum*. Physiol Mol Plant Pathol. 69:230–239. https://doi.org/10.1016/j.pmpp.2007.05.002.

Lu Y *et al.* 2021. DNA methylation dynamics of sperm cell lineage development in tomato. Plant J. 105:565–579. https://doi.org/10.1111/tpj.15098.

Lüdke D *et al.* 2025. A root-specific NLR network mediates immune signaling of resistance genes against plant parasitic nematodes. Plant Cell. 37:koaf145. https://doi.org/10.1093/plcell/koaf145.

Madeo G *et al.* 2023. Coconat: a novel method based on deep learning for coiled-coil prediction. Bioinformatics. 39:btad495. https://doi.org/10.1093/bioinformatics/btad495.

Martin EC *et al.* 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.

Naish M *et al.* 2021. The genetic and epigenetic landscape of the Arabidopsis centromeres. Science. 374:eabi7489. https://doi.org/10.1126/science.abi7489.

Neumann P *et al.* 2011. Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mob DNA. 2:4. https://doi.org/10.1186/1759-8753-2-4.

Ni P *et al.* 2021. Genome-wide detection of cytosine methylations in plant from nanopore data using deep learning. Nat Commun. 12:5976. https://doi.org/10.1038/s41467-021-26278-9.

Open2C *et al.* 2023 Feb 15. Pairtools: from sequencing data to chromosome contacts [preprint]. bioRxiv, p. 2023.02.13.528389. https://doi.org/10.1371/journal.pcbi.1012164.

Paajanen P *et al.* 2019. A critical comparison of technologies for a plant genome sequencing project. GigaScience. 8:giy163. https://doi.org/10.1093/gigascience/giy163.

Percival-Alwyn L *et al.* 2024. 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.

Pham GM *et al.* 2020. Construction of a chromosome-scale long-read reference genome assembly for potato. GigaScience. 9:giaa100. https://doi.org/10.1093/gigascience/giaa100.

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.

Salcedo-Sánchez R, Cruz-Zamora Y, Cruz-García F. 2023. The SC10-*RNase* promoter displays changes in DNA methylation patterns through pistil development in self-incompatible *Nicotiana alata*. Plant Physiol Biochem. 205:108161. https://doi.org/10.1016/j.plaphy.2023.108161.

Shen W *et al.* 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 11:e0163962. https://doi.org/10.1371/journal.pone.0163962.

Shimada A *et al.* 2024. Retrotransposon addiction promotes centromere function via epigenetically activated small RNAs. Nat Plants. 10:1304–1316. https://doi.org/10.1038/s41477-024-01773-1.

Smith M, Jones JT, Hein I. 2025. Resistify: a novel NLR classifier that reveals helitron-associated NLR expansion in Solanaceae. Bioinform Biol Insights. 19:11779322241308944. https://doi.org/10.1177/11779322241308944.

Stoiber M *et al.* 2017 Apr 10. De novo identification of DNA modifications enabled by genome-guided nanopore signal processing [preprint]. bioRxiv, p. 094672. https://doi.org/10.1101/094672.

Tegenfeldt F *et al.* 2025. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. Nucleic Acids Res. 53:D516–D522. https://doi.org/10.1093/nar/gkae987.

Tillich M *et al.* 2017. Geseq—versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45:W6–W11. https://doi.org/10.1093/nar/gkx391.

von Dahlen JK *et al.* 2023. Global expression patterns of R-genes in tomato and potato. Front Plant Sci. 14:1216795. https://doi.org/10.3389/fpls.2023.1216795.

Wlodzimierz P *et al.* 2023b. Cycles of satellite and transposon evolution in Arabidopsis centromeres. Nature. 618:557–565. https://doi.org/10.1038/s41586-023-06062-z.

Wlodzimierz P, Hong M, Henderson IR. 2023a. TRASH: tandem repeat annotation and structural hierarchy. Bioinformatics. 39:btad308. https://doi.org/10.1093/bioinformatics/btad308.

Yin Z, Liu J, Dou D. 2024. RLKdb: a comprehensively curated database of plant receptor-like kinase families. Mol Plant. 17:513–515. https://doi.org/10.1016/j.molp.2024.02.014.

Zeng Y, Dawe RK, Gent JI. 2023. Natural methylation epialleles correlate with gene expression in maize. Genetics. 225:iyad146. https://doi.org/10.1093/genetics/iyad146.

Zhang H *et al.* 2014. Boom-bust turnovers of megabase-sized centromeric DNA in *Solanum* species: rapid evolution of DNA sequences associated with centromeres. Plant Cell. 26:1436–1447. https://doi.org/10.1105/tpc.114.123877.

Zhang R-G *et al.* 2022. TEsorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. Hortic Res. 9:uhac017. https://doi.org/10.1093/hr/uhac017.

Zhang L *et al.* 2024. *Solanum aculeatissimum* and *Solanum torvum* chloroplast genome sequences: a comparative analysis with other Solanum chloroplast genomes. BMC Genomics. 25:412. https://doi.org/10.1186/s12864-024-10190-9.

Zhang K *et al.* 2025. The gap-free assembly of pepper genome reveals transposable-element-driven expansion and rapid evolution of pericentromeres. Plant Commun. 6:101177. https://doi.org/10.1016/j.xplc.2024.101177.

Zhou C *et al.* 2025. Oatk: a de novo assembly tool for complex plant organelle genomes. Genome Biol. 26:235. https://doi.org/10.1186/s13059-025-03676-6.

Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 39:btac808. https://doi.org/10.1093/bioinformatics/btac808.

Editor: J. Birchler