Complex Genomic Structural Variation Underlies Climate Adaptation among *Eucalyptus* species

Abstract

Climate change threatens forest ecosystems globally, yet our understanding of adaptive genomic variation in wild forest trees remains limited. Here, we present the first comprehensive pangenome analysis of a widely adapted *Eucalyptus* tree species, *Eucalyptus viminalis*, using haplotype-resolved genomes, population-level long-read sequencing of a rangewide set, and short-read sequencing across larger populations and closely related species. We discovered complex and simple structural variants (SVs) which together effectively double the length of a single haploid genome (530Mb haploid vs. 1Gb pan-genome). SVs show significant enrichment near transposable elements, including a novel retrotransposon-derived polyprotein with active expression. Genome-environment association analyses of SVs in 49 *E. viminalis* identified *CHILL1*, a 400-kb cold-adaptation locus on chromosome 1. Validation of this major locus with short-reads of 434 wild accessions of three related species (*E. viminalis*, *E. dalrympleana*, *E. rubida*) confirmed that *CHILL1* genotype predicts cold adaptation better than species classification, with effect sizes equivalent to 440km in latitudinal or 307m altitudinal translocation. These findings demonstrate that structural variants harbor critical adaptive variation in wild tree populations that can be harnessed to understand plant abiotic stress tolerance and to design climate-informed restoration and conservation programs for forests of the future.

Introduction

Forests regulate global climate systems and harbor exceptional biodiversity (Rogger et al., 2024), yet climate change is converting them from carbon sinks to emission sources, driven partly by biodiversity loss (Brancalion et al., 2025). Understanding the genomic basis of climate adaptation in forest trees is crucial for conservation, yet fundamental knowledge of adaptive genetic variation in wild populations remains surprisingly limited. This knowledge gap constrains our ability to identify climate-resilient genotypes at continental and global scales —a critical need for global forest restoration and management.

In Australia, *Eucalyptus* is the most prevalent and abundant genus of forest trees, containing more than 840 species defined by morphology (Nicolle et al., 2025). These species exhibit complex evolutionary history, characterized by both deep divergences among subgenera and frequent hybridizations between related species (Thornhill et al., 2019). Most extant species diversified within the past 2 million years (Thornhill et al., 2019) — an exceptionally rapid radiation that has been linked to continental-scale changes in climate and soil conditions across the Australian continent (Li et al., 2018). Among these, *Eucalyptus viminalis* occupies one of the widest environmental niches among extant *Eucalyptus* species, being distributed along the south east coast and hinterland of Australia (-10°C to 25°C; 500-1,700 mm annual rainfall, 0-2000m altitude) where it serves as a primary food source for koalas, conferring both ecological and cultural significance (Moore and Foley, 2005).

Previous genome-wide association studies (GWAS) in *Eucalyptus* have relied exclusively on single nucleotide polymorphisms (SNPs) identified by mapping short reads to a single reference genome (*E. grandis*) (Ballesta et al., 2019; Costa e Silva et al., 2009; Murray et al., 2019; Myburg et al., 2014). These studies revealed remarkably high levels of genetic diversity in *Eucalyptus* species, even though these approaches systematically exclude larger genomic variations known as structural variants (SVs). SVs are rearrangements over 50 bp that include insertions, deletions, inversions, and duplications. Recent pangenome studies in humans (Gustafson et al., 2024; Liao et al., 2023; Reis et al., 2023; Silcocks et al., 2023), primates (Höps et al., 2024; Kuderna et al., 2023; Yoo et al., 2025), birds (Edwards et al., 2025; Lai et al., 2019), model plants (Kang et al., 2023; Lian et al., 2024), and crop species (Jayakodi et al., 2024; Jiao et al., 2024; Z. Zhang et al., 2025) reveal that SVs profoundly impact phenotypic variation and frequently underlie adaptive traits that SNPs cannot explain. In plants, SVs are often associated with active transposable element (TE) behaviours that are understudied, even in model species (Igolkina et al., 2025; Teasdale et al., 2025; Wlodzimierz et al., 2023). Even though some TEs can be very large, functional, and in ways that are important for genome evolution (Urquhart et al., 2024).

Recent comparative genomics of 33 *Eucalyptus* species demonstrated that large-scale SVs have driven interspecific genome evolution (Ferguson et al., 2022, 2024a) but the prevalence, segregation and adaptive significance of SVs within wild *Eucalyptus* populations remain unknown. This represents a critical gap: understanding intraspecific SV variation is essential for identifying adaptive alleles and predicting population responses to climate change (Ferguson et al., 2024b).

In this study, we perform the first sequence and orthogroup-based pangenome analyses of a naturally evolving *Eucalyptus* species. We discover both simple SVs (sSVs) and complex SVs (cSVs) and associate them with mobile genetic element activities. Using long- and short-read population resequencing, we identify a complex SV locus with strong evidence of a role in cold adaptation in *E. viminalis* and two closely related species. Our findings highlight the importance of individual adaptive loci, rather than species or local provenance, in climate adaptation and paves the way for smart design of forests in the face of rapid climate change.

Results

De novo assemblies of haplotype-resolved wild *Eucalyptus* trees

We generated chromosome-scale, haplotype-resolved genome assemblies for our target species *Eucalyptus viminalis*. We sequenced a single tree using PacBio HiFi reads (30.3 Gbp, 55× coverage), Oxford Nanopore ultra-long reads (filtered to 32.7 Gbp, 60× coverage, N50 36.6 kb), and Hi-C chromatin conformation data. We assembled two high-quality and highly complete haplotypes with hifiasm, achieving QVs of 56.41 and 55.52 48.26 and 53.27 for haplotypes 1 and 2, respectively. The assemblies were 547 Mbp (haplotype 1) and 551 Mbp (haplotype 2), with Benchmarking Universal Single-Copy Orthologs (BUSCO) scores of 99.61% and 99.68% respectively. To complement this reference diploid assembly, we generated four additional high base accuracy (QV > 60), near-gapless diploid genomes (one gap each; haplotype BUSCO range from 99.61% to 99.72%) using high-coverage (50x – 60x), ultra long (read length N50: 46 kbp), ONT data. Contigs from several high-coverage environmental samples were also assembled by scaffolding against the reference genome, generating near-complete supportive genomes (BUSCO > 92%). In total, we chose 10 high-quality representative haplotypes for a pangenome analysis than spanned four states and territories of Australia (Supplementary Table 1-5) that help us understand the underlying structural variation complexities that entail both genome evolution and adaptive capacity within wild *Eucalyptus* species.

Pangenome analyses reveal complexity and conservation of genome architecture in an undomesticated *Eucalyptus* species

Genome architecture and genomic polymorphism diversity in undomesticated plant species are often poorly characterised. In order to address this issue, we applied both sequence-based and gene-based pangenome analyses to better understand the genome complexity and organisation in *E. viminalis* (Fig. 1). First, using Pangenome Graph Builder (PGGB) (Garrison et al., 2024), we constructed a graph-based pangenome incorporating 10 haplotypes. To do so, PGGB first identified sequentially homologous haploblocks through pairwise genome mapping. Among them, 20.3% (17,065,267) of graph nodes (homologous sequence blocks) were conserved across all haplotypes, suggesting maintenance of core genomic architecture at the sequence level (Fig. 1b). Moreover, PGGB also revealed 1.1 Gbp of non-redundant pangenomic sequences, 2x larger than a single haploid genome (Fig. 1d). Surprisingly, novel sequences increased linearly with each additional genome— an uncommon pattern among studies of model, crop and primate species (Jayakodi et al., 2024; Lian et al., 2024; Liao et al., 2023; Teasdale et al., 2025). This demonstrates unusually abundant genomic sequence diversity in wild *Eucalyptus*, likely a result of their life history.

These variable genomic sequences could encompass all functional or non-function regions of the genomes. To more explicitly identify the variabilities in the protein-coding sequences, we first annotated the reference genome using Breaker3 and Isoquant, combining *ab initio* approaches with extrinsic evidence from ONT direct RNA sequencing data (mature leaves; 18 million filtered reads; N50: 1.62 kb; Supplementary Note 3). This captured a

comprehensive genome annotation of 33,818 genes and 4,352 isoforms (total = 1,373 +4,352 + 28,093 = 33, 818). These 33,818 genes were grouped into 16,770 orthogroups according to their sequence relatedness. Then, gene orthologs were described with functional annotations with EggNog. Other haplotype genomes were annotated with Liftoff from the reference genome.

In contrast to the linearly increasing sequence space of this genome, we only observed logarithmically gain of new orthogroups per haplotype input, saturating at 22,000 orthogroups with 10 haplotypes (Fig. 1e). Thus, the presence and absence variations (PAV) of orthogroups alone could not explain the rapid sequence expansion seen in PGGB. However, orthogroups with high copy number variation (CNVs) may also contribute to between-haplotype variation. Among the 16,770 annotated orthogroups, most had two to three gene copies per haplotype, with some reaching an average of 15 copies per haplotype (Supplementary Note 5). These results may suggest that tandem duplication events affect gene copies, which is consistent with previous studies in forestry species of *Eucalyptus* (Myburg et al., 2014). Four orthogroups exhibited high variance (SD > 10) in copy number among haplotypes (Fig. 1i): these orthogroups were functionally related to DNA damage tolerance, kinase activity, transport and regulation of cell walls via Cobra-like proteins (COBL) (Fig. 1i, Supplementary Note 5). CNVs in these orthogroups may therefore contribute to diversification of gene function through gradual sequence divergence among gene copies.

Haplotype genomes reveal complex structural variation related to transposable element activities

Since gene orthogroup and their copy number only partially explained the rapid sequence expansion within this *Eucalyptus* species, we next investigated the potential contribution of transposable element (TE) activity, one of the most well-known sources of sequence generation and genome perturbation mechanisms that allows rapid genome evolution and adaptation (Fedoroff, 2012). However, TEs are often poorly described in non-model plant species where they may underpin genome evolution. Here, we studied TE activity via structural variants (SVs), initially by identifying insertions, deletions, duplications, and inversions by comparing haplotypes, then studying sequence homology, repeat structure, and open reading frames to draw connections to TEs. Using three *E. viminalis* individuals with distributed provenances as references (ACT, NSW, TAS), we performed detailed SV characterisation using the 10 haplotype-resolved genomes of wild *E. viminalis* with Pannagram. In doing so, we identified both simple (biallelic) structural variants (sSVs) and complex (multi-allelic) structural variants (cSVs), resulting in 59,428 sSVs and 26,688 cSVs, reaching an average gain of 8,610 SVs per haplotype input (Fig. 2a,b). The differentiation of sSVs and cSVs may be confounded by other SV callers in similar studies (Hu et al., 2025). In this study, we identified that 20% of the total SVs are complex which could be due to nested events (Fig. 2b). Counterintuitively, longer SVs did not tend to be more complex by having higher chances of nesting smaller variants (Fig. 2b). Some SVs (13%) were associated with TEs, suggesting recent expansion events that could further explain the rapidly expanding pangenomic sequences in wild *Eucalyptus* (Supplementary Note 7).

Wild *Eucalyptus* harbour structural variation with functional implications

Although genome-based SV discovery methods offer detailed insights into pangenome complexity in undomesticated *Eucalyptus* species, such complexity also creates challenges to their biological interpretations. In order to identify SVs for functional screening, we shifted our discovery strategy to long-read reference-mapping. To do so, 71 *E. viminalis* individuals were sequenced with ONT long-reads (R10.4.1 chemistry) and basecalled with a species-specific basecaller. We selected 34 high coverage samples (68 haplotypes; >12× ONT coverage) for SV calling. After mapping long-reads to the reference genome, Sniffles2 initially identified 883,116 SVs. After singleton filtering and merging, 573,889 nonredundant SVs > 50 bp were consolidated. This corresponds to an average gain of 8,439 SVs per haplotype input, which is remarkably consistent with the genome-based results from Pannagram. Sniffles-discovered SVs covered 38% of the reference genome, leaving the rest of the genome positions as structurally conserved (Supplementary Table 10; Supplementary Note 9).

Across the genome, SV events had a range of size and frequencies. Insertions and deletions were the most frequent (> 99%) SV types, as observed in other plant pangenomes (Benoit et al., 2025; L. Cheng et al., 2025). However, inversions and duplications (median sizes: 3 kb and 12 kb respectively; Fig. 2c) were much larger than insertions and deletions (median sizes: 119 bp and 231 bp respectively) and therefore modified many more bases in single events. Overall, 15.4% (133,379) SVs were found within 5 kbp of gene annotations (Fig. 2e). These SVs were enriched in several disease resistance gene pathways (Supplementary Note 9). This was hypothesised previously to support the long life history of oak trees (Plomion et al., 2018), which could also be applicable to *Eucalyptus*. In particular, (tandem) duplications were sixfold enriched for events directly affecting exons (6% vs. 1% genome-wide), indicating that exon duplication is better tolerated than deletions or inversions.

SV genomic isolation by distance and environment

Genomic structural variation segregation in wild species is a rarely studied phenomenon, yet could be important for understanding their genome evolution and adaptation. To explore geography-driven genetic variation in SVs, we used the SV genotype matrix to calculate pairwise genetic distance (573,889 variants genotyped in 49 accessions; > 8x ONT coverage). Overall, PC1 of the SV genotype matrix only explained 5.49% of the variance (Supplementary Note 10), in line with previous SNP and short-read SV studies in *Eucalyptus* (Ferguson et al., 2024c; Murray et al., 2019). When we plotted the pairwise genetic distance against the pairwise geographic distance, genome-wide genetic distance continuously increased with geographic distance over a 620 km range (Fig. 3c). This lack of discrete population structure (or isolation by distance; IBD) is likely attributed to the long-range gene flows of *Eucalyptus* inferred from SNP data (Supple et al., 2018), which we now show to be applicable to SV polymorphisms too.

Heterogeneous environmental conditions are probable drivers of genome diversification or even speciation (Balkenhol et al., 2019). To explore genetic population structure driven by

environment, we generated pairwise environmental distances using BioClim variables at each geo-referenced sample location to identify genetic isolation by environment (IBE). Since environmental conditions are often confounded with geography (Murray et al., 2019), we fitted additive generalised dissimilarity models (GDMs) to adjust for variation due to geography. Over and above genographic distances, BioClim6 (minimum temperature of the coldest month) stood out as having robust positive covariation with SV allele frequencies (Fig. 3d). These analyses demonstrate general genetic divergence associated with cold environments. Next, we searched for individual cold-adaptive alleles through genome-environment association analyses.

Discovery of probable climate adaptive loci using structural variants

In order to further investigate the genetic basis of potential adaptation to cold environments, we performed a genome-wide association study (GWAS) to associate BioClim6 with genotypes at SV loci across the genome. In order to control for population relatedness, we used a SV-based genetic relationship matrix as a random effect in the model. We identified a notably large effect adaptive locus, *CHILL1*, associated with cold temperature extremes as represented by BioClim6 (minimum temperature in coldest month). SV-defined *CHILL1* is located in a 129 kb window between 27.7 and 27.8 Mb on chromosome 1 (Fig. 4c). All three most significantly associated SVs within *CHILL1* were all small indels (58 - 323 bp; Fig. 4c). The estimated effect size of *CHILL1/chill1* is -2°C (intercept at 0°C), consistent with it conferring frost tolerance. All trees that experienced average minimum temperatures of the coldest month below -1°C possessed the cold-adaptive *CHILL1* allele (Fig. 4b).

Within the SV-defined *CHILL1* locus, we identified two putatively implicated genes: GPI anchor protein (two copies) and Chlorophyll a reductase, both related to cold adaptation albeit through different ontogenetic pathways. GPI-anchor proteins modulate Detergent-Resistant Membranes (DRMs) in the plant apoplast that ultimately alter cell wall function, thus being essential for mediating freeze tolerance in plants (Takahashi et al., 2016). The *Arabidopsis* homologue of this protein (At1g61900) is primarily expressed in flowering tissues and seeds (Supplementary Note 11), suggesting protection of plant fecundity. Interestingly, GPI-anchor protein is a class of COBL proteins which were found to have the highest amount of CNV in the pangenome analyses and thus likely to be both sequence and functionally diverse across the genome. By contrast, Chlorophyll a reductase likely relates to physiological photoprotection under colder temperatures. Altering chlorophyll a/b ratio is one of the most common strategies for plants to protect the vulnerable photosystem I (PSI) from light damage under cold conditions (Niinemets, 2010). Once damaged, the plants need to regenerate the entire photosystem, and systemically decrease carbon assimilation which are both deleterious to vegetative survival and reproduction (Hallik et al., 2017). Interestingly, we also identified a root-specific phosphate transporter, and a phosphatase (both expressed), in the *CHILL* locus of chromosome 8, which may relate to phosphate uptake (*CHILL8*; expanded in Supplementary Note 12).

Cross-species validation of a complex, climate adaptive locus

Given that *CHILL1* is a large-effect locus covering >100 kbp, we expected many linked SNPs in this region (Sirén et al., 2021). To further explore genetic variation at *CHILL1*, we used genome-wide SNPs called from 128 *E. viminalis* samples and two related species (Supplementary Note 13). The SNP data within *E. viminalis* confirmed that *CHILL1* is an extensive region of tightly linked SNPs that resembles a recombination-reduced haploblock (supergene) of ~400kb with three distinct haplotypes that strongly associate with BioClim6 (Fig. 5b). SNPs within the haploblock were highly correlated to SV genotype, especially in the sub-region where the structural variants were located (positions 27.7-27.8Mb; Supplementary Note 13).

Remarkably, the *CHILL1* adaptive locus was not only present in *E. viminalis,* it was also polymorphic and associated with cold conditions (BioClim6) in two related species, *E. dalrympleana* and *E. rubida,* which form ecological assemblages due to shared ancestry and overlapping ranges (Fig. 5a; Supplementary Note 13). Among 134 *E. dalrympleana* samples, 114 carried a copy of *CHILL1*, most in samples from the colder southern and central regions, while *E. rubida* showed a similar pattern (Fig. 5b). These results validate those from long-read SV data. Moreover, they reveal that *CHILL1* likely emerged before the divergence of these species. Persistence of segregation at *CHILL1* across multiple species highlights its importance to adaptation in the *Maidenaria* section of *Eucalyptus*. Principal component analyses based on genome-wide SNP markers confirmed relatively low (9.5% of genome-wide SNP genetic variation explained by PC1 and PC2) genetic divergence of the three species as well as considerable hybridisation between them according to admixture results (Supplementary Note 13). This resulted in most of the neutral and adaptive variation as unexplained, emphasizing the large amount of natural diversity yet to be characterised. By contrast, PC analyses using *CHILL1* SNPs explained 65% of the variation across species, more so than the species label (9.5%). This illustrates that individual loci segregating within species may be able to explain adaptation more effectively than species themselves (Supplementary Note 13).

Multiple regression modelling of latitude, elevation, *CHILL1* genotype and species estimated a *CHILL1* effect size equivalent to that of translocating from one extreme of these species' frost environmental range to the other, namely, 440km in latitudinal distance or 307m in altitude (Supplementary Table 15).

Further characterisation of *CHILL1*

After discovery using long-read mapping, and genetic validation using short-read of *CHILL1*, we sought to use the haplotype assemblies and their annotations to further characterise this important climate adaptive locus. Using haplotype-resolved genome assemblies with homozygous *CHILL1* and *chill1* alleles, a physical map showed diverged (less than 80% sequence similarity) sequences inserted into the *CHILL1* locus (Fig. 6a). Counterintuitively, the region where the leading SV (and candidate genes) was located remained mostly syntenic, suggesting preservation of genomic structure. Additionally, we observed clustering of both LTRs (retrotransposons) and terminal inverted repeat (TIR; DNA transposons; Fig 6c) in the locus, suggesting that TE activity may have played a role in the evolution and function of this locus. Further analyses using haplotype-resolved genomes (n = 10) identified no

major inversions in the *CHILL1* region as defined by synteny-resolved orthogroups (Supplementary Note 14). The presence of complex SVs adds challenges in extracting function even with very high quality data. Pannagram identified the leading Sniffles2 58 bp SV (Chr1:27713800) as a cSV of 60 bp (Chr1:27713769-27713828). This simplified SV representation from Sniffles2 might have allowed us to detect associations with GWAS using the current framework.

Discussion

Genomic structural variation studied with long-read technologies is a powerful way to explore both genome evolution and functional adaptation. Here, we present one of the first comprehensive analyses of structural variation in naturally evolving *Eucalyptus*, a genus of forest trees with regional and global climate importance. In doing so, we identified a doubling (530Mb to 1.1Gb) in pan genome total sequence with 10 haplotypes over a single reference. Through studying structural variation, we observed extensive gene PAVs and CNVs, potentially leading to large functional differences within species. In this paper, we effectively combined high-quality long-read genomics with wide range short-read datasets to provide an optimal power for detecting, characterising, and validating large adaptive loci.

In particular, we identified a cross-species climate adaptive locus, *CHILL1,* which is functionally associated with temperature in the coldest quarter and on its own explains 65% of the genetic variation. The *CHILL1* haplotype likely pre-dates the species divergence and is a better predictor than morphology-based taxonomic classifications. *CHILL1* is a large, non-recombining haploblock that is broadly syntenic among haplotypes, despite containing many complex SVs. This contrasts with previous studies in sunflowers (Todesco et al., 2020) and deer mice (Harringmeyer and Hoekstra, 2022), where large inversions were the primary mechanism behind recombination barriers over extensive genomic regions. To confer with the largely syntenic *CHILL1* locus, other mechanisms, such as conical repeats (Logsdon et al., 2025), changes in methylation patterns (Huang et al., 2025), or simply very strong selection pressure on Interpunctuated SVs may also lead to extensive functional haploblocks.

Active TEs can generate new sequences, and interrupt existing gene introns/exons and open reading frames (Ardern, 2023). These changes are important for understanding genome evolution and adaptation (Alonge et al., 2020; Hof et al., 2016). In this paper, we found several independent lines of evidence that support ongoing TE expansions within *Eucalyptus*, generating many shared polymorphisms among related species. These include expression of TE-associated Pol poly-proteins in leaf tissues, clusters of TE-associated SVs with high sequence homology, and large families of TEs with no sequence divergence. Additionally, we also found extensive TE presence (both class I and class II) in our climate adaptive locus *CHILL1*. These TE insertions would have probable implications in changing the accessibility of transcriptional machinery, therefore altering expression of genes within this locus through modifications of chromatin structure (Bubb et al., 2025).

Within the *CHILL1* locus, we identified several significantly implicated genes, which include GPI-anchor protein and Chlorophyll a reductase, complementing existing molecular results regarding plant cold acclimation (Takahashi et al., 2016). Additionally, at the *CHILL8* locus, we identified expressed phosphate transporter and phosphatase genes (preferentially expressed in root tissues in *Arabidopsis* homologues), suggesting co-selection with GPI-anchor proteins and Chlorophyll a reductase under realistic environmental conditions, considering low soil phosphate conditions in Australia (Hou et al., 2020; Wang et al., 2025).

Although we identified several traits related to cold adaptation in *Eucalyptus* through scanning landscape genomic covariation with environments, future trait studies involving long-term trials of tree growth may still be needed to further characterise the adaptive traits underpinning *CHILL1*. Nonetheless, adaptive alleles such as *CHILL1* are readily deployable now in targeted tree planting for forestry, habitat restoration or carbon farming programs without the need for genetic modification. Natural selection has already filtered *CHILL1* to high frequency across three key species across tablelands and mountain landscapes. Utilising cutting-edge genomics to harness the existing natural adaptive capacity can address the urgent need to conserve global forests and biodiversity under rapidly changing climates.

Methods

Eucalyptus viminalis reference genome assembly

A *de novo* telomere-to-telomere *Eucalyptus viminalis* genome was assembled for use as the reference genome in all analyses in this study. Fresh leaf tissue of *E. viminalis* (subsp. *viminalis*) tree was sourced from the Australian National Botanic Gardens, Canberra Australia (accession number CBG 7902535.8, located in section 111). Tissue was kept cool and hydrated until cryogenically stored at -80°C as soon as possible. Approximately 10 g of leaf tissue was ground to a fine powder using a mortar and pestle cooled under liquid nitrogen, and DNA was extracted according to (Jones et al., 2021). DNA was subsequently size selected for fragments ≥ 20 kb using a BluPippin (Sage Science). A Pacific Biosciences (PacBio) SMRTbell DNA library was prepared (SMRTbell Express Template Prep Kit 3.0) and sequenced on the Sequel II platform using an 8M SMRT cell. High-fidelity (HiFi) circular consensus reads were generated using DeepConsensus (version: 1.2.0) (Baid et al., 2023). To obtain ultra-long reads, an Oxford Nanopore Technologies (ONT) native DNA library was prepared (ligation kit SQK-LSK114) and sequenced on the PromethION P24 platform using a FLO-PRO114M R10.4.1 flow cell. When sequencing declined (low active pore count), the flow cell was treated with ONT Flow Cell Wash Kit (EXP-WSH004), then re-primed and reloaded with more libraries following the manufacturer's instructions. Sequencing data (pod5) was basecalled to fastq with ONT Dorado (version: 0.3.4), using the super accurate model. A chromosome conformation capture Hi-C library was prepared with a Phase Genomics Proximo Hi-C Plant Kit (version 4, document KT3040B) and sequenced on an Illumina NovaSeq 6000 using an S4 flow cell with a 300 cycle kit (150 bp paired-end sequencing).

All PacBio HiFi reads were used in assembly (≥ Q20). ONT reads were filtered for ≥ Q7 and ≥ 55 kb using NanoFilt (version: 2.8.0) (De Coster and Rademakers, 2023). Hi-C reads were filtered with Trim Galore (version: 0.6.10) (Krueger et al., 2023), to remove any Illumina adapters and validate the read pairs. The *de novo* genome assembly was performed with Hifiasm ultra-long (UL) (version: 0.19.6-r597) (Cheng et al., 2024), incorporating the PacBio HiFi, ONT ultra-long (--ul) and Hi-C reads (--h1 --h2).

After assembly, Hi-C contact maps were independently created for each haplotype using Juicer (version 1.6) (Neva C. Durand et al., 2016) and subsequently scaffolded with 3D-DNA (parameters: --editor-repeat-coverage 5 -r 1; version: 190716) (Dudchenko et al., 2017). Juicer used BWA-MEM for read alignment (version: 0.7.17) (Neva C. Durand et al., 2016). After initial scaffolding, each haplotype was visually inspected and, when necessary, manually edited using Juicebox (version 1.11.08) (Neva C Durand et al., 2016).

For quality assessment of the genome, telomeres, single copy genes, and base accuracy was calculated. The telomere sequence for *E. viminalis* was discovered using tidk v0.2.65 (Brown et al., 2025) (using explore parameters: --distance 0.02 --minimum 5 --maximum 12; version: 0.2.4). The resulting list of candidate telomere k-mers was then processed with seqtk telo (version 1.4; https://github.com/lh3/seqtk), identifying the true telomeric kmer sequence and annotating all chromosomes for their telomere regions. Assembly completeness was assessed by BUSCO with compleasm v0.2.7 (Huang and Li, 2023), using

the Eudicotyledons OrthoDB v12 dataset (Tegenfeldt et al., 2025). To assess base accuracy of the genome, yak v0.1 (r56) was used to compare k-mer spectra (k=31) between the assembly and sequencing reads to calculate an assembly quality value (QV) (Cheng et al., 2021).

Additional T2T genome assemblies

In order to confirm the conservation of karyotype, we sampled two more interstate samples for T2T genome assembly. These included a sample from the Tasmanian Herbarium, Sandy Bay, Tasmania (first draft genome reported in Ferguson et al., 2024) and a sample from Trentham, Victoria (accession GDH178). HMW DNA was extracted, size-selected (≥ 20 kb), and sequenced using FLO-PRO114M R10.4.1 PromethION flow cells as described above. Basecalling was performed with Dorado v7.6.8 (sup v4.3.0, 400 bps). Raw reads were filtered for ≥ Q20 and ≥ 20 kb using chopper v0.8.0 (De Coster and Rademakers, 2023). *De novo* genome assemblies were performed with hifiasm (ONT) v0.25.0-r726 (Cheng et al., 2025) with ONT simplex reads (--ont), adding the telomere sequence (--telo-m AAACCCT), and dual scaffolding (--dual-scaf). Only chromosome two remained broken. Final scaffolding and ordering of chromosomes was performed with RagTag v2.1.0 (Alonge et al., 2022), using the *E. viminalis* ACT genome as the reference. Quality assessment of the genomes was performed as described above for the ACT *E. viminalis* reference.

Haplotype-resolved genome synteny analyses

To analyse the conservation of genome architecture and identify structural variants across *E. viminalis*, the highest quality genome assemblies were selected for phylogenetic and comparative genomic analyses. Samples were collected from four states and territories across Australia to capture geographic and genetic diversity. Only haplotype-resolved, chromosome-scale assemblies with > 92% BUSCO completeness were included to ensure accurate phylogenetic inference and detection of syntenic regions and structural rearrangements. These assemblies were subsequently used for pangenome construction and analysis.

A phylogenetic tree was constructed using highly conserved single-copy BUSCO genes via the BUSCO Phylogenomics workflow (https://github.com/jamiemcg/BUSCO_phylogenomics). The genome of closely related *E. globulus* was included as an outgroup (Ferguson et al., 2024c). BUSCO v5.8.2 was used to identify the single-copy genes and evaluate completeness of the assemblies, using the eudicotyledons odb12 dataset (Tegenfeldt et al., 2025). Protein sequences present in all genomes were aligned with Muscle5 v5.1 (Edgar, 2022), and poorly aligned regions trimmed using trimAl v1.4.rev15 (Capella-Gutiérrez et al., 2009). A total of 2,710 complete and single-copy BUSCO sequences in ≥4 genomes remained (97% of the 2,805 BUSCOs). Maximum likelihood gene trees were inferred with IQ-TREE v2.3.6 (Minh et al., 2020), and the species tree estimated under the multi-species coalescent model using ASTRAL-IV v1.23.4.6 (C. Zhang et al., 2025). The tree was generated and manually rooted on *E. globulus* (not displayed) using Figtree v1.4.5 (http://tree.bio.ed.ac.uk/software/figtree/).

To assess synteny and identify structural variations between genomes, pairwise alignments were performed using the MUMmer4 v4.0.1 DNA alignment tool nucmer (--maxmatch -l 50 -b 500 -c 200) (Marçais et al., 2018). Alignments were filtered to retain only those ≥200 bp in length and ≥95% sequence identity using the MUMmer4 delta-filter tool. Synteny and structural variants were identified with SyRI v1.6.3 (Goel et al., 2019), and visualised with plotsr v1.1.1 (Goel and Schneeberger, 2022). Across the distributed provinces of these assembled individuals, some levels of non-synteny were found but the geography did not seem to underpin any genetic isolations.

Direct RNA sequencing and genome annotation

Total RNA was extracted from the cryogenically frozen leaf tissue of the *Eucalyptus viminalis* reference tree (described above), using a cetrimonium bromide lysis buffer and lithium chloride to precipitate RNA, described in (Chang et al., 1993). The RNA was quality checked on a NanoDrop 1000 Spectrophotometer (Thermo Fisher), a TapeStation 4200 automated electrophoresis instrument with a RNA ScreenTape (Agilent Technologies, 5067-5576), and a Qubit 4 Fluorometer with a Qubit RNA Integrity and Quality (IQ) assay (Thermo Fisher, Q33221), following the manufacturer's instructions. Poly(A) transcripts were then purified with Oligo $(dT)_{25}$ magnetic beads (NEB S1419S), to enrich intact mRNA. An ONT library direct RNA sequencing library was prepared (SQK-RNA004), using Induro Reverse Transcriptase (NEB M0681) at 60°C for 15 min, to create stable full-length RNA-cDNA molecules (to enhance length and output of the RNA that is sequenced). Direct RNA sequencing was performed on the ONT PromethION P24 platform using a FLO-PRO004RA flow cell. Only the RNA strand of the RNA-cDNA hybrid is sequenced. When sequencing declined, the flow cell was treated once with ONT Flow Cell Wash Kit (EXP-WSH004), using 5 µL DNAse I (NEB M0303) and 5 µL RNase A (Thermo Fisher 12091039). The flow cell was then re-primed and reloaded with more libraries following the manufacturer's instructions. Sequencing data (pod5) was basecalled to fastq with ONT Dorado v7.6.8, using the super accurate model (sup v5.0.0, 130 bps). Sequence reads were filtered for read lengths ≥ 200 bp and mean read quality ≥ Q10 using Chopper v0.10.0 (De Coster and Rademakers, 2023). Quality checks were performed before and after filtering with NanoStat 1.6.0 (De Coster et al., 2018).

Before annotating genes, both haplotypes were repeat-masked to aid in gene prediction. First, *de novo* repeat libraries were created for each haplotype using EDTA (version: 1.9.6) (Ou et al., 2019), and subsequently repeat annotation was performed with RepeatMasker (version: 4.0.9; https://www.repeatmasker.org). Using these repeat annotations, both haplotypes were soft-masked. Initial transcript models were constructed using filtered long-read RNA data and IsoQuant (version 3.3.1; parameters: --report_novel_unspliced true --stranded forward --model_construction_strategy sensitive_ont (Prjibelski et al., 2023). Based on these RNA-predicted transcripts, gene annotation was performed using BRAKER3 (version 3.0.8) (Gabriel et al., 2024).

Species-specific basecaller training and basecalling

Training a species-specific basecaller improves the accuracy of ONT sequencing reads as well as the accuracy of their Q-scores, particularly in plant species (Ferguson et al., 2022).

As such, a basecaller for *E. viminalis* was trained for population analyses. The *E. viminalis* basecaller model was trained using the ONT assembly reads and our combinatorial PacBio HiFi, ONT and HiC genome as the truth set. To aid in distinguishing true polymorphisms from sequencing errors, our two haplotypes were joined, creating a diploid truth set.

As ONT sequencing chemistry advanced, we had both 5 kHz and 4 kHz ONT data, accordingly two basecallers were trained, one for each data type. All pod5 reads were randomly sampled to capture variable read lengths. For the 4 kHz model, Bonito was initially used to basecall three read batches of 220 fast5 files (model: dna_r10.4.1_e8.2_400bps_hac@v4.1.0; version: 0.7.2; https://github.com/nanoporetech/bonito). Next, each read subset was used to iteratively train a basecaller model using Bonito with the following parameters: --epochs 16 --batch 160 --chunks 0 --lr 15e-5 --save-optim-every 2 --valid-chunks 112. The 5 kHz model was trained similarly, basecalling with three read subsets of 148 fast5 files using model dna_r10.4.1_e8.2_400bps_hac@v4.2.0 and the same Bonito parameters, with the exception of --epochs 8 and --batch 170. Different training parameters were used because 5 kHz reads contained more data and had a different read length distribution.

Raw population sequencing files in pod5 file format were basecalled with the *E. viminalis* basecaller model and demultiplexed with Dorado (version 0.5.0, linux-x64) (https://github.com/nanoporetech/dorado). The output read FASTQ files were filtered to remove fragmented and low-quality reads using Fastq-filter (version 0.3.0) (https://github.com/LUMC/fastq-filter) (-q 7 -l 1000). Sequencing results were summarised in Supplementary table 7.

Pangenome analyses

In order to compare sequences and genome synteny across multiple haplotype-resolved genomes, T2T genomes described above and several high-quality genome assemblies of environmental samples (described in later sections) were annotated with Liftoff (version 1.6.3) (Shumate and Salzberg, 2021) using the reference genome annotations. The annotations of each haplotype genomes were then grouped into orthogroups with OrthoFinder (version 2.5.5) (Emms and Kelly, 2019) based on peptide sequence similarity and the DIAMOND database (Buchfink et al., 2021). Then, using the orthogroups as anchors across the genome, gene-centric synteny was analysed with GENESPACE (version 1.2.3) (https://github.com/jtlovell/GENESPACE). Additionally, saturations of novel orthogroups with the number of included genomes were also visualised.

Population sampling

Leaf tissues from 71 wild *E. viminalis* individuals were collected across the natural range of this species from New South Wales, Australia, between 2019 and 2021 for ONT long-read sequencing. These 71 were a subset of a total of 434 samples of *E. viminalis* and two closely related species - *E. dalrympleana* and *E. rubida* - collected for short-read sequencing. Samples derived from three geographically separated populations (populations A-C) delineated by latitude breakpoints of -34.42 and -31.80. Notably, population A is located near the Monaro region where a large-scale dieback event recently occurred (Ross and

Brack, 2015). The collected leaf samples were silica-dried and stored in air-tight containers for long-term storage. All samples collected in this study originated from woodlands as characterised by their rainfall and temperature conditions.

High-molecular weight DNA extractions of landscape leaf samples

To extract high-molecular weight DNA from the silica-dried leaf samples, a custom protocol was followed. Approximately 40 x 3 mm leaf Uni-Core punches (Qiagen, WB100078) and three 1/8" (3.175 mm) ball bearings were added to a 2 mL microcentrifuge tube for each sample. Samples were frozen under liquid nitrogen and homogenised using a TissueLyser II (Qiagen). The DNA was then extracted with a magnetic bead-based protocol, described in (Jones et al., 2021). In brief, the homogenised tissues were washed with a 0.35 M D-sorbitol solution (DTT freshly added when use), lysed with sodium dodecyl sulphate (SDS) based lysis buffer, washed with Chloroform: Isoamyl alcohol (24:1), and the DNA was bound to Sera-Mag SpeedBead Carboxylate-Modified beads (Cytiva 65152105050250) in the presence of binding buffer (20% PEG 8,000 + 3 M NaCl). A NaCl-based Long fragment buffer (LFB) wash step was introduced (0.75M NaCl + 5% PEG 8,000) to remove small DNA fragments (selecting for ~3 kb and above). This greatly aided barcoding efficiency and read lengths. The HMW DNA was eluted from beads with 50 µL of 10 mM Tris-HCl pH 8. The extracted HMW DNA was quantified with Qubit Broad Range Assay Kits (Thermo Fisher Scientific) and their purities were evaluated with the NanoDrop spectrophotometer (Thermo Fisher Scientific). The HMW DNA was stored at 4°C until sequencing. Quality check of DNA was performed on six representative samples, using a Femto Pulse System (Agilent Technologies) in gDNA mode.

Long-read library preparation and sequencing

Barcoding was performed with the Oxford Nanopore Technologies (ONT) SQK-NBD114.96 according to the manufacturer's instructions. In brief, 2 µg DNA was aliquoted for each sample and added Ultra II End-prep enzyme and FFPE DNA Repair enzyme (New England BioLabs). The incubation of the repair step was doubled due to the DNA being highly damaged from the silica drying process and storage condition. A barcode was added to each sample (2.5 µL), after which every 12 samples were pooled to be sequenced in batches.The pooled samples were size-selected with BluePippin (Sage Science) for fragments 8 kb and above. An ONT adapter was ligated to the barcoded pool and DNA sequencing was performed on the PromethION platform using a FLO-PRO114M R10.4.1 flow cell following the manufacturer’s instructions. Each flow cell was flushed three times (or until the pore viability had diminished), where more sequencing libraries were added.

To assemble genomes of *E. viminalis* sourced across the landscape, ONT long-reads sequenced with the R10.4.1 chemistry was passed to Hifiasm (version 0.24.0-r702) (H. Cheng et al., 2025). Hifiasm 0.24.0-r702 is able to take the raw reads from ONT sequenced with R10 chemistry or later (--ont --hom-cov [0.8 x estimated coverage] --telo-m AAACCCT --dual-scaf). Hifiasm then performs error corrections in adaptor and homopolymer regions as well as base corrections according to ONT sequencing error profiles. The completeness of

the assembled contigs were then assessed with Compleasm (version 0.2.6) (Huang and Li, 2023) and reported in Supplementary table 8.

Structural variant discovery and genotyping

The filtered long-read sequences from each sample were mapped to the reference genome (haplotype 1) with Minimap2 (version 2.24) (Li, 2018) to generate Binary Alignment Map (BAM) files for all 71 sequenced samples. The output BAM files are then sorted and indexed with SAMtools (version 1.9) (Danecek et al., 2021b). After the appropriate level of genome-wide coverage was determined for both SV discovery and genotyping (Supplementary Note 8), 34 samples with genome-wide coverage >12x were included for SV discovery with Sniffles2 (version 2.0.7) (Smolka et al., 2024) (n = 34; population A: 18, B: 1, C: 15), and 49 samples were included in the population genetic studies of SV genotype allele frequency as well as subsequent GEA analyses.

SVs were merged based on breakpoint locations, percentage overlaps (relative to SV size), and in some cases, sequence similarities. Testing different SV merging criteria revealed that the number of mergeable SVs monotonically decreases as merging criteria becomes increasingly stringent (Supplemental Note 8). To retain SV diversity without inflating allele frequencies of certain SVs, intermediate levels of merging criteria were chosen. For the majority of SVs, a threshold of 0.2 (20%) maximum linear distance (in relation to SV size) between breakpoint positions was chosen. For insertions > 1 kb (SVTYPE == 'INS', SVLEN >= 1000), a sequence similarity of 0.75 (75%) were added in addition to the maximum distance criterion to separate large insertions with similar breakpoint positions, but have different sequence content novel to the reference genome. Merged SVs were genotyped with CuteSV (version 2.0.3; -Ivcf) (Jiang et al., 2020) in all 71 sequenced individuals in order to remove singletons while retaining maximum number of SVs, resulting. Singleton SVs were removed with BCFtools (version 1.12) (Danecek et al., 2021b) by variant ID (-m id). Then, variants found in fewer than 2 individuals (i.e., singletons) were filtered out, resulting in 573,889 long-read discovered structural variants that are 1) found by both Sniffles2 and CuteSV, and 2) found in at least two biological samples.

Associating SVs with environmental variables

Prior to performing a genomic-environment association (GEA) study with structural variants, SVs were filtered at high stringency to ensure quality. First, only SVs supported by a minimum of three reads during SV genotyping were maintained: the rest were treated as missing data. Second, SV loci with more than 10% missing data were filtered out. Last, rare SVs (minor allele frequency < 0.02) were filtered out. In order to keep the sample individuals comparable, only the 49 trees with > 8x genome-wide coverage were used in the GEA.

Tests for associations between SV polymorphisms and environmental variables were implemented in GEMMA (version 0.98.3) (Zhou and Stephens, 2012). The sample relatedness matrix (or genomic relation matrix, GRM) was calculated using genome-wide SV genotypes (option -gk). For each of the 49 tree locations, long-term climate variables at 2.5 minutes spatial resolution were extracted from the BioClim data base (Fick and Hijmans, 2017) using the R package geospatial (https://github.com/rspatial/geodata). Then, a linear

mixed effect model (-lmm) was fitted by using SV allele frequencies to predict BioClim variables using the GRM information as the random effect in order to account for neutral variation in allele frequencies. After the GEA was performed, the Bonferroni threshold (p-value < 10^{-12}) was applied to identify putatively adaptive loci.

Short-read validation of climate adaptive loci

Leaf tissue was collected from 178 *E. viminalis*, 134 *E. dalrympleana*, and 122 *E. rubida* (total of 434) samples (metadata in Supplementary Table 16). These included the 49 samples that were long-read sequenced above. Leaves were added to envelopes and dried on silica gel for long-term storage. Approximately twenty 3 mm leaf punches (Qiagen Uni-Core, WB100078) and one 3 mm ball bearing per sample were added to 96-well arrayed 1.1 mL tubes (Axygen Scientific, MTS-11-8-C-R-10 and MTS-8CP-C-1). Samples were frozen in liquid nitrogen and homogenized using TissueLyser II (Qiagen). DNA was extracted with a NucleoSpin 96 Plant II 24×96 kit using CTAB lysis buffer PL1 according to the manufacturer's instructions (Macherey-Nagel, 740663.24).

Whole-genome short-read sequencing libraries were prepared using our cost-optimized transposase protocol (Jones et al., 2023), based on Illumina DNA Prep (M) Tagmentation (24 Samples, IPB) (Cat# 20060060) with bead-linked transposase BLT (Illumina document# 1000000033561 v05). Fragmented DNA was amplified using custom index primers with 14 PCR cycles. Libraries were pooled and size-selected for 300-500 bp inserts using PippinHT (Sage Science). Sequencing was performed on NovaSeq 6000 (Illumina) using an S4 flow cell with a 300-cycle kit (150 bp paired-end) at the Biomolecular Resource Facility, Australian National University, Canberra, Australia.

Variants were called using Acanthophis (Murray et al., 2024), using AdapterRemoval2 (Schubert et al., 2016), BWA MEM (Li, 2013), bcftools (Danecek et al., 2021a), DeepVariant (Poplin et al., 2018), GLNexus (Yun et al., 2020), and Snakemake (Köster and Rahmann, 2012) against the reference genome. Variants were filtered with SNPSnip (https://pypi.org/project/snpsnip/), VCFtools (Danecek et al., 2011) and R scripts. SNPs were filtered out if they had a quality score of <24, a minor allele frequency of <0.10 and a missingness rate of >0.7 across the 434 samples.

For the 49 samples that had both long-read and short-read data, correlations between SV-based *CHILL1* genotype and SNP genotypes at individual sites within the 27.0 to 27.9 Mb wider *CHILL1* region were computed in order to discover potentially diagnostic SNPs for *CHILL1*. Principal components (PCs) based on SNPs within the region containing 5 adjacent SVs (positions 27.8 to 27.8 Mb) and where the highest correlations occurred, were computed using the pcaMethods package in R (Stacklies et al., 2007)hreshold values of the first PC that reliably distinguished *CHILL1* genotypes(Supplementary Note 13).

References

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z.B., Schatz, M.C., Soyk, S., 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 23, 258.

Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell 182, 145–161.

Ardern, Z., 2023. Alternative reading frames are an underappreciated source of protein sequence novelty. J. Mol. Evol. 91, 570–580.

Baid, G., Cook, D.E., Shafin, K., Yun, T., Llinares-López, F., Berthet, Q., Belyaeva, A., Töpfer, A., Wenger, A.M., Rowell, W.J., Yang, H., Kolesnikov, A., Ammar, W., Vert, J.-P., Vaswani, A., McLean, C.Y., Nattestad, M., Chang, P.-C., Carroll, A., 2023. DeepConsensus improves the accuracy of sequences with a gap-aware sequence transformer. Nat. Biotechnol. 41, 232–238. https://doi.org/10.1038/s41587-022-01435-7

Balkenhol, N., Dudaniec, R.Y., Krutovsky, K.V., Johnson, J.S., Cairns, D.M., Segelbacher, G., Selkoe, K.A., von der Heyden, S., Wang, I.J., Selmoni, O., Joost, S., 2019. Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations, in: Rajora, O.P. (Ed.), Population Genomics: Concepts, Approaches and Applications. Springer International Publishing, Cham, pp. 261–322. https://doi.org/10.1007/13836_2017_2

Ballesta, P., Maldonado, C., Pérez-Rodríguez, P., Mora, F., 2019. SNP and Haplotype-Based Genomic Selection of Quantitative Traits in Eucalyptus globulus. Plants 8, 331. https://doi.org/10.3390/plants8090331

Benoit, M., Jenike, K.M., Satterlee, J.W., Ramakrishnan, S., Gentile, I., Hendelman, A., Passalacqua, M.J., Suresh, H., Shohat, H., Robitaille, G.M., 2025. Solanum pan-genetics reveals paralogues as contingencies in crop engineering. Nature 1–11.

Brancalion, P.H.S., Hua, F., Joyce, F.H., Antonelli, A., Holl, K.D., 2025. Moving biodiversity from an afterthought to a key outcome of forest restoration. Nat. Rev. Biodivers. 1, 248–261. https://doi.org/10.1038/s44358-025-00032-1

Brown, M.R., Manuel Gonzalez de La Rosa, P., Blaxter, M., 2025. Tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. Bioinformatics 41, btaf049.

Bubb, K.L., Hamm, M.O., Tullius, T.W., Min, J.K., Ramirez-Corona, B., Mueth, N.A., Ranchalis, J., Mao, Y., Bergstrom, E.J., Vollger, M.R., 2025. The regulatory potential of transposable elements in maize. Nat. Plants 1–12.

Buchfink, B., Reuter, K., Drost, H.-G., 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods 18, 366–368.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Chang, S., Puryear, J., Cairney, J., 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Report. 11, 113–116. https://doi.org/10.1007/BF02670468

Cheng, H., Asri, M., Lucas, J., Koren, S., Li, H., 2024. Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. Nat. Methods 21, 967–970. https://doi.org/10.1038/s41592-024-02269-8

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., Li, H., 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods 18, 170–175.

Cheng, H., Qu, H., McKenzie, S., Lawrence, K.R., Windsor, R., Vella, M., Park, P.J., Li, H., 2025. Efficient near telomere-to-telomere assembly of Nanopore Simplex reads. https://doi.org/10.1101/2025.04.14.648685

Cheng, L., Wang, N., Bao, Z., Zhou, Q., Guarracino, A., Yang, Y., Wang, P., Zhang, Z., Tang, D., Zhang, P., Wu, Y., Zhou, Yao, Zheng, Y., Hu, Y., Lian, Q., Ma, Z., Lassois, L., Zhang, C., Lucas, W.J., Garrison, E., Stein, N., Städler, T., Zhou, Yongfeng, Huang, S., 2025. Leveraging a phased pangenome for haplotype design of hybrid potato. Nature 640, 408–417. https://doi.org/10.1038/s41586-024-08476-9

Costa e Silva, J., Borralho, N.M.G., Araújo, J.A., Vaillancourt, R.E., Potts, B.M., 2009. Genetic parameters for growth, wood density and pulp yield in Eucalyptus globulus. Tree Genet. Genomes 5, 291–305. https://doi.org/10.1007/s11295-008-0174-9

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., 2021a. Twelve years of SAMtools and BCFtools. Gigascience 10, giab008.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021b. Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. https://doi.org/10.1093/gigascience/giab008

De Coster, W., D'hert, S., Schultz, D.T., Cruts, M., Van Broeckhoven, C., 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34, 2666–2669.

De Coster, W., Rademakers, R., 2023. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics 39, btad311. https://doi.org/10.1093/bioinformatics/btad311

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356, 92–95.

Durand, Neva C, Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., Aiden, E.L., 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3, 99–101.

Durand, Neva C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., Aiden, E.L., 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

Edgar, R.C., 2022. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat. Commun. 13, 6968.

Edwards, S.V., Fang, B., Khost, D., Kolyfetis, G.E., Cheek, R.G., DeRaad, D.A., Chen, N., Fitzpatrick, J.W., McCormack, J.E., Funk, W.C., 2025. Multispecies pangenomes reveal a pervasive influence of population size on structural variation. Science 390, eadw1931.

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238. https://doi.org/10.1186/s13059-019-1832-y

Fedoroff, N.V., 2012. Transposable Elements, Epigenetics, and Genome Evolution. Science 338, 758–767. https://doi.org/10.1126/science.338.6108.758

Ferguson, S., Jones, A., Murray, K., Andrew, R., Schwessinger, B., Borevitz, J., 2024a. Plant genome evolution in the genus Eucalyptus is driven by structural rearrangements that promote sequence divergence. Genome Res. 34, 606–619. https://doi.org/10.1101/gr.277999.123

Ferguson, S., Jones, A., Murray, K., Andrew, R.L., Schwessinger, B., Bothwell, H., Borevitz, J., 2024b. Exploring the role of polymorphic interspecies structural variants in reproductive isolation and adaptive divergence in Eucalyptus. GigaScience 13, giae029. https://doi.org/10.1093/gigascience/giae029

Ferguson, S., Jones, A., Murray, K., Andrew, R.L., Schwessinger, B., Bothwell, H., Borevitz, J., 2024c. Exploring the role of polymorphic interspecies structural variants in reproductive isolation and adaptive divergence in Eucalyptus. GigaScience 13, giae029. https://doi.org/10.1093/gigascience/giae029

Ferguson, S., Jones, A., Murray, K., Schwessinger, B., Borevitz, J.O., 2022. Interspecies genome divergence is predominantly due to frequent small scale rearrangements in Eucalyptus. Mol. Ecol.
Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315.
Gabriel, L., Brůna, T., Hoff, K.J., Ebel, M., Lomsadze, A., Borodovsky, M., Stanke, M., 2024. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. 34, 769–777.
Garrison, E., Guarracino, A., Heumos, S., Villani, F., Bao, Z., Tattini, L., Hagmann, J., Vorbrugg, S., Marco-Sola, S., Kubica, C., 2024. Building pangenome graphs. Nat. Methods 21, 2008–2012.
Goel, M., Schneeberger, K., 2022. Plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics 38, 2922–2926.
Goel, M., Sun, H., Jiao, W.-B., Schneeberger, K., 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20, 277.
Gustafson, J.A., Gibson, S.B., Damaraju, N., Zalusky, M.P., Hoekzema, K., Twesigomwe, D., Yang, L., Snead, A.A., Richmond, P.A., De Coster, W., 2024. High-coverage nanopore sequencing of samples from the 1000 Genomes Project to build a comprehensive catalog of human genetic variation. Genome Res. 34, 2061–2073.
Hallik, L., Kazantsev, T., Kuusk, A., Galmés, J., Tomás, M., Niinemets, Ü., 2017. Generality of relationships between leaf pigment contents and spectral vegetation indices in Mallorca (Spain). Reg. Environ. Change 17, 2097–2109.
Harringmeyer, O.S., Hoekstra, H.E., 2022. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. Nat. Ecol. Evol. 6, 1965–1979. https://doi.org/10.1038/s41559-022-01890-0
Hof, A.E. van't, Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby, A.C., Saccheri, I.J., 2016. The industrial melanism mutation in British peppered moths is a transposable element. Nature 534, 102–105. https://doi.org/10.1038/nature17951
Höps, W., Rausch, T., Jendrusch, M., Korbel, J.O., Sedlazeck, F.J., 2024. Impact and characterization of serial structural variations across humans and great apes. Nat. Commun. 15, 8007.
Hou, E., Luo, Y., Kuang, Y., Chen, C., Lu, X., Jiang, L., Luo, X., Wen, D., 2020. Global meta-analysis shows pervasive phosphorus limitation of aboveground plant production in natural terrestrial ecosystems. Nat. Commun. 11, 637.
Hu, M., Wan, P., Chen, C., Tang, S., Chen, Jiahao, Wang, L., Chakraborty, M., Zhou, Y., Chen, Jinfeng, Gaut, B.S., 2025. Accurate, scalable structural variant genotyping in complex genomes at population scales. Mol. Biol. Evol. 42, msaf180.
Huang, N., Li, H., 2023. compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics 39, btad595. https://doi.org/10.1093/bioinformatics/btad595
Huang, Y., Gao, Z.Y., Ly, K., Lin, L., Lambooij, J.-P., King, E.G., Janssen, A., Wei, K.H.-C., Lee, Y.C.G., 2025. Polymorphic transposable elements contribute to variation in recombination landscapes. Proc. Natl. Acad. Sci. 122, e2427312122.
Igolkina, A.A., Vorbrugg, S., Rabanal, F.A., Liu, H.-J., Ashkenazy, H., Kornienko, A.E., Fitz, J., Collenberg, M., Kubica, C., Mollá Morales, A., Jaegle, B., Wrightsman, T., Voloshin, V., Bezlepsky, A.D., Llaca, V., Nizhynska, V., Reichardt, I., Bezrukov, I., Lanz, C., Bemm, F., Flood, P.J., Nemomissa, S., Hancock, A., Guo, Y.-L., Kersey, P., Weigel, D., Nordborg, M., 2025. A comparison of 27 Arabidopsis thaliana genomes and the path toward an unbiased characterization of genetic polymorphism. Nat. Genet. 57, 2289–2301. https://doi.org/10.1038/s41588-025-02293-0
Jayakodi, M., Lu, Q., Pidon, H., Rabanus-Wallace, M.T., Bayer, M., Lux, T., Guo, Y., Jaegle, B., Badea, A., Bekele, W., Brar, G.S., Braune, K., Bunk, B., Chalmers, K.J., Chapman, B., Jørgensen, M.E., Feng, J.-W., Feser, M., Fiebig, A., Gundlach, H., Guo, W., Haberer, G., Hansson, M., Himmelbach, A., Hoffie, I., Hoffie, R.E., Hu, H.,

Isobe, S., König, P., Kale, S.M., Kamal, N., Keeble-Gagnère, G., Keller, B., Knauft, M., Koppolu, R., Krattinger, S.G., Kumlehn, J., Langridge, P., Li, C., Marone, M.P., Maurer, A., Mayer, K.F.X., Melzer, M., Muehlbauer, G.J., Murozuka, E., Padmarasu, S., Perovic, D., Pillen, K., Pin, P.A., Pozniak, C.J., Ramsay, L., Pedas, P.R., Rutten, T., Sakuma, S., Sato, K., Schüler, D., Schmutzer, T., Scholz, U., Schreiber, M., Shirasawa, K., Simpson, C., Skadhauge, B., Spannagl, M., Steffenson, B.J., Thomsen, H.C., Tibbits, J.F., Nielsen, M.T.S., Trautewig, C., Vequaud, D., Voss, C., Wang, P., Waugh, R., Westcott, S., Rasmussen, M.W., Zhang, R., Zhang, X.-Q., Wicker, T., Dockter, C., Mascher, M., Stein, N., 2024. Structural variation in the pangenome of wild and domesticated barley. Nature 1–9. https://doi.org/10.1038/s41586-024-08187-1

Jiang, T., Liu, Yongzhuang, Jiang, Y., Li, J., Gao, Y., Cui, Z., Liu, Yadong, Liu, B., Wang, Y., 2020. Long-read-based human genomic structural variation detection with cuteSV. Genome Biol. 21, 189. https://doi.org/10.1186/s13059-020-02107-y

Jiao, C., Xie, X., Hao, C., Chen, L., Xie, Y., Garg, V., Zhao, L., Wang, Z., Zhang, Y., Li, T., Fu, J., Chitikineni, A., Hou, J., Liu, H., Dwivedi, G., Liu, X., Jia, J., Mao, L., Wang, X., Appels, R., Varshney, R.K., Guo, W., Zhang, X., 2024. Pan-genome bridges wheat structural variations with habitat and breeding. Nature 1–10. https://doi.org/10.1038/s41586-024-08277-0

Jones, A., Stanley, D., Ferguson, S., Schwessinger, B., Borevitz, J., Warthmann, N., 2023. Cost-conscious generation of multiplexed short-read DNA libraries for whole-genome sequencing. PLoS One 18, e0280004.

Jones, A., Torkel, C., Stanley, D., Nasim, J., Borevitz, J., Schwessinger, B., 2021. High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing. PLOS ONE 16, e0253830. https://doi.org/10.1371/journal.pone.0253830

Kang, M., Wu, H., Liu, H., Liu, Wenyu, Zhu, M., Han, Y., Liu, Wei, Chen, C., Song, Y., Tan, L., Yin, K., Zhao, Y., Yan, Z., Lou, S., Zan, Y., Liu, J., 2023. The pan-genome and local adaptation of Arabidopsis thaliana. Nat. Commun. 14, 6259. https://doi.org/10.1038/s41467-023-42029-4

Köster, J., Rahmann, S., 2012. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28, 2520–2522.

Krueger, F., James, F., Ewels, P., Afyounian, E., Weinstein, M., Schuster-Boeckler, B., Hulselmans, G., sclamons, 2023. FelixKrueger/TrimGalore: v0.6.10 - add default decompression path. https://doi.org/10.5281/zenodo.7598955

Kuderna, L.F., Gao, H., Janiak, M.C., Kuhlwilm, M., Orkin, J.D., Bataillon, T., Manu, S., Valenzuela, A., Bergman, J., Rousselle, M., 2023. A global catalog of whole-genome diversity from 233 primate species. Science 380, 906–913.

Lai, Y.-T., Yeung, C.K.L., Omland, K.E., Pang, E.-L., Hao, Y., Liao, B.-Y., Cao, H.-F., Zhang, B.-W., Yeh, C.-F., Hung, C.-M., Hung, H.-Y., Yang, M.-Y., Liang, W., Hsu, Y.-C., Yao, C.-T., Dong, L., Lin, K., Li, S.-H., 2019. Standing genetic variation as the predominant source for adaptation of a songbird. Proc. Natl. Acad. Sci. 116, 2152–2157. https://doi.org/10.1073/pnas.1813597116

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

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv Prepr. ArXiv13033997.

Li, X., Blackman, C.J., Choat, B., Duursma, R.A., Rymer, P.D., Medlyn, B.E., Tissue, D.T., 2018. Tree hydraulic traits are coordinated and strongly linked to climate-of-origin across a rainfall gradient. Plant Cell Environ. 41, 646–660. https://doi.org/10.1111/pce.13129

Lian, Q., Huettel, B., Walkemeier, B., Mayjonade, B., Lopez-Roques, C., Gil, L., Roux, F., Schneeberger, K., Mercier, R., 2024. A pan-genome of 69 Arabidopsis thaliana accessions reveals a conserved genome structure throughout the global species range. Nat. Genet. 56, 982–991.

Liao, W.-W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J.K., Monlong, J., Abel, H.J., Buonaiuto, S., Chang, X.H., Cheng, H., Chu, J., Colonna, V., Eizenga, J.M., Feng, X., Fischer, C., Fulton, R.S., Garg, S., Groza, C., Guarracino, A., Harvey, W.T., Heumos, S., Howe, K., Jain, M., Lu, T.-Y., Markello, C., Martin, F.J., Mitchell, M.W., Munson, K.M., Mwaniki, M.N., Novak, A.M., Olsen, H.E., Pesout, T., Porubsky, D., Prins, P., Sibbesen, J.A., Sirén, J., Tomlinson, C., Villani, F., Vollger, M.R., Antonacci-Fulton, L.L., Baid, G., Baker, C.A., Belyaeva, A., Billis, K., Carroll, A., Chang, P.-C., Cody, S., Cook, D.E., Cook-Deegan, R.M., Cornejo, O.E., Diekhans, M., Ebert, P., Fairley, S., Fedrigo, O., Felsenfeld, A.L., Formenti, G., Frankish, A., Gao, Y., Garrison, N.A., Giron, C.G., Green, R.E., Haggerty, L., Hoekzema, K., Hourlier, T., Ji, H.P., Kenny, E.E., Koenig, B.A., Kolesnikov, A., Korbel, J.O., Kordosky, J., Koren, S., Lee, H., Lewis, A.P., Magalhães, H., Marco-Sola, S., Marijon, P., McCartney, A., McDaniel, J., Mountcastle, J., Nattestad, M., Nurk, S., Olson, N.D., Popejoy, A.B., Puiu, D., Rautiainen, M., Regier, A.A., Rhie, A., Sacco, S., Sanders, A.D., Schneider, V.A., Schultz, B.I., Shafin, K., Smith, M.W., Sofia, H.J., Abou Tayoun, A.N., Thibaud-Nissen, F., Tricomi, F.F., Wagner, J., Walenz, B., Wood, J.M.D., Zimin, A.V., Bourque, G., Chaisson, M.J.P., Flicek, P., Phillippy, A.M., Zook, J.M., Eichler, E.E., Haussler, D., Wang, T., Jarvis, E.D., Miga, K.H., Garrison, E., Marschall, T., Hall, I.M., Li, H., Paten, B., 2023. A draft human pangenome reference. Nature 617, 312–324. https://doi.org/10.1038/s41586-023-05896-x

Logsdon, G.A., Ebert, P., Audano, P.A., Loftus, M., Porubsky, D., Ebler, J., Yilmaz, F., Hallast, P., Prodanov, T., Yoo, D., 2025. Complex genetic variation in nearly complete human genomes. Nature 644, 430–441.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., Zimin, A., 2018. MUMmer4: A fast and versatile genome alignment system. PLoS Comput. Biol. 14, e1005944.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534.

Moore, B.D., Foley, W.J., 2005. Tree use by koalas in a chemically complex landscape. Nature 435, 488–490. https://doi.org/10.1038/nature03551

Murray, K.D., Borevitz, J.O., Weigel, D., Warthmann, N., 2024. Acanthophis: a comprehensive plant hologenomics pipeline. J. Open Source Softw. 9, 6062.

Murray, K.D., Janes, J.K., Jones, A., Bothwell, H.M., Andrew, R.L., Borevitz, J.O., 2019. Landscape drivers of genomic diversity and divergence in woodland Eucalyptus. Mol. Ecol. 28, 5232–5247. https://doi.org/10.1111/mec.15287

Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., Goodstein, D.M., Dubchak, I., Poliakov, A., Mizrachi, E., Kullan, A.R.K., Hussey, S.G., Pinard, D., van der Merwe, K., Singh, P., van Jaarsveld, I., Silva-Junior, O.B., Togawa, R.C., Pappas, M.R., Faria, D.A., Sansaloni, C.P., Petroli, C.D., Yang, X., Ranjan, P., Tschaplinski, T.J., Ye, C.-Y., Li, T., Sterck, L., Vanneste, K., Murat, F., Soler, M., Clemente, H.S., Saidi, N., Cassan-Wang, H., Dunand, C., Hefer, C.A., Bornberg-Bauer, E., Kersting, A.R., Vining, K., Amarasinghe, V., Ranik, M., Naithani, S., Elser, J., Boyd, A.E., Liston, A., Spatafora, J.W., Dharmwardhana, P., Raja, R., Sullivan, C., Romanel, E., Alves-Ferreira, M., Külheim, C., Foley, W., Carocha, V., Paiva, J., Kudrna, D., Brommonschenkel, S.H., Pasquali, G., Byrne, M., Rigault, P., Tibbits, J., Spokevicius, A., Jones, R.C., Steane, D.A., Vaillancourt, R.E., Potts, B.M., Joubert, F., Barry, K., Pappas, G.J., Strauss, S.H., Jaiswal, P., Grima-Pettenati, J., Salse, J., Van de Peer, Y., Rokhsar, D.S., Schmutz, J., 2014. The genome of Eucalyptus grandis. Nature 510, 356–362. https://doi.org/10.1038/nature13308

Nicolle, D., Ritter, M.K., Jones, R.C., Phillips, G.P., French, M.E., Cumming, R., Bell, S.A., 2025. The genus problem–Eucalyptus as a model system for minimising taxonomic disruption. Taxon 74, 495–506.

Niinemets, Ü., 2010. Responses of forest trees to single and multiple environmental stresses from seedlings to mature plants: past stress history, stress interactions, tolerance and acclimation. For. Ecol. Manag. 260, 1623–1639.
Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20, 275.
Plomion, C., Aury, J.-M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillonne, N., Labadie, K., Le Provost, G., 2018. Oak genome reveals facets of long lifespan. Nat. Plants 4, 440–452.
Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., 2018. A universal SNP and small-indel variant caller using deep neural networks. Nat. Biotechnol. 36, 983–987.
Prjibelski, A.D., Mikheenko, A., Joglekar, A., Smetanin, A., Jarroux, J., Lapidus, A.L., Tilgner, H.U., 2023. Accurate isoform discovery with IsoQuant using long reads. Nat. Biotechnol. 41, 915–918. https://doi.org/10.1038/s41587-022-01565-y
Reis, A.L., Rapadas, M., Hammond, J.M., Gamaarachchi, H., Stevanovski, I., Ayuputeri Kumaheri, M., Chintalaphani, S.R., Dissanayake, D.S., Siggs, O.M., Hewitt, A.W., 2023. The landscape of genomic structural variation in Indigenous Australians. Nature 624, 602–610.
Rogger, J., Mills, B.J., Gerya, T.V., Pellissier, L., 2024. Speed of thermal adaptation of terrestrial vegetation alters Earth's long-term climate. Sci. Adv. 10, eadj4408.
Ross, C., Brack, C., 2015. Eucalyptus viminalis dieback in the Monaro region, NSW. Aust. For. 78, 243–253. https://doi.org/10.1080/00049158.2015.1076754
Schubert, M., Lindgreen, S., Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes 9, 88.
Shumate, A., Salzberg, S.L., 2021. Liftoff: accurate mapping of gene annotations. Bioinformatics 37, 1639–1643.
Silcocks, M., Farlow, A., Hermes, A., Tsambos, G., Patel, H.R., Huebner, S., Baynam, G., Jenkins, M.R., Vukcevic, D., Easteal, S., 2023. Indigenous Australian genomes show deep structure and rich novel variation. Nature 624, 593–601.
Sirén, J., Monlong, J., Chang, X., Novak, A.M., Eizenga, J.M., Markello, C., Sibbesen, J.A., Hickey, G., Chang, P.-C., Carroll, A., 2021. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. Science 374, abg8871.
Smolka, M., Paulin, L.F., Grochowski, C.M., Horner, D.W., Mahmoud, M., Behera, S., Kalef-Ezra, E., Gandhi, M., Hong, K., Pehlivan, D., Scholz, S.W., Carvalho, C.M.B., Proukakis, C., Sedlazeck, F.J., 2024. Detection of mosaic and population-level structural variants with Sniffles2. Nat. Biotechnol. 1–10. https://doi.org/10.1038/s41587-023-02024-y
Stacklies, W., Redestig, H., Scholz, M., Walther, D., Selbig, J., 2007. pcaMethods—a bioconductor package providing PCA methods for incomplete data. Bioinformatics 23, 1164–1167.
Supple, M.A., Bragg, J.G., Broadhurst, L.M., Nicotra, A.B., Byrne, M., Andrew, R.L., Widdup, A., Aitken, N.C., Borevitz, J.O., 2018. Landscape genomic prediction for restoration of a Eucalyptus foundation species under climate change. eLife 7, e31835. https://doi.org/10.7554/eLife.31835
Takahashi, D., Kawamura, Y., Uemura, M., 2016. Cold acclimation is accompanied by complex responses of glycosylphosphatidylinositol (GPI)-anchored proteins in Arabidopsis. J. Exp. Bot. 67, 5203–5215.
Teasdale, L.C., Murray, K.D., Collenberg, M., Contreras-Garrido, A., Schlegel, T., van Ess, L., Jüttner, J., Lanz, C., Deusch, O., Fitz, J., Mencia, R., van Velthoven, R., Drost, H.-G., Weigel, D., Shirsekar, G., 2025. Pangenomic context reveals the extent of intraspecific plant NLR evolution. Cell Host Microbe 33, 1291-1305.e9. https://doi.org/10.1016/j.chom.2025.07.011

Tegenfeldt, F., Kuznetsov, D., Manni, M., Berkeley, M., Zdobnov, E.M., Kriventseva, E.V., 2025. OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. Nucleic Acids Res. 53, D516–D522.

Thornhill, A.H., Crisp, M.D., Külheim, C., Lam, K.E., Nelson, L.A., Yeates, D.K., Miller, J.T., Thornhill, A.H., Crisp, M.D., Külheim, C., Lam, K.E., Nelson, L.A., Yeates, D.K., Miller, J.T., 2019. A dated molecular perspective of eucalypt taxonomy, evolution and diversification. Aust. Syst. Bot. 32, 29–48. https://doi.org/10.1071/SB18015

Todesco, M., Owens, G.L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D.O., Huang, K., Ostevik, K.L., Drummond, E.B.M., Imerovski, I., Lande, K., Pascual-Robles, M.A., Nanavati, M., Jahani, M., Cheung, W., Staton, S.E., Muños, S., Nielsen, R., Donovan, L.A., Burke, J.M., Yeaman, S., Rieseberg, L.H., 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature 584, 602–607. https://doi.org/10.1038/s41586-020-2467-6

Urquhart, A., Vogan, A.A., Gluck-Thaler, E., 2024. Starships: a new frontier for fungal biology. Trends Genet. 40, 1060–1073.

Wang, S., Ciais, P., Reich, P.B., Cescatti, A., Ellsworth, D.S., Janssens, I.A., Sardans, J., Luo, Y., Smith, N.G., Du, E., 2025. Phosphorus constrains global photosynthesis more than nitrogen does. Nat. Ecol. Evol. 1–11.

Wlodzimierz, P., Rabanal, F.A., Burns, R., Naish, M., Primetis, E., Scott, A., Mandáková, T., Gorringe, N., Tock, A.J., Holland, D., 2023. Cycles of satellite and transposon evolution in Arabidopsis centromeres. Nature 618, 557–565.

Yoo, D., Rhie, A., Hebbar, P., Antonacci, F., Logsdon, G.A., Solar, S.J., Antipov, D., Pickett, B.D., Safonova, Y., Montinaro, F., Luo, Y., Malukiewicz, J., Storer, J.M., Lin, J., Sequeira, A.N., Mangan, R.J., Hickey, G., Monfort Anez, G., Balachandran, P., Bankevich, A., Beck, C.R., Biddanda, A., Borchers, M., Bouffard, G.G., Brannan, E., Brooks, S.Y., Carbone, L., Carrel, L., Chan, A.P., Crawford, J., Diekhans, M., Engelbrecht, E., Feschotte, C., Formenti, G., Garcia, G.H., de Gennaro, L., Gilbert, D., Green, R.E., Guarracino, A., Gupta, I., Haddad, D., Han, J., Harris, R.S., Hartley, G.A., Harvey, W.T., Hiller, M., Hoekzema, K., Houck, M.L., Jeong, H., Kamali, K., Kellis, M., Kille, B., Lee, C., Lee, Y., Lees, W., Lewis, A.P., Li, Q., Loftus, M., Loh, Y.H.E., Loucks, H., Ma, J., Mao, Y., Martinez, J.F.I., Masterson, P., McCoy, R.C., McGrath, B., McKinney, S., Meyer, B.S., Miga, K.H., Mohanty, S.K., Munson, K.M., Pal, K., Pennell, M., Pevzner, P.A., Porubsky, D., Potapova, T., Ringeling, F.R., Rocha, J.L., Ryder, O.A., Sacco, S., Saha, S., Sasaki, T., Schatz, M.C., Schork, N.J., Shanks, C., Smeds, L., Son, D.R., Steiner, C., Sweeten, A.P., Tassia, M.G., Thibaud-Nissen, F., Torres-González, E., Trivedi, M., Wei, W., Wertz, J., Yang, M., Zhang, P., Zhang, S., Zhang, Y., Zhang, Z., Zhao, S.A., Zhu, Y., Jarvis, E.D., Gerton, J.L., Rivas-González, I., Paten, B., Szpiech, Z.A., Huber, C.D., Lenz, T.L., Konkel, M.K., Yi, S.V., Canzar, S., Watson, C.T., Sudmant, P.H., Molloy, E., Garrison, E., Lowe, C.B., Ventura, M., O'Neill, R.J., Koren, S., Makova, K.D., Phillippy, A.M., Eichler, E.E., 2025. Complete sequencing of ape genomes. Nature 1–18. https://doi.org/10.1038/s41586-025-08816-3

Yun, T., Li, H., Chang, P.-C., Lin, M.F., Carroll, A., McLean, C.Y., 2020. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. Bioinformatics 36, 5582–5589.

Zhang, C., Nielsen, R., Mirarab, S., 2025. ASTER: a package for large-scale phylogenomic reconstructions. Mol. Biol. Evol. 42, msaf172.

Zhang, Z., Zhang, P., Ding, Y., Wang, Z., Ma, Z., Gagnon, E., Jia, Y., Cheng, L., Bao, Z., Liu, Z., 2025. Ancient hybridization underlies tuberization and radiation of the potato lineage. Cell.

Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821–824.

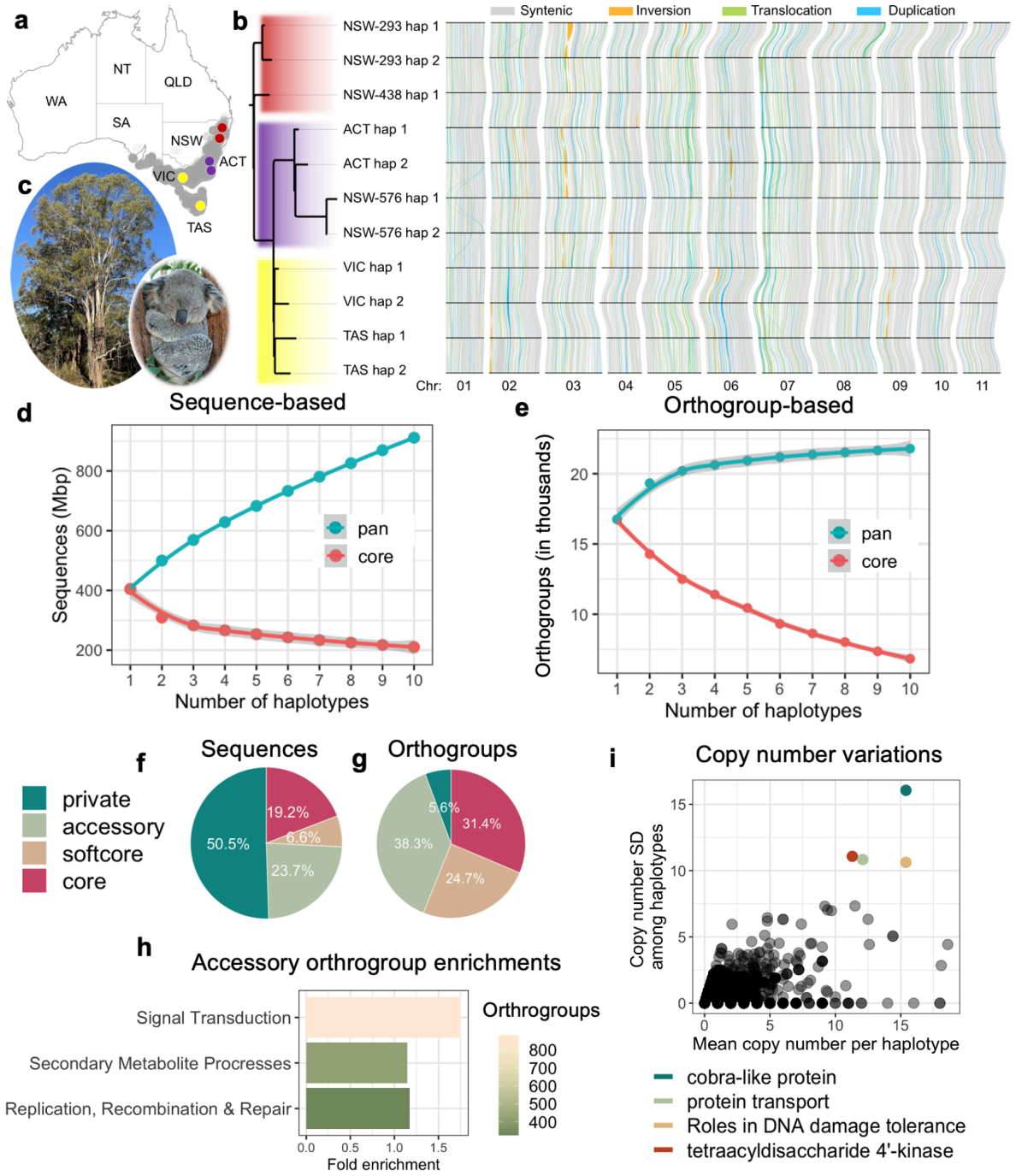

Figure 1. *Eucalyptus viminalis* is undergoing rapid sequence expansion while preserving core genome architecture. (a) Map of Australia showing the natural distribution of *E. viminalis* (grey shading) and locations of trees used for genome assembly and pangenome analysis (red, purple and yellow). ACT: Australian Capital Territory, NSW: New South Wales, NT: Northern Territory, QLD: Queensland, SA: South Australia, TAS: Tasmania, VIC: Victoria, and WA: Western Australia. (b) Phylogenetic relationships, synteny, and structural variation between genomes. Intrachromosomal structural variants ≥20 kb are displayed. (c) *Eucalyptus viminalis* ssp. *viminalis* in its natural environment at Bendoura, NSW. Inset: dependent arboreal marsupial, the koala (*Phascolarctos cinereus*). Photographs taken by Dean Nicolle and Gary Runn, respectively. (d) Linearly increasing novel sequences in sequence-based pangenomes. (e) Orthogroup expansion per genome input. (f-g) Private,

accessory, and core sequences and orthogroups. (h) Gene pathway enrichment in accessory orthogroups. (i) Orthogroups with high copy number variation.

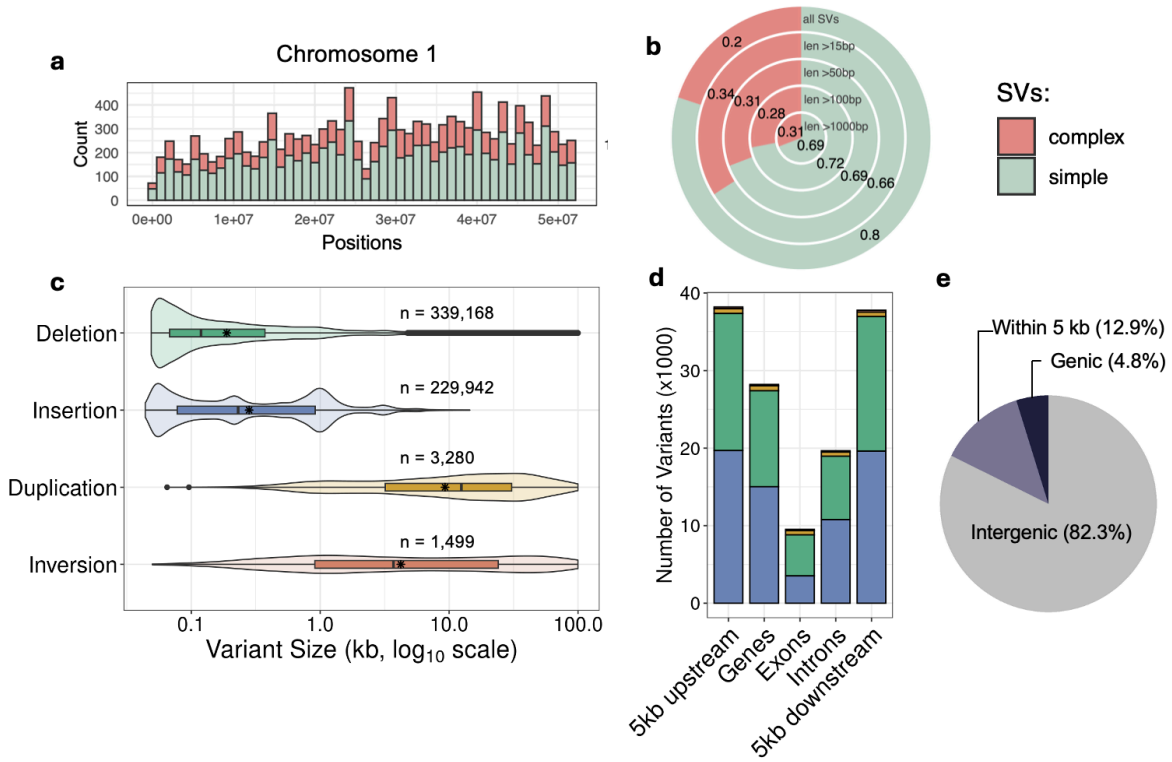

Figure 2. *Eucalyptus viminalis* harbours diverse structural variation with potential functional implications. (a) Simple (sSV) and complex (cSV) structural variants along chromosome 1 as an example (see Supplementary Note 6 for full genome). (b) Simple-to-complex SV ratio by size range. (c) Size distribution of non-redundant SVs discovered by Sniffles2. (d) SVs intersecting different genic regions. (e) Proportion of SVs intersecting intergenic and genic regions of the genome.

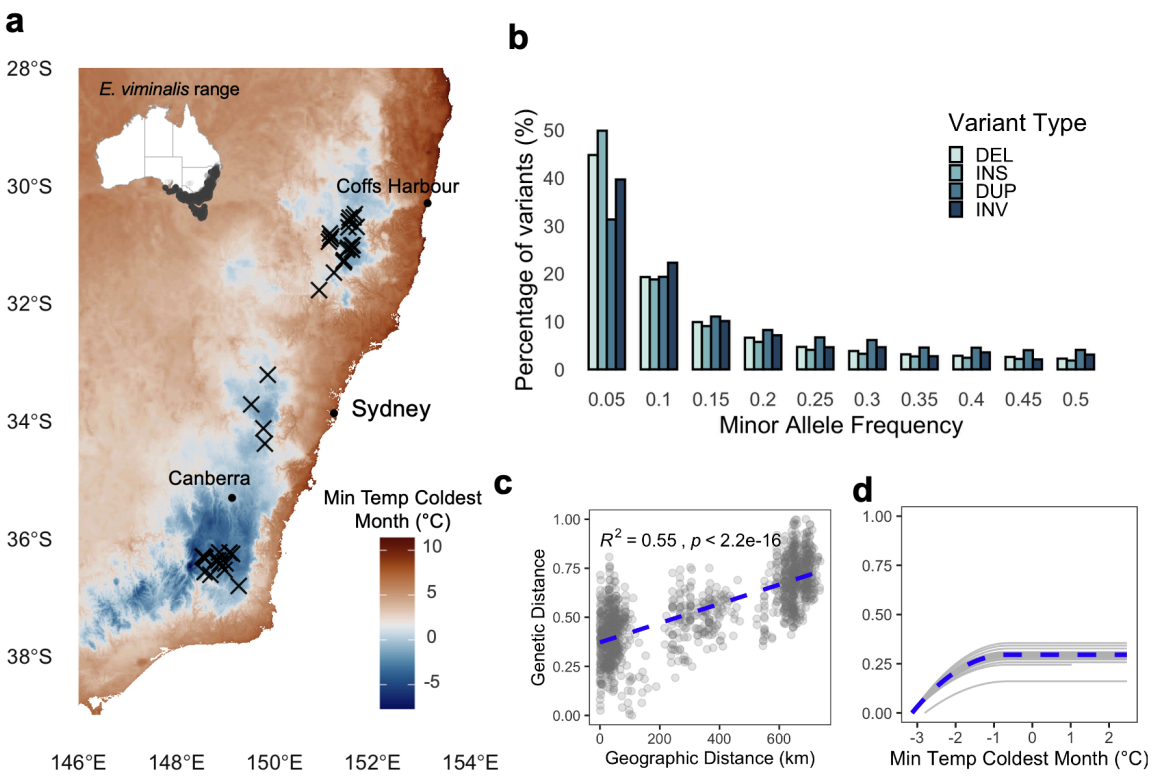

Figure 3. Intraspecific SV alleles segregate across geographic and environmental ranges of *Eucaltypus viminalis* (a) Spatial distribution of *E. viminalis* samples with long-read data (n = 49, 2n = 98). (b) Minor allele frequency spectrum of four structural variant (SV) types (deletions, insertions, duplications and inversions). (c) Covariation between pairwise geographic distance and pairwise genetic distance calculated with SV genotypes. (d) Genetic distance covariation with minimum temperature of the coldest month (BioClim6) over and above that which can be explained by geography alone using generalised dissimilarity modelling (GDM). The dotted line shows the spline curve fit using all samples. Grey lines show cross-validations by iteratively leaving one sample out.

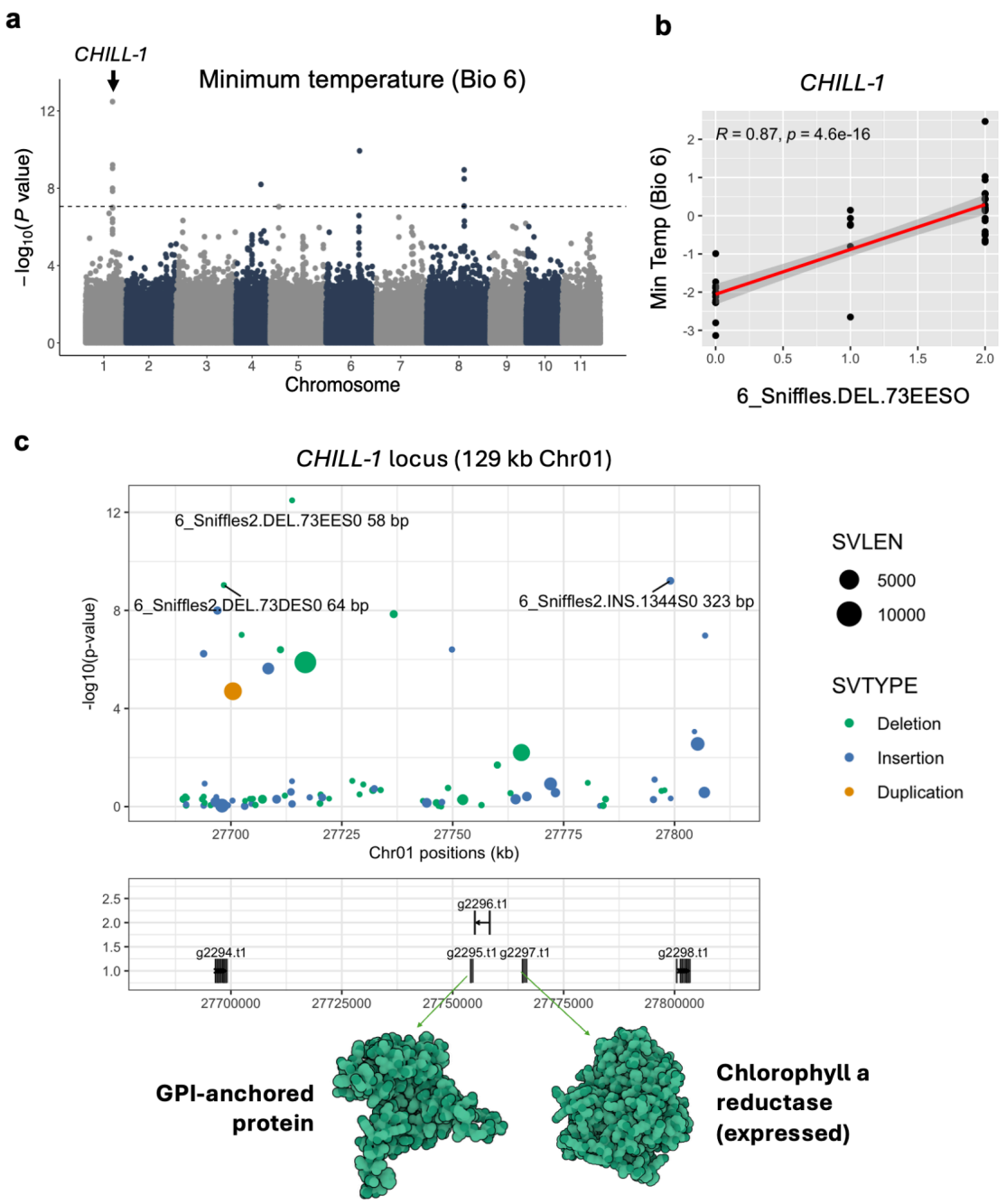

Figure 4. Complex structural variation underlie *CHILL1* — a cold-adaptation locus. (a) Manhattan plot of SV-based genome-wide association study (GWAS) with minimum temperature of the coldest month (BioClim6). (b) Genotype-environment association at the *CHILL1* locus. (c) Zoom-in of the *CHILL1* locus, including putatively associated structural variants and implicated genes and protein structure.

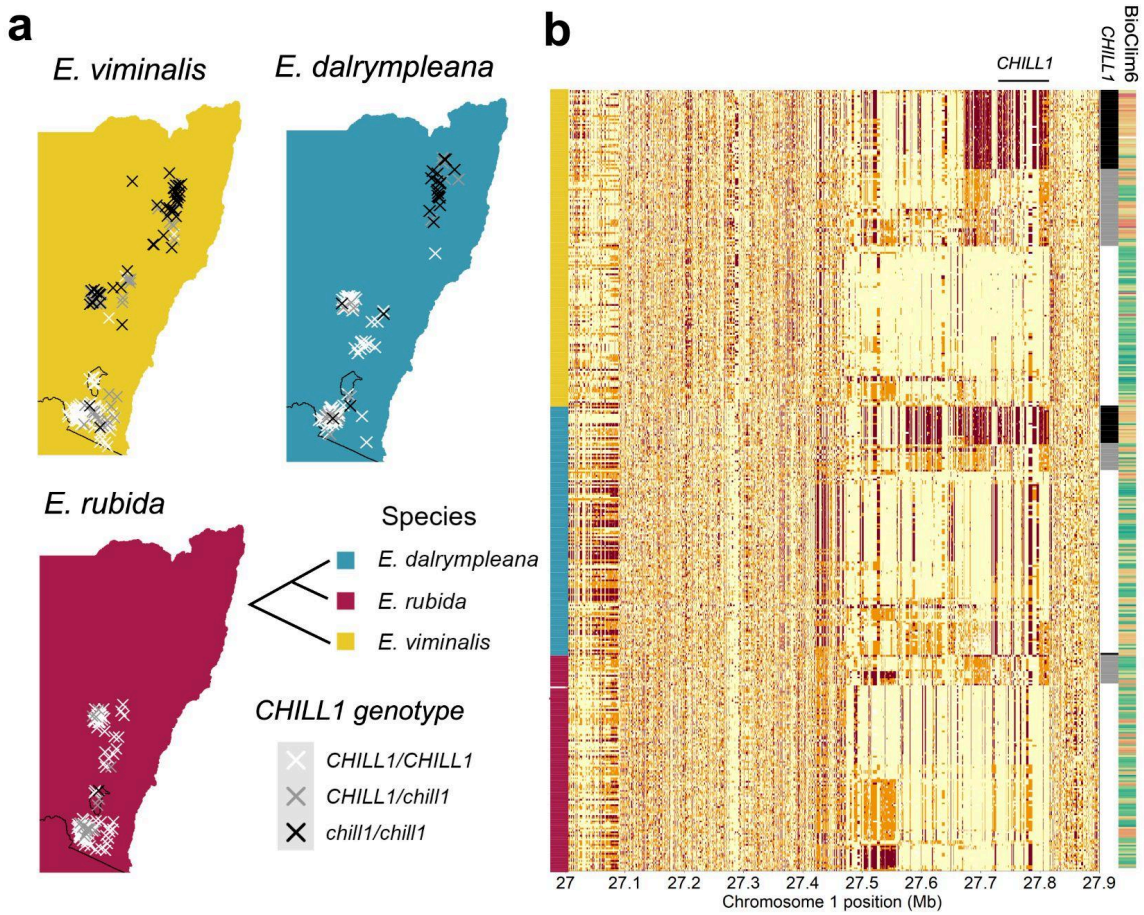

Figure 5. Short-read data shows that *CHILL1* is a conserved, across-species supergene. (a) Spatial distribution of *CHILL1* genotype across three *Eucalyptus* subg. *Symphyomyrtus* sect. *Maidenaria* Eucalyptus species (*E. dalrympleana*, *E. rubida*, and *E. viminalis*). (b) Short-read SNP genotypes in the *CHILL1* region for 434 trees from three species confirm the association of *CHILL1* allele frequency with minimum temperature and reveal large species-specific extended haplotypes that suggest its origin pre-dates species divergence.

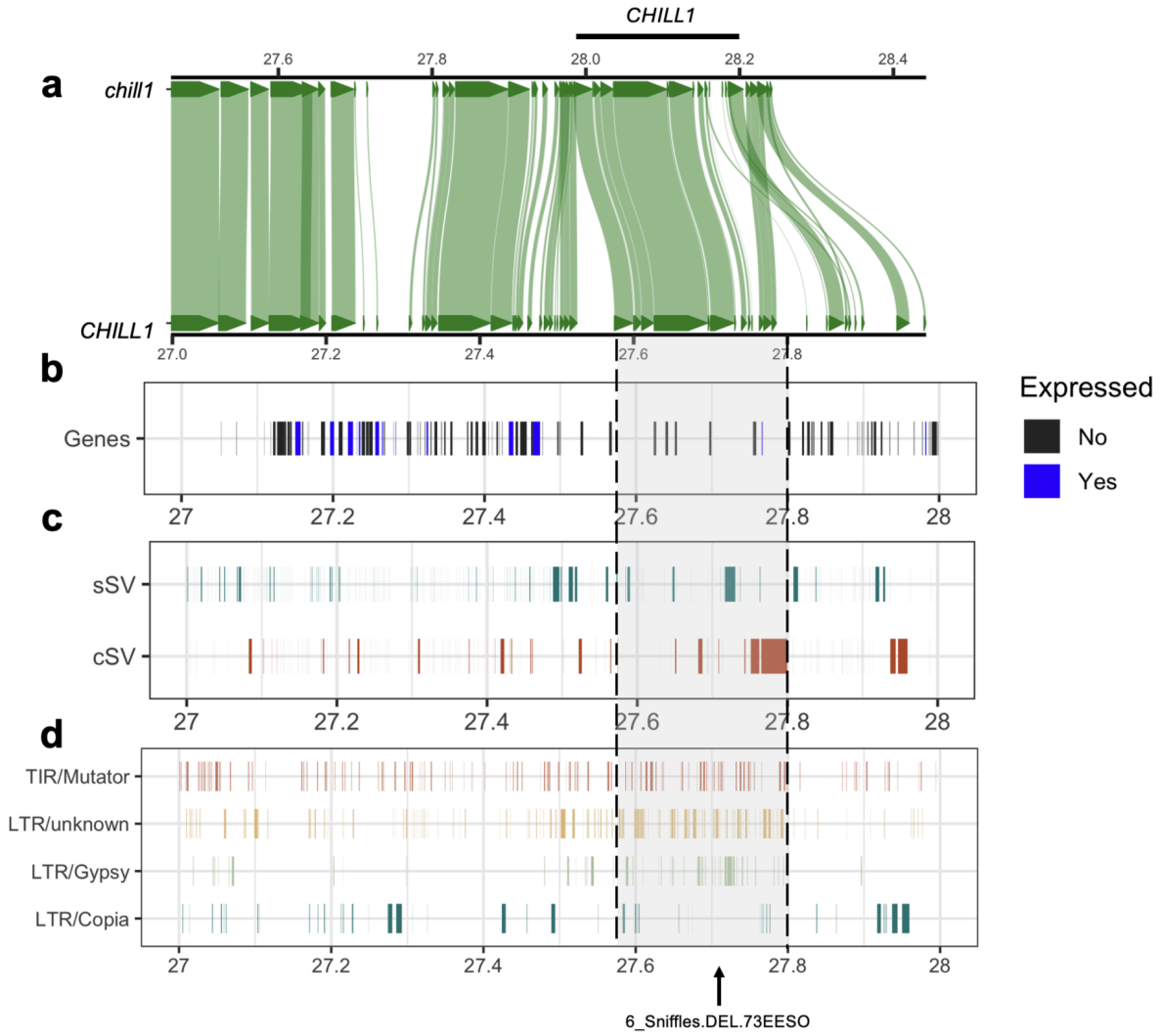

Figure 6. Characterisation of *CHILL1* with haplotype assemblies, genes, structural variants, and TEs. (a) Homology plot of two *Eucalyptus viminalis* haplotypes in the wider *CHILL1* region (chromosome 1: 27 Mb - 28 Mb) (b) Gene models annotated for the reference genome with expression validated by direct RNA sequencing, ≥1 transcript per million. (c) Simple and complex structural variants (sSVs, and cSVs) identified by Pannagram using all 10 haplotype-resolved assemblies of *Eucalyptus viminalis*. (d) Transposable element annotations across the *CHILL1* region.