

The pangenome of the flooded gum, *Eucalyptus grandis*

Anneri Lötter

`anneri.lotter@up.ac.za`

University of Pretoria <https://orcid.org/0000-0003-4502-7234>

Tomas Bruna

U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory

Sumaira Zama

Lawrence Berkeley National Laboratory

Rian Pierneef

Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

Kerrie Barry

Department of Energy - Joint Genome Institute <https://orcid.org/0000-0002-8999-6785>

Anna Lipzen

Lawrence Berkeley National Laboratory <https://orcid.org/0000-0003-2293-9329>

Chris Daum

U.S. Department of Energy <https://orcid.org/0000-0003-3895-5892>

Yuko Yoshinaga

Lawrence Berkeley National Lab <https://orcid.org/0000-0002-4978-9394>

Jane Grimwood

HudsonAlpha Institute for Biotechnology <https://orcid.org/0000-0002-8356-8325>

LoriBeth Boston

HudsonAlpha Institute for Biotechnology, 601 Genome Way Northwest, Huntsville, AL 35806, USA

Melissa Williams

HudsonAlpha Institute for Biotechnology

Jerry Jenkins

HudsonAlpha Institute for Biotechnology <https://orcid.org/0000-0002-7943-3997>

Chris Plott

HudsonAlpha Institute for Biotechnology <https://orcid.org/0000-0002-0109-5174>

Jayson Talag

Arizona Genomics Institute and BIO5 Institute

Lillian Padgitt-Cobb

HudsonAlpha Institute for Biotechnology, 601 Genome Way Northwest, Huntsville, AL 35806, USA

Justin Borevitz

Australian National University

Sanushka Naidoo

Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

John Lovell

HudsonAlpha Institute for Biotechnology <https://orcid.org/0000-0002-8938-1166>

Jeremy Schmutz

HudsonAlpha Institute for Biotechnology <https://orcid.org/0000-0001-8062-9172>

Jill Wegrzyn

University of Connecticut

Alexander Myburg

Department of Genetics, University of Stellenbosch, Private Bag X1, Stellenbosch 7600, South Africa.
Email: zandermyburg@sun.ac.za

Article

Keywords: Eucalyptus grandis, Phased assembly, Pangenome, Structural variants, Lignin pathway

Posted Date: July 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10148330/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

The pangenome of the flooded gum,

Eucalyptus grandis

Anneri Lötter¹, Tomas Bruna², Sumaira Zaman², Rian E. Pierneef³, Kerrie Barry², Anna Lipzen², Chris Daum², Yuko Yoshinaga², Jane Grimwood⁵, LoriBeth Boston⁴, Melissa Williams⁴, Jerry W. Jenkins⁵, Chris Plott⁴, Jayson Talag⁵, Lillian Padgitt-Cobb⁴, Justin Borevitz⁶, Sanushka Naidoo¹, John T. Lovell^{2,4}, Jeremy Schmutz^{2,4}, Jill L. Wegrzyn⁷ and Alexander A. Myburg^{1,8}

¹Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

²Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Mail Stop: 91R183, Berkeley, CA 94720, USA

³Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

⁴HudsonAlpha Institute for Biotechnology, 601 Genome Way Northwest, Huntsville, AL 35806, USA

⁵Arizona Genomics Institute, University of Arizona, 1657 E. Helen St., Tucson, AZ 85721, USA

⁶Research School of Biology and Centre for Biodiversity Analysis, ARC Centre of Excellence in Plant Energy Biology, Australian National University, Canberra, ACT 0200, Australia

⁷Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA

⁸Department of Genetics, University of Stellenbosch, Private Bag X1, Stellenbosch 7600, South Africa. Email: zandermyburg@sun.ac.za

Author contributions:

J.T. performed high-molecular weight DNA extractions. J.G. and K.B. performed RNA sequencing. J.J. and C.P. performed genome assembly and T.B. and S.Z. performed genome annotation. L.B. and M.W. were involved with data generation. A.A.M. and A.Lö. sampled tissues for DNA and RNA sequencing. L.P.C. was involved in panKmer analyses. RNA extractions and data analyses (all analyses using the assembled genomes and transcriptomes as input such as QC, BUSCO, QUAST, gFACs, SV analyses with SyRI, genome contiguity and gene content analyses with GENESPACE, functional annotation with EnTAP, GO enrichment analyses, lignin biosynthesis gene content analyses, population diversity analyses and all visualisations) were performed by A.Lö., with contributions from J.L. and supervision from R.P., T.D., J.T.L., J.L.W., J.S., S.N. and A.A.M. A.Lö. drafted the manuscript. A.A.M., J.B. and J.L.W. conceived the study. All authors read, edited and approved the manuscript.

Keywords: *Eucalyptus grandis*, Phased assembly, Pangenome, Structural variants, Lignin pathway

Abstract

Eucalyptus grandis serves as a model for understanding woody plant biology. We produced high-quality, *de novo* phased assemblies for 24 diverse *E. grandis* genotypes resulting in 48 haplotype-resolved assemblies (average 553.4 Mbp) and annotations (average 39,581 genes, BUSCO completeness >98.50%). We report a pangenome that catalogues structural variants (SVs) and pangene families in these individuals. The pangene catalogue has 123,938 pangene families (19,961 core, 1,730 softcore, 65,165 shell and 37,082 private gene families). Comparison to the phased reference (V4.2, TAG0014_HAP1) revealed 37,561 non-redundant SVs and 128,487 non-redundant non-aligned regions across the haplotype-resolved assemblies. Focusing on lignin biosynthesis, we show that while the pathway is highly conserved at the pangenome level, key enzymatic modules such as 4CL, CCR, CAD, and COMT exhibit substantial CNV and orthogroup diversification, revealing hidden functional plasticity within an otherwise core metabolic pathway. This resource provides a critical foundation for exploring functional biology and deciphering genome evolution.

Introduction

There are more than 900 species of “eucalypts”, a group of woody plants with members in the sister genera, *Eucalyptus*, *Corymbia* and *Angophora*¹, some of which contribute to over 20 million ha of the world’s hardwood plantations². As fast-growing woody perennials, eucalypt plantation species have been identified as candidates for global carbon drawdown³ and a renewable feedstock of woody biomass for bio-based materials. The use of *E. grandis* as genomic reference for the genus has led to the development of numerous genomic resources including a haplotype-resolved reference genome with much improved gene models⁴. As a species with high genetic diversity⁵⁻⁷ and genetic structuring spanning the east coast of Australia, *E. grandis* is an excellent candidate for pangenome diversity studies and understanding how this diversity contributes to adaptive processes in natural populations⁸. A few comparative genomic studies have been completed for eucalypt species⁹⁻¹³, showing that genome architecture is broadly conserved. However, the extent of haplotype structure and pangene variation have not been comprehensively studied in many eucalypt species.

We are now in an era where long-read sequencing enables a view of the entire genome and its structural diversity, breaking away from the conventional gene, gene variant and single nucleotide polymorphism (SNP)¹⁴ centred view. Pangenomics is a contemporary resource in understanding plant genome diversity, while providing new resources to develop molecular breeding tools aimed at understanding crop evolution, domestication and genetic improvement¹⁵⁻²⁷. This major step change stems from advances in long-read sequencing technologies that enable analysis of structural variants (SVs) and transposable elements (TEs) as major components of pangenome variation²⁸. SVs and TEs are known to influence gene expression, directly (via copy number alterations, duplications, and insertion-deletion events) or indirectly (via *cis*-regulatory interactions or epigenetic changes)^{15,21,23,24,29-33}. This work is exemplified in tomatoes where SVs and TEs influence production traits³⁴⁻³⁹. Although pangenomes are available for multiple crop species^{24,40,41}, comparable resources remain scarce for forest trees⁴². For *Eucalyptus*, only limited or preliminary genome diversity efforts have been reported^{10,11,13}. Current studies indicate that a genome diversity view will greatly improve understanding of environmental adaptation and genome diversity available for breeding and conservation efforts⁴²⁻⁴⁶.

Here, we analysed the pangenome of *E. grandis* reconstructed from the chromosome-level haplotypes of 25 diverse individuals, including TAG0014, representing the natural species range and early breeding material. We generate 48 haplotype assemblies from PacBio high-fidelity (HiFi), and proximity ligation (Omni-C) sequencing data. We annotate gene models with support from Iso-Seq and Illumina RNA-Seq data, and TEs across all 48 genomes. We report a wide range of sequence variants using TAG0014_HAP1 as the reference genome for assembly based comparative analyses⁴. We also generate a pangene catalogue which suggests that local environmental adaptations might involve shell genes rather than high-effect private genes. This is the first pangenome catalogue for a eucalypt species, capturing a large proportion of gene and sequence level variation in *E. grandis*. Due to the extreme genome diversity observed, a traditional pangenome graph approach is financially and

computationally inviable. Rather, we use the lignin pathway to showcase the use of this resource. These can provide future targets for functional validation studies.

Materials and Methods

Plant materials and diversity representation

The pangenome individuals selected for this study were chosen to represent most of the natural species' range sampled in a gene conservation park in the KwaZulu-Natal Midlands of South Africa originating from a landscape genomics study⁷. We selected individuals to represent as many provenances as possible concluding with 18 of the 33 provenances being represented⁷. The breeding material (RM01, RM02 and RM03) was selected from some of the most advanced breeding material available for the species, fifth-generation selections of *E. grandis* (Hans Merensky Timber, South Africa) grown in the Tzaneen area of the Mpumalanga province, South Africa (**Supplementary Table 1**).

The genetic representation of the 24 pangenome individuals relative to a range-wide collection of *E. grandis* was assessed using a 72K SNP Axiom 384-format array for *Eucalyptus* and *Corymbia* (Affymetrix Inc., Santa Clara, California, USA). Allelic intensity data were assessed using Axiom Analysis Suite v5.2.0.65 to recluster genotypic classes, where necessary, as described by⁴⁷. SNP data for natural population references was obtained from⁷ and the SNP calls were converted to Axiom SNPs using a custom R script in RStudio v2023.12.1. The converted calls were imported into SNP & Variation Suite™ v8.9.1 (SVS8; Golden Helix, Inc., Bozeman, MT). High confidence, informative SNPs were retained by filtering for minor allele frequency >0.02, genotyped in at least 80% of samples and Hardy Weinberg Equilibrium P-value >1e-5. Starting with 18,328 SNP markers that mapped to the v2.0 reference genome⁵, 17,036 SNPs remained after filtering. To determine the genetic relatedness of the 24 pangenome individuals with trees representing the natural range of *E. grandis*, we analysed the SNP data with that from⁷ using a Principal Component Analysis (PCA)⁴⁸ using SVS8. The 50 genomes were also evaluated for relatedness with PanKmer v0.20.4⁴⁹. First, the genomes were decomposed into 31-mers and indexed, followed by conversion to an adjacency matrix

containing counts of shared 31-mers. The shared 31-mer counts were transformed into Jaccard similarity scores and visualized by clustering.

Genome sequencing and haplotype-resolved assembly

All individuals were sequenced using PacBio HiFi (Pacific Biosciences Inc., CA, USA), chromosome conformation capture with Omni-C (Dovetail Genomics, CA, USA) and Illumina short-read sequencing technologies. PacBio Iso-Seq and Illumina RNA-Seq data were generated separately from each of the 24 individuals for genome annotation of the corresponding trees. High molecular weight (HMW) DNA was isolated from frozen leaf tissue at the Arizona Genomics Institute (AGI, Tucson, Arizona, USA). PacBio and Illumina sequence reads were generated at the HudsonAlpha Institute for Biotechnology (HA) in Huntsville, Alabama and at the Department of Energy Joint Genome Institute (DOE-JGI) in Berkeley, California. An Illumina fragment library (2×150 nt, 400 bp insert) was sequenced with 2×150 nt Dovetail Hi-C library (**Supplementary Table 2**). Illumina libraries were sequenced using the NovoSeq 6000 platform. PacBio sequence reads were generated using the Sequel IIe platform at HA.

Error-corrected PacBio reads were assembled into initial haplotype-resolved contigs using HiFiAsm+HiC⁵⁰. The assemblies were polished with RACON⁵¹ and misjoins were identified in the polished assemblies. This produced highly contiguous initial assemblies for all haplotypes. The assemblies were further scaffolded, ordered and oriented using Omni-C proximity ligation contacts generated with the JUICER pipeline⁵² and the *E. grandis* v2.0 genome. Contigs terminating in significant telomeric sequences were properly oriented in the assembly. Adjacent alternative haplotypes were identified on the joined contig set. The resulting assemblies were screened for contaminants using NCBI's FCS-GX pipeline⁵³. The mitochondrial and chloroplast genomes were assembled using OatK⁵⁴. Additional scaffolds were classified as redundant, repetitive or organellar (assembled with OatK before genome assembly).

HiFiAsm may occasionally misphase small regions within the chromosomes⁵⁵. Misphased regions were identified through alignment of Omni-C data to the combined HAP1/HAP2 chromosomes and manual identification of the misphased regions in the contact map. The corrected haplotype-specific resolved assemblies were selected as haplotype-resolved references (HAP1 and HAP2) and screened for PacBio linkers.

Homozygous SNPs and INDELs were corrected in both haplotypes using the corresponding Illumina reads (**Supplementary Table 2, Supplementary Data 1**). The reads were aligned to each haplotype using *bwa mem*⁵⁶ and identifying homozygous (haploid) SNPs and INDELs with the GATK's UnifiedGenotyper tool⁵⁷. Homozygous SNPs/indels were then corrected using an in-house developed pipeline, and the Illumina reads were re-aligned to the corrected sequence to verify the corrections. Heterozygous SNPs and INDELs were corrected using HiFi data generated for each individual haplotype-resolved assembly (**Supplementary Data 1**).

Genome assembly statistics were generated with QUAST v5.2.0^{58,59} and completeness was assessed with BUSCO v5.4.5⁶⁰⁻⁶² using the *embryophyta_odb10* database. Telomere sequences were identified for all genomes using the (CCCGAAA)*n* and (CCCTAAA)*n* repeats and genome contiguity was detected and visualised with GENESPACE⁶³.

Completeness of the euchromatic portion of the genome assembly was assessed by aligning 34,121 existing *E. grandis* v2.0 annotated primary transcripts to each haplotype-resolved genome assembly. The completeness analysis aims to obtain a measure of assembly completeness, rather than a comprehensive examination of the gene space. We retained genes that aligned at greater than 90% identity and 85% coverage (**Supplementary Data 1**).

Genome annotation

Total RNA was extracted from flower, leaf, phloem and xylem tissues collected from the pangenome individuals (flower tissue could not be collected for two samples, RM02 and RM03) using the Plant/Fungi total RNA purification kit (Norgen Biotek Corp.) per the manufacturer's protocol. RNA

purity was verified using spectroscopy on a NanoDrop ND1000 (Nano-Drop Technologies) and integrity was verified with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA-Seq data was generated by the JGI (Berkeley, CA) for the four tissue libraries prepared using Illumina's TruSeq Stranded mRNA HT sample prep kit using poly-A selection of mRNA for each individual. RNA-Seq libraries were sequenced on the NovaSeq S4 platform to produce 2x150 bp RNA-Seq reads. Transcripts were assembled from 2x150 bp stranded PE reads with PERTRAN⁶⁴, which conducts genome-guided transcriptome short read assembly via GSNAP⁶⁵ and builds splice alignment graphs after alignment validation, realignment, and correction.

PacBio Iso-Seq full-length RNA transcript sequences were generated using the Sequel IIe platform for two tissue pools (green/leaf and a pool of woody/cambium and flower tissues) generated from the RNA extracted above. A genome-guided correction pipeline was used to correct and collapse PacBio Iso-Seq CCS reads. The pipeline aligns CCS reads to the genome with GMAP⁶⁵, corrects small indels in splice junctions, and clusters alignments when all introns are the same or $\geq 95\%$ overlap for single-exon alignments. Transcript assembly sets were combined into a final set of HAP1 and HAP2 transcripts with PASA⁶⁶.

A *de novo* repeat library was created for the TAG0014_HAP1 v3.0 genome draft using RepeatModeler2⁶⁷. To identify repeats with significant hits to protein-coding domains, the repeat library was functionally analysed with InterProScan⁶⁸ based on the Pfam⁶⁹ and PANTHER⁷⁰ databases. Repeats with significant protein-coding domain hits were removed from the repeat library and each individual genome was soft-masked using RepeatMasker⁷¹ with the resulting species-specific repeat library.

To determine putative gene loci, transcript assembly alignments and/or EXONERATE⁷² alignments of proteins from *Corymbia citriodora*, *Arabidopsis thaliana*, *Glycine max*, *Fragaria vesca*, *Vitis vinifera*, *Liriodendron tulipifera*, *Punica granatum*, *Rhodamnia argentea*, *Syzygium oleosum*, *Brassica rapa*, *Citrus clementina*, *Gossypium raimondii*, *Populus trichocarpa*, *Oryza sativa*, *Beta*

vulgaris, and Swiss-Prot release 2022_04 of eukaryote proteomes were generated using repeat-soft-masked *E. grandis* var. TAG0014_HAP1 and HAP2 v4.0 genomes, with up to 2 kbp bidirectional extension unless extending into another locus on the same strand. Gene models in each locus were predicted by homology-based predictors, FGENESH+⁷³, FGENESH_EST (similar to FGENESH+, but using PASA-assembled transcripts to compute splice site and intron input instead of protein/translated ORF), EXONERATE, PASA assembly ORFs (in-house homology constrained ORF finder), and AUGUSTUS⁷⁴ trained on the high confidence PASA assembly ORFs and with intron hints from short-read alignments. The best-scored predictions for each locus were selected using multiple positive factors including transcriptome and protein support, and one negative factor: overlap with repeats. The selected gene predictions were improved by adding UTRs, splicing correction, and adding alternative transcripts with PASA.

To obtain a Cscore (protein BLASTP score ratio to the mutual best hit BLASTP score) and protein coverage (percentage of protein aligned to the best homolog). PASA-improved gene model proteins were subjected to protein homology analysis to the above-mentioned proteomes. PASA-improved transcripts were selected based on Cscore, protein coverage, transcriptome coverage, and their CDS overlap with repeats. The transcripts were selected if their Cscore and protein coverage were ≥ 0.5 or if covered by transcripts assembled by PASA. For gene models whose CDS overlapped with repeats predicted by RepeatMasker by more than 20%, the Cscore had to be at least 0.9 and homology coverage at least 70% to be selected. The selected gene models were subject to Pfam analysis and gene models without strong transcriptome and homology support and whose proteins overlapped more than 30% with Pfam TE domains were removed. Incomplete gene models, low homology supported without full transcriptome-supported gene models, short single exon (<300 bp CDS) without protein domains supported by less than 10 RNA-Seq reads, and genes with high copy numbers without strong homology support were manually filtered out.

A second gene call round was performed using first round annotations from all genomes in the pangenome through all-to-all mapping of gene models using Liftoff⁷⁵. Gene models in each genome

were added or revised when supported by consistent annotations across multiple genomes or by strong external evidence, such as high homology to proteins from other species or robust transcriptome mapping. When conflicting models were found between a genome's first-round annotations and mapped models from other genomes, the version with the strongest combined evidence - incorporating support from both the genome itself and other members of the pangenome - was selected. This process produced a refined and evidence-supported set of gene annotations for each genome, improving consistency and completeness across the entire pangenome. Detailed gene metrics were obtained for the haplotype-resolved genome assemblies with gFACs v1.1.2⁷⁶. Completeness of transcript and protein models was determined using BUSCO v5.4.5 and the *embryophyta_odb10* database.

Pangenome construction

All 49 *E. grandis* haplotype-resolved assemblies were aligned to the TAG0014_HAP1 reference using nucmer from the MUMmer4 toolbox⁷⁷. Resulting alignments were filtered with nucmer (“—maxmatch -c 100 -b 500 -l 50”). The alignments were further filtered for alignment length (>100) and identity (>90), then used to identify local variants (LVs) and structural rearrangements with SyRI v1.6.3⁷⁸. A summary of the results for all 49 comparisons was generated and visualised with custom scripts. A set of non-redundant SVs and LVs were identified with SURVIVOR with 1,000 bp maximum distance allowed between breakpoints, at least one caller genotype, not considering the type of SV or the strand and a minimum SV size of 100 bp (“1000 1 1 0 0 100”). To visualise SVs, the 50 haplotype assemblies were aligned and arranged alphabetically using the MiniMap2 v2.26⁷⁹ module within GENESPACE⁶³.

Homologous relationships among the 50 haplotype-resolved assemblies were predicted based on the longest transcript of each predicted gene model as a representative. We performed an all-vs-all comparison with Diamond v2.18^{80,81}. Gene clustering and identification of ortholog groups was performed using Orthofinder v2.5.5^{82,83} as implemented in GENESPACE. Genes were classified based on hierarchical orthologous groups (HOGs) from the ‘combed.txt’ file as core (shared among

all 50 haplotypes), softcore (shared by 49 haplotypes), shell (shared by two to 48 haplotypes) and private (only present in a single haplotype). To infer more specific gene ontology (GO) terms, all 50 genomes were functionally annotated with EnTAP⁸⁴ using the following databases: NCBI RefSeq plant, EMBL-EBI UniProt, to integrate similarity search and other annotation resources including gene family (eggNOG), protein domains (Pfam), gene ontology and KEGG pathway assignment. GO term enrichment analyses were performed for each class using Fisher's exact test, followed by the Benjamin–Hochberg correction to obtain adjusted *P* values for multiple testing with topGO⁸⁵. Given the importance of woody biomass accumulation, we examined which genes associated with lignin production were present within categories of the pangenome gene set. Genes were identified based on keywords in the lignin toolbox⁸⁶ annotations using GO, KEGG and Pfam IDs. The following genes were selected based on the core lignin toolbox⁸⁶: caffeoyl CoA 3-O-methyltransferase (CCoAOMT), caffeate/5-hydroxyferulate O-methyltransferase (COMT), caffeoyl shikimate esterase (CSE), cinnamate 4-hydroxylase (C4H), cinnamoyl CoA reductase (CCR), cinnamyl alcohol dehydrogenase (CAD), 4-coumarate:CoA ligase (4CL), ferulate 5-hydroxylase (F5H), p-coumarate 3-hydroxylase (C3'H), phenylalanine ammonia-lyase (PAL) and shikimate O-hydroxycinnamoyltransferase (HCT). This is not an exhaustive functional validation set, but an annotation-based gene module.

Since all alignments made with SyRI used TAG0014_HAP1 as reference, we used the TAG0014_HAP1 gene models and TE models to find overlap between SVs, TEs and genes. The number of overlaps between genes, SVs and TEs was calculated using the BEDTools v2.29.0⁸⁷ intersect function, with the number of overlaps per event capped at the total number of features in the TAG0014_HAP1 genome, since an event cannot exceed the number of available events. The results were summarized and visualised using R v2024.12.1.

Results

Population diversity

To maximize genetic diversity, we selected 21 individuals from the natural range and three genotypically contrasting fifth generation breeding individuals that span 18 of the 33 *E. grandis* provenances⁷ (**Figure 1, Supplementary Table 1**). A PCA (**Figure 1**) with 17,036 informative SNP markers confirmed that 7, 1 and 13 individuals from the North, Mackay, and South subpopulations, respectively, based on their clustering positions. We observed that TAG0014 and the breeding material (RM01, RM02 and RM03) clustered within the Southern subpopulation (**Figure 1**), as expected from the recorded breeding material origin⁸⁸. There was low relatedness between assemblies, confirming high genetic diversity.

Pangenome assembly and annotation

To characterize the pangenome of *E. grandis*, we assembled haplotype-resolved *de novo* genomes for the 24 individuals using a combination of Illumina (average coverage 50.2x), HiFi (average coverage of 53.1x) and Omni-C sequencing technologies (**Supplementary Table 2**). The haplotype-resolved assemblies (hereafter assemblies) range in size from 531.0 Mbp (RK17_HAP2) to 574.5 Mbp (RM02_HAP1) (scaffold N50 of 51.5 Mbp to 57.9 Mbp, **Figure 2**), consisting of 11 chromosomes (**Supplementary Table 3 and 4**) with 19 assemblies having all 22 telomeres assembled (**Table 1, Supplementary Figure 1 and 2**). Assembly completeness exceeded 98.1% with one exception (RK17_HAP2, 96.90%, **Figure 2, Table 1, Supplementary Table 5**). The chloroplast and mitochondrial genomes for each assembly ranged from 159.2 – 160.4 kbp and 442.5 – 490.8 kbp in size, respectively (**Table 1**).

Genome annotation integrating multiple evidence lines, including the 50 assemblies, predicted 38,555 – 40,292 protein-coding genes (average 39,581 genes) that had a BUSCO completion score of $\geq 98.50\%$ (**Table 1, Supplementary Table 6 and 8**). This indicated high quality annotations (and assemblies), laying a solid foundation for pangenome analysis. Repeat elements (REs) occupied between 43.1% - 45.1% of the genomes (**Figure 2, Table 1, Supplementary Figure 3, Supplementary Table 7**). Most of REs consisted of unknown elements (25.89% - 27.24%), followed by *Copia* (8.7% - 9.35%) and *Gypsy* (3.36% - 3.84%) LTR elements (**Supplementary Figure 3,**

Supplementary Table 7). All assemblies showed comparable genome size, N50s, number of annotated genes and REs (**Figure 2, Supplementary Figure 3, 7, and 8**) with no quality bias related to provenance or haplotype.

Pangenome analysis

To explore the SV landscape, we aligned 49 assemblies to TAG0014_HAP1. The majority of the assemblies were syntenic to TAG0014_HAP1 (358.8 – 396.3 Mbp equating to 62.9 – 69.4% of the TAG0014_HAP1 genome, **Figure 3 and 4, Supplementary Figure 4, Supplementary Table 9**). The average syntenic block size was 27.6 kbp. Total SV content per assembly ranged from 8.3 to 42.5 Mbp (inversions), 41.7 to 64.1 Mbp (translocations) and 43.8 to 56.8 Mbp (duplications) relative to the TAG0014_HAP1 assembly (**Figure 3 and 4, Supplementary Figure 4, Supplementary Table 9**). Although the total SV size was similar (**Figure 4**), SV events showed variation in their number and size (**Figure 4, Supplementary Table 9**). Inversions were the least frequent (76 – 116 events per assembly) and tended to be larger (average 135.3 kbp), followed by translocations (7,147 – 9,591 events per assembly, average 5.9 kbp) and duplications (13,513 – 18,115 events gained per assembly, average 3.2 kbp and 19,621 – 23,553 events lost per assembly, average 3.8 kbp, **Figure 4, Supplementary Table 9**). We also noted large inversions not shared by any other assembly in TAG0014_HAP1 (chromosome 2, approximately 8.01 Mbp), RK84_HAP2 (chromosome 7, 10.9 Mbp and chromosome 11, 12.8 Mbp), RK436_HAP1 (chromosome 5, 18.1 Mbp) and RK05_HAP1 (chromosome 6, 30.7 Mbp), which explains the higher-than-average inversion size for these individual haplotypes (**Figure 4**). Approximately 13.2 to 15.8 % (75.3 – 90.1 Mbp) of the assemblies could not be aligned to TAG0014_HAP1 (average block size 3.5 kbp, **Figure 3 and 4, Supplementary Table 9**)

Smaller local variants (LVs, <50 bp) are embedded within syntenic regions and SVs and contributed to the following events: 5.9 – 7.2 million SNPs, 4.3 – 5.4 Mbp insertions, 3.9 – 5.1 Mbp deletions, 3.2 – 6.99 Mbp copygains, 3.3 – 8.1 Mbp copylosses, 12.5 – 21.7 Mbp highly diverged regions and 0.7 – 1.7 Mbp tandem repeats (**Supplementary Figure 6 and 7, Supplementary Table 10**). These

results highlight that both SVs and LVs contribute to *E. grandis* genome variation, however SVs contribute to a larger absolute size proportion of genome changes. How SVs and LVs contribute to phenotypic variation is still to be determined.

We observed 4,761 inversions, 414,223 translocations, 760,907 duplication copygains, 1,059,479 duplication copylosses, 983,996 and 1,047,116 regions did not align in the query and TAG0014_HAP1 assemblies and 678,931 syntenic regions (4,949,413 total events, **Supplementary Table 9 and 11**). These were merged (not considering SV type) into a non-redundant set of 511 inversions, 32,659 duplications, 4,391 translocations, 90,926 non-aligned and 72,064 syntenic regions (total of 200,551 non-redundant regions, **Supplementary Figure 8, Supplementary Table 11**). This may be explained by the 32.6 – 67.3% of calls shared between alignments (**Supplementary Figure 8**), showing overlap between regions (breakpoint and event overlaps) where SVs are found despite them being different SV events, which may be due to genome complexity. Therefore, SVs are treated as a complementary structural layer rather than a fully saturating map of variation. LVs including copygains (62,982), copylosses (60,972), highly diverged regions (237,848) and tandem repeats (14,549, a total of 376,351 events) were merged into 2,924, 4,728, 29,437 and 1,499 non-redundant events, respectively (38,588 total events, **Supplementary Table 11**).

We performed pangenome analysis based on 50 annotations. A total set of 1,979,054 genes were used, of which 1,977,044 were assigned to 46,626 orthogroups by Orthofinder excluding private assignments (**Supplementary Table 12**). Using global phylogenetically hierarchical orthogroups ('globHOG', shortened to HOG from here) as defined by GENESPACE, the number of HOGs increased with the number of genomes added, leaving the pangenome open, with 123,938 HOGs (**Figure 5**). As such, the current dataset may not yet capture the full extent of gene diversity in the species, and some additional provenances could close the pangenome. HOGs were classified based on the number of genomes that have them as core (50 genomes), softcore (49 genomes), shell (2 – 48 genomes) and private (1 genome). We found 19,961 core, 1,730 softcore, 65,165 shell and 37,082 private HOGs representing 16.1%, 1.4%, 52.6% and 29.9% of the assigned HOGs (**Figure 5**). This

corresponded to 999,540 core, 85,277 softcore, 858,344 shell and 37,893 private genes which make up 50.1%, 4.3%, 43.4% and 1.9% of assigned genes (**Figure 5, Supplementary Table 13**). On average, 19,991 genes were assigned as core per genome along with 1,706 softcore, 17,127 shell and 758 private genes (**Figure 5, Supplementary Table 13**). An average of 1,761 genes were added to the pangenome by each additional genome, with 9,237 and 986 genes added by the first and last genomes, respectively (**Figure 5**). Private genes had the lowest assignment for a single HOG (~1 gene per HOG), followed by shell genes (13 genes per HOG). More genes were found per HOG for core/softcore (50 /49 genes per HOG). Core and softcore genes were enriched for processes such as regulation of flower development, transcription factor binding and activity related functions. Shell genes were enriched for sesquiterpene biosynthesis, salicylic acid response, and multiple synthase and receptors. Private genes were enriched for defence response and nucleocytoplasmic transport (**Supplementary Figure 9 - 11, Supplementary Table 14 - 16**).

We discovered 10,104 lignin-associated genes within the pangene catalogue, of which 1,744 genes occurred as copy number variants (CNVs). These were classified as 4,743 core, 502 softcore, 4,552 shell and 308 private genes, (**Figure 6**) that represent 8,537 HOGs (4,465 core, 443 softcore, 3,238 shell and 295 private HOGs, **Figure 6**). Consistent with the conserved nature of lignin biosynthesis, modules like C4H, CSE, and F5H (8, 1 and 2 HOGs) were restricted almost entirely to the core genome and were represented by relatively few gene families. In contrast, 4CL, CCR, CAD and COMT showed broad representation across pangenome compartments with much higher orthogroups diversity and CNV across genomes.

We found that inversion events overlapped (any shared genomic region between an SV and another feature regardless of intersect length) with 9,564 genes, duplications with 11,947 genes and translocations with 13,012 genes (**Supplementary Table 17**). Along a similar vein, 29,424 genes overlap RE space (**Supplementary Table 17**). Although this does not provide information on the effects these overlaps may have on gene regulation, expression and function, it does suggest a sizeable proportion of the genome and coding space are affected by rearrangements and REs. Overlap size is

not revealed, as any length of intersection was considered. However, these events could have large effects when interrupting or altering gene regulation and expression, although further validation and testing would be needed. Inversions, duplications and translocations also overlap 200,665, 295,844 and 323,685 of the 743,996 RE space, suggesting a role of REs in the evolution and loss and gain of rearrangement events (**Supplementary Table 17**).

Discussion

Here, we *de novo* assembled and annotated 48 high-quality, chromosome-level haplotype-resolved assemblies for 24 *E. grandis* individuals (**Figure 1, Supplementary Table 1**) representing broad species genetic diversity. We used the assemblies generated in this study alongside the TAG0014 reference⁴ and performed pangenome analyses of 50 genomes to gather gene diversity information (**Figure 5**) and generated an extensive SV catalogue for *E. grandis* (**Figure 4, Supplementary Figure 4 – 11, Supplementary Table 9 and 10**). All assemblies and annotations were of similar chromosome-level quality (**Figure 2, Supplementary Figure 1 – 6, Table 1, Supplementary Table 3 – 6**), with the added advantage that all assemblies and annotations were generated with the same data types and methodologies, leading to minimal methodology-based errors or biases when interpreting results. Similar studies depended on different types and amounts of sequencing data or multiple independently generated assemblies^{33-36,39,42,44,89-95}. The genomes of the 24 individuals will facilitate research and breeding of *E. grandis* and eucalypts, benefitting the forestry and conservation genomics industries⁹⁶. Our findings provide a solid foundation for further functional research that uses pangenome information to identify and validate genes of interest, and the methodologies can be extended to fruit and nut crops. We furthermore showcase an example application of the pangenome using the lignin biosynthesis pathway that has been partially functionally characterised⁸⁶.

In recent years pangenomes have become a “gold standard” in the field of crop genomics, however, few such studies exist for forest tree species^{40,42,44,46}. Synteny is the preservation of genome order and sequence⁹⁷. The average syntenic block size was 27.6 kbp, which is larger than previously reported

between different species of *Eucalyptus* (16.1 kbp¹⁰, 17.97 kbp¹¹ and 16.84 kbp⁹). This is likely due to the intra-specific comparisons and the high assembly uniformity for the 50 assemblies. Similar to previous studies including other eucalypt species^{9-12,98}, we observed that inversions events tend to be fewer in number but larger in size (**Figure 3 and 4, Supplementary Figure 5, Supplementary Table 11**). Studies in other crops have shown that inversions repress recombination within these regions maintaining a high level of conservation⁹⁹⁻¹⁰³ and may carry undesirable variants that can be difficult to remove through regular breeding methods. The functional implications of such events remain to be explored in the future.

Even though the majority of the assemblies are syntenic (62.9 – 69.4%, **Figure 3 and 4, Supplementary Figure 4 and 5, Supplementary Table 9**), this is lower than the 77.7% reported for *E. regnans*¹² and is more comparable to *E. viminalis*¹³. This is likely due to lower species diversity representation, as only one individual was represented¹² compared to ten¹³. We discovered a total of 2,239,370 SVs and 2,031,112 non-aligned regions overlapping a total of 27,967 and 8,452 genes in TAG0014_HAP1, respectively. Although we use a reference-based approach and acknowledge that there may be overlaps at the gene level due to complex SVs, this suggests that even though synteny is mostly conserved, many genes may potentially be affected and future studies should focus on validating the impact of SVs on gene regulation, expression and function (**Supplementary Table 17**). SVs calls showed varying levels of overlap (32.6 – 67.3%, **Supplementary Figure 8, Supplementary Table 11**), resulting in a set of 37,561 non-redundant SVs and 128,487 non-redundant non-aligned regions (a total of 200,551, **Supplementary Table 11**). This is higher than studies in poplar (142,202 SVs)⁴² and grape (132,518 SVs)⁹⁰, which reiterates the high level of genomic diversity of *E. grandis*.

Genome annotation in plants remains challenging, with long introns resulting in multiple gene models where a single model is the true gene^{4,104}. The use of full-length transcript sequencing technologies (Iso-Seq) and multi-tissue sampling of all individuals (**Supplementary Table 1**), in combination with multiple gene prediction methods and an integrated second round of pangenome informed annotation,

has allowed us to better predict and annotate genes for the 50 assemblies (**Figure 2, Table 1, Supplementary Table 6 and 8**). We have identified an average of 39,581 genes per genome (**Figure 2, Table 1, Supplementary Table 6 and 8**), with BUSCO completeness scores exceeding 98.50%.

The high quality of the gene annotations provides a preview of pangenome variation in *E. grandis*. At gene level, we have discovered 999,540 (50.1%) core, 85,277 (4.3%) softcore, 858,344 (43.4%) shell and 37,893 (1.9%) private genes (**Figure 5, Supplementary Table 13**) within 19,961 (16.1%) core, 1,730 (1.4%) softcore, 65,165 (52.6%) shell and 37,082 (29.9%) private HOGs (**Figure 5, Supplementary Table 13**). We found fewer HOGs in the core and softcore set than found by ¹¹ in a multi-species comparison, which likely represent key genes within the individual species. Most HOGs were classified as shell, with an average 43.3% (17,127/39,581) of the genes within any *E. grandis* individual exhibiting presence-absence variation (**Figure 5, Supplementary Table 13**). We found low numbers of private HOGs, 758 of the 39,581 genes (1.9% of genes), similar to¹¹. Ferguson *et al.*¹¹ suggested that shell and private genes may be the main contributors to local environmental adaptation, due to their numbers and functional classifications. This is somewhat supported by the observation that shell and private genes were enriched for gene ontologies pertaining known environment-specific terms (**Figure 6, Supplementary Figure 9, Supplementary Table 14 – 16**)^{35,39,42,90,91}, but needs to be functionally validated. Caution should be applied as to functional implications and biological conclusions as number of HOGs are not indicative of functional diversity.

Using the core lignin biosynthesis pathway⁸⁶ as a case study, we demonstrate how the pangenome can be used to resolve functional variation within a conserved metabolic pathway. Although we identified 10,104 lignin-associated genes, of which 1,744 were classified as CNVs, these genes were distributed across all pangenome compartments (4,743 core, 502 softcore, 4,552 shell and 308 private genes, corresponding to 8,537 HOGs 4,465 core, 443 softcore, 3,238 shell and 295 private HOGs). As expected for essential biosynthetic functions, key upstream and bottleneck enzymes such as CSE, C4H and F5H were largely restricted to the core genome and were represented by few orthogroups which may suggest strong evolutionary constraint. In contrast, enzyme families such as 4CL, CCR,

CAD and COMT showed broader distributions across pangenome compartments and higher HOG diversity, indicating localized gene family expansion and CNV. These patterns show that while the lignin biosynthesis pathway is conserved at functional level, individual enzymatic modules exhibit differential evolutionary plasticity, likely reflecting opportunities for fine-tuning pathway output without disruption of core metabolic integrity.

We provide a catalogue of SV and gene variation. Any one of the assemblies can serve as a reference for the species, although we encourage the use of TAG0014_HAP1 (V4.0) as reference of *E. grandis* and related species in the subgenus. Using a more appropriate and complete reference genome, or even a reference that includes gene, SV and TE level information, will reduce molecular marker representation biases that influence the accuracy of prediction strategies and completely overlook the importance of SVs^{23,24} and TEs¹⁰⁵. Future studies can investigate the effects of SVs and TEs on gene expression and functionally validate the findings produced here through genome editing⁴². In addition, this data can be used to develop better gene editing targets using more complete gene models⁴⁴, understand the process of early domestication and genome evolution and response to environmental stressors^{30,33,35,40,46,105-112}.

As the genomic blueprints and their regulatory mechanisms are uncovered in greater detail, allele-defined genome-based research is becoming an undeniable reality and will become crucial even for large complex genomes like pine¹¹³, and sugarcane⁵⁵. An unexplored avenue would be to integrate a genome architecture layer. This has delivered insights into how SVs and TEs affect the regulation of gene expression in traits of importance in grape⁹⁰, soybean³¹, pepper³³ and tomato^{34,35}. The resource can also be combined with other omics data to provide an extraordinarily rich resource for understanding interactions between multiple variables and their effects on growth, development and responses to the environment¹¹⁴.

Figures

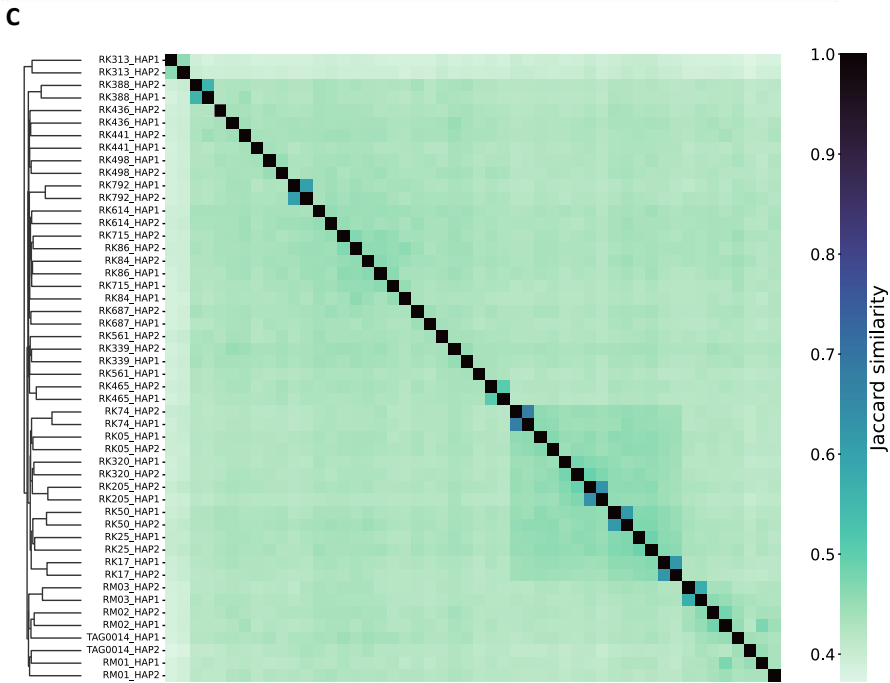
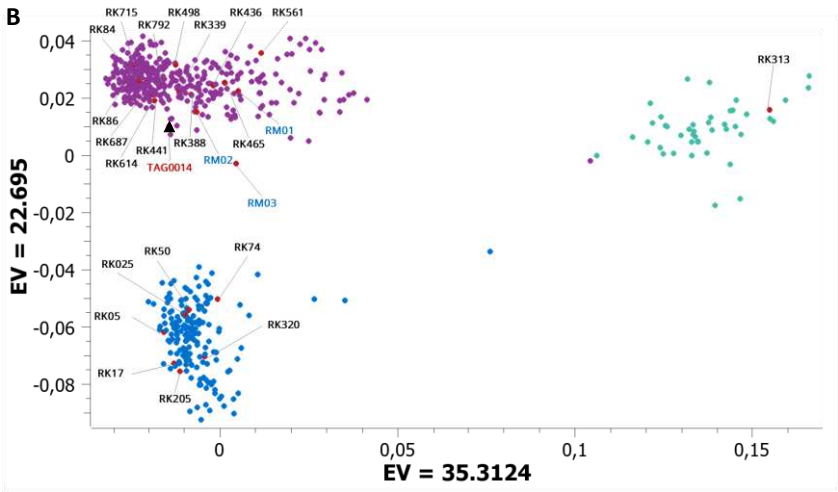
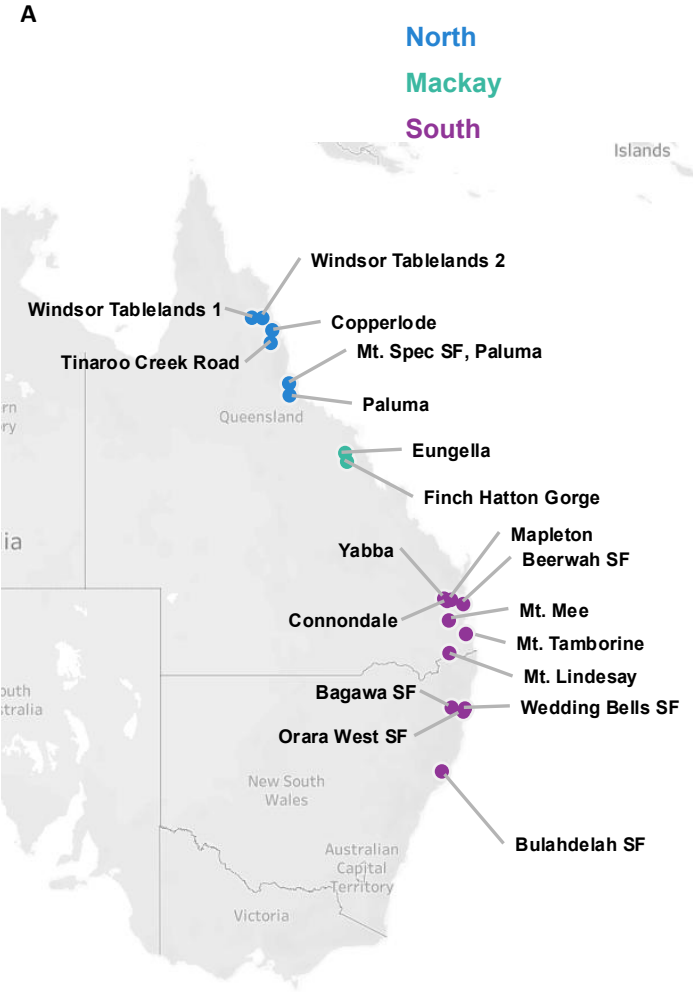


Figure 1 Genetic relatedness of the pangenome individuals compared to the natural distribution and population structure of *E. grandis*. A) Location of sampled provenances in the natural range of *E. grandis*. Provenances are coloured and grouped by broad geographic regions, North (blue), Mackay (green) and South (purple). B) Population structure based on Principal Component Analysis (PCA) using 17,036 informative SNPs. Sample IDs correspond to sample and provenance descriptions in Table 1. EV represents the Eigenvalue for the axis. The clustering position of the TAG0014 reference genotype is indicated by a black triangle. C) Relatedness heatmap of 50 haplotype resolved genome assemblies.

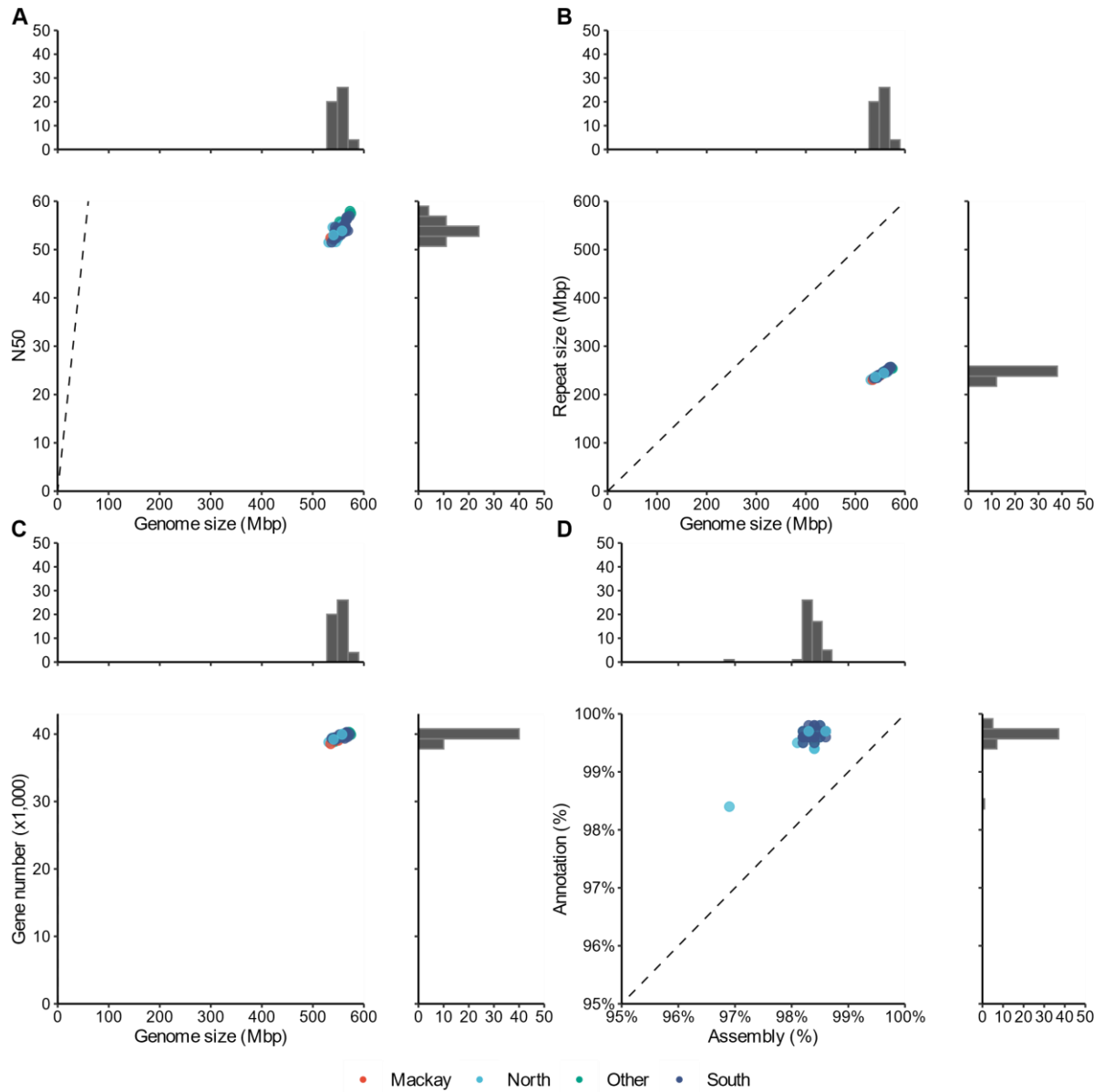


Figure 2 Genome assembly and annotation statistics of 25 diverse *Eucalyptus grandis* pangenome individuals. A) Genome size (x-axis) per genome assembly N50 (y-axis) in megabase pairs. B) Genome size (x-axis) occupied by repeat elements (y-axis) in megabase pairs. C) Total number of genes in thousands (y-axis) per genome assembly size (x-axis). D) Genome assembly BUSCO completeness score per (percentage, x-axis) vs annotation BUSCO completeness score (percentage, y-axis). Subpopulation colour is indicated at the bottom as North, Mackay, South and Other where Other represents the TAG0014 reference and the fifth generation breeding material (see **Table 1, Supplementary Table 1 – 6**) in A - D.

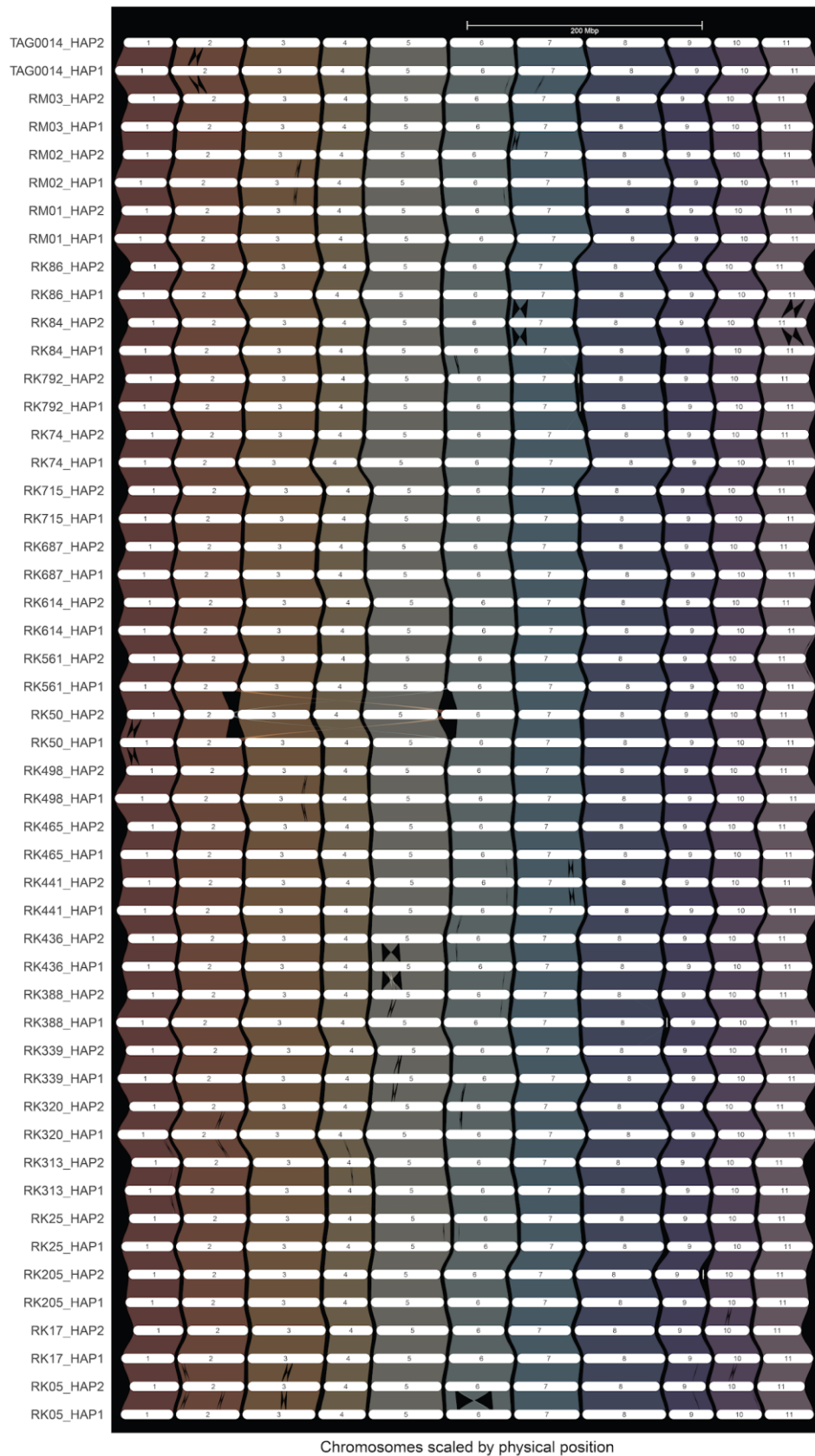


Figure 3 GENESPAC synteny map of 25 *E. grandis* individuals based on whole-genome alignments. Collinear blocks between the same chromosomes are shown as coloured braids and chromosomes are visualised as white blocks along the x-axis that are scaled by physical genome size in megabase pairs. Black wedges represent regions that do not align between the genome assemblies.

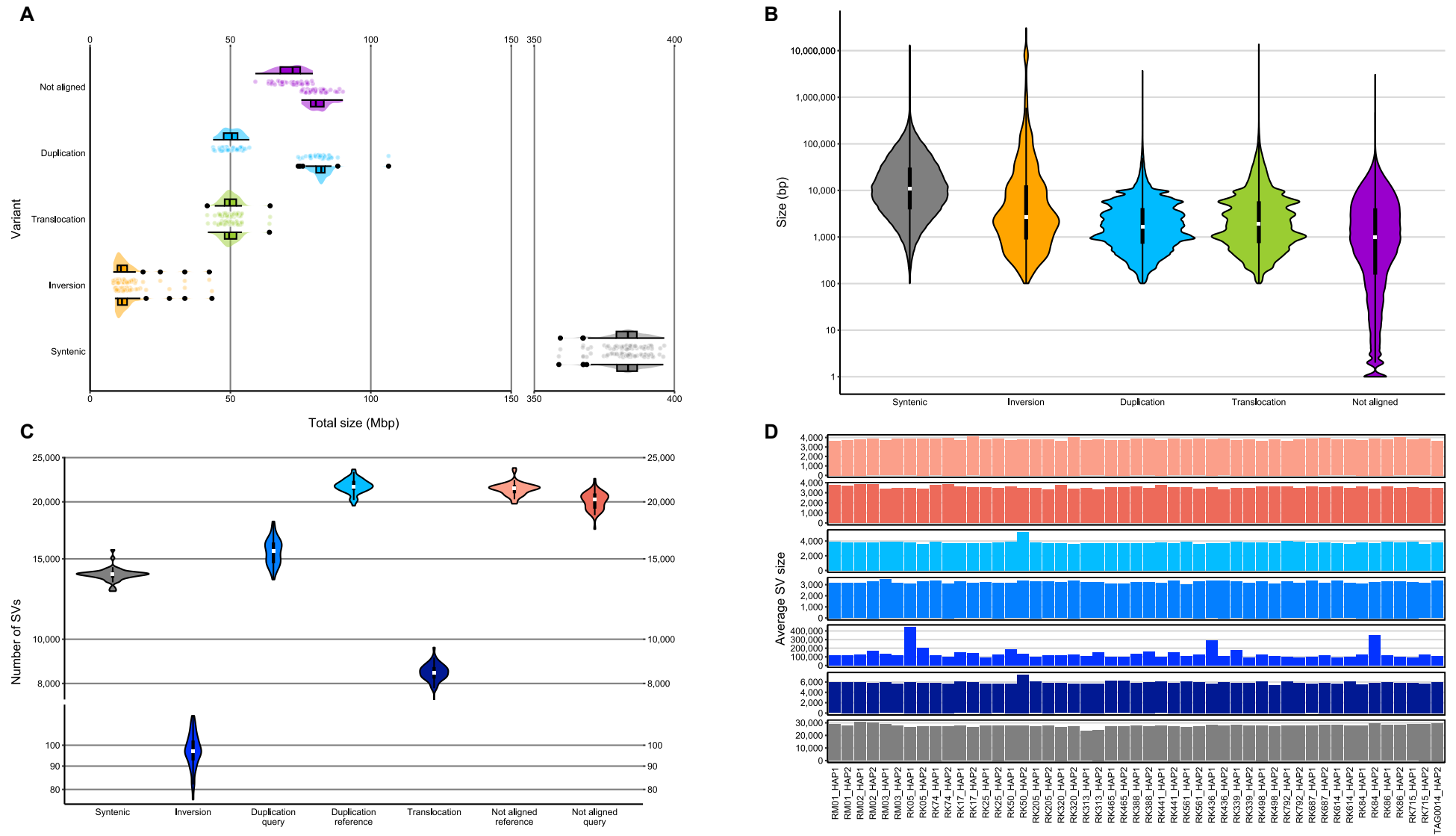


Figure 4 Genome-wide SV statistics generated for 25 *E. grandis* pangenome individuals using SyRI⁸⁷. A) Total size contribution of syntenic regions, SVs and regions that do not align to genome-wide variation. The variant type is indicated on the y-axis with the total contribution per comparison of the variant type indicated

in megabase pairs on the x-axis. Each dot in the raincloud plot indicates the total SV contribution for a single genome, where the top section of each raincloud indicates the reference genome (TAG0014_HAP1) contribution and the bottom the query genome contribution (see Supplementary Figure 4 and Supplementary Table 9 for the numbers). B) Size distribution of syntenic regions, SVs and regions that do not align for all assembly comparisons. The size (bp) of different variant types is indicated on the y-axis (on a logarithmic scale) and the variant type on the x-axis. See Supplementary Figure 5 and Supplementary Table 9 for more detail. C) Total number of syntenic regions, SVs and not-aligned regions per variant type. Variant type is indicated on the x-axis and the number of variants on the y-axis. See Supplementary Figure 5 and Supplementary Table 9 if more detail is required. D) Average size per syntenic region, SVs and non-aligned regions per variant type and genome queried. The query genome is indication on the x-axis and the average variant size on the y-axis. The colour scheme follows that of C.

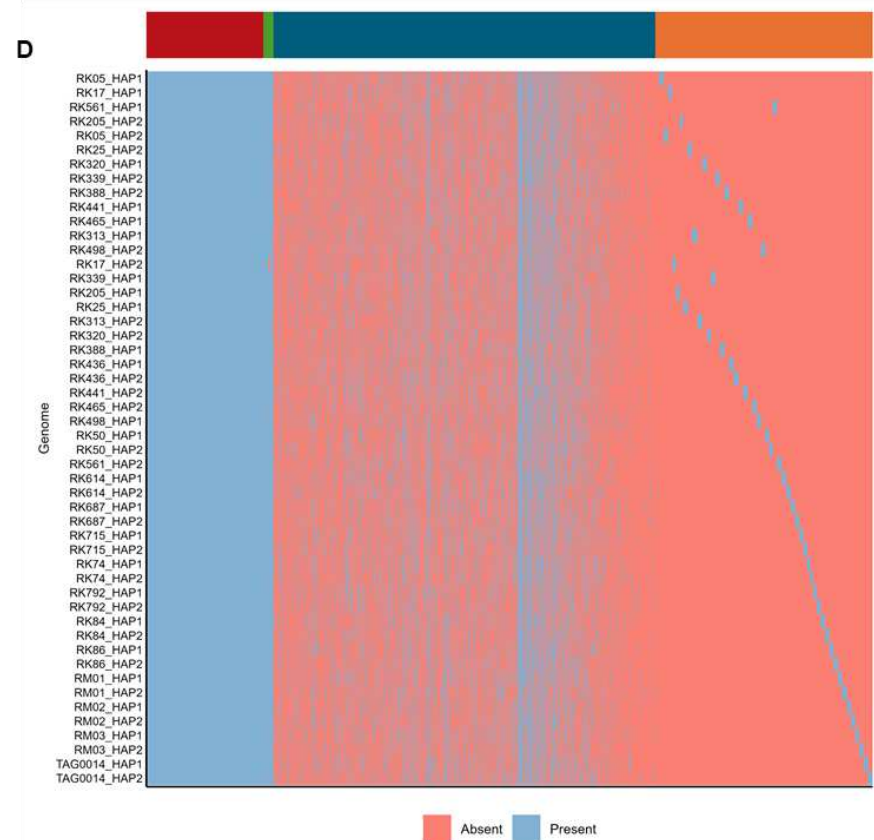
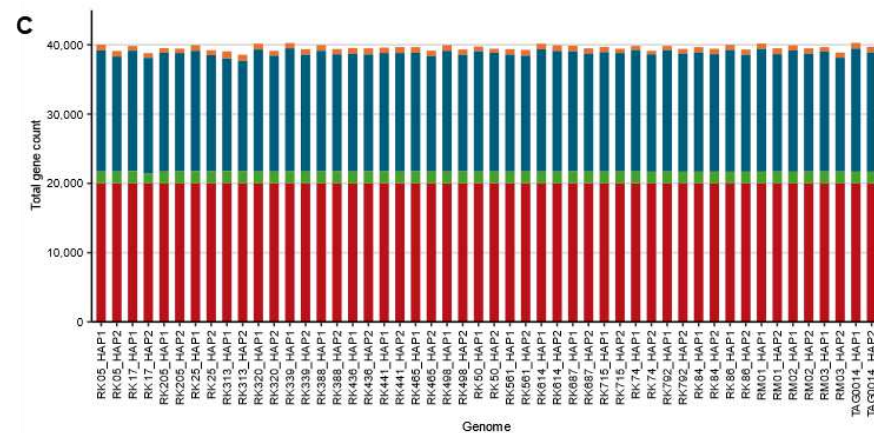
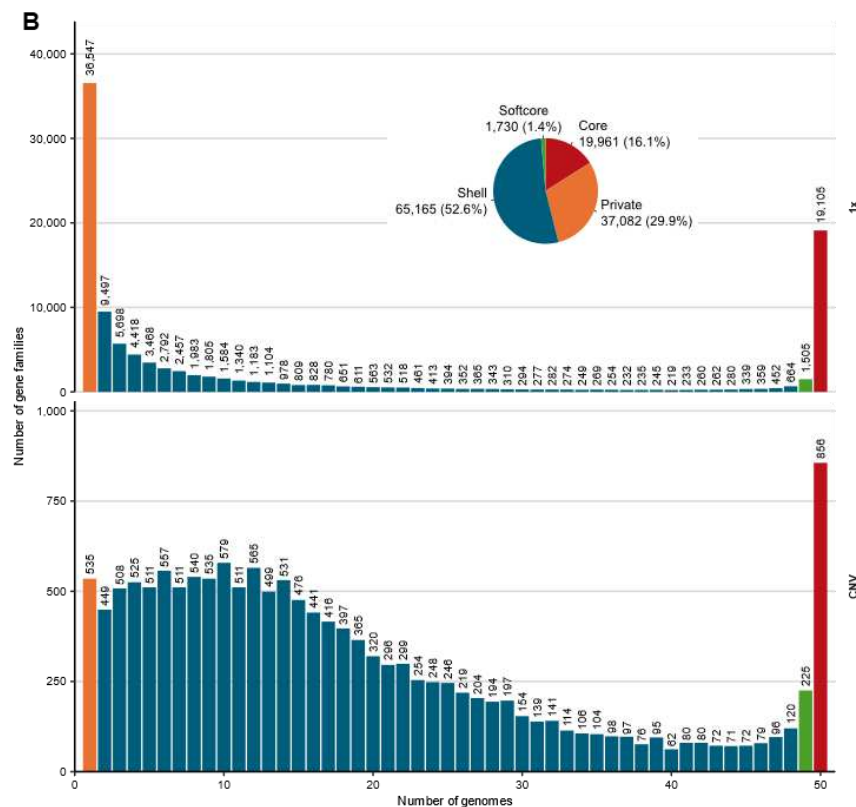
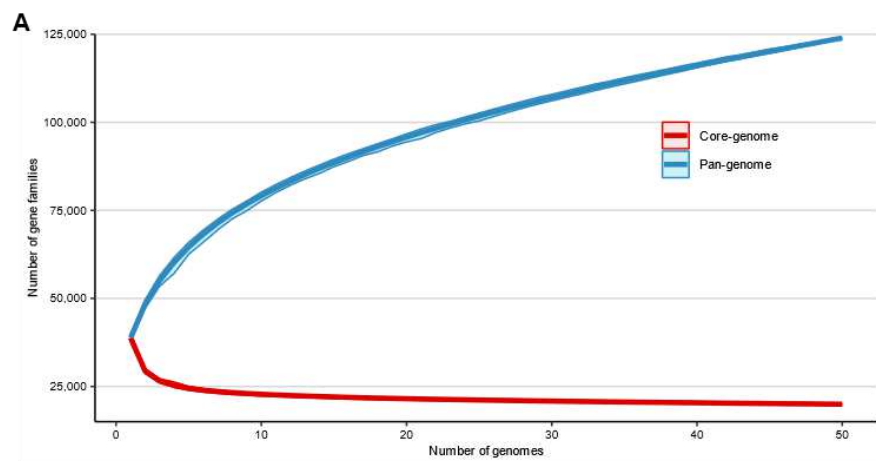


Figure 5 Pangene analysis of *E. grandis*. A) Variation of gene families in the pan- and core-genome with additional *E. grandis* genomes. At each given number of genomes (x-axis), the number of combinations is 100. The ribbon represents minimum and maximum number of gene families (y-axis) with the centre line indicating the mean. B) Number of core, softcore, shell and private genes (y-axis) in 50 haplotype-resolved genomes (x-axis) (see Supplementary Table 13). C) Composition of the *E. grandis* pangenome and individual haplotype-resolved genomes. The histogram shows the number of gene families (y-axis) in the 50 haploid genomes (x-axis) with different frequencies where histogram colours correspond to core, softcore, shell and private genes (see also Supplementary Table 12) in single copy or as copy number variants. The pie chart shows the proportion of the total gene families marked by each composition. D) Presence and absence variation of gene families in 50 haplotypes (see **Supplementary Data 2**). The genome is indicated on the y-axis.

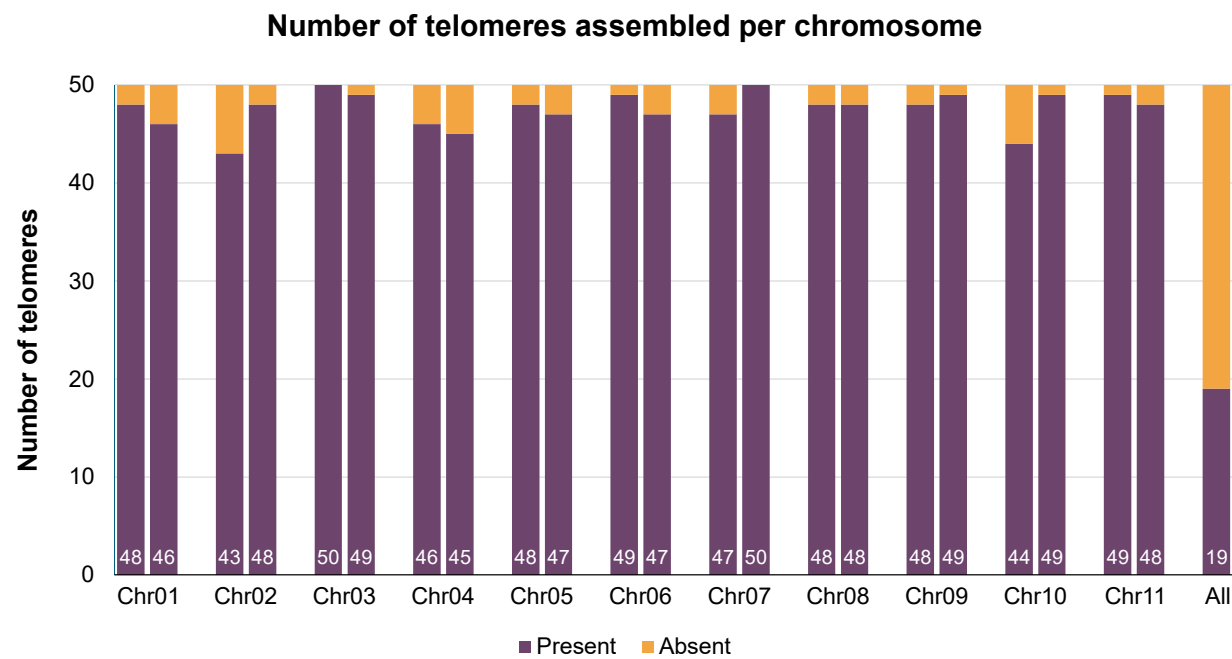
Tables

Table 1 Genome assembly and annotation statistics of 25 phased *Eucalyptus grandis* genome assemblies (50 haplotypes). Assemblies where all 22 telomeres have been assembled are shown in bold in the Number of telomeres column. The lowest annotation BUSCO completeness score is indicated in bold. HAP1 indicates Haplotype 1 and HAP2 indicates Haplotype 2 of each phased assembly. For more technical detail, see **Supplementary Figure 1 – 5, Supplementary Figure 3, Supplementary Table 3 – 6.**

Assembly	Number of contigs	Number of scaffolds	Total length (Mbp)	N50 (Mbp)	L50	Number of telomeres (/22)	Genome BUSCO	GC (%)	Chloroplast size (Kbp)	Mitochondrial size (Kbp)	Repeat content (%)	Number of genes	Protein BUSCO
RK05_HAP1	54	12	558.65	54.90	5	21	98.30%	39.49%	160.2	442.5	44.18%	40,029	99.70%
RK05_HAP2	63	11	539.04	54.62	5	22	98.20%	39.49%	160.2	442.5	43.49%	39,118	99.50%
RK17_HAP1	34	11	557.11	55.20	5	22	98.30%	39.46%	160.2	469.6	44.10%	39,846	99.60%
RK17_HAP2	37	11	531.00	51.51	5	21	96.90%	39.44%	160.2	469.6	43.42%	38,803	98.40%
RK205_HAP1	29	12	548.71	52.32	5	22	98.50%	39.44%	160.2	469.7	43.85%	39,516	99.40%
RK205_HAP2	33	14	538.03	52.09	5	22	98.50%	39.45%	160.2	469.7	43.42%	39,454	99.40%
RK25_HAP1	61	11	557.16	53.87	5	20	98.60%	39.49%	160.2	470.5	43.88%	39,948	99.70%
RK25_HAP2	65	11	540.54	53.02	5	20	98.30%	39.48%	160.2	470.5	43.73%	39,227	99.70%
RK313_HAP1	52	13	549.61	53.65	5	19	98.30%	39.47%	160.2	470.0	43.56%	39,053	99.70%
RK313_HAP2	43	11	534.88	52.44	5	21	98.30%	39.46%	160.2	470.0	43.14%	38,555	99.70%
RK320_HAP1	92	11	566.68	56.44	5	19	98.50%	39.46%	160.4	447.3	44.14%	40,158	99.60%
RK320_HAP2	96	11	539.53	52.55	5	19	98.60%	39.47%	160.4	447.3	43.70%	39,136	99.70%
RK339_HAP1	62	21	566.04	56.20	5	20	98.50%	39.48%	160.2	456.9	44.25%	40,272	99.70%
RK339_HAP2	54	15	548.06	54.45	5	20	98.30%	39.46%	160.2	456.9	43.88%	39,380	99.70%
RK388_HAP1	44	14	566.11	56.64	5	22	98.40%	39.47%	160.2	457.4	44.59%	39,963	99.80%
RK388_HAP2	46	14	544.84	53.20	5	21	98.30%	39.46%	160.2	457.4	43.76%	39,400	99.70%
RK436_HAP1	46	12	555.71	53.29	5	22	98.40%	39.46%	160.3	455.6	44.21%	39,541	99.60%
RK436_HAP2	57	14	543.51	52.56	5	22	98.60%	39.48%	160.3	455.6	43.27%	39,503	99.70%
RK441_HAP1	87	11	568.39	53.90	5	22	98.20%	39.48%	160.2	456.1	45.11%	39,612	99.70%
RK441_HAP2	82	11	554.69	53.05	5	22	98.60%	39.43%	160.2	456.1	44.19%	39,675	99.60%
RK465_HAP1	57	22	560.89	54.94	5	21	98.50%	39.44%	159.4	471.0	43.94%	39,666	99.60%
RK465_HAP2	41	11	543.82	54.66	5	20	98.20%	39.46%	159.4	471.0	43.22%	39,155	99.60%
RK498_HAP1	57	13	571.67	56.93	5	20	98.20%	39.45%	160.2	457.6	44.92%	39,974	99.60%
RK498_HAP2	52	11	546.87	53.67	5	21	98.40%	39.49%	160.2	457.6	43.59%	39,323	99.50%

RK50_HAP1	43	12	560.59	55.03	5	22	98.40%	39.48%	160.3	447.0	43.42%	39,750	99.40%
RK50_HAP2	44	11	544.84	51.59	5	22	98.40%	39.47%	160.3	447.0	44.38%	39,435	99.40%
RK561_HAP1	54	15	562.50	55.16	5	22	98.20%	39.45%	159.9	489.6	44.51%	39,370	99.50%
RK561_HAP2	42	12	541.07	52.07	5	21	98.30%	39.47%	159.9	489.6	43.64%	39,271	99.70%
RK614_HAP1	44	19	564.97	54.18	5	22	98.20%	39.49%	160.2	457.0	44.53%	40,154	99.60%
RK614_HAP2	33	13	552.19	53.37	5	22	98.50%	39.44%	160.2	457.0	43.94%	39,922	99.80%
RK687_HAP1	56	17	565.74	56.16	5	21	98.30%	39.45%	159.2	447.6	44.05%	39,882	99.80%
RK687_HAP2	49	12	548.03	53.84	5	21	98.30%	39.46%	159.2	447.6	43.85%	39,499	99.60%
RK715_HAP1	35	13	563.89	54.74	5	21	98.20%	39.46%	159.9	447.1	44.45%	39,685	99.60%
RK715_HAP2	42	15	542.26	52.44	5	22	98.30%	39.45%	159.9	447.1	43.76%	39,432	99.60%
RK74_HAP1	51	12	563.50	54.70	5	21	98.20%	39.50%	159.8	447.0	43.67%	39,843	99.70%
RK74_HAP2	59	11	547.39	53.98	5	21	98.10%	39.47%	159.8	447.0	43.57%	39,145	99.50%
RK792_HAP1	79	16	562.41	53.68	5	20	98.30%	39.45%	159.9	490.8	44.21%	39,850	99.70%
RK792_HAP2	68	14	545.56	52.38	5	19	98.30%	39.43%	159.9	490.8	43.96%	39,421	99.70%
RK84_HAP1	92	13	562.27	54.62	5	20	98.40%	39.51%	159.4	446.7	44.38%	39,667	99.60%
RK84_HAP2	94	18	543.02	53.78	5	22	98.60%	39.47%	159.4	446.7	43.28%	39,434	99.70%
RK86_HAP1	46	14	564.90	54.09	5	20	98.40%	39.49%	159.4	456.9	44.42%	40,010	99.80%
RK86_HAP2	40	14	537.08	51.53	5	21	98.30%	39.46%	159.4	456.9	43.61%	39,337	99.70%
RM01_HAP1	58	11	572.78	57.95	5	21	98.30%	39.47%	159.4	457.5	44.43%	40,173	99.60%
RM01_HAP2	57	11	556.63	54.76	5	22	98.40%	39.47%	159.4	457.5	44.45%	39,510	99.70%
RM02_HAP1	81	12	574.50	57.49	5	20	98.40%	39.48%	160.2	456.3	44.19%	39,915	99.50%
RM02_HAP2	82	13	556.78	55.53	5	17	98.20%	39.46%	160.2	456.3	44.06%	39,495	99.60%
RM03_HAP1	52	11	557.87	54.74	5	22	98.20%	39.46%	160.2	456.0	43.85%	39,668	99.70%
RM03_HAP2	54	11	542.81	53.41	5	22	98.40%	39.47%	160.2	456.0	43.86%	38,882	99.70%
TAG0014_HAP1	37	12	570.83	57.17	5	21	98.50%	39.50%	160.2	456.9	44.99%	40,292	99.60%
TAG0014_HAP2	46	11	552.41	55.77	5	20	98.30%	39.48%	160.2	456.9	43.78%	39,705	99.60%

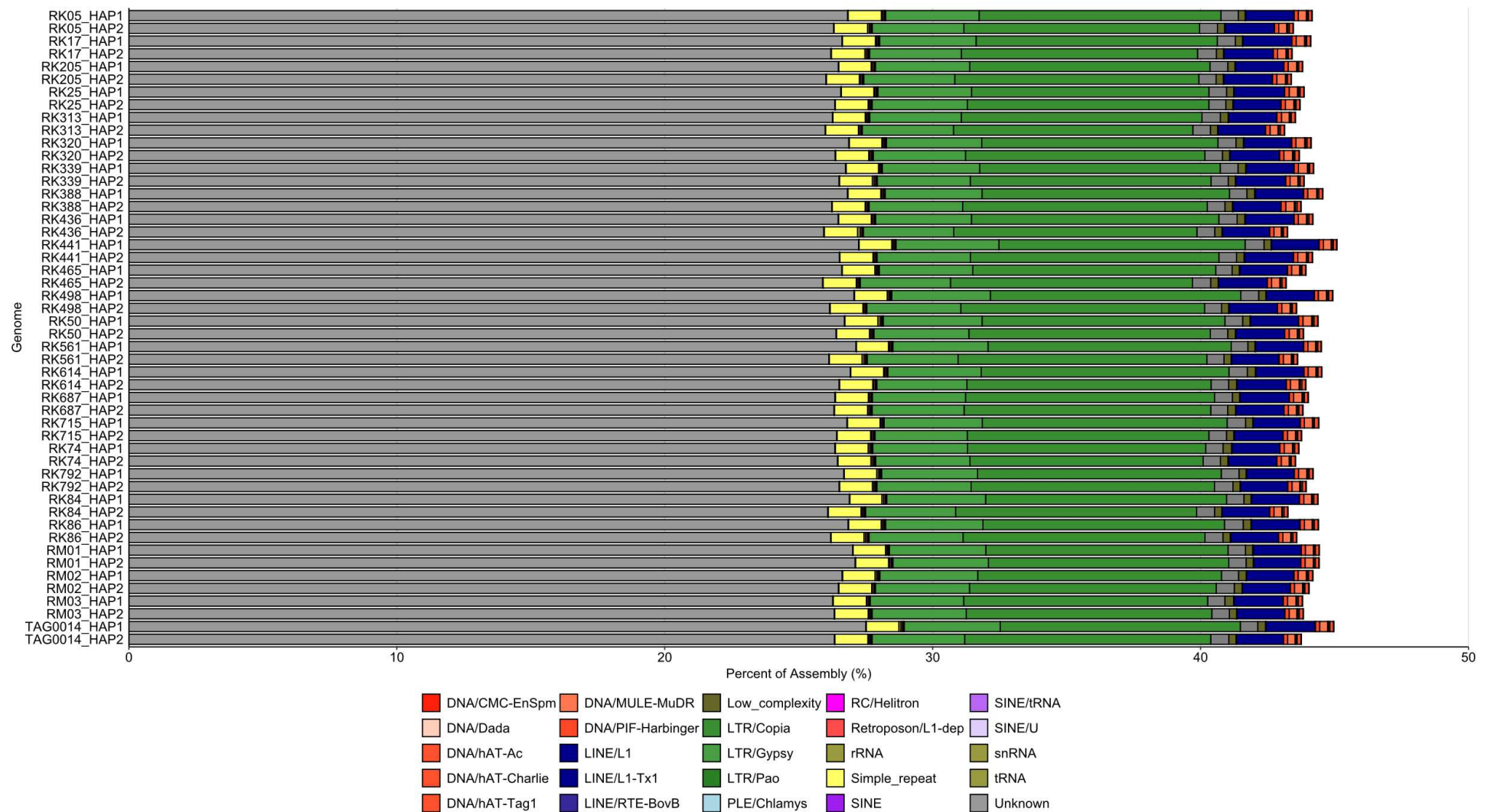
Supplementary figures



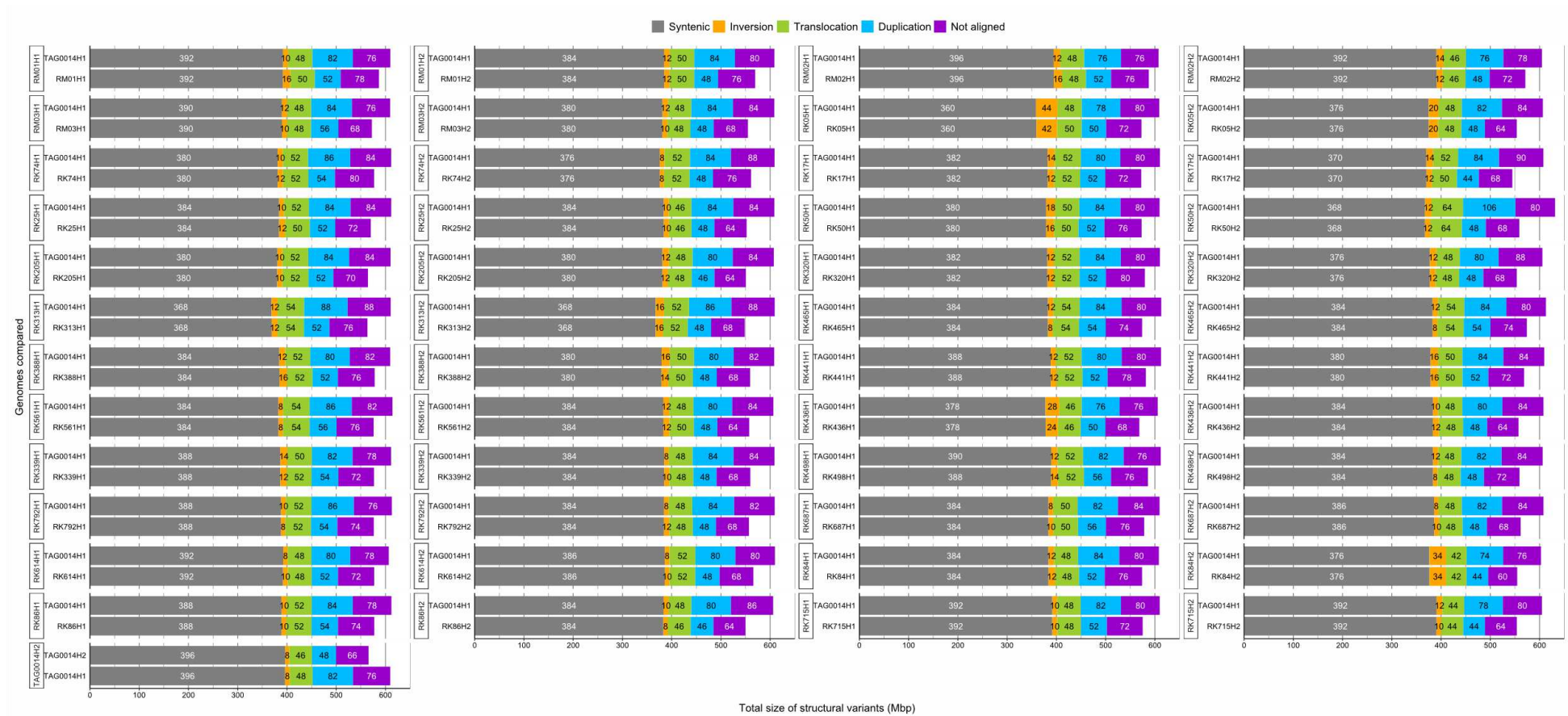
Supplementary Figure 1 Number of telomeres assembled per haplotype (11 chromosomes) out of 50 *E. grandis* haplotype-resolved assemblies. A complete set of 22 telomeres were assembled for 19 of the 50 haplotype assemblies (Green checks in **Supplementary Figure 2, Table 1**).



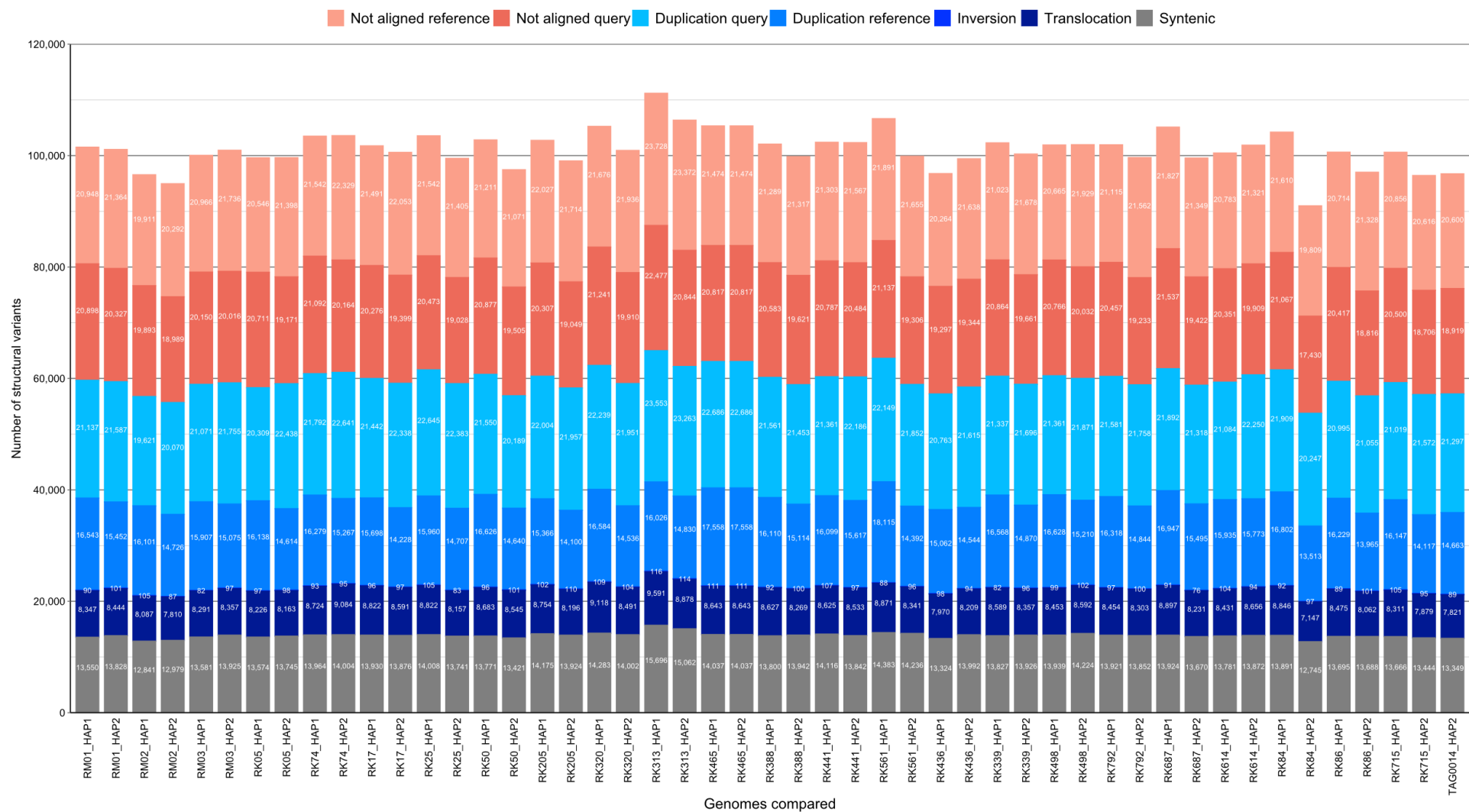
Supplementary Figure 2 Haplotype-resolved genome contiguity of the *E. grandis* genomes. Contig position of the genome assemblies were generated using GENESPACE⁶¹. A blue-yellow palette of 20 colours was selected for all assemblies i.e. each cycle through colours represent a set of 20 contigs. Telomere sequences are shown by a red asterisk and the numbers of contigs that make up individual chromosome scaffolds are shown on the right. Green tick marks indicate presence of all 22 telomeres in that assembly. The red box marks the TAG0014 HAP1 and HAP2 reference assemblies. Chromosome position and size are indicated on the x-axis, and chromosome number on the y-axis. See also **Table 1, Supplementary Figure 1, Supplementary Table 2 – 4.**



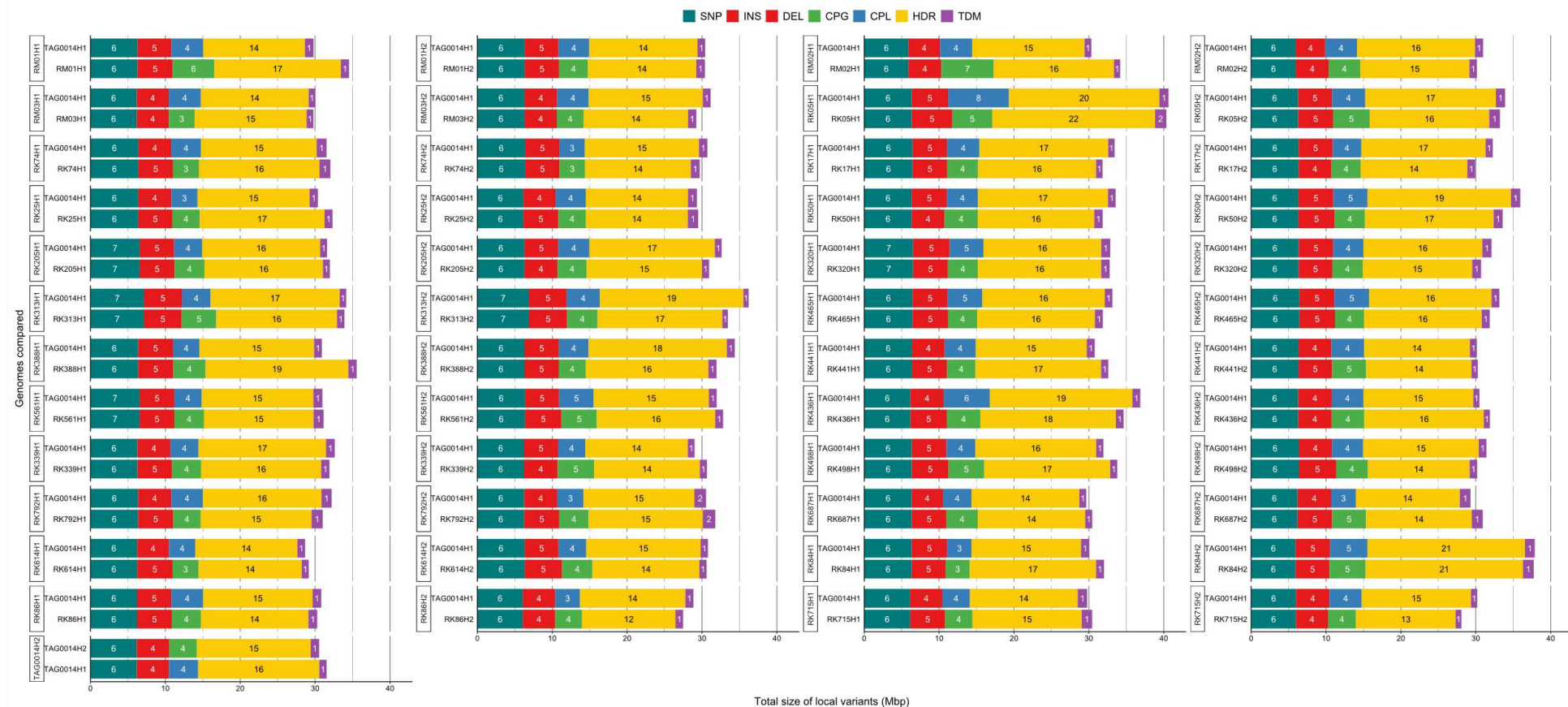
Supplementary Figure 3 *E. grandis* repeat content for haplotype assemblies. The total percentage of repeat elements is shown on the x-axis per haplotype assembly (y-axis). Colours of elements are indicated on the legend key. See Repeat content summary **Table 1**, and detailed repeat content reports in **Supplementary Table 7** and **extended Supplementary Table 7**.



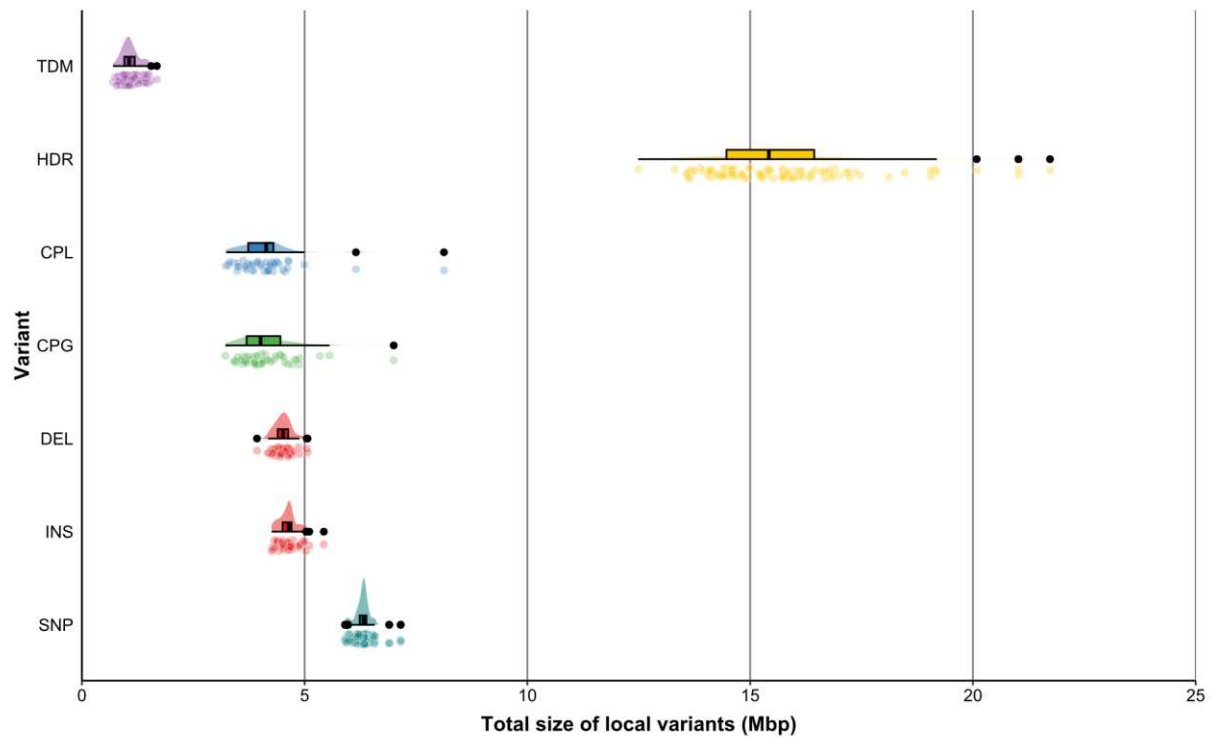
Supplementary Figure 4 Total contribution of SVs (in megabase pairs) per reference alignment. The reference genome used is TAG0014 HAP1. Haplotypes are denoted as H due to space constraints. Colour of the variant type is shown in the key. See also **Supplementary Table 9**.



Supplementary Figure 5 Number of SVs per genome alignment. The reference genome used is TAG0014_HAP1. Haplotype assemblies are indicated on the x-axis and the number of variants on the y-axis. Numbers inside bars are the actual number of events per indicated variant type. Variant type colours are indicated by the key. Also see **Figure 3**, **Figure 4** and **Supplementary Table 9**.

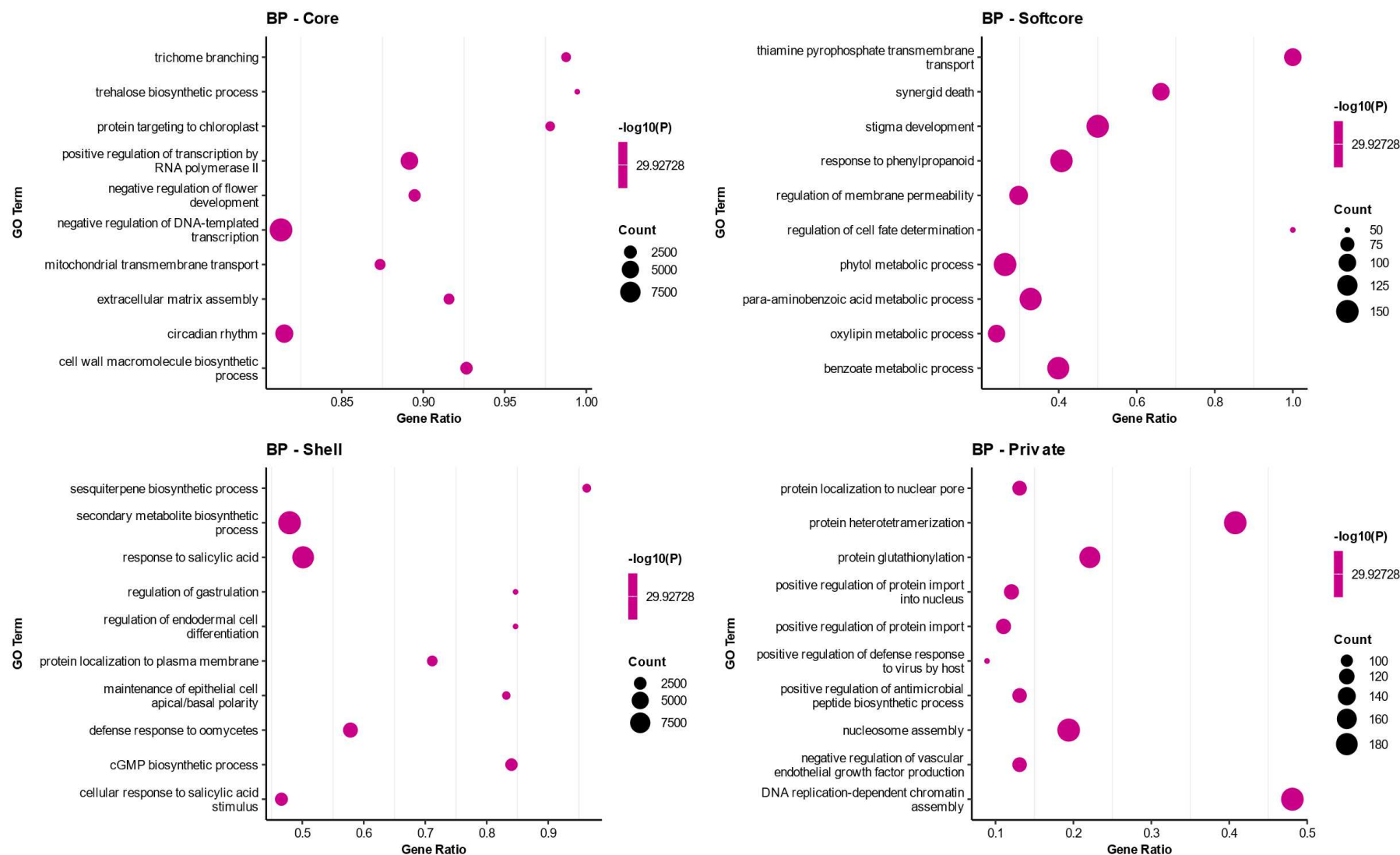


Supplementary Figure 6 Total contribution of local variants (in megabase pairs) per alignment compared. The reference genome used is TAG0014_HAP1. Haplotypes are denoted as H due to space constraints. Colour of the variant type is shown in the key. See also **Supplementary Table 10** and **Supplementary Table 11**.

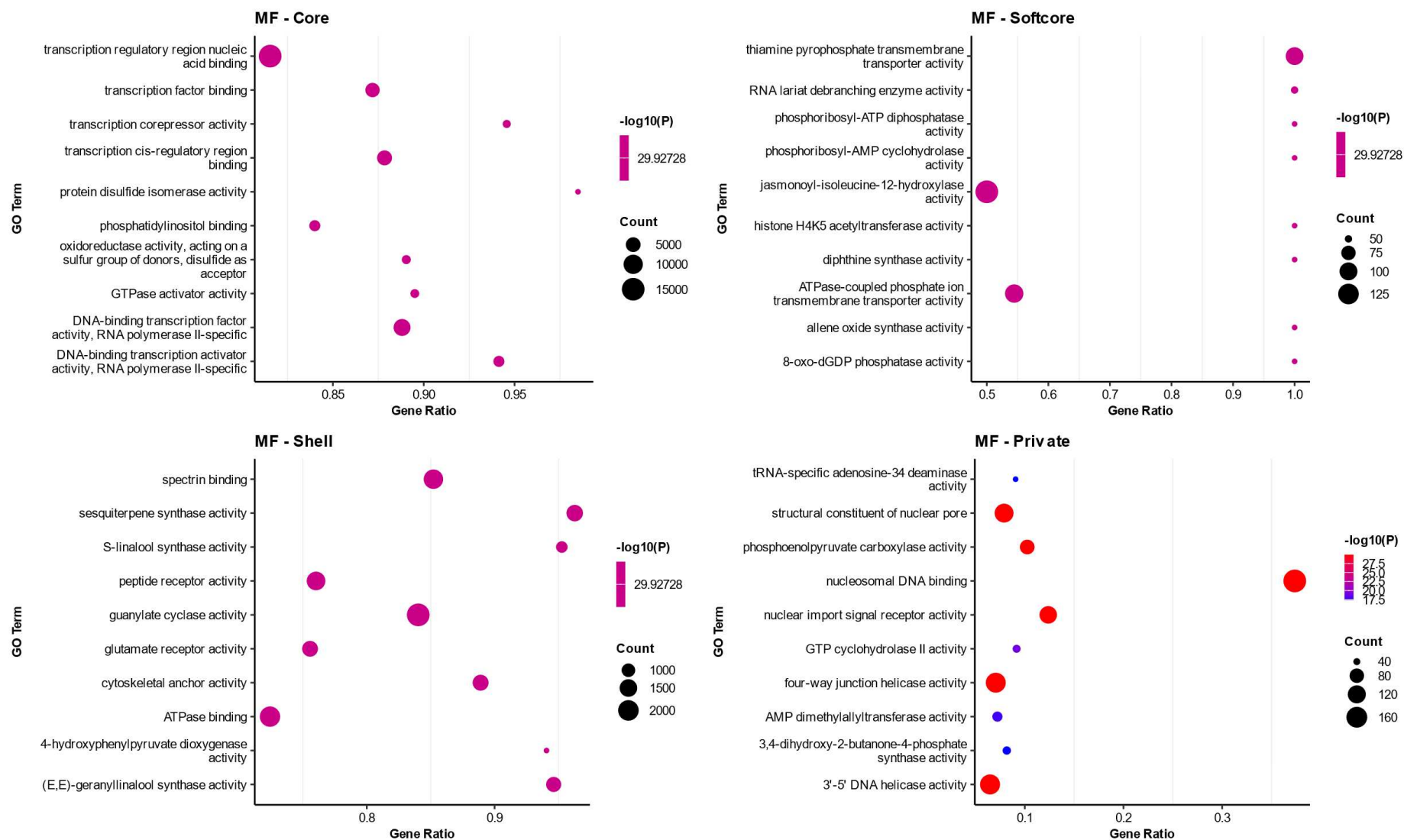


Supplementary Figure 7 Total size contribution of local variants (LV). The variant type is indicated on the y-axis with the total contribution per comparison of the variant type indicated in megabase pairs on the x-axis. Each dot in the raincloud plot indicates the total LV contribution for a single genome. See also **Supplementary Table 10** and **Supplementary Table 11**.

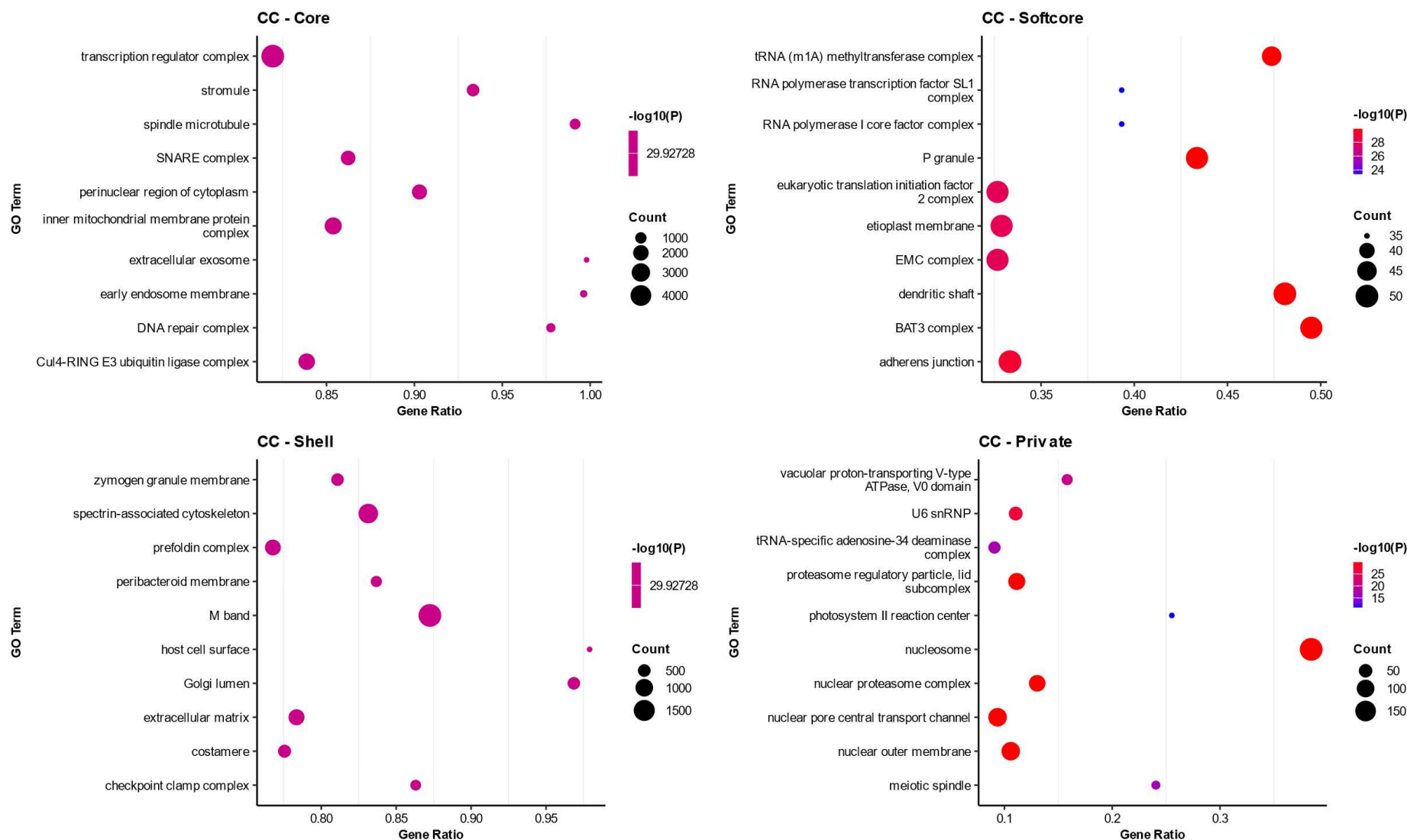
		RK05		RK17		RK205		RK25		RK313		RK320		RK339		RK388		RK436		RK441		RK465		RK498		RK50		RK561		RK614		RK687		RK715		RK74		RK792		RK84		RK86		RM01		RM02		RM03		TAG0014	
		HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2	HAP1	HAP2		
RK05	HAP1	1.00	0.44	0.42	0.42	0.43	0.42	0.43	0.42	0.41	0.40	0.43	0.42	0.39	0.40	0.39	0.38	0.38	0.39	0.39	0.39	0.39	0.39	0.38	0.39	0.42	0.41	0.40	0.38	0.39	0.38	0.39	0.39	0.37	0.37	0.43	0.43	0.38	0.38	0.39	0.36	0.38	0.38	0.38	0.37	0.37	0.36	0.38	0.39	0.35	
	HAP2		1.00	0.42	0.42	0.42	0.42	0.43	0.42	0.40	0.39	0.42	0.42	0.39	0.38	0.38	0.38	0.37	0.38	0.39	0.39	0.39	0.38	0.39	0.39	0.41	0.40	0.39	0.39	0.38	0.38	0.38	0.38	0.38	0.42	0.42	0.37	0.38	0.38	0.36	0.38	0.38	0.37	0.37	0.36	0.35	0.37	0.38	0.36		
RK17	HAP1			1.00	0.62	0.42	0.41	0.44	0.43	0.40	0.39	0.43	0.41	0.38	0.39	0.39	0.38	0.38	0.39	0.39	0.38	0.39	0.38	0.39	0.38	0.39	0.42	0.41	0.38	0.39	0.38	0.39	0.39	0.38	0.38	0.41	0.41	0.38	0.38	0.39	0.36	0.37	0.38	0.37	0.37	0.36	0.35	0.38	0.38	0.36	
	HAP2				1.00	0.42	0.41	0.43	0.42	0.40	0.40	0.42	0.41	0.38	0.39	0.38	0.38	0.37	0.38	0.40	0.40	0.39	0.39	0.39	0.39	0.41	0.40	0.39	0.39	0.38	0.38	0.39	0.38	0.38	0.38	0.41	0.42	0.38	0.38	0.39	0.36	0.38	0.38	0.37	0.38	0.36	0.36	0.38	0.39	0.36	
RK205	HAP1					1.00	0.61	0.42	0.42	0.40	0.39	0.44	0.44	0.38	0.39	0.38	0.38	0.38	0.38	0.39	0.39	0.38	0.38	0.38	0.39	0.42	0.41	0.39	0.38	0.38	0.38	0.38	0.39	0.38	0.37	0.37	0.41	0.41	0.38	0.38	0.38	0.35	0.37	0.37	0.37	0.37	0.36	0.35	0.38	0.39	0.35
	HAP2						1.00	0.43	0.42	0.41	0.40	0.45	0.46	0.39	0.40	0.39	0.38	0.38	0.39	0.40	0.40	0.38	0.39	0.38	0.39	0.42	0.41	0.40	0.38	0.38	0.38	0.39	0.38	0.38	0.38	0.41	0.42	0.38	0.38	0.39	0.36	0.37	0.38	0.37	0.38	0.36	0.36	0.38	0.39	0.36	
RK25	HAP1							1.00	0.46	0.41	0.40	0.42	0.43	0.39	0.40	0.39	0.39	0.38	0.39	0.40	0.39	0.40	0.39	0.38	0.39	0.43	0.42	0.39	0.39	0.39	0.40	0.39	0.38	0.38	0.38	0.42	0.42	0.38	0.39	0.39	0.36	0.38	0.38	0.38	0.37	0.38	0.36	0.37	0.38	0.40	0.36
	HAP2								1.00	0.42	0.41	0.43	0.42	0.40	0.40	0.39	0.39	0.39	0.39	0.40	0.39	0.39	0.39	0.40	0.39	0.43	0.42	0.40	0.39	0.39	0.39	0.40	0.39	0.38	0.38	0.41	0.42	0.38	0.38	0.39	0.36	0.39	0.38	0.38	0.38	0.37	0.36	0.39	0.40	0.36	
RK313	HAP1									1.00	0.46	0.37	0.37	0.36	0.35	0.35	0.36	0.35	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.37	0.37	0.37	0.37	0.36	0.36	0.36	0.35	0.35	0.34	0.36	0.35	0.36	0.35	0.36	0.35	0.35	0.34	0.35	0.33	0.33	0.35	0.36	0.33		
	HAP2										1.00	0.39	0.38	0.36	0.36	0.37	0.37	0.36	0.36	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.36	0.37	0.36	0.35	0.36	0.35	0.38	0.38	0.36	0.36	0.36	0.34	0.36	0.35	0.35	0.34	0.36	0.34	0.34	0.35	0.36	0.33	
RK320	HAP1											1.00	0.45	0.38	0.39	0.38	0.38	0.37	0.38	0.38	0.38	0.39	0.38	0.39	0.39	0.42	0.40	0.39	0.38	0.39	0.38	0.39	0.37	0.37	0.37	0.40	0.41	0.38	0.37	0.38	0.36	0.37	0.37	0.36	0.37	0.36	0.36	0.38	0.39	0.36	
	HAP2												1.00	0.39	0.40	0.39	0.38	0.38	0.39	0.40	0.40	0.39	0.39	0.39	0.39	0.42	0.41	0.40	0.38	0.38	0.38	0.40	0.38	0.38	0.41	0.43	0.38	0.39	0.38	0.36	0.39	0.39	0.38	0.38	0.36	0.36	0.38	0.39	0.37		
RK339	HAP1													1.00	0.41	0.39	0.39	0.39	0.40	0.40	0.40	0.39	0.39	0.39	0.40	0.39	0.38	0.41	0.39	0.40	0.40	0.40	0.39	0.39	0.38	0.39	0.40	0.39	0.39	0.37	0.39	0.39	0.38	0.38	0.36	0.36	0.38	0.38	0.37		
	HAP2														1.00	0.40	0.40	0.40	0.40	0.41	0.41	0.40	0.40	0.40	0.41	0.40	0.39	0.41	0.40	0.39	0.40	0.39	0.39	0.39	0.40	0.39	0.39	0.40	0.39	0.39	0.37	0.40	0.39	0.39	0.39	0.37	0.37	0.38	0.39	0.37	
RK388	HAP1																1.00	0.54	0.39	0.39	0.39	0.42	0.39	0.39	0.39	0.40	0.38	0.40	0.40	0.40	0.40	0.40	0.39	0.39	0.38	0.37	0.39	0.39	0.39	0.38	0.38	0.38	0.38	0.36	0.36	0.38	0.38	0.36			
	HAP2																	1.00	0.38	0.39	0.40	0.40	0.39	0.39	0.39	0.40	0.39	0.38	0.40	0.39	0.39	0.39	0.39	0.38	0.37	0.39	0.39	0.39	0.38	0.39	0.36	0.38	0.38	0.38	0.37	0.38	0.36	0.36	0.38	0.38	0.36
RK436	HAP1																		1.00	0.40	0.41	0.40	0.40	0.39	0.40	0.40	0.38	0.40	0.40	0.39	0.39	0.39	0.39	0.38	0.37	0.39	0.39	0.39	0.36	0.38	0.38	0.37	0.38	0.36	0.37	0.38	0.39	0.36			
	HAP2																			1.00	0.40	0.40	0.41	0.40	0.40	0.39	0.40	0.40	0.39	0.40	0.40	0.39	0.40	0.39	0.39	0.39	0.40	0.39	0.39	0.37	0.38	0.39	0.37	0.37	0.36	0.36	0.38	0.39	0.36		
RK441	HAP1																				1.00	0.41	0.40	0.39	0.41	0.41	0.39	0.39	0.41	0.39	0.40	0.40	0.40	0.39	0.39	0.38	0.39	0.40	0.39	0.41	0.38	0.39	0.39	0.38	0.38	0.36	0.37	0.38	0.38	0.37	
	HAP2																					1.00	0.40	0.39	0.40	0.40	0.39	0.38	0.41	0.40	0.39	0.41	0.40	0.39	0.38	0.39	0.39	0.40	0.39	0.39	0.37	0.40	0.39	0.39	0.39	0.37	0.37	0.38	0.38	0.37	
RK465	HAP1																						1.00	0.47	0.39	0.39	0.39	0.38	0.40	0.39	0.38	0.39	0.40	0.38	0.38	0.37	0.39	0.39	0.38	0.36	0.38	0.38	0.37	0.38	0.36	0.36	0.38	0.38	0.36		
	HAP2																							1.00	0.39	0.40	0.39	0.38	0.41	0.39	0.39	0.39	0.40	0.39	0.39	0.39	0.39	0.39	0.39	0.37	0.39	0.39	0.38	0.38	0.37	0.36	0.38	0.39	0.36		
RK498	HAP1																								1.00	0.41	0.39	0.38	0.40	0.38	0.39	0.40	0.38	0.38	0.39	0.38	0.39	0.38	0.39	0.38	0.39	0.38	0.39	0.38	0.37	0.38	0.36	0.38	0.38	0.37	
	HAP2																									1.00	0.40	0.39	0.40	0.40	0.40	0.40	0.41	0.40	0.39	0.39	0.39	0.39	0.40	0.39	0.39	0.40	0.37	0.40	0.39	0.37	0.39	0.37	0.37	0.38	0.39
RK50	HAP1																									1.00	0.59	0.40	0.38	0.39	0.39	0.38	0.38	0.38	0.41	0.42	0.38	0.38	0.39	0.37	0.38	0.38	0.38	0.36	0.36	0.38	0.39	0.36			



Supplementary Figure 9 Enriched GO (gene ontology) of pangenome core genes, softcore genes, shell genes and private genes generated with TopGO⁸³. The enriched GO categories were determined using Fisher's exact test, followed by the Benjamin–Hochberg correction to obtain adjusted *P* values for multiple testing. See **Supplementary Figure 10**. A full list of the top 20 GO terms is provided in **Supplementary Table 14 - 16**.



Supplementary Figure 10 Enriched GO (gene ontology) of pangenome core genes, softcore genes, shell genes and private genes generated with TopGO⁸³. The enriched GO categories were determined using Fisher's exact test, followed by the Benjamin–Hochberg correction to obtain adjusted P values for multiple testing. See **Supplementary Figure 11**. A full list of the top 20 GO terms is provided in **Supplementary Table 14 - 16**.



Supplementary Figure 11 Enriched GO (gene ontology) of pangenome core genes, softcore genes, shell genes and private genes generated with TopGO⁸³. The enriched GO categories were determined using Fisher's exact test, followed by the Benjamin–Hochberg correction to obtain adjusted *P* values for multiple testing. A full list of the top 20 GO terms is provided in **Supplementary Table 14 - 16**.

Supplementary tables

Supplementary Table 1 Information of 25 *E. grandis* individual used for pangenome analyses.

Sample	Species	Wild or Breeding	Location/Provenance	Region	Subpopulation	Ploidy	Sample_HAP1	Sample_HAP2
RK05	<i>E. grandis</i>	Wild	Windsor Tablelands 1	Windsor	North	2	RK05_HAP1	RK05_HAP2
RK74	<i>E. grandis</i>	Wild	Windsor Tablelands 2	Windsor	North	2	RK74_HAP1	RK74_HAP2
RK17	<i>E. grandis</i>	Wild	Copperlode	Copperlode	North	2	RK17_HAP1	RK17_HAP2
RK50	<i>E. grandis</i>	Wild	Tinaroo Creek Road	Atherton	North	2	RK50_HAP1	RK50_HAP2
RK205	<i>E. grandis</i>	Wild	Mt. Spec SF, Paluma	Paluma	North	2	RK205_HAP1	RK205_HAP2
RK320	<i>E. grandis</i>	Wild	Paluma	Paluma	North	2	RK320_HAP1	RK320_HAP2
RK313	<i>E. grandis</i>	Wild	Eungella	Eungella	Mackay	2	RK313_HAP1	RK313_HAP2
RK465	<i>E. grandis</i>	Wild	Yabba	Sunshine Coast	South	2	RK465_HAP1	RK465_HAP2
RK388	<i>E. grandis</i>	Wild	Mapleton	Sunshine Coast	South	2	RK388_HAP1	RK388_HAP2
RK561	<i>E. grandis</i>	Wild	Connondale	Sunshine Coast	South	2	RK561_HAP1	RK561_HAP2
RK441	<i>E. grandis</i>	Wild	Beerwah SF	Sunshine Coast	South	2	RK441_HAP1	RK441_HAP2
RK436	<i>E. grandis</i>	Wild	Mt. Mee	Sunshine Coast	South	2	RK436_HAP1	RK436_HAP2
RK339	<i>E. grandis</i>	Wild	Mt. Tamborine	Mt. Tamborine & Mt Lindsay	South	2	RK339_HAP1	RK339_HAP2
RK498	<i>E. grandis</i>	Wild	Mt. Lindesay	Mt. Tamborine & Mt Lindsay	South	2	RK498_HAP1	RK498_HAP2
RK792	<i>E. grandis</i>	Wild	Bagawa SF	Coffs Harbour	South	2	RK792_HAP1	RK792_HAP2
RK687	<i>E. grandis</i>	Wild	Wedding Bells SF	Coffs Harbour	South	2	RK687_HAP1	RK687_HAP2
RK614	<i>E. grandis</i>	Wild	Orara West SF (72)	Coffs Harbour	South	2	RK614_HAP1	RK614_HAP2
RK84	<i>E. grandis</i>	Wild	Bulahdelah SF	Bulahdelah	South	2	RK84_HAP1	RK84_HAP2
RK86	<i>E. grandis</i>	Wild	Bulahdelah SF	Bulahdelah	South	2	RK86_HAP1	RK86_HAP2
RK715	<i>E. grandis</i>	Wild	Bulahdelah SF	Bulahdelah	South	2	RK715_HAP1	RK715_HAP2
RK25	<i>E. grandis</i>	Wild	Copperlode	Copperlode	North	2	RK25_HAP1	RK25_HAP2
RM01	<i>E. grandis</i>	5 th generation breeding material	Tzaneen	Mpumalanga	NA	2	RM01_HAP1	RM01_HAP2
RM02	<i>E. grandis</i>	5 th generation breeding material	Tzaneen	Mpumalanga	NA	2	RM02_HAP1	RM02_HAP2
RM03	<i>E. grandis</i>	5 th generation breeding material	Tzaneen	Mpumalanga	NA	2	RM03_HAP1	RM03_HAP2
TAG0014	<i>E. grandis</i>	Clone	Zululand	KwaZulu-Natal	NA	2	TAG0014_HAP1	TAG0014_HAP2

Supplementary Table 2 Summary statistics of sequencing data generated and coverage estimate of generated data for 25 *E. grandis* individuals (PacBio HiFi and Illumina).

Sample	Basepairs	HiFi N50	HiFi coverage (x)	Illumina coverage (x)	Illumina Library
RK05	29,686,885,808	46.39x	16,837	49.2x	KWER
RK17	32,176,769,241	50.28x	18,793	51.9x	KWES
RK205	40,916,043,263	63.93x	16,883	53.8x	KWEX
RK25	29,413,814,731	52.53x	17,117	53.3x	KNMU
RK313	33,266,106,316	51.98x	19,304	51.2x	KWET
RK320	27,825,179,706	43.48x	18,297	25.8x	KLDF
RK339	30,779,259,512	54.96x	17,818	50.9x	KWEY
RK388	31,833,419,975	56.85x	18,194	49.9x	KWEZ
RK436	29,368,472,396	52.45x	16,632	50.7x	KWFA
RK441	31,614,315,806	49.40x	19,531	42.2x	KNMT
RK475	28,356,980,533	44.31x	18,524	51.0x	KWFB
RK498	33,071,905,867	59.05x	18,279	50.5x	KWFC
RK50	31,240,949,662	56.80x	17,799	52.0x	KWEU
RK561	32,587,117,596	58.19x	15,817	49.7x	KWFD
RK614	36,940,563,829	65.97x	16,505	51.3x	KWFH
RK687	31,912,996,676	56.98x	17,218	50.6x	KWFI
RK715	38,175,459,295	68.17x	16,118	49.4x	KWFJ
RK74	28,173,678,993	44.02x	18,169	54.9x	KWEW
RK792	22,634,181,187	40.42x	16,001	51.9x	KWFK
RK84	22,661,764,209	40.47x	14,830	48.0x	KWFL
RK86	36,801,598,729	65.71x	17,718	51.4x	KWFM
RM01	32,545,589,418	50.85x	18,060	53.8x	KLCA
RM02	33,601,800,650	52.50x	17,826	56.2x	KLCB
RM03	31,391,717,467	49.05x	17,001	55.6x	KLCC
TAG0014	41,920,934,055	65.50x	17,743	60.4x	KCUX

Supplementary Table 3 Number of contigs per chromosome for chromosome-level haplotype-resolved genome for 25 *E. grandis* individuals (50 haplotypes).

Sample	Chr01	Chr02	Chr03	Chr04	Chr05	Chr06	Chr07	Chr08	Chr09	Chr10	Chr11	Remaining
RK05_HAP1	3	5	9	5	4	6	6	7	4	3	1	1
RK05_HAP2	4	5	6	8	11	7	6	4	2	4	6	0
RK17_HAP1	3	4	6	3	1	3	2	3	1	1	7	0
RK17_HAP2	3	3	4	2	3	4	4	3	4	1	6	0
RK205_HAP1	4	5	2	1	2	1	3	2	2	2	4	1
RK205_HAP2	3	2	3	1	4	1	3	2	2	2	7	3
RK25_HAP1	5	7	7	5	6	12	5	4	3	3	4	0
RK25_HAP2	6	7	7	3	10	8	8	2	3	3	8	0
RK313_HAP1	5	7	3	6	5	2	5	7	1	4	5	2
RK313_HAP2	3	5	7	3	2	5	5	5	3	3	2	0
RK320_HAP1	8	7	7	10	9	12	10	9	9	5	6	0
RK320_HAP2	10	12	4	10	10	9	17	7	5	6	6	0
RK339_HAP1	6	4	6	2	5	6	6	5	4	2	6	10
RK339_HAP2	6	5	7	4	4	3	3	5	7	3	3	4
RK388_HAP1	3	5	4	2	6	2	4	6	5	2	2	3
RK388_HAP2	4	6	5	2	5	5	4	5	3	2	2	3
RK436_HAP1	2	3	4	5	8	7	2	6	1	2	5	1
RK436_HAP2	4	6	6	4	9	6	5	4	5	2	3	3
RK441_HAP1	9	11	5	5	7	11	9	4	6	18	2	0
RK441_HAP2	11	7	7	4	8	10	15	4	9	3	4	0
RK465_HAP1	3	7	7	2	6	3	3	7	3	3	2	11
RK465_HAP2	6	5	2	2	4	7	4	3	3	1	4	0
RK498_HAP1	10	4	5	2	5	5	2	7	9	2	4	2
RK498_HAP2	4	3	6	3	6	6	8	6	3	3	4	0
RK50_HAP1	5	4	7	1	2	5	1	3	7	1	6	1
RK50_HAP2	7	5	7	1	3	5	2	5	2	2	5	0
RK561_HAP1	8	5	4	2	5	8	3	7	1	4	3	4
RK561_HAP2	3	7	3	2	3	3	2	7	3	4	4	1
RK614_HAP1	4	2	4	2	4	3	4	2	2	7	2	8
RK614_HAP2	2	4	3	4	7	2	1	1	1	4	2	2

RK687_HAP1	3	4	7	3	4	8	3	5	6	5	2	6
RK687_HAP2	5	2	9	2	6	4	6	4	4	4	2	1
RK715_HAP1	4	4	3	2	2	5	5	3	1	2	2	2
RK715_HAP2	2	8	3	3	3	3	4	4	3	3	2	4
RK74_HAP1	2	1	4	3	6	6	7	10	1	8	2	1
RK74_HAP2	2	4	6	3	6	7	8	10	2	7	4	0
RK792_HAP1	7	10	4	5	6	8	6	13	5	3	7	5
RK792_HAP2	3	9	4	9	5	10	4	5	8	1	7	3
RK84_HAP1	7	10	7	11	10	7	13	6	7	5	7	2
RK84_HAP2	8	10	8	8	8	10	10	4	5	7	9	7
RK86_HAP1	3	5	7	1	6	5	5	3	2	2	4	3
RK86_HAP2	3	4	2	5	6	3	3	3	2	3	3	3
RM01_HAP1	3	9	6	4	6	6	8	5	5	2	4	0
RM01_HAP2	4	10	7	3	7	5	7	3	5	3	3	0
RM02_HAP1	6	10	8	7	10	12	11	5	5	4	2	1
RM02_HAP2	4	8	7	5	14	11	11	6	4	7	3	2
RM03_HAP1	6	5	4	3	5	9	8	4	2	2	4	0
RM03_HAP2	5	9	4	3	6	5	9	3	7	1	2	0
TAG0014_HAP1	5	6	3	3	2	4	5	3	2	1	2	1
TAG0014_HAP2	4	4	4	2	2	9	4	3	6	3	5	0

Supplementary Table 4 Chromosome size summary for chromosome-level haplotype-resolved genome for 25 *E. grandis* individuals (50 haplotypes).

Sample	Chr01	Chr02	Chr03	Chr04	Chr05	Chr06	Chr07	Chr08	Chr09	Chr10	Chr11	Remaining
RK05_HAP1	43,866,089	53,500,689	61,345,732	39,824,991	63,151,688	54,903,440	54,964,677	70,455,933	37,445,668	38,004,703	41,056,859	134,141
RK05_HAP2	41,339,368	51,953,843	61,255,524	36,870,401	59,647,924	54,615,551	54,692,296	65,658,572	35,710,736	37,285,060	40,007,234	0
RK17_HAP1	44,528,443	56,586,569	64,253,392	36,392,278	61,339,073	55,195,082	53,153,924	69,099,827	36,274,228	38,250,955	42,031,621	0
RK17_HAP2	42,091,882	51,505,262	62,339,297	36,218,596	60,971,976	43,699,440	52,874,303	67,574,077	35,795,011	37,361,719	40,565,831	0
RK205_HAP1	44,706,831	52,320,148	61,104,218	37,335,171	62,357,744	51,905,859	55,642,653	65,565,504	37,090,506	36,290,475	44,229,822	157,611
RK205_HAP2	43,250,146	52,085,732	60,249,612	37,023,274	58,461,093	51,729,826	53,665,662	63,736,490	36,992,321	35,926,977	44,218,868	694,797
RK25_HAP1	43,625,907	56,309,016	64,437,014	38,887,816	64,124,751	53,871,911	52,059,226	66,237,456	37,516,938	37,170,977	42,914,045	0
RK25_HAP2	42,828,743	53,022,725	63,275,617	36,143,060	64,002,783	53,138,552	50,251,668	64,193,395	36,669,192	36,783,366	40,227,089	0
RK313_HAP1	42,427,812	56,749,372	63,555,106	37,168,734	61,333,411	53,645,336	52,529,844	66,350,653	37,457,035	36,156,482	42,082,804	157,484
RK313_HAP2	40,387,023	55,640,127	60,180,406	35,613,549	59,567,991	52,031,669	52,435,972	65,412,028	37,300,333	35,984,326	40,330,269	0
RK320_HAP1	42,655,642	55,179,924	63,804,716	37,784,216	66,701,800	56,443,680	59,010,732	67,901,260	36,735,297	37,794,985	42,671,148	0
RK320_HAP2	42,036,757	52,209,244	60,139,461	37,635,197	60,341,668	54,483,390	52,550,186	65,606,249	34,661,804	37,722,949	42,138,259	0
RK339_HAP1	46,284,116	56,204,986	64,566,983	38,543,953	64,569,628	54,059,875	57,082,065	66,912,716	36,246,714	38,327,444	42,413,599	832,342
RK339_HAP2	44,926,421	54,972,102	63,125,025	37,658,370	61,406,734	49,573,818	54,447,884	65,108,946	36,026,655	38,225,425	42,005,730	584,848
RK388_HAP1	44,099,135	56,643,645	64,357,802	37,717,451	61,940,644	53,997,193	57,348,516	69,299,168	38,645,411	39,622,945	41,601,013	836,550
RK388_HAP2	43,120,589	53,195,363	60,522,694	37,492,715	61,859,287	52,080,280	54,046,376	65,361,277	37,158,914	39,141,059	40,601,482	264,832
RK436_HAP1	44,341,113	53,287,294	63,698,176	38,638,708	61,930,973	54,551,032	52,787,499	68,585,703	38,265,047	37,406,958	42,169,830	51,152
RK436_HAP2	41,774,203	52,552,369	62,082,487	36,737,731	60,867,942	53,550,600	52,563,555	67,910,951	36,206,542	37,274,541	41,150,291	843,189
RK441_HAP1	47,232,735	53,897,521	65,472,410	39,080,873	65,225,352	52,197,800	57,771,723	66,200,872	38,830,619	40,163,072	42,320,276	0
RK441_HAP2	44,799,151	53,050,136	64,771,165	38,622,671	64,179,070	51,890,312	54,745,898	64,521,408	36,459,498	39,456,764	42,195,659	0
RK465_HAP1	44,422,802	55,967,221	61,245,816	40,213,661	66,015,057	51,533,884	54,937,797	68,835,478	35,499,801	38,134,613	43,083,496	1,002,650
RK465_HAP2	41,980,019	54,677,699	60,681,737	38,286,129	63,337,048	49,696,222	54,660,869	66,048,480	35,268,155	37,338,873	41,842,173	0
RK498_HAP1	45,710,569	57,398,734	64,890,314	38,335,895	63,859,933	54,308,108	56,925,077	64,782,004	41,306,423	39,474,803	44,504,562	174,380
RK498_HAP2	42,907,096	53,668,391	61,837,958	35,978,062	62,457,435	53,401,891	55,873,937	64,780,359	37,529,256	36,516,540	41,922,792	0
RK50_HAP1	45,083,441	55,029,939	64,060,506	39,017,254	63,655,740	50,838,626	56,251,655	69,183,342	38,024,202	38,674,504	40,715,998	52,691
RK50_HAP2	44,799,520	42,595,211	60,870,217	38,893,577	63,413,676	62,257,908	51,589,665	65,801,295	36,605,471	37,608,269	40,406,328	0
RK561_HAP1	43,375,424	53,281,305	61,737,639	39,863,464	65,993,068	55,155,423	59,156,520	66,213,489	37,084,713	38,656,704	41,321,370	662,902
RK561_HAP2	42,428,611	52,072,121	58,777,900	37,263,409	64,968,785	51,524,833	53,739,507	65,871,122	37,000,456	36,743,790	40,542,987	131,853
RK614_HAP1	43,799,565	54,183,546	64,944,794	37,637,542	69,395,941	52,852,529	55,462,265	67,489,081	37,797,481	38,694,986	42,116,098	595,006
RK614_HAP2	43,244,885	54,142,015	64,715,093	37,304,167	64,156,493	51,793,772	53,368,176	66,834,882	36,900,197	37,753,463	41,455,576	521,918

RK687_HAP1	44,120,348	56,157,182	60,664,184	39,111,456	65,085,224	54,727,240	59,163,075	67,881,968	38,051,326	37,591,587	42,729,794	459,872
RK687_HAP2	41,125,186	52,852,485	60,277,077	37,929,977	61,895,300	53,835,574	56,461,750	67,490,756	37,278,883	36,417,166	42,377,241	83,908
RK715_HAP1	44,447,728	56,489,503	62,817,728	37,222,916	63,583,724	54,236,900	54,742,871	68,289,221	39,129,192	39,906,652	42,763,864	263,179
RK715_HAP2	42,672,986	53,460,348	61,734,588	36,961,679	60,049,824	50,797,531	52,439,094	66,324,107	38,498,055	37,441,818	41,518,534	356,768
RK74_HAP1	44,303,918	51,539,596	60,168,946	37,684,470	69,042,769	54,696,583	64,955,389	65,478,445	36,153,783	37,090,599	42,320,556	67,528
RK74_HAP2	44,268,684	50,883,936	58,385,604	37,276,382	65,924,785	53,977,466	59,313,962	64,716,287	35,315,762	37,020,161	40,306,132	0
RK792_HAP1	43,555,557	57,725,610	59,980,300	37,856,202	66,846,749	53,435,748	53,675,113	67,552,897	39,393,832	37,625,414	43,704,646	1,060,451
RK792_HAP2	42,873,820	55,916,018	58,454,678	36,292,183	64,847,326	52,383,154	51,479,072	65,490,318	38,018,702	37,596,206	41,480,074	729,387
RK84_HAP1	43,809,237	54,620,545	62,519,017	37,280,040	63,392,958	54,086,920	57,065,294	67,175,057	38,942,592	40,392,786	42,849,567	134,788
RK84_HAP2	42,183,567	54,313,465	58,768,418	36,451,793	59,140,906	52,633,554	53,781,581	66,894,008	37,982,808	38,806,516	41,359,996	698,658
RK86_HAP1	42,751,234	54,090,820	62,443,520	36,610,501	66,060,702	55,019,907	53,947,750	73,906,533	39,475,831	39,154,109	41,215,685	226,785
RK86_HAP2	40,324,993	51,527,914	61,436,118	35,633,620	60,342,669	51,277,359	53,567,773	66,454,493	36,943,090	38,248,240	40,950,774	371,318
RM01_HAP1	43,450,313	57,950,322	63,273,074	38,918,006	67,290,617	55,392,032	62,094,544	66,464,933	36,897,772	38,342,412	42,701,731	0
RM01_HAP2	42,210,311	54,756,610	59,748,321	37,798,566	65,254,885	54,145,123	61,482,995	66,231,216	35,878,365	37,974,976	41,144,251	0
RM02_HAP1	43,491,124	57,491,493	62,718,826	37,714,378	66,206,107	54,990,919	62,421,783	70,367,016	37,803,974	38,555,327	42,182,665	553,391
RM02_HAP2	41,509,460	55,531,415	60,982,529	37,128,994	62,624,302	54,978,083	60,966,028	66,156,572	36,617,188	38,120,651	41,871,895	288,513
RM03_HAP1	44,013,246	55,621,320	62,413,612	37,017,745	61,594,215	54,740,874	54,241,530	68,087,383	37,778,014	38,153,764	44,210,825	0
RM03_HAP2	43,381,688	52,318,792	61,015,965	36,005,048	60,154,630	53,408,080	53,497,468	65,913,137	36,964,352	37,282,011	42,870,264	0
TAG0014_HAP1	44,764,794	57,174,188	64,190,686	37,932,624	67,041,112	55,026,902	58,892,594	68,844,007	36,944,981	38,023,239	41,944,707	50,505
TAG0014_HAP2	41,459,349	56,979,195	61,673,418	36,908,141	64,932,641	53,665,386	55,768,775	66,164,710	36,522,489	37,128,649	41,205,020	0

Supplementary Table 5 BUSCO genome completeness statistics per haplotype-resolved assembly.

Sample	Complete BUSCOs (C)	Complete and single-copy BUSCOs (S)	Complete and duplicated BUSCOs (D)	Fragmented BUSCOs (F)	Missing BUSCOs (M)	Complete BUSCOs (C)	Complete and single-copy BUSCOs (S)	Complete and duplicated BUSCOs (D)	Fragmented BUSCOs (F)	Missing BUSCOs (M)
TAG0014_HAP1	1,589	1,538	51	16	9	98.50%	95.30%	3.20%	1.00%	0.60%
TAG0014_HAP2	1,588	1,547	41	18	8	98.30%	95.80%	2.50%	1.10%	0.60%
RM01_HAP1	1,586	1,543	43	21	7	98.30%	95.60%	2.70%	1.30%	0.40%
RM01_HAP2	1,589	1,545	44	17	8	98.40%	95.70%	2.70%	1.10%	0.50%
RM02_HAP1	1,589	1,550	39	18	7	98.40%	96.00%	2.40%	1.10%	0.50%
RM02_HAP2	1,586	1,545	41	21	7	98.20%	95.70%	2.50%	1.30%	0.50%
RM03_HAP1	1,585	1,545	40	20	9	98.20%	95.70%	2.50%	1.20%	0.60%
RM03_HAP2	1,587	1,541	46	21	6	98.40%	95.50%	2.90%	1.30%	0.30%
RK05_HAP1	1,587	1,546	41	18	9	98.30%	95.80%	2.50%	1.10%	0.60%
RK05_HAP2	1,585	1,537	48	18	11	98.20%	95.20%	3.00%	1.10%	0.70%
RK74_HAP1	1,585	1,548	37	18	11	98.20%	95.90%	2.30%	1.10%	0.70%
RK74_HAP2	1,584	1,544	40	18	12	98.10%	95.70%	2.50%	1.10%	0.70%
RK17_HAP1	1,586	1,539	47	20	8	98.30%	95.40%	2.90%	1.20%	0.50%
RK17_HAP2	1,564	1,520	44	20	30	96.90%	94.20%	2.70%	1.20%	1.90%
RK25_HAP1	1,591	1,548	43	16	7	98.60%	95.90%	2.70%	1.00%	0.40%
RK25_HAP2	1,586	1,544	42	20	8	98.30%	95.70%	2.60%	1.20%	0.50%
RK50_HAP1	1,588	1,551	37	18	8	98.40%	96.10%	2.30%	1.10%	0.50%
RK50_HAP2	1,588	1,548	40	18	8	98.40%	95.90%	2.50%	1.10%	0.50%
RK205_HAP1	1,589	1,546	43	18	7	98.50%	95.80%	2.70%	1.10%	0.40%
RK205_HAP2	1,590	1,551	39	17	7	98.50%	96.10%	2.40%	1.10%	0.40%
RK313_HAP1	1,587	1,547	40	18	9	98.30%	95.80%	2.50%	1.10%	0.60%
RK313_HAP2	1,587	1,547	40	17	10	98.30%	95.80%	2.50%	1.10%	0.60%
RK320_HAP1	1,589	1,544	45	17	8	98.50%	95.70%	2.80%	1.10%	0.40%

RK320_HAP2	1,590	1,544	46	16	8	98.60%	95.70%	2.90%	1.00%	0.40%
RK339_HAP1	1,590	1,545	45	15	9	98.50%	95.70%	2.80%	0.90%	0.60%
RK339_HAP2	1,586	1,548	38	19	9	98.30%	95.90%	2.40%	1.20%	0.50%
RK388_HAP1	1,587	1,544	43	20	7	98.40%	95.70%	2.70%	1.20%	0.40%
RK388_HAP2	1,587	1,542	45	19	8	98.30%	95.50%	2.80%	1.20%	0.50%
RK436_HAP1	1,589	1,545	44	19	6	98.40%	95.70%	2.70%	1.20%	0.40%
RK436_HAP2	1,591	1,548	43	16	7	98.60%	95.90%	2.70%	1.00%	0.40%
RK441_HAP1	1,585	1,547	38	21	8	98.20%	95.80%	2.40%	1.30%	0.50%
RK441_HAP2	1,591	1,545	46	17	6	98.60%	95.70%	2.90%	1.10%	0.30%
RK465_HAP1	1,590	1,547	43	16	8	98.50%	95.80%	2.70%	1.00%	0.50%
RK465_HAP2	1,584	1,541	43	21	9	98.20%	95.50%	2.70%	1.30%	0.50%
RK498_HAP1	1,585	1,542	43	20	9	98.20%	95.50%	2.70%	1.20%	0.60%
RK498_HAP2	1,588	1,546	42	18	8	98.40%	95.80%	2.60%	1.10%	0.50%
RK561_HAP1	1,584	1,539	45	21	9	98.20%	95.40%	2.80%	1.30%	0.50%
RK561_HAP2	1,586	1,546	40	20	8	98.30%	95.80%	2.50%	1.20%	0.50%
RK792_HAP1	1,587	1,548	39	18	9	98.30%	95.90%	2.40%	1.10%	0.60%
RK792_HAP2	1,586	1,541	45	20	8	98.30%	95.50%	2.80%	1.20%	0.50%
RK687_HAP1	1,587	1,540	47	20	7	98.30%	95.40%	2.90%	1.20%	0.50%
RK687_HAP2	1,587	1,546	41	20	7	98.30%	95.80%	2.50%	1.20%	0.50%
RK614_HAP1	1,586	1,542	44	18	10	98.20%	95.50%	2.70%	1.10%	0.70%
RK614_HAP2	1,589	1,543	46	17	8	98.50%	95.60%	2.90%	1.10%	0.40%
RK84_HAP1	1,587	1,541	46	17	10	98.40%	95.50%	2.90%	1.10%	0.50%
RK84_HAP2	1,591	1,553	38	16	7	98.60%	96.20%	2.40%	1.00%	0.40%
RK86_HAP1	1,587	1,541	46	19	8	98.40%	95.50%	2.90%	1.20%	0.40%
RK86_HAP2	1,587	1,542	45	18	9	98.30%	95.50%	2.80%	1.10%	0.60%
RK715_HAP1	1,585	1,544	41	22	7	98.20%	95.70%	2.50%	1.40%	0.40%
RK715_HAP2	1,586	1,541	45	18	10	98.30%	95.50%	2.80%	1.10%	0.60%

Supplementary Table 6 BUSCO gene completeness statistics per haplotype-resolved annotation.

Sample	Complete BUSCOs (C)	Complete and single-copy BUSCOs (S)	Complete and duplicated BUSCOs (D)	Fragmented BUSCOs (F)	Missing BUSCOs (M)	Complete BUSCOs (C)	Complete and single-copy BUSCOs (S)	Complete and duplicated BUSCOs (D)	Fragmented BUSCOs (F)	Missing BUSCOs (M)
RK05_HAP1	1,609	1,558	51	4	1	99.70%	96.50%	3.20%	0.20%	0.10%
RK05_HAP2	1,607	1,553	54	4	3	99.50%	96.20%	3.30%	0.20%	0.30%
RK74_HAP1	1,610	1,563	47	3	1	99.70%	96.80%	2.90%	0.20%	0.10%
RK74_HAP2	1,606	1,557	49	3	5	99.50%	96.50%	3.00%	0.20%	0.30%
RK17_HAP1	1,608	1,556	52	5	1	99.60%	96.40%	3.20%	0.30%	0.10%
RK17_HAP2	1,588	1,536	52	6	20	98.40%	95.20%	3.20%	0.40%	1.20%
RK25_HAP1	1,609	1,556	53	4	1	99.70%	96.40%	3.30%	0.20%	0.10%
RK25_HAP2	1,608	1,552	56	5	1	99.70%	96.20%	3.50%	0.30%	0.00%
RK50_HAP1	1,604	1,551	53	7	3	99.40%	96.10%	3.30%	0.40%	0.20%
RK50_HAP2	1,605	1,555	50	8	1	99.40%	96.30%	3.10%	0.50%	0.10%
RK205_HAP1	1,610	1,558	52	3	1	99.70%	96.50%	3.20%	0.20%	0.10%
RK205_HAP2	1,610	1,562	48	3	1	99.80%	96.80%	3.00%	0.20%	0.00%
RK320_HAP1	1,608	1,551	57	4	2	99.60%	96.10%	3.50%	0.20%	0.20%
RK320_HAP2	1,609	1,554	55	4	1	99.70%	96.30%	3.40%	0.20%	0.10%
RK313_HAP1	1,610	1,555	55	3	1	99.70%	96.30%	3.40%	0.20%	0.10%
RK313_HAP2	1,608	1,560	48	3	3	99.70%	96.70%	3.00%	0.20%	0.10%
RK465_HAP1	1,607	1,561	46	6	1	99.60%	96.70%	2.90%	0.40%	0.00%
RK465_HAP2	1,607	1,554	53	5	2	99.60%	96.30%	3.30%	0.30%	0.10%
RK388_HAP1	1,611	1,557	54	2	1	99.80%	96.50%	3.30%	0.10%	0.10%
RK388_HAP2	1,610	1,555	55	3	1	99.70%	96.30%	3.40%	0.20%	0.10%
RK561_HAP1	1,606	1,557	49	6	2	99.50%	96.50%	3.00%	0.40%	0.10%
RK561_HAP2	1,609	1,562	47	4	1	99.70%	96.80%	2.90%	0.20%	0.10%
RK441_HAP1	1,609	1,560	49	4	1	99.70%	96.70%	3.00%	0.20%	0.10%

RK441_HAP2	1,609	1,555	54	4	1	99.60%	96.30%	3.30%	0.20%	0.20%
RK436_HAP1	1,607	1,556	51	6	1	99.60%	96.40%	3.20%	0.40%	0.00%
RK436_HAP2	1,609	1,557	52	4	1	99.70%	96.50%	3.20%	0.20%	0.10%
RK339_HAP1	1,609	1,551	58	4	1	99.70%	96.10%	3.60%	0.20%	0.10%
RK339_HAP2	1,608	1,552	56	4	2	99.70%	96.20%	3.50%	0.20%	0.10%
RK498_HAP1	1,608	1,553	55	4	2	99.60%	96.20%	3.40%	0.20%	0.20%
RK498_HAP2	1,606	1,554	52	5	3	99.50%	96.30%	3.20%	0.30%	0.20%
RK792_HAP1	1,609	1,553	56	4	1	99.70%	96.20%	3.50%	0.20%	0.10%
RK792_HAP2	1,609	1,557	52	4	1	99.70%	96.50%	3.20%	0.20%	0.10%
RK687_HAP1	1,610	1,559	51	3	1	99.80%	96.60%	3.20%	0.20%	0.00%
RK687_HAP2	1,608	1,554	54	5	1	99.60%	96.30%	3.30%	0.30%	0.10%
RK614_HAP1	1,607	1,552	55	4	3	99.60%	96.20%	3.40%	0.20%	0.20%
RK614_HAP2	1,610	1,557	53	3	1	99.80%	96.50%	3.30%	0.20%	0.00%
RK84_HAP1	1,607	1,552	55	4	3	99.60%	96.20%	3.40%	0.20%	0.20%
RK84_HAP2	1,608	1,560	48	4	2	99.70%	96.70%	3.00%	0.20%	0.10%
RK86_HAP1	1,610	1,562	48	3	1	99.80%	96.80%	3.00%	0.20%	0.00%
RK86_HAP2	1,610	1,555	55	3	1	99.70%	96.30%	3.40%	0.20%	0.10%
RK715_HAP1	1,608	1,559	49	4	2	99.60%	96.60%	3.00%	0.20%	0.20%
RK715_HAP2	1,608	1,551	57	5	1	99.60%	96.10%	3.50%	0.30%	0.10%
RM01_HAP1	1,609	1,555	54	4	1	99.60%	96.30%	3.30%	0.20%	0.20%
RM01_HAP2	1,610	1,558	52	3	1	99.70%	96.50%	3.30%	0.20%	0.10%
RM02_HAP1	1,607	1,553	54	5	2	99.50%	96.20%	3.30%	0.30%	0.20%
RM02_HAP2	1,608	1,553	55	4	2	99.60%	96.20%	3.40%	0.20%	0.20%
RM03_HAP1	1,609	1,558	51	3	2	99.70%	96.50%	3.20%	0.20%	0.10%
RM03_HAP2	1,609	1,550	59	3	2	99.70%	96.00%	3.70%	0.20%	0.10%
TAG0014_HAP1	1,607	1,549	58	5	2	99.60%	96.00%	3.70%	0.30%	0.10%
TAG0014_HAP2	1,607	1,561	46	5	2	99.60%	96.70%	2.90%	0.30%	0.10%

Supplementary Table 7 Summary statistics of RE content for chromosome-level haplotype-resolved genome for 25 *E. grandis* accessions (50 haplotypes). A full list is provided in the supplementary excel sheet.

Sample	Type	DNA MULE- MuDR	DNA PIF- Harbinger	DNA hAT-Ac	LINE L1	LTR Caulimovirus	LTR Copia	LTR Gypsy	Low complexity	Simple repeat	Unknown	Total
RM01_HAP1	Length (bp)	1,718,832	859,748	802,375	10,215,668	3,750,574	51,775,637	20,629,796	1,509,181	6,952,761	154,715,381	254,521,165
	Percentage (%)	0.3	0.15	0.14	1.78	0.65	9.04	3.6	0.26	1.21	27.01	44.43
RM01_HAP2	Length (bp)	1,694,542	876,008	751,952	9,784,052	3,614,770	49,940,343	19,862,919	1,463,952	6,847,268	150,895,654	247,318,970
	Percentage (%)	0.3	0.16	0.14	1.76	0.65	8.97	3.57	0.26	1.23	27.11	44.45
RM02_HAP1	Length (bp)	1,795,106	870,906	839,026	10,080,331	3,685,750	52,250,250	20,992,838	1,622,375	6,941,278	152,953,089	253,907,490
	Percentage (%)	0.31	0.15	0.15	1.75	0.64	9.09	3.65	0.28	1.21	26.62	44.19
RM02_HAP2	Length (bp)	1,683,262	858,371	798,168	9,893,249	3,790,761	51,268,060	19,554,130	1,581,095	6,803,778	147,449,243	245,290,363
	Percentage (%)	0.3	0.15	0.14	1.78	0.68	9.21	3.51	0.28	1.22	26.48	44.06
RM03_HAP1	Length (bp)	1,760,339	885,698	778,440	10,139,874	3,604,582	50,810,716	19,464,066	1,789,128	6,861,959	146,566,846	244,408,018
	Percentage (%)	0.32	0.16	0.14	1.82	0.65	9.11	3.49	0.32	1.23	26.27	43.85
RM03_HAP2	Length (bp)	1,571,887	860,340	761,733	9,651,344	3,575,190	49,753,669	19,045,995	1,447,048	6,788,118	142,893,482	237,970,799
	Percentage (%)	0.29	0.16	0.14	1.78	0.66	9.17	3.51	0.27	1.25	26.32	43.86
RK05_HAP1	Length (bp)	1,709,133	857,329	713,669	10,045,531	3,632,299	50,425,635	19,501,865	1,420,071	6,945,374	149,875,375	246,751,803
	Percentage (%)	0.31	0.15	0.13	1.8	0.65	9.03	3.49	0.25	1.24	26.83	44.18
RK05_HAP2	Length (bp)	1,606,455	891,602	738,719	9,866,554	3,607,126	47,412,046	18,421,660	1,423,863	6,720,294	141,771,652	234,293,362
	Percentage (%)	0.3	0.17	0.14	1.83	0.67	8.8	3.42	0.26	1.25	26.3	43.49
RK74_HAP1	Length (bp)	1,727,492	900,652	840,794	10,023,934	3,680,338	50,098,390	19,905,295	1,757,180	6,874,264	148,497,960	246,114,202
	Percentage (%)	0.31	0.16	0.15	1.78	0.65	8.89	3.53	0.31	1.22	26.35	43.67
RK74_HAP2	Length (bp)	1,624,454	893,848	811,104	9,889,163	3,534,949	47,637,762	19,275,633	1,523,356	6,721,542	144,781,042	238,385,221
	Percentage (%)	0.3	0.16	0.15	1.81	0.65	8.7	3.52	0.28	1.23	26.45	43.57
RK17_HAP1	Length (bp)	1,754,653	874,934	784,406	10,166,416	3,780,394	50,148,668	20,066,127	1,464,522	6,869,974	148,278,196	245,775,911
	Percentage (%)	0.31	0.16	0.14	1.82	0.68	9	3.6	0.26	1.23	26.62	44.1
RK17_HAP2	Length (bp)	1,679,533	853,526	713,910	9,742,006	3,747,041	46,814,578	18,197,254	1,398,795	6,645,598	139,135,191	230,555,939
	Percentage (%)	0.32	0.16	0.13	1.83	0.71	8.82	3.43	0.26	1.25	26.2	43.42
RK50_HAP1	Length (bp)	1,860,683	897,015	763,869	10,050,111	3,749,779	50,805,936	20,681,173	1,514,712	6,899,477	149,728,593	248,846,498
	Percentage (%)	0.32	0.16	0.13	1.83	0.71	8.82	3.43	0.26	1.25	26.2	43.42
RK50_HAP2	Length (bp)	1,651,017	882,888	748,078	9,927,227	3,560,813	49,104,563	19,295,335	1,527,801	6,698,728	143,813,827	238,881,762
	Percentage (%)	0.33	0.16	0.14	1.79	0.67	9.06	3.69	0.27	1.23	26.71	44.38

RK205_HAP1	Length (bp)	1,745,856	873,919	741,521	9,872,763	3,666,088	49,207,886	19,317,543	1,446,745	6,643,300	145,280,700	240,412,151
	Percentage (%)	0.32	0.16	0.14	1.8	0.67	8.97	3.52	0.26	1.21	26.48	43.85
RK205_HAP2	Length (bp)	1,747,898	862,391	720,605	9,756,396	3,485,116	49,014,532	18,279,832	1,427,157	6,618,657	139,977,217	233,458,913
	Percentage (%)	0.32	0.16	0.13	1.81	0.65	9.11	3.4	0.27	1.23	26.02	43.42
RK320_HAP1	Length (bp)	1,878,736	887,599	778,617	10,069,031	3,836,215	49,953,103	20,170,742	1,553,526	6,888,084	152,281,219	250,087,721
	Percentage (%)	0.33	0.16	0.14	1.78	0.68	8.81	3.56	0.27	1.22	26.87	44.14
RK320_HAP2	Length (bp)	1,749,048	845,800	805,603	9,919,742	3,561,156	48,211,239	18,572,055	1,421,012	6,658,091	142,232,303	235,715,071
	Percentage (%)	0.32	0.16	0.15	1.84	0.66	8.94	3.44	0.26	1.23	26.36	43.7
RK313_HAP1	Length (bp)	1,593,754	925,125	785,459	9,869,127	3,780,539	49,368,990	18,796,460	1,581,931	6,638,598	144,329,898	239,333,544
	Percentage (%)	0.29	0.17	0.14	1.8	0.69	8.98	3.42	0.29	1.21	26.26	43.56
RK313_HAP2	Length (bp)	1,591,726	864,340	805,811	9,471,948	3,537,161	47,777,383	18,208,903	1,418,281	6,549,323	139,000,530	230,753,770
	Percentage (%)	0.3	0.16	0.15	1.77	0.66	8.93	3.4	0.27	1.22	25.99	43.14
RK465_HAP1	Length (bp)	1,667,418	848,690	761,320	9,949,162	3,441,910	50,874,968	19,549,367	1,482,947	6,841,244	149,243,153	246,434,125
	Percentage (%)	0.3	0.15	0.14	1.77	0.61	9.07	3.49	0.26	1.22	26.61	43.94
RK465_HAP2	Length (bp)	1,673,485	857,947	779,932	9,880,838	3,726,321	49,114,106	18,296,611	1,483,319	6,723,994	140,818,791	234,924,361
	Percentage (%)	0.31	0.16	0.14	1.82	0.69	9.03	3.36	0.27	1.24	25.89	43.22
RK388_HAP1	Length (bp)	1,919,456	858,129	784,776	10,172,844	3,697,793	52,297,483	20,476,104	1,605,457	6,928,661	151,850,569	252,304,801
	Percentage (%)	0.34	0.15	0.14	1.8	0.65	9.24	3.62	0.28	1.22	26.82	44.59
RK388_HAP2	Length (bp)	1,735,325	924,432	768,206	9,719,924	3,609,988	49,747,108	19,037,187	1,502,577	6,693,130	142,926,151	238,391,900
	Percentage (%)	0.32	0.17	0.14	1.78	0.66	9.13	3.49	0.28	1.23	26.23	43.76
RK561_HAP1	Length (bp)	1,600,717	870,486	696,077	10,057,594	3,522,414	51,054,655	19,991,197	1,458,176	6,734,708	152,655,922	250,384,229
	Percentage (%)	0.28	0.15	0.12	1.79	0.63	9.08	3.55	0.26	1.2	27.14	44.51
RK561_HAP2	Length (bp)	1,618,423	854,128	720,213	9,535,233	3,467,696	50,245,548	18,351,995	1,434,097	6,596,783	141,348,286	236,052,363
	Percentage (%)	0.3	0.16	0.13	1.76	0.64	9.29	3.39	0.27	1.22	26.12	43.64
RK441_HAP1	Length (bp)	1,629,888	891,166	709,831	10,084,061	4,045,884	52,238,171	21,802,585	1,457,800	6,913,691	154,816,026	256,315,296
	Percentage (%)	0.29	0.16	0.12	1.77	0.71	9.19	3.84	0.26	1.22	27.24	45.11
RK441_HAP2	Length (bp)	1,786,606	904,967	768,741	10,101,193	3,656,959	51,441,355	19,325,127	1,541,778	6,875,665	147,099,157	245,073,273
	Percentage (%)	0.32	0.16	0.14	1.82	0.66	9.27	3.48	0.28	1.24	26.52	44.19
RK436_HAP1	Length (bp)	1,707,230	874,155	738,172	10,078,040	3,791,921	51,315,937	19,896,911	1,587,568	6,783,871	147,116,609	245,611,939
	Percentage (%)	0.31	0.16	0.13	1.81	0.68	9.23	3.58	0.29	1.22	26.47	44.21
RK436_HAP2	Length (bp)	1,491,207	826,494	781,828	9,561,501	3,639,239	49,332,460	18,324,321	1,432,253	6,752,872	140,975,851	235,078,212
	Percentage (%)	0.27	0.15	0.14	1.76	0.67	9.08	3.37	0.26	1.24	25.94	43.27
RK339_HAP1	Length (bp)	1,896,885	915,011	773,403	10,038,257	3,789,311	50,827,862	20,552,145	1,602,861	6,837,272	151,435,273	250,317,475

RK339_HAP2	Percentage (%)	0.34	0.16	0.14	1.77	0.67	8.98	3.63	0.28	1.21	26.75	44.25
	Length (bp)	1,590,576	876,919	735,382	10,194,821	3,525,643	49,246,962	19,035,376	1,448,021	6,684,905	145,279,861	240,416,934
RK498_HAP1	Percentage (%)	0.29	0.16	0.13	1.86	0.64	8.99	3.47	0.26	1.22	26.51	43.88
	Length (bp)	1,743,494	838,553	819,411	10,170,385	3,845,153	53,438,312	21,033,481	1,503,869	6,963,550	154,738,461	256,914,255
RK498_HAP2	Percentage (%)	0.3	0.15	0.14	1.78	0.67	9.35	3.68	0.26	1.22	27.07	44.92
	Length (bp)	1,675,744	861,765	787,031	9,865,507	3,471,031	49,764,475	19,167,780	1,435,615	6,704,740	143,045,135	238,378,928
RK792_HAP1	Percentage (%)	0.31	0.16	0.14	1.8	0.63	9.1	3.5	0.26	1.23	26.16	43.59
	Length (bp)	1,714,835	877,724	739,884	9,992,805	3,753,335	51,143,338	20,120,474	1,466,766	6,820,570	150,062,369	248,609,630
RK792_HAP2	Percentage (%)	0.3	0.16	0.13	1.78	0.67	9.09	3.58	0.26	1.21	26.68	44.21
	Length (bp)	1,577,288	848,875	721,733	9,598,628	3,762,524	49,566,046	19,172,127	1,439,158	6,668,789	144,636,477	239,761,761
RK687_HAP1	Percentage (%)	0.29	0.16	0.13	1.76	0.69	9.09	3.51	0.26	1.22	26.51	43.96
	Length (bp)	1,791,992	891,948	772,978	10,352,958	3,788,921	52,591,815	19,658,752	1,499,027	6,951,221	149,096,763	249,064,953
RK687_HAP2	Percentage (%)	0.32	0.16	0.14	1.83	0.67	9.3	3.47	0.26	1.23	26.35	44.05
	Length (bp)	1,638,764	867,607	750,120	9,852,154	3,482,066	50,456,560	18,808,096	1,585,443	6,754,508	144,231,346	240,172,702
RK614_HAP1	Percentage (%)	0.3	0.16	0.14	1.8	0.64	9.21	3.43	0.29	1.23	26.32	43.85
	Length (bp)	1,662,578	876,947	721,634	10,139,196	3,893,565	52,253,837	19,674,933	1,668,133	6,914,600	152,105,660	251,575,207
RK614_HAP2	Percentage (%)	0.29	0.16	0.13	1.79	0.69	9.25	3.48	0.3	1.22	26.92	44.53
	Length (bp)	1,818,279	848,307	747,633	10,099,609	3,632,331	50,315,874	18,588,221	1,615,979	6,818,539	146,392,083	242,588,139
RK84_HAP1	Percentage (%)	0.33	0.15	0.14	1.83	0.66	9.11	3.37	0.29	1.23	26.51	43.94
	Length (bp)	1,623,131	911,123	825,329	10,042,021	3,666,512	50,567,463	20,757,061	1,475,726	6,707,239	151,219,885	249,576,974
RK84_HAP2	Percentage (%)	0.29	0.16	0.15	1.79	0.65	8.99	3.69	0.26	1.19	26.89	44.38
	Length (bp)	1,584,036	856,826	735,608	9,606,679	3,656,541	48,837,324	18,253,606	1,438,207	6,642,587	141,661,485	234,934,307
RK86_HAP1	Percentage (%)	0.29	0.16	0.14	1.77	0.67	8.99	3.36	0.26	1.22	26.09	43.28
	Length (bp)	1,635,342	889,819	819,763	10,232,296	3,979,160	50,916,890	20,538,233	1,612,913	6,842,705	151,655,809	250,830,697
RK86_HAP2	Percentage (%)	0.29	0.16	0.15	1.81	0.7	9.01	3.64	0.29	1.21	26.85	44.42
	Length (bp)	1,561,992	871,081	703,451	9,520,057	3,625,913	48,507,680	18,855,990	1,431,638	6,606,385	140,660,571	234,133,350
RK715_HAP1	Percentage (%)	0.29	0.16	0.13	1.77	0.68	9.03	3.51	0.27	1.23	26.19	43.61
	Length (bp)	1,720,559	840,458	829,939	9,864,372	3,904,494	51,542,598	20,653,027	1,494,907	6,869,778	151,145,614	250,467,417
RK715_HAP2	Percentage (%)	0.31	0.15	0.15	1.75	0.69	9.14	3.66	0.27	1.22	26.8	44.45
	Length (bp)	1,618,494	885,036	709,314	9,836,025	3,597,978	48,884,063	18,702,576	1,443,537	6,777,072	143,255,256	237,364,738
RK25_HAP1	Percentage (%)	0.3	0.16	0.13	1.81	0.66	9.01	3.45	0.27	1.25	26.42	43.76
	Length (bp)	1,750,639	922,964	784,569	10,476,428	3,704,065	49,350,376	19,478,479	1,455,193	6,764,377	148,050,667	244,408,057
	Percentage (%)	0.31	0.17	0.14	1.88	0.66	8.86	3.5	0.26	1.21	26.57	43.88

RK25_HAP2	Length (bp)	1,712,403	876,870	719,535	9,581,909	3,442,796	48,708,765	19,200,886	1,399,178	6,627,640	142,425,422	236,303,706
	Percentage (%)	0.32	0.16	0.13	1.77	0.64	9.01	3.55	0.26	1.23	26.35	43.73
TAG0014_HAP1	Length (bp)	1,689,454	923,841	781,621	10,472,401	3,727,706	51,161,910	20,457,810	1,628,973	6,917,998	157,020,984	256,742,534
	Percentage (%)	0.3	0.16	0.14	1.83	0.65	8.96	3.58	0.29	1.21	27.51	44.99
TAG0014_HAP2	Length (bp)	1,554,936	831,417	761,165	9,669,214	3,744,898	50,735,113	19,075,600	1,472,798	6,793,321	145,479,922	241,754,195
	Percentage (%)	0.28	0.15	0.14	1.75	0.68	9.18	3.45	0.27	1.23	26.34	43.78

Supplementary Table 8 Gene model statistics as calculated with gFACs.

	Number of genes	Number of monoexonic genes	Number of multiexonic genes	Average overall gene size	Average overall CDS size	Average overall exon size	Average number of exons per multiexonic gene	Largest multiexonic exon	Most exons in one gene	Largest intron	Number of complete models
RK05_HAP1	40,029	8,379	31,650	3,682.86	2,886.00	346.79	10.26	9,225	978	47,796	40,001
RK05_HAP2	39,118	8,214	30,904	3,702.84	2,917.01	345.33	10.43	9,342	861	56,719	39,090
RK17_HAP1	39,846	8,078	31,768	3,694.05	3,027.74	352.35	10.52	9,463	798	84,811	39,816
RK17_HAP2	38,803	7,853	30,950	3,689.41	3,027.19	350.07	10.59	18,100	798	44,068	38,779
RK205_HAP1	39,516	8,191	31,325	3,695.55	3,008.19	346.93	10.68	17,382	940	56,436	39,488
RK205_HAP2	39,454	8,204	31,250	3,701.47	3,030.10	348.02	10.73	12,945	847	48,429	39,421
RK25_HAP1	39,948	8,260	31,688	3,749.01	3,131.51	352.85	10.93	7,890	1,000	41,583	39,915
RK25_HAP2	39,227	8,050	31,177	3,761.85	3,158.20	351.79	11.04	13,981	993	39,852	39,198
RK313_HAP1	39,053	7,966	31,087	3,719.42	3,090.04	347.83	10.90	10,070	945	45,463	39,022
RK313_HAP2	38,555	7,791	30,764	3,745.65	3,093.15	346.32	10.94	8,879	963	58,234	38,523
RK320_HAP1	40,158	8,266	31,892	3,755.29	3,313.66	358.46	11.38	11,443	889	56,767	40,129
RK320_HAP2	39,136	8,067	31,069	3,732.69	3,330.86	356.19	11.52	12,630	1,008	71,969	39,106
RK339_HAP1	40,272	8,689	31,583	3,692.04	2,621.57	340.46	9.54	10,581	581	60,871	40,231
RK339_HAP2	39,380	8,321	31,059	3,695.20	2,647.33	341.52	9.56	8,602	581	43,797	39,350
RK388_HAP1	39,963	8,359	31,604	3,754.73	2,849.16	354.32	9.90	11,818	523	56,738	39,934
RK388_HAP2	39,400	8,201	31,199	3,746.84	2,863.32	352.82	9.99	7,890	914	64,875	39,375
RK436_HAP1	39,541	8,120	31,421	3,716.48	3,159.47	351.45	11.06	9,038	568	63,474	39,507
RK436_HAP2	39,503	8,066	31,437	3,738.13	3,184.64	352.73	11.09	9,485	615	46,607	39,466
RK441_HAP1	39,612	7,982	31,630	3,817.27	3,353.19	358.50	11.46	7,890	1,018	56,723	39,579
RK441_HAP2	39,675	8,090	31,585	3,778.29	3,343.84	357.50	11.49	8,838	1,012	74,618	39,645
RK465_HAP1	39,666	8,558	31,108	3,698.28	2,492.52	349.33	8.82	9,314	793	70,205	39,637
RK465_HAP2	39,155	8,376	30,779	3,699.98	2,505.57	348.37	8.88	11,532	450	47,808	39,135
RK498_HAP1	39,974	8,453	31,521	3,723.82	2,940.32	344.84	10.55	8,883	765	64,802	39,941
RK498_HAP2	39,323	8,228	31,095	3,717.40	2,979.89	345.09	10.66	8,019	877	48,292	39,287
RK50_HAP1	39,750	8,734	31,016	3,701.18	3,090.94	355.61	10.86	12,301	1,011	47,962	39,720
RK50_HAP2	39,435	8,631	30,804	3,704.71	3,093.35	354.42	10.89	12,803	1,011	47,962	39,410
RK561_HAP1	39,370	8,026	31,344	3,760.73	3,272.80	349.94	11.49	8,433	541	63,475	39,337
RK561_HAP2	39,271	8,046	31,225	3,732.38	3,244.38	346.65	11.51	9,918	510	47,936	39,247
RK614_HAP1	40,154	8,325	31,829	3,736.05	3,136.53	353.70	10.93	13,833	1,001	64,801	40,122

RK614_HAP2	39,922	8,333	31,589	3,718.57	3,122.81	352.30	10.94	10,860	883	68,718	39,891
RK687_HAP1	39,882	8,325	31,557	3,714.03	3,073.30	350.13	10.83	10,598	914	60,897	39,852
RK687_HAP2	39,499	8,202	31,297	3,708.03	3,080.45	348.01	10.91	12,511	992	42,839	39,480
RK715_HAP1	39,685	8,492	31,193	3,700.57	2,405.06	343.88	8.63	9,155	399	47,119	39,646
RK715_HAP2	39,432	8,569	30,863	3,666.55	2,350.90	343.12	8.48	7,890	394	61,923	19,408
RK74_HAP1	39,843	8,553	31,290	3,658.18	2,652.04	346.87	9.46	7,890	955	40,472	39,812
RK74_HAP2	39,145	8,295	30,850	3,678.21	2,681.06	345.57	9.58	7,890	822	52,591	39,112
RK792_HAP1	39,850	8,545	31,305	3,721.00	2,891.25	352.56	10.17	13,991	773	57,694	39,823
RK792_HAP2	39,421	8,517	30,904	3,689.39	2,816.05	349.13	10.01	18,101	773	47,649	39,386
RK84_HAP1	39,667	8,567	31,100	3,700.50	2,733.58	349.93	9.69	10,258	418	57,154	39,637
RK84_HAP2	39,434	8,547	30,887	3,703.81	2,794.74	349.35	9.94	8,045	445	43,787	39,399
RK86_HAP1	40,010	7,954	32,056	3,809.11	3,432.44	361.19	11.61	9,277	1,005	60,910	39,973
RK86_HAP2	39,337	7,803	31,534	3,799.27	3,431.45	362.85	11.55	11,799	1,010	44,003	39,305
RM01_HAP1	40,173	8,450	31,723	3,714.32	2,869.38	340.67	10.40	11,941	869	47,853	40,141
RM01_HAP2	39,510	8,321	31,189	3,694.02	2,896.87	340.75	10.50	10,099	808	55,949	39,479
RM02_HAP1	39,915	8,725	31,190	3,667.12	2,452.52	327.88	9.29	10,099	1,006	47,262	39,881
RM02_HAP2	39,495	8,623	30,872	3,650.62	2,489.95	328.96	9.40	10,330	999	55,023	39,454
RM03_HAP1	39,668	8,612	31,056	3,642.89	2,546.17	340.45	9.28	10,099	607	56,716	39,642
RM03_HAP2	38,882	8,337	30,545	3,639.44	2,553.72	339.84	9.29	10,099	623	46,357	38,844
TAG0014 HAP1	40,292	7,967	32,325	3,795.22	3,460.93	351.76	12.02	10,064	1,013	165,631	40,268
TAG0014 HAP2	39,705	7,860	31,845	3,771.46	3,465.84	351.81	12.04	13,880	1,010	61,943	39,678

Supplementary Table 9 Summary of structural variants in genome comparisons to the TAG0014_HAP1 reference.

	Syntenic regions	Inversions	Translocations	Duplications (reference)	Duplications (query)	Not aligned (reference)	Not aligned (query)
Query vs Reference							
RK05_HAP1 vs TAG0014_HAP1							
Count	13,574	97	8,226	20,309	16,138	20,546	20,711
Length reference	358,794,672	43,413,506	49,420,540	78,349,294	-	78,907,584	-
Length query	359,326,906	42,458,733	49,616,396	-	49,785,583	-	71,555,905
RK05_HAP2 vs TAG0014_HAP1							
Count	13,745	98	8,163	22,438	14,614	21,398	19,171
Length reference	374,781,650	19,972,331	47,461,037	81,564,324	-	82,821,190	-
Length query	374,977,613	18,880,097	47,578,663	-	47,609,740	-	64,525,350
RK17_HAP1 vs TAG0014_HAP1							
Count	13,930	96	8,822	21,442	15,698	21,491	20,276
Length reference	382,275,463	14,296,333	53,361,821	79,958,898	-	80,351,352	-
Length query	382,350,897	12,821,278	53,144,723	-	51,214,504	-	72,789,700
RK17_HAP2 vs TAG0014_HAP1							
Count	13,876	97	8,591	22,338	14,228	22,053	19,399
Length reference	370,146,712	13,588,618	50,792,752	83,104,632	-	90,062,893	-
Length query	370,226,563	11,473,940	50,457,675	-	44,542,470	-	67,924,336
RK205_HAP1 vs TAG0014_HAP1							
Count	14,175	102	8,754	22,004	15,366	22,027	20,307
Length reference	380,521,186	9,713,867	53,027,020	83,059,159	-	83,968,012	-
Length query	380,191,290	10,235,772	53,172,960	-	50,811,590	-	69,971,646
RK205_HAP2 vs TAG0014_HAP1							
Count	13,924	110	8,196	21,957	14,100	21,714	19,049
Length reference	381,343,600	13,197,538	48,016,335	81,481,352	-	83,256,651	-
Length query	381,354,046	10,991,832	48,515,881	-	46,095,193	-	63,637,600
RK25_HAP1 vs TAG0014_HAP1							
Count	14,008	105	8,822	22,645	15,960	21,542	20,473
Length reference	384,093,199	9,889,942	50,638,367	84,486,783	-	82,515,106	-
Length query	383,877,843	12,434,140	50,276,010	-	51,494,033	-	72,029,200
RK25_HAP2 vs TAG0014_HAP1							
Count	13,741	83	8,157	22,383	14,707	21,405	19,028
Length reference	383,997,757	10,281,469	45,906,412	85,078,629	-	83,203,904	-
Length query	384,145,324	10,428,219	45,761,594	-	46,547,374	-	65,270,518
RK313_HAP1 vs TAG0014_HAP1							
Count	15,696	116	9,591	23,553	16,026	23,728	22,477
Length reference	368,710,061	12,235,377	54,005,237	88,230,480	-	87,337,215	-

Length query	369,148,766	11,764,688	53,670,091	-	51,724,757	-	77,085,001
RK313_HAP2 vs TAG0014_HAP1							
Count	15,062	114	8,878	23,263	14,830	23,372	20,844
Length reference	367,440,391	17,027,220	50,931,001	85,902,905	-	87,843,869	-
Length query	367,331,152	14,736,291	50,539,733	-	47,550,977	-	68,161,026
RK320_HAP1 vs TAG0014_HAP1							
Count	14,283	109	9,118	22,239	16,584	21,676	21,241
Length reference	381,683,926	13,134,748	52,525,169	83,446,754	-	79,379,718	-
Length query	381,790,729	12,925,351	52,537,906	-	53,396,414	-	78,997,611
RK320_HAP2 vs TAG0014_HAP1							
Count	14,002	104	8,491	21,951	14,536	21,936	19,910
Length reference	377,303,237	13,183,277	47,692,719	78,671,731	-	88,835,304	-
Length query	377,087,367	11,939,400	47,784,104	-	48,997,584	-	67,787,411
RK339_HAP1 vs TAG0014_HAP1							
Count	13,827	82	8,589	21,337	16,568	21,023	20,864
Length reference	386,685,402	14,177,510	49,910,063	82,421,836	-	78,119,006	-
Length query	386,557,280	13,011,388	50,528,998	-	54,092,212	-	72,490,410
RK339_HAP2 vs TAG0014_HAP1							
Count	13,926	96	8,357	21,696	14,870	21,678	19,661
Length reference	384,852,571	8,935,890	48,729,767	83,196,336	-	82,992,944	-
Length query	384,983,897	10,229,852	48,891,261	-	47,030,792	-	68,527,785
RK388_HAP1 vs TAG0014_HAP1							
Count	13,800	92	8,627	21,561	16,110	21,289	20,583
Length reference	384,381,831	12,309,051	50,555,563	80,261,742	-	82,043,760	-
Length query	384,753,873	15,558,384	51,414,382	-	51,690,847	-	74,610,073
RK388_HAP2 vs TAG0014_HAP1							
Count	13,942	100	8,269	21,453	15,114	21,317	19,621
Length reference	379,500,010	16,366,844	49,532,825	80,534,425	-	82,166,046	-
Length query	379,273,893	14,038,613	49,826,174	-	48,475,415	-	67,665,288
RK436_HAP1 vs TAG0014_HAP1							
Count	13,324	98	7,970	20,763	15,062	20,264	19,297
Length reference	377,730,781	28,281,636	45,577,742	76,845,440	-	77,335,669	-
Length query	378,321,957	25,028,108	45,917,735	-	50,496,122	-	68,677,405
RK436_HAP2 vs TAG0014_HAP1							
Count	13,992	94	8,209	21,615	14,544	21,638	19,344
Length reference	383,551,742	10,458,651	49,262,286	81,079,132	-	83,458,507	-
Length query	383,391,014	12,056,509	49,165,498	-	48,999,522	-	63,576,380
RK441_HAP1 vs TAG0014_HAP1							
Count	14,116	107	8,625	21,361	16,099	21,303	20,787
Length reference	388,430,047	10,695,208	52,190,802	81,334,475	-	79,773,345	-

Length query	388,631,374	12,037,584	52,456,622	-	50,973,900	-	77,567,172
RK441_HAP2 vs TAG0014_HAP1							
Count	13,842	97	8,533	22,186	15,617	21,567	20,484
Length reference	378,667,908	14,866,112	49,999,290	82,899,680	-	82,974,964	-
Length query	378,851,593	14,788,843	49,999,330	-	52,299,999	-	72,526,378
RK465_HAP1 vs TAG0014_HAP1							
Count	14,037	111	8,643	22,686	17,558	21,474	20,817
Length reference	382,830,651	10,964,289	53,534,259	85,202,466	-	80,344,265	-
Length query	382,900,549	9,402,039	53,669,293	-	54,290,318	-	73,818,443
RK465_HAP2 vs TAG0014_HAP1							
Count	13,965	91	8,444	21,684	14,873	21,673	19,738
Length reference	381,469,817	9,922,208	47,233,749	83,369,964	-	83,647,098	-
Length query	381,558,967	9,511,139	47,843,340	-	46,747,009	-	70,630,511
RK498_HAP1 vs TAG0014_HAP1							
Count	13,939	99	8,453	21,361	16,628	20,665	20,766
Length reference	389,558,646	12,417,687	52,308,397	82,372,110	-	75,268,951	-
Length query	389,296,203	13,927,446	52,817,005	-	54,965,359	-	74,862,740
RK498_HAP2 vs TAG0014_HAP1							
Count	14,224	102	8,592	21,871	15,210	21,929	20,032
Length reference	383,458,231	10,724,238	46,952,818	81,951,852	-	83,171,798	-
Length query	383,741,275	8,954,957	47,109,286	-	47,211,736	-	72,165,026
RK50_HAP1 vs TAG0014_HAP1							
Count	13,771	96	8,683	21,550	16,626	21,211	20,877
Length reference	379,329,302	17,921,175	49,555,487	83,509,146	-	79,200,325	-
Length query	379,293,659	16,043,518	49,739,753	-	52,518,071	-	75,442,511
RK50_HAP2 vs TAG0014_HAP1							
Count	13,421	101	8,545	20,189	14,640	21,071	19,505
Length reference	367,537,327	13,184,939	63,896,963	106,325,777	-	80,424,599	-
Length query	367,477,317	10,686,142	64,141,246	-	48,708,127	-	67,740,338
RK561_HAP1 vs TAG0014_HAP1							
Count	14,383	88	8,871	22,149	18,115	21,891	21,137
Length reference	382,688,213	9,293,846	54,107,265	85,618,505	-	82,429,522	-
Length query	382,882,262	9,175,606	54,120,518	-	54,514,230	-	75,232,699
RK561_HAP2 vs TAG0014_HAP1							
Count	14,236	96	8,341	21,852	14,392	21,655	19,306
Length reference	383,173,015	11,823,832	49,373,126	78,580,105	-	83,488,632	-
Length query	383,292,934	12,333,472	49,549,950	-	47,597,785	-	64,627,807
RK614_HAP1 vs TAG0014_HAP1							
Count	13,781	104	8,431	21,084	15,935	20,783	20,351
Length reference	392,352,387	9,190,326	48,383,214	78,580,337	-	78,180,187	-

Length query	391,942,860	9,766,893	48,779,153	-	53,135,810	-	73,470,860
RK614_HAP2 vs TAG0014_HAP1							
Count	13,872	94	8,656	22,250	15,773	21,321	19,909
Length reference	386,035,781	9,432,169	52,714,919	81,288,383	-	80,165,656	-
Length query	385,958,670	9,737,354	52,274,933	-	49,353,709	-	68,489,261
RK687_HAP1 vs TAG0014_HAP1							
Count	13,924	91	8,897	21,892	16,947	21,827	21,537
Length reference	383,899,629	9,375,171	50,018,292	81,695,139	-	84,502,215	-
Length query	383,933,929	9,730,704	50,340,626	-	56,752,392	-	77,410,302
RK687_HAP2 vs TAG0014_HAP1							
Count	13,670	76	8,231	21,318	15,495	21,349	19,422
Length reference	385,731,153	8,856,201	47,647,912	81,686,852	-	83,934,386	-
Length query	385,996,818	9,765,969	47,949,447	-	49,395,671	-	68,357,054
RK715_HAP1 vs TAG0014_HAP1							
Count	13,666	105	8,311	21,019	16,147	20,856	20,500
Length reference	391,689,964	9,744,293	47,986,803	81,733,423	-	78,597,798	-
Length query	391,655,439	10,201,640	48,190,672	-	52,514,686	-	72,632,429
RK715_HAP2 vs TAG0014_HAP1							
Count	13,444	95	7,879	21,572	14,117	20,616	18,706
Length reference	390,797,269	11,844,639	44,386,217	78,370,996	-	79,362,122	-
Length query	390,595,469	9,987,738	44,429,078	-	44,073,723	-	64,623,366
RK74_HAP1 vs TAG0014_HAP1							
Count	13,964	93	8,724	21,792	16,279	21,542	21,092
Length reference	381,061,508	10,462,526	51,368,031	85,604,208	-	83,059,950	-
Length query	381,232,128	10,739,111	51,382,302	-	54,199,316	-	79,356,830
RK74_HAP2 vs TAG0014_HAP1							
Count	14,004	95	9,084	22,641	15,267	22,329	20,164
Length reference	375,892,348	9,311,128	52,167,751	82,980,286	-	88,526,118	-
Length query	375,934,389	8,424,304	52,074,948	-	47,562,881	-	77,195,229
RK792_HAP1 vs TAG0014_HAP1							
Count	13,921	97	8,454	21,581	16,318	21,115	20,457
Length reference	388,065,652	9,639,710	52,210,674	86,017,852	-	76,989,409	-
Length query	388,055,536	8,340,975	52,380,742	-	53,599,090	-	73,909,820
RK792_HAP2 vs TAG0014_HAP1							
Count	13,852	100	8,303	21,758	14,844	21,562	19,233
Length reference	384,431,219	9,418,758	47,953,090	85,032,074	-	82,305,971	-
Length query	384,405,170	10,701,293	47,959,974	-	46,793,959	-	67,000,789
RK84_HAP1 vs TAG0014_HAP1							
Count	13,891	92	8,846	21,909	16,802	21,610	21,067
Length reference	383,907,429	11,427,110	48,807,102	83,603,322	-	80,305,330	-

Length query	384,143,277	13,217,583	48,750,168	-	52,256,617	-	75,603,704
RK84_HAP2 vs TAG0014_HAP1							
Count	12,745	97	7,147	20,247	13,513	19,809	17,430
Length reference	376,167,865	33,760,654	41,977,139	74,131,087	-	76,800,084	-
Length query	376,352,110	33,746,783	41,737,723	-	43,835,977	-	58,886,782
RK86_HAP1 vs TAG0014_HAP1							
Count	13,695	89	8,475	20,995	16,229	20,714	20,417
Length reference	389,244,129	10,318,945	51,008,133	82,970,708	-	78,467,062	-
Length query	389,217,662	9,661,594	51,178,167	-	53,625,700	-	73,544,316
RK86_HAP2 vs TAG0014_HAP1							
Count	13,688	101	8,062	21,055	13,965	21,328	18,816
Length reference	383,349,301	10,099,804	46,606,340	80,251,787	-	85,660,390	-
Length query	383,015,986	9,242,475	46,383,213	-	46,238,900	-	64,812,991
RM01_HAP1 vs TAG0014_HAP1							
Count	13,550	90	8,347	21,137	16,543	20,948	20,898
Length reference	392,085,015	10,247,293	49,338,850	82,167,079	-	75,967,340	-
Length query	391,900,321	14,897,041	49,690,866	-	52,548,889	-	77,684,570
RM01_HAP2 vs TAG0014_HAP1							
Count	13,828	101	8,444	21,587	15,452	21,364	20,327
Length reference	384,566,295	11,380,863	50,069,030	82,533,719	-	79,964,522	-
Length query	384,763,658	10,991,232	50,246,978	-	48,282,502	-	75,252,644
RM02_HAP1 vs TAG0014_HAP1							
Count	12,841	105	8,087	19,621	16,101	19,911	19,893
Length reference	394,676,767	13,051,071	48,701,983	74,784,132	-	76,236,203	-
Length query	394,920,992	16,293,764	48,962,885	-	50,892,679	-	76,519,225
RM02_HAP2 vs TAG0014_HAP1							
Count	12,979	87	7,810	20,070	14,726	20,292	18,989
Length reference	390,591,306	14,332,568	45,800,517	75,774,838	-	78,197,631	-
Length query	391,005,053	12,361,417	46,465,033	-	48,823,757	-	72,378,341
RM03_HAP1 vs TAG0014_HAP1							
Count	13,581	82	8,291	21,071	15,907	20,966	20,150
Length reference	389,633,878	10,693,283	48,930,737	82,767,111	-	77,441,426	-
Length query	389,812,793	10,316,284	48,929,302	-	55,430,587	-	67,896,889
RM03_HAP2 vs TAG0014_HAP1							
Count	13,925	97	8,357	21,755	15,075	21,736	20,016
Length reference	381,420,497	11,567,644	47,004,095	84,640,521	-	83,734,908	-
Length query	381,274,099	9,989,404	46,866,478	-	47,313,575	-	69,353,800
TAG0014_HAP2 vs TAG0014_HAP1							
Count	13,349	89	7,827	21,289	14,659	20,602	18,917
Length reference	396,000,960	9,417,568	46,538,476	82,095,122	-	75,546,270	-

Length query	396,333,150	8,282,210	46,282,393	-	48,917,664	-	65,828,782
--------------	-------------	-----------	------------	---	------------	---	------------

Supplementary Table 10 Summary of local variants in genome comparisons to the TAG0014_HAP1 reference.

	SNPs	Insertions	Deletions	Copygains	Copylosses	Highly diverged	Tandem repeats
Query vs Reference							
RK05_HAP1 vs TAG0014_HAP1							
Count	6,333,002	656,754	494,692	1,450	1,384	5,209	299
Length reference	6,333,002	-	4,868,678	-	8,126,971	20,084,107	1,229,447
Length query	6,333,002	5,430,671	-	5,343,003	-	21,729,465	1,508,025
RK05_HAP2 vs TAG0014_HAP1							
Count	6,268,350	586,737	493,094	1,323	1,289	5,121	292
Length reference	6,268,350	-	4,530,707	-	4,427,127	17,463,644	1,227,859
Length query	6,268,350	4,674,588	-	4,891,107	-	15,958,989	1,437,895
RK17_HAP1 vs TAG0014_HAP1							
Count	6,455,562	585,042	504,362	1,307	1,285	4,974	299
Length reference	6,455,562	-	4,632,428	-	4,298,783	17,193,025	852,650
Length query	6,455,562	4,631,418	-	4,118,141	-	15,747,662	869,833
RK17_HAP2 vs TAG0014_HAP1							
Count	6,325,116	574,488	496,182	1,277	1,268	4,957	291
Length reference	6,325,116	-	4,565,683	-	3,822,866	16,595,164	963,088
Length query	6,325,116	4,413,076	-	3,895,673	-	14,236,746	1,051,898
RK205_HAP1 vs TAG0014_HAP1							
Count	6,573,463	584,030	512,514	1,327	1,298	5,036	295
Length reference	6,573,463	-	4,600,640	-	3,734,018	15,769,820	912,988
Length query	6,573,463	4,648,435	-	4,018,388	-	15,795,390	940,781
RK205_HAP2 vs TAG0014_HAP1							
Count	6,293,109	561,034	494,305	1,323	1,239	4,903	307
Length reference	6,293,109	-	4,513,235	-	4,162,565	16,744,120	918,437
Length query	6,293,109	4,424,765	-	3,878,988	-	15,400,982	951,231
RK25_HAP1 vs TAG0014_HAP1							
Count	6,386,184	571,535	496,748	1,230	1,245	4,838	314
Length reference	6,386,184	-	4,417,245	-	3,497,527	14,963,425	1,113,090
Length query	6,386,184	4,554,131	-	3,653,981	-	16,672,550	1,059,492
RK25_HAP2 vs TAG0014_HAP1							
Count	6,154,989	556,308	483,958	1,208	1,180	4,616	302
Length reference	6,154,989	-	4,296,871	-	4,031,318	13,640,043	1,200,281
Length query	6,154,989	4,664,750	-	3,663,904	-	13,619,844	1,384,404
RK313_HAP1 vs TAG0014_HAP1							
Count	7,155,797	639,475	555,265	1,403	1,345	5,397	293
Length reference	7,155,797	-	5,064,682	-	3,820,841	17,238,188	881,556

Length query	7,155,797	5,031,857	-	4,620,729	-	16,117,367	995,472
RK313_HAP2 vs TAG0014_HAP1							
Count	6,898,700	628,166	539,827	1,339	1,279	5,242	314
Length reference	6,898,700	-	5,047,313	-	4,394,796	19,190,264	698,238
Length query	6,898,700	5,103,544	-	4,054,176	-	16,657,071	738,845
RK320_HAP1 vs TAG0014_HAP1							
Count	6,527,876	591,954	503,999	1,371	1,251	5,090	320
Length reference	6,527,876	-	4,882,702	-	4,512,298	15,732,424	1,188,869
Length query	6,527,876	4,648,636	-	4,010,136	-	16,456,280	1,128,434
RK320_HAP2 vs TAG0014_HAP1							
Count	6,338,479	577,349	497,488	1,319	1,310	4,878	309
Length reference	6,338,479	-	4,624,618	-	4,020,613	15,887,050	1,233,814
Length query	6,338,479	4,508,634	-	4,071,508	-	14,599,716	1,190,410
RK339_HAP1 vs TAG0014_HAP1							
Count	6,281,632	572,709	492,667	1,230	1,231	4,735	303
Length reference	6,281,632	-	4,405,987	-	3,740,509	17,027,816	1,173,442
Length query	6,281,632	4,623,108	-	3,893,915	-	16,009,350	1,110,277
RK339_HAP2 vs TAG0014_HAP1							
Count	6,229,859	564,079	488,392	1,301	1,199	4,705	281
Length reference	6,229,859	-	4,556,160	-	3,620,303	13,695,006	913,446
Length query	6,229,859	4,496,120	-	4,873,151	-	14,112,719	948,346
RK388_HAP1 vs TAG0014_HAP1							
Count	6,364,710	578,149	498,049	1,280	1,211	4,913	292
Length reference	6,364,710	-	4,666,719	-	3,527,873	15,303,833	1,044,990
Length query	6,364,710	4,705,065	-	4,310,349	-	19,060,286	1,110,482
RK388_HAP2 vs TAG0014_HAP1							
Count	6,313,869	581,327	497,947	1,306	1,257	4,846	312
Length reference	6,313,869	-	4,567,803	-	3,956,178	18,468,444	1,071,943
Length query	6,313,869	4,548,982	-	3,615,402	-	16,380,003	1,083,869
RK436_HAP1 vs TAG0014_HAP1							
Count	6,171,004	602,200	484,843	1,291	1,269	4,894	287
Length reference	6,171,004	-	4,446,000	-	6,152,044	19,048,587	1,042,849
Length query	6,171,004	4,866,086	-	4,456,176	-	18,113,920	1,027,761
RK436_HAP2 vs TAG0014_HAP1							
Count	6,273,975	573,814	492,549	1,297	1,219	4,853	301
Length reference	6,273,975	-	4,492,663	-	4,217,225	14,688,579	810,966
Length query	6,273,975	4,481,023	-	4,355,331	-	15,962,100	811,078
RK441_HAP1 vs TAG0014_HAP1							
Count	6,385,625	572,165	500,507	1,321	1,281	4,911	289
Length reference	6,385,625	-	4,364,489	-	4,137,885	14,826,113	1,067,810

Length query	6,385,625	4,678,981	-	3,786,257	-	16,777,342	959,075
RK441_HAP2 vs TAG0014_HAP1							
Count	6,307,061	589,471	493,195	1,327	1,271	4,845	286
Length reference	6,307,061	-	4,412,349	-	4,363,526	14,140,889	914,247
Length query	6,307,061	4,507,215	-	4,557,267	-	14,052,736	849,857
RK465_HAP1 vs TAG0014_HAP1							
Count	6,451,719	579,314	503,761	1,306	1,265	4,983	288
Length reference	6,451,719	-	4,676,473	-	4,633,807	16,349,947	1,034,530
Length query	6,451,719	4,739,001	-	3,887,881	-	15,719,687	1,067,165
RK465_HAP2 vs TAG0014_HAP1							
Count	6,298,147	567,629	494,593	1,249	1,277	4,699	294
Length reference	6,298,147	-	4,246,861	-	4,635,754	13,857,796	1,072,377
Length query	6,298,147	4,612,027	-	3,388,054	-	14,777,128	1,264,572
RK498_HAP1 vs TAG0014_HAP1							
Count	6,349,732	574,811	495,102	1,382	1,200	4,919	287
Length reference	6,349,732	-	4,629,199	-	3,848,932	16,155,312	963,737
Length query	6,349,732	4,905,535	-	4,772,766	-	16,812,827	950,433
RK498_HAP2 vs TAG0014_HAP1							
Count	6,388,410	565,853	497,181	1,275	1,243	4,743	280
Length reference	6,388,410	-	4,396,992	-	4,170,036	15,438,216	1,030,699
Length query	6,388,410	4,966,520	-	4,240,463	-	13,585,528	1,006,611
RK50_HAP1 vs TAG0014_HAP1							
Count	6,331,894	591,335	493,786	1,300	1,253	4,894	304
Length reference	6,331,894	-	4,719,065	-	4,137,309	17,364,405	1,012,670
Length query	6,331,894	4,394,168	-	4,459,753	-	15,526,245	1,124,453
RK50_HAP2 vs TAG0014_HAP1							
Count	6,293,452	600,104	491,969	1,295	1,289	4,882	308
Length reference	6,293,452	-	4,698,710	-	4,558,527	19,143,183	1,242,707
Length query	6,293,452	4,873,448	-	4,038,956	-	17,171,525	1,214,741
RK561_HAP1 vs TAG0014_HAP1							
Count	6,563,257	589,095	510,549	1,349	1,289	5,028	309
Length reference	6,563,257	-	4,637,581	-	3,661,775	14,903,212	1,207,706
Length query	6,563,257	4,639,655	-	3,956,858	-	14,628,352	1,350,947
RK561_HAP2 vs TAG0014_HAP1							
Count	6,382,825	576,896	500,026	1,337	1,234	4,760	301
Length reference	6,382,825	-	4,538,367	-	4,609,550	15,384,293	1,053,011
Length query	6,382,825	4,863,457	-	4,709,487	-	15,817,170	1,046,643
RK614_HAP1 vs TAG0014_HAP1							
Count	6,247,212	560,483	487,125	1,211	1,219	4,739	294
Length reference	6,247,212	-	4,247,946	-	3,506,936	13,629,950	1,034,874

Length query	6,247,212	4,705,262	-	3,499,142	-	13,760,508	940,992
RK614_HAP2 vs TAG0014_HAP1							
Count	6,328,423	561,016	492,161	1,290	1,302	4,901	310
Length reference	6,328,423	-	4,514,199	-	3,702,622	15,275,836	978,767
Length query	6,328,423	4,995,177	-	4,049,788	-	14,271,921	969,571
RK687_HAP1 vs TAG0014_HAP1							
Count	6,340,498	567,038	495,283	1,247	1,217	4,787	291
Length reference	6,340,498	-	4,174,627	-	3,812,520	14,397,545	919,453
Length query	6,340,498	4,657,569	-	4,158,244	-	14,350,110	941,104
RK687_HAP2 vs TAG0014_HAP1							
Count	6,195,266	554,686	484,577	1,274	1,175	4,773	274
Length reference	6,195,266	-	4,489,002	-	3,289,812	13,892,940	1,428,459
Length query	6,195,266	4,617,840	-	4,520,574	-	14,129,450	1,472,568
RK715_HAP1 vs TAG0014_HAP1							
Count	6,129,931	551,338	482,381	1,253	1,198	4,584	300
Length reference	6,129,931	-	4,300,765	-	3,678,487	14,397,009	1,265,704
Length query	6,129,931	4,675,773	-	3,652,657	-	14,576,866	1,404,971
RK715_HAP2 vs TAG0014_HAP1							
Count	6,009,031	541,707	470,181	1,190	1,199	4,587	278
Length reference	6,009,031	-	4,452,456	-	4,314,563	14,576,802	839,772
Length query	6,009,031	4,256,818	-	3,733,287	-	13,306,601	779,056
RK74_HAP1 vs TAG0014_HAP1							
Count	6,391,426	571,095	498,263	1,282	1,252	4,914	315
Length reference	6,391,426	-	4,395,686	-	3,976,470	15,442,479	1,325,877
Length query	6,391,426	4,633,733	-	3,462,399	-	16,100,948	1,450,772
RK74_HAP2 vs TAG0014_HAP1							
Count	6,394,128	574,816	499,303	1,212	1,221	4,876	326
Length reference	6,394,128	-	4,564,655	-	3,404,952	15,236,598	1,120,402
Length query	6,394,128	4,577,061	-	3,411,387	-	14,130,867	1,180,638
RK792_HAP1 vs TAG0014_HAP1							
Count	6,353,016	559,633	495,056	1,309	1,251	4,822	295
Length reference	6,353,016	-	4,459,112	-	4,189,774	15,848,985	1,363,227
Length query	6,353,016	4,707,774	-	3,667,530	-	14,803,267	1,485,409
RK792_HAP2 vs TAG0014_HAP1							
Count	6,257,905	553,314	488,808	1,216	1,207	4,734	302
Length reference	6,257,905	-	4,429,352	-	3,482,755	14,820,060	1,553,975
Length query	6,257,905	4,676,449	-	3,928,975	-	15,235,346	1,682,574
RK84_HAP1 vs TAG0014_HAP1							
Count	6,337,807	572,625	494,802	1,200	1,183	4,698	304
Length reference	6,337,807	-	4,746,546	-	3,238,200	14,625,906	1,073,950

Length query	6,337,807	4,526,144	-	3,220,711	-	16,869,808	1,062,796
RK84_HAP2 vs TAG0014_HAP1							
Count	5,942,336	597,893	469,318	1,258	1,290	4,762	309
Length reference	5,942,336	-	4,607,934	-	4,996,038	21,021,528	1,307,790
Length query	5,942,336	4,520,270	-	4,821,451	-	21,021,173	1,442,459
RK86_HAP1 vs TAG0014_HAP1							
Count	6,234,719	563,762	487,627	1,247	1,223	4,770	297
Length reference	6,234,719	-	4,572,946	-	4,241,159	14,617,091	1,156,477
Length query	6,234,719	4,638,323	-	3,881,395	-	14,350,781	1,178,734
RK86_HAP2 vs TAG0014_HAP1							
Count	6,068,643	547,177	477,480	1,243	1,197	4,610	306
Length reference	6,068,643	-	4,325,196	-	3,314,025	14,077,096	1,075,183
Length query	6,068,643	4,331,111	-	3,573,113	-	12,485,743	1,019,034
RM01_HAP1 vs TAG0014_HAP1							
Count	6,284,147	557,580	486,661	1,303	1,222	4,708	287
Length reference	6,284,147	-	4,533,811	-	4,245,864	13,569,387	1,153,554
Length query	6,284,147	4,702,293	-	5,559,356	-	16,906,357	1,075,554
RM01_HAP2 vs TAG0014_HAP1							
Count	6,329,112	565,164	495,355	1,280	1,258	4,928	283
Length reference	6,329,112	-	4,506,721	-	4,103,977	14,463,278	1,025,007
Length query	6,329,112	4,595,596	-	3,826,451	-	14,484,119	1,154,859
RM02_HAP1 vs TAG0014_HAP1							
Count	5,906,910	544,648	461,720	1,237	1,166	4,720	269
Length reference	5,906,910	-	4,212,219	-	4,290,407	15,009,612	902,678
Length query	5,906,910	4,386,886	-	6,998,941	-	16,061,317	823,935
RM02_HAP2 vs TAG0014_HAP1							
Count	5,973,993	539,354	465,360	1,251	1,189	4,641	287
Length reference	5,973,993	-	3,929,606	-	4,243,163	15,744,528	1,083,507
Length query	5,973,993	4,387,591	-	4,201,688	-	14,589,140	992,429
RM03_HAP1 vs TAG0014_HAP1							
Count	6,190,597	549,234	483,153	1,259	1,196	4,727	281
Length reference	6,190,597	-	4,259,156	-	4,283,711	14,406,999	864,389
Length query	6,190,597	4,277,775	-	3,440,840	-	14,954,510	909,899
RM03_HAP2 vs TAG0014_HAP1							
Count	6,309,164	573,277	493,891	1,223	1,227	4,847	298
Length reference	6,309,164	-	4,361,363	-	4,202,869	15,233,383	1,040,853
Length query	6,309,164	4,302,455	-	3,598,130	-	13,932,877	1,102,203
TAG0014_HAP2 vs TAG0014_HAP1							
Count	6,197,893	550,834	482,009	1,274	1,219	4,849	286
Length reference	6,197,893	-	4,267,853	-	3,916,910	16,182,833	970,007

Length query	6,197,893	4,318,059	-	3,698,389	-	15,199,906	1,121,248
--------------	-----------	-----------	---	-----------	---	------------	-----------

Supplementary Table 11 Total number of variants before and after merging with SURVIVOR.

	Syntenic	Inversions	Trans	Duplications (reference)	Duplications (query)	Not aligned (reference)	Not aligned (query)	Copygains	Copylosses	Highly diverged	Tandem repeats
Total events from SyRI	678,931	4,761	414,223	1,059,479	760,907	1,047,116	983,996	62,982	60,972	237,848	14,549
Total merged events	72,064	511	4,391	32,659		90,926		2,924	4,728	29,437	1,499

Supplementary Table 12 Summary statistics of orthogroups assignments from Orthofinder.

Number of species	50
Number of genes	1,979,054
Number of genes in orthogroups	1,977,044
Number of unassigned genes	2,010
Percentage of genes in orthogroups	99.9
Percentage of unassigned genes	0.1
Number of orthogroups	46,626
Number of species-specific orthogroups	124
Number of genes in species-specific orthogroups	381
Percentage of genes in species-specific orthogroups	0
Mean orthogroup size	42.4
Median orthogroup size	50
G50 (assigned genes)	50
G50 (all genes)	50
O50 (assigned genes)	14,178
O50 (all genes)	14,198
Number of orthogroups with all species present	22,387
Number of single-copy orthogroups	18,461

Supplementary Table 13 Number of genes per genome and gene family across pangenome categories.

Genome	Core	Softcore	Shell	Private
RK05_HAP1	19,993	1,716	17,486	834
RK05_HAP2	19,992	1,699	16,618	809
RK17_HAP1	19,989	1,748	17,430	679
RK17_HAP2	19,982	1,394	16,750	677
RK205_HAP1	19,982	1,722	17,139	673
RK205_HAP2	19,986	1,729	17,114	625
RK25_HAP1	19,987	1,724	17,425	812
RK25_HAP2	20,002	1,717	16,781	727
RK313_HAP1	20,055	1,689	16,277	1,032
RK313_HAP2	20,022	1,690	15,979	864
RK320_HAP1	19,986	1,719	17,632	821
RK320_HAP2	19,995	1,708	16,741	692
RK339_HAP1	20,004	1,720	17,767	781
RK339_HAP2	19,986	1,708	16,893	793
RK388_HAP1	20,000	1,728	17,415	820
RK388_HAP2	19,998	1,719	16,908	775
RK436_HAP1	19,995	1,697	17,042	807
RK436_HAP2	19,997	1,730	16,889	887
RK441_HAP1	20,014	1,704	17,103	791
RK441_HAP2	19,990	1,709	17,120	856
RK465_HAP1	20,005	1,712	17,156	793
RK465_HAP2	19,992	1,700	16,687	776
RK498_HAP1	19,995	1,710	17,433	836
RK498_HAP2	20,004	1,711	16,847	761
RK50_HAP1	19,995	1,711	17,374	670
RK50_HAP2	19,990	1,709	17,157	579
RK561_HAP1	19,995	1,712	16,887	776
RK561_HAP2	19,988	1,716	16,741	826
RK614_HAP1	19,982	1,715	17,650	807
RK614_HAP2	19,982	1,727	17,408	805
RK687_HAP1	20,013	1,713	17,335	821
RK687_HAP2	19,989	1,716	17,060	734
RK715_HAP1	19,984	1,714	17,258	729
RK715_HAP2	20,000	1,709	17,100	623
RK74_HAP1	19,988	1,724	17,544	587
RK74_HAP2	19,981	1,693	16,960	511
RK792_HAP1	19,982	1,711	17,541	616
RK792_HAP2	19,971	1,709	17,069	672
RK84_HAP1	19,978	1,700	17,235	754
RK84_HAP2	19,975	1,704	16,968	787
RK86_HAP1	19,974	1,711	17,573	752
RK86_HAP2	19,976	1,706	16,891	764
RM01_HAP1	19,980	1,705	17,733	755
RM01_HAP2	19,982	1,714	16,959	855
RM02_HAP1	19,974	1,713	17,500	728
RM02_HAP2	19,982	1,720	17,015	778
RM03_HAP1	19,976	1,717	17,311	664

RM03_HAP2	19,975	1,717	16,446	744
TAG0014_HAP1	19,982	1,703	17,772	835
TAG0014_HAP2	19,995	1,685	17,225	800
Average	19,991	1,706	17,127	758
Total	979,545	83,592	839,119	37,093
Percentage	50.5%	4.3%	43.3%	1.9%

Supplementary Table 14 Enriched Biological Processes GO (gene ontology) of pangenome core genes, softcore genes, shell genes and private genes generated with TopGO⁸³. The enriched GO categories were determined using Fisher's exact test, followed by the Benjamin–Hochberg correction to obtain adjusted *P* values for multiple testing. Also see **Supplementary Figure 9**.

GO ID	GO Term	Annotated	Significant	Expected	Fisher.p	Adjust.p	Pangenome	GO category
GO:0006355	regulation of DNA-templated transcription	84,631	66,579	55,924.90	1,00E-30	1.18e-30	Core	Biological processes
GO:0010628	positive regulation of gene expression	27,454	20,501	18,141.84	1,00E-30	1.18e-30	Core	Biological processes
GO:0010629	negative regulation of gene expression	26,506	19,307	17,515.40	1,00E-30	1.18e-30	Core	Biological processes
GO:0045944	positive regulation of transcription by RNA polymerase II	5,825	5,193	3,849.21	1,00E-30	1.18e-30	Core	Biological processes
GO:0030154	cell differentiation	40,774	31,250	26,943.82	1,00E-30	1.18e-30	Core	Biological processes
GO:0045892	negative regulation of DNA-templated transcription	11,856	9,636	7,834.55	1,00E-30	1.18e-30	Core	Biological processes
GO:0007623	circadian rhythm	6,775	5,520	4,476.98	1,00E-30	1.18e-30	Core	Biological processes
GO:0010091	trichome branching	1,453	1,435	960.16	1,00E-30	1.18e-30	Core	Biological processes
GO:0044038	cell wall macromolecule biosynthetic process	2,504	2,320	1,654.67	1,00E-30	1.18e-30	Core	Biological processes
GO:0045036	protein targeting to chloroplast	1,446	1,414	955.53	1,00E-30	1.18e-30	Core	Biological processes
GO:0098771	inorganic ion homeostasis	13,016	8,906	8,601.09	1,00E-30	1.18e-30	Core	Biological processes
GO:0015833	peptide transport	23,230	16,972	15,350.59	1,00E-30	1.18e-30	Core	Biological processes
GO:0051128	regulation of cellular component organization	23,844	17,108	15,756.32	1,00E-30	1.18e-30	Core	Biological processes
GO:0009059	macromolecule biosynthetic process	169,600	126,656	112,073.16	1,00E-30	1.18e-30	Core	Biological processes
GO:0051726	regulation of cell cycle	12,583	10,159	8,314.96	1,00E-30	1.18e-30	Core	Biological processes
GO:0009910	negative regulation of flower development	2,507	2,243	1,656.65	1,00E-30	1.18e-30	Core	Biological processes
GO:1990542	mitochondrial transmembrane transport	2,032	1,775	1,342.76	1,00E-30	1.18e-30	Core	Biological processes
GO:0085029	extracellular matrix assembly	1,912	1,751	1,263.47	1,00E-30	1.18e-30	Core	Biological processes
GO:0016567	protein ubiquitination	22,512	16,718	14,876.13	1,00E-30	1.18e-30	Core	Biological processes
GO:0005992	trehalose biosynthetic process	903	898	596.71	1,00E-30	1.18e-30	Core	Biological processes
GO:0030974	thiamine pyrophosphate transmembrane transport	99	99	4.40	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0090559	regulation of membrane permeability	370	110	16.43	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0048480	stigma development	296	148	13.14	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0002213	defense response to insect	1,229	273	54.57	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0055080	monoatomic cation homeostasis	12,380	558	549.69	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0008152	metabolic process	445,709	20,142	19,789.96	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0080184	response to phenylpropanoid	361	147	16.03	1,00E-30	1.18e-30	Softcore	Biological processes

GO:0044550	secondary metabolite biosynthetic process	20,377	1,677	904.76	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0018874	benzoate metabolic process	361	144	16.03	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0010198	synergid death	148	98	6.57	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0009694	jasmonic acid metabolic process	2,114	347	93.86	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0009058	biosynthetic process	243,422	10,402	10,808.20	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0046482	para-aminobenzoic acid metabolic process	439	144	19.49	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0044070	regulation of monoatomic anion transport	1,440	146	63.94	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0010043	response to zinc ion	3,128	388	138.89	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0033306	phytol metabolic process	575	151	25.53	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0030243	cellulose metabolic process	3,050	200	135.42	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0032259	methylation	13,777	961	611.71	1,00E-30	1.18e-30	Softcore	Biological processes
GO:1905933	regulation of cell fate determination	50	50	2.22	1,00E-30	1.18e-30	Softcore	Biological processes
GO:0031407	oxylin metabolic process	407	98	18.07	1,00E-30	1.18e-30	Softcore	Biological processes
GO:1903224	regulation of endodermal cell differentiation	1,221	1,034	345.00	1,00E-30	1.18e-30	Shell	Biological processes
GO:0031347	regulation of defense response	19,309	8,123	5,455.92	1,00E-30	1.18e-30	Shell	Biological processes
GO:0007165	signal transduction	93,573	30,919	26,439.82	1,00E-30	1.18e-30	Shell	Biological processes
GO:0007166	cell surface receptor signaling pathway	19,912	7,402	5,626.30	1,00E-30	1.18e-30	Shell	Biological processes
GO:0044550	secondary metabolite biosynthetic process	20,377	9,766	5,757.69	1,00E-30	1.18e-30	Shell	Biological processes
GO:0071446	cellular response to salicylic acid stimulus	5,714	2,663	1,614.54	1,00E-30	1.18e-30	Shell	Biological processes
GO:0072659	protein localization to plasma membrane	2,536	1,804	716.57	1,00E-30	1.18e-30	Shell	Biological processes
GO:0006468	protein phosphorylation	53,039	22,621	14,986.60	1,00E-30	1.18e-30	Shell	Biological processes
GO:0045088	regulation of innate immune response	8,378	3,243	2,367.27	1,00E-30	1.18e-30	Shell	Biological processes
GO:0051762	sesquiterpene biosynthetic process	1,394	1,342	393.89	1,00E-30	1.18e-30	Shell	Biological processes
GO:0006182	cGMP biosynthetic process	2,892	2,430	817.16	1,00E-30	1.18e-30	Shell	Biological processes
GO:0009311	oligosaccharide metabolic process	4,763	1,611	1,345.82	1,00E-30	1.18e-30	Shell	Biological processes
GO:0010073	meristem maintenance	7,733	2,392	2,185.02	1,00E-30	1.18e-30	Shell	Biological processes
GO:0045199	maintenance of epithelial cell apical/basal polarity	1,528	1,271	431.75	1,00E-30	1.18e-30	Shell	Biological processes
GO:0002229	defense response to oomycetes	6,230	3,603	1,760.34	1,00E-30	1.18e-30	Shell	Biological processes
GO:0009751	response to salicylic acid	17,562	8,804	4,962.29	1,00E-30	1.18e-30	Shell	Biological processes
GO:0045184	establishment of protein localization	23,134	5,849	6,536.70	1,00E-30	1.18e-30	Shell	Biological processes
GO:0009725	response to hormone	95,724	29,349	27,047.60	1,00E-30	1.18e-30	Shell	Biological processes
GO:0046677	response to antibiotic	24,217	10,528	6,842.71	1,00E-30	1.18e-30	Shell	Biological processes
GO:0010470	regulation of gastrulation	1,221	1,034	345.00	1,00E-30	1.18e-30	Shell	Biological processes

GO:0006335	DNA replication-dependent chromatin assembly	395	190	6.27	1,00E-30	1.18e-30	Private	Biological processes
GO:0051290	protein heterotetramerization	466	190	7.40	1,00E-30	1.18e-30	Private	Biological processes
GO:0006334	nucleosome assembly	985	191	15.64	1,00E-30	1.18e-30	Private	Biological processes
GO:0010731	protein glutathionylation	778	172	12.35	1,00E-30	1.18e-30	Private	Biological processes
GO:0071103	DNA conformation change	7,063	400	112.16	1,00E-30	1.18e-30	Private	Biological processes
GO:0071446	cellular response to salicylic acid stimulus	5,714	223	90.74	1,00E-30	1.18e-30	Private	Biological processes
GO:0010468	regulation of gene expression	106,623	1,579	1,693.17	1,00E-30	1.18e-30	Private	Biological processes
GO:0007165	signal transduction	93,573	1,693	1,485.94	1,00E-30	1.18e-30	Private	Biological processes
GO:0090204	protein localization to nuclear pore	863	113	13.70	1,00E-30	1.18e-30	Private	Biological processes
GO:1904046	negative regulation of vascular endothelial growth factor production	863	113	13.70	1,00E-30	1.18e-30	Private	Biological processes
GO:0002807	positive regulation of antimicrobial peptide biosynthetic process	863	113	13.70	1,00E-30	1.18e-30	Private	Biological processes
GO:0042307	positive regulation of protein import into nucleus	971	117	15.42	1,00E-30	1.18e-30	Private	Biological processes
GO:0006749	glutathione metabolic process	3,715	218	58.99	1,00E-30	1.18e-30	Private	Biological processes
GO:1904591	positive regulation of protein import	1,061	117	16.85	1,00E-30	1.18e-30	Private	Biological processes
GO:0006518	peptide metabolic process	17,608	582	279.62	1,00E-30	1.18e-30	Private	Biological processes
GO:0015977	carbon fixation	1,043	81	16.56	1,00E-30	1.18e-30	Private	Biological processes
GO:0070417	cellular response to cold	3,914	190	62.15	1,00E-30	1.18e-30	Private	Biological processes
GO:0034250	positive regulation of amide metabolic process	2,298	137	36.49	1,00E-30	1.18e-30	Private	Biological processes
GO:0006952	defense response	67,174	1,538	1,066.72	1,00E-30	1.18e-30	Private	Biological processes
GO:0002230	positive regulation of defense response to virus by host	942	84	14.96	1,00E-30	1.18e-30	Private	Biological processes

Supplementary Table 15 Enriched Cellular Component GO (gene ontology) of pangenome core genes, softcore genes, shell genes and private genes generated with TopGO⁸³. The enriched GO categories were determined using Fisher’s exact test, followed by the Benjamin–Hochberg correction to obtain adjusted *P* values for multiple testing. Also see **Supplementary Figure 10**.

GO ID	GO Term	Annotated	Significant	Expected	Fisher.p	Adjust.p	Pangenome	GO category
GO:0005634	nucleus	176,593	130,609	117,645.87	1,00E-30	1.18E-30	Core	Cellular component
GO:0031984	organelle subcompartment	45,616	34,772	30,389.28	1,00E-30	1.18E-30	Core	Cellular component
GO:0009535	chloroplast thylakoid membrane	11,952	9,418	7,962.40	1,00E-30	1.18E-30	Core	Cellular component
GO:0000139	Golgi membrane	6,497	5,310	4,328.29	1,00E-30	1.18E-30	Core	Cellular component
GO:0000785	chromatin	7,337	5,911	4,887.89	1,00E-30	1.18E-30	Core	Cellular component
GO:0005667	transcription regulator complex	6,070	4,974	4,043.82	1,00E-30	1.18E-30	Core	Cellular component
GO:0005802	trans-Golgi network	11,474	8,987	7,643.95	1,00E-30	1.18E-30	Core	Cellular component
GO:0010319	stromule	1,336	1,247	890.04	1,00E-30	1.18E-30	Core	Cellular component
GO:0048471	perinuclear region of cytoplasm	2,070	1,869	1,379.03	1,00E-30	1.18E-30	Core	Cellular component
GO:0005622	intracellular anatomical structure	496,704	345,259	330,903.12	1,00E-30	1.18E-30	Core	Cellular component
GO:1990391	DNA repair complex	715	699	476.33	1,00E-30	1.18E-30	Core	Cellular component
GO:0009941	chloroplast envelope	25,301	19,003	16,855.47	1,00E-30	1.18E-30	Core	Cellular component
GO:0031901	early endosome membrane	541	539	360.41	1,00E-30	1.18E-30	Core	Cellular component
GO:0005876	spindle microtubule	933	925	621.56	1,00E-30	1.18E-30	Core	Cellular component
GO:0031201	SNARE complex	2,012	1,735	1,340.39	1,00E-30	1.18E-30	Core	Cellular component
GO:0009570	chloroplast stroma	31,743	23,111	21,147.12	1,00E-30	1.18E-30	Core	Cellular component
GO:0080008	Cul4-RING E3 ubiquitin ligase complex	2,772	2,325	1,846.70	1,00E-30	1.18E-30	Core	Cellular component
GO:0009534	chloroplast thylakoid	15,108	12,106	10,064.92	1,00E-30	1.18E-30	Core	Cellular component
GO:0070062	extracellular exosome	491	490	327.10	1,00E-30	1.18E-30	Core	Cellular component
GO:0098800	inner mitochondrial membrane protein complex	2,935	2,506	1,955.29	1,00E-30	1.18E-30	Core	Cellular component
GO:0005839	proteasome core complex	957	148	42.75	1,00E-30	1.18E-30	Softcore	Cellular component
GO:0031463	Cul3-RING ubiquitin ligase complex	522	126	23.32	1,00E-30	1.18E-30	Softcore	Cellular component
GO:0005930	axoneme	396	98	17.69	1,00E-30	1.18E-30	Softcore	Cellular component
GO:0071818	BAT3 complex	99	49	4.42	1,00E-30	1.18E-30	Softcore	Cellular component
GO:0043198	dendritic shaft	104	50	4.65	1,00E-30	1.18E-30	Softcore	Cellular component
GO:0043186	P granule	113	49	5.05	1,00E-30	1.18E-30	Softcore	Cellular component
GO:0031515	tRNA (m1A) methyltransferase complex	95	45	4.24	1,00E-30	1.18E-30	Softcore	Cellular component

GO:0005912	adherens junction	150	50	6.70	7.10E-30	8.35E-30	Softcore	Cellular component
GO:0010287	plastoglobule	3,317	298	148.19	3.60E-29	4.21E-29	Softcore	Cellular component
GO:0034426	etioplast membrane	149	49	6.66	5.30E-29	6.17E-29	Softcore	Cellular component
GO:0005850	eukaryotic translation initiation factor 2 complex	150	49	6.70	7.50E-29	8.65E-29	Softcore	Cellular component
GO:0072546	EMC complex	150	49	6.70	7.50E-29	8.65E-29	Softcore	Cellular component
GO:0005652	nuclear lamina	157	49	7.01	8,00E-28	9.19E-28	Softcore	Cellular component
GO:0005642	annulate lamellae	185	49	8.27	2.90E-24	3.27E-24	Softcore	Cellular component
GO:0070860	RNA polymerase I core factor complex	89	35	3.98	3.50E-24	3.91E-24	Softcore	Cellular component
GO:0005668	RNA polymerase transcription factor SL1 complex	89	35	3.98	3.50E-24	3.91E-24	Softcore	Cellular component
GO:0031011	Ino80 complex	189	49	8.44	8.20E-24	9.07E-24	Softcore	Cellular component
GO:0034245	mitochondrial DNA-directed RNA polymerase complex	189	49	8.44	8.20E-24	9.07E-24	Softcore	Cellular component
GO:0071007	U2-type catalytic step 2 spliceosome	562	87	25.11	1.20E-23	1.32E-23	Softcore	Cellular component
GO:0070937	CRD-mediated mRNA stability complex	200	50	8.94	1.70E-23	1.86E-23	Softcore	Cellular component
GO:0016323	basolateral plasma membrane	2,362	1,813	653.66	1,00E-30	1.18E-30	Shell	Cellular component
GO:0016020	membrane	256,391	86,215	70,953.21	1,00E-30	1.18E-30	Shell	Cellular component
GO:0005886	plasma membrane	160,044	60,748	44,290.30	1,00E-30	1.18E-30	Shell	Cellular component
GO:0031430	M band	2,055	1,793	568.70	1,00E-30	1.18E-30	Shell	Cellular component
GO:0014731	spectrin-associated cytoskeleton	1,530	1,272	423.41	1,00E-30	1.18E-30	Shell	Cellular component
GO:0005694	chromosome	21,513	6,910	5,953.47	1,00E-30	1.18E-30	Shell	Cellular component
GO:0048226	Casparian strip	1,751	1,206	484.57	1,00E-30	1.18E-30	Shell	Cellular component
GO:0005796	Golgi lumen	509	493	140.86	1,00E-30	1.18E-30	Shell	Cellular component
GO:0031012	extracellular matrix	1,030	807	285.04	1,00E-30	1.18E-30	Shell	Cellular component
GO:0016272	prefoldin complex	1,029	790	284.76	1,00E-30	1.18E-30	Shell	Cellular component
GO:0005640	nuclear outer membrane	1,068	776	295.56	1,00E-30	1.18E-30	Shell	Cellular component
GO:0035649	Nrd1 complex	874	664	241.87	1,00E-30	1.18E-30	Shell	Cellular component
GO:0044613	nuclear pore central transport channel	1,218	821	337.07	1,00E-30	1.18E-30	Shell	Cellular component
GO:0042589	zymogen granule membrane	608	493	168.26	1,00E-30	1.18E-30	Shell	Cellular component
GO:0043034	costamere	686	532	189.84	1,00E-30	1.18E-30	Shell	Cellular component
GO:0043661	peribacteroid membrane	484	405	133.94	1,00E-30	1.18E-30	Shell	Cellular component
GO:0030896	checkpoint clamp complex	438	378	121.21	1,00E-30	1.18E-30	Shell	Cellular component
GO:0005773	vacuole	51,600	16,121	14,279.70	1,00E-30	1.18E-30	Shell	Cellular component
GO:0005576	extracellular region	27,726	9,815	7,672.85	1,00E-30	1.18E-30	Shell	Cellular component
GO:0044228	host cell surface	239	234	66.14	1,00E-30	1.18E-30	Shell	Cellular component

GO:0000786	nucleosome	494	190	7.98	1,00E-30	1.18E-30	Private	Cellular component
GO:0005694	chromosome	21,513	710	347.73	1,00E-30	1.18E-30	Private	Cellular component
GO:0005640	nuclear outer membrane	1,068	113	17.26	1,00E-30	1.18E-30	Private	Cellular component
GO:0044613	nuclear pore central transport channel	1,218	114	19.69	1,00E-30	1.18E-30	Private	Cellular component
GO:0000228	nuclear chromosome	9,603	350	155.22	1,00E-30	1.18E-30	Private	Cellular component
GO:0031595	nuclear proteasome complex	652	85	10.54	1,00E-30	1.18E-30	Private	Cellular component
GO:0008541	proteasome regulatory particle, lid subcomplex	817	91	13.21	1,00E-30	1.18E-30	Private	Cellular component
GO:0005654	nucleoplasm	29,829	692	482.15	1,00E-30	1.18E-30	Private	Cellular component
GO:0005688	U6 snRNP	480	53	7.76	1.70E-27	1.94E-27	Private	Cellular component
GO:0000151	ubiquitin ligase complex	9,884	242	159.76	1.50E-24	1.70E-24	Private	Cellular component
GO:0000220	vacuolar proton-transporting V-type ATPase, V0 domain	196	31	3.17	2.40E-21	2.61E-21	Private	Cellular component
GO:0052718	tRNA-specific adenosine-34 deaminase complex	430	39	6.95	1.10E-17	1.14E-17	Private	Cellular component
GO:0072687	meiotic spindle	79	19	1.28	3.20E-17	3.30E-17	Private	Cellular component
GO:0014731	spectrin-associated cytoskeleton	1,530	73	24.73	1.30E-15	1.33E-15	Private	Cellular component
GO:0046540	U4/U6 x U5 tri-snRNP complex	1,003	53	16.21	1.90E-13	1.94E-13	Private	Cellular component
GO:0005737	cytoplasm	376,896	6,314	6,092.01	2.20E-12	2.24E-12	Private	Cellular component
GO:0022625	cytosolic large ribosomal subunit	4,362	134	70.51	5.80E-12	5.87E-12	Private	Cellular component
GO:0031430	M band	2,055	79	33.22	5.90E-12	5.95E-12	Private	Cellular component
GO:0009539	photosystem II reaction center	47	12	0.76	9.80E-12	9.84E-12	Private	Cellular component
GO:0000932	P-body	2,483	89	40.13	1.10E-11	1.10E-11	Private	Cellular component

Supplementary Table 16 Enriched Molecular Function GO (gene ontology) of pangenome core genes, softcore genes, shell genes and private genes generated with TopGO⁸³. The enriched GO categories were determined using Fisher’s exact test, followed by the Benjamin–Hochberg correction to obtain adjusted *P* values for multiple testing. Also see **Supplementary Figure 11**.

GO ID	GO Term	Annotated	Significant	Expected	Fisher.p	Adjust.p	Pangenome	GO category
GO:1901363	heterocyclic compound binding	118,704	86,820	78,588.74	1,00E-30	1.18E-30	Core	Molecular function
GO:0003700	DNA-binding transcription factor activity	61,205	48,828	40,521.16	1,00E-30	1.18E-30	Core	Molecular function
GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	2,593	2,441	1,716.71	1,00E-30	1.18E-30	Core	Molecular function
GO:0000981	DNA-binding transcription factor activity, RNA polymerase II-specific	8,050	7,149	5,329.55	1,00E-30	1.18E-30	Core	Molecular function
GO:0043565	sequence-specific DNA binding	30,718	23,920	20,337.05	1,00E-30	1.18E-30	Core	Molecular function
GO:0001067	transcription regulatory region nucleic acid binding	18,480	15,070	12,234.80	1,00E-30	1.18E-30	Core	Molecular function
GO:0003682	chromatin binding	6,163	4,904	4,080.25	1,00E-30	1.18E-30	Core	Molecular function
GO:0008134	transcription factor binding	5,321	4,639	3,522.80	1,00E-30	1.18E-30	Core	Molecular function
GO:0042803	protein homodimerization activity	12,424	9,600	8,225.39	1,00E-30	1.18E-30	Core	Molecular function
GO:0003756	protein disulfide isomerase activity	993	978	657.42	1,00E-30	1.18E-30	Core	Molecular function
GO:0003714	transcription corepressor activity	1,363	1,289	902.38	1,00E-30	1.18E-30	Core	Molecular function
GO:0043168	anion binding	22,025	16,265	14,581.79	1,00E-30	1.18E-30	Core	Molecular function
GO:0004842	ubiquitin-protein transferase activity	20,246	15,028	13,403.99	1,00E-30	1.18E-30	Core	Molecular function
GO:0000976	transcription cis-regulatory region binding	5,898	5,181	3,904.81	1,00E-30	1.18E-30	Core	Molecular function
GO:0035091	phosphatidylinositol binding	2,901	2,437	1,920.63	1,00E-30	1.18E-30	Core	Molecular function
GO:0004519	endonuclease activity	26,254	19,111	17,381.63	1,00E-30	1.18E-30	Core	Molecular function
GO:0044877	protein-containing complex binding	11,388	7,186	7,539.50	1,00E-30	1.18E-30	Core	Molecular function
GO:0005096	GTPase activator activity	1,668	1,493	1,104.31	1,00E-30	1.18E-30	Core	Molecular function
GO:0016788	hydrolase activity, acting on ester bonds	54,284	39,162	35,939.07	1,00E-30	1.18E-30	Core	Molecular function
GO:0016671	oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor	1,743	1,552	1,153.96	1,00E-30	1.18E-30	Core	Molecular function
GO:0090422	thiamine pyrophosphate transmembrane transporter activity	99	99	4.59	1,00E-30	1.18E-30	Core	Molecular function
GO:0052694	jasmonoyl-isoleucine-12-hydroxylase activity	296	148	13.71	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0052640	salicylic acid glucosyltransferase (glucoside-forming) activity	361	144	16.72	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0052641	benzoic acid glucosyltransferase activity	361	144	16.72	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0052639	salicylic acid glucosyltransferase (ester-forming) activity	361	144	16.72	1,00E-30	1.18E-30	Softcore	Molecular function

GO:0090704	nicotinate-O-glucosyltransferase activity	361	144	16.72	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0080002	UDP-glucose:4-aminobenzoate acylglucosyltransferase activity	361	144	16.72	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0004518	nuclease activity	29,461	941	1,364.84	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0015415	ATPase-coupled phosphate ion transmembrane transporter activity	191	104	8.85	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0016709	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen	10,407	976	482.12	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0052638	indole-3-butyrate beta-glucosyltransferase activity	210	102	9.73	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0001067	transcription regulatory region nucleic acid binding	18,480	1,178	856.12	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0022821	solute:potassium antiporter activity	896	191	41.51	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0008419	RNA lariat debranching enzyme activity	50	50	2.32	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0009978	allene oxide synthase activity	49	49	2.27	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0044715	8-oxo-dGDP phosphatase activity	49	49	2.27	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0004635	phosphoribosyl-AMP cyclohydrolase activity	49	49	2.27	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0004636	phosphoribosyl-ATP diphosphatase activity	49	49	2.27	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0004164	diphthine synthase activity	49	49	2.27	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0043995	histone H4K5 acetyltransferase activity	49	49	2.27	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0080013	(E,E)-geranylinalool synthase activity	1,227	1,161	342.97	1,00E-30	1.18E-30	Softcore	Molecular function
GO:0030507	spectrin binding	2,110	1,798	589.79	1,00E-30	1.18E-30	Shell	Molecular function
GO:0004383	guanylate cyclase activity	2,892	2,430	808.38	1,00E-30	1.18E-30	Shell	Molecular function
GO:0010334	sesquiterpene synthase activity	1,394	1,342	389.65	1,00E-30	1.18E-30	Shell	Molecular function
GO:0034007	S-linalool synthase activity	865	824	241.79	1,00E-30	1.18E-30	Shell	Molecular function
GO:0009378	four-way junction helicase activity	2,082	1,461	581.97	1,00E-30	1.18E-30	Shell	Molecular function
GO:0008093	cytoskeletal anchor activity	1,424	1,266	398.04	1,00E-30	1.18E-30	Shell	Molecular function
GO:0008066	glutamate receptor activity	1,653	1,249	462.05	1,00E-30	1.18E-30	Shell	Molecular function
GO:0004674	protein serine/threonine kinase activity	42,903	19,598	11,992.35	1,00E-30	1.18E-30	Shell	Molecular function
GO:0030246	carbohydrate binding	6,790	3,198	1,897.96	1,00E-30	1.18E-30	Shell	Molecular function
GO:0051117	ATPase binding	2,709	1,962	757.23	1,00E-30	1.18E-30	Shell	Molecular function
GO:0005516	calmodulin binding	12,213	6,443	3,413.81	1,00E-30	1.18E-30	Shell	Molecular function
GO:0016709	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen	10,407	5,747	2,908.99	1,00E-30	1.18E-30	Shell	Molecular function

GO:0001653	peptide receptor activity	2,190	1,665	612.15	1,00E-30	1.18E-30	Shell	Molecular function
GO:0051119	sugar transmembrane transporter activity	5,855	2,977	1,636.60	1,00E-30	1.18E-30	Shell	Molecular function
GO:0080043	quercetin 3-O-glucosyltransferase activity	8,780	4,040	2,454.21	1,00E-30	1.18E-30	Shell	Molecular function
GO:0080044	quercetin 7-O-glucosyltransferase activity	8,780	4,040	2,454.21	1,00E-30	1.18E-30	Shell	Molecular function
GO:0043138	3'-5' DNA helicase activity	2,284	1,466	638.43	1,00E-30	1.18E-30	Shell	Molecular function
GO:0016887	ATP hydrolysis activity	28,177	10,579	7,876.10	1,00E-30	1.18E-30	Shell	Molecular function
GO:0003868	4-hydroxyphenylpyruvate dioxygenase activity	640	602	178.89	1,00E-30	1.18E-30	Shell	Molecular function
GO:0031492	nucleosomal DNA binding	514	192	8.06	1,00E-30	1.18E-30	Shell	Molecular function
GO:0061608	nuclear import signal receptor activity	913	113	14.31	1,00E-30	1.18E-30	Private	Molecular function
GO:0004364	glutathione transferase activity	3,266	204	51.20	1,00E-30	1.18E-30	Private	Molecular function
GO:0017056	structural constituent of nuclear pore	1,658	131	25.99	1,00E-30	1.18E-30	Private	Molecular function
GO:0009378	four-way junction helicase activity	2,082	147	32.64	1,00E-30	1.18E-30	Private	Molecular function
GO:0043138	3'-5' DNA helicase activity	2,284	148	35.81	1,00E-30	1.18E-30	Private	Molecular function
GO:0008964	phosphoenolpyruvate carboxylase activity	800	82	12.54	1,00E-30	1.18E-30	Private	Molecular function
GO:0005516	calmodulin binding	12,213	389	191.48	1,00E-30	1.18E-30	Private	Molecular function
GO:0016887	ATP hydrolysis activity	28,177	668	441.76	1.30E-25	1.48E-25	Private	Molecular function
GO:0000166	nucleotide binding	16,827	337	263.82	8.10E-22	8.84E-22	Private	Molecular function
GO:0051119	sugar transmembrane transporter activity	5,855	206	91.80	4.60E-21	4.97E-21	Private	Molecular function
GO:0016759	cellulose synthase activity	2,326	104	36.47	1.70E-20	1.83E-20	Private	Molecular function
GO:0008515	sucrose transmembrane transporter activity	1,480	78	23.20	1,00E-19	1.07E-19	Private	Molecular function
GO:0080013	(E,E)-geranylinalool synthase activity	1,227	70	19.24	1.20E-19	1.28E-19	Private	Molecular function
GO:0003935	GTP cyclohydrolase II activity	458	42	7.18	1.40E-19	1.49E-19	Private	Molecular function
GO:0031267	small GTPase binding	3,718	138	58.29	1.50E-19	1.59E-19	Private	Molecular function
GO:0004888	transmembrane signaling receptor activity	13,203	242	207.00	9.60E-19	1.01E-18	Private	Molecular function
GO:0009824	AMP dimethylallyltransferase activity	692	50	10.85	1.40E-18	1.47E-18	Private	Molecular function
GO:0008686	3,4-dihydroxy-2-butanone-4-phosphate synthase activity	526	43	8.25	3.90E-18	4.07E-18	Private	Molecular function
GO:0052717	tRNA-specific adenosine-34 deaminase activity	430	39	6.74	4,00E-18	4.16E-18	Private	Molecular function
GO:1901363	heterocyclic compound binding	118,704	86,820	78,588.74	1,00E-30	1.18E-30	Private	Molecular function

Supplementary Table 17 Number of gene feature overlaps with SVs. The Max events represent the maximum number of those features annotated and the theoretical maximum number of overlaps that a single event can occur.

Feature	Max events	Inversions	Duplications (TAG0014_HAP1)	Not aligned (TAG0014_HAP1)	Translocation	Syntenic	Repeat elements
3' UTR	70,751	14,201	12,593	10,904	14,792	61,586	12,963
5' UTR	78,966	15,490	12,502	11,797	15,197	67,998	22,377
2 kbp upstream	40,292	9,513	12,116	14,151	13,372	39,547	36,311
2 kbp downstream	40,291	9,517	12,453	12,886	13,442	39,528	33,029
Gene	40,292	9,564	11,947	10,206	13,012	39,540	29,424
mRNA	67,367	15,025	14,879	12,921	16,920	64,921	48,768
Exon	397,397	55,503	50,295	41,004	58,825	242,232	49,225
CDS	366,385	45,062	43,112	33,719	49,857	195,776	19,774
Introns	330,027	38,010	35,389	26,980	40,914	166,038	35,754
Repeat elements	743,996	200,665	295,844	411,402	323,685	710,161	-

Data availability

The TAG0014 HAP1 and HAP2 genomes are available at <https://phytozome-next.jgi.doe.gov>, along with all assemblies generated in this article. Scripts for genome comparisons, summary statistics and to generate images are available on GitLab: <https://gitlab.com/Anneri/pangenome-of-egr>.

Acknowledgements

The work (proposal: 10.46936/10.25585/60008105) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xml1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. The work conducted at University of Pretoria was supported by the University of Connecticut (UConn) Computational Biology Core for bioinformatics and computational biology support. Mondi and Hans Merensky Timber kindly provided the plant materials used in this study. ALö Acknowledges the University of Pretoria Postgraduate Research Bursary Programme and Sappi South Africa for bursary support. The Forest Sector Innovation Fund (FSIF) for Climate change and risk to the forestry sector in South Africa: approaches to enhance resilience, work package 3 is also acknowledged for career development support. JTL was supported by the Center for Bioenergy Innovation supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research under Contract Number ERKP886.

References

- 1 Brooker, M. I. H. A new classification of the genus *Eucalyptus* L'Hér. (Myrtaceae). *Australian Systematic Botany* **13**, 79–148 (2000). <https://doi.org/10.1071/sb98008>
- 2 Iglesias-Trabado, G. & Wilstermann, D. *Eucalyptus universalis*. *Global cultivated eucalypt forests map 2008*. GIT Forestry Consulting's EUCALYPTOLOGICS: Information Resources on Eucalyptus Cultivation Worldwide., <<https://www.git-forestry.com/>> (2008).
- 3 Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on earth. *Proc Natl Acad Sci U S A* **115**, 6506–6511 (2018). <https://doi.org/10.1073/pnas.1711842115>
- 4 Lötter, A. *et al.* A haplotype-resolved reference genome for *Eucalyptus grandis*. *G3 (Bethesda)* (2025). <https://doi.org/10.1093/g3journal/jkaf112>
- 5 Myburg, A. A. *et al.* The genome of *Eucalyptus grandis*. *Nature* **510**, 356–362 (2014). <https://doi.org/10.1038/nature13308>
- 6 Silva-Junior, O. B. & Grattapaglia, D. Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. *New Phytol* **208**, 830–845 (2015). <https://doi.org/10.1111/nph.13505>
- 7 Mostert-O'Neill, M. M. *et al.* Genomic evidence of introgression and adaptation in a model subtropical tree species, *Eucalyptus grandis*. *Mol Ecol* **30**, 625–638 (2021). <https://doi.org/10.1111/mec.15615>
- 8 Harwood. in *Developing a eucalypt resource: learning from Australia and elsewhere*. 43–54.
- 9 Lötter, A. *et al.* Haplogenome assembly reveals structural variation in *Eucalyptus* interspecific hybrids. *Gigascience* **12**, giad064 (2023). <https://doi.org/10.1093/gigascience/giad064>
- 10 Ferguson, S., Jones, A., Murray, K., Schwessinger, B. & Borevitz, J. O. Interspecies genome divergence is predominantly due to frequent small scale rearrangements in *Eucalyptus*. *Mol Ecol* (2022). <https://doi.org/10.1111/mec.16608>
- 11 Ferguson, S. *et al.* Plant genome evolution in the genus *Eucalyptus* is driven by structural rearrangements that promote sequence divergence. *Genome Res* **34**, 606–619 (2024). <https://doi.org/10.1101/gr.277999.123>
- 12 Ferguson, S., Bar-Ness, Y. D., Borevitz, J. & Jones, A. A telomere-to-telomere *Eucalyptus regnans* genome: unveiling haplotype variance in structure and genes within one of the world's tallest trees. *BMC Genomics* **25**, 913 (2024). <https://doi.org/10.1186/s12864-024-10810-4>
- 13 Zhuang, Z. *et al.* Complex genomic structural variation underlies climate adaptation among *Eucalyptus* species. *bioRxiv* (2026). <https://doi.org/10.64898/2025.12.18.695066>
- 14 Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**, 30 (2020). <https://doi.org/10.1186/s13059-020-1935-5>
- 15 Danilevicz, M. F., Tay Fernandez, C. G., Marsh, J. I., Bayer, P. E. & Edwards, D. Plant pangenomics: approaches, applications and advancements. *Curr Opin Plant Biol* **54**, 18–25 (2020). <https://doi.org/10.1016/j.pbi.2019.12.005>
- 16 Golicz, A. A., Bayer, P. E., Bhalla, P. L., Batley, J. & Edwards, D. Pangenomics comes of age: from bacteria to plant and animal applications. *Trends Genet* **36**, 132–145 (2020). <https://doi.org/10.1016/j.tig.2019.11.006>
- 17 Khan, A. W. *et al.* Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends Plant Sci* **25**, 148–158 (2020). <https://doi.org/10.1016/j.tplants.2019.10.012>
- 18 Zanini, S. F. *et al.* Pangenomics in crop improvement-from coding structural variations to finding regulatory variants with pangenome graphs. *Plant Genome*, e20177 (2021). <https://doi.org/10.1002/tpg2.20177>

- 19 Gasparini, K., Moreira, J. D. R., Peres, L. E. P. & Zsogon, A. *De novo* domestication of wild species to create crops with increased resilience and nutritional value. *Curr Opin Plant Biol* **60**, 102006 (2021). <https://doi.org/10.1016/j.pbi.2021.102006>
- 20 Ebler, J. *et al.* Pangenome-based genome inference allows efficient and accurate genotyping across a wide spectrum of variant classes. *Nat Genet* **54**, 518–525 (2022). <https://doi.org/10.1038/s41588-022-01043-w>
- 21 Schreiber, M., Jayakodi, M., Stein, N. & Mascher, M. Plant pangenomes for crop improvement, biodiversity and evolution. *Nat Rev Genet* (2024). <https://doi.org/10.1038/s41576-024-00691-4>
- 22 Kaur, H., Shannon, L. M. & Samac, D. A. A stepwise guide for pangenome development in crop plants: an alfalfa (*Medicago sativa*) case study. *BMC Genomics* **25**, 1022 (2024). <https://doi.org/10.1186/s12864-024-10931-w>
- 23 Secomandi, S. *et al.* Pangenome graphs and their applications in biodiversity genomics. *Nat Genet* **57**, 13–26 (2025). <https://doi.org/10.1038/s41588-024-02029-6>
- 24 Hu, H., Zhao, J., Thomas, W. J. W., Batley, J. & Edwards, D. The role of pangenomics in orphan crop improvement. *Nat Commun* **16**, 118 (2025). <https://doi.org/10.1038/s41467-024-55260-4>
- 25 Tay Fernandez, C. G. *et al.* Expanding gene-editing potential in crop improvement with pangenomes. *Int J Mol Sci* **23** (2022). <https://doi.org/10.3390/ijms23042276>
- 26 Bevan, M. W. *et al.* Genomic innovation for crop improvement. *Nature* **543**, 346–354 (2017). <https://doi.org/10.1038/nature22011>
- 27 Garg, V. *et al.* Unlocking plant genetics with telomere-to-telomere genome assemblies. *Nat Genet* (2024). <https://doi.org/10.1038/s41588-024-01830-7>
- 28 Morgante, M., De Paoli, E. & Radovic, S. Transposable elements and the plant pan-genomes. *Curr Opin Plant Biol* **10**, 149–155 (2007). <https://doi.org/10.1016/j.pbi.2007.02.001>
- 29 Mascher, M., Jayakodi, M., Shim, H. & Stein, N. Promises and challenges of crop translational genomics. *Nature* (2024). <https://doi.org/10.1038/s41586-024-07713-5>
- 30 Jayakodi, M. *et al.* Structural variation in the pangenome of wild and domesticated barley. *Nature* **636**, 654–662 (2024). <https://doi.org/10.1038/s41586-024-08187-1>
- 31 Ni, L. & Tian, Z. Toward *cis*-regulation in soybean: a 3D genome scope. *Molecular Breeding* **43** (2023). <https://doi.org/10.1007/s11032-023-01374-4>
- 32 He, X. *et al.* Genetic-epigenetic interplay in the determination of plant 3D genome organization. *Nucleic Acids Res* **52**, 10220–10234 (2024). <https://doi.org/10.1093/nar/gkae690>
- 33 Zhang, K. *et al.* Transposon proliferation drives genome architecture and regulatory evolution in wild and domesticated peppers. *Nat Plants* **11**, 359–375 (2025). <https://doi.org/10.1038/s41477-025-01905-1>
- 34 Alonge, M. *et al.* Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**, 145–161 e123 (2020). <https://doi.org/10.1016/j.cell.2020.05.021>
- 35 Benoit, M. *et al.* *Solanum* pan-genetics reveals paralogues as contingencies in crop engineering. *Nature* **640**, 135–145 (2025). <https://doi.org/10.1038/s41586-025-08619-6>
- 36 Wang, X. *et al.* Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nat Commun* **11**, 5817 (2020). <https://doi.org/10.1038/s41467-020-19682-0>
- 37 Dominguez, M. *et al.* The impact of transposable elements on tomato diversity. *Nat Commun* **11**, 4058 (2020). <https://doi.org/10.1038/s41467-020-17874-2>
- 38 Zhou, Y. *et al.* Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* (2022). <https://doi.org/10.1038/s41586-022-04808-9>
- 39 Li, N. *et al.* Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat Genet* (2023). <https://doi.org/10.1038/s41588-023-01340-y>

- 40 Jiang, M. *et al.* PlantPan: A comprehensive multi-species plant pan-genome database. *Plant J* **122**, e70144 (2025). <https://doi.org/10.1111/tpj.70144>
- 41 Yang, Q. *et al.* High-quality population-specific haplotype-resolved reference panel in the genomic and pangenomic eras. *Genomics Proteomics Bioinformatics* (2025). <https://doi.org/10.1093/gpbjnl/qzaf022>
- 42 Shi, T. *et al.* The super-pangenome of *Populus* unveils genomic facets for its adaptation and diversification in widespread forest trees. *Mol Plant* **17**, 725–746 (2024). <https://doi.org/10.1016/j.molp.2024.03.009>
- 43 Wei, S., Yang, Y. & Yin, T. The chromosome-scale assembly of the willow genome provides insight into Salicaceae genome evolution. *Hortic Res* **7**, 45 (2020). <https://doi.org/10.1038/s41438-020-0268-6>
- 44 Shi, T. L. *et al.* High-quality genome assembly enables prediction of allele-specific gene expression in hybrid poplar. *Plant Physiol* **195**, 652–670 (2024). <https://doi.org/10.1093/plphys/kiae078>
- 45 Ferguson, S. *et al.* Exploring the role of polymorphic interspecies structural variants in reproductive isolation and adaptive divergence in *Eucalyptus*. *Gigascience* **13** (2024). <https://doi.org/10.1093/gigascience/giae029>
- 46 Feng, J. *et al.* Integrating evolutionary genomics of forest trees to inform future tree breeding amid rapid climate change. *Plant Commun*, 101044 (2024). <https://doi.org/10.1016/j.xplc.2024.101044>
- 47 Silva-Junior, O. B., Faria, D. A. & Grattapaglia, D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol* **206**, 1527–1540 (2015). <https://doi.org/10.1111/nph.13322>
- 48 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006). <https://doi.org/10.1038/ng1847>
- 49 Aylward, A. J., Petrus, S., Mamerto, A., Hartwick, N. T. & Michael, T. P. PanKmer: k-mer-based and reference-free pangenome analysis. *Bioinformatics* **39** (2023). <https://doi.org/10.1093/bioinformatics/btad621>
- 50 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021). <https://doi.org/10.1038/s41592-020-01056-5>
- 51 Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* **27**, 737–746 (2017). <https://doi.org/10.1101/gr.214270.116>
- 52 Durand, N. C. *et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst* **3**, 99–101 (2016). <https://doi.org/10.1016/j.cels.2015.07.012>
- 53 Astashyn, A. *et al.* Rapid and sensitive detection of genome contamination at scale with FCS-GX. *bioRxiv* (2023). <https://doi.org/10.1101/2023.06.02.543519>
- 54 Zhou, C. *et al.* Oatk: a *de novo* assembly tool for complex plant organelle genomes. *Genome Biol* **26**, 235 (2025). <https://doi.org/10.1186/s13059-025-03676-6>
- 55 Healey, A. L. *et al.* The complex polyploid genome architecture of sugarcane. *Nature* (2024). <https://doi.org/10.1038/s41586-024-07231-4>
- 56 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010). <https://doi.org/10.1093/bioinformatics/btp698>
- 57 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010). <https://doi.org/10.1101/gr.107524.110>
- 58 Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013). <https://doi.org/10.1093/bioinformatics/btt086>

- 59 Mikheenko, A., Prijbelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142–i150 (2018). <https://doi.org/10.1093/bioinformatics/bty266>
- 60 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015). <https://doi.org/10.1093/bioinformatics/btv351>
- 61 Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol Biol* **1962**, 227–245 (2019). https://doi.org/10.1007/978-1-4939-9173-0_14
- 62 Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* **38**, 4647–4654 (2021). <https://doi.org/10.1093/molbev/msab199>
- 63 Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife* **11** (2022). <https://doi.org/10.7554/eLife.78526>
- 64 Shu, S., Goodstein, D. & Rokhsar, D. PERTRAN: Genome-guided RNA-seq read assembler. (Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2013).
- 65 Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010). <https://doi.org/10.1093/bioinformatics/btq057>
- 66 Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654–5666 (2003). <https://doi.org/10.1093/nar/gkg770>
- 67 Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* **117**, 9451–9457 (2020). <https://doi.org/10.1073/pnas.1921046117>
- 68 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014). <https://doi.org/10.1093/bioinformatics/btu031>
- 69 Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412–D419 (2021). <https://doi.org/10.1093/nar/gkaa913>
- 70 Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47**, D419–D426 (2019). <https://doi.org/10.1093/nar/gky1038>
- 71 Smit, A. F. A., Hubley, R. & Green, P. (2013).
- 72 Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005). <https://doi.org/10.1186/1471-2105-6-31>
- 73 Salamov, A. A. & Solovyev, V. V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res* **10**, 516–522 (2000). <https://doi.org/10.1101/gr.10.4.516>
- 74 Stanke, M. *et al.* AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res* **34**, W435–439 (2006). <https://doi.org/10.1093/nar/gkl200>
- 75 Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021). <https://doi.org/10.1093/bioinformatics/btaa1016>
- 76 Caballero, M. & Wegrzyn, J. gFACs: Gene Filtering, Analysis, and Conversion to unify genome annotations across alignment and gene prediction frameworks. *Genomics Proteomics Bioinformatics* **17**, 305–310 (2019). <https://doi.org/10.1016/j.gpb.2019.04.002>
- 77 Marcais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol* **14**, e1005944 (2018). <https://doi.org/10.1371/journal.pcbi.1005944>
- 78 Goel, M. & Schneeberger, K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**, 2922–2926 (2022). <https://doi.org/10.1093/bioinformatics/btac196>

- 79 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018). <https://doi.org/10.1093/bioinformatics/bty191>
- 80 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59–60 (2015). <https://doi.org/10.1038/nmeth.3176>
- 81 Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368 (2021). <https://doi.org/10.1038/s41592-021-01101-x>
- 82 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157 (2015). <https://doi.org/10.1186/s13059-015-0721-2>
- 83 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019). <https://doi.org/10.1186/s13059-019-1832-y>
- 84 Hart, A. J. *et al.* EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Mol Ecol Resour* **20**, 591–604 (2020). <https://doi.org/10.1111/1755-0998.13106>
- 85 Alexa, A. & Rahnenfuhrer, J. topGO: enrichment analysis for gene ontology. *R package version 2*, 2010 (2010). <https://doi.org/https://doi.org/doi:10.18129/B9.bioc.topGO>
- 86 Carocha, V. *et al.* Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *New Phytol* **206**, 1297–1313 (2015). <https://doi.org/10.1111/nph.13313>
- 87 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). <https://doi.org/10.1093/bioinformatics/btq033>
- 88 Mostert-O'Neill, M. M. *et al.* Genomic consequences of artificial selection during early domestication of a wood fibre crop. *New Phytol* (2022). <https://doi.org/10.1111/nph.18297>
- 89 Cochetel, N. *et al.* A super-pangenome of the North American wild grape species. *Genome Biol* **24**, 290 (2023). <https://doi.org/10.1186/s13059-023-03133-2>
- 90 Guo, L. *et al.* Super pangenome of *Vitis* empowers identification of downy mildew resistance genes for grapevine improvement. *Nat Genet* (2025). <https://doi.org/10.1038/s41588-025-02111-7>
- 91 Satterlee, J. W. *et al.* Convergent evolution of plant prickles by repeated gene co-option over deep time. *Science* **385**, eado1663 (2024). <https://doi.org/10.1126/science.adol663>
- 92 Sun, H. *et al.* The phased pan-genome of tetraploid European potato. *Nature* (2025). <https://doi.org/10.1038/s41586-025-08843-0>
- 93 Sreedasyam, A. *et al.* Genome resources for three modern cotton lines guide future breeding efforts. *Nat Plants* **10**, 1039–1051 (2024). <https://doi.org/10.1038/s41477-024-01713-z>
- 94 He, X. *et al.* Pangenome analysis reveals transposon-driven genome evolution in cotton. *BMC Biol* **22**, 92 (2024). <https://doi.org/10.1186/s12915-024-01893-2>
- 95 Li, J. *et al.* Convergence and divergence of diploid and tetraploid cotton genomes. *Nat Genet* **56**, 2562–2573 (2024). <https://doi.org/10.1038/s41588-024-01964-8>
- 96 Chakrabarty, S. & Külheim, C. Trends in tree improvement methods: from classical breeding to genomic technologies. *Tree Genetics & Genomes* **21** (2025). <https://doi.org/10.1007/s11295-025-01698-6>
- 97 Jiao, W. B. & Schneeberger, K. Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat Commun* **11**, 989 (2020). <https://doi.org/10.1038/s41467-020-14779-y>
- 98 Shen, C., Li, L., Ouyang, L., Su, M. & Guo, K. *E. urophylla* x *E. grandis* high-quality genome and comparative genomics provide insights on evolution and diversification of *Eucalyptus*. *BMC Genomics* **24**, 223 (2023). <https://doi.org/10.1186/s12864-023-09318-0>
- 99 Schwartz, C. *et al.* CRISPR-Cas9-mediated 75.5-Mb inversion in maize. *Nat Plants* **6**, 1427–1431 (2020). <https://doi.org/10.1038/s41477-020-00817-6>
- 100 Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16**, 351–358 (2001). [https://doi.org/10.1016/s0169-5347\(01\)02187-5](https://doi.org/10.1016/s0169-5347(01)02187-5)
- 101 Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006). <https://doi.org/10.1534/genetics.105.047985>

- 102 Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol* **8** (2010). <https://doi.org/10.1371/journal.pbio.1000500>
- 103 Saxena, R. K., Edwards, D. & Varshney, R. K. Structural variations in plant genomes. *Brief Funct Genomics* **13**, 296–307 (2014). <https://doi.org/10.1093/bfpg/elu016>
- 104 Kress, W. J. *et al.* Green plant genomes: what we know in an era of rapidly expanding opportunities. *Proc Natl Acad Sci U S A* **119** (2022). <https://doi.org/10.1073/pnas.2115640118>
- 105 Cai, X. *et al.* Transposable element insertion: a hidden major source of domesticated phenotypic variation in *Brassica rapa*. *Plant Biotechnol J* (2022). <https://doi.org/10.1111/pbi.13807>
- 106 Yu, H. & Li, J. Breeding future crops to feed the world through *de novo* domestication. *Nat Commun* **13**, 1171 (2022). <https://doi.org/10.1038/s41467-022-28732-8>
- 107 Li, W. *et al.* Plant pan-genomics: recent advances, new challenges, and roads ahead. *J Genet Genomics* (2022). <https://doi.org/10.1016/j.jgg.2022.06.004>
- 108 Fawcett, J. A. *et al.* Genome sequencing reveals the genetic architecture of heterostyly and domestication history of common buckwheat. *Nat Plants* **9**, 1236–1251 (2023). <https://doi.org/10.1038/s41477-023-01474-1>
- 109 He, Q. *et al.* A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nat Genet* **55**, 1232–1242 (2023). <https://doi.org/10.1038/s41588-023-01423-w>
- 110 Chen, X. *et al.* A wild melon reference genome provides novel insights into the domestication of a key gene responsible for melon fruit acidity. *Theor Appl Genet* **137**, 144 (2024). <https://doi.org/10.1007/s00122-024-04647-4>
- 111 Chen, R. *et al.* Genomic analyses reveal the stepwise domestication and genetic mechanism of curd biogenesis in cauliflower. *Nat Genet* **56**, 1235–1244 (2024). <https://doi.org/10.1038/s41588-024-01744-4>
- 112 Yin, M. *et al.* The haplotype-resolved genome assembly of an ancient citrus variety provides insights into the domestication history and fruit trait formation of loose-skin mandarins. *Genome Biol* **26**, 61 (2025). <https://doi.org/10.1186/s13059-025-03535-4>
- 113 Jang, M. J. *et al.* Haplotype-resolved genome assembly and resequencing analysis provide insights into genome evolution and allelic imbalance in *Pinus densiflora*. *Nat Genet* **56**, 2551–2561 (2024). <https://doi.org/10.1038/s41588-024-01944-y>
- 114 Fang, Y. *et al.* Pan-genome and phylogenomic analyses highlight *Hevea* species delineation and rubber trait evolution. *Nat Commun* **15**, 7232 (2024). <https://doi.org/10.1038/s41467-024-51031-3>