

RESEARCH PAPER

Fine-mapping, candidate gene identification, and marker development for the apple scab resistance gene *Rvi2*

Elena López-Girona^{1,*}, Chris Kirk¹, Cecilia H. Deng², Anže Švara^{3,4}, Awais Khan³, Vincent G.M. Bus⁵, David Chagné¹, and Richard K. Volz⁵

¹ New Zealand Institute for Bioeconomy Science Limited (Bioeconomy Science Institute), Food Innovation Research Centre, Palmerston North 4442, New Zealand

² Bioeconomy Science Institute, Mount Albert Research Centre, Auckland 1142, New Zealand

³ School of Integrative Plant Science, Cornell University, Geneva, NY 14456, USA

⁴ Department of Plant Sciences, University of Saskatchewan, Saskatoon, Canada

⁵ Bioeconomy Science Institute, Hawkes Bay Research Centre, Havelock North 4157, New Zealand

* Correspondence: Elena.Lopez-Girona@plantandfood.co.nz

Received 4 July 2025; Accepted 3 November 2025

Editor: Fabrizio Costa, University of Trento, Italy

Abstract

Breeding elite apple cultivars with scab resistance is a key global goal, as reliance on fungicides is unsustainable. The causal fungus, *Venturia inaequalis*, evolves rapidly, threatening cultivars with single-gene resistance. Since the 1980s, breeding programmes have introduced novel resistance sources via backcrossing. Here, we generated a haplotype-phased genome assembly of Russian apple R12740-7A and an Oxford Nanopore assembly of the *Rvi2*-resistance accession TSR34T15, enabling detailed dissection of the *Rvi2* resistance locus. Fine-mapping using a ‘Royal Gala’ × TSR34T15 segregating family delimited *Rvi2* to a narrow genomic interval, within which we identified a 10 041 bp long terminal repeat retrotransposon (LTR-RT) insertion—an insert-based structural variant (SV) strongly linked with *Rvi2*. Notably, this LTR-RT harbours an *FPPS* gene, a member of the farnesyl pyrophosphate/geranylgeranyl pyrophosphate (FPP/GGPP) synthase family, located 2 kb from a key candidate defence gene. Although the *FPPS* gene exhibits stable expression, its integration within the retrotransposon suggests a *cis*-regulatory role, potentially priming adjacent defence genes for robust up-regulation upon pathogen attack. We validated the marker derived from this SV in diverse germplasms and successfully implemented it in marker-assisted selection across extensive seedling cohorts. This marker will streamline the development of scab-resistant apple varieties.

Keywords: Apple scab, long-terminal repeats, *Malus*, marker-assisted selection, molecular marker, resistance gene, *Venturia inaequalis*.

Introduction

Apple scab, caused by the fungus *Venturia inaequalis*, is a major disease of apple (*Malus domestica*) and requires intensive fungicide management. Breeding elite cultivars with scab resistance

is the preferred long-term strategy to reduce pesticide use and control the disease. However, resistance can be overcome by rapid evolution of the pathogen’s virulence (Caffier *et al.*, 2014),

which can accumulate through its annual sexual cycle, posing a significant threat to current cultivars that predominantly rely on single-gene resistance (Bastiaanse *et al.*, 2016). Therefore, developing elite cultivars with multiple resistance genes as well as desirable fruit qualities is a high priority for pome fruit breeding programmes worldwide.

To date, 19 resistance genes for *V. inaequalis* (*Rvi* genes) have been described (Patocchi *et al.*, 2020; Khan and Korban, 2022; Khan *et al.*, 2024). However, candidate genes have been identified only for *Rvi1*, *Rvi6*, *Rvi12*, *Rvi15*, *Rvi17*, and *Vhc1* (Parisi *et al.*, 1993; Belfanti *et al.*, 2004; Malnoy *et al.*, 2008; Bus *et al.*, 2009, 2011; Joshi *et al.*, 2011; Jänsch *et al.*, 2014; Schouten *et al.*, 2014; Cova *et al.*, 2015; Padmarasu *et al.*, 2018; Papp *et al.*, 2020; Švara *et al.*, 2024). Among these, only *Rvi6* (*Vf*) and *Rvi15* (*Vr2*), recently shown to be the same gene as *Rvi4* (*Vh4*) (Peil *et al.*, 2023), have been functionally validated as a receptor kinase and a Toll Interleukin-1 Receptor–Nucleotide-Binding Site–Leucine-Rich Repeats (TIR–NBS–LRP) gene, respectively (Bus *et al.*, 2011; Padmarasu *et al.*, 2018).

The apple breeding programme at the Bioeconomy Science Institute, New Zealand (formerly Plant & Food Research and HortResearch), began developing scab-resistant cultivars in the early 1980s by introducing founders carrying *Rvi6* resistance (Peil *et al.*, 2023). This led to the release of two cultivars carrying only this resistance gene, namely ‘PremA34’ and the open-release ‘PremA25’. However, many scab resistance genes, including *Rvi6*, show differential interactions with pathogen races: some virulences persist over time while others have emerged more recently (Parisi *et al.*, 1993; Xu *et al.*, 2008; Caffier *et al.*, 2014; Papp *et al.*, 2020; Patocchi *et al.*, 2020; Peil *et al.*, 2023). To address this, and to identify new resources of disease resistance, a pre-breeding programme was launched at the Institute in 1993 to identify new sources of disease resistance, leading to the introgression and pyramiding of multiple resistance genes through backcrossing (Bus *et al.*, 2000). Developing more durable resistance via gene pyramiding requires genetic mapping of resistance loci (Williams and Kuc, 1969) and the validation of markers for marker-assisted selection (MAS) (Khan and Korban, 2022; Khan *et al.*, 2024).

The *Rvi2* (*Vh2*) gene was first introduced into the Bioeconomy Science Institute breeding programme via pollen from accession PRI 2375-5 of the Purdue–Rutgers–Illinois programme in the USA (Bus *et al.*, 2000), which was used to create family A068. This family was an F₃ from the non-differential accession OR42T175 (PRI 442-23) (Williams and Kuc, 1969) and an F₅ of Russian apple R12740-7A (*M. pumila*), a key source of *Rvi4* (*Vh4*) and *Rvi2* for scab resistance in apple breeding programmes worldwide. It also contains *Rvi19* (Gardiner *et al.*, 2007) and might have more genes associated with scab resistance (Dayton *et al.*, 1953). The full-sib A068T03T079 × A068R03T013 family led to two key selections, A185R09T174 and A185R09T179 (Bus *et al.*, 2005b), with which gene pyramiding with *Rvi6* commenced in the Institute breeding programme in 2003 (Volz *et al.*, 2023).

In apple–*Venturia* interactions, stellate necrosis (SN) is a characteristic resistance phenotype in which browning or necrotic spots radiate outward in a star-shaped or stellate pattern from the infection site, often accompanied by autofluorescence of surrounding epidermal cells when observed under blue-range fluorescence (Bus *et al.*, 2011). The SN-conditioning *Rvi2* gene was first mapped to the distal end of linkage group (LG) 2 using two seedling progenies, ‘Sciglo’ × A068R03T057 and ‘Royal Gala’ × TSR34T15. TSR34T15 (PRI 384-1; OR42T173) is an F₂ selection from Russian apple R12740-7A (Bus *et al.*, 2005a), received as differential host X2250 from Y. Lespinasse, INRAE, Angers, France, and therefore carrying the *Rvi2* resistance. *Rvi2* mapped above the simple sequence repeat (SSR) marker CH05e03 (Liebhard *et al.*, 2002) and the OPL19 SCAR marker in TSR34T15, but below them in A068R03T057 (Bus *et al.*, 2005b). Later, Jänsch *et al.* (2015) positioned *Rvi2* below the CH05e03 SSR, but above OPL19 SCAR using a ‘Golden Delicious’ × TSR34T15 segregating population (Bus *et al.*, 2005a) and developed two single-nucleotide polymorphism (SNP) markers (FBsnRvi2-7 and FBsnRvi2-8), located ~1.2 cM from the locus. The SNP FBsnRvi2-7 was later converted into a quantitative PCR (qPCR) assay by Chagné *et al.* (2019), who reported co-segregation with *Rvi2* in the tested germplasm. However, the 1.2 cM distance suggests that these markers from Jänsch *et al.* (2015) and Chagné *et al.* (2019) might not be universally effective for all germplasm, as recombination could occur within the targeted locus. Thus, a more accurate, high-throughput, and broadly applicable marker remains necessary for efficient screening and MAS of *Rvi2*.

Advances in long-read sequencing have made whole-genome sequencing more feasible and accurate, enabling complete assembly and haplotyping, even in heterozygous species such as apple (Zhang *et al.*, 2019; Sun *et al.*, 2020; Khan *et al.*, 2022; Mansfeld *et al.*, 2023). Haplotyped whole-genome assembly methodology detects all variant types, including structural variants (SVs), which is crucial when identifying causative variants for traits, whether they are SNPs or SVs. Resistance genes are often found in tandem clusters, which are prone to gene loss or gain through natural and/or artificial selection (McHale *et al.*, 2012; Barragan and Weigel, 2021). Recombination and unequal crossing-over can generate novel resistance (R) genes that provide fitness advantage against new pathogens. At the same time, in the absence of a pathogen, redundant R genes can be lost to avoid fitness costs. Consequently, R gene clusters are hotspots for SVs owing to their dynamic evolution under changing pathogen pressures. Similar structural diversity patterns have been documented in R gene clusters in Arabidopsis (Botella *et al.*, 1998; Guo *et al.*, 2011; Barragan *et al.*, 2019; Jiao and Schneeberger, 2020), soybean (Ashfield *et al.*, 2012), barley (Muñoz-Amatrián *et al.*, 2013), and humans (Conrad *et al.*, 2010). In apple, a cluster of 19 copy number variants (CNVs) has been identified at the top of chromosome 2 (Boocock *et al.*, 2015), overlapping

with major apple scab resistance loci (Bus *et al.*, 2004, 2005b; Patocchi *et al.*, 2004) such as *Rvi4/Rvi15* (Peil *et al.*, 2023). Additionally, *Rvi2* is part of a cluster with three other major scab R genes [*Rvi2*, *Rvi11* (*Vhj*), *Rvi8* (*Vh8*); Jiao and Schneeberger, 2020] along with race-specific QTLs (Bus *et al.*, 2004). Another *V. inaequalis* resistance cluster is the *Vf* locus, comprising four R genes (*HcrVf1*, *HcrVf2*, *HcrVf3*, and *HcrVf4*) (Vinatzer *et al.*, 2001), *Vf2ARD* (Boudichevskaia *et al.*, 2009), and *Vd3* (Soriano *et al.*, 2009). Haplotyped whole-genome assembly enables detection and characterization of such variants and opens new avenues for understanding genetic resistance in apple and other complex genomes.

Here, we fine-mapped the apple *Rvi2* locus using long-read sequencing to generate a haplotyped-phased genome assembly of Russian apple R12740-7A, the source of the *Rvi2* resistance. An Oxford Nanopore assembly of the *Rvi2*-resistance accession TSR34T15 was also produced. We identified a long terminal repeat retrotransposon associated with the *Rvi2* resistance phenotype, offering a valuable tool for implementing MAS.

Materials and methods

Plant material and scab phenotyping

For the fine-mapping of *Rvi2*, a segregating F₁ population of 361 seedlings from a *Malus domestica* ‘Royal Gala’ × TSR34T15 cross was grown in a glasshouse at the Bioeconomy Science Institute, Havelock North, New Zealand (39°40'S, 176°53'E). The seedlings were inoculated with a conidial suspension of *Venturia inaequalis* isolate MNH120 and incubated as described by Bus *et al.* (2005b). Disease phenotyping was conducted in the third week post-inoculation using the scale developed by Chevalier *et al.* (1991), with an additional class for stellate necrosis (SN). The resistance phenotypes were classified as: scab resistant, when SN conditioned by the *Rvi2* gene was observed (R = Class 2, SN); scab susceptible, when extensive sporulation occurred (S = Class 4); and asymptomatic (no symptoms = Class 0), probably due to inoculation failure.

To validate the marker assay, trees from 239 advanced selections and commercial apple cultivars within the Bioeconomy Science Institute breeding programme in Havelock North were tested (Supplementary Table S1). Among these, 116 were confirmed to have the *Rvi2* resistance gene based on their pedigree and previous observation of leaf SN resistance reactions on 1-month-old seedlings under optimized glasshouse conditions (Bus *et al.*, 2005b), and/or segregation for the SN reaction in progenies produced by crossing with a scab-susceptible parent. For the remaining 123 selections/cultivars, the absence of the *Rvi2* gene was assumed because their parents and ancestors were also known to lack it. Additionally, an F₁ population of 110 seedlings derived from a cross between selection A (an *Rvi2* + *Rvi6* Bioeconomy Science Institute breeding selection) and ‘Royal Gala’ was genotyped and phenotyped for *Rvi2* resistance as previously described (Bus *et al.*, 2005b).

Genotyping and genetic linkage mapping

SNP array

Genomic DNA was extracted from leaves of 188 out of the total of 361 seedlings and progenitors from the ‘Royal Gala’ × TSR34T15 cross using a NucleoSpin®96 Plant II Kit (Macherey-Nagel). The DNA quality was checked using Nanodrop 1000 (ThermoFisher Scientific), quantified

using PicoGreen® (Invitrogen) on an Infinite® 200 PRO plate reader (Tecan), and then hybridized to the International RosBREED SNP Consortium (IRSC) apple 8k SNP array v1 (Chagné *et al.*, 2012), following the Illumina Infinium® HD Assay Ultra protocol. The array was scanned using an Illumina HiScan, and the Genotyping Module of the Illumina GenomeStudio™ software was used to normalize the sample intensity values, transform them into allele calls, and perform genotype clustering and filtering. The automatic genotype clustering mode determined the cluster positions for AA, AB, and BB genotypes for each marker. Markers that showed more than three well-defined clusters (probably because of the annealing to paralog genomic regions) were removed. To ensure high-quality SNP calling, stringent quality parameters were applied: GenCall ≥0.15, GenTrain score ≥0.50, minor allelic frequency (MAF) ≥0.15, and call rate ≥80%. Manual editing was performed for markers showing significant (*P* < 0.05) deviations from the expected Mendelian ratios or for which one of the parental genotypes was missing.

Linkage map construction

A genetic linkage for the *Rvi2* resistant parent TSR34T15 derived from the segregating F₁ ‘Royal Gala’ × TSR34T15 population was constructed using JoinMap v5.0 (Stam, 1993), prioritizing the use of the backcross-type markers (ab × aa) over the intercross markers (ab × ab). Linkage groups were determined with a minimum LOD score of 5, and map distances were calculated using the Kosambi function (Kosambi, 1943; Vinod, 2011). The resulting genetic linkage map was visualized with MapChart v2.32 (Voorrips, 2002).

Genome sequencing of Russian apple R12740-7a and TSR34T15

Sample collection and DNA isolation

Intact, young, fully expanded leaves (20 g) were collected from Russian apple (*M. pumila*) R12740-7A and its F₂ selection derivative TSR34T15 for sequencing as they both carry the *Rvi2* locus. Samples for R12740-7A were obtained from the research orchard at the USDA ARS Plant Germplasm Resources Unit at Cornell AgriTech in Geneva, NY, USA, while TSR34T15 samples were collected from the orchard at the Bioeconomy Science Institute, Havelock North, New Zealand. For TSR34T15, nuclear genomic DNA was extracted from isolated nuclei as described previously (López-Girona *et al.*, 2020). The R12740-7A samples were shipped on dry ice to the DNA Sequencing and Genotyping Center at the University of Delaware, USA, where high-molecular-weight DNA was extracted.

Genome sequencing and assembly of TSR34T15

The extracted nuclear genomic DNA (~45 µg) was sent to the Biomolecular Resource Facility at the John Curtin School of Medical Research (Australian National University, Canberra) for Oxford Nanopore PromethION sequencing. After purification with AMPure XP beads (0.3 volume ratio; Beckman Coulter, Inc.), ~7.4 µg of DNA was used to prepare a library with a SQK-LSK110 Kit (Oxford Nanopore Technologies, ONT). The final library had a concentration of 28.6 ng µl⁻¹ (686.40 ng of DNA molecules), a mean size of 59.4 kb and 17.51 fmol, and was sequenced on a single flow cell (FLO-PRO002) for 1d 19 h, and 50 min, with base-calling performed by Guppy (v5.0.7; ONT).

Quality assessment was conducted before and after base-calling using MinIONQC (Lanfear *et al.*, 2019) and NanoPlot (De Coster *et al.*, 2018). Reads were corrected with Canu v2.1.1 (Koren *et al.*, 2017) and used to generate 21-mers for genome size and heterozygosity estimates with Jellyfish 2.2.10 (Marçais and Kingsford, 2011) and GenomeScope2 (Ranallo-Benavidez *et al.*, 2020). *De novo* assembly was performed using Flye v2.8.3 (Kolmogorov *et al.*, 2019), Shasta v0.7.0 (Shafin *et al.*, 2020), and Canu v2.1.1. Flye and Shasta assemblies were

polished with Medaka (v1.4.3; <https://github.com/nanoporetech/medaka>) and Racon v1.4.7 (Vaser *et al.*, 2017). Scaffold statistics were produced using the `assemblathon_stats_v1.1.pl` script (<https://github.com/ucdavis-bioinformatics/assemblathon2-analysis>) (Bradnam *et al.*, 2013), and genome completeness was assessed with the embryophyte_odb10 (1614 genes) dataset from Benchmarking Universal Single-Copy Orthologs (BUSCO). Contigs were re-sorted into primary and haploid sets using Purge Haplotigs (Roach *et al.*, 2018) (version 1.1.1; settings 'LOW=5 MID=27 HIGH=72'). The curated assembly was scaffolded with RagTag (Alonge *et al.*, 2022) using alignments against the 'Golden Delicious' reference genome (GDDH13v1.1) (Daccord *et al.*, 2017) and the homozygous HFTH1 genome (Zhang *et al.*, 2019) via nucmer (Kurtz *et al.*, 2004) to produce 'agp' files, which were then processed with agptools (<https://github.com/WarrenLab/agptools>) to generate the final assembly. Final quality checks were performed on the assembly, using `assemblathon_stats` and BUSCO both before and after scaffolding.

Sequencing and phased assembly of Russian apple R12740-7a

High-molecular-weight DNA was used to prepare a SMRTbell library following protocols outlined previously (Khan *et al.*, 2022) and sequenced on a Pacific BioSciences (PacBio) Sequel II platform. Additionally, Illumina Hi-C chromatin conformation capture was performed using a Dovetail Genomics Omni-C Kit, with sequencing on an Illumina NovaSeq 6000 platform (PE150 reads) at the DNA Sequencing & Genotyping Center, University of Delaware's. HiFi data quality was assessed with FastQC v0.12.1. (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and read length distribution was checked with 'gt seqstat' in GenomeTools (Alonge *et al.*, 2022). K-mer analysis and heterozygosity estimation were conducted using Jellyfish (Marçais and Kingsford, 2011) and GenomeScope (Ranallo-Benavidez *et al.*, 2020). Genome assembly was performed with hifiasm (Cheng *et al.*, 2021), and gfa files were converted into fasta using awk (Aho *et al.*, 1979) and compressed with pigz (<http://zlib.net/pigz>). Quality checks followed the AssemblyQC pipeline (Rashid *et al.*, 2024), validating file formats, removing adapters, detecting duplicate sequences, screening contaminants, assessing telomeres, and evaluating completeness using BUSCO (Simão *et al.*, 2015) and LAI (Ou *et al.*, 2018). PacBio adaptor sequences were removed, and contigs classified as *Buchnera aphidicola* and *Aphis gossypii* were discarded. Synteny analysis against the HFTH1 genome (Zhang *et al.*, 2019) was done using Ragtag (Alonge *et al.*, 2022), with chromosome-scale scaffolds renamed accordingly. The final assembly was annotated using the GenePal pipeline (<https://github.com/Plant-Food-Research-Open/genepal>), including genome soft-masking with RepeatMasker (Tarailo-Graovac and Chen, 2009) and a pre-build pan-*Malus* TE library. Gene prediction utilized homologous amino acid sequences from TAIR10 (<https://www.arabidopsis.org/>) and the high-quality apple assemblies 'Honeycrisp' (Khan *et al.*, 2022), HFTH1 (Zhang *et al.*, 2019), GDDH13v1.1 (Daccord *et al.*, 2017), 'Hanfu' (Qin *et al.*, 2023), *M. fusca* (Mansfeld *et al.*, 2023), *M. sieversii*, *M. sylvestris* (Sun *et al.*, 2020), and *M. prunifolia* (Li *et al.*, 2022) retrieved from GDR (Jung *et al.*, 2014). Functional annotation of gene models was performed using EggNOG-mapper (Cantalapiedra *et al.*, 2021) with the eggNOG database 5.0.2 (Huerta-Cepas *et al.*, 2019).

Variant calling for fine-mapping of the *Rvi2* locus

ONT long reads from TSR34T15 were corrected with Canu v2.1.1 (Koren *et al.*, 2017) and 'Gala' HiFi reads were obtained from NCBI's SRA database (SRR12031344; <https://www.ncbi.nlm.nih.gov/sra/>). Both datasets along with the R12740-7A HiFi reads were aligned to the apple genome GDDH13v1.1 (Daccord *et al.*, 2017) using minimap2 v2.9 (Li, 2018) with a minimum mapping quality of 30 and at least 10 alternate allele observations. Alignments were visualized in IGV (Robinson

et al., 2017), and per-nucleotide coverage was calculated with samtools (v1.9) depth (Danecek *et al.*, 2021).

SNPs and short indels were called with Freebayes v1.3.6 (Garrison and Marth, 2012, Preprint) over the *Rvi2* locus as defined by flanking markers from the SNP array-based linkage map. Heterozygous variants unique to TSR34T15 (absent in 'Royal Gala') were selected using bcftools (Danecek *et al.*, 2021).

Structural variants (deletions, duplications, inversions, insertions, and translocations) in TSR34T15 ONT reads were detected using Sniffles (Sedlazeck *et al.*, 2018) (sniffles -m file.bam -min_support 3 -vcf file.vcf), and variants within the refined *Rvi2* locus were selected using the bcftools software (Danecek *et al.*, 2021).

Marker development for fine-mapping the *Rvi2* locus

SNP markers

SNPs that were heterozygous in TSR34T15 and homozygous in 'Royal Gala' against the 'Golden Delicious' genome GDDH13v1.1 (Daccord *et al.*, 2017) were screened within the *Rvi2* region. Primer pairs for high-resolution melting (HRM) analysis were designed as described by Chagné *et al.*, (2008). PCR reactions (10 µl total volume) contained 20 ng genomic DNA (extracted via the CTAB method; Doyle and Doyle, 1987), 2.5 mM MgCl₂, 200 nM of each primer, and 1× HRM master mix (Roche). These were run on a LightCycler[®] 480 (Roche Diagnostics) following the PCR programme outlined in López-Girona *et al.* (2021). Initially, the parental accessions of the 'Royal Gala' × TSR34T15 cross were screened, and the identified polymorphic markers were then applied to 173 of 361 seedlings to fine-map the *Rvi2* locus.

SSR markers

The 'Royal Gala' × TSR34T1 population was screened with SSRs markers, including the CH05e03 marker (Liebhard *et al.*, 2002) and the newly developed marker SSR_MDC016328_315, and the SNP-based AHI15L TaqMan[™] assay (Chagné *et al.*, 2019) derived from the SNP *Rvi2*-7_W242 (Jänsch *et al.*, 2015). PCR conditions were as described by López-Girona *et al.* (2021).

INDEL markers

To explore the associations between two specific long terminal repeat retrotransposon (LTR-RT) insertions (7110 bp at Chr02_31670599 and 10 041 bp at Chr02_31730132) and the *Rvi2* resistance, we analysed 'Royal Gala', TSR34T15, 16 seedlings from their cross (nine susceptible and seven resistant), and R12740-7A. Long-range PCRs were carried out using the following specific primer pairs. For the start of the 7110 bp insertion, 5'-ACCATAGACTTCTGAGGAAT-3' and 5'-CGATTGC GTGATGATATG-3'; at the end, 5'-CCAACACGTAGAC ATTGGA-3' and 5'-GCACTCTTGTGAATGCATAT-3'. For the start of the 10 041 bp insertion, 5'-TCGGATTCTCTCTTCAGTCT-3' and 5'-CCTTCAAAGATCAAGTCCCA-3'; at the end, 5'-CTTGATCTT GAAGTGGTGGGA-3' and 5'-ACGTGATTGTGAGAAATGGA-3'. The PCR reactions (total volume 10 µl) contained 3 µl 5× PS GXL Buffer, 1.2 µl dNTP mix (2.5 mM), 0.2 µl of each primer (10 µM), template DNA (10 ng µl⁻¹), 0.3 µl PrimerSTAR[®] GXL Taq polymerase (2.5 U µl⁻¹), and 8.1 µl sterile water. Cycling conditions were: initial denaturation at 98 °C for 10 s; 30 cycles of 98 °C, 10 s; annealing (53 °C for the 7110 bp insertion, 55 °C for the 10 025 bp insertion) for 15 s; extension 68 °C (10 s kb⁻¹); and a final extension at 68 °C for 2 min. Amplicons were separated by electrophoresis on 1% agarose in 1× TAE buffer.

High-throughput marker assay for *Rvi2*

To validate the association between the 10 041 bp (Chr02_31730132) LTR-RT insertion and *Rvi2* resistance, we developed a 5'-fluorogenic hydrolysis probe assay (termed *Rvi2*-ELG-10kb, for

Enhanced Locus Genotyping for *Rvi2* Resistance) to screen 239 apple selections (Supplementary Table S1), 361 seedlings from the ‘Royal Gala’ × TSR34T15 cross, and 110 seedlings from an *Rvi2*+*Rvi6* Bioeconomy Science Institute breeding selection A (*Rvi2*+*Rvi6*) × ‘Royal Gala’ cross. The assay employs two probes, namely ‘PrIns_*Rvi2* (FAM)’ that targets the insert by spanning the 5′ reference/insert boundary, and ‘*Rvi2*_ref_Pr (YAK)’ that detects the flanking sequence immediately adjacent to the insert. A primer pair (‘F_2Ins’ and ‘R_2Ins’) amplifies the reference (lacking the insertion), while ‘F_2Ins’, together with ‘R_2Ins’ amplify the insertion allele, which is detected by the insert probe ‘PrIns_*Rvi2* (FAM)’ (Fig. 1; Table 1). This assay distinguishes the three possible genotypes: resistant homozygous (AA), resistant heterozygous (Aa), and susceptible (aa).

PCR reactions were carried out in a total volume of 5 µl, containing 1× LightCycler 480 Probes Master Mix (Roche Diagnostics), 750 nM of each primer (‘F_2Ins’, ‘R_2Ins’, and ‘R_2Ins’), and 250 nM of each probe [‘*Rvi2*_ref_Pr (YAK)’ and ‘PrIns_*Rvi2* (FAM)’]. The PCR programme used was: 10 min at 95 °C, followed by 45 cycles of 10 s at 95 °C and 90 s at 60 °C. Reactions were run on BioRad Hard-Shell 480 PCR plates (384-well) sealed with BioRad Microseal ‘C’ Film in conjunction with fluorescence measurements using the Roche LightCycler 480 II System in ‘Dual Color Hydrolysis Probe’ detection format [(FAM (465–510) and VIC/HEX/Yellow555 (533–580) filters] and analysed via the end-point genotyping.

Targeted-sequencing enrichment and *de novo* assembly of *Rvi2*-associated 10 041 bp LTR-RT insertion

Long-range PCR was performed with PrimerSTAR[®] GXL DNA Polymerase (TaKaRa) following the protocol in the ‘Marker development’ section but using 5 min extension time. The 10 041 bp fragment was extracted from a 2% agarose gel using a Qiagen QIAquick Gel Extraction Kit (#QIAG28704). Nanopore sequencing libraries were prepared with a SQK-LSK109 ligation kit (ONT). The workflow began with an end-repair and nickase treatment using 3.5 µl Ultra II end-prep reaction buffer, 1.5 µl Ultra II end-prep enzyme mix (NEB), and 24.5 µl of PCR product (>50 ng µl⁻¹ or 50–100 fmol), incubated at 20 °C for 5 min then 65 °C for 5 min. After a bead-based clean-up with AMPure XP beads, sequencing adapters were ligated in a reaction containing 30 µl PCR product, 12.5 µl Ligation Buffer (ONT), 5 µl NEBNext Quick T4 DNA Ligase, and 2.5 µl AMX Adapter Mix (ONT), incubated at room temperature for 10 min. A final clean-up using Long Fragment Buffer (ONT) preceded library elution in 7 µl Elution Buffer (ONT) after a 10 min incubation at 37 °C. Libraries were sequenced on a MinION device using a Flongle Flowcell (R.9.4.1).

Base-calling from FAST5 to FASTQ was performed using Guppy v.6.4.4 (ONT) and Dorado v.0.3.0 (<https://github.com/nanoporetech/dorado>). Read quality was evaluated using NanoPlot v1.1.1 (De Coster *et al.*, 2018), and BAM to FASTQ conversion was done using samtools v.1.9 (Danecek *et al.*, 2021). Reads were filtered for a minimum length of 5 kb and an average quality ≥10 using seqkit v2.4.0, (Shen *et al.*, 2016). Guppy- and Dorado-derived reads were independently *de novo* assembled using Canu v.2.1.1 using parameters: minReadLength=2000, contigFilter=“2 0 1.0 0.5 0”, useGrid=false, genomeSize=11k with -nanopore-raw flag. Contigs from both assemblies were pairwise aligned using Geneious Primer[®] 2022.0.1. Additionally, both sets of base-called reads were aligned to their respective Canu contigs using minimap2 v2.9 (Li, 2018) and visualized with IGV 2.10.0 (Robinson *et al.*, 2017).

Haplotype-based characterization of the *Rvi2* locus in the Russian apple R12740-7a genome

To identify the haplotype linked to *Rvi2* scab resistance, the *de novo*-assembled 10 041 bp LTR-RT insertion was aligned to both

haplotypes of the R12740-7A genome using the Nucmer (Kurtz *et al.*, 2004). Alignment outputs were filtered with delta-filter and converted to a tab-separated file using show-coords. SyRI (Goel *et al.*, 2019) was then employed to pinpoint the location of the insertion between the two haplotypes.

For further characterization of the insertion, HiFi reads from R12740-7A and ‘Gala’, as well as ONT reads from TSR34T15, were aligned to the R12740-7A haplotype carrying the insertion using minimap2 (Li, 2018). The *Rvi2* locus coordinates were defined by BLASTing the flanking marker primer sequences (*Rvi2*_31 573 921 and *Rvi2*_31 934 769), against R12740-7A.

Small and structural variants within the *Rvi2* locus were called, and candidate variants common to R12740-7A and TSR34T15 but absent in ‘Gala’ were identified using Freebayes v1.3.6 (Garrison and Marth, 2012, Preprint). SnpEff v4.3 (Cingolani *et al.*, 2012) was used to predict the functional impact of these variants, classifying them as low (e.g. synonymous), moderate, or high impact (affecting splice-sites or start or stops codons).

Additionally, resistance genes within the *Rvi2* locus were annotated using the GenePal pipeline in both R12740-7A haplotypes and in the ‘Gala’ genome. Genes unique to the R12740-7A haplotypes were identified, and the coding sequences of shared genes were compared using multiple sequence alignment in Geneious prime 2022.01.1 (<https://www.geneious.com>).

Gene expression analysis

Plant material and inoculation

Seedlings at ~1 month old from a ‘Royal Gala’ × TSR34T15 cross and from a cross between *Rvi2* heterozygous parents were inoculated with *V. inaequalis* isolate MNH120 at 3×10⁵ conidia ml⁻¹. Inoculation was performed via mist spraying and incubation under conditions previously described by Gardiner *et al.* (1996). Based on the *Rvi2*-ELG-10kb assays, we selected 15 seedlings that were susceptible (lacking the LTR-insertion), 15 that were heterozygous resistant, and 15 that were homozygous resistant. Each sample consisted of a single young leaf (~100 mg) from a different seedling. Three biological replicates were collected per genotype during the scab inoculation test. Leaf samples were taken at four time points: just before inoculation (0 h) and at 3, 24, and 120 h post-inoculation. These samples were used for expression analysis (4 time points × 3 genotypes × 3 biological replicates = 36 seedlings).

RNA isolation and cDNA synthesis

The collected leaf samples were snap-frozen in an ethanol-dry ice bath and stored at -80 °C. RNA was extracted using a Spectrum Plant Total RNA Kit (Sigma-Aldrich) with an on-column DNase I digestion step, following the manufacturer’s protocol. RNA concentration and quality were assessed using a Qubit RNA BR assay (ThermoFisher) and capillary electrophoresis using an Agilent DNF-471 RNA Kit (15nt). Samples with a quality score >8 were processed for cDNA synthesis using a Superscript IV Kit (ThermoFisher) with Oligo dT priming. Control reactions with no RNA and no RT were included to monitor contamination.

Quantitative real-time PCR analyses

Gene expression analysis was performed using a Roche LightCycler[®] 480 with Roche SYBR Green 1 Master (Cat. #04707516001). Cycling conditions were 95 °C for 5 min, followed by 45 cycles of 95 °C for 10 s, 64–57 °C (−0.5 °C per cycle) for 10 s and 72 °C for 10 s, with a single acquisition per cycle. Dissociation curve analysis consisted of 95 °C for 5 s, 65 °C for 60 s, and gradual heating to 97 °C with continuous acquisition. The reference and target coding sequences were confirmed by cloning and Sanger sequencing from the seedling background. Intron–exon boundaries were

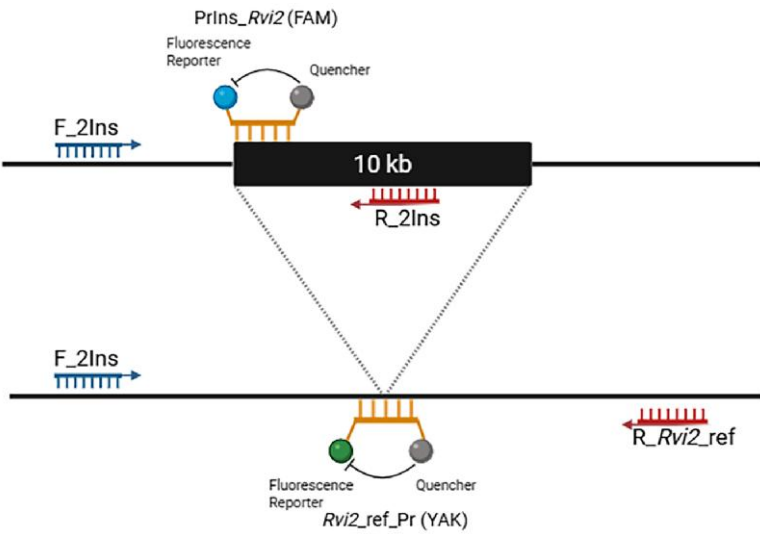


Fig. 1. Graphic representation of the *Rvi2*-ELG-10kb assay. This 5' fluorogenic hydrolysis probe assay targets a large insert highly associated with *Rvi2* resistance using two probes and three primer sequences. The latter consist of one forward-primer ('F_2Ins') and two reverse-primer sequences ('R_2Ins' and 'R_Rvi2_ref'). The primer pair 'F_2Ins' and 'R_Rvi2_ref' amplifies the reference genome allele, while 'F_2Ins' and 'R_2Ins' amplifies the insert allele. The 'PrIns_Rvi2 (FAM)' probe is specific for the 5' reference/insert boundary and the 'Rvi2_ref_Pr (YAK)' probe is specific for the immediate 3' and 5' sequences flanking the insert. The probes are designed so that the reporter fluorophores are on the 5'-end and the quencher is on the 3'-end.

Table 1. Primer sequences of the 5'-fluorogenic hydrolysis probe *Rvi2*-ELG-10kb assay

Name	Sequence	Tm (°C)
F_2Ins	GGACTTACCTTTTTTCGGCTTATC	61.9
R_Rvi2_ref	CCAGGTAGACGTGATTGTGAG	61.9
R_2Ins	CAACACCTTACGGAGATCGAATC	62.6
Rvi2_ref_Pr (YAK)	CGTCTGATATAGT+A+TT+AA+TC+TCG+AT	64.8
PrIns_Rvi2 (FAM)	TGATATAG+T+T+G+T+GAACACG	67.1

Common sequences between the wild type and the insert probes are underlined. Bold indicates bases different to the wild-type probe. '+' indicates the base has a locked nucleic acid modification. Tm, melting temperature.

based on the 'Golden Delicious' genome GDDH13v1.1 (Daccord et al., 2017). Primers were designed using PrimerQuest (IDT, USA) with at least one primer spanning an intron-exon junction. Negative controls (no cDNA and no RT) were included, and gene expression was normalized using two or three reference genes: *LIPID TRANSPORT LIKE-1* (*LTL1*; MDP0000173025), *PHYTOCHROME-ASSOCIATED PROTEIN PHOSPHATASE-3* (MDP0000060858), and *CASEIN KINASE SUBUNIT-4* (*CKB4*; MDP0000095375) (Bowen et al., 2014). Primer efficiencies and relative expression values for three technical replicates per sample were calculated using the Roche 480 Light Cycling software.

The effects of two factors on candidate gene expression were analysed, namely time after *V. inaequalis* inoculation (0, 3, 24, and 120 h) and the presence of the *Rvi2* allele (homozygous/heterozygous by *Rvi2*-ELG-10kb assay). A two-way ANOVA was performed in R v.2023.12.1 to assess the effects on normalized expression levels. Log-transformation ensured normality and homoscedasticity, with residuals checked using the 'predictmeans' R package (<http://dx.doi.org/10.32614/CRAN.package.predictmeans>). Pairwise comparisons against the baseline (0 h) were conducted using multivariate *t*-tests and Dunnett's test.

For genes with non-detectable expression, a binomial generalized linear model was applied, considering *Rvi2* allele presence and time as fixed factors. Expression data were visualized using the R package ggplot2 (Wickham, 2016).

Results

Scab phenotyping

Following inoculation with the *V. inaequalis* isolate MNH120, both the 'Royal Gala' × TSR34T15 and the Selection A (*Rvi2* + *Rvi6*) × 'Royal Gala' populations, segregating for *Rvi2* and *Rvi2*+*Rvi6*, respectively, exhibited the distinct SN resistance response characteristic of *Rvi2*. In the 'Royal Gala' × TSR34T15 population (*n*=361), 159 individuals were classified as resistant, 195 as susceptible, and seven as asymptomatic (Table 2). In the Selection A (*Rvi2*+*Rvi6*) × 'Royal Gala' population (*n*=110), 50 showed SN for *Rvi2*, and 60 were susceptible or showed chlorotic symptoms conditioned by *Rvi6*. Chi-square analysis confirmed that *Rvi2* resistance segregation in both populations followed a 1:1 resistant/susceptible ratio (*P* < 0.05), consistent with the monogenic inheritance of *Rvi2* as previously reported (Bus et al., 2005b).

Linkage map construction using the 'Royal Gala' × TSR34T15 cross

A subset of 188 seedlings from the total population of 361 of the 'Royal Gala' × TSR34T15 progeny were selected, comprising 90 resistant, 94 susceptible, and four asymptomatic individuals together with both parental cultivars, and were genotyped using the International RosBREED (IRSC) apple

Table 2. Mendelian segregation analysis of *Rvi2*-mediated resistance in two apple families inoculated with *Venturia inaequalis* isolate MNH120

Female parent	Male parent	No. of seedlings Total	Segregation					P
			R	S	0	χ^2		
'Royal Gala'	TSR34T15	361	159	195	7	3.661		0.055
Selection A (<i>Rvi2</i> + <i>Rvi6</i>)	'Royal Gala'	110	50	60	0	0.90909		0.340

R, resistant; S, susceptible; 0, asymptomatic.

8k SNP array v1 (Chagné *et al.*, 2012), which includes 7867 *Malus* and 921 *Pyrus* SNPs. Among the SNPs identified, 646 and 515 were polymorphic and segregated in a backcross fashion in 'Royal Gala' and TSR34T15, respectively. Additionally, 1748 SNPs were heterozygous in both parents. The polymorphic markers were assigned to the expected 17 linkage groups (LGs), with LG2 in TSR34T15 containing 73 loci, including the *Rvi2* trait locus, spanning a genetic distance of 95.3 cM (Fig. 2). The *Rvi2* locus was flanked by markers RosBREEDSNP_SNP_GA_30846235_Lg2_263337_MAF20_263337_exon_2 and RosBREEDSNP_SNP_CT_34033084_Lg2_RosCOS1019_MAF20_MDP0000184243_exon. This region corresponds to a physical interval between 27 962 381 bp and 32 257 043 bp on the 'Golden Delicious' genome (GDDH13v1.1; Daccord *et al.*, 2017).

Genome assembly of Russian apple R12740-7A

A total of 2 697 310 HiFi reads (38 Gb) were generated, with a read N50 of 14 578 nucleotides (NTs) and a maximum read length of 50 031 NTs. Genome assembly metrics are summarized in Table 3, showing the key characteristics of the primary assembly and phased haplotypes. These include the number of scaffolds, scaffold size range, total assembly size, and the N50 values.

The final curated assembly was free of adapters and non-*Malus* contaminations. It consisted of 17 chromosomes along with unanchored scaffolds in Rhap1 (682.2 Mb, including 11 gaps of 100 NTs each), Rhap2 (655.7 Mb, 10 gaps) and Rp (683.6 Mb, five gaps). Using the eudicot-s_odb10 database, BUSCO completeness scores were 98.6, 99.0, and 98.9% for Rhap1, Rhap2, and Rp, respectively. The LAI values were 21.0 for Rhap1, 22.0 for Rhap2, and 21.3 for Rp.

Gene annotation was performed using GenePal and is summarized in Table 3.

Genome assembly of the TSR34T15 *Rvi2*-resistance accession

We generated a *de novo* genome assembly for TSR34T15, which is a *Rvi2* scab differential host (Bus *et al.*, 2005b). A total

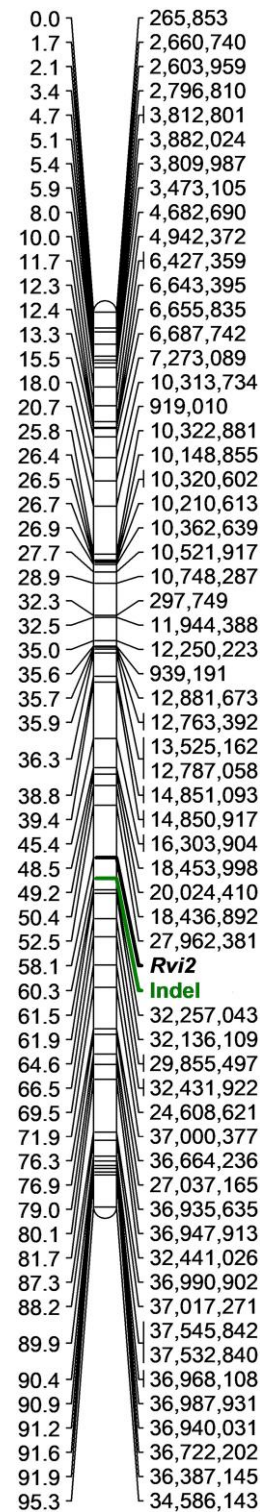
**Fig. 2.** Linkage group 2 of TSR34T15, an F₂ selection from Russian apple R12740-7A (*Malus pumila*), showing the genetic map position of the *Rvi2* locus. The map was constructed using polymorphic SNP markers from the International RosBREED SNP Consortium (IRSC) apple 8K SNP array v1. Genetic distances (cM) are indicated along the linkage group. 'Indel' indicates the long terminal repeat retrotransposon insertion associated with *Rvi2* in several apple selections.

Table 3. Summary of genome assembly characteristics of the Russian apple R12740-7A

Metric	Rhap1	Rhap2	Rp
No. of seqs	17 chromosomes + 467 contigs	17 chromosomes + 455 contigs	17 chromosomes + 79 contigs
N50	39 037 833	38 314 341	37 988 491
Shortest	10 351	11 771	23 137
Longest	14 330 285.00	14 127 199.00	14 330 285.00
Total length	683 574 724.00	655 742 921.00	682 154 938.00
LAI	21.30	21.00	22.00
Count of genes	50 270	48 426	49 541
mRNA	57 150	55 115	55 902
rRNA	89	12	92
tRNA	1016	421	1010
% genome that is exons	9.43	9.42	9.28
% genome that is genes	15.74	15.96	15.41
% genome that is introns	10.03	10.31	9.57
BUSCO eudi- cots n:2326	C:96.1% [S:50.9%, D:45.2%], F:1.0%, M:2.9%	C:96.0% [S:52.0%, D:44.0%], F:0.9%, M:3.1%	C:95.7% [S:51.7%, D:44.0%], F:1.2%, M:3.1%
BUSCO em- briophyta n:1614	C:96.5% [S:52.9%, D:43.6%], F:1.2%, M:2.3%	C:96.3% [S:53.2%, D:43.1%], F:1.1%, M:2.6%	C:95.7% [S:53.5%, D:42.2%], F:1.5%, M:2.8%

Rhap1 and Rhap2 are the phased haplotype assemblies, Rp is the primary genome assembly. N50 is the scaffold length for which 50% of the total length is in scaffolds of equal or greater size. LAI is the LTR Assembly Index.

of 2.29 million ONT raw sequencing reads were generated, with an average length of 27.12 kb and an N50 of 36.34 kb. Approximately 80% of the reads exceeded 10 kb, providing a total of 62.29 Gb of sequence data. After read correction, 667 233 reads were retained, with an average length of 43.23 kb and an N50 of 45.54 kb, resulting in 28.88 Gb of data. A 21-mer analysis showed 56.63-fold genome coverage, with an estimated genome size of 510 Mb, a heterozygosity rate of 0.69% and a repeat content of 41.56%.

We compared the performance of three widely used genome assemblers, Flye, Shasta, and Canu, using Canu-corrected ONT data. Raw Flye and Shasta assemblies were polished with ONT reads using Medaka or Racon. Assembly statistics are summarized in [Supplementary Table S2](#). Medaka polishing slightly improved the BUSCO scores for the raw Flye assembly, reducing its size by ~13 Mb without altering scaffolds numbers, while having no effect on the raw Shasta assembly. In contrast, Racon polishing produced shorter assemblies for both Flye and Shasta, with lower BUSCO scores than Medaka. Canu generated the longest assembly (1.62 Gb across 8664 contigs) with the highest N50 (187.24 kb) and BUSCO completeness (98.8%), making it the most complete and accurate choice for further analysis. The

Canu assembly was then scaffolded into 17 chromosomes using Ragtag ([Alonge et al., 2022](#)), aligning it with the ‘Golden Delicious’ GDDH13v1.1 ([Daccord et al., 2017](#)) and HFTH1 genomes ([Zhang et al., 2019](#)). The scaffolded genomes measured 560 Mb and 635 Mb, respectively, with BUSCO completeness scores of 83.4% and 88.1%, respectively. The final TSR34T15 assembly, based on the HFTH1 genome was selected for its higher completeness and lower missing BUSCO scores (9.4%). It also showed greater contiguity, with an N50 of 34.1 Mb, consolidating half the genome into just six chromosomes.

Fine mapping of the *Rvi2* locus

Our analysis of small variants within the *Rvi2* locus (27.9 Mb to 32.3 Mb on chromosome 2 of the ‘Golden Delicious’ GDDH13v1.1 genome) identified 2445 *in silico* SNPs present in TSR34T15 and R12740-7A but absent in ‘Gala’ alignment files. We selected evenly distributed SNPs across the locus as candidates for SNP-HRM-markers. To validate these markers, we tested the corresponding primer pairs in both parents and a subset of the F₁ ‘Royal Gala’ × TSR34T15 seedlings. Primer pairs that failed to amplify efficiently or did not yield the expected heterozygous genotype in the scab-resistant TSR34T15 parent (consistent with dominant inheritance) were discarded. Six HRM markers with optimal performance and even distribution within the *Rvi2* interval were selected for further genotyping. Additionally, we incorporated two previously established SSR markers (CH05e03 and SSR_MDC016328_315) and one TaqMan marker (AHI15IL). These markers were used to genotype 173 F₁ seedlings.

Recombination was observed in 12 seedlings, with five classified as resistant and seven as susceptible. Re-phenotyping confirmed their correct classification. The recombination break-points localized the *Rvi2* genomic region to a ~303 kb interval (31 630 783 bp to 31 934 769 bp) specifically between marker SSR_MDC016328_315 and marker *Rvi2*_31 934 769 on the ‘Golden Delicious’ GDDH13v1.1 genome ([Table 4](#)). The lower boundary was defined by five susceptible individuals showing recombination between *Rvi2*_31 473 921 and *Rvi2*_31 630 783, while the upper boundary was defined by five resistant and five out of seven susceptible individuals showing recombination between *Rvi2*_31 780 921 and *Rvi2*_31 934 769.

We identified a total of 156 structural variants (SVs) within the fine-mapped *Rvi2* region on TSR34T15 read alignments against the ‘Golden Delicious’ GDDH13v1.1 reference genome ([Daccord et al., 2017](#)). Most were deletions (80, 51.28%) and insertions (54, 34.62%) over 20 base pairs. Additionally, we detected translocation break-points (16, 10.26%), inversions (5, 3.21%), and one duplication (0.64%).

The region contains 41 annotated genes in the GDDH13v1.1 genome, including four with predicted functions in plant defence, one with a structural function, 24 involved in metabolism, and 11 of unknown function ([Supplementary Table S3](#)).

Table 4. Fine-mapping of the *Rvi2* locus through the identification of recombination break-points (underlined markers)

Marker	Position	Genotype										
<i>CH05e03</i>	30 748 692	h	h	m	m	m	m	m	m	m	h	m
<i>Rvi2_30 960 778</i>	30 960 778	h	h	h	m	m	m	m	m	m	h	m
<i>Rvi2_31 055 136</i>	31 055 136	h	h	h	m	m	h	h	m	m	m	m
<u><i>Rvi2_31 573 921</i></u>	31 573 921	h	h	h	h	h	h	h	h	h	m	m
SSR_MDC016328_315	31 630 783	h	h	h	h	h	m	m	m	m	m	m
AHI15IL	31 686 648	h	h	h	h	h	m	m	m	m	m	m
<i>Rvi2</i> -ELG-10 K	31 730 000	h	h	h	h	h	m	m	m	m	m	m
<i>Rvi2_31 777 875</i>	31 777 875	h	h	h	h	h	m	m	m	m	m	m
<u><i>Rvi2_31 780 921</i></u>	31 780 921	h	h	h	h	h	m	m	m	m	m	m
<i>Rvi2_31 934 769</i>	31 934 769	m	m	m	m	m	h	h	m	h	h	h
Phenotype		R	R	R	R	R	S	S	S	S	S	S

m, allele from susceptible parent 'Royal Gala'; **h**, allele from heterozygote individual. R, resistant, S, susceptible.

The defence-related genes were *MD02G1262400*, encoding a Leucine-Rich Repeat Receptor-Like Kinase (LRR-RLK) involved in immune signalling; *MD02G1263000*, encoding an amino phospholipid ATPase1 (ALA1), also known as a lipid flippase; *MD02G1263700*, encoding a Tumour-necrosis Receptor-Associated Factor (TRAF) protein, which regulates immune responses; and *MD02G1264800*, encoding a TIR-NBS-LRR plant resistance gene associated with pathogen recognition and defence activation. Additionally, *MD02G1262900*, encoding a DNA helicase, although not a resistance gene, contributes to metabolism and can influence plant defence mechanisms.

Among the 156 SVs identified within the *Rvi2* locus for the TSR34T15 ONT-read-alignments against the 'Golden Delicious' GDDH13v1.1 reference genome, 16 were predicted to affect four of the five annotated genes related to resistance. *MD02G1262900* harboured a 9.8 kb upstream insertion with a predicted modifier effect. *MD02G1263000* had multiple variants, namely a 51 bp deletion resulting in a high-impact frameshift, a 32 bp downstream insertion with a modifier effect, and a 3649 bp insertion located between this and another gene (*MD02G1263100*), also with a modifier effect. *MD02G1263700* carried five upstream SVs with modifier effects and a 152 bp coding insertion predicted to cause a high impact stop-gained mutation. Finally, *MD02G1264800* showed five upstream SVs with modifier effects and a 72 bp intragenic deletion that induced high-impact disruption (Supplementary Table S4). However, only the 9.8 kb upstream insertion in *MD02G1262900* was also detected in the Russian apple accession R12740-7A, where it measured 10 041 bp, and it was absent in non-*Rvi2* resistant cultivars such as 'Gala', used as a negative control.

Variant-calling was also conducted on the fine-mapped *Rvi2* locus using the newly phased diploid Russian apple R12740-7A genome, which is the source of *Rvi2* resistance, as the reference. The *Rvi2* region spanned chromosome 2, from 32 418 227 bp to 32 693 564 bp on haplotype 1 and 32 748 312 bp to 33 277 842 bp on haplotype 2, with marker positions from the

GDDH13v1.1 genome converted to the corresponding coordinates in both haplotypes (Supplementary Table S5). Genotyping analysis using these fine-mapping markers indicated that the *Rvi2* allele was present in haplotype 2. Consequently, variant-calling focused solely on this haplotype, which was used as reference genome. Variants were selected based on being heterozygous (0/1) in R12740-7A, either homozygous reference (0/0) or heterozygous (0/1) in TSR34T15, and homozygous alternative (1/1) for 'Gala'.

A total of 2216 variants were identified in TSR34T15, R12740-7A, and 'Gala'. Most (2119, 95.62%) were predicted to have a modifier effect (Supplementary Table S6), primarily affecting non-coding regions. In contrast, only two variants (0.09%) were classified as high-impact, 38 (1.72%) as moderate-impact, and 57 (2.57%) as low-impact. The two high-impact variants are predicted to disrupt protein function. The first was a heterozygous 59 bp deletion located at position 32 843 623 on haplotype 2 of Russian apple R12740-7A, resulting in three predicted effects, namely frameshift_variant, stop_lost, and splice_region_variant. This variant affected gene 4884, annotated as an F-box protein. The second was a heterozygous 10 041 bp insertion at position 32 992 843 on haplotype 2, resulting in the predicted loss of two genes (4895 and 4896) in accessions where the insertion is absent. These genes were annotated as SWI SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like and an FPP/GGPP synthase family member, respectively.

Development and validation of a new *Rvi2* marker assay

To investigate the association between the presence of the two high-impact variants and the *Rvi2*-resistance, we performed *in silico* analyses of several publicly available phased genomes. This ruled out the 59 bp deletion as being linked to resistance and confirmed that only the 10 041 bp insertion was exclusively present in the R12740-7A and TSR34T15 genomes.

Subsequently, we developed a 5' fluorogenic hydrolysis probe assay, termed *Rvi2*-ELG-10kb, to detect the 10 041 bp insertion. Screening of 239 apple selections and cultivars with this assay revealed a perfect genotype–phenotype correlation: all 116 *Rvi2*-resistance selections were heterozygous, while all 123 *Rvi2*-susceptible selections were homozygous for the reference allele (Supplementary Table S1).

The marker assay was further applied to an F₁ population of 137 seedlings derived from a selection A (*Rvi2*+*Rvi6*) × ‘Royal Gala’ cross previously phenotyped for scab resistance. Among these seedlings, 54 exhibited a chlorotic *Rvi6*-resistance response corresponding to a heterozygous genotype for the *Rvi6* AHZAHRG marker (Chagné *et al.*, 2019). Of the 77 seedlings showing SN resistance for *Rvi2*, 50 were heterozygous for the *Rvi2*-ELG-10kb assay. Combining both markers allowed us to identify 27 SN seedlings that also carried the *Rvi6* gene, with their phenotype showing only the SN response owing to the epistasis over the chlorotic response of the *Rvi6* plants. All six seedlings with strong sporulating symptoms had homozygous susceptible genotypes for both the *Rvi6* and *Rvi2* markers.

Sequencing and characterization of an *Rvi2*-associated LTR-RT insertion

To characterize the sequence content of the insertion, we conducted long-range PCRs to amplify the entire 10 041 bp sequence in TSR34T15, R12740-7A, ‘Royal Gala’, an *Rvi2*-susceptible seedling, and an *Rvi2*-resistance and homozygous seedling, as determined by the *Rvi2*-ELG-10kb assay. TSR34T15 and R12740-7A produced two distinct bands (~10 kb and ~200 bp), confirming the presence of the same 10 041 bp insertion sequence on one haplotype and its absence on the other. In contrast, ‘Royal Gala’ and the susceptible seedlings (*rvi2/rvi2*) showed only the ~200 bp band, while the homozygous *Rvi2*-resistance seedling (*Rvi2/Rvi2*) showed only the ~10 kb band (Fig. 3), indicating the presence of the insertion on both haplotypes. The *Rvi2* homozygous seedlings were identified from a cross between *Rvi2* heterozygous parents from the breeding programme using the *Rvi2*-ELG-10kb assay.

The larger bands from TSR34T15, R12740-7A, and the homozygous *Rvi2*-resistance seedling (*Rvi2/Rvi2*) were excised and sequenced individually using separate Flongle ONT flow cells. Multiple-alignment analyses, using two base-callers per sample, confirmed identical sequence content across all three genotypes, and the consensus sequence was validated by alignment to the R12740-7A genome assembly, locating it on haplotype 2.

The 10 041 bp sequence corresponded to a LTR-RT with two flanking 1001 bp Ty3 LTRs at its 5'- and 3'-ends (TE_00009794 LTRs) and a poorly conserved motif ‘QGXXEXXXFXLXXH’ typical of Retroviridae gag-proteins following the 5'-LTR. Additionally, the two captured

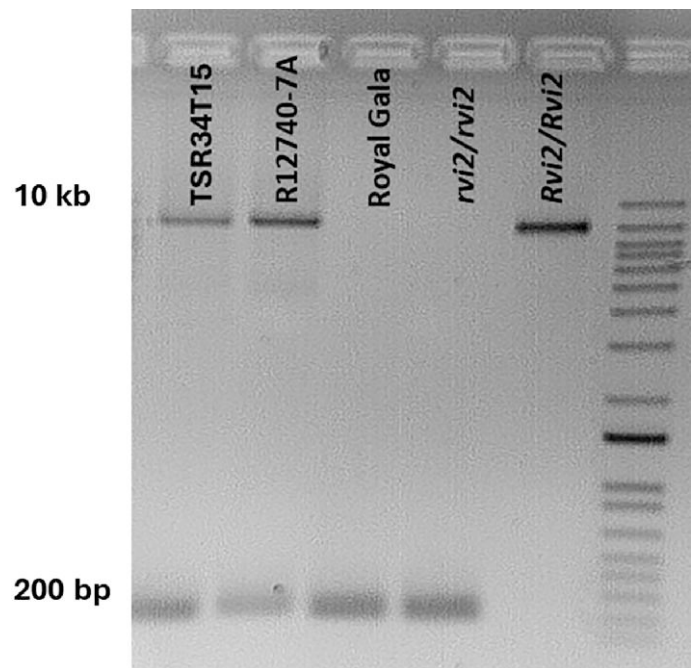


Fig. 3. Agarose gel visualization of the long-range PCR products of the long terminal repeat retrotransposon (LRT-RT) insertion associated with *Rvi2* in several apple selections. TSR34T15, Russian apple R12740-7A (*Malus pumila*); ‘Royal Gala’ (*M. domestica*), and *rvi2/rvi2* (susceptible) and *Rvi2/Rvi2* (resistant) seedlings. The standards in the last lane range from 100 bp to 15 kb. The 10 kb band corresponds to the presence of the *Rvi2*-associated LRT-RT and the 200 bp band to its absence.

genes (gene 4895 and gene 4896) within the insertion were annotated as encoding a Switch/Sucrose on-fermentable (SWI/SNF) ATP-dependent chromatin remodelling complex subunit and a farnesyl pyrophosphate and/or geranylgeranyl pyrophosphate FPP/GGPP synthase family (FPPS), respectively. Three extra single LTRs were also annotated within the insertion (Fig. 4).

Expression analysis of candidate genes

Based on the functional annotations suggesting roles in disease response within the *Rvi2* region of both the ‘Golden Delicious’ genome GDDH13v1.1 (Daccord *et al.*, 2017) and haplotype 2 of the newly assembled R12740-7A genome, we evaluated the relative normalized expression profiles of six candidate genes and identified their physical positions in both genomes (Supplementary Table S7). The analysis considered the effects of time post-scab inoculation and the presence or absence of the *Rvi2* allele, as determined by the *Rvi2*-ELG-10kb assay.

Two candidate loci were gene 4895 and gene 4896, located within the high-impact 10 041 bp insertion. Because gene 4895 spans only 482 bp and encodes a truncated SWI/SNF-related, matrix-associated, actin-dependent chromatin-remodelling factor, it lacks the full complement of functional



Fig. 4. Sequence of the long terminal repeat retrotransposon (LRT-RT) insertion site associated with *Rvi2* in haplotypes 1 and 2 of the Russian apple R12740-7A genome assembly.

domains necessary to influence phenotype and was therefore excluded from further analysis, leaving the *FPPS* gene 4896 as the focus. As expected, in susceptible (*rvi2/rvi2*) ‘Royal Gala’ × TSR34T15 seedlings this gene was absent and not expressed, while resistant seedlings with heterozygous (*Rvi2/rvi2*) and homozygous (*Rvi2/Rvi2*) genotypes retained one and two copies, respectively. Three of the candidate genes (*MD02G1263000*, *MD02G1264800*, and *FPPS*) were significantly influenced by time post-inoculation, with the most notable changes at 120 h ($P < 0.001$; Fig. 5). *MD02G1263000* was down-regulated at 3 and 24 h followed by robust up-regulation at 120 h ($P < 0.0001$) in all the genotypes. In contrast, *MD02G1264800* showed pronounced early down-regulation in the *rvi2/rvi2* and *Rvi2/rvi2* genotypes (0–24 h) but only *Rvi2/Rvi2* seedlings showed significant up-regulation at 120 h. The *FPPS* gene maintained stable expression across all time-points in both the *Rvi2/rvi2* or *rvi2/rvi2* genotypes, with negligible changes.

Discussion

The development of cultivars with durable resistance to apple scab (*V. inaequalis*) is essential to reduce reliance on fungicide and ensure sustainable apple production. In line with this goal, the global apple and pear industry is striving for spray-free orchards through the adoption of sustainable practices. A key strategy involves breeding elite cultivars with multiple resistance genes to enhance durability and minimize the risk of break-down of single-gene resistance (Khan and Korban, 2022; Khan *et al.*, 2024).

This study provides comprehensive insights into the genetic architecture of *Rvi2*-mediated scab resistance. By integrating phenotyping, high-density linkage mapping, genome assembly, fine-mapping, and candidate gene expression analyses, we have refined the candidate region and elucidated the potential mechanisms.

The *Rvi2* gene was initially mapped to a 1.2 cM region on chromosome 2 (Chagné *et al.*, 2019). Using the IRSC *Malus* SNP array, we confirmed its location on linkage group 2 in a ‘Royal Gala’ × TSR34T15 cross (Fig. 2). We first localized *Rvi2* to a 4.3 Mb interval in the ‘Golden Delicious’ genome GDDH13v1.1 (Daccord *et al.*, 2017) and further narrowed it to a ~330 kb region using fine-mapping with HRM, SSR, and TaqMan markers. This precise localization was critical for identifying candidate genes associated with the resistance.

To explore structural variations at the *Rvi2* locus, we generated high-quality genome assemblies for Russian apple R12740-7A and the *Rvi2*-resistance accession TSR34T15 using PacBio HiFi and Oxford Nanopore Technologies, respectively. These assemblies, validated with BUSCO and LAI scores, provided a robust framework for variant discovery. We identified a 59 bp deletion affecting an F-box gene and a notably large 10 041 bp insertion. This insertion contained two genes, including one from the *FPP/GGPP synthase* (*FPPS*) family. This insertion was exclusively found in resistant genotypes (*Rvi2/rvi2* and *Rvi2/Rvi2*) and was validated through a newly developed fluorogenic hydrolysis probe assay (*Rvi2*-ELG-10kb), which showed perfect correlation with *Rvi2*-mediated resistance. This marker-assisted selection assay will enhance breeding efficiency and, when combined with markers for *Rvi6* or other *Rvi* genes discovered in the future, it will facilitate gene pyramiding to strengthen resistance durability.

Long-range PCR and sequencing revealed that the insertion is an LTR-RT with two flanking *Ty3* LTRs (Fig. 4). Retrotransposons can influence genome evolution and gene regulation by inserting near or within genes and altering local chromatin structure (Tossolini *et al.*, 2025). In this case, the insertion captured two genes: a SWI/SNF ATP-dependent chromatin-remodelling factor and a gene that is part of the farnesyl pyrophosphate/geranylgeranyl pyrophosphate (*FPP/GGPP*) synthase family (*FPPS*). Notably, the *FPPS* gene is located only ~2 kb from the 3'-UTR of *MD02G1263000*, suggesting

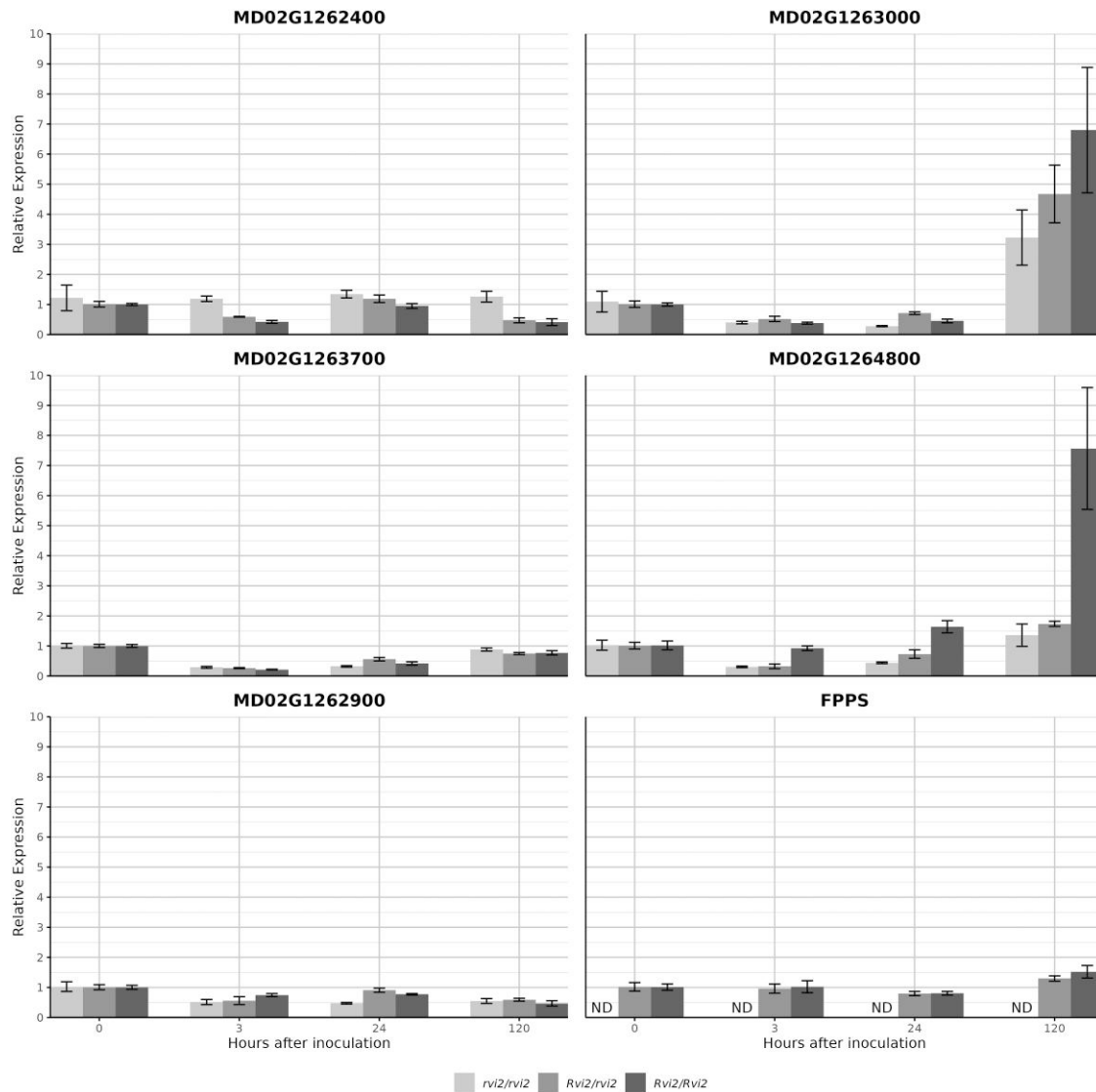


Fig. 5. Gene expression profiles in leaves of apple seedlings of different *Rvi2* genotypes following inoculation with *Venturia inaequalis*. Susceptible (*rvi2/rvi2*), heterozygous resistant, (*Rvi2/rvi2*), and homozygous resistant (*Rvi2/Rvi2*) plants were used and candidate genes found within the *Rvi2* locus on haplotype 2 of Russian apple R12740-7A genome were analysed by quantitative real-time PCR. Expression is relative to that in the resistant genotypes, the values of which were set as 1. ND, not detected. FPPS, farnesyl pyrophosphate/geranylgeranyl pyrophosphate (FPP/GGPP) synthase. Data are means ($\pm$ SE), $n=3$ biological replicates.

a potential *cis*-regulatory influence. The *MD02G1263000* gene is an amino phospholipid ATPase1 (*ALA1*), one of the 12 members within the P4 ATPase family; the largest subgroup of P-type ATPases are also called lipid flippases, with early studies indicating their role in plant fitness and tolerance to cold temperatures (Gomès *et al.*, 2000; Poulsen *et al.*, 2008; Nintemann *et al.*, 2019) and more recent studies have reported their involvement in development, reproduction, and signalling events as well as adaptation responses to biotic and abiotic stress (Pichersky and Gershenzon, 2002). Our expression analysis showed that *MD02G1263000* was significantly down-regulated at 3 h and 24 h post-inoculation across all genotypes, followed by robust

up-regulation at 120 h (Fig. 5). This late-stage up-regulation was strongest in homozygous resistant (*Rvi2/Rvi2*) seedlings, indicating a dosage effect and its potential role in a late-defence response. *MD02G1264800*, predicted to belong to the TIR-NBS-LRR class of plant resistance genes, also showed early down-regulation in susceptible (*rvi2/rvi2*) and heterozygous (*Rvi2/rvi2*) genotypes; however, significant up-regulation at 120 h post-inoculation was observed only in the homozygous resistant (*Rvi2/Rvi2*) seedlings. While *MD02G1264800* might contribute to defence signalling, its expression pattern does not fully align with the dominant inheritance of *Rvi2* resistance, suggesting it is not the primary determinant of resistance, but it might

enhance or sustain defence signalling in fully resistant plants. Genes of the FPP/GGPP synthase (*FPPS*/*GGPS*) family are central to the synthesis of terpenoid precursors, specifically farnesyl pyrophosphate/diphosphate (FPP) and geranylgeranyl pyrophosphate (GGPP). FPP serves as the direct substrate to produce sesquiterpenes involved in multiple biological functions in plants, including growth regulation and development, signalling in response to biotic and abiotic stress, and defence responses (Pichersky and Gershenzon, 2002). Beyond secondary metabolism, FPP and GGPP are also substrates for protein prenylation (Hála and Žárský, 2019), a post-translational modification that affects the localization and function of numerous proteins, including those involved in signal transduction pathways and biotic and abiotic stress responses. Recent studies support a defence-related role for FPP/GPP-derived metabolites. For example, in tea plants, the alternatively spliced terpene synthase (*TPS1-AS*) catalyses the formation of geraniol, a monoterpene. Its expression is strongly induced by pathogen infection, and silencing *TPS1-AS* reduces geraniol accumulation and increases susceptibility to disease, along with down-regulation of the pathogenesis-related defence genes *PR1* and *PR2*, demonstrating a direct role of monoterpene synthesis in plant defence (Jiang *et al.*, 2023). In potato, terpene synthase *TPS18* converts FPP into (*E,E*)-farnesol, a sesquiterpenoid that functions as an elicitor by activating phytosterol biosynthesis and defence gene expression, and transgenic expression of *StTPS18* also confers pathogen resistance in tobacco (Dwivedi *et al.*, 2022). In apple, suppression of the α -farnesene synthase gene *AFS1*, which converts FPP into α -farnesene, reduces α -farnesene accumulation and is associated with lower rates of disease initiation by post-harvest fungal pathogens, although lesion development in successful infections remains comparable (Souleyre *et al.*, 2019).

Although our expression analyses revealed that *FPPS* remained stably expressed over time in resistant genotypes (Fig. 5), its genomic location within the retrotransposon and its constitutive expression support a model in which it acts in *cis* to modulate local gene expression, possibly via changes in chromatin structure, signalling, or metabolite production for a rapid activation upon infection. Given that retrotransposons can restructure chromatin and influence gene accessibility, the *FPPS*-containing insertion could act as a regulatory hub priming neighbouring genes, particularly *MD02G1263000*, for a rapid and sustained defence response.

Our findings suggest that the *Rvi2*-associated 10 041 bp insertion, particularly the *FPPS* gene embedded within it, could play a role in shaping the regulatory landscape of resistance. While *MD02G1264800* probably contributes to defence activation, its expression dynamics suggest it is not the primary determinant of resistance. Instead, *MD02G1263000* emerges as a strong candidate for mediating *Rvi2*-dependent scab resistance through its role in membrane dynamics and stress adaptation.

Further functional validation, including use of CRISPR-knockouts, overexpression studies, and chromatin

accessibility assays, are needed to elucidate the precise regulatory mechanisms. Metabolomic profiling could also determine whether *FPPS* influences resistance via isoprenoid-derived signalling molecules. These insights will be invaluable for refining breeding strategies, ultimately accelerating the development of durable scab-resistant apple cultivars.

Conclusion

Breeding elite apple cultivars with durable scab resistance is critical because fungicide reliance is unsustainable, and *V. inaequalis* rapidly evolves to overcome single-gene resistance. Our study has provided a detailed dissection of the *Rvi2* locus by generating haplotype-phased and Oxford Nanopore genome assemblies. Fine-mapping identified a 10 041 bp LTR-RT insertion strongly linked to *Rvi2* that harbours an *FPPS* gene with a potential *cis*-regulatory role in defence gene priming. We have validated a structural variant marker derived from this insertion, integrating it into marker-assisted selection for breeding resistance to scab. This marker contributes to the rapid identification of scab-resistant seedlings, thereby accelerating the development of resistant apple cultivars and supporting more sustainable apple production.

Supplementary data

The following supplementary data are available at [JXB](#) online.

Table S1. Apple genotypes used to validate the *Rvi2*-ELG-10kb assay.

Table S2. Genome assembly statistics for TSR34T15 using different ONT assemblers.

Table S3. Annotated genes at the *Rvi2* locus on the GDDH13v1.1 reference genome.

Table S4. Structural variants (SVs) identified between TSR34T15 and the GDDH13v1.1 reference genome.

Table S5. Primer sequences and marker positions for *Rvi2* fine-mapping on the GDDH13v1.1 reference and R12740-7A genomes.

Table S6. Effects of variants detected in the *Rvi2* locus.

Table S7. Annotated and candidate genes at the *Rvi2* locus in R12740-7A.

Acknowledgements

This research used NCRIS-enabled Bioplatforms Australia infrastructure, and we thank the Biomolecular Resource Facility at the John Curtin School of Medical Research for performing the PromethION Oxford Nanopore sequencing. We thank Bruce Kingham and his team at the University of Delaware DNA Sequencing & Genotyping Center for assistance in generating high-quality PacBio and Hi-C data.

Author contributions

ELG, RKV, and VGMB secured funding for the project; ELG, RKV, VGMB, and DC conceived and designed the work; ELG, CK, and CHD contributed to the acquisition and analysis of data; AS and AK provided the tissue samples for Russian apple R12740-7A; ELG, RKV, VGMB, DC, and CHD wrote and prepared the manuscript; all authors read and approved the final manuscript.

Conflict of interest

The authors declare that they have no conflicts of interest in relation to this work.

Funding

This work was funded by Prevar Limited (<https://prevar.co.nz>) and the New Zealand Ministry of Business, Innovation and Employment (MBIE) Strategic Science Investment Fund (SSIF) allocated to the Bioeconomy Science Institute, New Zealand ('Technology Development' project).

Data availability

The datasets generated and/or analysed are available in the NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra>) under accession number SRR34368262 and the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA1286121.

References

- Aho AV, Kernighan BW, Weinberger PJ. 1979. Awk – a pattern scanning and processing language. Software: Practice and Experience **9**, 267–279.
- Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biology **23**, 258.
- Ashfield T, Egan AN, Pfeil BE, *et al.* 2012. Evolution of a complex disease resistance gene cluster in diploid *Phaseolus* and tetraploid *Glycine*. Plant Physiology **159**, 336–354.
- Barragan AC, Weigel D. 2021. Plant NLR diversity: the known unknowns of pan-NLRomes. The Plant Cell **33**, 814–831.
- Barragan CA, Wu R, Kim ST, *et al.* 2019. RPW8/HR repeats control NLR activation in *Arabidopsis thaliana*. PLoS Genetics **15**, e1008313.
- Bastiaanse H, Bassett HCM, Kirk C, Gardiner SE, Deng C, Groenworld R, Chagné D, Bus VGM. 2016. Scab resistance in 'Geneva' apple is conditioned by a resistance gene cluster with complex genetic control. Molecular Plant Pathology **17**, 159–172.
- Belfanti E, Silfverberg-Dilworth E, Tartarini S, Patocchi A, Barbieri M, Zhu J, Vinatzer BA, Gianfranceschi L, Gessler C, Sansavini S. 2004. The *HcrVf2* gene from a wild apple confers scab resistance to a transgenic cultivated variety. Proceedings of the National Academy of Sciences, USA **101**, 886–890.
- Boocock J, Chagné D, Merriman TR, Black MA. 2015. The distribution and impact of common copy-number variation in the genome of the domesticated apple, *Malus × domestica* Borkh. BMC Genomics **16**, 848.
- Botella MA, Parker JE, Frost LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, Holub EB, Jones JD. 1998. Three genes of the *Arabidopsis RPP1* complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. The Plant Cell **10**, 1847–1860.
- Boudichevskaia A, Flachowsky H, Dunemann F. 2009. Identification and molecular analysis of candidate genes homologous to *HcrVf* genes for scab resistance in apple. Plant Breeding **128**, 84–91.
- Bowen J, Ireland HS, Crowhurst R, *et al.* 2014. Selection of low-variance expressed *Malus × domestica* (apple) genes for use as quantitative PCR reference genes (housekeepers). Tree Genetics & Genomes **10**, 751–759.
- Bradnam KR, Fass JN, Alexandrov A, *et al.* 2013. Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. GigaScience **2**, 10.
- Bus V, Bradley S, Hofstee M, Alspach P, Brewer L, Luby J. 2000. Increasing genetic diversity in apple breeding to improve the durability of pest and disease resistance. Acta Horticulturae **538**, 185–190.
- Bus VGM, Laurens FND, Weg WEVD, Rusholme RL, Rikkerink EHA, Gardiner SE, Bassett HCM, Kodde LP, Plummer KM. 2005a. The *Vh8* locus of a new gene-for-gene interaction between *Venturia inaequalis* and the wild apple *Malus sieversii* is closely linked to the *Vh2* locus in *Malus pumila* R12740-7A. New Phytologist **166**, 1035–1049.
- Bus V, Rikkerink E, Aldwinckle HS, *et al.* 2009. A proposal for the nomenclature of *Venturia inaequalis* races. Acta Horticulturae **814**, 739–746.
- Bus VGM, Rikkerink EHA, Caffier V, Durel C-E, Plummer KM. 2011. Revision of the nomenclature of the differential host–pathogen interactions of *Venturia inaequalis* and *Malus*. Annual Review of Phytopathology **49**, 391–413.
- Bus VGM, Rikkerink EHA, Van De Weg WE, *et al.* 2005b. The *Vh2* and *Vh4* scab resistance genes in two differential hosts derived from Russian apple R12740-7A map to the same linkage group of apple. Molecular Breeding **15**, 103–116.
- Bus V, van de Weg WE, Durel CE, *et al.* 2004. Delineation of a scab resistance gene cluster on linkage group of apple. Acta Horticulturae **663**, 57–62.
- Caffier V, Lasserre-Zuber P, Giraud M, *et al.* 2014. Erosion of quantitative host resistance in the apple *Venturia inaequalis* pathosystem. Infection, Genetics and Evolution **27**, 481–489.
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Molecular Biology and Evolution **38**, 5825–5829.
- Chagné D, Crowhurst RN, Troggio M, *et al.* 2012. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS ONE **7**, e31745–e31745.
- Chagné D, Gasic K, Crowhurst RN, *et al.* 2008. Development of a set of SNP markers present in expressed genes of the apple. Genomics **92**, 353–358.
- Chagné D, Vanderzande S, Kirk C, Profitt N, Weskett R, Gardiner SE, Peace CP, Volz RK, Bassil NV. 2019. Validation of SNP markers for fruit quality and disease resistance loci in apple (*Malus × domestica*, Borkh.) using the OpenArray® platform. Horticulture Research **6**, 30.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. Nature Methods **18**, 170–175.
- Chevalier M, Lespinasse Y, Renaudin SC. 1991. A microscopic study of the different classes of symptoms coded by the *Vf* gene in apple for resistance to scab (*Venturia inaequalis*). Plant Pathology **40**, 249–256.
- Cingolani P, Platts A, le Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. SNPs in the genome of *Drosophila melanogaster* strain *w*¹¹¹⁸; *iso-2*; *iso-3*. Fly **6**, 80–92.
- Conrad DF, Pinto D, Redon R, *et al.* 2010. Origins and functional impact of copy number variation in the human genome. Nature **464**, 704–712.
- Cova V, Lasserre-Zuber P, Piazza S, Cestaro A, Velasco R, Durel CE, Malnoy M. 2015. High-resolution genetic and physical map of the *Rvi1* (*Vg*) apple scab resistance locus. Molecular Breeding **35**, 16.
- Daccord N, Celton J, Linsmith G, *et al.* 2017. High-quality *de novo* assembly of the apple genome and methylation dynamics of early fruit development. Nature Genetics **49**, 1099–1106.
- Danecek P, Bonfield JK, Liddle J, *et al.* 2021. Twelve years of SAMtools and BCFtools. GigaScience **10**, giab008.
- Dayton DF, Shay JR, Hough LF. 1953. Apple scab resistance from R12740-7A, a Russian apple. Proceedings. American Society for Horticultural Science **62**, 334–340.
- De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics **34**, 2666–2669.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochemical Bulletin **19**, 11–15.
- Dwivedi V, Kumar SR, Shilpashree HB, Krishna R, Rao S, Shasany AK, Olsson SB, Nagegowda DA. 2022. An inducible potato (*E*,

E)-farnesol synthase confers tolerance against bacterial pathogens in potato and tobacco. *The Plant Journal* **111**, 1308–1323.

Gardiner SE, Bassett HCM, Noiton DAM, Bus VG, Hofstee ME, White AG, Ball RD, Forster RLS, Rikkerink EHA. 1996. A detailed linkage map around an apple scab resistance gene demonstrates that two disease resistance classes both carry the *Vf* gene. *Theoretical and Applied Genetics* **93**, 485–493.

Gardiner SE, Bus VGM, Rusholme RL, Chagné D, Rikkerink EHA. 2007. Apple. In: Kole C, ed. *Fruits and nuts. Genome mapping and molecular breeding in plants*, vol 4. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–62.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv* doi: [10.48550/ARXIV.1207.3907](https://doi.org/10.48550/ARXIV.1207.3907). [Preprint].

Goel M, Sun H, Jiao W-B, Schneeberger K. 2019. SyRl: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology* **20**, 277.

Gomès E, Jakobsen MK, Axelsen KB, Geisler M, Palmgren MG. 2000. Chilling tolerance in *Arabidopsis* involves *ALA1*, a member of a new family of putative aminophospholipid translocases. *The Plant Cell* **12**, 2441–2453.

Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D. 2011. Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. *Plant Physiology* **157**, 757–769.

Hála M, Žárský V. 2019. Protein prenylation in plant stress responses. *Molecules* **24**, 3906.

Huerta-Cepas J, Szklarczyk D, Heller D, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309–D314.

Jänsch M, Broggini GAL, Weger J, Bus VGM, Gardiner SE, Bassett H, Patocchi A. 2015. Identification of SNPs linked to eight apple disease resistance loci. *Molecular Breeding* **35**, 45.

Jänsch M, Paris R, Amoako-Andoh F, Keulemans W, Davey MW, Pagliarini G, Tartarini S, Patocchi A. 2014. A phenotypic, molecular and biochemical characterization of the first cisgenic scab-resistant apple variety ‘Gala’. *Plant Molecular Biology Reporter* **32**, 679–690.

Jiang H, Zhang M, Yu F, et al. 2023. A geraniol synthase regulates plant defense via alternative splicing in tea plants. *Horticulture Research* **10**, uhad184.

Jiao W-B, Schneeberger K. 2020. Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nature Communications* **11**, 989.

Joshi SG, Schaart JG, Groenwold R, Jacobsen E, Schouten HJ, Krens FA. 2011. Functional analysis and expression profiling of *HcrVf1* and *HcrVf2* for development of scab resistant cisgenic and intragenic apples. *Plant Molecular Biology* **75**, 579–591.

Jung S, Ficklin SP, Lee T, et al. 2014. The genome database for Rosaceae (GDR): year 10 update. *Nucleic Acids Research* **42**, D1237–D1244.

Khan A, Carey SB, Serrano A, Zhang H, Hargarten H, Hale H, Harkess A, Honaas L. 2022. A phased, chromosome-scale genome of ‘Honeycrisp’ apple (*Malus domestica*). *Gigabyte* 2022, <https://doi.org/10.46471/gigabyte.69>.

Khan A, Korban SS. 2022. Breeding and genetics of disease resistance in temperate fruit trees: challenges and new opportunities. *Theoretical and Applied Genetics* **135**, 3961–3985.

Khan A, Švara A, Wang N. 2024. Comparing apples and oranges: advances in disease resistance breeding of woody perennial fruit crops. *Annual Review of Phytopathology* **62**, 263–287.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**, 540–546.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Research* **27**, 722–736.

Kosambi DD. 1943. The estimation of map distances from recombination values. *Annals of Eugenics* **12**, 172–175.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biology* **5**, R12–R12.

Lanfeer R, Schalamun M, Kainer D, Wang W, Schwessinger B. 2019. MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics* **35**, 523–525.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100.

Li Z, Wang L, He J, et al. 2022. Chromosome-scale reference genome provides insights into the genetic origin and grafting-mediated stress tolerance of *Malus prunifolia*. *Plant Biotechnology Journal* **20**, 1015–1017.

Liebhart R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van De Weg E, Gessler C. 2002. Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.). *Molecular Breeding* **10**, 217–241.

López-Girona E, Bowatte DR, Smart M, Alvares S, Brancher L, Chagné T, Volz D, K R. 2021. A high-throughput *S-RNase* genotyping method for apple. *Fruit Research* **1**, 10.

López-Girona E, Davy MW, Albert NW, Hilario E, Smart MEM, Kirk C, Thomson SJ, Chagné D. 2020. CRISPR-Cas9 enrichment and long read sequencing for fine mapping in plants. *Plant Methods* **16**. <https://doi.org/10.1186/s13007-020-00661-x>

Malnoy M, Xu M, Borejsza-Wysocka E, Korban SS, Aldwinckle HS. 2008. Two receptor-like genes, *Vfa1* and *Vfa2*, confer resistance to the fungal pathogen *Venturia inaequalis* inciting apple scab disease. *Molecular Plant-Microbe Interactions* **21**, 448–458.

Mansfeld BN, Yocca A, Ou S, Harkess A, Burchard E, Gutierrez B, van Nocker S, Gottschalk C. 2023. A haplotype resolved chromosome-scale assembly of North American wild apple *Malus fusca* and comparative genomics of the fire blight *Mfu10* locus. *The Plant Journal* **116**, 989–1002.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770.

McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddelloh JA, Stupar RM. 2012. Structural variants in the soybean genome localize to clusters of biotic stress-response genes. *Plant Physiology* **159**, 1295–1308.

Muñoz-Amatriáin M, Eichten SR, Wicker T, et al. 2013. Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biology* **14**, R58.

Nintemann SJ, Palmgren M, López-Marqués RL, et al. 2019. Catch you on the flip side: a critical review of flippase mutant phenotypes. *Trends in Plant Science* **24**, 468–478.

Ou S, Chen J, Jiang N. 2018. Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Research* **46**, e126.

Padmarasu S, Sargent DJ, Patocchi A, et al. 2018. Identification of a leucine-rich repeat receptor-like serine/threonine-protein kinase as a candidate gene for *Rvi2* (*Vb*)-based apple scab resistance. *Molecular Breeding* **38**, 73.

Papp D, Singh J, Gadoury D, Khan A. 2020. New North American isolates of *Venturia inaequalis* can overcome apple scab resistance of *Malus floribunda* 821. *Plant Disease* **104**, 649–655.

Parisi L, Lespinasse Y, Guillaumès J, Krüger J. 1993. A new race of *Venturia inaequalis* virulent to apples with resistance due to the *Vf* gene. *Phytopathology* **83**, 533–537.

Patocchi A, Bigler B, Koller B, Kellerhals M, Gessler C. 2004. *Vr2*: a new apple scab resistance gene. *Theoretical and Applied Genetics* **109**, 1087–1092.

Patocchi A, Wehrli A, Dubuis PH, et al. 2020. Ten years of VINQUEST: first insight for breeding new apple cultivars with durable apple scab resistance. *Plant Disease* **104**, 2074–2081.

Peil A, Howard NP, Bühlmann-Schütz S, Hiller I, Schouten H, Flachowsky H, Patocchi A. 2023. *Rvi4* and *Rvi15* are the same apple scab resistance genes. *Molecular Breeding* **43**, 74.

- Pichersky E, Gershenzon J.** 2002. The formation and function of plant volatiles: perfumes for pollinator attraction and defense. *Current Opinion in Plant Biology* **5**, 237–243.
- Poulsen LR, López-Marqués RL, Palmgren MG.** 2008. Flippases: still more questions than answers. *Cellular and Molecular Life Sciences* **65**, 3119–3125.
- Qin S, Xu G, He J, Li L, Ma H, Lyu D.** 2023. A chromosome-scale genome assembly of *Malus domestica*, a multi-stress resistant apple variety. *Genomics* **115**, 110627.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC.** 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* **11**, 1432.
- Rashid U, Wu C, Shiller J, et al.** 2024. AssemblyQC: a Nextflow pipeline for reproducible reporting of assembly quality. *Bioinformatics* **40**, btad477.
- Roach MJ, Schmidt SA, Borneman AR.** 2018. Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460.
- Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP.** 2017. Variant review with the integrative genomics viewer. *Cancer Research* **77**, e31–e34.
- Schouten HJ, Brinkhuis J, van der Burgh A, Schaart JG, Groenwold R, Brogini GAL, Gessler C.** 2014. Cloning and functional characterization of the *Rvi15* (*Vr2*) gene for apple scab resistance. *Tree Genetics & Genomes* **10**, 251–260.
- Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC.** 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods* **15**, 461–468.
- Shafin K, Pesout T, Lorig-Roach R, et al.** 2020. Nanopore sequencing and the Shasta toolkit enable efficient *de novo* assembly of eleven human genomes. *Nature Biotechnology* **38**, 1044–1053.
- Shen W, Le S, Li Y, Hu F.** 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM.** 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212.
- Soriano JM, Joshi SG, van Kaauwen M, Noordijk Y, Groenwold R, Henken B, van de Weg WE, Schouten HJ.** 2009. Identification and mapping of the novel apple scab resistance gene *Vd3*. *Tree Genet Genomes* **5**, 475–482. <https://doi.org/10.1007/s11295-009-0201-5>
- Souleyre EJF, Bowen JK, Matich AJ, et al.** 2019. Genetic control of α -farnesene production in apple fruit and its role in fungal pathogenesis. *The Plant Journal* **100**, 1148–1162.
- Stam P.** 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *The Plant Journal* **3**, 739–744.
- Sun X, Jiao C, Schwaninger H, et al.** 2020. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nature Genetics* **52**, 1423–1432.
- Švara A, Sun H, Fei Z, Khan A.** 2024. Chromosome-level phased genome assembly of ‘Antonovka’ identified candidate apple scab resistance genes highly homologous to *HcrVf2* and *HcrVf1* on linkage group 1. *G3 Genes|Genomes|Genetics* **14**, jkad253.
- Tarailo-Graovac M, Chen N.** 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* **25**, 4.10.1–4.10.14.
- Tossolini I, Mencia R, Arce AL, Manavella PA.** 2025. The genome awakens: transposon-mediated gene regulation. *Trends in Plant Science* **30**, 857–871.
- Vaser R, Sović I, Nagarajan N, Šikić M.** 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Research* **27**, 737–746.
- Vinater BA, Patocchi A, Gianfranceschi L, Tartarini S, Zhang HB, Gessler C, Sansavini S.** 2001. Apple contains receptor-like genes homologous to the *Cladosporium fulvum* resistance gene family of tomato with a cluster of genes cosegregating with *Vf* apple scab resistance. *Molecular Plant-Microbe Interactions* **14**, 508–515.
- Vinod KK.** 2011. Kosambi and the genetic mapping function. *Resonance* **16**, 540–550.
- Volz RK, Proffitt N, Marshall C, Orcheski B, Bowatte D, Chagné D, López-Girona E, Bus VGM.** 2023. Apple cultivar breeding for multigenic resistance to multiple diseases: the New Zealand experience. *Acta Horticulturae* **1362**, 205–212.
- Voorrips RE.** 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *The Journal of Heredity* **93**, 77–78.
- Wickham H.** 2016. ggplot2. Elegant graphics for data analysis. Cham, Switzerland: Springer International Publishing.
- Williams EB, Kuc J.** 1969. Resistance in *Malus* to *Venturia inaequalis*. *Annual Review of Phytopathology* **7**, 223–246.
- Xu X, Yang J, Thakur V, Roberts A, Barbara DJ.** 2008. Population variation of apple scab (*Venturia inaequalis*) isolates from Asia and Europe. *Plant Disease* **92**, 247–252.
- Zhang L, Hu J, Han X, et al.** 2019. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nature Communications* **10**, 1494.