

Supplementary Figures and Tables are available online at the website of the PanGraphRNA project (<https://github.com/cma2015/PanGraphRNA>).

Supplementary Table Legends

Table S1. Data availability of reference genomes.

Table S2. Sample details and alignment results of 54 RNA-seq datasets.

Table S3. Detailed information for 1,050 Arabidopsis population sample.

Table S4. DEGs with ED > 1 between the Asia_graph and SLR methods.

Table S5. Alignment performance of Iberian_graph and Col0_linear genomes in 800 Arabidopsis samples.

Table S6. Details of eQTLs linked to GWAS-related loci.

Table S7. Details of read alignment and error mapping rate of 54 RNA-seq datasets using VG.

Table S8. Details of primers used for RT-qPCR validation.

Supplementary Figure Legends

Figure S1. Comparison of computational consumption for building graph pangenome index files using HISAT2, minigraph (+VG), minigraph-cactus (+VG), and PGGB (+VG).

Figure S2. Intersection of variation information among An1, Cvi0, Ler and Sha accessions.

Figure S3. Comparison of overall alignment rates between graph pangenome and Col0_linear genome methods.

Figure S4. The proportion of reads in variant-containing genomic region based on different graph pangenomes.

Figure S5. Comparison of mapping error rates between graph pangenome and Col0_linear genome methods.

Figure S6. Comparison of quantification results between graph pangenomes and Col0_linear genome methods.

Figure S7. Unique alignment rates and F1 scores of different simulated RNA-seq datasets aligned to Col0_linear and eleven graph pangenomes.

Figure S8. Alignment performance between Iberian_graph and Col0_linear references across 800 accessions.

Figure S9. Comparison of read alignment metrics across 800 accessions using Iberian_graph and Col0_linear reference genomes.

Figure S10. Comparison of alignment performance between graph pangenomes and the IRGSP_linear reference across three rice accessions.

Figure S11. Comparison of RNA-seq alignment performance between B73_linear and 510G_graph genomes across thirteen maize inbred lines.

Figure S12. Generation of simulated RNA-seq reads for six *Arabidopsis* accessions.

Figure S13. Illustration of the definitions for five types of mapping errors.

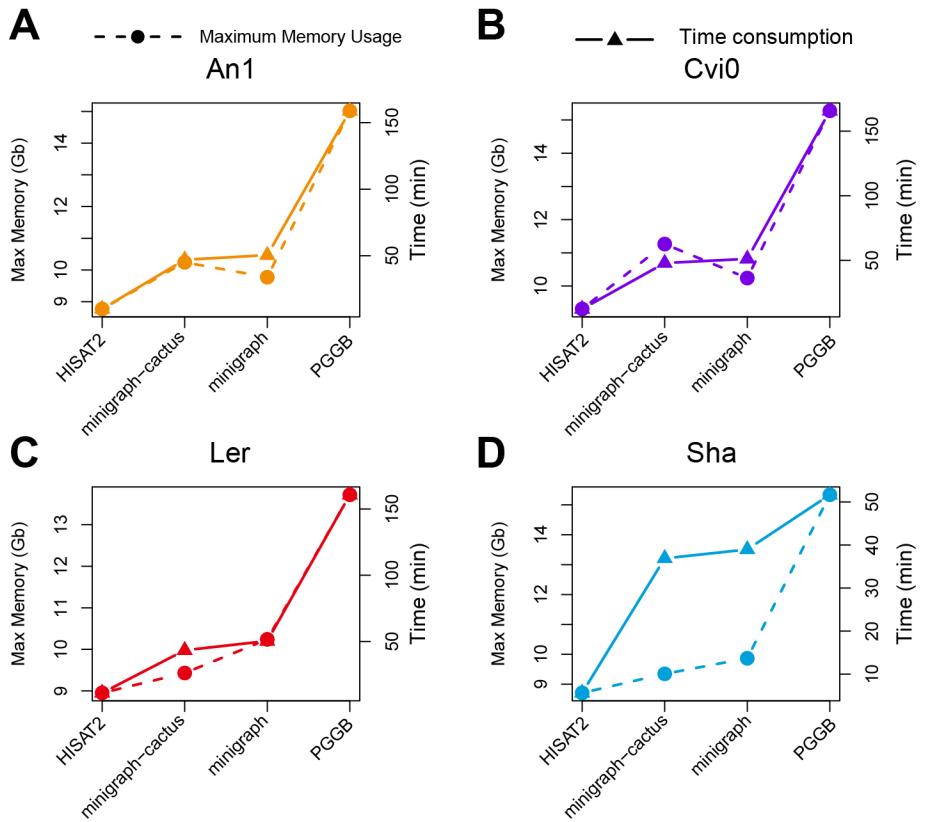


Figure S1. Comparison of computational consumption for building graph pangenome index files using HISAT2, minigraph (+VG), minigraph-cactus (+VG), and PGGB (+VG). (A), (B), (C), and (D) display the memory and runtime consumption for constructing graph pangenomes of An1, Cvi0, Ler, and Sha accessions. The x-axis represents the graph construction tools, and the y-axis indicates the computational resources consumed (memory in GB and time in minutes).

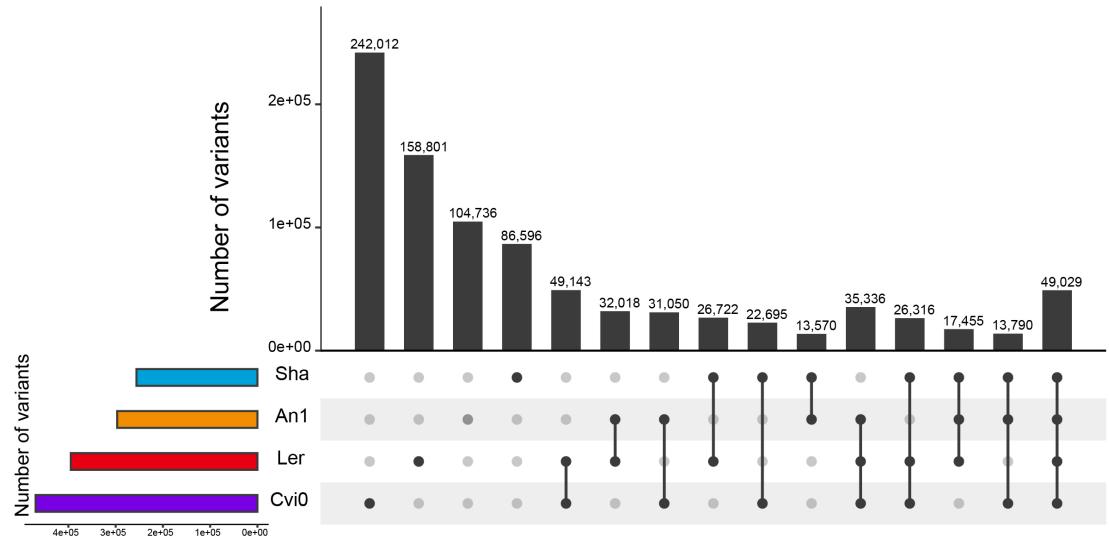


Figure S2. Intersection of variation information among An1, Cvi0, Ler and Sha accessions.

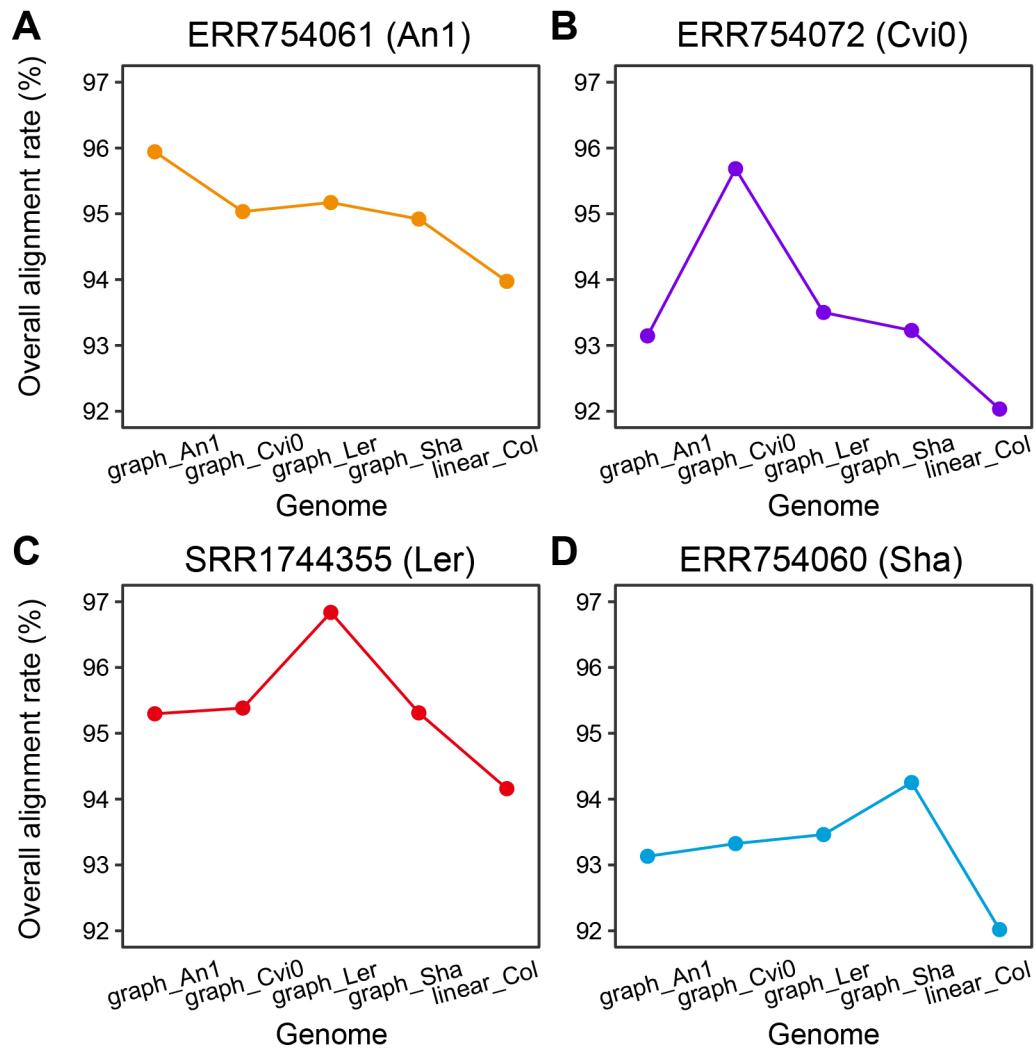


Figure S3. Comparison of overall alignment rates between graph pangenome and Col0_linear genome methods. (A), (B), (C) and (D) display the overall alignment rates of RNA-seq for ERR754061, ERR754072, SRR1744355, and ERR754060 samples, respectively. The x-axis denotes the genome using the reference, and the y-axis indicates the overall alignment rate of RNA-seq reads.

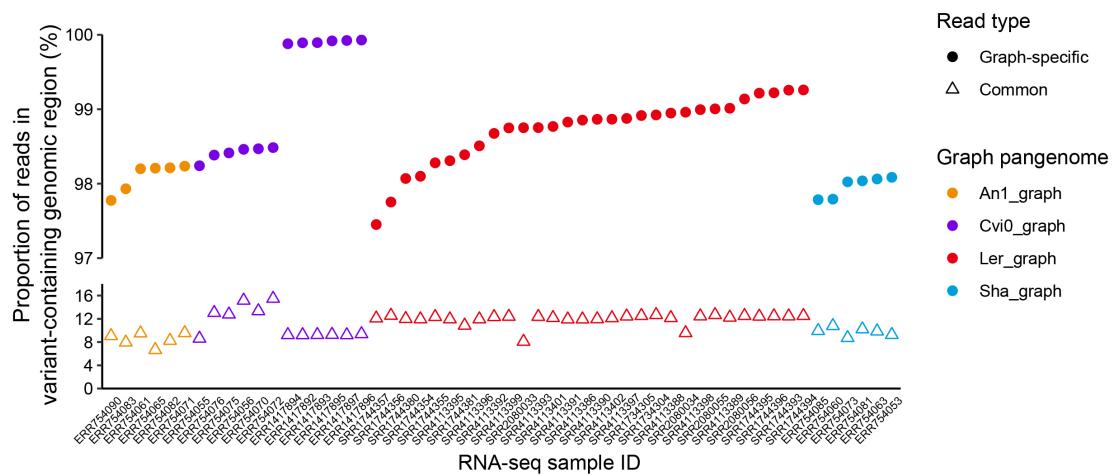


Figure S4. The proportion of reads in variant-containing genomic region based on different graph pangenomes. The plots show the proportion of reads mapped to variant-containing genomic region. This includes reads mapped exclusively by the graph pangenome method ("Graph-specific" reads) and those shared between the graph pangenome and Col0_linear methods ("Common" reads). The x-axis represents RNA-seq samples, the y-axis shows the proportion of reads, with colors indicating different graph pangenome methods and dots representing different read types.

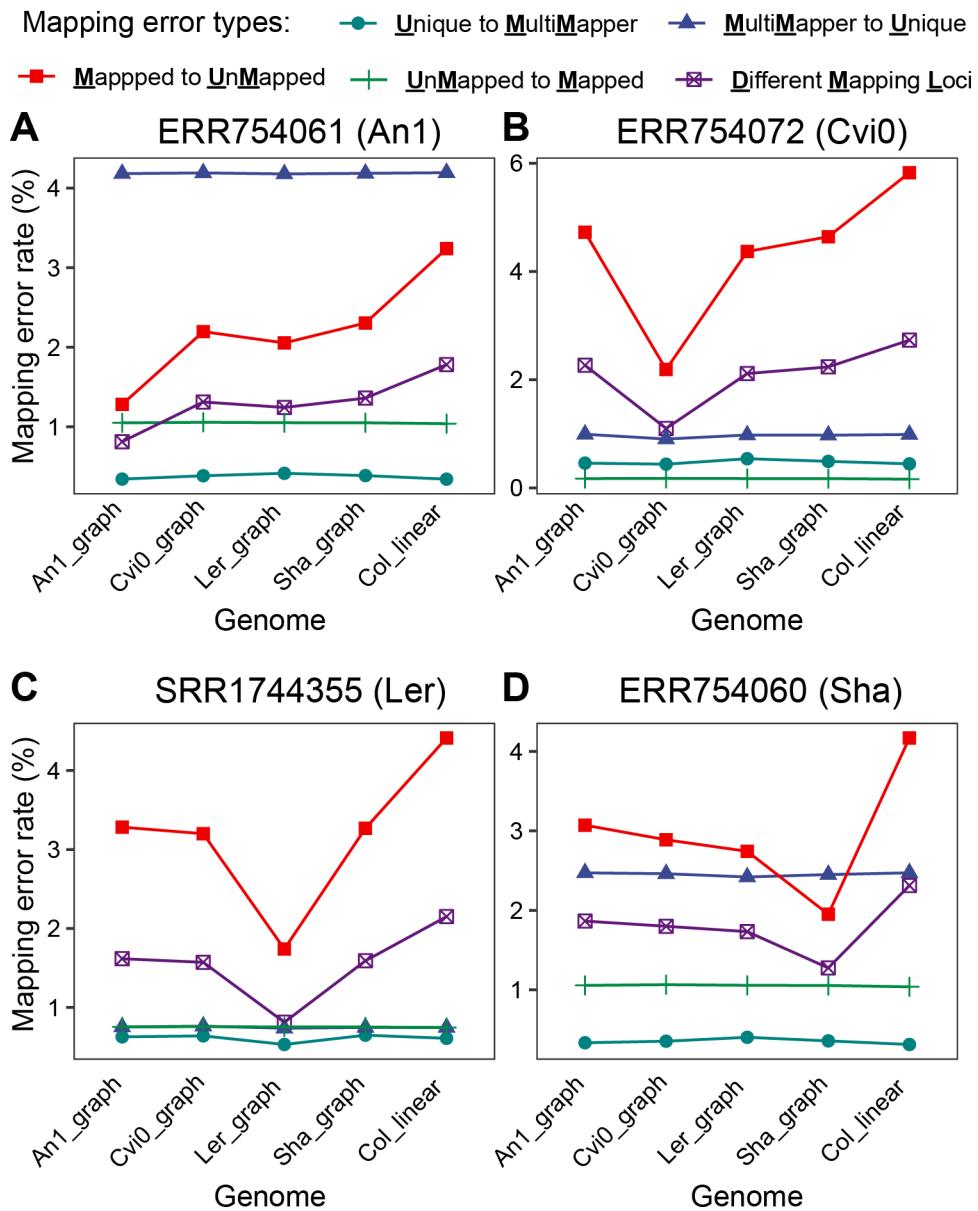


Figure S5. Comparison of mapping error rates between graph pangenome and Col0_linear genome methods. (A), (B), (C) and (D) illustrate the mapping error rate of RNA-seq for ERR754061, ERR754072, SRR1744355, ERR754060 samples. The x-axis represents the genome used as the reference, and the y-axis shows the mapping error rate.

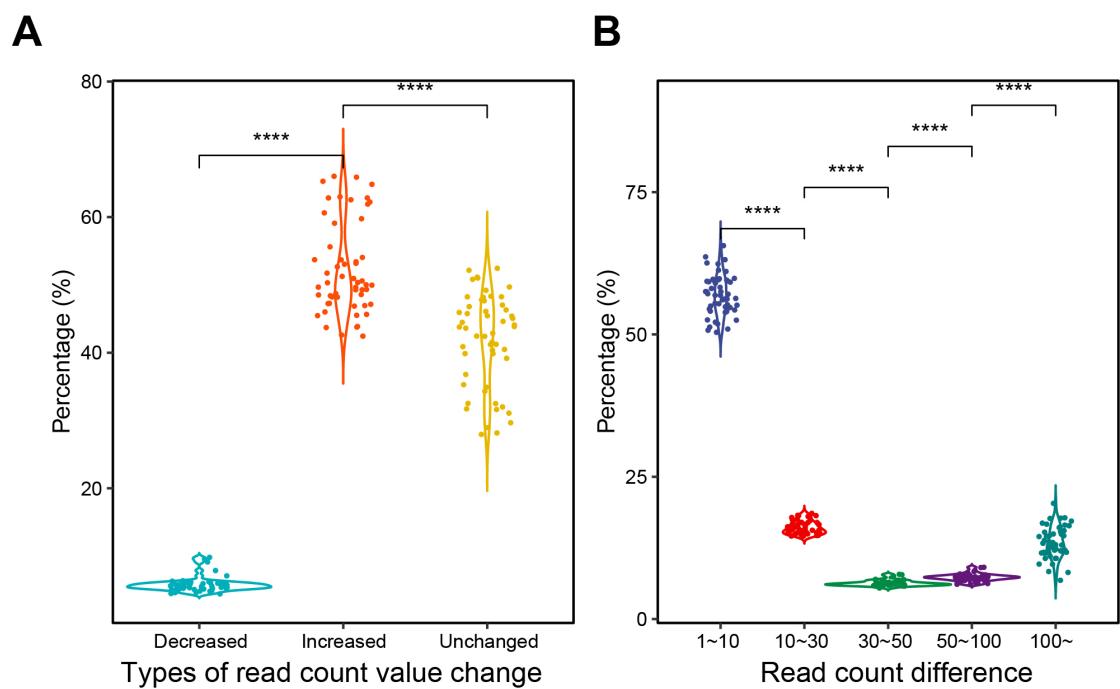


Figure S6. Comparison of quantification results between graph pangenomes and Col0_linear genome methods. **(A)** Proportion of genes with increased, decreased, and unchanged read count values in single-accession-based graph pangenome method compared to Col0_linear method. **(B)** Proportion of genes with differences in read count values between graph pangenome and Col0_linear methods.

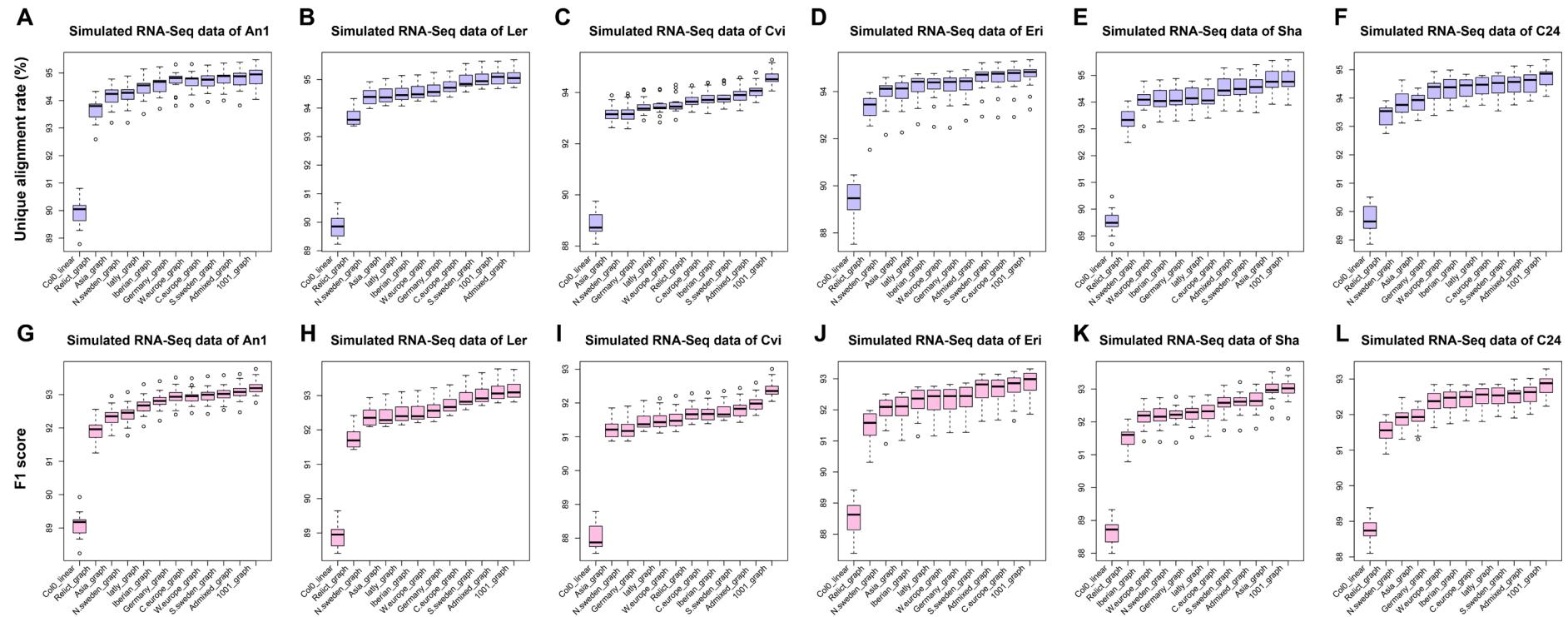


Figure S7. Unique alignment rates and F1 scores of different simulated RNA-seq datasets aligned to Col0_linear and eleven graph pangenomes. Boxplots (A), (B), (C), (D), (E) and (F) represent unique alignment rates of simulated RNA-seq data for An1, Ler, Cvi0, Eri, Sha, and C24, respectively. Boxplots (G), (H), (I), (J), (K) and (L) denote F1 scores of simulated RNA-seq data for An1, Ler, Cvi0, Eri, Sha, and C24, respectively.

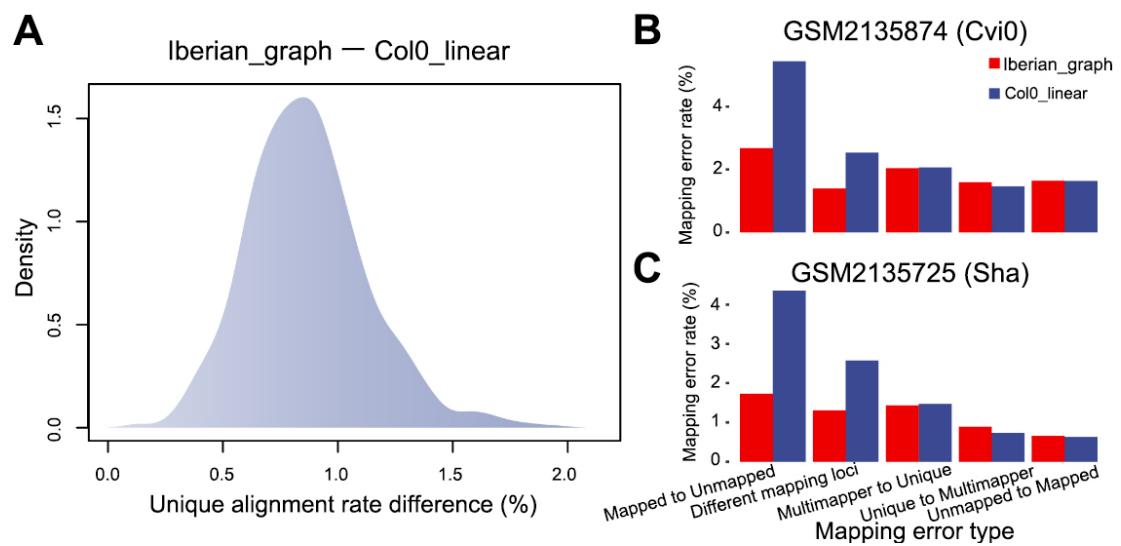


Figure S8. Alignment performance between Iberian_graph and Col0_linear references across 800 accessions. (A) Distribution of differences in unique alignment rate between Iberian_graph and Col0_linear genomes. (B) Mapping error rate for a Cvi0 RNA-seq dataset. (C) Mapping error rate for a Sha RNA-seq dataset.

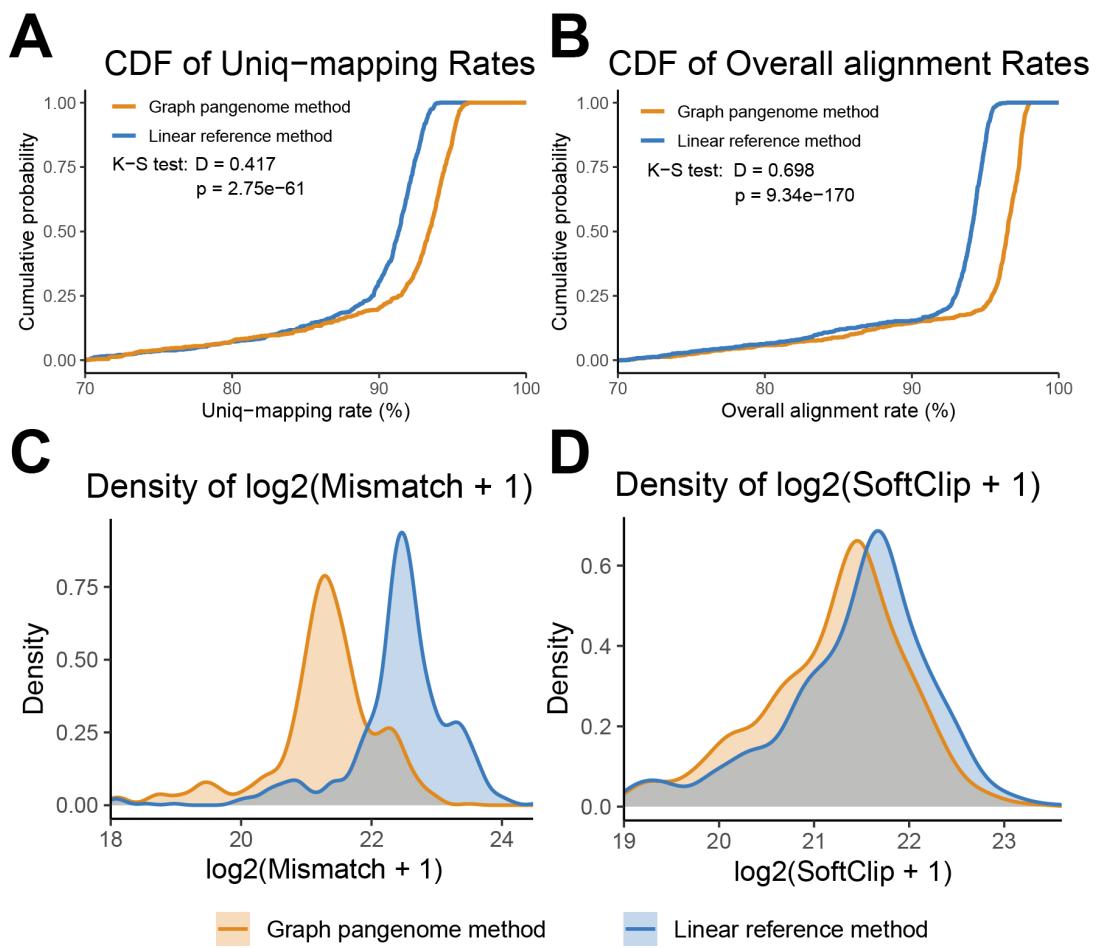


Figure S9. Comparison of read alignment metrics across 800 accessions using Iberian_graph and Col0_linear reference genomes. (A) Cumulative distribution function (CDF) of unique mapping rates. (B) CDF of overall alignment rates. (C) Density distribution of log2-transformed (+1) mismatch bases. (D) Density distribution of log2-transformed (+1) softclip bases.

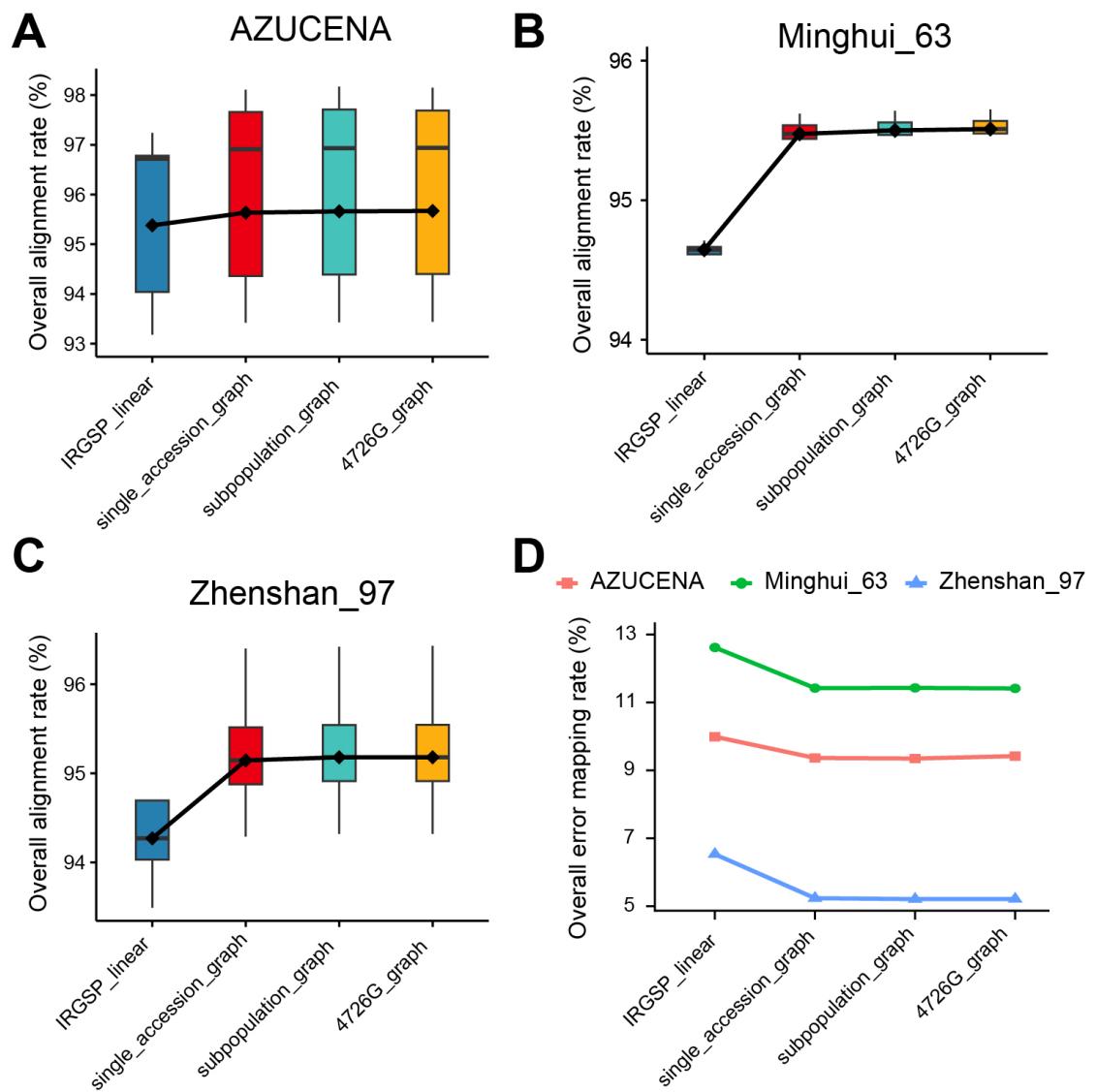


Figure S10. Comparison of alignment performance between graph pangenomes and the IRGSP_linear reference across three rice accessions. (A) Overall alignment rates for the AZUCENA accession using graph pangenomes versus the IRGSP_linear reference. (B) Overall alignment rates for Minghui_63. (C) Overall alignment rates for Zhenshan_97. (D) Overall error mapping rates for three rice accessions.

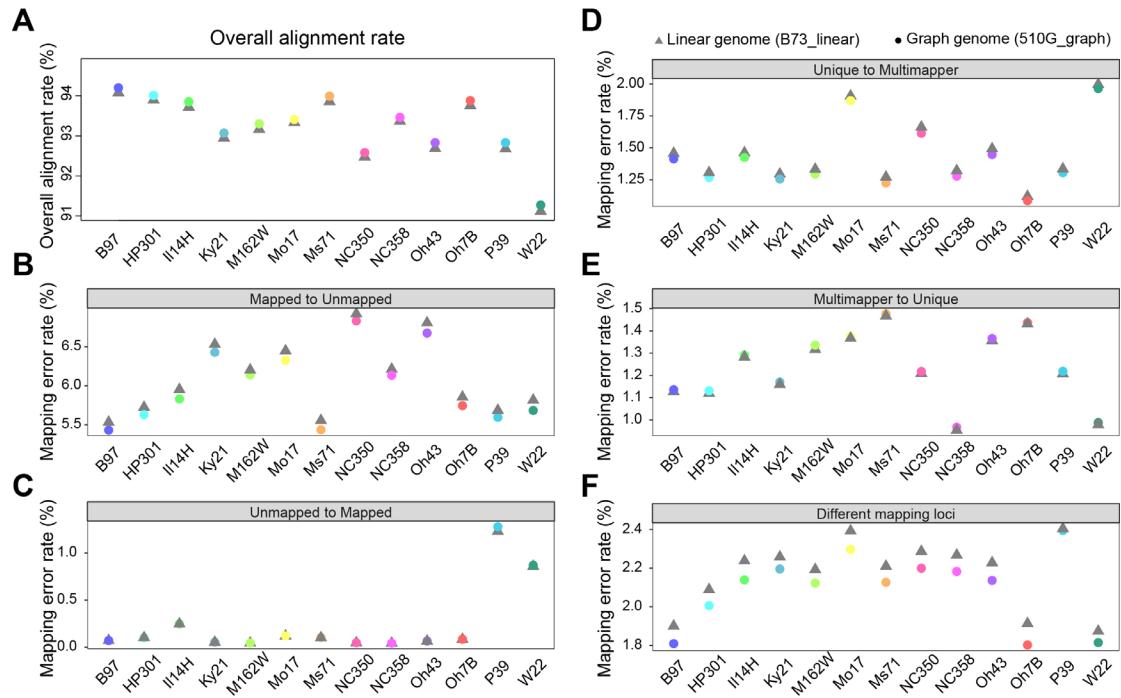


Figure S11. Comparison of RNA-seq alignment performance between B73_linear and 510G_graph genomes across 13 maize inbred lines. **(A)** Overall alignment rates for RNA-seq data from 13 inbred lines when mapped to the B73_linear reference versus the 510G_graph genome. **(B-F)** Error rates (M-UM, UM-M, U-MM, MM-U, DML) for each inbred line when aligned to the 510G_graph.

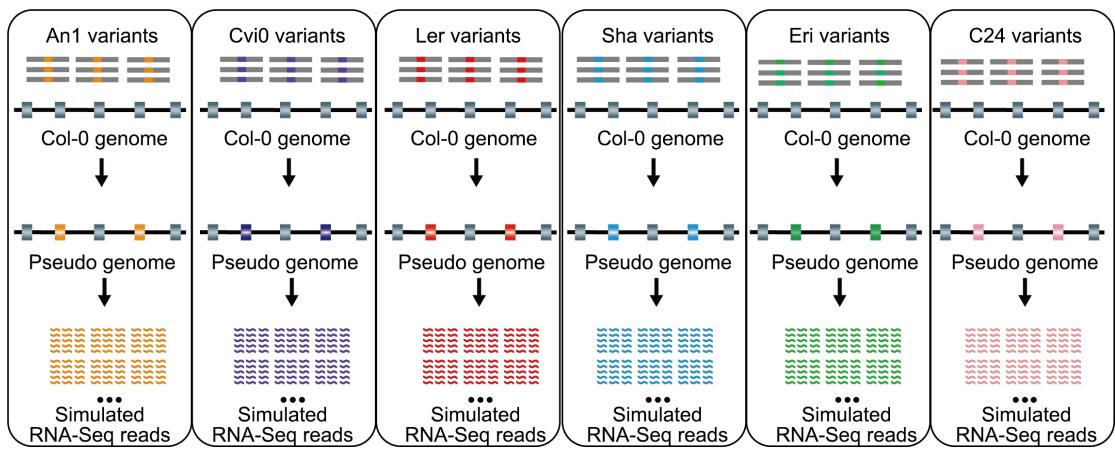


Figure S12. Generation of simulated RNA-seq reads for six *Arabidopsis* accessions.

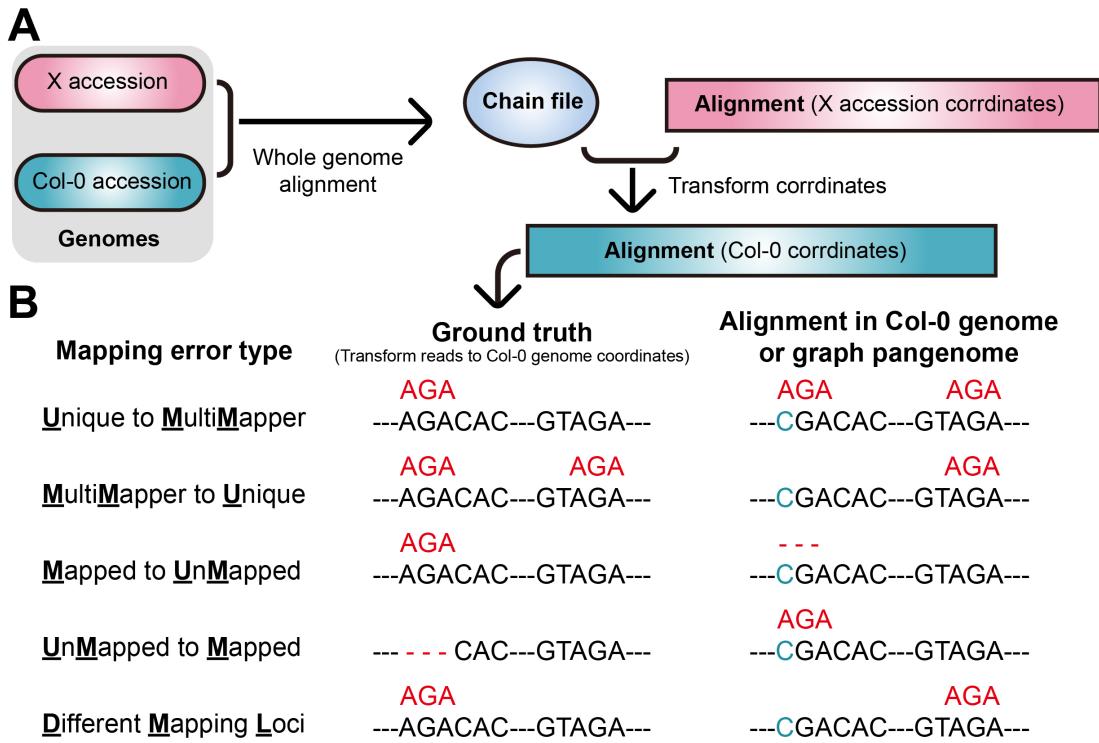


Figure S13. Illustration of the definitions for five types of mapping errors. (A) Generation of Col-0 reference genome coordinate-based ground truth. (B) Definition of five types of read-genome alignment errors.