Author version

Liu et al 2026, Genomics**,** Volume 118, Issue 2

https://doi.org/10.1016/j.ygeno.2026.111210

Original Research

Haplotype-resolved genome assembly of *Musella lasiocarpa* reveals the critical role of structural variations in chromosomal and genome evolution

Qing Liu[a, b, c, *, 1], Dongli Cui[a, b, d, 1], Yaqi Tian[a, b, d], Yehan Wang[a, b, d], Mathieu Rouard[e], John Seymour Heslop-Harrison[a, b, f, *], Trude Schwarzacher[a, b, f, *], Ziwei Wang[a, b, g, *]

[a] Key Laboratory of National Forestry and Grassland Administration on Plant Conservation and Utilization in Southern China / Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

[b] South China National Botanical Garden, Guangzhou 510650, China

[c] Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

[d] College of Life Sciences, University of the Chinese Academy of Sciences, Beijing 100049, China

[e] Biodiversity International, Parc Scientifique Agropolis II, 34397 Montpellier Cedex 5, France

[f] University of Leicester, Institute for Environmental Futures, Department of Genetics and Genome Biology, Leicester, LE1 7RH, UK

[g] Present address: School of Biology and Agriculture, Shaoguan University, Shaoguan 512005, China

[1] These authors contributed equally to this work.

* Corresponding authors' E-mails: liuqing@scib.ac.cn; phh4@le.ac.uk; ts32@le.ac.uk; and wangziwei@sgu.edu.cn

Running head: Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa*

ABSTRACT

Musella lasiocarpa (MLA, 2*n*=18, Musaceae) is an endangered species native to south-western China. We assembled its haplotype-resolved, telomere-to-telomere genomes with a genome size of 503.6 Mb consisting of 52.8% repetitive DNA. A 134 bp tandem repeat, Mlcen, was identified at all centromeres, and telomere sequences were present at 30 of 36 assembled pseudo-chromosome ends. The distal gene-rich regions display high synteny, whereas retrotransposon polymorphisms between haplotypes occurred throughout chromosomes, contributing to diversity. Phylogenetic analysis shows MLA diverged from *Ensete* 42 million years ago, and together they share a common ancestor with *Musa*. Among 35,312 protein-coding genes, 14 up-regulated and 34 down-regulated genes were identified under cold treatment. This high-quality genomic resource advances our understanding of MLA chromosomal evolution characterized by structural variations, repetitive DNA dynamics, and cold-responsive genes at both haplotype and species levels; and enables genome-assisted improvement of more resilient crops such as bananas and *Ensete*.

Keywords:

Musella lasiocarpa, telomere-to-telomere genome assembly, centromere, repetitive DNA, haplotypes, structural variation, cold adaptation

Introduction

Musella lasiocarpa (Franch.) C.Y. Wu ex H.W. Li (MLA, 2*n* = 2*x* =18, family Musaceae), is an endangered species (Plummer et al., 2022) endemic to Yunnan and Sichuan provinces of China (Fig. 1A-D). Commonly known as Dì Yǒng Jīn Lián, MLA is valued both as an ornamental plant and a feed crop. It has a feature of striking bracts, vibrant yellow in var. *lasiocarpa* and orange-red in var. *rubribracteata* (Fig. 1B-D). In addition, its pseudo-stems are used as a starch source for animals in Yunnan (Long et al., 2008). In the family Musaceae (order Zingiberales), MLA and its sister group *Ensete* (*x* = 9) are closely related to bananas (*Musa* spp.; *x* = 7, 9, 10, and 11) (Heslop-Harrison and Schwarzacher, 2007). MLA can be distinguished from *Ensete* and *Musa* by its erect, compact rosette-shaped inflorescences, and by chloroplast genome analyses (Fu et al., 2022). Notably, the species demonstrates higher cold tolerance than other members of the Musaceae; it can withstand the severe winter conditions prevalent in Sichuan (Ma et al., 2011).

More than ten genomes of species within Musaceae (Droc et al., 2022) have been assembled using long-read sequencing technologies (Oxford Nanopore Technologies, ONT, and Pacific Biosciences, PacBio), and with Hi-C chromatin conformation capture. Chromosome-scale assemblies are available for *Musa acuminata* (Belser et al., 2021), *M. schizocarpa* (Belser et al., 2018), *M. balbisiana* (Wang et al., 2019), *Ensete ventricosum* (GCA_029747655.1), *M. textilis* (Galvez et al., 2021), *E. glaucum* (Wang et al., 2022), *M. beccarii* (Wang et al., 2023), diploid *M. acuminata* (Liu X et al., 2023; Huang et al., 2023), and *M. velutina* and *M. ornata* (Xiao et al., 2024) and in comparative studies of wild and cultivated *Musa* species (Martin et al., 2025). The haploid genome sizes of these species range from 428 Mb (Xiao et al., 2024) to 603 Mb (Li et al., 2022). A recent publication by Zhao et al. (2024) reported the genome of *Musella lasiocarpa* but it does not have an in depth analysis of repetitive sequences, haplotype resolved assemblies and structural variations (SVs) as we report. Our genome assembly, annotation and raw data for *M. lasiocarpa* are on NCBI and the Banana Genome Hub (https://banana-genome-hub.southgreen.fr/) with tools for comparative analysis in the Musaceae (Droc et al., 2022).

Haplotype-resolved, telomere-to-telomere genome assemblies provide valuable insights into chromosome structural variations and allelic differences within a species.. These have been generated from *Musa acuminata* accessions (a diploid; Liu X et al., 2023), the triploid banana cultivar AAA ‘Baxijiao’ from the Cavendish cultivar group (Huang et al., 2023), two additional AAA genome cultivars from the Cavendish and Gros Michel cultivar groups (Li X et al., 2024), and two AAB genome triploid hybrids representing the Plantain and Silk cultivar groups (Xie et al., 2024). The haplotypes reveal the chromosomal contributions of ancestral species to the genomes of cultivated bananas within the genus *Musa* (Martin et al., 2025). Repetitive DNAs constitute a dynamic and rapidly evolving component that differentiates haplotypes within the same species. Analyses of chromosomal rearrangements provide insight into the complex genomic organization of repetitive sequences, extending beyond the implications of chromosome number variation (*x* = 9 or 11) observed among three families of Musaceae (Droc et al., 2022; Heslop-Harrison et al., 2023; Liu et al., 2023a). Haplotype assemblies are able to elucidate the relationships between repetitive components in species such as, for example, lemon (Bao et al., 2023), melon (Li G et al., 2023), and jujube (Li K et al., 2024). Moreover, it is also

essential to investigate structural variations, including retroelement insertion polymorphisms and large chromosomal rearrangements (inversions, translocations, duplications, and deletions).

Cold stress is a key factor restricting the geographical distribution and productivity of Musaceae species (Yang et al., 2015; Joshi et al., 2023). Among them, *Musella lasiocarpa* exhibits remarkable tolerance to low temperatures. By analyzing the transcriptome data, the candidate genes and regulatory pathways associated with cold tolerance in MLA can be identified. Overall, our study presents the first haplotype-resolved, near-telomere-to-telomere genome assembly including intensive manual curation (as suggested by Yang et al., 2025) of *M. lasiocarpa*, characterized by structural variations, repetitive DNA dynamics, and cold-responsive genes. The high-quality genome assembly serves as a valuable genomic resource for breeding programs aimed at developing resilient crops, including bananas and *Ensete* species.

Results

De novo haplotype-resolved genome assembly

We generated a *de novo* haplotype resolved, chromosome-level genome assembly of *Musella lasiocarpa* var. *lasiocarpa* (MLA). Using long-single-molecule PacBio HiFi and ONT reads with 62 × to 204 × sequencing depth, along with Hi-C and Illumina data (Table S1), we assembled nine pseudo-chromosomes for each haplotype (Figs. 1E and S1), yielding genome sizes of 500.05 Mb for MLAh1 and 498.37 Mb for MLAh2 with individual chromosomes ranging from 40.91 Mb to 59.46 Mb in length (Tables 1 and S2A). Chromosome identity and orientation were determined based on their homology to the *Ensete glaucum* (EGL) genome assembly (Wang et al., 2022; itself based on homology to *Musa*). Because the parental origin of each chromosome is unknown, the pseudo-chromosomes with fewer contigs and higher N50 values were designed as MLAh1, which was used as the reference genome for subsequent analyses (Tables 1 and S2A).

The haploid genome size of MLA was initially estimated to be 436 Mb using the 21-mer distribution visualized by GenomeScope (Vurture et al., 2017), also showing a heterozygosity of 1.24% (Fig. S2A, Table S2B). Based on sequence coverage distribution and mapped nucleotide (backmap.pl v.0.5; Pfenninger et al., 2022) results, the genome size was refined to 503.58 Mb (Table S2B), which is slightly smaller than the 535 Mb reported by Zhao et al. (2024).

Quality assessment of genome assembly

Our haplotype-resolved genome assemblies captured 99% of the MLA genome (Table 1, Table S2), organized into nine pairs of gapless pseudo-chromosomes. Genome completeness and the predicted protein sequences were accessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) (Manni et al., 2021), yielding scores greater than 98% (Table S3), which were higher than those reported for other related monocotyledonous assemblies (Fig. S2B). In Hi-C interaction heatmap, signal intensity along the diagonal, representing interactions between neighboring sequences, was stronger than that of non-diagonal positions (Fig. S3), indicating that the haploid assemblies are complete, and no structural variations were caused by assembly algorithms. Telomeric repeats $(TTTAGGG)_n$ were detected at 30 of the 36 chromosome ends (Table S4). A 134 bp centromeric repeat (Mlcen) was detected at all centromere positions (Figs. 1E and S1; Table S5) and showed high sequence homology to Egcen from *E. glaucum* (Wang et al., 2022). Additionally, 5S and 45S rDNA arrays were localized on chromosomes MLA05, MLA06, and MLA08 (Table S6). These results collectively provide high-quality, haplotype-resolved, telomere-to-telomere assemblies of the MLA genome.

Gene annotation and expression analysis

Gene prediction and functional annotation and ontology

The protein-coding genes (PCGs) of the MLA haplotypes were annotated using an integrated strategy that combined RNA-seq-assisted and *ab initio* prediction with homology-based annotation. In total, 35,387 and 35,237 PCGs were predicted in MLAh1 and MLAh2, and about 94% of these genes were functionally annotated using seven major protein databases (NCBI NR, Pruitt et al., 2005; InterPro, Paysan-Lafosse et al., 2023; GO, Ashburner et al., 200; KOG, Tatusov et al., 2003; KEGG, Kanehisa et al., 2023; TrEMBL and SwissProt, Bairoch and Apweiler, 2000) (Table 2 and S3C). The number of annotated genes represents roughly a 3% increase compared to the 34,391 genes reported by Zhao et al. (2024) in a previous assembly and annotation. Predicted gene density was lower near centromeric regions and gradually increased toward the chromosome ends (Figs. 1E and S1).

Gene family analysis of Musaceae species (*Musa acuminata*, *M. textilis*, *M. beccarii*, *Musella lasiocarpa*, *Ensete ventricosum*, and *E. glaucum*) clustered 195,076 genes (89.4%) of the total 218,084

into 31,235 ortho-groups (Table S7). The remaining 23,008 genes (10.6%) were unassigned, representing putative singletons with no detectable orthologs in the other species (Table S7). Among the six species, 18,453 ortho-groups (59.1%) identified by UpSetR (Conway et al. 2017) were shared, corresponding to 22,670 genes (72.6%) that constitute the core or pseudo-core gene sets (Fig. 2A), assuming that orthogroups absent in only one species are likely due to annotation artefacts, while 212 orthogroups (0.7%) were specific to *Musella*, with no homologs identified in *Musa* or *Ensete* (Fig. 2A). Synonymous substitution rate (Ks) analysis showed a peak at ~0.53 in both intra and interspecific comparisons involving MLA, *E. glaucum*, and *M. acuminata* (Figs. 2B and S4), supporting the occurrence of the α/β whole genome duplication (WGD) events during the evolution of the Musaceae genome. Additionally, a Ks peak around 0.75 for the MLA-*Zingiber officinale* comparison (Li H et al. 2021) supports the γ WGD event associated with the Zingiberales lineage.

Gene family expansion and contraction

Protein sequences from MLA and 13 other monocotyledonous species were clustered into 33,044 gene families. In the MLA genome, 16,595 orthologous genes were identified - fewer than in the eight other Musaceae species but greater than in the other Zingiberales species (Table S8). Among these, 994 single-copy orthologs were shared across 14 species while 155 orthologs were unique to MLA (Fig. S5A, left; Table S9). Among 154 gene families (consisting of 1,618 genes) showing significant expansion, and 162 (with 346 genes) showing significant contraction (Fig. 2C; Table S8B-C) in MLA, the enrichment analysis of the expanded gene families indicated 55 genes with the GO term "photosynthesis/light reaction" (Fig. S5A, right; Table S10). Phylogenetic reconstruction and divergence time estimation based on the shared single-copy orthologs indicated that *Musella* and *Ensete* form sister taxa that diverged approximately 42.0 million years ago (Mya), and that both lineages diverged from *Musa* around 57.0 Mya (Fig. 2C).

Differential gene expression (DEG) under cold treatment

When subject to a 10°C temperature drop from 25°C (normal conditions) to 15°C (cold treatment) for 48 h, *M. lasiocarpa* var. *lasiocarpa* (yellow bracts) and var. *rubribracteata* (red bracts) showed significant transcriptomic responses. Approximately 10% of genes were differentially expressed, with 3,151 differentially expressed genes (DEGs) in yellow bracts and 3,726 in red bracts, showing at least a two-fold change in expression ($P < 0.05$) between normal- and cold-treated samples (Fig. 2D, upper left). Of these, 1,470 genes were down-regulated in both varieties, whereas only 88 genes were up-regulated (Fig. 2D, upper right; Table S11). KEGG pathway and GO enrichment analyses of DEGs revealed that the up-regulated genes were enriched in pathways related to environmental adaptation and DNA repair and recombination (Fig. 2D, bottom left; Table S12A). Conversely, the down-regulated genes were predominantly involved in fundamental cellular processes, transcription, metabolism and particularly photosynthesis (Fig. 2D, bottom right; Table S12B). Notably, genes associated with environmental adaptation showed significant expansion in *M. lasiocarpa* (Fig. S5A, right; Table S10).

Among 1,470 down-regulated genes and 88 up-regulated genes in both MLA varieties, the 48 differentially expressed transcription factors (DETFs) were identified. Of these, 14 DETFs were up-regulated and 34 were down-regulated, all of which are involved in cold acclimation (Fig. 2E, right; Table S13). Previous studies (Li Y et al., 2021; Pashapu et al., 2024) reported that those 14 up-regulated genes belong to 10 transcription factor (TF) families, while the 34 down-regulated genes belong to 13 TF families (Table S13). These TFs are associated with the major cold stress-responsive pathways previously characterized in grasses and Arabidopsis as well as *Musa* (Yang et al., 2015; Joshi et al., 2023). The DEGs are distributed across all chromosomes except MLA01, with notable clustering on MLA02, MLA05, and MLA06. However, they are not linked to any structural variations identified in the MLA genome (Fig. S6).

Synteny and structural variations between haplotypes

Pairwise whole-genome comparisons revealed largely conserved synteny between the two haplotypes. Structural variations including inversions, translocations, and duplications, were abundant in the polymorphic proximal pericentromeric regions (Fig. 3; Table S14), and some large SVs were present in gene-rich regions of chromosome arms. Chromosomes MLA01, MLA03, and MLA05 exhibited the highest proportion of rearrangements, whereas MLA02, MLA07, and MLA09 maintained the greatest synteny. Synteny was most pronounced at the chromosome ends as evidenced by the SyRI collinearity plots (Goel et al., 2019) by dense, parallel grey lines in the distal regions (Fig. 3) that are gene-rich (Fig. 1E). In contrast, synteny was increasingly disrupted near the

centromeres, due to accumulation of repetitive sequences causing SVs. Furthermore, the 45S rDNA arrays on MLA06 and MLA08 displayed substantial differences between the two haplotypes (Fig. 3).

Detailed chromosome sequence dotplots together with SyRI analysis (Fig. S7) showed that each chromosome contained between two to seven regions, ranging 0.1 to 2.2 Mb, that harbored major structural rearrangements, such as inversions, translocations, and duplications including tandem and mixed repetitive sequences (Table S14). Interestingly, some SVs included genes or coding regions (e.g., Ml01i, Ml04ii, Ml05iii, Ml09i; Fig. S7A, D, E, J), while others either lacked genes entirely (Ml06ii; Fig. S7F), or contained gaps of uncharacterized sequence (Ml02ii; Fig. S7B). Several tandem arrays contributed to the polymorphisms observed between haplotypes, and many of the arrays were rich in transposable elements (e.g., Ml02iii, Ml03ii, Ml06iii, Ml08iii; Fig. S7B, C, F, I). For instance, a 600 kb insertion (Ml09i; Fig. S7J) consisted of eight copies of a 12 kb monomer, which included a gene (polygalacturonase), fragments of chloroplast DNA, and some retroelement domains (APR, aspartic protease), along with additional other retroelement fragments located outside the repetitive monomer motif (Figs. 4A and S7J). The transfer of organellar genes to the nucleus has been well documented, and the presence of their fragments within these repeat monomers is noteworthy (Liu et al., 2023b).

Prominent inversions (2–5 Mb) were detected outside the centromeric regions of MLA03, MLA04, and MLA08 (Fig. S7C, D, I). ONT reads spanning the inversion breakpoints confirmed these real SVs were not caused by assembly artefacts (Fig. S8). Analysis of breakpoints using dotplots revealed small inverted repeats at the sites, indicative of LTR retroelement activities, implying that haplotype diversity in *Musella* is primarily driven by LTR retrotransposon amplification and insertion (Fig. S8Aii-iv, Bii-iv, Cii-iv).

Haplotype SNPs, short *InDels, and heterozygosity*

Analysis of single nucleotide polymorphisms (SNPs) identified 3,352,140 variants, accounting for 0.67% of the genome with 563,104 insertions and deletions (InDels < 50 bp) within syntenic blocks (Fig. 4B, top). Most SNPs and InDels were located in intergenic regions (43.82%), followed by upstream and downstream regulatory regions and introns (Fig. 4B, bottom). Only a small fraction (4.81%) occurred in exonic regions, frameshift mutations or alterations resulting in gain or loss of stop codons were rare (0.70–1.26%).

Repeat analysis

In MLA, 263,543,055 bp (52.81%) of repetitive sequences were identified by RepeatMasker (Tables 2 and S15), consistent with the results from RepeatExplorer2 result (51.08%; Fig. S9B) and similar to other banana genomes (52.62% in *M. acuminata*, Belser et al., 2021; 55.02% in *E. glaucum*, Wang et al., 2022). The most abundant class of repeats were long terminal repeat retroelements (LTRs), which accounted for 34.40% of the MLA haplotypes (Table 2). In addition, multiple short tandem arrays of various repeat monomers, including microsatellites (simple sequence repeats, SSRs), were detected in both *Musa* and *Ensete* assemblies using RepeatExplorer2 (Novák et al., 2020).

Retrotransposons

The most abundant classes of retrotransposons were *Copia* (14.5%) and *Gypsy* (11.9%) (Tables 2 and S15), with proportions slightly lower than those reported for *E. glaucum* (Wang et al., 2022). In *Musa acuminata*, *Copia* elements were notably more prevalent (29%) than *Gypsy* elements, which accounted for 11% and 19% in D'Hont et al. (2012) and Martin et al. (2025), respectively. The chromosome distribution of both *Copia* and *Gypsy* elements showed higher densities in proximal regions, contrasting with the distal enrichment of genes (Figs. 1E, S1, and S9A), a pattern consistent with observations across other Musaceae species. Analysis of LTR retroelement insertion times in two haplotypes indicate two major episodes of LTR activity: a recent peak around 0.46 Mya and older peak around 3 Mya (Fig. 4C). In comparison, *Ensete glaucum* e exhibits a predominant insertion peak between 3.5 and 5.5 Mya with a secondary peak at 0.5 Mya (Wang et al., 2022), whereas *M. balbisiana* and *M. acuminata* display major peaks at 0.5 Mya and 1.5 Mya, respectively (Wang et al., 2019). These results suggest that retroelement insertion timings are species-specific within Musaceae, reflecting at least two distinct bursts of retroelement activity and genome instability (Fig. 4C).

Our haplotype assemblies enabled a detailed examination of Presence Absence Variations (PAVs) between homologous chromosome pairs (Fig. 4D). Genomic variations >50 bp were distributed relatively uniformly across the chromosomes, with localized peaks on MLA02 (Fig. 4D, red density maps), that correspond to regions identified in the synteny plots (Fig. 3). Extraction of polymorphisms between haplotypes that were annotated as transposable elements (Fig. 4D, blue density maps) also showed an overall uniform distribution across

nine chromosomes, with additional enrichment in centromeric regions where transposable elements were more abundant (Fig. 1E).

Telomeres

MLAh1 and MLAh2 each contained arrays of 160–6,000 copies of the telomeric monomer (TTTAGGG/CCCTAAA), located at 30 of 36 pseudo-chromosome ends (Table S4). Telomeric sequences were absent from the right ends of MLA06 and MLA08, both of which terminate with 45S rDNA arrays, as well as from the left ends of MLA05.1 and MLA08.2 (Table S6A). Consistent with previously published genome assemblies, telomeres were not assembled on chromosomes bearing 45S sites. Additionally, no ONT reads > 50 kb were detected that contained both 45S rDNA loci and telomeric monomers. Fluorescence *in situ* hybridization (FISH), however, detected the telomeric $(TTTAGGG)_7$ at the ends of all chromosomes, appearing as double dots of variable intensity, including at sites distal to the nucleolar organizing regions (NORs; Figs. 4E and S10A). We propose that the extension of the transcribed rDNA array within the interphase nucleus may cause DNA breakage during extraction. Since the chromosomal segment distal to the NOR appears to be very small based on FISH results (Fig. S10A, B), such segments containing the telomeres may have been lost during sequencing library preparation and, consequently, could not be assembled.

Centromeres

A 134 bp centromeric monomer (Egcen) from *E. glaucum* was used to identify a homologous centromeric repeat, Mlcen, in MLA. Mlcen accounts for approximately 2.33% of the MLA genome (86,889 copies; Tables 2 and S5) and shares 98% sequence identity with Egcen. Notably, Mlcen was not detected in the centromeric region of *Musa* chromosomes (Wang et al., 2022). Cluster analysis of Illumina sequencing reads using RepeatExplorer2 (Novák et al., 2020), identified Mlcen as the first cluster representing about 2.3% of the genome, and displaying a characteristic 'star-shaped' pattern typical of tandem satellite repeats (Fig. 4F). In addition, *k*-mer analysis using TAREAN program (Novák et al., 2020) confirmed Mlcen as a high confidence satellite repeat). Mlcen is a GC-rich sequence (46.6%) that contains part of the functional CENP-B box as well as 11-mer motif, suggesting low sequence variability within its 134 bp monomer. Arrays of Mlcen were detected on nine chromosome assemblies (Table S5B), and FISH revealed strong, though variably intense signals at centromeric regions (Figs. 4E, 5A, and S10A, B). Based on these Mlcen arrays, centromeric locations were defined within the assemblies (Figs. 1E, S1), and chromosome arm lengths were estimated (Table S5C, D).

Comparison of MLA haplotypes reveals differences in centromeric organization. The Mlcen arrays range from 1.0 to 10.6 Mb, with the Mlcen region of MLA09.1 containing less than half the copy number of MLA09.2 (Table S5B). In some chromosomes, the Mlcen arrays were interrupted by genes (MLA01), LTR retroelements (MLA04), or inversions MLA05 (Fig. 3). The regions surrounding these arrays show variable structural features: some harbor genes and few variations on MLA03 and MLA08 (Fig. S7C, I), while others exhibit extensive inversions, translocations, and duplications, such as on MLA05, MLA06, MLA07 (Fig. S7E-H).

rDNA

In the MLA assemblies, two 45S rDNA loci were identified at the distal ends of the right arm of MLA06 and MLA08, and a single 5S rDNA locus was located in the middle of the right arm of MLA05 (Figs. 1E and S1), consistent with FISH results (Figs. 5A-D and S10C-F). The 5S rDNA monomer was 658 bp in length, comprising a 119 bp of gene (typical for all plants) and a 539 bp the intergenic spacer (IGS), with 2,069 copies representing 0.27% of MLA genome (Tables 2 and S6). The 5S rDNA locus on MLA05 was organized into two major arrays separated by roughly 10 Mb unrelated sequences (Fig. 5E, F). FISH revealed the organization of 5S rDNA on ml05, exhibiting a discontinuous signal along extended prometaphase chromosomes (Fig. 5A, right). This observation supports the presence of two distinct 5S rDNA arrays on ml05, analogous to the fragmented loci on EGL05 and the paired 5S rDNA sites reported in *Musa* spp. (Čížková et al., 2013; Wang et al., 2022).

FISH analysis revealed that 45S rDNA sites at the ends of chromosome ml06 and ml08 co-localized with the tandem repeat MuTR (GenBank: AM905888, Teo and Schwarzacher unpubl. data) (Fig. 5B, D). Sequence analysis confirmed that each 45S rDNA monomer typically contains four copies of a degenerate MuTR tandem repeat. The arrangement on MLA06 and MLA08 follows the structure: 18S rRNA gene-ITS1-5.8S rRNA gene-ITS2-26S rRNA gene-NTSL-STR-MuTR-NTSR (Fig. 5G). The 45S rDNA monomer in MLA is 8,347 bp in length (Table S6B), slightly longer than in *Musa acuminata* (7,553 bp; Hřibová et al., 2010) but shorter than in *E. glaucum* (9,984 bp) (Wang et al., 2022). The arrays are occasionally interrupted by LTR retroelement insertions (Fig. 5H), indicating a complex organization of the 45S rDNA loci (Figs. S11 and S12). In total, MLA contains 1,465 copies of 45S rDNA monomer, accounting for 2.58% of the genome

(Tables 2 and S6B). This pattern differs from *Musa* spp. (1, 2, or 4 sites) and EGL (1 site) (Osuji et al., 1998; Čížková et al., 2013; Wang et al., 2022).

Synteny and karyotype evolution

Syntenic analyses were performed to compare the genomic structures of *Musella lasiocarpa* (x = 9, MLAh1) with other Musaceae species (Fig. 6), including *E. glaucum* (EGL, x = 9, Wang et al., 2022), *Musa acuminata* v.4 (MAC, x = 11) and *M. beccarii*, (MBE, x = 9, Wang et al., 2023). The assemblies showed similar numbers of gene orthogroups (Fig. 2A). Using MCScan (Wang et al., 2012), the number of syntenic genes of MLA-MAC (19,683) was higher than MLA-EGL (17,529) and EGL-MBE (14,722). The highest number of syntenic genes were present between MBE-MAC (21,214) in Musaceae (Tables S7B).

There was a high level of similarity between MLA-EGL (Fig. 6A). Six chromosome pairs–EGL02/MLA02, EGL03/MLA03, EGL05/MLA05 (containing 5S rDNA arrays), EGL07/MLA07, EGL08/MLA08, and EGL09/MLA09 shared both gene order and content. This was supported by parallel lines in SynViso (Bandi and Gutwin, 2020) and dotplot results (Fig. S13). However, synteny was less conserved in pericentromeric regions rich in retroelements and repetitive DNAs (Fig. 1E), despite the presence of the shared centromeric repeats (Mlcen and Egcen). The remaining three chromosomes showed a small number of SVs: MLA01 consisted of the intercalary region of EGL01, with additional segments derived from translocations and internal inversions involving EGL04 and EGL07. MLA04 contained three large segments from EGL01, EGL04, and EGL06, along with smaller parts from EGL07. MLA06 had segments originating from EGL04 and EGL06. Among these, MLA07 showcased the highest similarity to EGL07 in terms of gene content (Fig. 6A).

The positions of rDNA loci underwent substantial shifts during species radiation, often accompanied by a depletion of PCGs on chromosome regions immediately adjacent to 45S rDNA loci in Musaceae (Wang et al., 2022, 2023; Bartoš et al., 2005). A comparison of the karyotypes of MLA and EGL revealed changes in both individual chromosome sizes and centromere positions, particularly in chromosomes 01 and 04 (Fig. 6B). Notably, the chromosome arm carrying the 45S rDNAs in EGL06 showed depletion of genes (Wang et al., 2022), while the 45S rDNA bearing arms in MLA06 and MLA08 maintained gene density similar to other chromosome arms (Fig. 6A). Furthermore, the 45SrDNA genes in MLA are located within a different genomic context compared to EGL: on MLA06, they are on the opposite arm relative to EGL06, and on MLA08, they are in region where EGL had no 45S rDNA sequences. As mentioned previously, both 45S rDNA sites in MLA are positioned at the end of the genome assemblies; however, the arrays are not completely assembled, as indicated by the absence of the telomere sequence TTTAGGG that is observed as signals by FISH (Figs. 4E and S10B).

A larger number of chromosomal rearrangements were observed between MLA (x = 9) with MAC (x = 11) (Fig. 6C), beyond the consequence of chromosome number differences. Except for MLA05 (entirely syntenic with MAC05), all chromosomes showed multiple rearrangements. The inversions and translocations mainly involved large segments except for MLA04, which is composed of regions derived from MAC03, MAC04, MAC06 and MAC10. MLA09 includes all of MAC11 chromosome and part of MAC04, indicating possible centromere gain or loss. An interesting scenario is found for EGL-MLA-MBE (x = 9); in comparison with the x = 9 EGL and MLA, no single chromosome remained syntenic in their entirety and showed extensive rearrangements. Hence, we postulate that MBE (x = 9) possesses a derived karyotype, as evidenced by fewer rearrangements revealed in MBE-MAC synteny (Fig. 6D).

Discussion

Haplotype assembly

MLA haplotype assembly is critical for revealing the nature of polymorphisms within a single non-inbred plant, whose heterozygous genome is rich in repetitive sequences (Figs. 1, 3, and S1). Studies of allelic differences between genes have formed the foundation of genetics for decades; however, consensus and mosaic assemblies often omit important allelic variations that influence key traits. In contrast, haplotype assemblies of polymorphic species show allelic variation across all genes within a single accession (Huang et al., 2023; Li W et al., 2024; Usai et al., 2025). The existing genome assemblies constitute valuable resources for allelic studies, including genes associated with disease resistance (Zhao et al., 2024), secondary metabolites (Cui et al., 2024), and cold tolerance in the Musaceae. Our haplotype assemblies further enhance this understanding by revealing structural variation including inversions, translocations, sequence amplifications, and duplicatons, between haplotypes.

Genes and cold adaptation

Cold is a major abiotic factor that markedly affects banana production (Joshi et al., 2023). The

plants' response to cold is complex, involving both rapid "early" (< 6hr) and slow "late" (> 24 hr) responses (Yang et al., 2015) with genes either up- or down-regulated. In our cold treatment experiments, we identified genes and transcription factors known to play roles in cold tolerance of MLA (Fig. 2E). The C-repeat/DREB binding factor (CBF)-dependent responsive pathway (involving ERF, Ethylene Response Factors, Fig. 2D) has been identified as important for survival of MLA, *Musa* (Yang et al., 2015), grasses (Pashapu et al., 2024), and Arabidopsis (Li Y et al., 2021). These genes, along with other stress-related transcription factors (such as WRKY, zinc-finger, and basic helix-loop-helix HLH, with the PIF phytochrome subfamily; or lectin receptor-like kinases, Xiong et al., 2024), contribute to the differential expression in cold-treated MLA (Fig. 2D, bottom) as well as in previous studies (Li Y et al., 2021; Pashapu et al., 2024). These TFs are candidate genes to support breeding strategies for cold-resilient crops in Musaceae (Yang et al., 2015; Luo et al., 2022).

Genome assemblies and transcriptome data are proving useful for revealing the regulation of secondary metabolite production and the variation. Zhao et al. (2024) show the O-methyltransferases involved in phytoalexin biosynthesis, providing valuable genetic resources related to anti-fungal compounds. Flavonoid biosynthesis is critical for plant adaptation to extreme environments (Cui et al., 2024). As examples, the assembly and transcriptome analyses of the *Phanera championii* (Lu et al., 2024) reveal how alleles among haplotypes influence flavonoid biosynthesis. Cui et al. (2024) show that in MLA genotypes, cold stress modulates the expression of flavonoid biosynthesis genes and TFs, altering the relative abundance of different anthocyanins and their biosynthetic pathways.

Haplotype SNPs and Heterozygosity

The haplotype-based SNP comparison revealed a heterozygosity level of 0.78%, which is lower than the 1.24% inferred by GenomeScope (Fig. S2A). This estimate falls within a two-fold range of values reported for other Musaceae species by the similar approach, e.g., *Musa acuminata* (0.59%, Liu X et al., 2023), *M. acuminata* ssp. *banksii* (0.02–0.34%, Sardos et al., 2022), *Ensete glaucum* (0.164%, Wang et al., 2022), and *E. ventricosum* (0.73%, Haile et al., 2024). Our result aligned with the expected analysis differences, as GenomeScope often over-estimations heterozygosity due to the presence of pseudogene and retroelement fragments in *k*-mer profiles. SNPs occurred throughout the genic and intergenic regions of the genome (Fig. 4D), and potential inbreeding events in some lineages. The reduced density of SNPs in exonic region likely reflects stronger evolutionary constraints that limit the tolerance of variation in protein-coding sequences. Within these exonic SNPs, the proportions of synonymous and non-synonymous variants were nearly equal (Fig. 4B), indicating a balance between mutations that preserve protein function.

Structural variations between haplotypes exhibited a range of differences (Fig. 3) in broad, repeat-rich, pericentromeric regions (Figs. 1E, S1), as elucidated in Musaceae genomes (Martin et al., 2017, 2025; Droc et al., 2022). Polymorphisms were confirmed by long-single-molecule ONT reads (e.g., Fig. S8). The broad distribution of retroelements between haplotypes (Fig. 4D) is consistent with the polymorphisms identified among *Musa* accessions, identified through Inter-Retroelement Amplified Polymorphisms (IRAPs) (Nair et al., 2005). This consistency demonstrates the reliability of the polymorphisms for variety identification (Beránková et al., 2022). Our genome assemblies will serve as a useful resource for future studies aimed at understanding the impact of SVs, including retroelement insertions, on gene expression.

In *Musa*, phased haplotype assemblies have been generated from diploid (Liu X et al., 2023), triploid *M. acuminata* (Huang et al., 2023; Li X et al., 2024), and *Musa* AAB triploid hybrids (Xie et al., 2024). In wild diploid *M. acuminata*, 47 translocations (cumulative size 2.7 Mb), 23 inversions (11.3 Mb) and 53 duplications (1.33 Mb) were detected. These values are substantially lower than in MLA, where 459 translocations (7.4 Mb), 140 inversions (23 Mb) and 1,108 duplications (13.5 Mb) were observed. In *Musella*, the average recombination rate was 1.2 events per Mb, representing 26–37 recombination events per chromosome arm (Table S5C) per meiosis (Table S14). Interestingly, the relatively large inversion fragments (8,045,525 bp) of MLA03 did not alter the recombination rate, unlike MLA01, which displayed the highest number and sizes of translocations and duplications. These inversions likely suppress crossover formation in the affected chromosome regions (as shown by Li H. et al., 2023). In triploid *M. acuminata*, the synteny plots revealed additional rearrangements, particularly in haplotype BXJ2 (Huang et al., 2023: their fig. 1f). Overall, comparisons of multiple *Musa* genomes indicate that cultivated *Musa* are genetically derived from multiple wild progenitors (Martin et al., 2025). In contrast, in diploid *Musella*, no inter-chromosomal exchanges or gaps flanking inversions were observed. Both phased assemblies were continuous (Fig. S8), as confirmed by single-molecular reads (Fig. S7; S9). This suggests that homologous exchanges involving loss of

segments between chromosomes as seen in MLA are more likely associated with speciation events (Martin et al 2025).

Three haplotype-resolved assemblies of the apple genome, each approximately 650 Mb in size, provide a useful comparator (Li W et al., 2024). As observed in Musaceae, gene density was highest near the chromosome ends (their Fig. 3a circle 6 of Li W et al., 2024), while transposable elements tended to be in proximal regions (also their circle 6). SVs were distributed across chromosomes (their circle 2), including approximately ten large inversions of multiple-megabase size (their Fig. 3c). In apple, 4.76% of SNPs were located in the coding sequences, similar to the 4.81% in *Musella*. However, unlike MLA, the distribution of SNPs and InDels in the apple genome was uneven, with certain chromosomes or large segments exhibiting very few variations (their Fig. 3a circles 3 and 4 in Li et al., 2022). This pattern raises the possibility that uniparental disomy, potentially arising from meiotic from non-disjunction in inbred parents (a phenomenon best known in animals), may have occurred in this system.

The breeding system and natural pollinators of *M. lasiocarpa* remains poorly understood. In the horticultural trade and possibly in the wild, the species primarily propagates through vegetative side-suckers (Liu A et al., 2002), with fruit production being rare in cultivation. Female sterility promotes clonal growth (Xue et al., 2007), restricts gene flow, and helps maintain heterozygosity. Additionally, SVs in the genome can disrupt meiosis, leading to reduced fertility (Xue et al., 2007) and inhibits recombination within the species, complicating efforts to produce interspecific or intergeneric hybrids or generate desirable recombinant genotypes.

Repeat analysis and polymorphisms

Variation in the copy number of tandem repeats, including rDNA arrays, is commonly observed between homologous chromosomes, both in species and within a single individual (Liu et al., 2019). In *M. lasiocarpa*, the 45S rDNA arrays on MLA06 and MLA08 exhibited pronounced differences between haplotypes (Figs 3 and 5F). Small tandem arrays also contribute to haplotypic polymorphisms; however, many of these arrays were degenerate and interspersed with other sequences including chloroplast fragments, genes, and retroelements. These mixed arrays tend to occur in genomic regions that are rich in either genes or retroelements (Figs 3 and S7). The polymorphisms located in and around centromeric regions, which were generally depleted in genes, are likely generated through uneven crossing-over. Such SVs have been reported in various species. For instance, in poplar, a ~200 kb tandem-repeat array is inserted at the centromere of Chr04A relative to Chr04T (Zhou et al., 2023). In plants, chromosome pairing during meiosis is conventionally initiated at the telomeres, with recombination occurring more frequently in distal regions (Sepsi et al., 2017). Consequently, the highly syntenic MLA chromosome ends likely facilitate accurate homologous pairing and recombination during meiosis. Conversely, the high density of repeats, frequent SVs and inversions around centromeres (Figs. 1E and 3), are expected to suppress meiotic recombination in these pericentromeric regions.

Evolution, diversification and synteny in Musaceae

Our genome assembly confirmed *Musella* to be a distinct genus within Musaceae, exhibiting a closer relationship to *Ensete* than to *Musa*. *Musella* can be distinguished from *Musa* (Franchet, 1889) and *Ensete* (Cheesman, 1947) by its erect, compact rosette inflorescences, and its genome assembly provides an opportunity to identify the genes underlying these unique traits. Syntenic comparison revealed extensive rearrangements in EGL and MLA relative to *Musa acuminata* (x = 11) (Fig. 6D), and even more so compared to *M. beccarii*, which we propose as a derived x = 9 species. A previous study by D'Hont et al. (2012) showed that *Musa* underwent three rounds of WGD events. In this study, we also detected evidence for the α and β WGDs at the Cretaceous-Paleocene boundary (Fig. 2B), as well as the more ancient γ WGD event at ~100.0 Mya (D'Hont et al., 2012) and *E. glaucum* (Wang et al., 2022), with no further additional MLA-specific WGD events. Dating based on single-copy orthologs indicated that the *Musella* and *Ensete* genera diverged approximately 42 Mya, while they split from *Musa* around 57 Mya (Fig. 2C). These estimates align with findings from recent nuclear and chloroplast genome studies (Fu et al., 2022; Zhou et al., 2024).

Conclusion

Our publicly available, high-quality, haplotype-resolved, telomere-to-telomere (T2T) genome assembly of MLA, generated through the integration of long- and short-read sequencing technologies, provides detailed insights into SVs between MLA haplotypes and the role repetitive sequences play to generate diversity. In general, the detailed analysis of the nature of the haplotype variations suggests the mechanisms for recombination-lead generation of polymorphisms, often involving repetitive DNA, including between genotypes of inbred lines. In addition to gene-level differences, these SVs likely contribute to phenotypic variation, limit

recombination within the species, and influence the formation of wide hybrids. Such insights are critical for effectively utilizing crop genetic resources (in Musaceae, Jenny et al., 2024, and more widely) to enhance the agronomically important traits through breeding and crop improvement.

Materials and methods

Additional information about material and methods is provided in supplementary file (Data S1).

Collection of plant material and extraction of DNA

Fresh leaves of *Musella lasiocarpa* var. *lasiocarpa* and var. *rubribracteata* were collected from the South China Botanical Garden (originally sourced from Nanhua city, Chuxiong Yi Autonomous Prefecture, Yunnan, China). Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) , and its quality was assessed through gel electrophoresis, a NanoDrop One UV-Vis spectrophotometer and a Qubit 3.0 fluorometer (Thermo Fisher Scientific, USA).

Library preparation and whole genome sequencing

Multiple sequencing libraries were generated using different technologies: the PacBio HiFi library was prepared with the SMRTbell Express Template Prep Kit 2.0, the Oxford Nanopore library was created using the SQK-LSK109 kit on PromethION, and the Hi-C library (Belton et al., 2012) was prepared with modified protocols on the Illumina NovaSeq 6000 platform (Illumina, USA). For the Hi-C library preparation, cells were cross-linked with 2% formaldehyde, digested using DpnII (New England Biolabs), and biotinylated with biotin-14-dCTP, and generating 2 × 150 bp paired-end Illumina reads. For ONT transcriptome sequencing, RNA was extracted using the TRNzol Universal Kit (Tiangen). Quality control of the sequencing data was performed by fastp v.0.23.3 (Chen et al., 2018).

Transcriptome sequencing for cold response

MLA varieties were initially cultivated at 25°C and then subjected to cold treatment of 15°C for 48h. RNA was extracted using the Plant RNA Kit (OMEGA-R6827) and sequenced on the DNBSEQ-T7 platform (BGI, Shenzhen, China) with three biological replicates. DEGs were identified using DESeq2 v.1.54.0 (Love et al., 2014) with thresholds of fold-change >2 and the adjusted P-value (padj) < 0.05. TFs associated with cold adaptation were identified using PlantTFDB (Jin et al., 2009) with default setting (without exclusive selection of "best hit in *Arabidopsis thaliana*"). A heatmap of gene expression levels was generated by TBtools (Chen et al., 2023), applying row scale normalization.

k-mer analysis and genome assembly

The genome size was estimated using GenomeScope v.1.0 (Ding et al., 2020) With 21-mers generated by Jellyfish v.2.3.0 (Marçais and Kingsford, 2011). The phased genome assembly was generated by combining PacBio HiFi reads, HiC reads, and ONT reads [filtered using Filtlong v.0.2.4 (Wick et al., 2017) and assembled with Hifiasm v.0.16.1 (Cheng et al., 2021)]. The resulting genome assemblies were then manually refined by Juicer v. 1.5 (Durand et al., 2016a) and Juicebox (Durand et al., 2016b).

Genomic evaluation and annotation

Among other methods, the quality of the assembly was assessed using BUSCO v.5.5.0 (Simão et al., 2015). Repeats were annotated with RepeatModeler (Flynn et al., 2020) and EDTA v.2.1.0 (Ou et al., 2019), and LTR elements were classified using TEsorter (Zhang et al., 2022). Genome masking was then performed with RepeatMasker (http://www.repeatmasker.org). Gene prediction and functional annotation were carried out on the soft-masked genomes by integrating *ab initio*, homology-based, and transcriptome-assisted prediction approaches. *Ab initio* prediction was performed by Augustus v.3.5.0 (Keilwagen et al., 2018) and SNAP (Johnson et al., 2008). The homologous protein sequences were aligned by Tblastn v.2.7.1 (Camacho et al., 2009) and GeMoMa (Keilwagen et al., 2018), the alignment results were used to predict gene structures. For transcriptome prediction, RNA-seq data and ONT read alignment were performed by STAR v.2.7 (Dobin et al., 2013), StringTie v.2.7 (Kovaka et al., 2019) and PASA (Haas et al., 2003). All evidence was weighted and merged with EVidenceModeler (Haas et al, 2008) to produce the final non-redundant gene set (Full details in Data S1).

Gene functions were annotated by Diamond v.2.0.11.149 (Buchfink et al., 2011), based on comparisons against major databases including NCBI NR (Pruitt, Tatusova, and Maglott, 2005), InterPro (Paysan-Lafosse et al., 2023), GO (Ashburner et al., 2001), KOG (Tatusov et al., 2003), KEGG (Kanehisa et al., 2023), TrEMBL, and SwissProt (Bairoch and Apweiler, 2000). Conserved sequences, motifs, and domains of proteins were identified using InterProScan v.5.68 (Jones et al., 2014) and Hmmscan v3.3.2 (Finn et al., 2011) based on comparisons with the InterPro and Pfam databases (Finn et al., 2014). The MLA genome features were visualized using Circos (Krzywinski et al., 2009).

Gene family and genome synteny analysis

Gene families were analyzed using OrthoFinder v.2.4.0 (Emms and Kelly, 2019), and divergence

times were estimated with PAML v.4.9j (Yang, 2007). R package UpSetR (Conway et al., 2017) was used for the orthogroup distribution visualization. WGD events were analyzed using WGDI v.0.6.5 (Sun et al., 2022). Structural variations were detected with Mummer v.4.0.0 (Marçais et al., 2018) and SyRI v.1.6 (Goel et al., 2019). To examine genomic synteny among MLA and EGL/MAC/MBE, MCScanX (Wang et al., 2012) was employed, and resulting synteny blocks were visualized using SynVisio (Bandi and Gutwin, 2020).

FISH analysis

Root tip preparation and FISH were carried out following Schwarzacher and Heslop-Harrison (2000). Details of probe mixtures were provided in Table S16. Images were acquired using Nikon Eclipse 80i microscope (Nikon, Japan) equipped with UV-2A, FITC, and TRITC filters. Image overlays were created with Nikon NIS-elements software and adjusted in Photoshop v.25.6.0.

Data availability

The telomere-to-telomere genome assemblies of *Musella lasiocarpa* have been deposited in NCBI (PRJNA1117752 for MLAh1; PRJNA1117751 for MLAh2) and the website of https://banana-genome-hub.southgreen.fr/node/50/16444789 (Droc et al., 2022). The raw sequence data including Illumina, PacBio HiFi, and ONT genome sequences, as well as RNAseq data, can be accessed in the NCBI SRA library (SAMN41579419), and the China National Center for Bioinformation at https://ngdc.cncb.ac.cn/gsa/browse/CRA014572.

CRediT authorship contribution statement

Qing Liu: Conceptualization, Data Curation, Funding acquisition, Project administration, Supervision, Writing-Original draft, Writing-Review & Editing. **Dongli Cui**: Methodology, Writing-Original draft, Writing-Review & Editing. **Yaqi Tian, Yehan Wang**: Resources, Investigation, Methodology, Writing-Original draft. **Ziwei Wang** and **Mathieu Rouard**: Formal analysis, Investigation, Data curation, Software, Validation. **John Seymour Heslop-Harrison** and **Trude Schwarzacher**: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing-Review & Editing.

Conflict of interest

The authors declare no competing interests.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32070359, 32370402), Guangdong Flagship Project of Basic and Applied Research (2023B0303050001), Chinese Academy of Sciences (CAS) President's International Fellowship Initiative (2024PVA0028), Compilation Project for Flora of Nanling Grasses (Forestry Administration of Guangdong Province), UK Research and Innovation (UKRI) via the Engineering and Physical Sciences Research Council (EPSRC; EP/Y00597X/1-project RP13W471907), and UK Biotechnology and Biological Sciences Research Council (BB/R022828/1). We would like to thank Jing Li for MLA sample collection, Benagen Technology Company Ltd. (Wuhan, China) for Illumina, Nanopore transcriptome sequencing and Beijing Genomics Institute-Shenzhen (BGI, Shenzhen) for PacBio HiFi sequencing, and T2T genome assembly.

Supplementary data

Additional Supporting Information can be found in the online version of this article.

Data S1. Extended Materials and Methods.

Fig. S1. *Musella lasiocarpa* haplotype 2 (MLAh2) genomic features.

Fig. S2. Genome size and BUSCO comparison of MLA and 13 other monocotyledonous plants. **A** Genome size and heterozygosity estimated from *k*-mer frequency distribution of MLA Illumina data. **B** Comparison of genome assembly completeness by BUSCO scores of 14 monocotyledonous plant genome assemblies.

Fig. S3. Hi-C interaction heatmap of MLA haplotypes.

Fig. S4. Synonymous substitutions per site (Ks) plot of orthologous gene pairs between MLAh1 as compared to itself (paralogs) and other species.

Fig. S5. GO-terms of MLA gene families. **A** GO-terms of the unique (Top left; Table S9) and expanded (Top right; Table S10) gene families of MLA. **B** GO-terms of up- (Bottom left; Table S12C) and down-regulated (Bottom right; Table S12D) DEGs of MLA varieties under cold treatment.

Fig. S6. Localization of transcription factors identified from up- and down-regulated DEGs of two MLA varieties under cold treatment and structural variations between MLA haplotypes.

Fig. S7. Chromosomal synteny between MLAh1 and MLAh2. **A** Chromosome MLA01. **B** Chromosome MLA02. **C** Chromosome MLA03. **D** Chromosome MLA04. **E** Chromosome MLA05. **F** Chromosome MLA06. **G** Chromosome MLA07 left arm. **H** Chromosome MLA07 right arm. **I** Chromosome MLA08. **J** Chromosome MLA09.

Fig. S8. The ONT long-reads evidence for breakpoints of the longest inversions on chromosomes 3, 4 and 8 between MLA haplotypes. **A** The inversion structure of MLA03 region. **B** The

inversion structure of MLA04 region. **C** The inversion structure of MLA08 region.

Fig. S9. Abundance and copy number of MLA repetitive DNAs. **A** MLA haplotype sequence pairwise comparison by D-Genies. **B** Distribution of RepeatExplorer clusters.

Fig. S10. FISH distribution of telomere, MLcen, and 45S rDNA probes on MLA metaphase chromosomes. **A** MLA metaphase chromosomes stained by the telomere and MLcen probes. **B** MLA metaphase chromosomes stained by Mlcen probe. **C-F** Distribution of 45S rDNA and telomere probes on MLA metaphase chromosomes. **D** The 45S rDNA loci labeled by green points **E** The telomere sequences labeled by red points. **F** DAPI image of chromosomes is shown in white.

Fig. S11. Dotplots of the 45S rDNA array on chromosome MLA06.

Fig. S12. Dotplots of the 45S rDNA array on chromosome MLA08.

Fig. S13. Dotplot between *Musella lasiocarpa* and *Ensete glaucum* genome assemblies.

Table S1. Statistics of genome sequencing and assembly of *Musella lasiocarpa* (MLA) haplotypes. **A** Summary of genome sequencing and transcriptome data of MLA haplotypes. **B** Statistics for MLA haplotype assemblies.

Table S2. Chromosome and genome sizes of MLA haplotypes. **A** Chromosome size of MLA. **B** Genome size of MLA.

Table S3. BUSCO accessment and gene function annotation of MLA haplotype assemblies. **A** BUSCO of MLA haplotype assemblies. **B** BUSCO assessment of genes of MLA haplotype assemblies. **C** Functional annotation of genes of MLA haplotype assemblies. **D** Genome assemblies of investigated species in Fig. S2B.

Table S4. Telomere characterization of MLA haplotypes.

Table S5. Centromere characterization of MLA haplotypes. **A** Abundance of Mlcen tandem repeat array in MLA. **B** Centromere identification by tidk. **C** Centromere positions and chromosome arm lengths in MLA haplotypes. **D** Data used for drawing the karyotype of Fig. 6B.

Table S6. The 5S and 45S rDNA arrays in MLA haplotypes. **A** Position and length of rDNA arrays in MLA genome. **B** The rDNA copy number estimation using map to reference in Illumina reads.

Table S7. Statistics of orthogroups and gene families in Musaceae. **A** Shared orthogroups and gene clustering of the investigated species in Musaceae. **B** Pairwise gene numbers in syntenic region between the species pairs in Musaceae. **C** Database of 31,235 orthogroups in Musaceae.

Table S8. Statistics and database of orthologous gene families for the investigated species in Fig. 2C. **A** Gene family categories in genomes of MLA and 13 other monocotyledonous species. **B** Statistics of gene family expansion and contraction for each clade. **C** Significant expanded and contracted gene families and genes of MLA. **D** Database of orthologous gene families for the investigated species.

Table S9. GO enrichment results of unique gene families in MLA (Fig. S5A).

Table S10. GO enrichment results of significantly expanded gene families in MLA (Fig. S5B).

Table S11. A total of 88 upregulated and 1,470 down-regulated DEGs shared by MLA varieties under cold treatment.

Table S12. KEGG and GO enrichment results of MLA varieties under cold treatment. **A** KEGG of 38 genes among 88 up-regulated DEGs (Fig. 2D, bottom left). **B** KEGG of 410 genes among 1,470 down-regulated DEGs (Fig. 2D, bottom right). **C** GO-terms of up-regulated DEGs (Fig. S5C). **D** GO-terms of down-regulated DEGs (Fig. S5D).

Table S13. Expression levels before row scale normalization for up-regulated and down-regulated DEGs of MLA varieties under cold treatment.

Table S14. Statistics of the structural variations detected in MLA haplotypes by SyRI.

Table S15. Repetitive DNA composition of MLA haplotypes. **A** Repetitive DNA proportion of MLA haplotypes. **B** Statistics of repetitive DNA proportion of MLA haplotypes.

Table S16. Fluorescence *in situ* hybridization (FISH) probes used in FISH experiments.

References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A. P., Dolinski, K., Dwight, S.S., Eppig J.T., et al., 2001. Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.

Bairoch, A., Apweiler, R., 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28, 45–48.

Bandi, V., Gutwin, C., 2020. Interactive exploration of genomic conservation. In, Proceedings of the 46th Graphics Interface Conference on Proceedings of Graphics Interface 2020 (GI'20). Waterloo, Canada, Canadian Human-Computer Communications Society.

Bao, Y., Zeng, Z., Yao, W., Chen, X., Jiang, M.W., Sehrish, A., Wu, B., Powell C.A., Chen, B.S., Xu J.L., et al., 2023. A gap-free and haplotype-resolved lemon genome provides insights into flavor synthesis and huan-glongbing (HLB) tolerance. Hortic. Res. 10, uhad020.

Bartoš, J., Alkhimova, O., Doleželová, M., De Langhe, E., Dolezel, J., 2005. Nuclear genome size and genomic

distribution of ribosomal DNA in *Musa* and *Ensete* (Musaceae): taxonomic implications. Cytogenet. Genome Res. 109, 50–57.

Belser, C., Baurens, F.C., Noel, B., Martin, G., Cruaud, C., Istace, B., Yahiaoui, N., Labadie, K., Hřibová, E., Doležel, J., et al., 2021. Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. Commun. Biol. 4, 1047.

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.C., Falentin, C., Genete, M., Berrabah, W., Chèvre, A.M., Delourme, R., et al., 2018. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat. Plants 4, 879–887.

Belton, J.M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., Dekker, J., 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58, 268–276.

Beránková, D., Čížková, J., Majzlíková, G., Doležalová, A., Mduma, H., Brown, A., Swennen, R., Hřibová, E., 2024. Striking variation in chromosome structure within *Musa acuminata* subspecies, diploid cultivars, and F1 diploid hybrids. Front. Plant Sci. 15, 1387055.

Buchfink, B., Reuter, K., Drost, H.G., 2011. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods 18, 366–368.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421.

Cheesman, E.E., 1947. Classification of the bananas. Kew Bull. 2, 97–117.

Chen, C.J., Wu, Y., Li, J.W., Wang, X., Zeng, Z.H., Xu, J., Liu Y.L, Feng J.T., Che, H., Ye, Y.H., et al., 2023. Tbtools-II: a "one for all, all for one" bioinformatics platform for biological big-data mining. Mol. Plant 16, 1733–1742.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.

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

Čížková, J., Hřibová, E., Humplikova, L., Christelová, P., Suchánková, P., Doležel, J., 2013. Molecular analysis and genomic organization of major DNA satellites in banana (*Musa* spp.). PLoS One 8, e54808.

Conway, J.R., Lex, A., Gehlenborg, N., 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics, 33, 2938–2940.

Cui, D.L., Xiong, G., Ye, L.H., Gornall, R., Schwarzacher, T., Heslop-Harrison, J.S., Liu, Q., 2024. Genome-wide analysis of flavonoid biosynthetic genes in Musaceae (*Ensete*, *Musella*, and *Musa* species) reveals amplification of flavonoid 3'5'-hydroxylase. AoB Plants, 16, plae049.

D'Hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., et al., 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488, 213.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., Aiden, E.L., 2016a. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98.

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

Droc, G., Martin, G., Guignon, V., Summo, M., Sempéré, G., Durant, E. et al., 2022. The banana genome hub, a community database for genomics in the Musaceae. Hortic. Res. 9, uhac221.

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al., 2014. Pfam: the protein families database. Nucleic Acids Res. 42, D222–D230.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., et al., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. 117, 9451–9457.

Franchet, M.Z., 1889. Un nouveau type de *Musa lasiocarpa*. J. Bot. 3, 329–331.

Fu, N., Ji, M., Rouard, M., Yan, H.F., Ge, X.J., 2022. Comparative plastome analysis of Musaceae and new insights into phylogenetic relationships. BMC Genom. 23, 223.

Galvez, L.C., Koh, R.B.L., Barbosa, C.F.C., Asunto, J.C., Catalla, J.L., Atienza, R.G. et al., 2021. Sequencing and *de novo* assembly of abaca (*Musa textilis* Née) var. *abuab* genome. Genes 12, 1202.

Goel, M., Sun, H., Jiao, W.B., Schneeberger, K., 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20, 1–13.

Haile, A.T., Kovi, M.R., Johnsen, S.S., Hvoslef-Eide, T., Tesfaye, B., Rognli, O.A., 2024. Limited genetic diversity found among genotypes of the Entada landrace (*Ensete ventricosum*, (Welw.) Chessman) from Ethiopia. Front. Plant Sci. 15, 1336461.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith Jr, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al., 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31, 5654–5666.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C, R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9, R7d.

Heslop-Harrison, J.S., Schwarzacher, T., Liu, Q., 2023. Polyploidy: its consequences and enabling role in plant diversification and evolution. Ann. Bot. 131, 1–10.

Heslop-Harrison, J.S., Schwarzacher, T., 2007. Domestication, genomics and the future for banana. Ann. Bot. 100, 1073–1084.

Hřibová, E., Neumann, P., Matsumoto, T., Roux, N., Macas, J., Dolezel, J., 2010. Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. BMC Plant Biol. 10, 204.

Huang, H.R., Liu, X., Arshad, R., Wang, X., Li, M.W., Zhou, Y.F. et al., 2023. Telomere-to-telomere haplotype-resolved reference genome reveals subgenome divergence and disease resistance in triploid Cavendish banana. Hortic. Res.10, uhad153.

Jenny, C., Guignon, V., Manyer, I., Ballester, F., Ruas, M., Rouard, M., 2024. Collecting and managing *in situ* banana genetic resources information (*Musa* spp.) using online resources and citizen science. Database, baae036.

Jin, J.P., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J.C. et al., 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research, 45, D1040–D1045.

Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., Donnell, C.J.O., de Bakker, P.I.W., 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24, 2938–2939.

Joshi, R.U., Singh, A.K., Singh, V.P., Rai, R., Joshi, P., 2023. A review on adaptation of banana (*Musa* spp.) to cold in subtropics. Plant Breed. 142, 269–283.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W.Z., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka G., et al., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., Ishiguro-Watanabe, M., 2023. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 51, D587–D592.

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., Grau, J., 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinf. 19, 189.

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., Pertea, M., 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 20, 278.

Krzywinski, M., Schein, J., Birol, I., Gascoyne, R., Horsman, D., Jones, S.J., Jones, S.J., Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645.

Li, G., Tang, L., He, Y., Xu, Y.Y., Bendahmane, A., Garcia-Mas, J., Lin, T., Zhao, G.W., 2023. The haplotype-resolved T2T reference genome highlights structural variation underlying agronomic traits of melon. Hortic. Res. 10, uhad182.

Li, H., Berent, E., Hadjipanteli, S., Galey, M., Muhammad-Lahbabi, N., Miller, D.E., Crown, K.N., 2023. Heterozygous inversion breakpoints suppress meiotic crossovers by altering recombination repair outcomes. PLoS Genet. 3, 1010702.

Li, H.L., Dong, Z.M., Jiang, Y.S., Jiang, Y.S., Jiang, S.J., Xing, H.T., Li, Q., Liu, G.C., Tian, S.M., Wu, Z.Y., et al., 2021. Haplotype-resolved genome of diploid ginger (*Zingiber officinale*) and its unique gingerol biosynthetic pathway. Hortic. Res. 8, 189.

Li, K., Chen, R.H., Abudoukayoumu, A., Wei, Q., Ma, Z.B., Wang, Z.Y., Hao, Q., Huang, J., 2024. Haplotype-resolved T2T reference genomes for wild and domesticated accessions shed new insights into the domestication of jujube. Hortic. Res. 11, uhae071.

Li, W., Chu, C., Li, H., Zhang, H.T., Sun, H.C., Wang, S.Y., Wang, Z.J., Li, Y.Q., Foster, T.M., López-Girona, E., et al., 2024. Near-gapless and haplotype-resolved apple genomes provide insights into the genetic basis of rootstock-induced dwarfing. Nat. Genet. 56, 505–516.

Li, X.X., Yu, S., Cheng, Z.H., Chang XJ, Yun YZ, Jiang, M.W., Chen, X.Q., Wen, X.H., Li, H., Zhu, W.J., et al., 2024. Origin and evolution of the triploid cultivated banana genome. Nat. Genet. 56, 136–142.

Li, Y.P., Shi, Y.T., Li, M.Z., Fu, D.Y., Wu, S.F., Li, J.G., Gong, Z.Z., Liu, H.T., Yang, S.H., 2021. The CRY2–COP1–HY5–BBX7/8 module regulates blue light-dependent cold acclimation in *Arabidopsis*. Plant Cell 33, 3555–3573.

Li, Z., Wang, J., Fu, Y., Jing, Y.L., Huang, B.L., Chen, Y., Wang, Q.L., Wang, X.B., Meng, C.Y., Yang, Q.Q., et al., 2022. The *Musa troglodytarum* L. genome provides insights into the mechanism of non-climacteric behaviour and enrichment of carotenoids. BMC Biol., 20, 186.

Liu, A.Z., Kress, W.J., Li, D.Z., 2002. Insect pollination of *Musella lasiocarpa* (Musaceae): a monotypic genus endemic to China. Plant Syst. Evol. 235, 135–146.

Liu, Q., Li, X.Y., Zhou, X.Y., Li, M.Z., Zhang, F.J., Schwarzacher, T., Heslop-Harrison, P., 2019. The DNA landscape in *Avena*: chromosome and genome evolution defined by major repetitive DNA classes in whole-genome sequence reads. BMC Plant Biol. 19, 226.

Liu, Q., Ye, L.H., Li, M.Z., Wang, Z.W., Xiong, G., Ye, Y.S., Tu, T.Y., Schwarzacher, T., Heslop-Harrison, J.S., 2023a. Genome-wide expansion and reorganization during grass evolution: from 30 Mb chromosomes in rice and *Brachypodium* to 550 Mb in *Avena*. BMC Plant Biol. 23, 627.

Liu, Q., Yuan, H.Y., Xu, J.X., Cui, D.L., Xiong, G., Schwarzacher, T., Heslop-Harrison, J.S., 2023b. The mitochondrial genome of the diploid oat *Avena longiglumis*. BMC Plant Biol. 23, 218.

Liu, X., Arshad, R., Wang, X., Li, W.M., Zhou, Y.F., Ge, X.J., Huang, H.R., 2023. The phased telomere-to-

telomere reference genome of *Musa acuminata*, a main contributor to banana cultivars. Sci. Data 10, 631.

Long, C.L., Ahmed, S., Wang, X.Y., Liu, Y.T., Long, B., Yang, C.Y., Shi, Y.N., Li, X.Y., Guo, R., 2008. Why *Musella lasiocarpa* (Musaceae) is used in Southwest China to feed pigs. Econ. Bot. 62, 182–186.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Lu, Y.B., Chen, X., Yu, H., Zhang, C., Xue, Y.J., Zhang, Q., Wang, H.F., 2024. Haplotype-resolved genome assembly of *Phanera championii* reveals molecular mechanisms of flavonoid synthesis and adaptive evolution. Plant J. 118, 488–505.

Luo, Z.Y., Zhou, Z.C., Li, Y.Y., Tao, S.T., Hu, Z.Y., Yang, J.S., Cheng, XJ., Hu, R.S., Zhang, W.L., 2022. Transcriptome-based gene regulatory network analyses of differential cold tolerance of two tobacco cultivars. BMC Plant Biol. 22, 369.

Ma, H., Pan, Q.J., Wang, L., Li, Z.H., Wan, Y.M., Liu, X.X., 2011. *Musella lasiocarpa* var. *rubribracteata* (Musaceae), a new variety from Sichuan, China. Novon 21, 349–353.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. BMC Plant Biol. 38, 4647–4654.

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

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

Martin, G., Carreel, F., Coriton, O., Hervouet, C., Cardi, C., Derouault, P., Roques, D., Salmon, F., Rouard, M., Sardos, J., et al., 2017. Evolution of the banana genome (*Musa acuminata*) is impacted by large chromosomal translocations. Mol. Biol. Evol. 34, 2140–2152.

Martin, G., Istace, B., Baurens, F.C., Belser, C., Hervouet, C., Labadie, K., Cruaud, C., Noel, B., Chantal, G., Salmon, F., et al., 2025. Unravelling genomic drivers of speciation in *Musa* through genome assemblies of wild banana ancestors. Nat. Commun. 16, 961.

Nair, A.S., Teo, C.H., Schwarzacher, T., Heslop-Harrison, J.S., 2005. Genome classification of banana cultivars from South India using IRAP markers. Euphytica 144, 285–290.

Novák, P., Neumann, P., Macas, J., 2020. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nat. Protoc. 15, 3745–3776.

Osuji, J.O., Crouch, J., Harrison, G., Heslop-Harrison, J.S., 1998. Molecular cytogenetics of *Musa* species, cultivars and hybrids, localisation of 18S-5.8S-25S and 5S rDNA and telomere-like sequences. Ann. Bot. 82, 243–248.

Ou, S.J., Su, W.J., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al., 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20, 275.

Pashapu, A.R, Statkevičiūtė, G., Sustek-Sánchez, F., Kovi, M.R., Rognli, O.A., Sarmiento, C., Rostoks, N., Jaškūnė, K., 2024. Transcriptome profiling reveals insight into the cold response of perennial ryegrass genotypes with contrasting freezing tolerance. *Plant Stress*, 14, 100598.

Paysan-Lafosse, T., Blum, T., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al., 2023. InterPro in 2022. Nucleic Acids Res. 51, D418–D427.

Pfenninger, M., Schönenbeck, P., Schell, T., 2022. ModEst: accurate estimation of genome size from next-generation sequencing data. Mol. Ecol. Resour. 22, 1454–1464.

Plummer, J., Allen, R., Kallow, S., 2022. *Musella lasiocarpa*. The IUCN Red List of Threatened Species. e.T98249468A.

Pruitt, K.D., Tatusova, T., Maglott, D.R., 2005. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts, and proteins. Nucleic Acids Res. 33, D501–D504.

Sardos, J., Breton, C., Perrier, X., den Houwe, I.V., Carpentier, S., Paofa, J., Rouard, M., Roux, N., 2022. Hybridization, missing wild ancestors and the domestication of cultivated diploid bananas. Front. Plant Sci. 13, 969220.

Schwarzacher, T., Heslop-Harrison, P., 2000. Practical *in situ* hybridization. BIOS Scientific Publishers Ltd.

Sepsi, A., Higgins, J.D., Heslop-Harrison, J.S., Schwarzacher, T., 2017. CENH3 morphogenesis reveals dynamic centromere associations during synaptonemal complex formation and the progression through male meiosis in hexaploid wheat. Plant J. 89, 235–249.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Sun, P.C., Jiao, B.B., Yang, Y.Z., Shan, L., Li, T., Li, X.N., Xi, Z.X., Wang, X.Y., Liu, J.Q., 2022. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Mol. Plant 15, 1841–1851.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al., 2003. The COG database: an updated version includes eukaryotes. BMC Bioinf. 4, 41.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M. C., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204.

Wang, Z., Miao, H.X., Liu, J.H., Xu, B. Y., Yao, X.M., Xu, C.Y., Zhao, S.C., Fang, X.D., Jia, C.H., Wang, J.Y., et al., 2019. *Musa balbisiana* genome reveals subgenome evolution and functional divergence. Nat. Plants 5, 810–821.

Wang, Z., Rouard, M., Biswas, M.K., Droc, G., Cui, D.L., Roux, N., Baurens, F. C., Ge, X.J., Schwarzacher, T., Heslop-Harrison, J. S., et al., 2022. A chromosome-level reference genome of *Ensete glaucum* gives insight into diversity and chromosomal and repetitive sequence evolution in the Musaceae. GigaScience 11, giac027.

Wang, ZF., Rouard, M., Droc, G., Heslop-Harrison, J.S., Ge, X.J., 2023. Genome assembly of *Musa beccarii* shows extensive chromosomal rearrangements and genome expansion during evolution of Musaceae genomes. GigaScience 12, giad005.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J.P., Wang, X.Y., Lee, T.H., Jin, H.Z., Marler, B., Guo, H., et al., 2012. MCScanX, a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40, e49.

Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb. Genom. 3, e000.132.

Xiao, T.W., Liu, X., Fu, N., Liu, T.J., Wang, Z.F., Ge, X.J., Huang, H.R., 2024. Chromosome-level genome assemblies of *Musa ornata* and *Musa velutina* provide insights into pericarp dehiscence and anthocyanin biosynthesis in banana. Hortic. Res. 11, uhae079.

Xie, Z.W., Zheng, Y.Y., He, W.D., Bi, F.C., Li, Y.Y., Dou, T.X., Zhou, R., Guo, Y.X., Deng, G.M., Zhang, W.H., et al., 2024. Two haplotype-resolved genome assemblies for AAB allotriploid bananas provide insights into banana subgenome asymmetric evolution and *Fusarium wilt* control. Plant Commun. 5, 100766.

Xiong, G., Cui D.L., Tian, Y.Q., Schwarzacher, T., Heslop-Harrison, J.S., Liu, Q., 2024. Genome-wide identification of lectin receptor-like kinases gene family in *Avena sativa* and their roles in salt stress tolerance. Int. J. Mol. Sci. 25, 12754.

Xue, C.Y., Wang, H., Li, D.Z., 2007. Female gametophyte and seed development in *Musella lasiocarpa* (Musaceae), a monotypic genus endemic to Southwestern China. Can. J. Bot. 85, 964–975.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. BMC Plant Biol. 24, 1586–1591.

Yang, Q.S., Gao, J., He, W.D., Dou, T.X., Ding, L.J., Wu, J.H., Li, C.Y., Peng, X.X., Zhang, S., Yi, G.J., 2015. Comparative transcriptomics analysis reveals difference of key gene expression between banana and plantain in response to cold stress. BMC Genom. 16, 1–8.

Yang Y, Du W, Li Y, Lei J, Pan W. 2025. Recent advances and challenges in *de novo* genome assembly. Genomics Comm. 2, e014.

Zhang, L.Y., Jiang, X.C., Liu, Q.Y., Ahammed, G.J., Lin, R., Wang, L.Y., Shao, S.J., Yu, J.Q., Zhou, Y.H., 2020. The HY5 and MYB15 transcription factors positively regulate cold tolerance in tomato via the CBF pathway. Plant Cell Environ. 43, 2712–2716.

Zhang, R.G., Li, G.Y., Wang, X.L., Dainat, J., Wang, Z.X., Ou, S.J., Ma, Y.P., 2022. TEsorter, an accurate and fast method to classify LTR-retrotransposons in plant genomes. Hortic. Res. 9, uhac017.

Zhao, W.L., Wu, J.Z., Tian, M., Xu, S., Hu, S.Y., Wei, Z.Y., Lin, G.Y., Tang L., Wang, R.Y., Feng, B.Y., et al., 2024. Characterization of O-methyltransferases in the biosynthesis of phenylphenalenone phytoalexins based on the telomere-to-telomere gapless genome of *Musella lasiocarpa*. Hortic. Res. 11, uhae042.

Zhou, R., Jenkins, J.W., Zeng, Y., Shu, S.Q., Jang, H.S., Harding, S.A., Williams, M., Plott, C., Barry, K.W., Koriabine, M., et al., 2023. Haplotype-resolved genome assembly of *Populus tremula* × *P. alba* reveals aspen-specific megabase satellite DNA. Plant J. 116, 1003–1017.

Zhou, R., Wang, S., Zhan, N., He, W.D., Deng, G.M., Dou, T.X., Zhu, X.T., Xie, W.Z., Zheng, Y.Y., Hu, C.H., et al., 2024. High-quality genome assemblies for two *Australimusa* bananas (*Musa* spp.) and insights into regulatory mechanisms of superior fiber properties. Plant Commun. 5, 100681.

Tables

Table 1. Summary of genome assembly statistics and functional annotation of MLA haplotypes.

Genome assembly	**MLAh1 (Reference)**	**MLAh2**
Estimated genome size by backmap [a)]	503.58 Mb	
Assembly size	500,050,725 bp	498,373,817 bp
Percent assembled of estimated genome	99.30%	98.97%
GC Ratio	40%	41%
Total number of contigs	365	295
Contig N50[b)]	49,080,785 bp	18,553,585 bp
Total number of scaffolds	357	274
Scaffold N50	51,603,189 bp	53,045,455 bp
Anchored into pseudo-chromosomes	474,465,207 bp	462,245,547 bp
Number of pseudo-chromosomes	9	9
Quality of genome assembly		
Hi-C mounting rate	94.88%	92.75%
BUSCO completeness of assembly	98.51%	98.82%
Functional annotation		
Number of predicted genes [c)]	35,387	35,237
BUSCO completeness of annotation	97.96%	97.83%
Proportion of function-annotated gene number (%)	33,232 (93.91%)	33,172 (94.14%)

a) MLAh1 genome assembly analyzed by backmap pl v.0.5 (Pfenninger et al., 2022) see Table S2

b) Longer N50 assigned to haplotype h1

c) Gene number predicted by *ab initio*, transcriptome, and homolog-based structure prediction

Table 2. Summary of *Musella lasiocarpa* genome assembly—gene prediction and repetitive DNA proportions.

a. Protein-coding genes (PCGs)	
Number of PCGs	35,387 (MLAh1) 35,237 (MLAh2)
Average length per gene	4498.07bp
Average number of exons per gene	5.20
Average length of exon per gene	237.02 bp
Average length of intron per gene	777.77 bp
Average length of proteins coded	410.75 aa
b. Repetitive DNA proportion by RepeatMasker (Average of MLAh1 and MLAh2)	**52.81%**
Class I retroelements	34.85%
LTR retroelements	34.40%
Copia	14.52%
Gypsy	11.92%
Unknown or mixture	7.87%
Non-LTR LINES	0.46%
Class II DNA transposons	9.01%
MITEs	3.69%
Unknown interspersed repeats	3.98%
Simple repeats and low complexity	1.29%
c. Tandem repeat proportion in Illumina reads	**5.68%**
45S rDNA	2.58%
5S rDNA	0.27%
Centromeric sequence Mlcen	2.33%
Microsatellites (<8 bp motif) [a)]	0.50%

a) Data summary from EDTA (Ou et al., 2019).

Figure legends

Fig. 1. Plant morphology, habitat and genomic features of *Musella lasiocarpa* (MLA). **A**: Cliff habitat (Nanhua County, Chuxiong Yi Autonomous Prefecture, Yunnan, China). **B**: *M. lasiocarpa* var. *rubibracteata* in the wild. **C**-**D**: *M. lasiocarpa* var. *lasiocarpa* in the greenhouse of South China National Botanical Garden. **E**: MLAh1 (chromosomes designated haplotype 1) genomic features. (**a**) Nine chromosomes (scale: 5.0 Mb) with green, pink, and black boxes representing 5S (MLA05.1), 45S (MLA06.1 and MLA08.1) rDNA and centromere positions; (**b**) Gene density; (**c**) Repeat density; (**d**) *Copia* density; (**e**) *Gypsy* density; (**f**) DNA transposon density; (**g**) Simple repeat density; (**h**) Syntenic genomic blocks, connected with curves. b–h: 10,000 bp bins. For MLAh2 genomic features see Fig. S1.

Fig. 2. Comparative analysis of gene families between *Musella lasiocarp*a (MLA), five Musaceae species, and other monocotyledonous species, and identification of differentially expressed genes (DEGs) of MLA varieties under cold treatment. **A**: UpSetR diagram of shared orthogroups in MLA and five Musaceae species (MAC: *Musa acuminata*; MTE: *M. textilis*; MBE: *M. beccarii;* EVE: *Ensete ventricosum*; EGL: *E. glaucum*). The shared orthogroups (OGs; ≥ two sequences/OG) are listed for each species set. **B**: Synonymous substitutions per site (Ks) plot showing whole genome duplication (WGD) of MLA_self (MLAh1) and divergence events between MLA-EGL and MLA-MAC (~0.53 inset), and the divergence of MLA-ZOF (~0.75; ZOF, *Zingiber officinalis*). **C**: Phylogenetic relationships and timescale of 14 monocotyledonous species inferred on the basis of 994 single-copy orthologous groups by PAML using *Oryza sativa* as outgroup species. Black numbers after branch nodes represent divergence times (Mya, million years ago) with confidence intervals of each node in square brackets, and pink (+) and blue (-) numbers near branches and species names represent expansion (pink) and contraction (blue) orthologous gene families. The green and pink boxes indicate α/β and γ WGD events, respectively. Pie diagrams on the right show the proportion of gene families undergoing expansion (pink), contraction (blue) or unchanged (yellow) (Database in Table S8D). **D**: Venn diagrams and KEGG enrichment of DEGs of MLA varieties under cold treatment. **Top left** Venn diagrams of 3,151 DEGs in var. *lasiocarpa* (Y) and 3,726 DEGs in var. *rubibracteata* (R). **Top right** Venn diagrams (Table S11) of the 88 up-regulated (Upper orange square) and 1470 down-regulated (Bottom blue square) DEGs shared by MLA varieties. **Bottom left** (Table S12A) shows KEGG enrichment analysis of up-regulated DEG proteins shared by MLA varieties. **Bottom right** (Table S12B) shows KEGG enrichment analysis of down-regulated DEG proteins shared by MLA varieties. **E**: Transcription factor (TF) identification under cold treatment. **Left**: Model of TF function under cold treatment, with the black boxes denoting downstream genes inferred from literature (Li Y et al., 2021; Pashapu et al., 2024). HY5, COP1, and PIF genes were determined based on KEGG enrichment analysis. The central box contains identified TFs, with ovals representing TFs and squares representing non-TF genes. Red words represent up-regulated, and green words represent down-regulated genes. Dashed arrows represent the regulatory pathways await to be proven. Abbreviations: CBFs, C-repeat binding factors; COP1, constitutive photomorphogenic1 protein, HY5, elongated hypocotyl 5; BBXs, B-box domain protein; ERF, ethylene response factors; PIFs, phytochrome-interacting factors; phyB, phytochrome B; COR, cold-responsive genes. **Right**: Expression levels of 14 up-regulated and 34 down-regulated DEGs in MLA varieties under cold treatment (Table S13). The relative expression levels were generated by TBtools (Chen et al., 2023) with row scale normalization. Green bars to –1.5 represent the relative gene expression level below the average, red bars to 1.5 represent gene expression level above the average.

Fig. 3. Comparison maps of MLAh1 (yellow line) and MLAh2 (blue line) by SyRI (Goel et al., 2019) plots. Syntenic regions, inversions, translocations, and duplications are indicated by gray, orange, green, and blue curves, respectively; Black boxes represent centromeres. The 5S and 45S rDNA loci are indicated by red words. Since the largest InDels occurred around centromeres, while synteny was highest along distal arm regions,

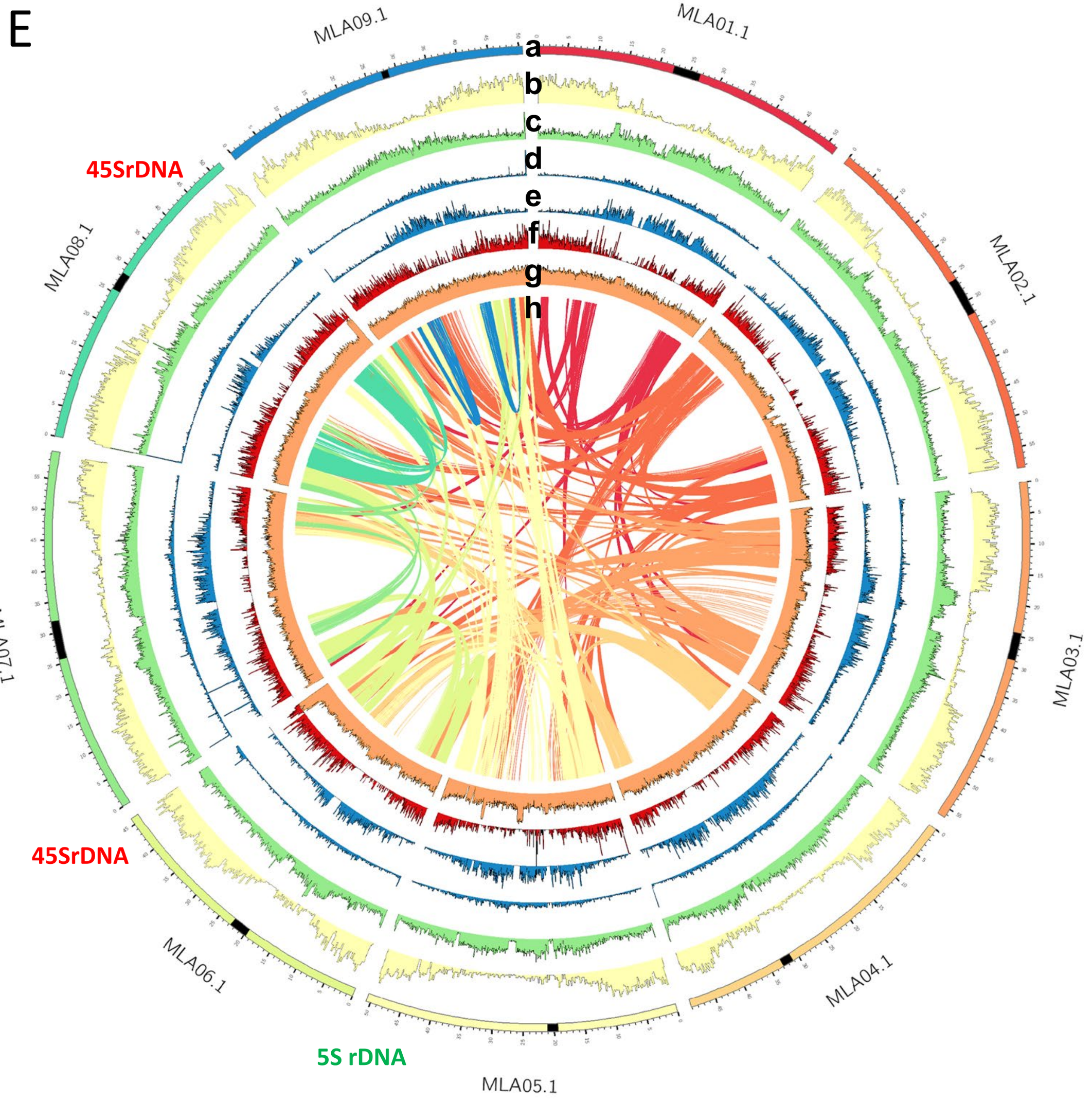
except for the more structural variations (SVs) on MLA06 left ends and on MLA08 right ends due to 45S rDNAs. To present all SVs, MLAh1 (yellow line) was placed in the central position with MLAh2 (blue line) above and below; the alignments started from the left side for the above pairs, and from the right side for the below pairs in each plot to show clearly synteny along chromosome arms. Statistics of all SVs are summarized in Table S14.

Fig. 4. MLA structural variation comparison and repeat analyses. **A**: Structural variations of MLA09 dotplot showing Ml09i, Ml09ii and Ml09iii and the centromere repeats. **B**: SNP variation between MLA haplotypes. Variation types are shown on the top, and SNP positions within genes are shown on the bottom. **C**: Retroelement insertion time with a peak value of 0.46 Mya for MLA haplotypes. **D**: MLA haplotype structural variation comparison. Left density maps represent the distribution of presence-absence variations (PAVs, red), right density maps represent retroelement variations (TEs, blue) with 0.10 Mb bins. **E**: Fluorescence *in situ* hybridization (FISH) on MLA metaphase chromosomes (blue with DAPI): (i) and (ii) Centromeric Mlcen probe (green) and telomere probe (red). Arrows represent the weak nucleolar organizer regions (NORs), and the broken satellite repeats are indicated by dotted line; (iii) 45S rDNA probe (green) and telomere probe (red). For complete images of (ii) and (iii), see Fig. S10, where enhancements showing that the telomere signals are present on all chromosome ends including those with 45S rDNA, bar = 5μm. **F**: RepeatExplorer cluster 1 (CL1) graphs and consensus sequence of 134 bp Mlcen in MLAh1. **Top left**: 'Star-shaped' tandem CL1; **Top right**: CL1 graph from TAREAN pipeline (Novák et al., 2020); **Bottom**: The consensus sequence of 134 bp Mlcen monomer variation. The black rectangle representing the functional CENP-B box motif.

Fig. 5. Distribution and structure of 5S and 45S ribosomal DNA (rDNA) in MLA genome. **A-C**: FISH on MLA metapahse chromosomes (blue with DAPI). **A**: 5S rDNA probe (red) on chromosome ml05 and Mlcen probe (green). Complete metaphase on the left and enlarged pro-metaphase chromosomes on the right showing the double 5S rDNA site. **B**: 5S (green) and 45S (red) rDNA probes. **C**: MuTR (red, arrows) and 5S rDNA (green, arrows) probes. Bar = 5 μm. **D**: Diagram showing locations of 5S rDNAs (orange squares) on ml05 and 45S rDNA loci (purple squares) on ml06 and ml08. **E**: Dotplot of MLA05.1 and MLA05.2 showing two 5S rDNA loci are located on the right arm of ml05 with the 10 Mb separation. **F**: Structural variations between MLA05.1 and MLA05.2. Syntenic regions, inversions, translocations, duplications, Mlcen, and 5S rDNA are indicated by grey, yellow, green, blue, dark blue, and red boxes. Bars above and below chromosome represent distributions of genes, coding regions (green, yellow grey), and LTR-retroelements (light red). **G**: The 45S rDNA monomer and 1,465 copy arrays. The left monomer consists of tandem repetitive units of 18S-5.8S-28S rRNA and non-transcribed spacer (NTS) regions (NTSL-STR-MuTR-NTSR). **H**: Dotplot of 45S rDNA arrays (8,347 bp) against MLAh1 (Left top). The MuTR monomers are denoted by grey dotted squares (Left up and bottom). Right blue square: Self dotplot of the inserted LTRs (6,100 bp) within 45S rDNA arrays.

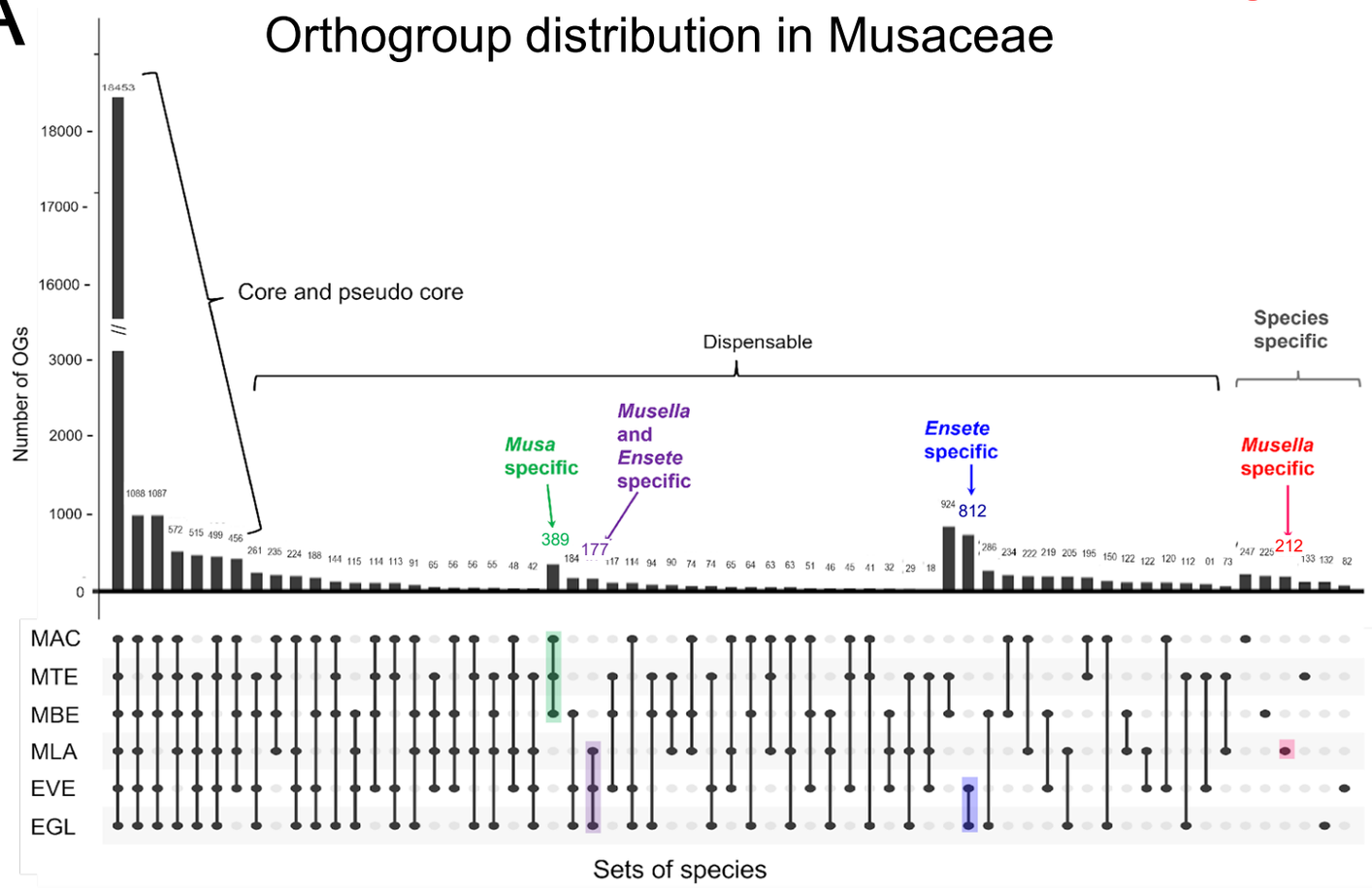
Fig. 6. Genome synteny plots of MLA and other Musaceae species. **A**: Synteny plot of MLA-*Ensete glaucum* (EGL). Centromeres are indicated by red and black dots, and 5S and 45S rDNA sites are indicated by green and yellow boxes, respectively. **B**: Synteny blocks on EGL adopted from Wang et al. (2022) and MLA karyotypes (Data from Table S5D). Syntenic regions on short and long arms are indicated by darker and lighter colors, and 5S, 45S rDNAs, and satellites are indicated by black, white, and blue boxes, respectively. **C**: Synteny plot of MLA-*Musa acuminata* (MAC). Four MAC chromosomes (MAC01/02/05/11) correspond to parts of four MLA chromosomes (MLA01/02/05/09), while the remaining seven chromosomes are split into multiple fragments, constituting nine MLA chromosomes. The syntenic regions of MLA-MAC (Fig. 6C) are few than those of MLA-EGL (Fig. 6A). **D**: Synteny plot of Musaceae showing extensive rearrangements between EGL-MBE and MBE-MAC (x = 11).

Liu et al 2026 Figure 1

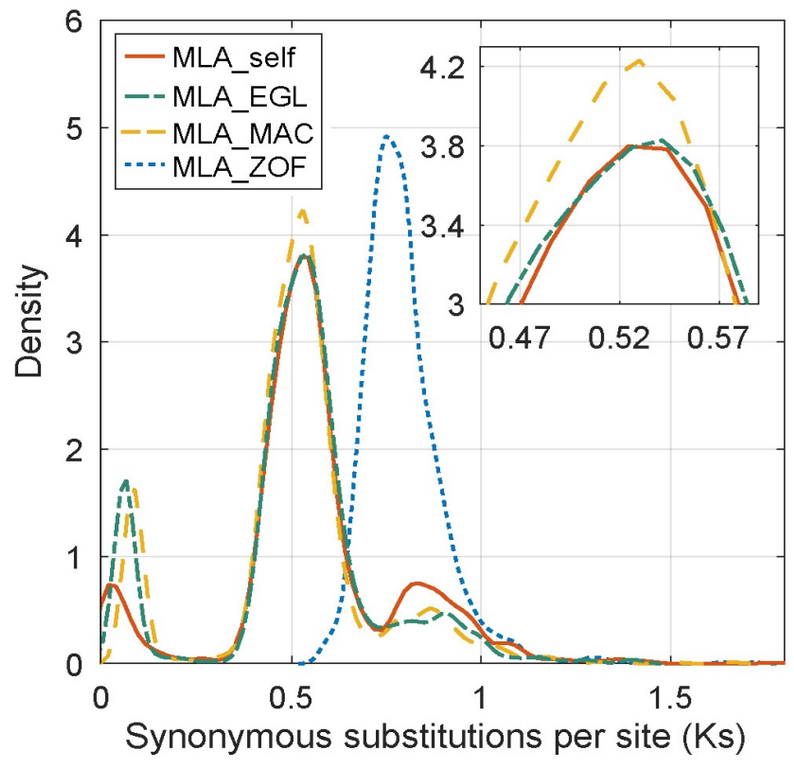


Liu et al 2026 Figure 2

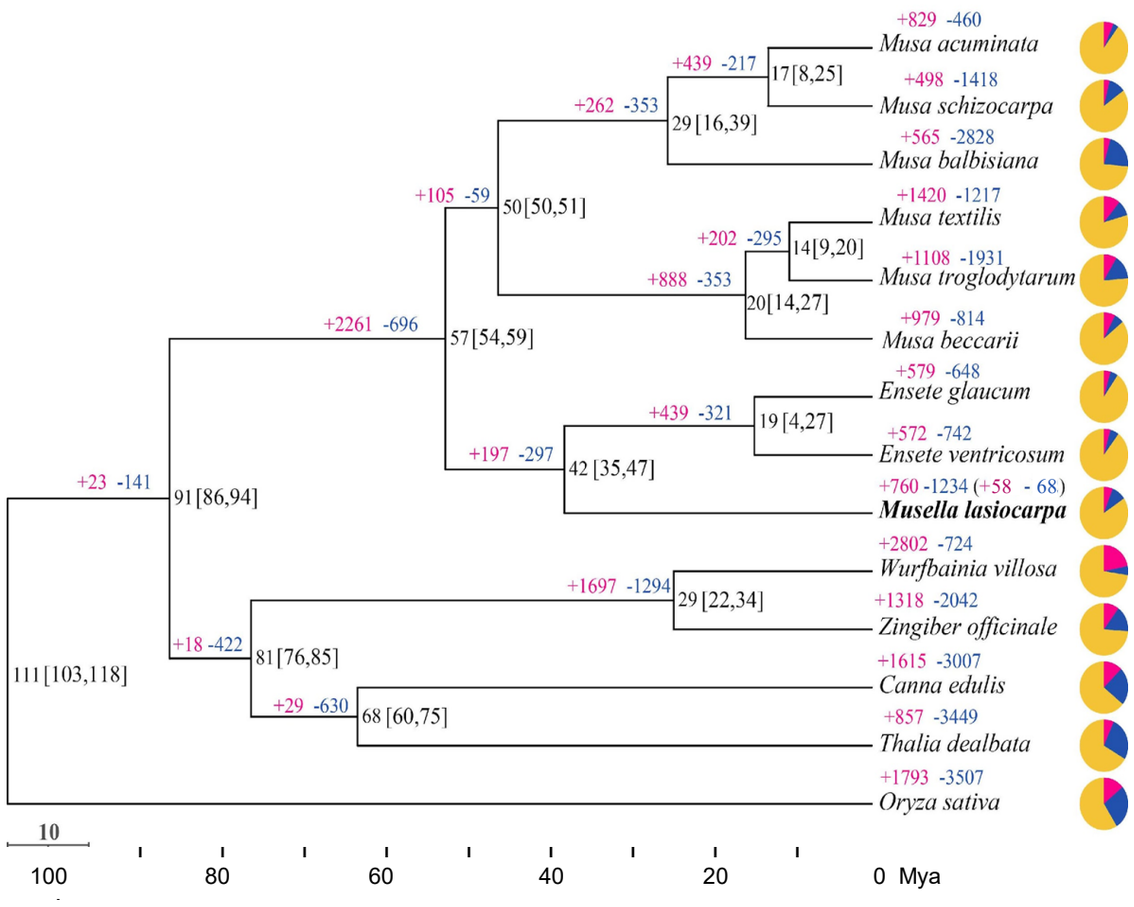
A Orthogroup distribution in Musaceae



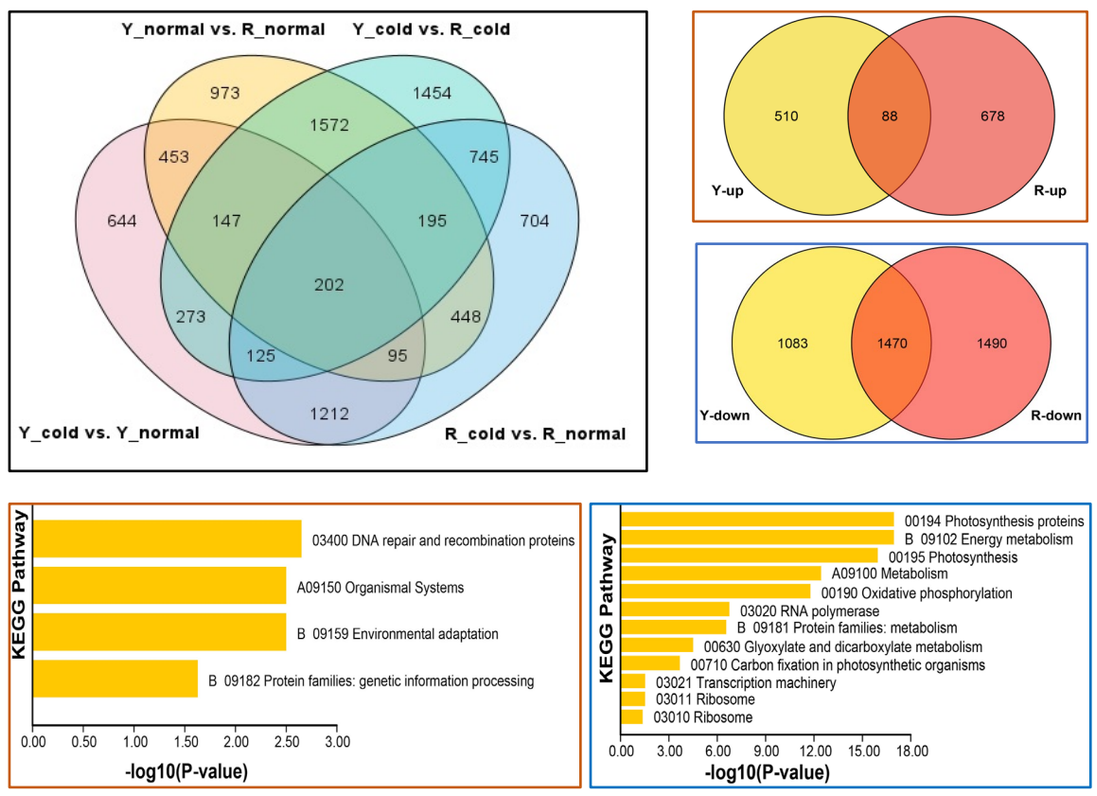
B Whole genome duplications



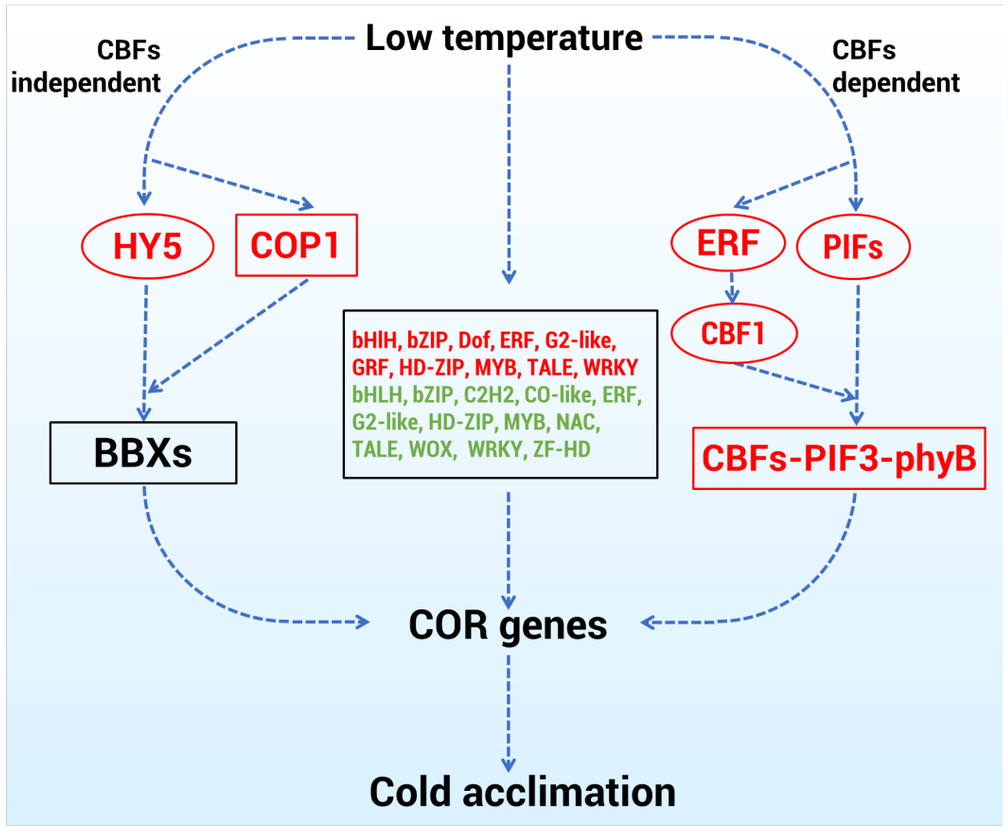
C Gene family expansions and contractions



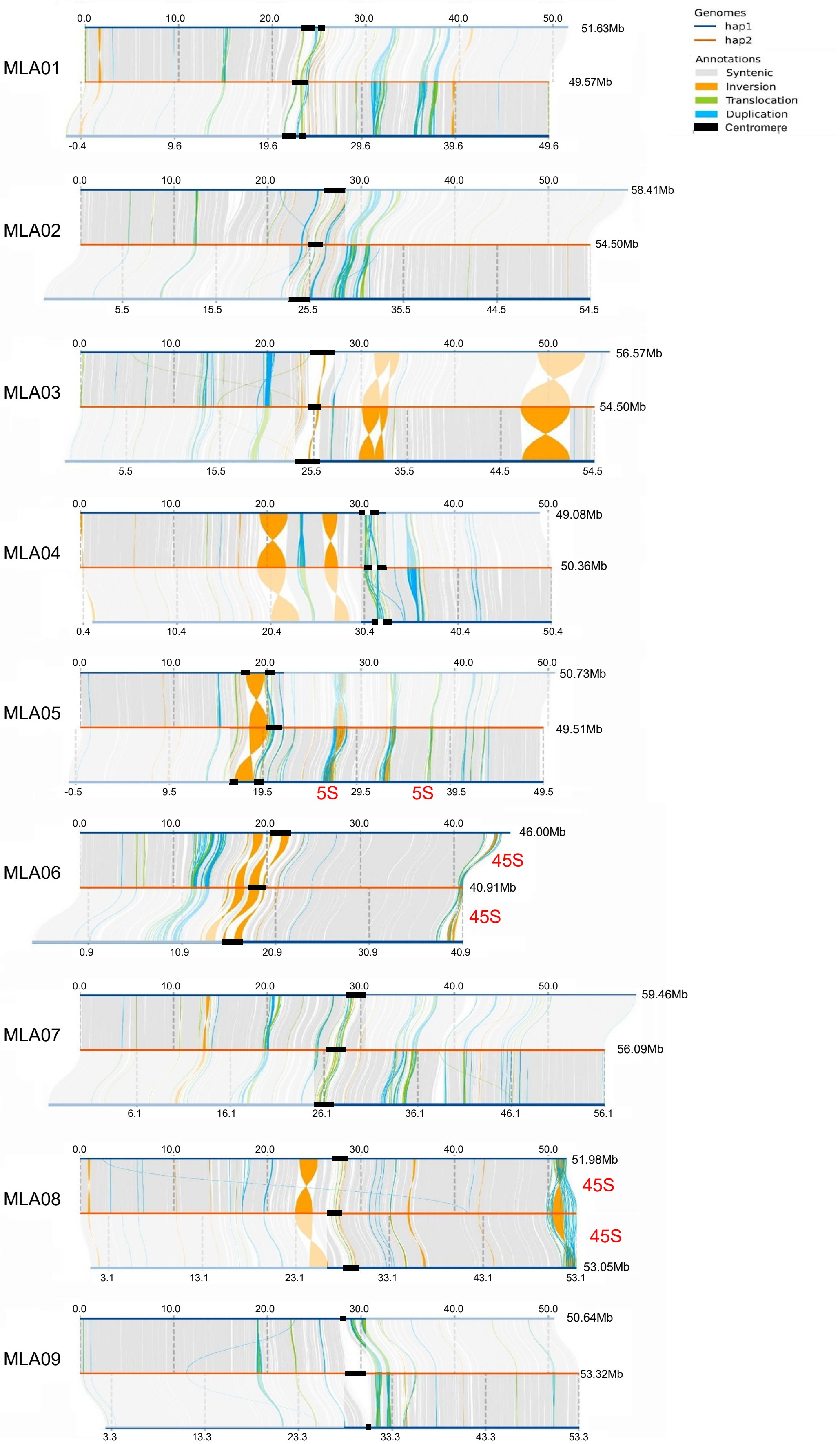
D Differential gene expression and upregulated genes under cold treatment



E Transcription factor functions and expression under cold treatment

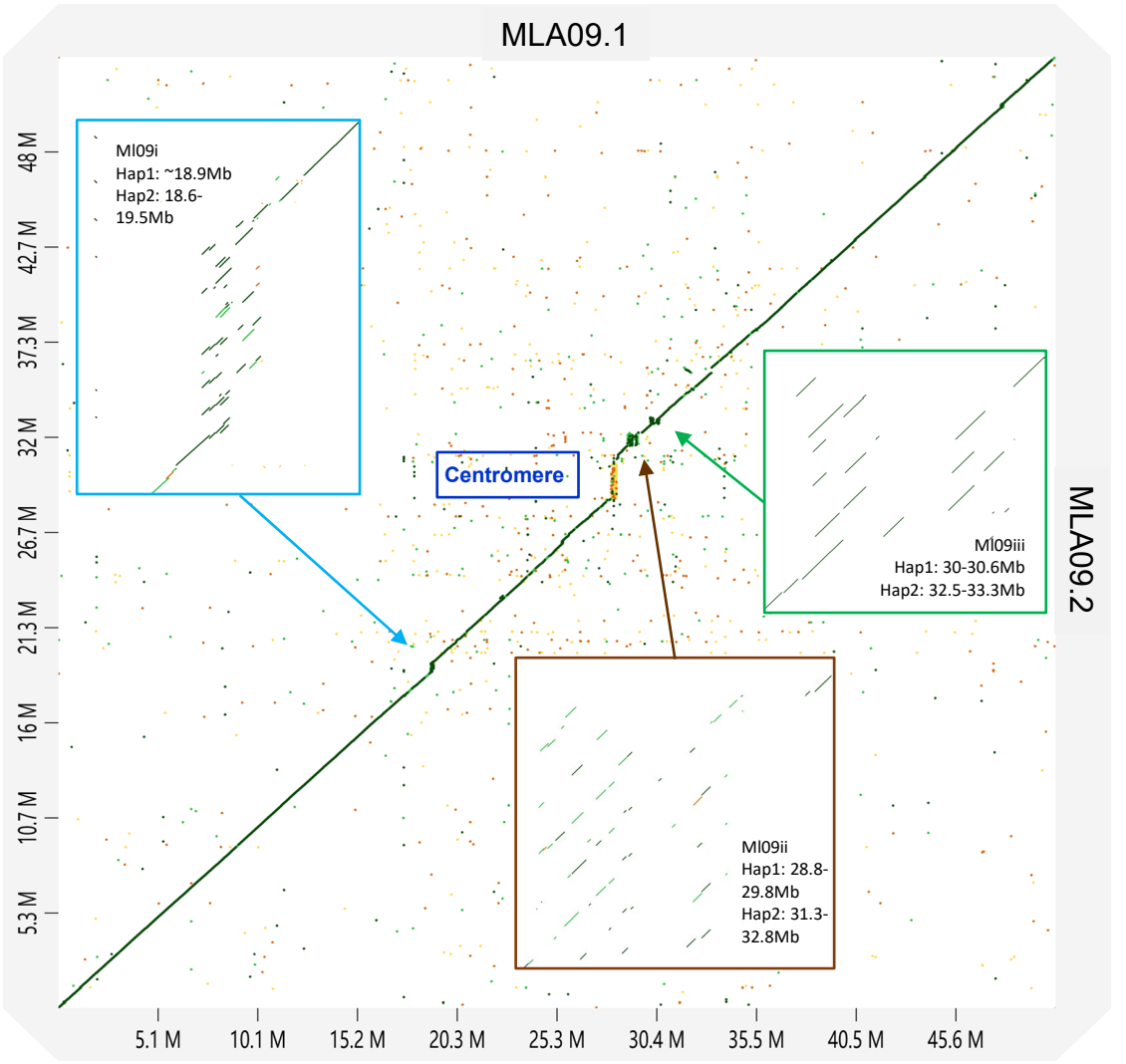


Liu et al 2026 Figure 3

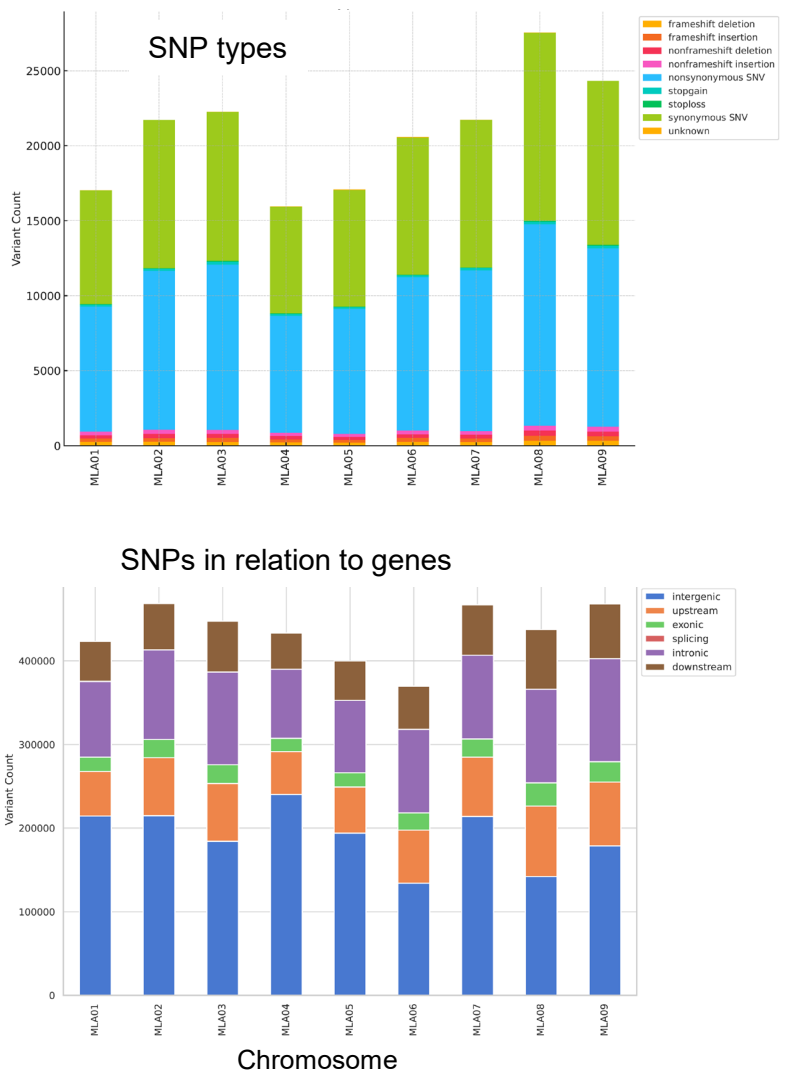


Liu et al 2026 Figure 4

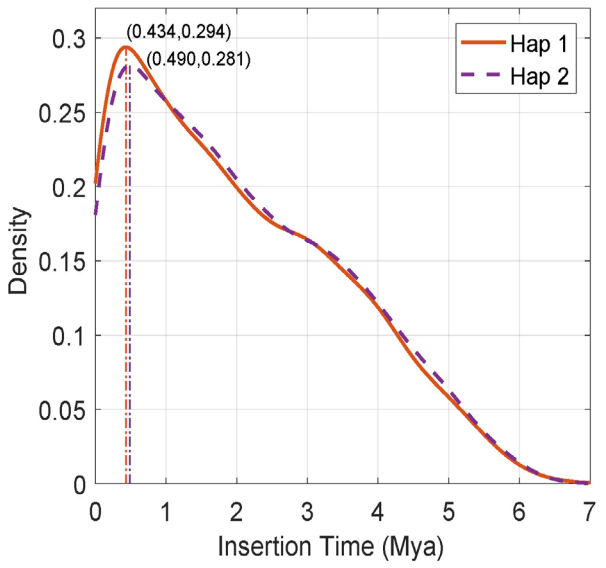
A Dot plot chromosome MLA09



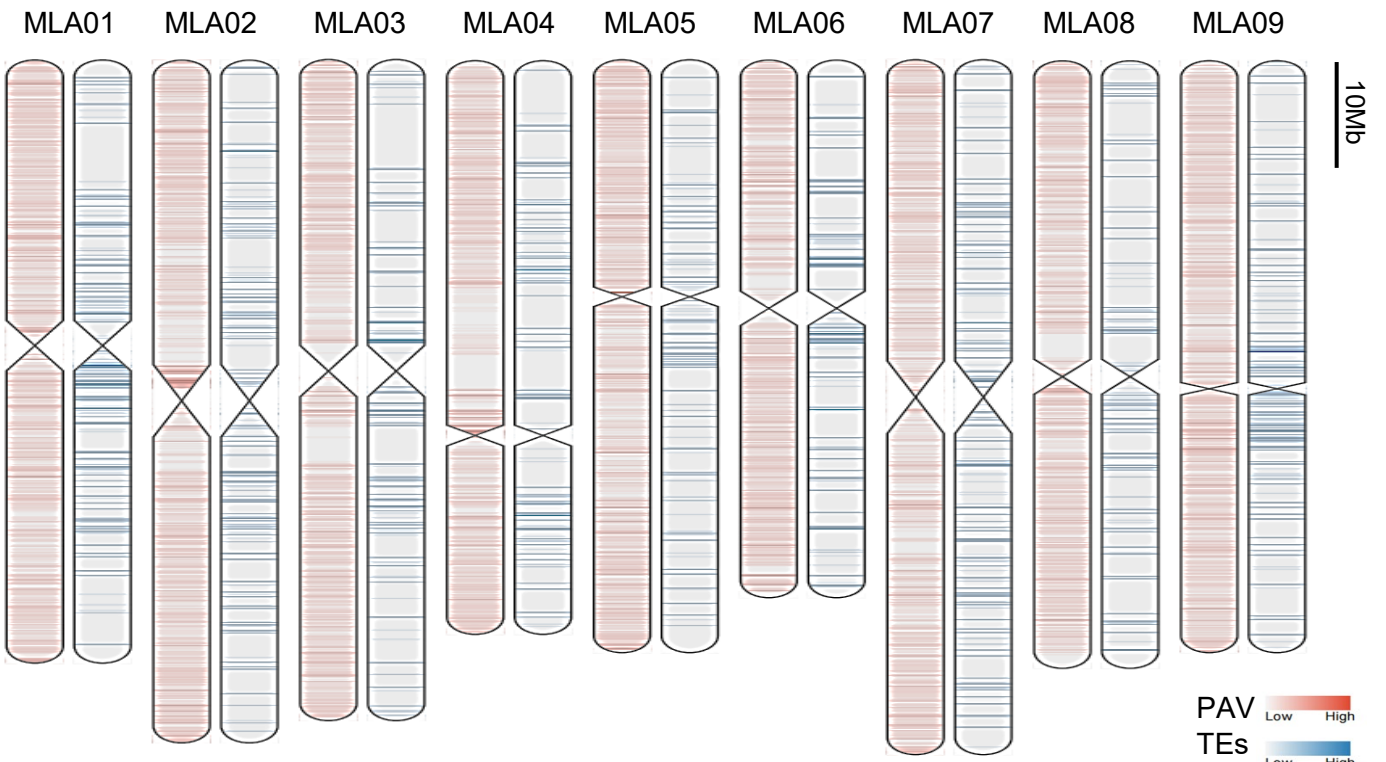
B SNP variation and distribution



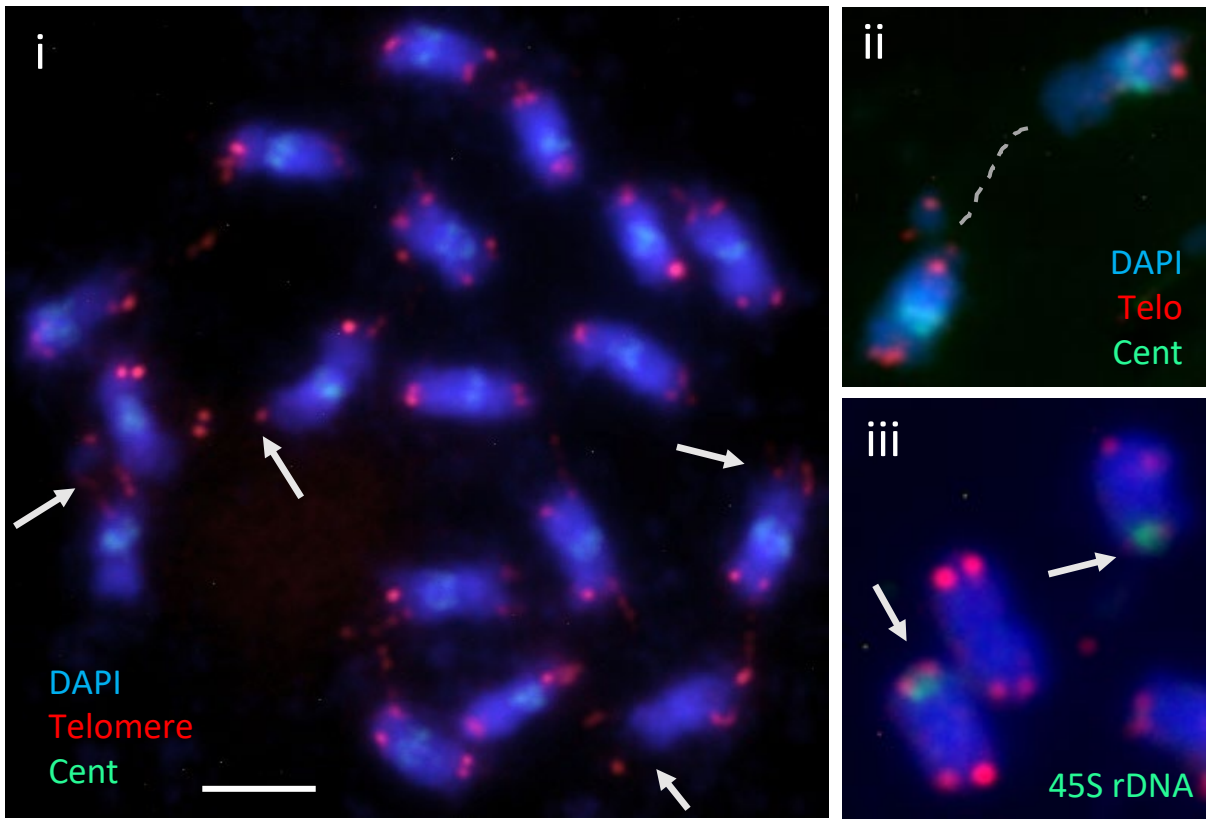
C Retroelement Insertion



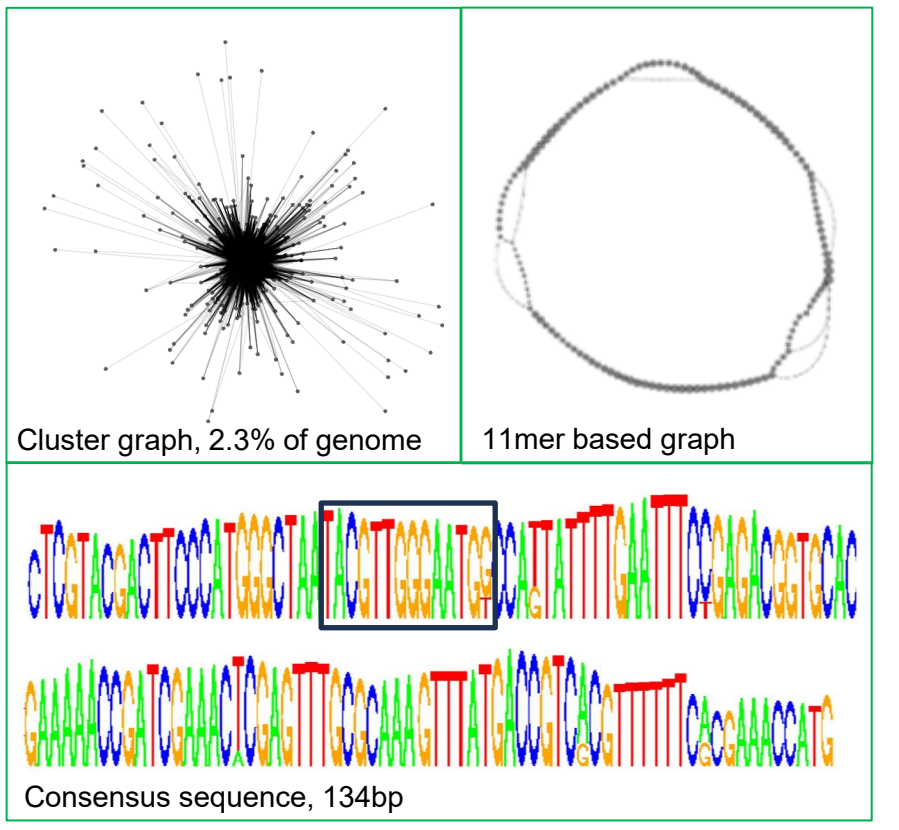
D Structural variant comparisons



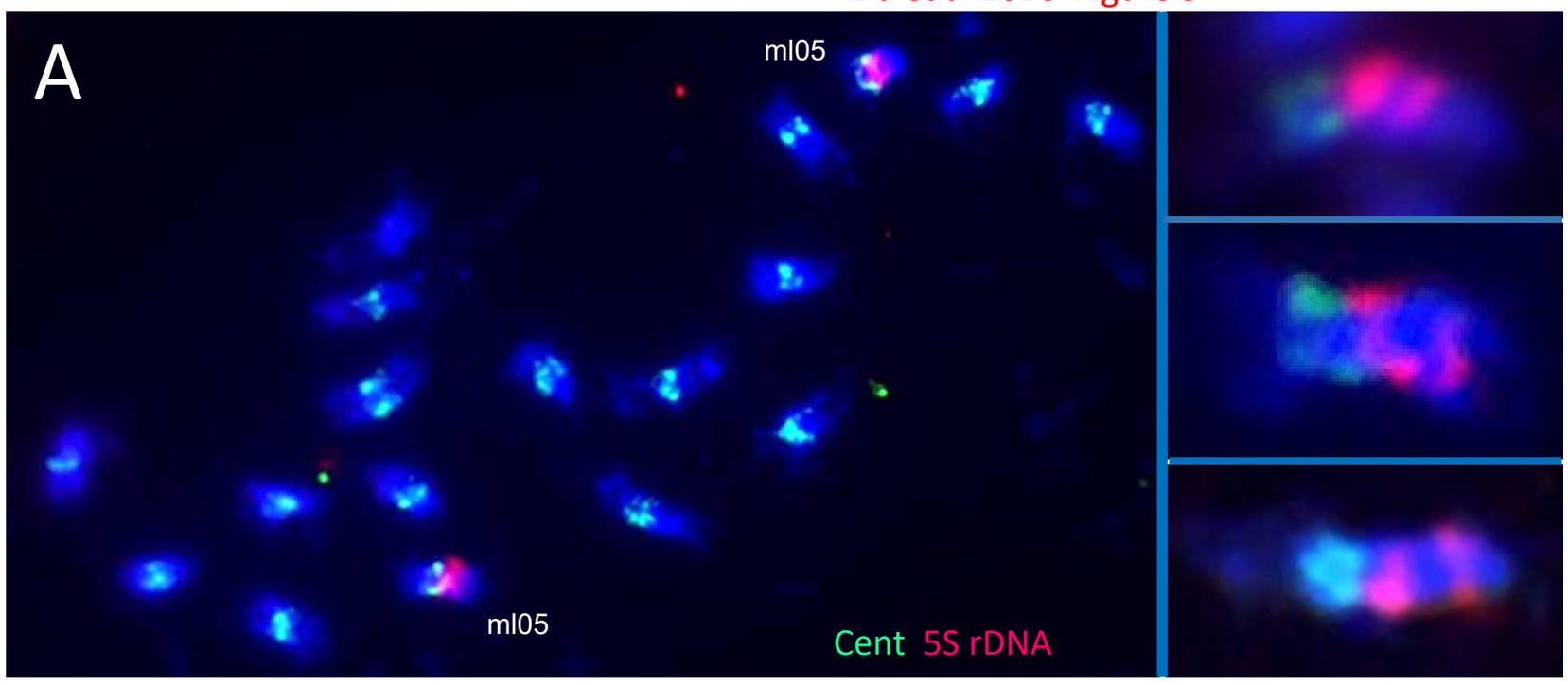
E Telomere FISH

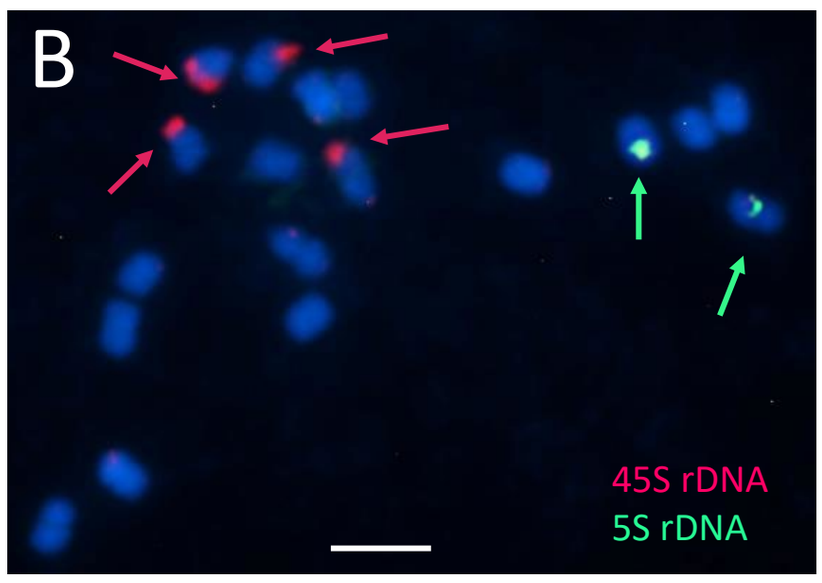


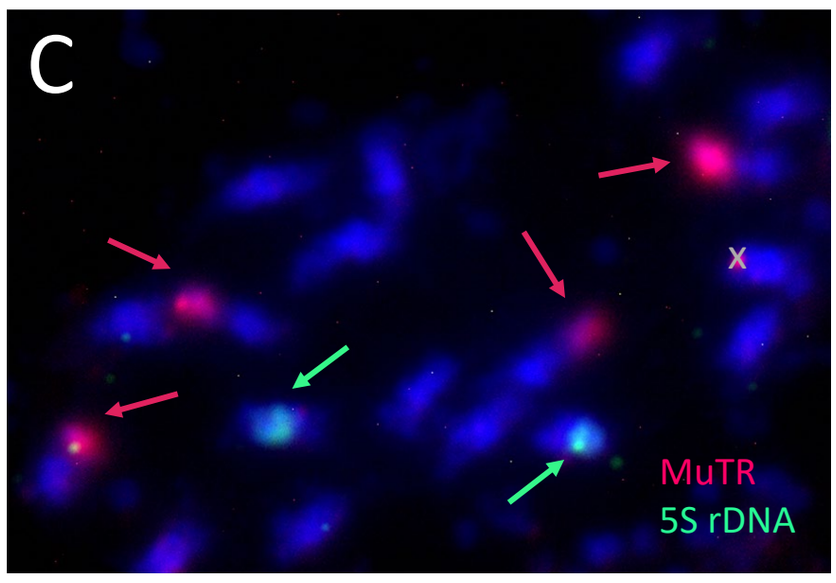
F RepeatExplorer Cluster 1, Mlcen

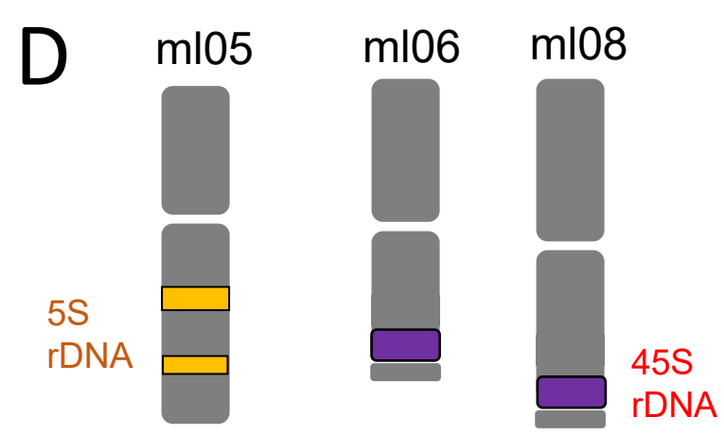


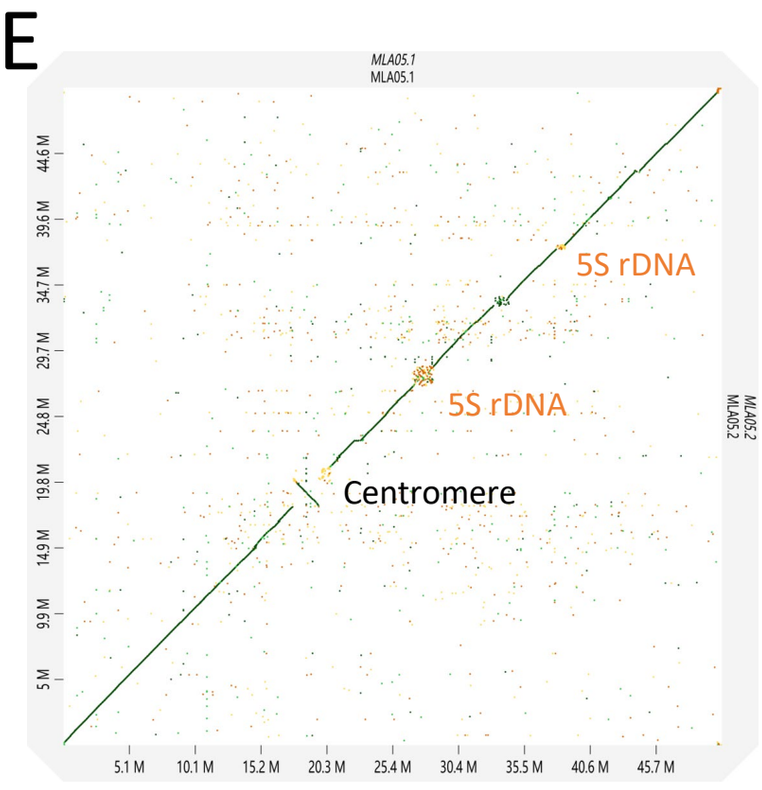
Liu et al 2026 Figure 5

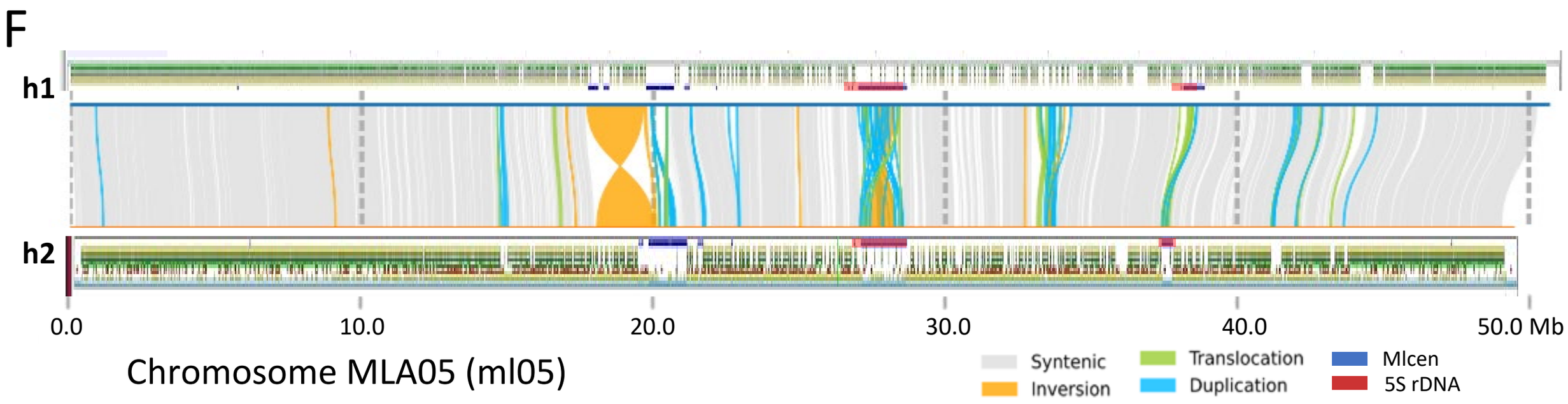


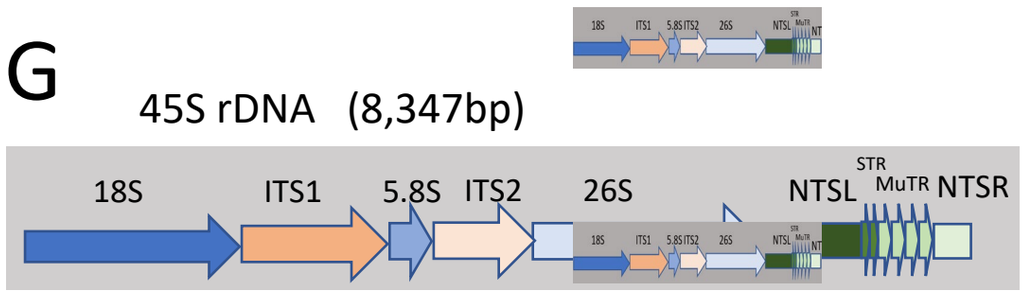


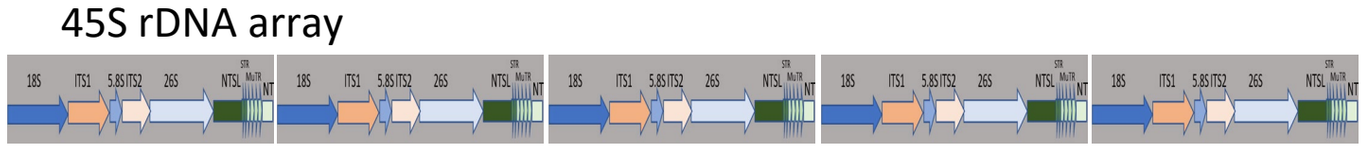






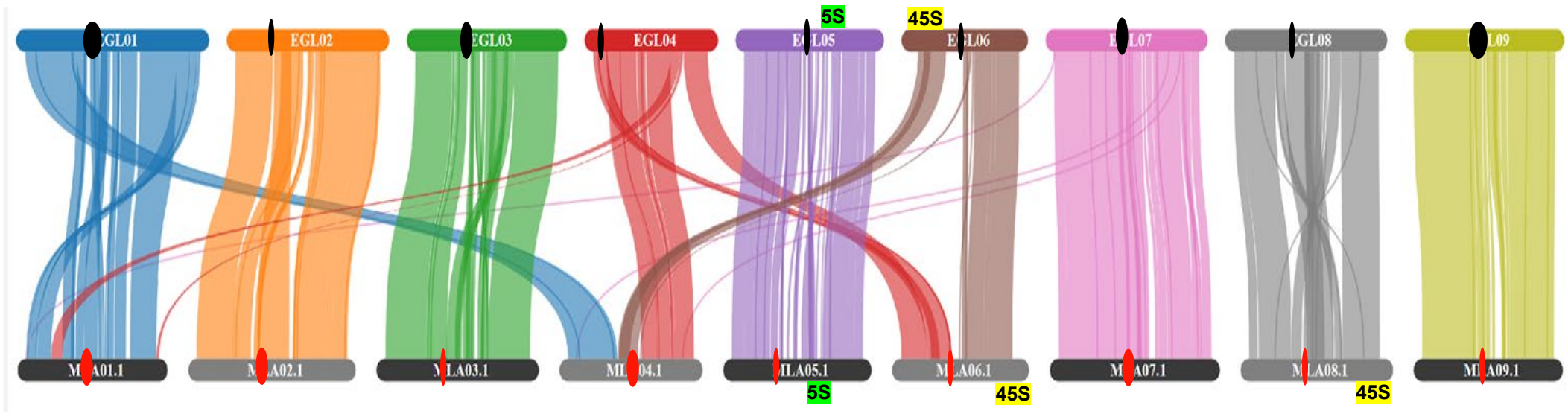




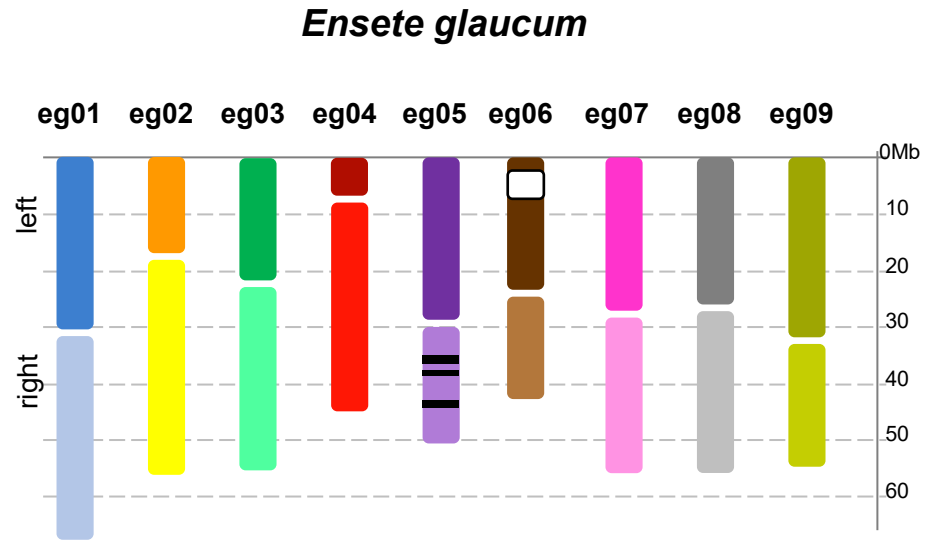


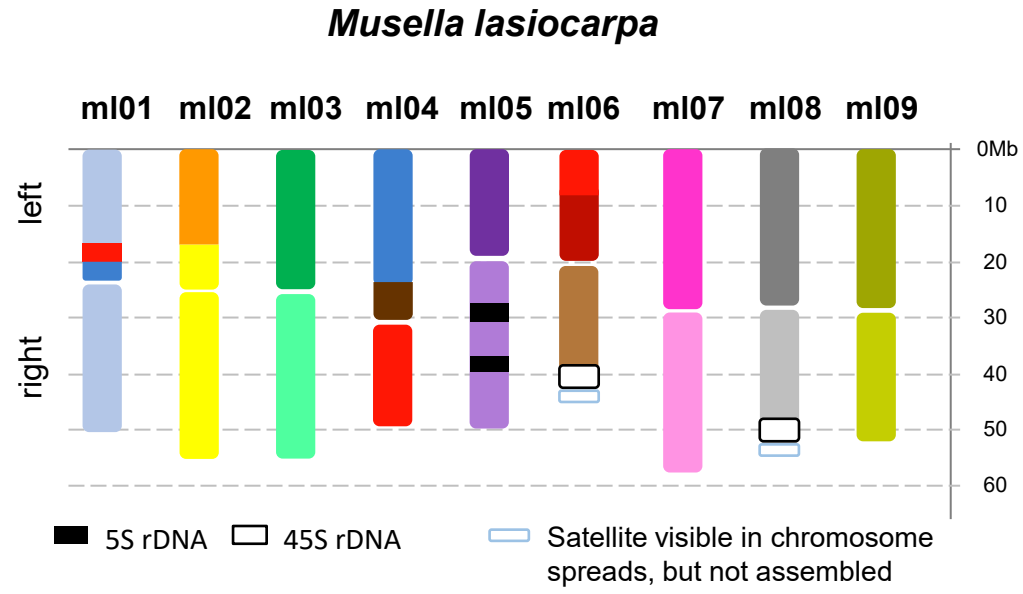
Liu et al 2026 Figure 6

A

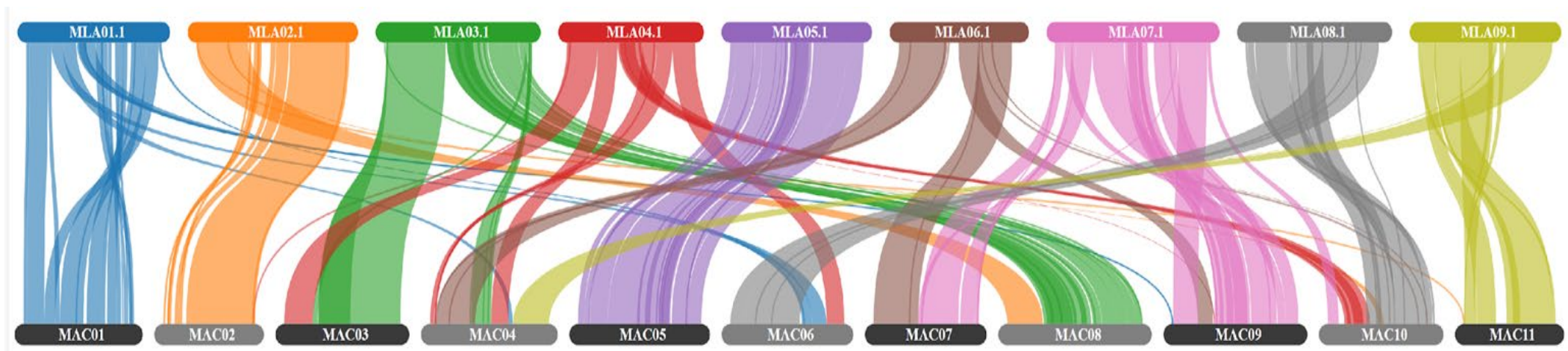


B

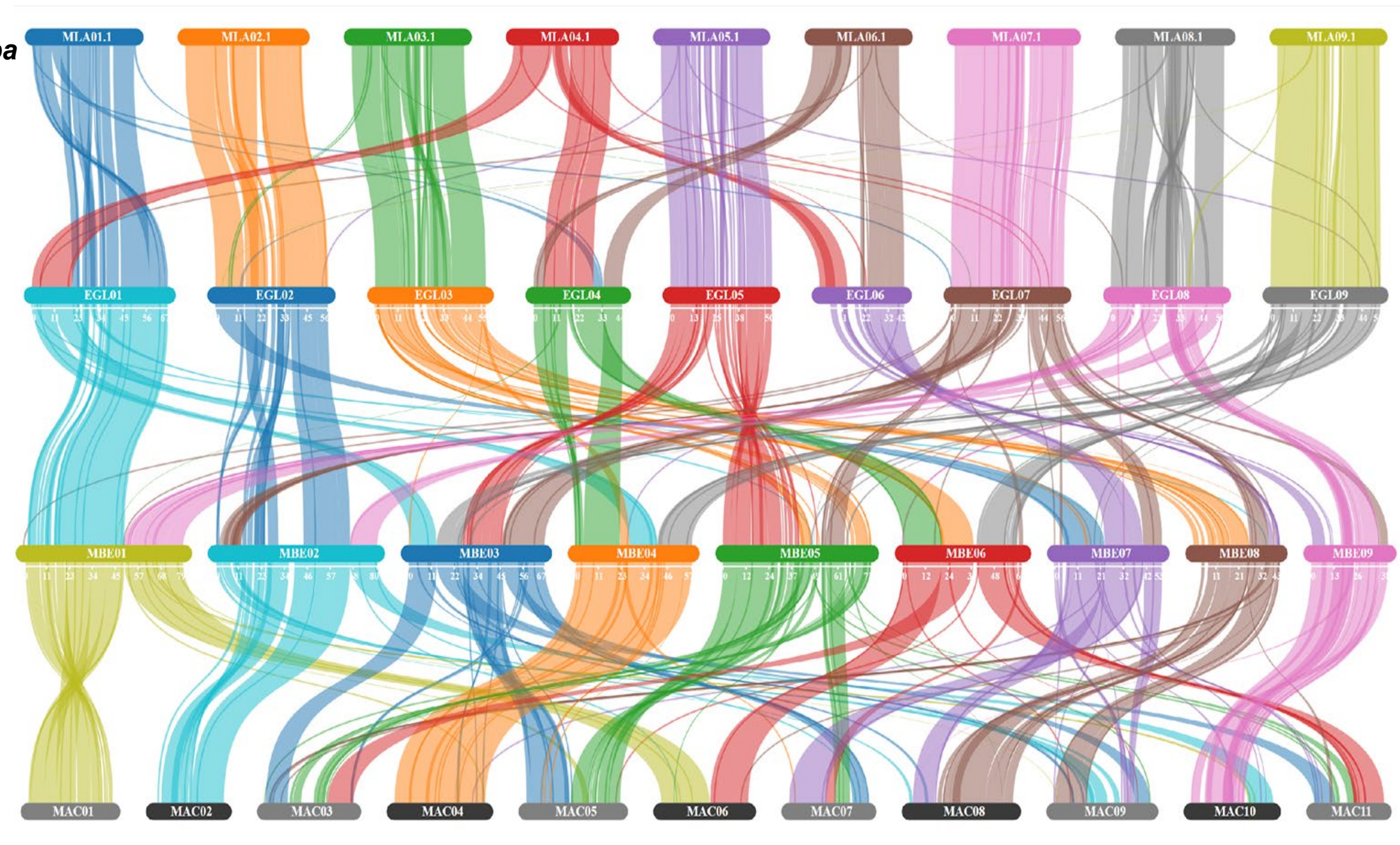




C



D



Supplementary materials 1-4
Data S1: pp1-4
Figures S1-S13: pp 5-11
Tables S1-S16: pp12-36

Haplotype resolved genome assembly of *Musella lasiocarpa* reveals a critical role for structural variation in chromosomal and genome evolution

Qing Liu, Dongli Cui, Yaqi Tian, Yehan Wang, Mathieu Rouard, John Seymour Heslop-Harrison, Trude Schwarzacher and Ziwei Wang

Extended Material and Methods

Plant material collection and DNA extraction

Fresh and young leaves of *Musella lasiocarpa* var. *lasiocarpa* and var. *Rubribracteata* were collected from the greenhouse of South China Botanical Garden, Chinese Academy of Sciences [originally collected from Nanhua city (E 113.36°, N 23.18°), Chuxiong Yi Autonomous Prefecture, Yunnan, China] from and subjected to genomic DNA extraction following procedures of the Qiagen Genomic Kit. Degradation of extracted DNA was assessed by 0.75% gel electrophoresis; DNA purity was evaluated using a NanoDrop One UV-Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA); DNA concentration was measured by Qubit 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). High-quality DNA was used to prepare short and long read whole-genome sequencing (WGS) libraries.

Library preparation and whole genome sequencing

For PacBio HiFi sequencing, the libraries were sequenced on one 8M-SMRT cell on a PacBio Sequel II system (Pacific Biosciences, USA) with Circular Consensus Sequencing (CCS) sequencing mode, by SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, USA) following the manufacturer's instructions.

For Oxford Nanopore sequencing, the libraries were prepared using the Ligation Sequencing Kit (Q-SQK-LSK109, Oxford Nanopore Technologies, Oxford, UK) according to the standard protocol. The purified library was loaded on Nanopore R9.4.1 Flow Cell and sequenced using a PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK).

For Hi-C library was generated based on the method (Belton et al., 2012) with some modifications. Briefly, young and fresh leaves were fixed in the nuclei isolation buffer with 2% formaldehyde. The cross-linked DNA was subsequently digested with 100 units of DpnII (New England Biolabs, USA). The digested fragments were biotinylated with biotin-14-dCTP and ligated using T4 DNA polymerase (New England BioLabs, MA, USA). The ligated DNA was enriched, sheared into 300–600 bp fragments, blunt-end repaired, and further processed. The final paired-end (2 × 150 bp) Hi-C library was sequenced on the Illumina NovaSeq 6000 platform (Illumina, San Diego, USA).

For Illumina sequencing, qualified DNA samples were interrupted to prepare a 400 bp inserted fragment library. Then, the Agilent 2100 Bioanalyzer system (Agilent Technologies, CA, USA) was used to verify the insertion size of the library, and qPCR was performed to ensure the effective quantitative centration of the library. The library was sequenced on the Illumina NovaSeq 6000 platform (Illumina, San Diego, USA). The resulting short reads were used for genome survey analysis.

RNA library of leaves, bracts, and root tissues of *Musella lasiocarpa* var. *lasiocarpa* was extracted using the TRNzol Universal RNA Extraction Kit (Tiangen, Beijing, China) for ONT full-length transcriptome sequencing. ONT full-length transcriptome data was used to assist gene structure annotation process.

After sequencing, fastp v.0.23.3 (Chen et al., 2018) was used to remove adapters and low-quality reads with default parameters for Illumina, ONT, Hi-C, and RNA reads.

Transcriptome sequencing for MLA varieties under cold treatment

Plants of *M. lasiocarpa* var. *lasiocarpa* and *M. lasiocarpa* var. *rubribracteata* (Originally collected from Puge city, E 102.68°, N 23.45°, Liangshan Yi Autonomous Prefecture, Sichuan, China) were grown in climatic chamber (Hengmin PRX-450D, Nanjing, China) conditions (16 h light/8 h dark and day/night temperature of 25°C with relative humidity 70%) in South China Botanical Garden for two weeks. After bracts were collected in liquid nitrogen, low temperature treatment was set at day/night temperature of 15°C for 48 h. Then MLA bracts under cold treatment were collected. Total RNA was extracted using Plant RNA Kit (OMEGA-R6827, Guangzhou, China). RNA sequencing libraries were prepared using MGIEasy RNA Library Prep Kit v.3.0 (Beijing Genomics Institution, Shenzhen, China) on UW Genetics DNBSEQ-T7 platform (BGI, Shenzhen, China) with three replicates set up for each treatment.

K-mer analysis and genome assembly

Genome size was estimated using Illumina reads via Jellyfish v.2.3.0 (Marçais and Kingsford, 2011) and GenomeScope v.1.0 (Vurture et al., 2017) with *k*-mer lengths of 17 and 21. A high-quality T2T genome was constructed by means of gap filling, telomere extension, and error correction by a hybrid assembly method on ONT and PacBio HiFi data. The working process was as follows. Primarily, for the PacBio raw subreads, consensus reads (HiFi reads) were generated using CCS v.8.0.1 (https://ccs.how/; parameters: -min-passes 3 -min-snr 2.5 -top-passes60; https://github.com/PacificBiosciences/ccs). For ONT raw data, Filtlong v.0.2.4 (https://github.com/rrwick/Filtlong) was used to filter fragments smaller than 10 kb. We used Porechop v.0.2.4 (Wick et al., 2018) to filter the joint sequence and reads with length ≥ 30 kb and mean Q score > 7 were retained for assembly by Filtlong v.0.2.4 (Wick et al., 2017). Subsequently, the ONT pass reads and the PacBio HiFi clean reads were used for mosaic assembly. For genome haplotype assembly, the HiFi + ONT + Hi-C assembly mode of Hifiasm v.0.19.5 (Cheng et al., 2021) was used to conduct preliminary MLA assembly. Assembly contamination, including mitochondrial, chloroplast, and bacterial genome, were filtered by Minimap2 (Li, 2016) then deleted. To further optimize the assembly result, Juicer v. 1.5 (Durand et al., 2016a) was used to process Hi-C raw data, 3D-DNA v.180419 (Dudchenko et al., 2017) were used to divide contigs into different clusters according to the closeness of association between different contigs in the valid data. Contigs within the chromosome group were oriented, and removing redundancies to obtain a chromosome-level genome sequence. Juicebox (Durand et al., 2016b) was used for visualization and manual corrections.

Lastly, we utilized D-GENIES (Cabanettes and Klopp, 2018) to compare MLA haplotype assemblies with the released *Ensete glaucum* genome (Wang et al., 2022) to determine MLA chromosome count and orientation.

Genomic evaluation and repeat annotation

The integrity of MLA haplotype assemblies was accessed by BUSCO v.5.5.0 (Simão et al., 2015) with the embryophyta_2024-01-08 database. The genome size of MLAh1 was estimated by Backmap.pl v.0.5 (Pfenninger et al., 2022). A Hi-C interaction heatmap was generated by HiCExplorer v.3 (Wolff et al., 2020). A *de novo* repeat sequence library was constructed by RepeatModeler (Flynn et al., 2020) and Extensive *de novo* TEAnnotator (EDTA) v.2.1.0 (Ou et al.,

2019). LTR (Long terminal repeat) sequences in the repeat sequence library were classified using TEsorter (Zhang et al., 2022). Subsequently, the repeat sequence proportions were estimated by RepeatMasker v.4.2.1 (Tarailo-Graovac and Chen, 2009). The insertion time of LTR retrotransposons was calculated as $T = K/(2r)$ with $r = 4.5 \times 10^{-9}$ (Ma and Bennetzen, 2004). Additionally, the graph-based clustering analysis was performed by RepeatExplorer2 (Novák et al., 2020) in order to classify repeats from Illumina raw reads. Insertion dates of full-length long terminal repeat retrotransposons were determined as described in Ma et al., 2004) with a substitution rate of 9×10^{-9} per site per year, which is twofold higher than the substitution rate determined for *Musa* genes (Lescot et al., 2008).

The TeloExplorer function of quarTeT (Lin et al., 2023) was used to search the telomere monomer sequences within haplotype genome assemblies. The centromeric 134 bp *Egcen* monomer of *Ensete glaucum* (Wang et al., 2022) was used to identify tandem repeat *MLcen* in MLA haplotypes. *MLcen* repeat was visualized by the graph-based clustering of Illumina reads on RepeatExplorer2 (Novák et al., 2020) and TAREAN pipeline (Novák et al., 2017). The position of *MLcen* arrays were taken as centromere locations in MLA haplotype assemblies to calculate the length of the first/left and second/right arm for nine chromosomes.

Gene structure and function annotation

Haplotype assemblies were masked using RepeatMasker v.4.1.2 (Tarailo-Graovac and Chen, 2009) based on the above repeat library. Gene prediction and functional annotation were performed on the masked genomes by an integration of *de novo* prediction, homology-based prediction, and transcriptome-assisted prediction. *De novo* prediction was performed by Augustus v.3.4.044 (Keilwagen et al., 2018), SNAP (Johnson et al., 2008), and GlimmerHMM v.3.0.4 (Delcher et al., 2007). For homology-based identification, protein sequences from *Ensete glaucum* (Wang et al., 2022), *Musa accuminata* (D'Hont et al., 2012), *M. balbisiana* (Wang et al., 2019), *M. schizocarpa* (Belser et al., 2018), *Phoenix dactylifera* (Hazzouri et al., 2019), *Oryza sativa* (Kawahara et al., 2013) were used for homologous prediction. The homologous protein sequences were aligned by Tblastn v.2.7.1 (Camacho et al., 2009) and GeMoMa (Keilwagen et al., 2018) to predict transcripts and coding regions. Then gene structures were predicted by the alignment results. For transcriptome prediction, Illumina clean reads and Oxford Nanopore cDNA clean reads were aligned by STAR v.2.7 (Dobin et al., 2013). StringTie v.2.7 (Kovaka et al., 2019) and PASA (Haas et al., 2003) were used to reconstruct the transcript. TransDecoder v.5.7.1 (https://github.com/TransDecoder/TransDecoder) was used to predict the gene models. Finally, the gene set predicted by the above methods were integrated into a more complete and non-redundant gene set by EVidenceModeler (Haas et al, 2008).

Function annotation of the protein coding genes was implemented by Diamond v.2.0.11.149 (Buchfink et al., 2011) searches against the following public databases: NCBI NR (Pruitt et al., 2005), InterPro (Paysan-Lafosse et al., 2023), GO (Ashburner et al., 2001), KOG (Tatusov et al., 2003), KEGG (Kanehisa et al., 2023), TrEMBL (Bairoch and Apweiler, 2000) and SwissProt (Bairoch and Apweiler, 2000). InterProScan v.5.68–100.0 (Jones et al., 2014) and Hmmscan v.3.3.2 (Finn et al., 2011) were used to obtain conserved sequences, motifs, and domains of proteins by comparison with databases InterPro and Pfam (Finn et al., 2014). MLA genome features were visualized by Circos (Krzywinski et al., 2009).

Gene family expansion, contraction, and GO enrichment analysis

Orthology identification within and between six Musaceae species was inferred by OrthoFinder v.2.4.0 (Emms and Kelly, 2019) with default parameters. R package UpSetR (Conway et al., 2017) was used for the orthogroup distribution visualization (Fig. 2A).

Single copy genes identification and the phylogenetic relationship reconstruction were conducted by OrthoFinder for MLA and other 13 monocotyledon species (Table S8C). Divergence time was estimated using a OrthoFinder model implemented by MCMCtree in PAML v.4.9j (Yang, 2007). The crown ages of the *Musa*, Musaceae and Zingiberales were calibrated to 50.3 (49.9–50.7), 55.7 (45.1–69.1) Mya (Million years ago). The 64.1 (50.6–90.1) Mya were used to constrain the split of *Musella-Ensete*/*Musa* by TIMETREE 5 (Kumar et al., 2022). Gene family expansion or contraction were detected by CAFE v.5.1 (Mendes et al., 2021). GO and KEGG enrichment analyses were performed for unique and expansion gene families using the enricher function in the R package clusterProfiler v.4.8.2 (Wu et al., 2021).

Genome synteny, duplications, and whole-genome comparisons

The WGD events in MLA_self, MLA-EGL, MLA_MAC and MLA_ZOF (*Zingiber officinale*) were inferred by WGDI v.0.6.5 pipeline (Sun et al., 2022). The Gaussian fitting curve parameters of each Ks peak were used to produce the Ks distribution map (Fig. 2B) with the '-kf' option.

To detect structural variations between MLAh1-MLAh2, Mummer v.4.0.0 (Marçais et al., 2018) was used to align MLA haplotypes using EGL genome as a reference (Wang et al., 2022). A minimum alignment length of 100 bp was further filtered by the delta-filter (delta-filter -i 95 -l 10 000). Secondly, SyRI v.1.6 (Goel et al., 2019) was employed to identify collinear regions, structural rearrangements (inversions, translocations and duplications) and local variations (SNPs, InDels). Lastly, structural variations were visualized by Plotsr (Goel and Schneeberger, 2022) (Fig. 3). Protein-coding genes were processed to identify reciprocal best hits with BLASTP (e-value 1e-05) followed by MCScanX (Wang et al., 2012) with default parameters.

The results were imported into SynVisio (Bandi and Gutwin, 2020) for syntenic block visualization of MLA-EGL, MLA-MAC and MLA1-EGL-MAC-MBE (Fig. 6). The nucleotide-based comparative analysis was aligned by minimap2 (Li, 2016) and visualzied results by D-Genies (Cabanettes and Klopp, 2018) to generate dot plots of MLA1-MLA2 and MLA-EGL (Figs S7A and S8). ONT and PacBio HiFi long reads mapped to MLA haplotype assemblies by minimap2 (Li, 2016), the reads covered the inversion breakpoints was identified by igv (Robinson et al., 2011), and the dotplots of long reads mapping MLA haplotypes were captured from Geneious Primer v.2022.0.2 (https://www.geneious.com/).

Differential expressed gene (DEG) analysis of MLA

RNA of both normal and cold treatment samples was sequenced for *M. lasiocarpa* var. *lasiocarpa* (yellow bract, Y_normal), var. *rubribracteata* (red bract, R_normal), Y_cold and R_cold. RNA reads mapped to MLA assemblies by STAR (Dobin et al., 2013). The reads for gene exons were counted by featureCounts v.2.0.6 (Liao et al., 2014). DEGs were identified using the DESeq2 function including DESeq() and results()in the R package DESeq2 v.1.54.0 (Love et al., 2014), which required biological replicates for robust differential expression analysis. The DEGs between Y_normal and Y_cold as well as between R_normal and R_cold were selected with the criterion of absolute normalized $\log_2$-transformed fold-change > 2 and P-value < 0.001. KEGG enrichment analysis of the upregulated DEGs was performed using the enricher function in the R package clusterProfiler (Wu et al., 2021).

Identification of cold adaptation-related transcription factors in MLA

Identification of plant transcription factors was performed by PlantTFDB (Jin et al., 2009) with default setting (without exclusive selection of "best hit in Arabidopsis thaliana"). HY5, COP1, and PIFs genes were determined based on KEGG annotation results (Table S13). To investigate whether the identified genes were highly expressed in cold treated bracts, a heatmap displaying gene expression levels was generated with TBtools (Chen et al., 2023) with Row Scale normalization applied and ChiPlot (https://www.chiplot.online/). The chromosomal location of CIB genes was illustrated with TBtools. Potential TFBSs in the promoter sequences of CBF genes were predicted using the online program JASPAR (https://jaspar.elixir.no/) with relative profile score threshold > 90% (Sandelin et al., 2004). Upstream 2,000 bp sequences of the cold-related genes were extracted for the analysis.

Fluorescent *in situ* hybridization (FISH)

MLA plants (purchased commercially) were grown at the University of Leicester (UK) greenhouse at 16 h light/8 h dark and day/night temperature of 28°C/22°C. Young root tips (2 cm) were incubated in aerated 2 mM 8-hydroxyquinoline for 3 h at growing temperature followed by 4–6 h at 4°C, fixed in 96% ethanol:glacial acetic acid (3:1) and stored at –20°C. Chromosome preparations were obtained as described by Schwarzacher and Heslop-Harrison (2000) with minor modifications. Briefly, roots were digested with a cellulose and pectinase mixture containing (32U/ml cellulose, Sigma-Aldrich C1184; 20U/ml 'Onozuka' RS cellulose; 35 U/ml pectinase from *Aspergillus niger*, Sigma-Aldrich P4716; 20U/ml Viscozyme, Sigma-Alderich V2010 in 10 mM citric acid/sodium citrate buffer) for 3–5 h at 37°C, transferred to buffer for 12–30 h at 4°C. Root tip meristems were filtered after squash in a drop of 60% acetic acid with a cover slip and stored at –20°C until FISH.

FISH followed Schwarzacher and Heslop-Harrison (2000) with minor modifications. Oligonucleotide FISH probes are listed in Table S17. Additionally, random priming (Invitrogen) with digoxigenin dUTP and biotin dUTP (Roche) was used to label the PCR amplified insert of clone pTa794 (containing 5S rDNA repeat unit of *Triticum aestivum*; Gerlach and Dyer 1980) and the MuTR repeat sequence amplified from gDNA. Hybridization probes were prepared using 100–500 ng labelled probes in 30% (v/v) formamide, 2 × SSC (Sodium chloride sodium citrate), 10% (w/v) dextran sulphate, 0.025 µg/µl salmon sperm DNA, 1.25 mM EDTA (Ethylenediamine-tetra acetic acid) and 0.13% SDS (Sodium dodecyl sulphate). Probe matrix of 40 µl was applied to each slide, which was covered by a plastic cover slip and together denatured at 72°C for 8 mins and 37°C overnight.

Directly labeled oligonucleotides did not need a detection step. Biotin or Digoxinen labeled probes were incubated with Fluorescein-conjugated anti-digoxigenin (Roche) and Alexa 594 conjugated streptavidin (Molecular Probes). Chromosomes were counterstained with 4µg/ml DAPI (4',6-diamidino-2-phenylindole) and mounted in Citiflour AF1 (17970-25, EMS Acquisition Corp., PA, USA). Images were captured using Nikon Eclipse 80i microscope equipped with filter blocks UV-2A (EX330/380, DM400, BA410), FITC (EX465-495, DM505, BA512-558), TRITC (EX527-553, DM585, BA577-633) and a DS-QiMc monochrome camera controlled by NIS-Elements v2.34 (Nikon, Tokyo, Japan). Captured images were assigned colour and enhanced with Adobe Photoshop v.25.6.0 (Adobe, USA), using functions that treat all pixels equally.

References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A. P., Dolinski, K., Dwight, S.S., Eppig J.T., et al., 2001. Gene Ontology, tool for the unification of biology. Nat. Genet. 25, 25–29.

Bairoch, A., Apweiler, R., 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28, 45–48.

Bandi, V., Gutwin, C., 2020. Interactive exploration of genomic conservation. In, Proceedings of the 46th Graphics Interface Conference on Proceedings of Graphics Interface 2020 (GI'20). Waterloo, Canada, Canadian Human-Computer Communications Society.

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.C., Falentin, C., Genete, M., Berrabah, W., Chèvre, A. M., Delourme, R., et al., 2018. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat. Plants 4, 879–887.

Belton, J.M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., Dekker, J., 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58, 268–276.

Buchfink, B., Reuter, K., Drost, H.G., 2011. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods 18, 366–368.

Cabanettes, F., Klopp, C., 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6, e4958.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421.

Chen, C.J., Wu, Y., Li, J.W., Wang, X., Zeng, Z.H., Xu, J., Liu Y.L, Feng J.T., Che, H., Ye, Y.H., et al., 2023. Tbtools-II: a "one for all, all for one" bioinformatics platform for biological big-data mining. Mol. Plant 16, 1733–1742.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.

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

Conway, J.R., Lex, A., Gehlenborg, N., 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics, 33, 2938–2940.

Delcher, A.L., Bratke, K.A., Powers, E.C., Salzberg, S.L., 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics, 23, 673–679.

D'Hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., et al., 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488, 213.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

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

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., Aiden, E.L., 2016a. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98.

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

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al., 2014. Pfam: the protein families database. Nucleic Acids Res. 42, D222–D230.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37.

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. 117, 9451–9457.

Gerlach, W.L., Dyer, T.A., 1980. Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. Nucleic Acids Res. 8, 4851–4865.

Goel, M., Schneeberger, K., 2022. plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics 38, 2922–2926.

Goel, M., Sun, H., Jiao, W.B., Schneeberger, K., 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20, 1–13.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith Jr, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al., 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31, 5654–5666.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C, R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 9, R7d.

Hazzouri, K.M., Gros-Balthazard, M., Flowers, J.M., Copetti, D., Lemansour, A., Lebrun, M. Masmoudi, K., Ferrand, S., Dhar, M.I., Fresquez, Z.A., et al., 2019. Genome-wide association mapping of date palm fruit traits. Nat. Commun., 10, 4680.

Jin, J.P., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J.C. et al., 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research, 45, D1040–D1045.

Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., Donnell, C.J.O., de Bakker, P.I.W., 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24, 2938–2939.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W.Z., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka G., et al., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240.

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., Ishiguro-Watanabe, M., 2023. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 51, D587–D592.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J.Z., Zhou, S.G., et al., 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice 6, 4.

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., Grau, J., 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinf. 19, 189.

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., Pertea, M., 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 20, 278.

Kumar, S., Suleski, M., Craig, C.M., Kasprowicz, A.E., Sanderford, M., Li, M., Stecher, G., Hedges, S.B., 2022. TimeTree 5: an expanded resource for species divergence times. Mol. Biol. Evol. 39, msac174.

Krzywinski, M., Schein, J., Birol, I., Gascoyne, R., Horsman, D., Jones, S.J., Jones, S.J., Marra, M. A., 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645.

Lescot, M., Piffanelli, P., Ciampi, A.Y., Ruiz, M., Blanc, G., Leebens-Mack, J., da Silva, F.R., Santos, C.M.R., D'Hont, A., Garsmeur, O., et al., 2008. Insights into the *Musa* genome: syntenic relationships to rice and between *Musa* species. BMC Genom. 9, 58.

Li, H., 2016. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. Bioinformatics 32, 2103–2110.

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.

Lin, Y.Z., Ye, C., Li, X.Z., Chen, Q.Y., Wu, Y., Zhang, F., Pan, R., Zhang, S.J., Chen, S.X., Wang, X., et al., 2023. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic. Res. 10, uhad127.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Ma, J., Bennetzen, J.L., 2004. Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. U. S. A. 101, 12404–12410.

Ma, J., Devos, K.M., Bennetzen, J. L., 2004. Analyses of LTR-retrotransposon structures

reveal recent and rapid genomic DNA loss in rice. Genome Res. 14, 860–869.

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

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

Mendes, F.K., Vanderpool, D., Fulton, B., Hahn, M.W., 2021. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics 36, 5516–5518

Novák, P., Neumann, P., Macas, J., 2020. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nat. Protoc. 15, 3745–3776.

Novák, P., Robledillo, L.A., Koblížková, A., Vrbová, I., Neumann, P., Macas, J., 2017. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res. 45, e111.

Ou, S.J., Su, W.J., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al., 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20, 275.

Paysan-Lafosse, T., Blum, T., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al., 2023. InterPro in 2022. Nucleic Acids Res. 51, D418–D427.

Pfenninger, M., Schönenbeck, P., Schell, T., 2022. ModEst: accurate estimation of genome size from next-generation sequencing data. Mol. Ecol. Resour. 22, 1454–1464.

Pruitt, K.D., Tatusova, T., Maglott, D.R., 2005. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts, and proteins. Nucleic Acids Res. 33, D501–D504.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P, 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.

Sandelin A, Alkema W, Engström P. Wasserman, W.W., Lenhard, B., 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. 32, D91–D94.

Schwarzacher, T., Heslop-Harrison, P., 2000. Practical *in situ* hybridization. BIOS Scientific Publishers Ltd.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Sun, P.C., Jiao, B.B., Yang, Y.Z., Shan, L., Li, T., Li, X.N., Xi, Z.X., Wang, X.Y., Liu, J.Q., 2022. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Mol. Plant 15, 1841–1851.

Tarailo-Graovac, M., Chen, N., 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinform. 4, 1–14.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al., 2003. The COG database: an updated version includes eukaryotes. BMC Bioinf. 4, 41.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M. C., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204.

Wang, Z., Miao, H.X., Liu, J.H., Xu, B. Y., Yao, X.M., Xu, C.Y., Zhao, S.C., Fang, X.D., Jia, C.H., Wang, J.Y., et al., 2019. *Musa balbisiana* genome reveals subgenome evolution and functional divergence. Nat. Plants 5, 810–821.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J.P., Wang, X.Y., Lee, T.H., Jin, H.Z., Marler, B., Guo, H., et al., 2012. MCScanX, a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40, e49.

Wang, Z., Rouard, M., Biswas, M.K., Droc, G., Cui, D.L., Roux, N., Baurens, F. C., Ge, X.J., Schwarzacher, T., Heslop-Harrison, J. S., et al., 2022. A chromosome-level reference genome of *Ensete glaucum* gives insight into diversity and chromosomal and repetitive sequence evolution in the Musaceae. GigaScience 11, giac027.

Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb. Genom. 3, e000.132.

Wick, R.R., 2018. Porechop: an adapter trimmer for Oxford Nanopore reads. https://github.com/rrwick/Porechop Google Scholar.

Wolff, J., Rabbani, L., Gilsbach, R., Richard, G., Manke, T., Backofen, R., Grüning, B.A., 2020. Galaxy HiC Explorer3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res. 48, W177–W184.

Wu, T, Hu E, Xu S. Chen, M.J., Guo, P.F., Dai, Z.H., Fengm T.Z., Zhou, L., Tang, W.L., Zhan, L., et al., 2021. clusterProfiler4.0: a universal enrichment tool for interpreting omics data. Innovation, 2, 100–141.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Zhang, R.G., Li, G.Y., Wang, X.L., Dainat, J., Wang, Z.X., Ou, S.J., Ma, Y.P., 2022. TEsorter, an accurate and fast method to classify LTR-retrotransposons in plant genomes. Hortic. Res. 9, uhac017.

Supplementary Data

Supplementary Figures

Haplotype-resolved genome assemblies of *Musella lasiocarpa* (Musaceae) reveals a critical role for structural variation in chromosome and genome evolution

Qing Liu, Dongli Cui, Yaqi Tian, Yehan Wang, Mathieu Rouard, John Seymour (Pat) Heslop-Harrison, Trude Schwarzacher and Ziwei Wang

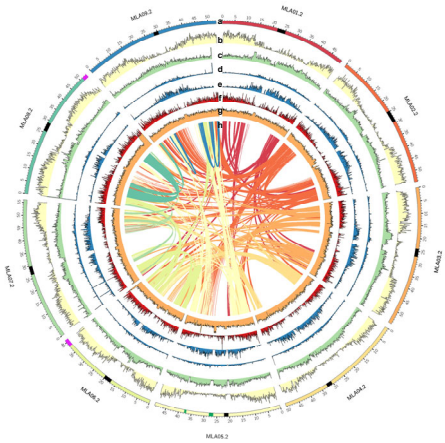
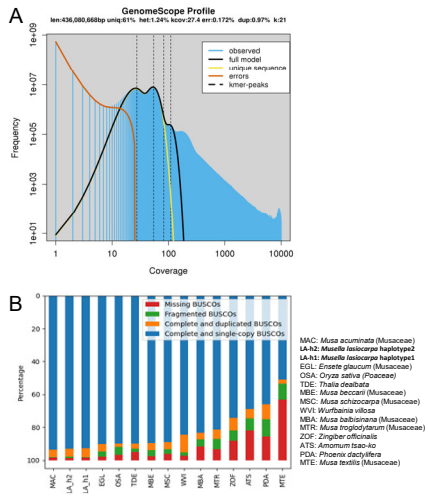


Fig. S1. *Musella lasiocarpa* haplotype 2 (MLAh2) genomic features of **a** Nine pseudo-molecules (MLA01 to MLA09) with green, pink, and black boxes denoting. 5S and (MLA05.2), 45S (MLA06.2 and MLA08.2) rDNA and centromere positions, scale in 5.0 Mb; **b** Gene density; **c** Repeat density; **d** *Copia* LTR retroelement density; **e** *Gypsy* LTR retroelement density; **f** DNA transposon density; **g** Simple sequence repeat (microsatellite) density; **h** Syntenic genomic blocks, connected by curved lines. b–h: 10,000 bp bins.



MAC: *Musa acuminata* (Musaceae)
LA-h2: *Musella lasiocarpa* haplotype2
LA-h1: *Musella lasiocarpa* haplotype1
EGL: *Ensete glaucum* (Musaceae)
OSA: *Oryza sativa (Poaceae)*
TDE: *Thalia dealbata*
MBE: *Musa beccarii* (Musaceae)
MSC: *Musa schizocarpa* (Musaceae)
WVI: *Wurfbainia villosa*
MBA: *Musa balbisiana* (Musaceae)
MTR: *Musa troglodytarum* (Musaceae)
ZOF: *Zingiber officinalis*
ATS: *Amomum tsao-ko*
PDA: *Phoenix dactylifera*
MTE: *Musa textilis* (Musaceae)

Fig. S2. Genome size and BUSCO comparison of MLA and other 13 monocotyledonous plants.
A: Genome size and heterozygosity estimated from *k*-mer frequency distribution of *Musella lasiocarpa* Illumina data using GenomeScope. *K*-mer (*k* = 21) analysis result showed that the estimated MLA genome size (length) is 436,080,668 bp (The middle peak); the heterozygosity value (heterozygosity) is 1.24% (the left peak); repetitive sequences accounted for 52.81% (the right peak) in the *k*-mer distribution diagram.
B: Comparison of genome assembly completeness by BUSCO scores of 14 monocotyledonous plant genome assemblies. In MLAh1 and MLAh2, 35,387 and 35,237 genes were annotated, respectively. The integrity of gene sets assessed by BUSCO was 97.96% and 97.83%, indicating high integrity of annotated genes (Table S3B). Similar in quality to the annotations of *Musa accumina* (MAC with 98.8% BUSCO; Belser et al., 2021). MAC: *Musa accuminata* (D'Hont et al., 2012); LA-h2: *Musella lasiocarpa* haplotype 2; LA_h1: *Musella lasiocarpa* haplotype 1; EGL: *Ensete glaucum* (Wang et al., 2022); OSA: *Oryza sativa* (Kawahara et al., 2013); TDE: *Thalia dealbata* (Tang et al., 2023); MBE: *Musa beccarii* (Wang et al., 2019); MSC: *Musa schizocarpa* (Besler et al., 2018); WVI: *Wurfbainia villosa* (Yang et al., 2022); MBA: *Musa balbisiana* (Wang et al., 2019); MTR: *Musa troglodytarum* (Zhou et al., 2024); ZOF: *Zingiber officinalis* (Li HL et al., 2021); ATS: *Amomum tsao-ko* (Li et al., 2022); PDA: *Phoenix dactylifera* (Hazzouri et al., 2019); MTE: *Musa textilis* (Galvez et al., 2021) (Table S3C).

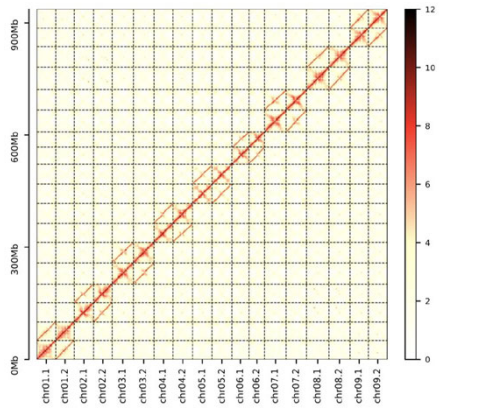


Fig. S3. Hi-C interaction heatmap of MLA haplotypes. MLAh1 and MLAh2 were divided by 300 kb partitions, and the number of Hi-C read pairs between each partition was converted into the intensity signal of the interaction between partitions. The interaction signal strength between adjacent sequences (diagonal positions) was high, while the interaction signal strength between non-adjacent sequences (non-diagonal positions) was relatively low, indicating high genome assembly accuracy and no algorithm-induced structural variation detected.

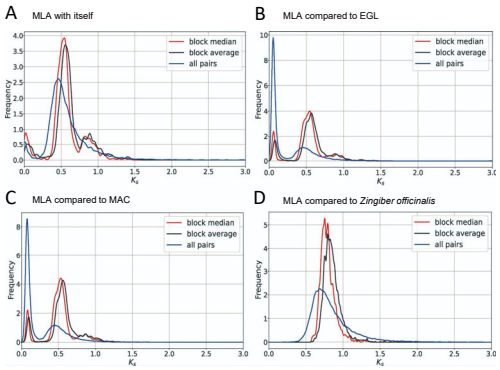


Fig. S4. Synonymous substitutions per site (Ks) plot of orthologous gene pairs between MLAh1 as compared to itself (paralogs) and other species. **A** MLA-MLA comparison. **B** MLA - *Ensete glaucum* (EGL) comparison. **C** MLA-*Musa acuminata* (MAC) comparison. **D** MLA compared to *Zingiber officinalis* (ZOF). The average *Ks* distribution showed that the peak values of MLA-MLA, MLA-EGL and MLA-ZOF were all around 0.53, which indicated the alpha/beta WGD event around 65 Myr ago. The peak value of MLA-ZOF was around 0.75, which the indicated gamma WGD event around 109–123 Mya in Zingiberales, consistent with the WGDs occurred in monocotyledons (D'Hont *et al.*, 2012). For overlay of block median curves see Figure 2B.

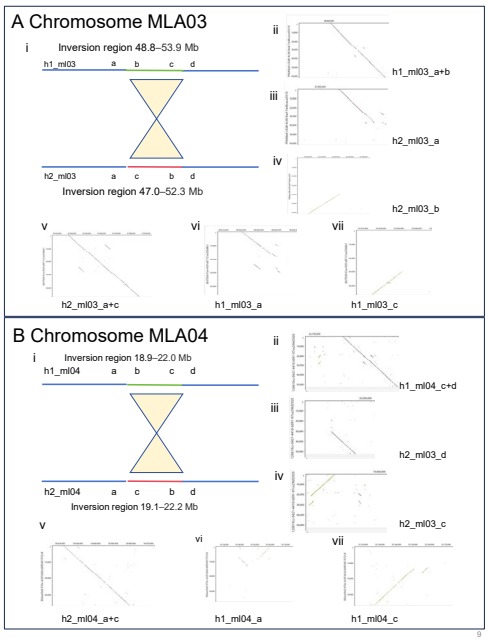
Fig. S5. GO-terms of MLA gene families.
A GO-terms of the unique gene families of MLA. **Left**: The top 20 GO enriched biological processes for the unique gene families in MLA (Table S9). GO enrichment results showed that the MLA unique gene families were related to amylopectin metabolic process and nitrate transmembrane transport. **Right**: The top 20 GO enriched biological processes for expanded gene families in MLA (Table S10). GO enrichment results showed that the MLA expansion gene families were significantly enriched in 169 biological processes, among which the most enriched biological processes were related to photosynthesis/light response and response to herbicide.
B GO-terms of up- and down-regulated DEGs of MLA varieties under cold treatment. **Left**: 38 genes among 88 up-regulated DEGs (Table S12A). **Right**: 410 genes among 1,470 down-regulated DEGs (Table S12B). According to Gene Ontology (GO) enrichment analysis, only 22 up-regulated proteins enriched in biological process "root morphogenesis", "root development", "root system development" and "plant organ morphogenesis (Table S12C)", and 1,315 down-regulated proteins were classified into three groups, i.e., calcium ion binding and oxidoreductase activity in Molecular function, and energy derivation by oxidation of organic compounds, obsolete oxidation-reduction process, response to toxic substance and defense response in Biological process (Table S12D), which are related to the response to cold tolerance.

Fig. S6. Localization of transcription factors (TFs, the position indicated by the TF name) identified from up- and down-regulated DEGs of two MLA varieties under cold treatment and structural variations (colored bars) between MLA haplotypes (Top: MLAh1, bottom: MLAh2 in the rectangle).

Fig. S7. Chromosomal synteny between MLAh1 and MLAh2.

See separate file MLASupplFigS7
Chromosomal synteny between MLA haplotype1 and haplotype 2 assemblies. Dot-plots between haplotype1 (Haplotype 1, x-axis) and haplotype 2 (Haplotype 2, y-axis) assemblies shown for each chromosome separately. Underneath each dot-plot, the PlotSR synteny comparison shown bars representing genes as well as retroelements. Centromere repeats MIcen and rDNAs are identified.

A: Chromosome MLA01.
B: Chromosome MLA02.
C: Chromosome MLA03.
D: Chromosome MLA04.
E: Chromosome MLA05.
F: Chromosome MLA06.
G: Chromosome MLA07 left arm.
H: Chromosome MLA07 right arm.
I: Chromosome MLA08.
J: Chromosome MLA09.



Supplementary Figure S9 continued

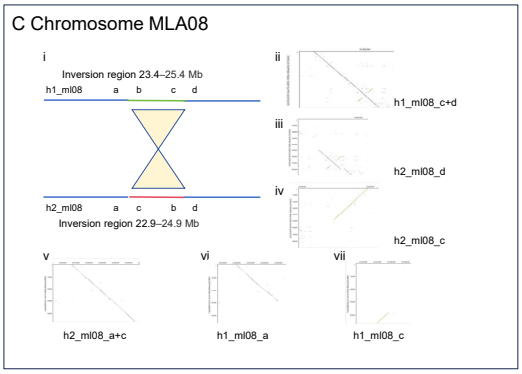


Figure S8. The ONT long-reads evidence for breakpoints of the longest inversions on chromosomes 3,4 and 8 between MLA haplotypes (Fig. 3) by igv.js (Robinson et al., 2023) and the dotplots visualized by Geneious Primer v.2022.0.2.

A The inversion structure of chromosome MLA03 (i) region 48.8–53.9 Mb and MLA03.2 region 47.0–52.1 Mb. (ii) Dotplot between the read named "f99dfda4-c5d8-4c38-8adf-7a48caca5510" (composed of regions **a**+**b**) and MLA03.1. (iii) Dotplot between this read region **a** and MLA03.2. (iv) Dotplot between this read region **b** and MLA03.2 in inversion direction. (v) Dotplot between the read named "d0cf7b58-01ee-455f-a0f7-7c7ae330d8e1" (composed of regions **a**+**c**) and MLA03.2. (vi) Dotplot between this read region **a** and MLA03.1. (vii) Dotplot between this read region **c** and MLA03.1 in inversion direction.

B The inversion structure of chromosome MLA04 (i) region 18.9–22.0 Mb and MLA04.2 region 19.1–22.2 Mb. (ii) Dotplot between the read named "120619cc-0421-4410-b091-97ee29bf2920" (composed of regions **c**+**d**) and MLA04.1. (iii) Dotplot between this read region **d** and MLA04.2. (iv) Dotplot between this read region **c** and MLA04.2 in inversion direction. (v) Dotplot between the read named "86bee9e0-97bc-4c0f-b60d-b085401972c8" (composed of regions **a**+**c**) and MLA04.2. (vi) Dotplot between this read region **a** and MLA04.1. (vii) Dotplot between this read region **c** and MLA04.1 in inversion direction.

C The inversion structure of chromosome MLA08 (i) region 23.4–25.4 Mb and MLAh2 region 22.9–24.9 Mb. (ii) Dotplot between the read named "6329c03f-8a70-4f03-909b-98a69c325462" (composed of regions **c**+**d**) and MLA08.1. (iii) Dotplot between this read region **d** and MLA08.2. (iv) Dotplot between this read region **c** and MLA08.2 in inversion direction. (v) Dotplot between the read named "1da3fd0f-02e7-4e14-9489-804ac02298f1" (composed of regions **a**+**c**) and MLA08.2. (vi) Dotplot between this read region **a** and MLA08.1. (vii) Dotplot between this read region **c** and MLA08.1 in inversion direction. Green lines indicated inversion regions of MLA03.1 48.8–53.9 Mb, MLA04.1 18.9–22.0 Mb, and MLA08.1 23.4–25.4 Mb on chromosome; Red lines indicated inversion regions of MLA03.2 47.0–52.1 Mb, MLA04.2 19.1–22.2 Mb, and MLA08.2 22.9–24.9 Mb on chromosome; Two neighbour regions (**a** and **b**, **a** and **c**, **c** and **d**, or **b** and **d**) in the breakpoints are composed of one ONT long read; **a** and **d** located on the non-inversion regions. **b** and **c** located on the inversion region. One pair of blue border triangles denoted reverse complementary sequences of the inversion structural variation.

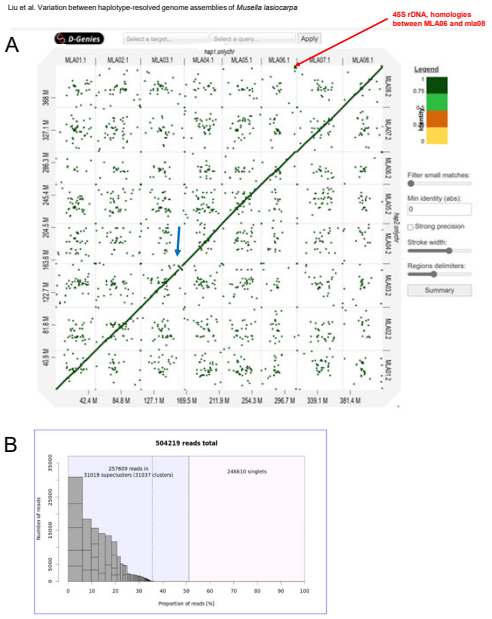


Figure S9. Abundance and copy number of repetitive DNAs in MLA genome.

A MLA haplotype sequence pairwise comparison by D-Genies (Cabanettes and Klopp, 2018). Homologies are mainly near the centromeres rich in repeats. The 45S rDNA array homologies on MLA06.1 and MLA08.2 are indicated by a red arrow. The longest inversions on MLA03 is indicated by a blue arrow.

B Distribution of graph-based clusters from hierarchical agglomeration using RepeatExplorer of Illumina reads. Bar chart (with reads ≥ 0.01%) of the most abundant individual graph-based read clusters (gray bars). Y-axis showing the read number in clusters while X-axis showing their cumulative DNA content.

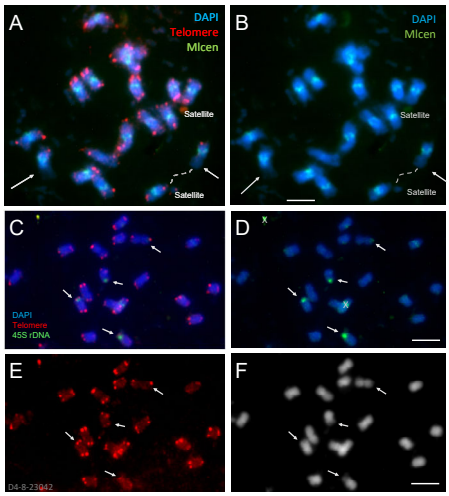
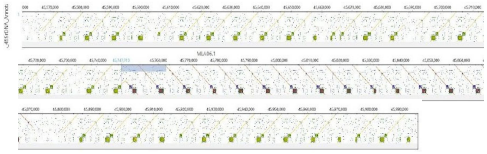
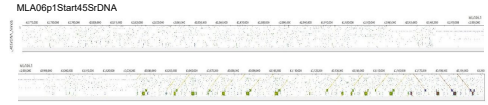


Fig. S10. FISH distribution of telomere, MIcen and 45S rDNA sequences on MLA metaphase chromosomes. **A** MLA metaphase chromosomes (DAPI stained, blue) with the telomere sequences (red signal) and the MIcen loci (green signals). **B** Same metaphase as in A, showing only blue DAPI staining and green MIcen loci. White arrows point to the nucleolus organizing regions (NORs formed by 45S rDNA arrays) and the dotted white line connects to the detached cytological satellite (shown in Fig 2Eii). **C**–**F** Distribution of 45S rDNA (green signal, arrows in all images) and telomere sequences (red signal) on DAPI stained (blue) MLA metaphase chromosomes. **C** merged image of A-F, . **D** The 45S rDNA loci labeled by green signals on blue DAPI chromosomes. **E** The telomere sequences labeled by red signal; this enhanced image also shows the outline of the chromosomes, and telomere signal on all chromosome ends including the 45SrDNA loci (arrows). **F** DAPI image of chromosomes in gray-scale, showing reduced staining at 45S rDNA sites (arrows). Bars = 5 μm.

MLA06p1End45SrDNAEndChrom06



MLA06p1Start45SrDNA



MLA06p1SelfEndwith45SrDNA

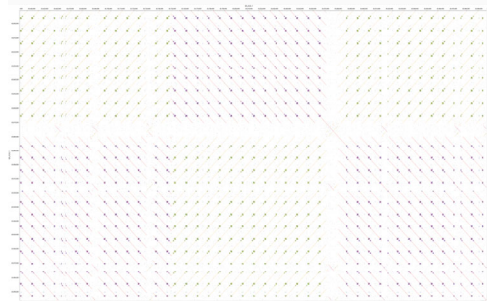


Fig. S11. Dotplots of the 45S rDNA array on chromosome MLA06. Top rows show start and end of the array, bottom shows a self dotplot indicating the long range organization.

MLA08p1rDNAstartArray50Mb

Enlarged

MLA08p1End45SrDNAEndChrom8

MLA08p1endwithy45SrDNA

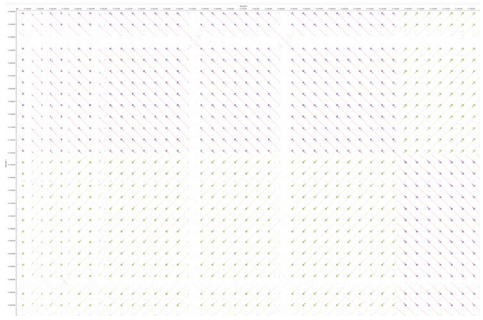


Fig. S12. Dotplots of the 45S rDNA array on chromosome MLA08. Top rows show start and end of the array, bottom shows a self dotplot indicating the long range organization.

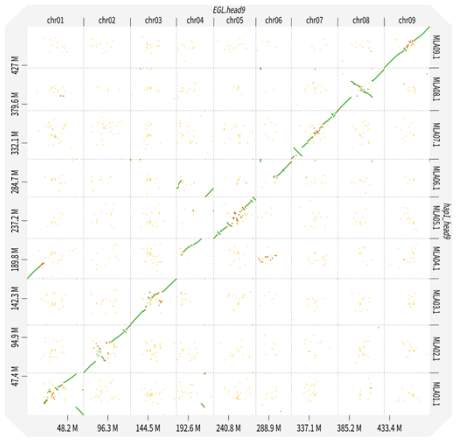


Fig. S13. Dotplot between *Musella lasiocarpa* and *Ensete glaucum* genome assemblies. The major translocation of chromosome 01, 04 and 08 are visible. At centromeres the synteny is disrupted except for the shared Egcen and Mlcen repeats.

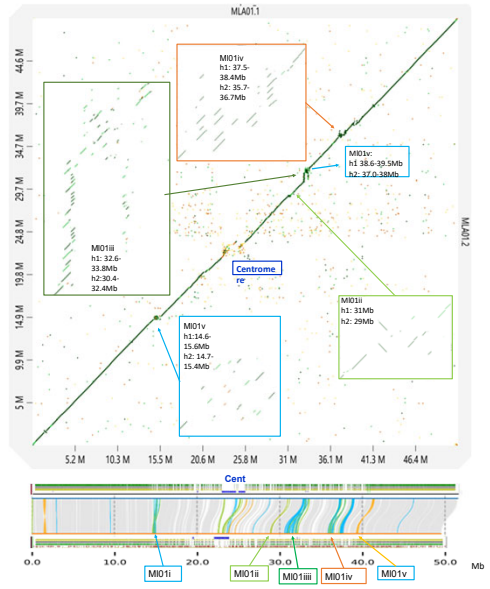
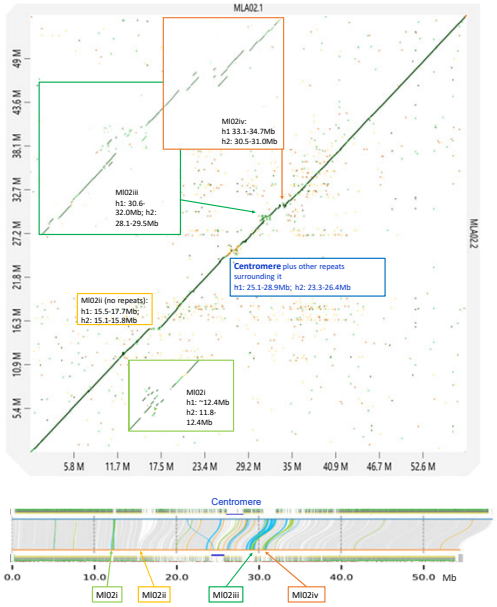


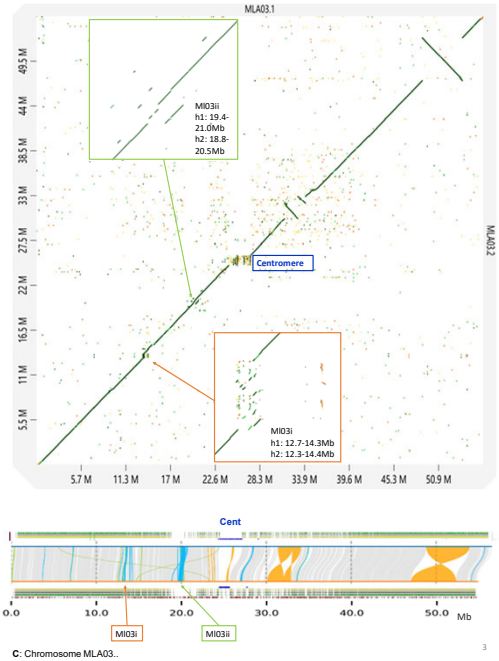
Fig. S7. Chromosomal synteny between MLA haplotype1 and haplotype 2 assemblies. **A**: Chromosome MLA01. Dot-plots between haplotype1 (Haplotype 1, x-axis) and haplotype 2 (Haplotype 2, y-axis) assemblies shown for each chromosome separately. Underneath each dot-plot, the PlotSR synteny comparison shown bars representing genes (yellow, green and grey) as well as retroelements (red). Centromere repeats MIcen are indicated in blue blocks and rDNA are labelled in red. The corresponding colored squares are enlargements of the indicated regions in the upper dot-plot.

Supplementary Fig. S7 continued.



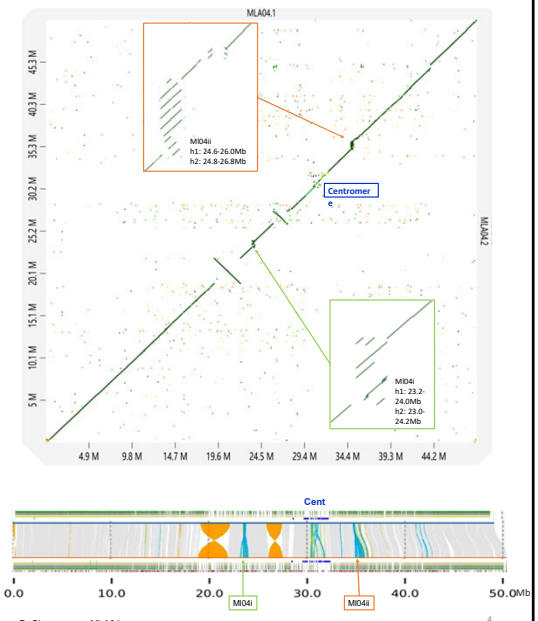
B: Chromosome MLA02.

Supplementary Fig. S7 continued.



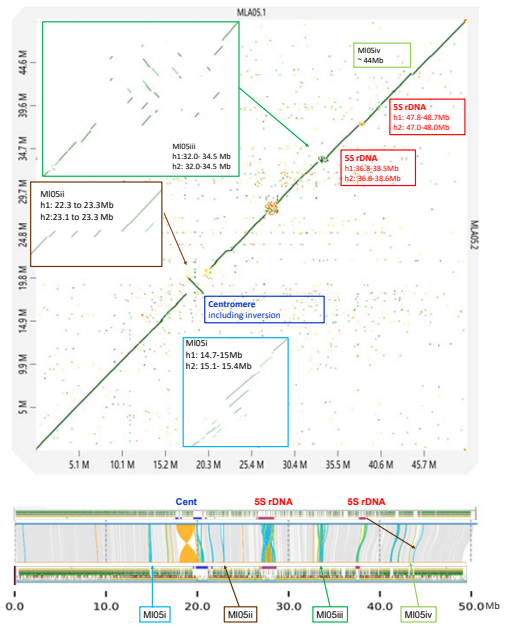
C: Chromosome MLA03..

Supplementary Fig. S7 continued.



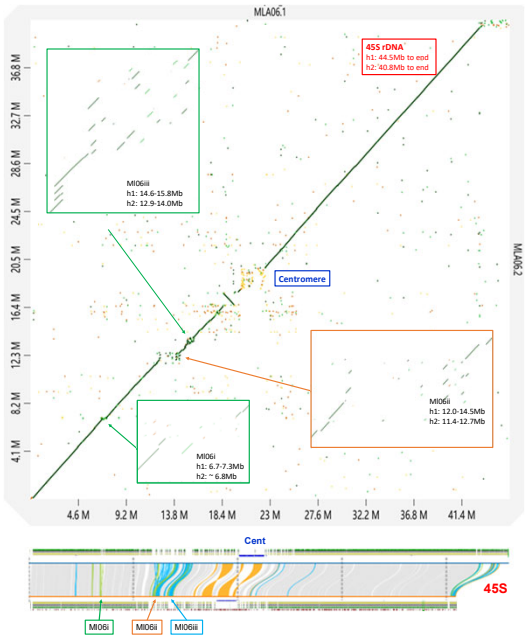
D: Chromosome MLA04.

Supplementary Fig. S7 continued.



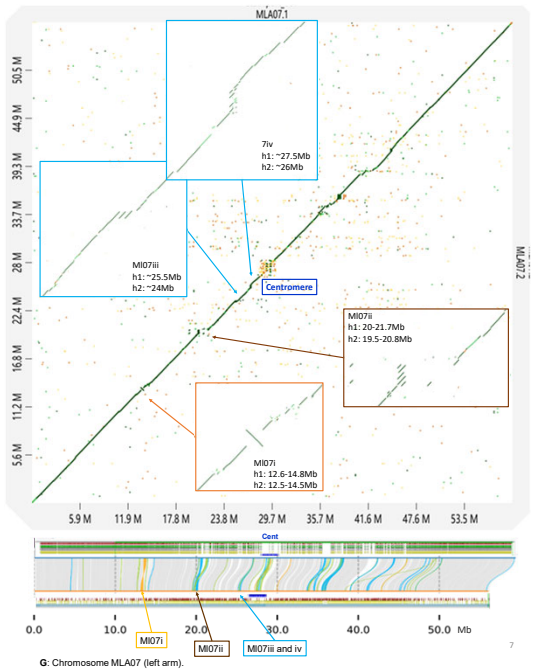
E: Chromosome MLA05.

Supplementary Fig. S7 continued.



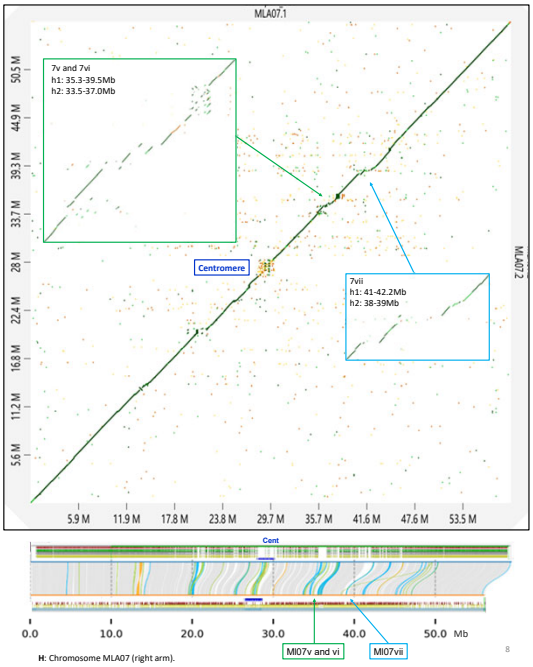
F: Chromosome MLA06.

Supplementary Fig. S7 continued.



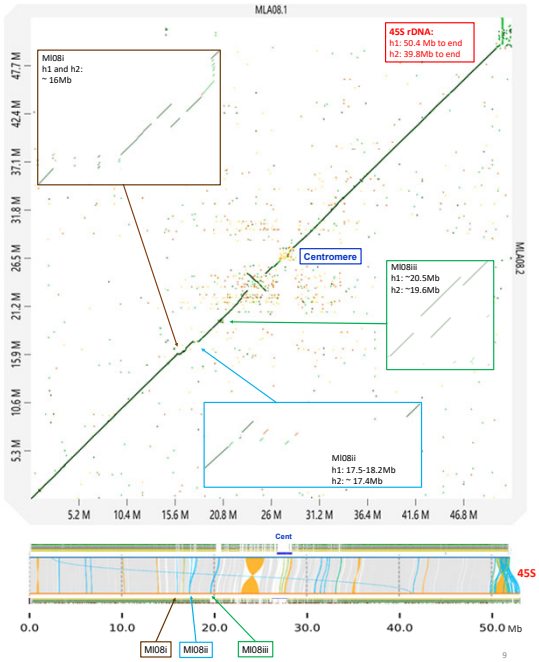
G: Chromosome MLA07 (left arm).

Supplementary Fig. S7 continued.



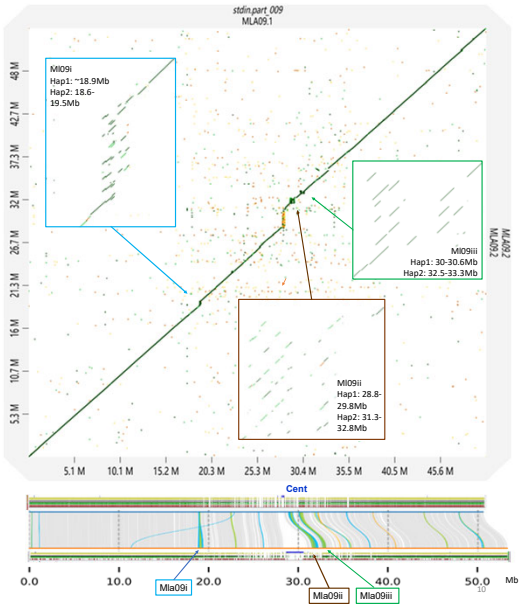
H: Chromosome MLA07 (right arm).

Supplementary Fig. S7 continued.



I: Chromosome MLA08.

Supplementary Fig. S7 continued.



J: Chromosome MLA09.

Supplementary Table S1. Statistics of genome sequencing and assembly of *Musella lasiocarpa* (MLA) haplotypes.

Table S1A. Summary of genome sequencing and transcription data of MLA haplotypes.

Type	Sequencing method	Raw data (clean reads)	Sequencing depth*)	Average read length	Number of reads	Read N50
Whole genome assembly	PacBio HiFi sequencing	31.46 GB	62.53	15,950bp	1,972,689	16,314 bp
	Nanopore (ONT)	103.11 Gb	204.75	24,252.82 bp	7,090,253	36,010bp
	Hi-C sequencing	57.04Gb	113.27	150bp x2	190,128,946 × 2	150bp
Genome survey and repeat analysis	Illumina sequencing	33.00Gb	65.53	150bp x2	110,009,047 × 2	150bp
Transcriptome	Illumina (leaf)	3.69 Gb	-	150bp × 2	12,291,105 × 2	150bp
	Nanopore (Mixed tissue)	11.51 Gb	-	150bp × 2	9,553,530	150bp
	Illumina (bract yellow)	9.99Gb × 3	-	150bp × 2	66,701,732 66,685,686 66,648,608	150bp
	Illumina (bract red)	9.99Gb × 3	-	150bp × 2	66,710,840 66,684,092 66,570,714	150bp
Differential gene expression analysis	Illumina (cold treatment, bract yellow)	9.99Gb × 3	-	150bp × 2	66,653,672 66,652,570 66,677,610	150bp
	Illumina (cold treatment, bract red)	9.99Gb × 3	-	150bp × 2	66,698,974 66,689,662 66,699,054	150bp

* A genome size of 503.58 Mb (Table S2) was used for sequencing depth calculation

This pdf file has been made from the supplementary Excel file that is available from the publishers site at https://doi.org/10.1016/j.ygeno.2026.111210

Please check the xls file for full data and functionality

Table S1B. Statistics for MLA haplotype genome assemblies.

Note: The more complete assembled chromosome with larger contigs was assigned to MLAh1, and hence has a higher N50. It was used a reference for further analysis

Haplotype 1 genome assembly of *Musella lasiocarpa* (MLAh1)

	Contig		Scaffold	
	Size(bp)	**Number**	**Size(bp)**	**Number**
N10	58,414,654	1	59,458,675	1
N20	56,565,741	2	58,414,654	2
N30	51,603,189	3	56,565,741	3
N40	50,644,119	4	51,979,136	4
N50	**49,080,785**	5	51,603,189	5
N60	45,199,150	6	50,726,621	6
N70	31,115,629	8	50,644,119	7
N80	16,816,717	10	49,080,785	8
N90	14,231,493	13	45,995,787	9
Max Length	58,414,654	--	59,458,675	--
Total Number	--	365	--	357
Total Length	500,046,725	--	500,050,725	--
GC Ratio	0.4	--	0.4	--

Haplotype 2 genome assembly of *Musella lasiocarpa* (MLAh2)

	Contig		Scaffold	
	Size(bp)	**Number**	Size(bp)	**Number**
N10	52,468,949	1	56,090,541	1
N20	49,567,952	2	54,952,116	2
N30	26,052,815	4	54,496,339	3
N40	23,677,898	6	53,318,599	4
N50	**18,553,585**	8	53,045,455	5
N60	15,324,019	11	50,361,345	6
N70	13,213,194	14	49,567,952	7
N80	11,401,996	18	49,514,395	8
N90	4,725,826	24	40,909,305	9
Max Length	52,468,949	--	56,090,541	--
Total Number	--	295	--	274
Total Length	498,363,317	--	498,373,817	--
GC Ratio	0.41	--	0.41	--

Number of scaffolds not assembled is about 200

Liu et al. Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa* .

Supplementary Table S2. Chromosome and genome sizes of MLA haplotypes.

Table S2A. Chromosome size of MLA.

MLAh1		
Chromosome ID	**Number of contig in chromosome**	**Size (bp)**
MLA01.1	1	51,603,189
MLA02.1	1	58,414,654
MLA03.1	1	56,565,741
MLA04.1	1	49,080,785
MLA05.1	3	50,726,621
MLA06.1	2	45,995,787
MLA07.1	4	59,458,675
MLA08.1	2	51,979,136
MLA09.1	1	50,644,119
Total		474,468,707

MLAh2		
Chromosome ID	**Number of contig in chromosome**	**Size (bp)**
MLA01.2	1	49,567,952
MLA02.2	2	54,496,339
MLA03.2	6	54,952,116
MLA04.2	6	50,361,345
MLA05.2	2	49,514,395
MLA06.2	4	40,909,305
MLA07.2	2	56,090,541
MLA08.2	3	53,045,455
MLA09.2	4	53,318,599
Total		462,256,047

Table S2B. Genome size of MLA.

Genomescope Results						
k=21	http://qb.cshl.edu/genomescope/analysis.php?code=lebuUC94k4w8UNs26owx					
bp len:	436,080,668	heterozygos:	1.24%	kcov:	27.4	(See Suppl Fig S1)
k=17	http://qb.cshl.edu/genomescope/analysis.php?code=yGKt6jt56OSJXSzxlQdD					
bp len:	421,656,041	heterozygos:	1.19%	kcov:	28.5	
Output Pacbio hifi reads that map to assembly to estimate genome size. The method is discribed in https://github.com/schellt/backmap HiFi (/mnt/d/wsl/backmap/MLA_chromosome/only_chr_MLA_h1.fa.hifi.sort.bam)						
Mapped nucleotides: 27.70Gb						
Peak coverage: 55						
Genome size estimate: 503.58Mb			503,580,000			
hap1 Chromosome Assembly length: 474.47Mb (94.22% of estimate)			474,465,207			
hap2 Chromosome Assembly length: 462.26Mb			462,245,547			
	Assembly size	hap1 (bp)	500,050,725	hap2 (bp)	498,373,817	
	% genome		99.30%		98.97%	

190128946
300
57038683800

Liu et al. Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa* . **Supplementary Table S3**. BUSCO accessment and gene function annotation of MLA haplotype assemblies. **Table S3A.** BUSCO of MLA haplotype assemblies.

Type	MLAh1		MLAh2	
	Number	Percentage(%)	Number	Percentage(%)
Complete BUSCOs (C)	1590	98.51	1595	98.82
Complete and single-copy BUSCOs (S)	1507	93.37	1503	93.12
Complete and duplicated BUSCOs (D)	83	5.14	92	5.7
Fragmented BUSCOs (F)	9	0.56	5	0.31
Missing BUSCOs (M)	15	0.93	14	0.87
Total BUSCO	1614	-	1614	-

Tables S3B. BUSCO accessment of genes of MLA haplotype assemblies.

Type	MLAh1		MLAh2	
	Number	Percentage(%)	Number	Percentage(%)
Complete BUSCOs (C)	1581	97.96	1579	97.83
Complete and single-copy BUSCOs (S)	1497	92.75	1499	92.87
Complete and duplicated BUSCOs (D)	84	5.2	80	4.96
Fragmented BUSCOs (F)	8	0.5	14	0.87
Missing BUSCOs (M)	25	1.55	21	1.3
Total BUSCO	1614	-	1614	-

Table S3C. Functional annotation of genes of MLA haplotype assemblies.

	MLAh1	MLAh2
	Percentage(%)	Percentage(%)
NCBI Nr (Non-redundant protein) InterPro	93.74%	93.92%
(Classification of protein families) GO (Gene	86.73%	87.06%
ontology)	67.63%	67.92%
KOG (EuKaryotic orthologous groups) KEGG	67.77%	68.12%
(Kyoto encyclopedia of genes and TrEMBL	69.08%	69.34%
(Translation of EMBL nucleotide SwissProt	93.31%	93.53%
(Swiss Institute of Bioinformatics	73.86%	73.99%

Table S3D. Genome assemblies of investigated species in Fig. S2B.

Species	Abbreviaiton	Genome assembly and protein	Reference
Amomum tsao-ko	ATS	https://www.ncbi.nlm.nih.gov/bioproject/322046	Li et al., 2022
Canna edulis	CED	https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_034359265.1/	Fu et al., 2022
Ensete glaucum	EGL	https://banana-genome-hub.southgreen.fr/node/50/16444768 https://	Wang et al., 2022
Ensete ventricosum	EVE	www.ncbi.nlm.nih.gov/datasets/genome/GCA_029747655.1/https://	Haile et al., 2024
Musa acuminata	MAC	banana-genome-hub.southgreen.fr/node/50/16444767 https://banana-	Belser et al., 2021
Musa balbisiana	MBA	genome-hub.southgreen.fr/node/50/16444779 https://banana-genome-	Wang et al., 2019
Musa beccarii	MBE	hub.southgreen.fr/node/50/16444778 https://banana-genome-	Wang et al., 2023
Musa schizocarpa	MSC	hub.southgreen.fr/node/50/16444771 https://banana-genome-	Besler et al., 2018
Musa textilis	MTE	hub.southgreen.fr/node/50/16444770 https://banana-genome-	Galvez et al., 2021
Musa troglodytarum	MTR	hub.southgreen.fr/node/50/16444766 https://banana-genome-	Zhou et al., 2024
Oryza sativa	OSA	hub.southgreen.fr/node/50/16444774 https://www.ncbi.nlm.nih.gov/	Kawahara et al., 2013
Phoenix dactylifera	PDA	bioproject/322046	Hazzouri et al., 2019
Thalia dealbata	TDE	https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030544415.1/	Tang et al., 2023
Wurfbainia villosa	WVI	https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_025175895.1/	Yang et al., 2022
Zingiber officinale	ZOF	https://banana-genome-hub.southgreen.fr/node/50/16444787	Li et al., 2021

References

Belser, C., Baurens, F.C., Noel, B., Martin, G., Cruaud, C., Istace, B., Yahiaoui, N., Labadie, Karine., Hřibová, E., Doležel, J., et al., 2021. Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. Commun. Biol. 4, 1047.

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.C., Falentin, C., Genete, M., Berrabah, W., Chèvre, A. M., Delourme, R., et al., 2018. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat. Plants 4, 879–887.

Fu, Y.H., Jiang, S., Zou, M.L., Xiao, J.J., Yang, L., Luo, C.F., Rao, P., Wang, W.Q., Ou, Z.G., Liu, F.Z., *et al* ., 2022. High-quality reference genome sequences of two Cannaceae species provide insights into the evolution of Cannaceae. Front. Plant Sci. 28, 955904.

Galvez, L.C., Koh, R.B.L., Barbosa, C.F.C., Asunto, J.C., Catalla, J.L., Atienza, R.G. et al., 2021. Sequencing and de novo assembly of abaca (Musa textilis Née) var. abuab genome. Genes 12, 1202.

Haile, A.T., Kovi, M.R., Johnsen, S.S., Hvoslef-Eide, T., Tesfaye, B., Rognli, O.A., 2024. Limited genetic diversity found among genotypes of the Entada landrace (Ensete ventricosum, (Welw.) Chessman) from Ethiopia. Front. Plant Sci. 15, 1336461.

Hazzouri, K.M., Gros-Balthazard, M., Flowers, J.M., Copetti, D., Lemansour, A., Lebrun, M. Masmoudi, K., Ferrand, S., Dhar, M.I., Fresquez, Z.A., et al., 2019. Genome-wide association mapping of date palm fruit traits. Nat. Commun., 10, 4680.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J.Z., Zhou, S.G., et al., 2013. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map Kawahara, Y., de la B

Li, H.L., Dong, Z.M., Jiang, Y.S., Jiang, Y.S., Jiang, S.J., Xing, H.T. et al., 2021. Haplotype-resolved genome of diploid ginger (Zingiber officinale) and its unique gingerol biosynthetic pathway. Hortic. Res. 8, 189.

Li, P., Bai, G.X., He, J.B.., Liu, B., Long, J.R., Morcol, T., Peng, W.Y., Quan, F., Luan, X.B., Wang, Z.Z., et al., 2022. Chromosome-level genome assembly of *Amomum tsao-ko* provides insights into the biosynthesis of flavor compounds. Hortic. Res. 9, uhad153.

Tang, M., Huang, J.L., Ma, X.L., Du, J., Bi, Y.F., Guo, P.W., Lu, H., Wang, L., 2023. A near-complete genome assembly of *Thalia dealbata* Fraser (Marantaceae). Front. Plant Sci. 14, 1183361.

Wang, Z., Miao, H.X., Liu, J.H., Xu, B. Y., Yao, X.M., Xu, C.Y., Zhao, S.C., Fang, X.D., Jia, C.H., Wang, J.Y., et al., 2019. Musa balbisiana genome reveals subgenome evolution and functional divergence. Nat. Plants 5, 810–821.

Wang, ZF., Rouard, M., Droc, G., Heslop-Harrison, J.S., Ge, X.J., 2023. Genome assembly of Musa beccarii shows extensive chromosomal rearrangements and genome expansion during evolution of Musaceae genomes. GigaScience 12, giad005.

Yang, P., Zhao, H.Y., Wei, J.S., *Zhao, Y-Y., Lin, X-J., Su, J., Li, F.P., Li, M., Ma, D.M., Tan, X.K., et al* ., 2022. Chromosome-level genome assembly and functional characterization of terpene synthases provide insights into the volatile terpenoid biosynthesis of *Wurfbainia villosa* . *Plant J.* 112,

Zhou, R., Wang, S., Zhan, N., He, W.D., Deng, G.M., Dou, T.X., Zhu, X.T., Xie, W.Z., Zheng, Y.Y., Hu, C.H., et al., 2024. High-quality genome assemblies for two Australimusa bananas (Musa spp.) and insights into regulatory mechanisms of superior fiber properties. Plant Zhou, R., Wang, S.

Liu et al. Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa* .

Supplementary Table S4. Telomere characterization of MLA haplotypes.

GitHub - aaranyue/quarTeT: A telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification

Haplotype 1 genome assembly (MLAh1)

Telomere repeat monomer: AAACCCT
Both telomere found: 6
Only one telomere found: 3
No telomere found: 0

Chromosome num	Chromosome length	Chromosome end(s)	Left number of te	Left direc	Right number of telomer	Right direction
MLA01.1	51603189	both	2912	+	2692	-
MLA02.1	58414654	both	928	+	1337	-
MLA03.1	56565741	both	2182	+	3435	-
MLA04.1	49080785	both	1162	+	2154	-
MLA05.1	50726621	right	0		6431	-
MLA06.1	45995787	left	843	+	0	
MLA07.1	59458675	both	162	+	162	-
MLA08.1	51979136	left	797	+	0	
MLA09.1	50644119	both	434	+	1139	-

Haplotype 2 genome assembly (MLAh2)

Telomere repeat monomer: AAACCCT
Both telomere found: 7
Only one telomere found: 1
No telomere found: 1

Chromosome num	Chromosome length	Chromosome ends	Left number of te	Left direc	Right number of telomer	Right direction
MLA01.2	49567952	both	602	+	509	-
MLA02.2	54496339	both	467	+	1893	-
MLA03.2	54952116	both	2103	+	1484	-
MLA04.2	50361345	both	784	+	1213	-
MLA05.2	49514395	both	2397	+	2382	-
MLA06.2	40909305	left	1014	+	0	
MLA07.2	56090541	both	207	-	735	-
MLA08.2	53045455	no	0		0	
MLA09.2	53318599	both	2365	+	4456	-

The smallest labled by light yellow
The highest labled by orange

Liu et al. Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa*.

Supplementary Table S5. Centromere characterization of MLA haplotypes.

Table S5A. Abundance of Mlcen tandem repeat array in MLA.

MLAh1 as reference	Mlcen
Monomer length (bp)	134
Reads mapped to monomer	5,123,388
Total reads	220 018 257
Genome proportion of Mlcen monom	2 33%
Copies of Mlcen monomer	86,889
Genome size (Mb)	500

Table S5B. Centromere identification by tidk (https://github.com/tolkit/telomeric-identifier.
GitHub - aaranyue/quarTeT: A telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification

MLAh1					
MLAh1	start	end	length	TRlength	TRcoverage
MLA01.1	22,296,831	26,582,398	4,285,568	1,992,734	46.50%
MLA02.1	26,051,045	32,236,618	6,185,574	1,835,383	29.67%
MLA03.1	24,467,567	28,849,955	4,382,389	2,294,528	52.36%
MLA04.1	31,187,337	32,955,062	1,767,726	863,422	48.84%
MLA05.1	19,445,213	21,115,759	1,670,547	927,847	55.54%
MLA06.1	19,773,953	22,699,615	2,925,663	1,917,592	65.54%
MLA07.1	25,778,238	31,974,145	6,195,908	1,687,182	27.23%
MLA08.1	25,541,949	28,518,640	2,976,692	1,456,735	48.94%
MLA09.1	27,539,194	28,587,590	1,048,397	193,566	18.46%

MLAh2					
Chr	start	end	length	TRlength	TRcoverage
MLA01.2	17,994,545	24,403,941	6,409,397	1,673,900	26.12%
MLA02.2	16,647,437	27,223,250	10,575,814	1,614,271	15.26%
MLA03.2	23,370,800	27,466,467	4,095,668	1,037,223	25.32%
MLA04.2	31,256,388	33,098,051	1,841,664	927,715	50.37%
MLA05.2	37,192,762	37,831,858	639,097	179,631	28.11%
MLA06.2	17,518,620	20,004,928	2,486,309	1,633,886	65.72%
MLA07.2	23,972,018	31,263,982	7,291,965	1,749,690	23.99%
MLA08.2	22,234,186	27,821,035	5,586,850	1,538,339	27.53%
MLA09.2	27,901,260	30,523,792	2,622,533	1,553,478	59.24%

Table S5C. Centromere positions and chromosome arm lengths in MLA haplotypes.
Centromere location taken as position of Mlcen tandem repeat array (Estimate of position from Geneious Primer v.2022.0.2; https://www.geneious.com/)

MLAh1	length (bp)	length (Mb)	cent (Mb) = 1st arm	2nd arm (Mb)	ratio (1st in % total)	**MLAh2**	length (bp)	length (Mb)	cent (Mb) = 1st arm	2nd arm (Mb)	ratio (1st in % total)
MLA1.1	51,603,189	51.60	24.80	26.80	48.1%	MLA1.2	49,567,952	49.57	22.70	26.87	45.8%
MLA2.1	58,414,654	58.41	27.10	31.31	46.4%	MLA2.2	54,496,339	54.50	25.00	29.50	45.9%
MLA3.1	56,565,741	56.57	25.80	30.77	45.6%	MLA3.2	54,952,116	54.95	25.20	29.75	45.9%
MLA4.1	49,080,785	49.08	31.20	17.88	63.6%	MLA4.2	50,361,345	50.36	31.10	19.26	61.8%
MLA5.1	50,726,621	50.73	19.20	31.53	37.8%	MLA5.2	49,514,395	49.51	20.20	29.31	40.8%
MLA6.1	45,995,787	46.00	21.50	24.50	46.7%	MLA6.2	40,909,305	40.91	18.80	22.11	46.0%
MLA7.1	59,458,675	59.46	29.30	30.16	49.3%	MLA7.2	56,090,541	56.09	27.50	28.59	49.0%
MLA8.1	51,979,136	51.98	27.50	24.48	52.9%	MLA8.2	53,045,455	53.05	27.00	26.05	50.9%
MLA9.1	50,644,119	50.64	28.20	22.44	55.7%	MLA9.2	53,318,599	53.32	29.50	23.82	55.3%
Average arm length (Mb)			26.07	26.65					25.22	26.14	
		6 of the 9 chr are larger in MLAh1		3 chr are larger in MLAh2		12 of the 18 chromosome arms larger in MLAh1		one arm is equal		5 arms are larger in MLAh2	

Table S5D. Data used for drawing the karyotype of Fig. 6B.

	Average length	Length (Mb)	Centromere	The 2nd arm	Average arm
MLA01	50,585,571	50.59	23.74	26.85	46.9%
MLA02	56,455,497	56.46	26.04	30.41	46.1%
MLA03	55,758,929	55.76	25.50	30.26	45.7%
MLA04	49,721,065	49.72	31.16	18.57	62.7%
MLA05	50,120,508	50.12	19.71	30.41	39.3%
MLA06	43,452,546	43.45	20.14	23.31	46.3%
MLA07	57,774,608	57.77	28.40	29.38	49.2%
MLA08	52,512,296	52.51	27.26	25.26	51.9%
MLA09	51,981,359	51.98	28.85	23.13	55.5%
Average of MLAh1 and MLAh2		468.4			

Liu et al. Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa*.

Supplementary Table S6. The 5S and 45S rDNA arrays of MLA haplotypes.

Table S6A. Position and length of rDNA arrays in MLA genome.

	Chromosome	Start (nt)	End (nt)	Repeated time	bp	of genome	of assembly
5S rDNA region *)	MLA05.1	26,669,614	28,529,189	1618	1,859,575	0.369%	0.372%
5S rDNA region*)	MLA05.1	37,948,026	38,655,005	808	706,979	0.140%	0.141%
SUM					**2,566,554**	**0.510%**	**0.513%**
45S rDNA**)	MLA06.1	43,033,008	45,995,787	end of assembly	2,962,779	0.588%	0.592%
45S rDNA**)	MLA08.1	50,221,697	51,979,136	end of assembly	1,757,439	0.349%	0.351%
Total					**4,720,218**	**0.937%**	**0.944%**
*) The 5S rDNA region contains 5SrDNA genes interrupted by other sequences, so is higher genomic proportion than map to reference data in Table S6B.							

**) The assembly has not assembled the entire 45S arrays, compare with numbers under Table S6B.

Table S6B. The rDNA copy number estimation using map to reference in Illumina reads.

		MLA		EGL comparator	
5S rDNA	Monomer (bp)	658		1,056	
(5S-ITS)	Read number mapped by monomer	594,837		191,703	
	Total Illumina read number	220,018,257		245,852,534	
	Sequenced 5S rDNA length (bp)			28,755,450	
	Genome proportion of 5S rDNA	0.27		0.08%	
	Genome size (Mb)	503.58		563	
	Copies of 5S rDNA monomer	2069		366	
		MLA		EGL comparator	MBE comparator
45S rDNA	Monomer (bp)	8,347		9,984	10402
(18S-ITS1-5.8S-ITS2-26S-NTS)	Read number mapped by monomer	5,676,003		2,982,590	
	Total Illumina read number	220,018,257		245,852,534	
	Sequenced 45S rDNA length (bp)			447,388,500	
	Genome proportion of 45S rDNA	2.58%		1.21%	5.0%
	Genome size (Mb)	503.58		563	570
	Copies of 45S rDNA monomer	1465		587	2740
	Number of sites	2		1	3
	Average copies per site	733		587	913

Supplementary Table S7. Statistics of orthogroups and gene families in Musaceae.

Table S7A. Shared orthogroups and gene clustering of the investigated species in Musaceae.

	Ensete glaucum (**EGL**)	*Ensete ventricosum* (**EVE**)	*Musa accuminata* (**MAC**)	*Musa beccarii* (**MBE**)	*Musella lasiocarpa* (**MLA**)	*Musa textilis* (**MTE**)	Total
Number of genes	36,836	34,701	36,979	39,112	35,387	35,069	218,084
Number of genes in orthogroups	32,003	31,468	33,505	33,492	32,418	32,190	195,076 (89.45%)
Number of unassigned genes	4,833	3,233	3,474	5,620	2,969	2,879	23,008 (10.55%)
Percentage of genes in orthogroups	86.9%	90.7%	90.6%	85.6%	91.6%	91.8%	
Percentage of unassigned genes	13.1%	9.3%	9.4%	14.4%	8.4%	8.2%	
Number of orthogroups	26,086	25,657	25,561	26,646	24,183	25,213	31,235 (Table S7C)
Percentage of orthogroups	83.5%	82.1%	81.8%	85.3%	77.4%	80.7%	
Number of species-specific orthogroups	132	82	247	225	212	133	
Number of genes in species-specific orthogroups	356	251	1135	589	1501	305	
Percentage of genes in species-specific orthogrou	1.00%	0.70%	3.10%	1.50%	4.20%	0.90%	

Table S7B. Pairwise gene numbers in syntenic region between the species pairs in Musaceae.

This table provides a summary of collinearity statistics between different genome pairs. It represents the total number of gene pairs identified between each unique genome pair, facilitating insights into genome-level interactions and syntenic relationships.

Genome pair	Gene number
MLA vs MAC	19,683
MLA vs MBE	18,768
MLA vs EGL	17,529
EGL vs MAC	14,248
EGL vs MBE	14,722
MAC vs MBE	21,214

Table S7C. Database of 31,235 orthogroups in Musaceae.

Orthogroup	*Ensete glaucum* (**EGL**)	*Ensete ventricosum* (**EVE**)	*Musa accuminata* (**MAC**)	*Musa beccarii* (**MBE**)	*Musella lasiocarpa* (**MLA**)	*Musa textilis* (**MTE**)	Total
OG0000000	0	0	0	0	521	0	521
OG0000001	0	0	3	0	300	1	304
OG0000002	23	24	26	29	24	18	144
OG0000003	0	0	136	0	0	0	136
OG0000004	21	20	21	20	20	12	114
OG0000005	1	0	1	0	108	0	110
OG0000006	24	1	23	18	21	22	109

Further lines deleted in pdf, please see xls file for all data

Liu et al. Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa*.

Supplementary Table S8. Statistics and database of orthologous gene families for the investigated species in Fig. 2C.

Table S8A. Gene family categories in genomes of MLA and 13 other monocotyledonous species.
MAC: Musa accuminata (MAC 98 8%; Belser et al 2021); MSC: Musa schizocarpa (Besler et al 2018); MBA: Musa balbisiana (Wang et al 2019); MTE: Musa textilis (Galvez et al 2021); MTR: Musa troglodytarum (Zhou et al 2024); MBE: Musa beccarii (Wang et al 2019); EGL: Ensete glaucum (Wang et al 2022); EVE: Ensete ventricosum (Borell et al 2019); MLA Musella lasiocarpa; WVI: Wurfbainia villosa (Yang et al 2022); ZOF: Zingiber officinalis (Li HL et al 2021); CED: Canna edulis (Fu et al 2022); TDE: Thalia dealbata (Tang et al 2023); OSA: Oryza sativa (Kawahara et al 2013)

Species (Abbreviation)	Single-copy orthologs	Multiple-copy orthologs	Unique orthologs	Other orthologs	Total orthologs
Musa accuminata (MAC)	994	5,109	59	13,456	19,618
Musa schizocarpa (MSC)	994	4,882	43	11,979	17,898
Musa balbisiana (MBA)	994	4,496	84	11,944	17,518
Musa textilis (MTE)	994	5,127	73	11,684	17,878
Musa troglodytarum (MTR)	994	4,852	189	13,052	19,087
Musa beccarii (MBE)	994	5,167	160	12,523	18,844
Ensete glaucum (EGL)	994	4,983	112	12,018	18,107
Ensete ventricosum (EVE)	994	4,987	46	11,817	17,844
Musella lasiocarpa (MLA)	994	4,926	155	10,520	16,595
Wurfbainia villosa (WVI)	994	5,052	650	9,170	15,866
Zingiber officinalis (ZOF)	994	4,451	663	8,945	15,053
Canna edulis (CED)	994	3,962	539	9,459	14,954
Thalia dealbata (TDE)	994	3,651	264	9,055	13,964
Oryza sativa (OSA)	994	3,175	2,218	8,689	15,076

Table S8B. Statistics of gene family expansion and contraction for each clade.

Species abbreviaiton<Clade numbe	Expansion gene families	Contraction gene families	Significant expansion ge	Significant contraction gene families (P<0.5)
MTR<0>	1108	1931		
MTE<1>	1420	1217		
MSC<2>	498	1418		
MAC<3>	829	460		
EVE<4>	572	742		
EGL<5>	579	648		
<6>=<1+10>	202	295		
MBE<7>	979	814		
MBA<8>	565	2828		
<9>=<2+3>	439	217		
<10>=<4+5>	439	321		
MLA<11>	760	1234	154 (Genes listed in Table S8C)	162 (Genes listed in Table S8C)
<12>=<0+1+7>	888	353		
<13>=<3+3+8>	262	353		
TDE<14>	857	3449		
CED<15>	1615	3007		
ZOF<16>	1318	2042		
WVI<17>	2802	724		
<18>=<4+5+11>	197	297		
<19>=<0+1+2+3+7+8>	105	59		
<20>=<14+15>	29	630		
<21>=<16+17>	1697	1294		
<22>=<0+1+2+3+4+5+7+8+11>	2261	696		
<23>=<14+15+16+17>	18	422		
<24>=<22+23>	23	141		
OSA<25>	1793	3507		

Table S8C. Significant expanded and contracted gene families and genes of MLA (Fig. 2C).

Significant expansion gene families	Genes of significant expansion gene fa	Significant contraction gene	Genes of significant
OG0000005	[ML1h01G00710 1 ML1h01G01985 1	OG0000016	[ML1h01G00025 1 ML1
OG0000009	[ML1h01G01071 1 ML1h01G02025 1	OG0000021	[ML1h05G02529 1 ML1
OG0000012	[ML1h02G03016 1 ML1h05G00366 1	OG0000029	[ML1h01G02343 1 ML1
OG0000015	[ML1h01G00276.1,ML1h01G00293.1,	OG0000032	[ML1h01G00290.1,ML1
OG0000017	[ML1h01G00752 1 ML1h01G00891 1	OG0000035	[ML1h01G01622 1 ML1
OG0000018	[ML1h01G02096 1 ML1h02G01364 1	OG0000040	[ML1h05G01678 1 ML1
OG0000019	[ML1h01G00043 1 ML1h02G01361 1	OG0000046	[ML1h03G03657 1 ML1
OG0000022	[ML1h01G00198.1,ML1h01G02117.1,	OG0000052	[ML1h01G01878.1,ML1
OG0000023	[ML1h01G03016.1,ML1h02G01070.1,	OG0000056	[ML1h03G00257.1,ML1
OG0000025	[ML1h01G00300 1 ML1h01G01557 1	OG0000057	[ML1h08G01719 1 ML1

Further lines deleted in pdf, please see xls file for all data

Supplementary Table S9. GO enrichment results of unique gene families in MLA (Fig. S5A).

Class	GO_Name	GO_ID	GO_Lev	P_value	EnrichmentSc	HitsGenes	HitsGenes	AllGene	AllGene	GenesOfSelectedSetInGOterm	corrected p-value(BH method)
Biological process	amylopectin metabolic process	GO:2000896	5	0	32.23273381	14	40	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	0
Biological process	nitrate transmembrane transport	GO:0015706	8	0.00E+00	28.65131894	14	45	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	0.00E+00
Biological process	inorganic ion transmembrane transport	GO:0098660	6	0.00E+00	7.027216781	38	498	139	12801	ML1h01G01140,ML1h01G01142,ML1h08G02300,ML1h01G0114	0.00E+00
Biological process	monosaccharide transmembrane transport	GO:0015749	8	0.00E+00	20.07166574	17	78	139	12801	ML1h08G02300,ML1h02G02987,ML1h02G02986,ML1h02G0298	0.00E+00
Biological process	monoatomic ion transmembrane transport	GO:0034220	6	0.00E+00	4.800485537	38	729	139	12801	ML1h01G01140,ML1h01G01142,ML1h08G02300,ML1h01G0114	0.00E+00
Biological process	nitrate import	GO:1902025	6	0.00E+00	28.65131894	14	45	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	0.00E+00
Biological process	intracellular glucose homeostasis	GO:0001678	6	0.00E+00	21.9270298	15	63	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	0.00E+00
Biological process	cellular response to monosaccharide stimulus	GO:0071326	7	0.00E+00	23.81729099	15	58	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	0.00E+00
Biological process	cellular response to hexose stimulus	GO:0071331	8	0.00E+00	23.81729099	15	58	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	0.00E+00
Biological process	cellular response to glucose stimulus	GO:0071333	9	0.00E+00	23.81729099	15	58	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	0.00E+00
Biological process	cellular response to fructose stimulus	GO:0071332	9	0.00E+00	47.75219824	14	27	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	0.00E+00
Biological process	amylopectin biosynthetic process	GO:0010021	6	0.00E+00	32.23273381	14	40	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	0.00E+00
Biological process	glucose transmembrane transport	GO:1904659	10	1.11E-16	20.33233673	17	77	139	12801	ML1h08G02300,ML1h02G02987,ML1h02G02986,ML1h02G0298	6.32E-15
Biological process	proton transmembrane transport	GO:1902600	8	1.11E-16	13.55978285	24	163	139	12801	ML1h08G02300,ML1h09G02109,ML1h09G02119,ML1h09G0211	6.32E-15
Biological process	glucose import	GO:0046323	11	1.11E-16	20.33233673	17	77	139	12801	ML1h08G02300,ML1h02G02987,ML1h02G02986,ML1h02G0298	6.32E-15
Biological process	hexose transmembrane transport	GO:0008645	9	1.11E-16	20.33233673	17	77	139	12801	ML1h08G02300,ML1h02G02987,ML1h02G02986,ML1h02G0298	6.32E-15
Biological process	glucose homeostasis	GO:0042593	5	2.22E-16	20.3147482	15	68	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	9.67E-15
Biological process	cellular response to disaccharide stimulus	GO:0071324	7	2.22E-16	39.06998038	14	33	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	9.67E-15
Biological process	cellular response to sucrose stimulus	GO:0071329	8	2.22E-16	39.06998038	14	33	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	9.67E-15
Biological process	transmembrane transport	GO:0055085	5	5.55E-16	4.186069326	53	1166	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01139,ML1h01G0115	2.05E-14
Biological process	response to fructose	GO:0009750	8	8.88E-16	22.2294716	14	58	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	3.13E-14
Biological process	response to nitrate	GO:0010167	5	5.00E-15	20.14545863	14	64	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	1.68E-13
Biological process	carbohydrate transmembrane transport	GO:0034219	7	6.66E-15	13.3811105	17	117	139	12801	ML1h08G02300,ML1h02G02987,ML1h02G02986,ML1h02G0298	2.14E-13
Biological process	carbohydrate homeostasis	GO:0033500	4	7.22E-15	17.05435651	15	81	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	2.23E-13
Biological process	cellular response to nitrate	GO:0071249	7	1.68E-14	40.04066312	10	23	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	4.96E-13
Biological process	carbohydrate transport	GO:0008643	6	1.90E-14	12.62572523	17	124	139	12801	ML1h08G02300,ML1h02G02987,ML1h02G02986,ML1h02G0298	5.40E-13
Biological process	response to glucose	GO:0009749	8	4.34E-14	15.1802514	15	91	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	1.19E-12
Biological process	monoatomic ion transport	GO:0006811	5	7.48E-14	3.972251937	38	881	139	12801	ML1h01G01140,ML1h01G01142,ML1h08G02300,ML1h01G0114	1.98E-12
Biological process	intracellular chemical homeostasis	GO:0055082	4	2.43E-13	7.798242052	21	248	139	12801	ML1h07G00976,ML1h09G02109,ML1h09G02119,ML1h09G0211	6.20E-12
Biological process	regulation of cellular respiration	GO:0043457	7	5.40E-13	71.62829736	7	9	139	12801	ML1h06G00021,ML1h06G00157,ML1h06G00123,ML1h182G000	1.33E-11
Biological process	inorganic anion transmembrane transport	GO:0098661	7	1.16E-12	13.86354142	14	93	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	2.78E-11
Biological process	response to hexose	GO:0009746	7	1.17E-12	12.22480423	15	113	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	2.71E-11
Biological process	response to monosaccharide	GO:0034284	6	1.52E-12	12.01219894	15	115	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	3.41E-11
Biological process	starch metabolic process	GO:0005982	7	2.82E-12	13.02332679	14	99	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	6.15E-11
Biological process	inorganic cation transmembrane transport	GO:0098662	7	6.32E-12	5.567366762	24	397	139	12801	ML1h08G02300,ML1h09G02109,ML1h09G02119,ML1h09G0211	1.34E-10
Biological process	cellular homeostasis	GO:0019725	3	8.53E-12	6.489812177	21	298	139	12801	ML1h07G00976,ML1h09G02109,ML1h09G02119,ML1h09G0211	1.75E-10
Biological process	cellular response to carbohydrate stimulus	GO:0071322	6	9.35E-12	10.62617598	15	130	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	1.87E-10
Biological process	monoatomic cation transmembrane transport	GO:0098655	7	1.53E-11	5.338755083	24	414	139	12801	ML1h08G02300,ML1h09G02109,ML1h09G02119,ML1h09G0211	2.98E-10
Biological process	response to disaccharide	GO:0034285	6	2.76E-11	9.867163412	15	140	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	5.24E-10
Biological process	response to sucrose	GO:0009744	7	2.76E-11	9.867163412	15	140	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	5.24E-10
Biological process	cellular response to reactive nitrogen species	GO:1902170	6	4.80E-11	20.02033156	10	46	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	8.67E-10
Biological process	lateral root development	GO:0048527	7	7.59E-11	7.637024039	17	205	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	1.34E-09
Biological process	inorganic anion transport	GO:0015698	5	1.66E-10	9.694055282	14	133	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	2.86E-09
Biological process	post-embryonic root development	GO:0048528	6	1.08E-09	6.442756906	17	243	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	1.81E-08
Biological process	monoatomic cation transport	GO:0006812	6	1.11E-09	4.33381295	24	510	139	12801	ML1h08G02300,ML1h09G02109,ML1h09G02119,ML1h09G0211	1.82E-08
Biological process	transport	GO:0006810	4	2.32E-09	2.20534384	54	2255	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01139,ML1h01G0115	3.72E-08
Biological process	cellular response to light stimulus	GO:0071482	7	2.70E-09	7.861642393	14	164	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	4.25E-08
Biological process	obsolete cellular glucan metabolic process	GO:0006073	0	2.80E-09	7.12063339	15	194	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	4.32E-08
Biological process	glucan metabolic process	GO:0044042	6	4.88E-09	6.838628107	15	202	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	7.37E-08
Biological process	establishment of localization	GO:0051234	3	5.26E-09	2.156569974	54	2306	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01139,ML1h01G0115	7.79E-08
Biological process	temperature compensation of the circadian clock	GO:0010378	5	7.66E-09	57.55845324	5	8	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	1.11E-07
Biological process	cellular response to radiation	GO:0071478	6	9.03E-09	7.162829736	14	180	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	1.28E-07
Biological process	localization	GO:0051179	2	2.01E-08	2.055659044	55	2464	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01139,ML1h01G0115	2.80E-07
Biological process	response to carbohydrate	GO:0009743	5	2.40E-08	5.624032072	16	262	139	12801	ML1h07G00976,ML1h06G02506,ML1h06G02507,ML1h01G0127	3.28E-07
Biological process	chemical homeostasis	GO:0048878	3	1.04E-07	3.852517986	21	502	139	12801	ML1h07G00976,ML1h09G02109,ML1h09G02119,ML1h09G0211	1.40E-06
Biological process	intracellular pH reduction	GO:0051452	9	1.28E-07	24.02439787	6	23	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	1.69E-06
Biological process	pH reduction	GO:0045851	7	1.28E-07	24.02439787	6	23	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	1.69E-06
Biological process	vacuolar acidification	GO:0007035	10	1.28E-07	24.02439787	6	23	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	1.69E-06
Biological process	response to nematode	GO:0009624	6	1.98E-07	8.770811922	10	105	139	12801	ML1h08G02300,ML1h08G02295,ML1h08G02296,ML1h08G0228	2.48E-06
Biological process	post-embryonic plant organ development	GO:0090696	5	2.20E-07	4.498821632	17	348	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	2.71E-06
Biological process	regulation of generation of precursor metabolites a	GO:0043467	6	3.56E-07	14.99196922	7	43	139	12801	ML1h06G00021,ML1h06G00157,ML1h06G00123,ML1h182G000	4.31E-06
Biological process	obsolete cellular polysaccharide metabolic process	GO:0044264	0	3.88E-07	4.916024476	15	281	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	4.63E-06
Biological process	cellular response to inorganic substance	GO:0071241	5	7.48E-07	7.611035139	10	121	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	8.78E-06
Biological process	cellular response to abiotic stimulus	GO:0071214	5	1.26E-06	4.810855793	14	268	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	1.45E-05
Biological process	cellular response to environmental stimulus	GO:0104004	4	1.26E-06	4.810855793	14	268	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	1.45E-05
Biological process	homeostatic process	GO:0042592	2	1.44E-06	3.283470337	21	589	139	12801	ML1h07G00976,ML1h09G02109,ML1h09G02119,ML1h09G0211	1.61E-05
Biological process	cellular respiration	GO:0045333	6	1.86E-06	7.969631987	9	104	139	12801	ML1h06G00021,ML1h04G01864,ML1h179G00005,ML1h06G001	2.05E-05
Biological process	regulation of intracellular pH	GO:0051453	8	3.05E-06	14.54108292	6	38	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	3.32E-05
Biological process	regulation of cellular pH	GO:0030641	7	3.05E-06	14.54108292	6	38	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	3.32E-05
Biological process	energy derivation by oxidation of organic compoun	GO:0015980	5	3.24E-06	6.48545952	10	142	139	12801	ML1h06G00021,ML1h04G01864,ML1h179G00005,ML1h01G012	3.42E-05
Biological process	polysaccharide metabolic process	GO:0005976	5	4.85E-06	4.01570604	15	344	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	5.05E-05
Biological process	positive regulation of long-day photoperiodism, flov	GO:0048578	8	6.27E-06	18.41870504	5	25	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	6.44E-05
Biological process	monoatomic anion transmembrane transport	GO:0098656	7	8.23E-06	4.093045564	14	315	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	8.35E-05
Biological process	cellular response to nitrogen compound	GO:1901699	5	8.40E-06	5.828704125	10	158	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	8.40E-05
Biological process	obsolete cellular monovalent inorganic cation home	GO:0030004	0	1.24E-05	11.51169065	6	48	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	1.22E-04
Biological process	regulation of pH	GO:0006885	6	1.24E-05	11.51169065	6	48	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	1.22E-04
Biological process	cellular response to oxygen-containing compound	GO:1901701	5	1.46E-05	2.463407052	26	972	139	12801	ML1h01G01140,ML1h01G01142,ML1h07G00976,ML1h01G0114	1.41E-04
Biological process	carbohydrate derivative biosynthetic process	GO:1901137	5	1.75E-05	3.606796025	15	383	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	1.66E-04
Biological process	obsolete cellular carbohydrate metabolic process	GO:0044262	0	3.35E-05	3.410871303	15	405	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	3.14E-04
Biological process	monoatomic anion transport	GO:0006820	6	6.03E-05	3.419918707	14	377	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	5.58E-04

Biological process	lateral root formation	GO:0010311	5	1.35E-04	6.198602656	7	104	139	12801	ML1h02G02987,ML1h02G02986,ML1h02G02989,ML1h02G0298	1.24E-03
Biological process	response to radiation	GO:0009314	4	1.63E-04	2.178181769	25	1057	139	12801	ML1h228G00002,ML1h07G01265,ML1h06G02506,ML1h06G025	1.47E-03
Biological process	obsolete monovalent inorganic cation homeostasis	GO:0055067	0	3.61E-04	6.351277599	6	87	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	3.22E-03
Biological process	lateral root morphogenesis	GO:0010102	7	4.01E-04	5.198828034	7	124	139	12801	ML1h02G02987,ML1h02G02986,ML1h02G02989,ML1h02G0298	3.53E-03
Biological process	response to oxygen-containing compound	GO:1901700	4	4.14E-04	1.682073519	40	2190	139	12801	ML1h01G01140,ML1h01G01142,ML1h07G00976,ML1h02G0180	3.61E-03
Biological process	post-embryonic root morphogenesis	GO:0010101	6	4.20E-04	5.15723741	7	125	139	12801	ML1h02G02987,ML1h02G02986,ML1h02G02989,ML1h02G0298	3.62E-03
Biological process	response to light stimulus	GO:0009416	5	4.21E-04	2.130936699	23	994	139	12801	ML1h07G01265,ML1h06G02506,ML1h06G02507,ML1h01G0127	3.58E-03
Biological process	regulation of circadian rhythm	GO:0042752	4	7.38E-04	6.976782211	5	66	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	6.20E-03
Biological process	response to nitrogen compound	GO:1901698	4	1.25E-03	2.538010536	14	508	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	1.03E-02
Biological process	vacuolar transport	GO:0007034	6	1.37E-03	4.933581706	6	112	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	1.12E-02
Biological process	response to far red light	GO:0010218	7	1.66E-03	5.828704125	5	79	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	1.34E-02
Biological process	regulation of long-day photoperiodism, flowering	GO:0048586	6	1.76E-03	5.755845324	5	80	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	1.40E-02
Biological process	lignin biosynthetic process	GO:0009809	7	2.42E-03	5.35427472	5	86	139	12801	ML1h08G02384,ML1h08G02385,ML1h08G02386,ML1h08G0238	1.90E-02
Biological process	carbohydrate derivative metabolic process	GO:1901135	4	2.69E-03	2.253512035	15	613	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	2.09E-02
Biological process	carbohydrate metabolic process	GO:0005975	4	3.18E-03	2.213786663	15	624	139	12801	ML1h06G02506,ML1h06G02507,ML1h01G01272,ML1h01G0127	2.45E-02
Biological process	intracellular monoatomic cation homeostasis	GO:0030003	6	3.39E-03	4.12359068	6	134	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	2.59E-02
Biological process	long-day photoperiodism, flowering	GO:0048574	8	3.72E-03	4.847027641	5	95	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	2.81E-02
Biological process	long-day photoperiodism	GO:0048571	7	4.25E-03	4.698649244	5	98	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	3.18E-02
Biological process	root development	GO:0048364	5	4.29E-03	2.022726005	17	774	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	3.18E-02
Biological process	root system development	GO:0022622	5	4.41E-03	2.017512794	17	776	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	3.23E-02
Biological process	lignin metabolic process	GO:0009808	6	4.84E-03	4.559085405	5	101	139	12801	ML1h08G02384,ML1h08G02385,ML1h08G02386,ML1h08G0238	3.51E-02
Biological process	intracellular monoatomic ion homeostasis	GO:0006873	5	5.87E-03	3.683741007	6	150	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	0.03985918
Biological process	generation of precursor metabolites and energy	GO:0006091	4	6.10E-03	2.544020033	10	362	139	12801	ML1h06G00021,ML1h04G01864,ML1h179G00005,ML1h01G012	0.040330631
Biological process	xenobiotic transmembrane transport	GO:0006855	6	7.73E-03	4.074934742	5	113	139	12801	ML1h06G02560,ML1h06G02541,ML1h06G02555,ML1h06G0255	0.050630477
Biological process	positive regulation of reproductive process	GO:2000243	5	8.02E-03	4.039189701	5	114	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	0.05204103
Biological process	xenobiotic transport	GO:0042908	5	8.61E-03	3.969548499	5	116	139	12801	ML1h06G02560,ML1h06G02541,ML1h06G02555,ML1h06G0255	0.055406374
Biological process	post-embryonic plant organ morphogenesis	GO:0090697	5	1.25E-02	2.815085923	7	229	139	12801	ML1h02G02987,ML1h02G02986,ML1h02G02989,ML1h02G0298	0.078851681
Biological process	response to blue light	GO:0009637	6	1.28E-02	3.597403327	5	128	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	0.079716749
Biological process	obsolete nucleic acid phosphodiester bond hydroly	GO:0090305	0	1.29E-02	2.081209609	12	531	139	12801	ML1h06G02263,ML1h07G01244,ML1h09G01863,ML1h09G0186	0.079380294
Biological process	regulation of photoperiodism, flowering	GO:2000028	5	1.49E-02	3.462162601	5	133	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	0.089783541
Biological process	positive regulation of post-embryonic development	GO:0048582	7	1.54E-02	3.436325566	5	134	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	0.091726845
Biological process	cellular process	GO:0009987	2	1.92E-02	1.059653532	132	11472	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01139,ML1h01G0115	0.113485095
Biological process	cellular response to acid chemical	GO:0071229	5	2.00E-02	2.03011779	11	499	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	0.117528813
Biological process	organic substance transport	GO:0071702	5	0.02071321	1.67781726	18	988	139	12801	ML1h08G02300,ML1h07G01244,ML1h02G02987,ML1h02G0298	0.120691157
Biological process	phenylpropanoid biosynthetic process	GO:0009699	6	0.02088789	3.175638799	5	145	139	12801	ML1h08G02384,ML1h08G02385,ML1h08G02386,ML1h08G0238	0.120758133
Biological process	nitrogen compound transport	GO:0071705	5	0.02280526	1.764243777	15	783	139	12801	ML1h01G01140,ML1h01G01142,ML1h01G01144,ML1h01G0114	0.128823627
Biological process	response to karrikin	GO:0080167	4	0.02499739	3.029392276	5	152	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	0.139083245
Biological process	regulation of biological quality	GO:0065008	3	0.02508854	1.52935096	23	1385	139	12801	ML1h07G00976,ML1h09G02109,ML1h09G02119,ML1h09G0211	0.138548671
Biological process	RNA modification	GO:0009451	8	0.02720843	2.015175606	10	457	139	12801	ML1h06G02263,ML1h09G01863,ML1h09G01862,ML1h09G0186	0.148045884
Biological process	response to hydrogen peroxide	GO:0042542	6	0.03029188	2.877922662	5	160	139	12801	ML1h05G00450,ML1h05G00443,ML1h05G00437,ML1h05G0042	0.163620351
Biological process	phenylpropanoid metabolic process	GO:0009698	5	0.03099961	2.860047366	5	161	139	12801	ML1h08G02384,ML1h08G02385,ML1h08G02386,ML1h08G0238	0.166229804
Biological process	monoatomic cation homeostasis	GO:0055080	5	0.03221928	2.523110279	6	219	139	12801	ML1h09G02109,ML1h09G02119,ML1h09G02118,ML1h09G0211	0.170301883
Biological process	plant organ formation	GO:1905393	4	0.03415891	2.28600949	7	282	139	12801	ML1h02G02987,ML1h02G02986,ML1h02G02989,ML1h02G0298	0.178011211
Biological process	photosynthesis	GO:0015979	4	0.04115258	2.646365666	5	174	139	12801	ML1h19G00003,ML1h04G01864,ML1h50G00003,ML1h57G0000	0.208581578
Biological process	flavonoid biosynthetic process	GO:0009813	5	0.04463787	2.586896775	5	178	139	12801	ML1h08G02384,ML1h08G02385,ML1h08G02386,ML1h08G0238	0.22021348
Biological process	response to inorganic substance	GO:0010035	4	0.04695564	1.484182517	20	1241	139	12801	ML1h01G01140,ML1h01G01142,ML1h07G00976,ML1h01G0114	0.230113752

Supplementary Table S10. GO enrichment results of significantly expanded gene families in MLA (Fig. S5B right).

Class	GO_Name	GO_ID	GO_Lev	P_value	EnrichmentSc	HitsGenesCo	HitsGenesCo	AllGenesCou	AllGene	GenesOfSelectedSetInGOterm	corrected p-value(BH method)
Biologica	photosynthesis, light reaction	GO:0019684	5	0	7.588483718	55	142	657	12872	ML52G00001.1,ML122G00006.1,ML84G00002.1,ML222G00002.1	0
Biologica	response to herbicide	GO:0009635	5	1.11E-16	15.86025948	17	21	657	12872	ML52G00001.1,ML222G00002.1,ML207G00003.1,ML58G00001.1	3.14E-14
Biologica	obsolete protein-chromophore linkage	GO:0018298	0	1.11E-16	9.056718647	49	106	657	12872	ML01G00001.1,ML122G00006.1,ML84G00002.1,ML19G00003.1,	3.14E-14
Biologica	proton export across plasma membrane	GO:0120029	9	2.22E-16	17.28713403	15	17	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	3.14E-14
Biologica	DNA-templated transcription termination	GO:0006353	9	3.33E-16	7.299012147	38	102	657	12872	ML05G00169.1,ML09G02047.1,ML03G03611.1,ML05G02094.1,M	3.77E-14
Biologica	calcium-mediated signaling	GO:0019722	8	3.33E-16	12.40832065	19	30	657	12872	ML06G02238.1,ML02G01194.1,ML01G00198.1,ML08G01316.1,M	3.77E-14
Biologica	electron transport chain	GO:0022900	5	3.33E-16	7.218136666	56	152	657	12872	ML52G00001.1,ML122G00006.1,ML84G00002.1,ML222G00002.1	3.77E-14
Biologica	photosynthetic electron transport chain	GO:0009767	6	3.33E-16	12.52982195	55	86	657	12872	ML52G00001.1,ML122G00006.1,ML84G00002.1,ML222G00002.1	3.77E-14
Biologica	photosynthesis	GO:0015979	4	3.33E-16	5.55596447	76	268	657	12872	ML52G00001.1,ML122G00006.1,ML84G00002.1,ML19G00003.1,	3.77E-14
Biologica	DNA integration	GO:0015074	8	3.33E-16	5.705387459	53	182	657	12872	ML05G03836.1,ML09G02003.1,ML09G01650.1,ML05G04887.1,M	3.77E-14
Biologica	intracellular signaling cassette	GO:0141124	7	4.44E-16	8.288959138	33	78	657	12872	ML08G00210.1,ML06G02238.1,ML02G01194.1,ML09G00060.1,M	2.29E-14
Biologica	photosynthetic electron transport in photosystem II	GO:0009772	7	5.55E-16	16.8491933	43	50	657	12872	ML52G00001.1,ML122G00006.1,ML84G00002.1,ML222G00002.1	2.62E-14
Biologica	generation of precursor metabolites and energy	GO:0006091	4	2.46E-14	2.863808283	63	431	657	12872	ML52G00001.1,ML122G00006.1,ML84G00002.1,ML09G01138.1,	1.07E-12
Biologica	proton transmembrane transport	GO:1902600	8	9.19E-13	9.499192842	16	33	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	3.72E-11
Biologica	regulation of intracellular pH	GO:0051453	8	8.94E-12	9.183789954	15	32	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	3.37E-10
Biologica	export across plasma membrane	GO:0140115	6	8.94E-12	9.183789954	15	32	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	3.37E-10
Biologica	regulation of cellular pH	GO:0030641	7	8.94E-12	9.183789954	15	32	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	3.37E-10
Biologica	response to toxic substance	GO:0009636	4	3.85E-11	7.240553239	17	46	657	12872	ML52G00001.1,ML222G00002.1,ML207G00003.1,ML58G00001.1	1.21E-09
Biologica	S-adenosylmethionine biosynthetic process	GO:0006556	6	1.12E-10	16.02988792	9	11	657	12872	ML01G00355.1,ML01G01109.1,ML09G00826.1,ML02G00623.1,M	3.33E-09
Biologica	establishment or maintenance of cell polarity	GO:0007163	3	1.14E-10	10.68659195	12	22	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	3.24E-09
Biologica	regulation of pH	GO:0006885	6	4.34E-10	7.347031963	15	40	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	1.17E-08
Biologica	S-adenosylmethionine metabolic process	GO:0046500	5	8.46E-09	11.75525114	9	15	657	12872	ML01G00355.1,ML01G01109.1,ML09G00826.1,ML02G00623.1,M	2.18E-07
Biologica	cortical cytoskeleton organization	GO:0030865	7	1.63E-08	7.584032995	12	31	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	4.02E-07
Biologica	carbohydrate metabolic process	GO:0005975	4	1.71E-08	1.715737239	105	1199	657	12872	ML09G00491.1,ML09G00416.1,ML04G02209.1,ML09G01138.1,M	4.03E-07
Biologica	small GTPase-mediated signal transduction	GO:0007264	8	1.96E-08	6.233845302	14	44	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	4.44E-07
Biologica	regulation of DNA-templated transcription	GO:0006355	9	2.73E-08	1.655669175	114	1349	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	5.93E-07
Biologica	regulation of RNA biosynthetic process	GO:2001141	8	2.73E-08	1.655669175	114	1349	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	5.93E-07
Biologica	cell-cell signaling involved in cell fate commitment	GO:0045168	6	1.16E-07	16.79321592	6	7	657	12872	ML06G00545.1,ML08G02514.1,ML06G01498.1,ML04G01083.1,M	2.34E-06
Biologica	hydrogen peroxide catabolic process	GO:0042744	6	1.27E-07	3.622066178	22	119	657	12872	ML05G03805.1,ML08G03399.1,ML05G02681.1,ML05G03334.1,M	2.48E-06
Biologica	hydrogen peroxide metabolic process	GO:0042743	5	1.48E-07	3.591882293	22	120	657	12872	ML05G03805.1,ML08G03399.1,ML05G02681.1,ML05G03334.1,M	2.79E-06
Biologica	adenine salvage	GO:0006168	11	2.41E-07	12.4676906	7	11	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	4.40E-06
Biologica	purine nucleobase salvage	GO:0043096	10	2.41E-07	12.4676906	7	11	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	4.40E-06
Biologica	cellular metabolic compound salvage	GO:0043094	4	2.79E-07	4.817725878	15	61	657	12872	ML04G02338.1,ML06G01053.1,ML46G00002.1,ML07G00436.1,M	4.79E-06
Biologica	regulation of cell shape	GO:0008360	7	3.10E-07	6.028333919	12	39	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	5.16E-06
Biologica	DNA-templated transcription	GO:0006351	9	5.40E-07	1.539378126	121	1540	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	8.73E-06
Biologica	adenine metabolic process	GO:0046083	9	5.53E-07	11.42871639	7	12	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	8.70E-06
Biologica	adenine biosynthetic process	GO:0046084	10	5.53E-07	11.42871639	7	12	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	8.70E-06
Biologica	AMP salvage	GO:0044209	13	5.53E-07	11.42871639	7	12	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	8.70E-06
Biologica	RNA biosynthetic process	GO:0032774	8	5.99E-07	1.536385168	121	1543	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	8.69E-06
Biologica	reactive oxygen species metabolic process	GO:0072593	4	6.29E-07	3.315583655	22	130	657	12872	ML05G03805.1,ML08G03399.1,ML05G02681.1,ML05G03334.1,M	8.90E-06
Biologica	regulation of actin cytoskeleton organization	GO:0032956	8	7.61E-07	5.597738639	12	42	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	1.05E-05
Biologica	regulation of actin filament-based process	GO:0032970	5	1.01E-06	5.46755867	12	43	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	1.36E-05
Biologica	regulation of RNA metabolic process	GO:0051252	7	1.25E-06	1.539281679	114	1451	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	1.64E-05
Biologica	cell-cell signaling	GO:0007267	5	1.27E-06	13.06139016	6	9	657	12872	ML06G00545.1,ML08G02514.1,ML06G01498.1,ML04G01083.1,M	1.64E-05
Biologica	pectin catabolic process	GO:0045490	8	1.66E-06	3.635644683	18	97	657	12872	ML07G00497.1,ML04G02769.1,ML03G00140.1,ML07G00493.1,M	2.09E-05
Biologica	DNA metabolic process	GO:0006259	7	1.90E-06	1.931119513	55	558	657	12872	ML05G03836.1,ML09G02003.1,ML09G01650.1,ML05G04887.1,M	2.34E-05
Biologica	purine ribonucleoside salvage	GO:0006166	11	2.19E-06	9.796042618	7	14	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	2.64E-05
Biologica	nucleoside salvage	GO:0043174	9	2.19E-06	9.796042618	7	14	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	2.64E-05
Biologica	monoatomic cation transmembrane transport	GO:0098655	7	3.73E-06	3.776787515	16	83	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	4.31E-05
Biologica	inorganic cation transmembrane transport	GO:0098662	7	3.73E-06	3.776787515	16	83	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	4.31E-05
Biologica	purine ribonucleoside biosynthetic process	GO:0046129	10	3.93E-06	9.14297311	7	15	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	4.36E-05
Biologica	purine nucleoside biosynthetic process	GO:0042451	9	3.93E-06	9.14297311	7	15	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	4.36E-05
Biologica	ribonucleoside biosynthetic process	GO:0042455	9	3.93E-06	9.14297311	7	15	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	4.36E-05
Biologica	nucleobase-containing small molecule biosynthetic	GO:0034404	7	3.93E-06	9.14297311	7	15	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	4.36E-05
Biologica	nucleoside biosynthetic process	GO:0009163	8	3.93E-06	9.14297311	7	15	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	4.36E-05
Biologica	cytoskeleton organization	GO:0007010	6	4.52E-06	2.419739611	31	251	657	12872	ML08G00210.1,ML07G02943.1,ML09G00060.1,ML06G02661.1,M	4.57E-05
Biologica	aromatic compound biosynthetic process	GO:0019438	5	5.75E-06	1.408454318	145	2017	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	5.71E-05
Biologica	glycosyl compound biosynthetic process	GO:1901659	6	6.67E-06	8.571537291	7	16	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	6.51E-05
Biologica	regulation of systemic acquired resistance	GO:0010112	7	6.67E-06	8.571537291	7	16	657	12872	ML02G00517.1,ML09G01913.1,ML02G00614.1,ML02G01138.1,M	6.51E-05
Biologica	purine ribonucleotide salvage	GO:0106380	12	6.67E-06	8.571537291	7	16	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	6.51E-05
Biologica	purine nucleobase biosynthetic process	GO:0009113	9	6.67E-06	8.571537291	7	16	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	6.51E-05
Biologica	regulation of nucleobase-containing compound met	GO:0019219	6	6.80E-06	1.486026425	114	1503	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	6.01E-05
Biologica	stomatal complex development	GO:0010374	6	7.46E-06	7.124394631	8	22	657	12872	ML03G01419.1,ML02G02041.1,ML02G01351.1,ML09G01581.1,M	6.49E-05
Biologica	purine ribonucleoside metabolic process	GO:0046128	9	1.08E-05	8.067329215	7	17	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	9.30E-05
Biologica	ribonucleoside metabolic process	GO:0009119	8	1.08E-05	8.067329215	7	17	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	9.30E-05
Biologica	photosynthesis, dark reaction	GO:0019685	6	1.09E-05	6.814638343	8	23	657	12872	ML46G00002.1,ML170G00002.1,ML57G00001.1,ML53G00005.1,	9.09E-05
Biologica	reductive pentose-phosphate cycle	GO:0019253	7	1.09E-05	6.814638343	8	23	657	12872	ML46G00002.1,ML170G00002.1,ML57G00001.1,ML53G00005.1,	9.09E-05
Biologica	regulation of cell morphogenesis	GO:0022604	6	1.10E-05	4.435943827	12	53	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	8.86E-05
Biologica	nucleobase-containing compound biosynthetic proc	GO:0034654	6	1.22E-05	1.422888872	130	1790	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	9.77E-05
Biologica	regulation of anatomical structure morphogenesis	GO:0022603	5	1.34E-05	4.353796719	12	54	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	1.06E-04
Biologica	peptide transport	GO:0015833	7	1.44E-05	5.155811904	10	38	657	12872	ML02G00961.1,ML04G02343.1,ML04G02322.1,ML07G00438.1,M	1.12E-04
Biologica	oligopeptide transport	GO:0006857	8	1.44E-05	5.155811904	10	38	657	12872	ML02G00961.1,ML04G02343.1,ML04G02322.1,ML07G00438.1,M	1.12E-04
Biologica	regulation of biological quality	GO:0065008	3	1.50E-05	2.247120887	32	279	657	12872	ML08G00210.1,ML01G00662.1,ML09G00060.1,ML06G02661.1,M	1.13E-04
Biologica	photosynthetic electron transport in photosystem I	GO:0009773	7	1.57E-05	6.530695079	8	24	657	12872	ML87G00001.1,ML26G00001.1,ML216G00001.1,ML44G00002.1,	1.17E-04
Biologica	photorespiration	GO:0009853	5	1.57E-05	6.530695079	8	24	657	12872	ML46G00002.1,ML170G00002.1,ML57G00001.1,ML53G00005.1,	1.17E-04

Biological AMP biosynthetic process	GO:0006167	12	1.70E-05	7.619144258	7	18	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	1.23E-04
Biological purine nucleotide salvage	GO:0032261	11	1.70E-05	7.619144258	7	18	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	1.23E-04
Biological AMP metabolic process	GO:0046033	11	1.70E-05	7.619144258	7	18	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	1.23E-04
Biological regulation of cytoskeleton organization	GO:0051493	7	1.99E-05	4.198303979	12	56	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	1.39E-04
Biological monoatomic ion transmembrane transport	GO:0034220	6	2.25E-05	3.299719619	16	95	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	1.55E-04
Biological purine nucleobase metabolic process	GO:0006144	8	2.57E-05	7.218136666	7	19	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	1.75E-04
Biological purine-containing compound salvage	GO:0043101	7	2.57E-05	7.218136666	7	19	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	1.75E-04
Biological inorganic ion transmembrane transport	GO:0098660	6	2.57E-05	3.265347539	16	96	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	1.71E-04
Biological intracellular monoatomic cation homeostasis	GO:0030003	6	2.63E-05	3.417224169	15	86	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	1.73E-04
Biological actin filament organization	GO:0007015	6	2.85E-05	2.988623172	18	118	657	12872	ML08G00210.1,ML07G02943.1,ML09G00060.1,ML06G02661.1,M	1.85E-04
Biological intracellular signal transduction	GO:0035556	6	3.83E-05	2.088184633	34	319	657	12872	ML08G00210.1,ML06G02238.1,ML02G01194.1,ML09G00060.1,M	2.46E-04
Biological amide transport	GO:0042886	6	4.64E-05	4.556298892	10	43	657	12872	ML02G00961.1,ML04G02343.1,ML04G02322.1,ML07G00438.1,M	2.95E-04
Biological systemic acquired resistance	GO:0009627	8	5.42E-05	6.530695079	7	21	657	12872	ML02G00517.1,ML09G01913.1,ML02G00614.1,ML02G01138.1,M	3.41E-04
Biological positive regulation of RNA biosynthetic process	GO:1902680	9	6.62E-05	2.145118821	30	274	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,M	4.12E-04
Biological positive regulation of DNA-templated transcription	GO:0045893	10	6.62E-05	2.145118821	30	274	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,M	4.12E-04
Biological intracellular monoatomic ion homeostasis	GO:0006873	5	6.77E-05	3.160013748	15	93	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	4.12E-04
Biological purine nucleoside metabolic process	GO:0042278	8	7.60E-05	6.233845302	7	22	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	4.58E-04
Biological actin cytoskeleton organization	GO:0030036	7	7.70E-05	2.776830978	18	127	657	12872	ML08G00210.1,ML07G02943.1,ML09G00060.1,ML06G02661.1,M	4.59E-04
Biological actin filament-based process	GO:0030029	3	8.55E-05	2.755136986	18	128	657	12872	ML08G00210.1,ML07G02943.1,ML09G00060.1,ML06G02661.1,M	5.04E-04
Biological carbon fixation	GO:0015977	4	9.49E-05	5.224556063	8	30	657	12872	ML46G00002.1,ML170G00002.1,ML57G00001.1,ML53G00005.1,	5.54E-04
Biological galacturonan metabolic process	GO:0010393	6	1.05E-04	2.712750263	18	130	657	12872	ML07G00497.1,ML04G02769.1,ML03G00140.1,ML07G00493.1,M	6.05E-04
Biological pectin metabolic process	GO:0045488	7	1.05E-04	2.712750263	18	130	657	12872	ML07G00497.1,ML04G02769.1,ML03G00140.1,ML07G00493.1,M	6.05E-04
Biological methylation	GO:0032259	3	1.21E-04	1.970801473	34	338	657	12872	ML08G00508.1,ML07G02177.1,ML07G02168.1,ML08G02374.1,M	6.87E-04
Biological monoatomic cation homeostasis	GO:0055080	5	1.25E-04	2.998788557	15	98	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	7.01E-04
Biological positive regulation of transcription by RNA polymer	GO:0045944	11	1.31E-04	2.511805799	20	156	657	12872	ML01G02334.1,ML04G01617.1,ML05G01251.1,ML04G03949.1,M	7.30E-04
Biological nucleotide salvage	GO:0043173	10	1.41E-04	5.714358194	7	24	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	7.75E-04
Biological heterocycle biosynthetic process	GO:0018130	5	2.02E-04	1.334228027	133	1953	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	0.001097087
Biological monoatomic ion homeostasis	GO:0050801	4	3.37E-04	2.746554005	15	107	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	0.001818936
Biological export from cell	GO:0140352	5	3.74E-04	2.721122949	15	108	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	0.001996049
Biological supramolecular fiber organization	GO:0097435	5	4.19E-04	2.432120926	18	145	657	12872	ML08G00210.1,ML07G02943.1,ML09G00060.1,ML06G02661.1,M	0.002215496
Biological plant epidermis development	GO:0090558	5	4.56E-04	4.236126537	8	37	657	12872	ML03G01419.1,ML02G02041.1,ML02G01351.1,ML09G01581.1,M	0.002390951
Biological positive regulation of macromolecule biosynthetic p	GO:0010557	8	5.17E-04	1.908319991	30	308	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,M	0.002686041
Biological response to oxidative stress	GO:0006979	4	6.05E-04	2.144407339	22	201	657	12872	ML05G03805.1,ML08G03399.1,ML05G02681.1,ML05G03334.1,M	0.00311481
Biological positive regulation of cellular biosynthetic process	GO:0031328	7	6.07E-04	1.889911759	30	311	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,M	0.003097511
Biological positive regulation of biosynthetic process	GO:0009891	6	6.40E-04	1.88385435	30	312	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,M	0.003236592
Biological intracellular chemical homeostasis	GO:0055082	4	7.36E-04	2.555489379	15	115	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	0.003685918
Biological regulation of gene expression	GO:0010468	8	8.40E-04	1.323946483	114	1687	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	0.004172058
Biological regulation of macromolecule biosynthetic process	GO:0010556	7	9.77E-04	1.31847563	114	1694	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	0.004807579
Biological polysaccharide catabolic process	GO:0000272	6	0.00101018	2.20266047	19	169	657	12872	ML07G00497.1,ML04G02769.1,ML03G00140.1,ML07G00493.1,M	0.004928988
Biological cell morphogenesis	GO:0000902	4	0.00109263	2.798869319	12	84	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	0.005285697
Biological positive regulation of RNA metabolic process	GO:0051254	8	0.00112136	1.819698319	30	323	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,M	0.005378725
Biological positive regulation of nucleobase-containing compc	GO:0045935	7	0.00142831	1.791959015	30	328	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,M	0.006793483
Biological one-carbon metabolic process	GO:0006730	4	0.00150323	3.265347539	9	54	657	12872	ML01G00355.1,ML01G01109.1,ML09G00826.1,ML02G00623.1,M	0.007090245
Biological regulation of cellular biosynthetic process	GO:0031326	6	0.0016096	1.300056878	114	1718	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	0.007529181
Biological regulation of nitrogen compound metabolic process	GO:0051171	5	0.00167608	1.298545184	114	1720	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	0.00777592
Biological regulation of biosynthetic process	GO:0009889	5	0.00174501	1.297037002	114	1722	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	0.008029885
Biological nucleobase biosynthetic process	GO:0046112	8	0.00199619	3.809572129	7	36	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	0.009111623
Biological cell fate commitment	GO:0045165	4	0.00203637	4.353796719	6	27	657	12872	ML06G00545.1,ML08G02514.1,ML06G01498.1,ML04G01083.1,M	0.009220663
Biological anatomical structure morphogenesis	GO:0009653	3	0.00203913	2.314025815	15	127	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	0.009159906
Biological cellular nitrogen compound biosynthetic process	GO:0044271	5	0.00210983	1.216060463	171	2755	657	12872	ML09G01796.1,ML118G00001.1,ML05G00169.1,ML09G02047.1,	0.009402844
Biological regulation of response to biotic stimulus	GO:0002831	5	0.00235551	3.70661072	7	37	657	12872	ML02G00517.1,ML09G01913.1,ML02G00614.1,ML02G01138.1,M	0.010415773
Biological nucleoside metabolic process	GO:0009116	7	0.00235551	3.70661072	7	37	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	0.010415773
Biological regulation of primary metabolic process	GO:0080090	5	0.00253711	1.282882089	114	1741	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	0.010878841
Biological ribosomal small subunit assembly	GO:0000028	8	0.00358641	3.918417047	6	30	657	12872	ML04G03354.1,ML04G02952.1,ML05G03636.1,ML08G01200.1,M	0.01526248
Biological glycosyl compound metabolic process	GO:1901657	5	0.00373665	3.428614916	7	40	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	0.015783177
Biological regulation of response to external stimulus	GO:0032101	5	0.00373665	3.428614916	7	40	657	12872	ML02G00517.1,ML09G01913.1,ML02G00614.1,ML02G01138.1,M	0.015783177
Biological organic cyclic compound biosynthetic process	GO:1901362	5	0.0037477	1.240241474	133	2101	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,M	0.01559705
Biological cell wall modification	GO:0042545	7	0.00419436	2.647579086	10	74	657	12872	ML07G00497.1,ML07G00493.1,ML04G03313.1,ML04G02406.1,M	0.017202972
Biological purine nucleoside monophosphate metabolic proce	GO:0009126	9	0.00431183	3.344990162	7	41	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	0.017557516
Biological purine nucleoside monophosphate biosynthetic pro	GO:0009127	10	0.00431183	3.344990162	7	41	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	0.017557516
Biological purine ribonucleoside monophosphate metabolic pr	GO:0009167	10	0.00431183	3.344990162	7	41	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	0.017557516
Biological purine ribonucleoside monophosphate biosynthetic	GO:0009168	11	0.00431183	3.344990162	7	41	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	0.017557516
Biological defense response	GO:0006952	4	0.00445166	1.83415266	22	235	657	12872	ML02G00517.1,ML03G02487.1,ML08G00539.1,ML03G02483.1,M	0.017619844
Biological microtubule cytoskeleton organization	GO:0000226	7	0.00533495	2.234185158	13	114	657	12872	ML05G01782.1,ML06G02725.1,ML05G04385.1,ML04G00687.1,M	0.020969325
Biological nucleic acid metabolic process	GO:0090304	6	0.00551456	1.18495086	176	2910	657	12872	ML09G01796.1,ML05G03836.1,ML09G02003.1,ML05G00169.1,M	0.021525809
Biological defense response to symbiont	GO:0140546	7	0.00644153	3.116922651	7	44	657	12872	ML02G00517.1,ML09G01913.1,ML02G00614.1,ML02G01138.1,M	0.024971952
Biological response to stress	GO:0006950	3	0.0065046	1.360561475	65	936	657	12872	ML52G00001.1,ML02G00517.1,ML03G02487.1,ML04G02943.1,M	0.025044937
Biological regulation of organelle organization	GO:0033043	6	0.00711821	2.158450068	13	118	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,M	0.02722236
Biological nucleobase metabolic process	GO:0009112	7	0.00730069	3.047657703	7	45	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,M	0.027732841
Biological biological regulation	GO:0065007	2	0.00732752	1.16819284	188	3153	657	12872	ML09G01796.1,ML02G00517.1,ML05G03836.1,ML08G00210.1,M	0.027649164
Biological carbohydrate catabolic process	GO:0016052	5	0.00907971	1.700445134	23	265	657	12872	ML09G01138.1,ML07G00782.1,ML07G00497.1,ML04G02769.1,M	0.034033889
Biological actin filament depolymerization	GO:0030042	8	0.00908112	3.265347539	6	36	657	12872	ML07G02943.1,ML01G00476.1,ML02G00427.1,ML06G01919.1,M	0.033815206
Biological protein depolymerization	GO:0051261	7	0.01181163	3.093487143	6	38	657	12872	ML07G02943.1,ML01G00476.1,ML02G00427.1,ML06G01919.1,M	0.04369531
Biological cellular homeostasis	GO:0019725	3	0.01218234	1.908319991	15	154	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	0.044774061
Biological positive regulation of cellular metabolic process	GO:0031325	6	0.01305083	1.534628086	30	383	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,M	0.047656568
Biological xyloglucan metabolic process	GO:0010411	8	0.01432849	2.487883839	8	63	657	12872	ML09G00491.1,ML09G01783.1,ML01G00742.1,ML05G00776.1,M	0.051986693
Biological chemical homeostasis	GO:0048878	3	0.0151413	1.860008092	15	158	657	12872	ML01G00662.1,ML01G01781.1,ML04G01243.1,ML09G00428.1,M	0.054585838
Biological metabolic process	GO:0008152	2	0.01519158	1.05422425	486	9032	657	12872	ML07G03763.1,ML09G00491.1,ML09G01796.1,ML52G00001.1,M	0.054420463
Biological cell wall organization	GO:0071555	6	0.0191449	1.5549274	25	315	657	12872	ML09G00491.1,ML07G00497.1,ML09G01561.1,ML06G01066.1,M	0.067725094

Biologica regulation of cellular metabolic process	GO:0031323	5	0.02180652	1.193373836	115	1888	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,N	0.076661416
Biologica external encapsulating structure organization	GO:0045229	5	0.02352347	1.525863336	25	321	657	12872	ML09G00491.1,ML07G00497.1,ML09G01561.1,ML06G01066.1,N	0.082186928
Biologica regulation of developmental process	GO:0050793	4	0.02449132	1.896008249	12	124	657	12872	ML08G00210.1,ML09G00060.1,ML06G02661.1,ML05G03885.1,N	0.085043489
Biologica ribonucleoside monophosphate biosynthetic proces	GO:0009156	10	0.02745682	2.364562011	7	58	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,N	0.09361784
Biologica ribonucleoside monophosphate metabolic process	GO:0009161	9	0.02745682	2.364562011	7	58	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,N	0.09361784
Biologica positive regulation of cellular process	GO:0048522	5	0.02801098	1.432440194	31	424	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,N	0.094370342
Biologica regulation of macromolecule metabolic process	GO:0060255	5	0.02848764	1.182370417	114	1889	657	12872	ML09G01796.1,ML05G00169.1,ML09G02047.1,ML03G03611.1,N	0.095408322
Biologica tissue development	GO:0009888	4	0.029797	2.17689836	8	72	657	12872	ML03G01419.1,ML02G02041.1,ML02G01351.1,ML09G01581.1,N	0.099206497
Biologica positive regulation of nitrogen compound metabolic	GO:0051173	6	0.03002825	1.433567212	30	410	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,N	0.099391757
Biologica positive regulation of macromolecule metabolic pro	GO:0010604	6	0.03088574	1.430079214	30	411	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,N	0.101635627
Biologica nucleoside monophosphate metabolic process	GO:0009123	8	0.03499748	2.248272076	7	61	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,N	0.114500442
Biologica nucleoside monophosphate biosynthetic process	GO:0009124	9	0.03499748	2.248272076	7	61	657	12872	ML04G02338.1,ML06G01053.1,ML07G00436.1,ML01G02929.1,N	0.114500442
Biologica regulation of biological process	GO:0050789	3	0.03888554	1.127746051	169	2936	657	12872	ML09G01796.1,ML02G00517.1,ML05G03836.1,ML08G00210.1,N	0.124345861
Biologica microtubule-based process	GO:0007017	3	0.04056575	1.586025948	17	210	657	12872	ML05G01782.1,ML03G00357.1,ML01G02382.1,ML06G02725.1,N	0.128989984
Biologica positive regulation of metabolic process	GO:0009893	5	0.04850985	1.373277003	30	428	657	12872	ML05G01755.1,ML02G00606.1,ML01G02334.1,ML05G01039.1,N	0.152536535

Supplementary Table S11. A total of 88 upregulated and 1,470 down-regulated DEGs shared by MLA varieties under cold treatment.

Upregulated DEGs shared by yellow and red bracts under cold treatment	Downregulated DEGs shared by yellow and red bracts under cold treatment
ML1h06G00605.1	ML1h15G00052.1
ML1h08G00589.1	ML1h61G00011.1
ML1h05G03655.1	ML1h05G04925.1
ML1h08G00807.1	ML1h03G00680.1
ML1h03G02107.1	ML1h05G05059.1
ML1h03G00560.1	ML1h82G00001.1
ML1h07G00481.1	ML1h18G00030.1
ML1h02G02974.1	ML1h12G00119.1
ML1h09G01055.1	ML1h28G00022.1
ML1h02G02068.1	ML1h15G00150.1
ML1h01G02100.1	ML1h06G00244.1
ML1h06G01645.1	ML1h186G00001.1
ML1h08G01223.1	ML1h180G00010.1
ML1h08G03089.1	ML1h05G05002.1
ML1h03G03135.1	ML1h13G00005.1
ML1h04G04182.1	ML1h06G00111.1
ML1h08G01729.1	ML1h05G05024.1
ML1h06G03716.1	ML1h16G00005.1
ML1h08G02169.1	ML1h11G00011.1
ML1h02G01942.1	ML1h06G00342.1
ML1h07G00515.1	ML1h185G00017.1
ML1h02G01080.1	ML1h12G00052.1
ML1h05G04192.1	ML1h67G00002.1
ML1h04G00769.1	ML1h15G00119.1
ML1h02G00119.1	ML1h15G00074.1
ML1h08G03395.1	ML1h05G04947.1
ML1h03G02777.1	ML1h05G05037.1
ML1h04G00030.1	ML1h189G00014.1
ML1h07G00779.1	ML1h06G00168.1
ML1h03G02449.1	ML1h06G00266.1
ML1h09G00614.1	ML1h09G01723.1
ML1h01G00141.1	ML1h05G04893.1
ML1h08G02733.1	ML1h06G02325.1
ML1h02G00037.1	ML1h06G00364.1
ML1h06G00722.1	ML1h06G00133.1
ML1h08G01821.1	ML1h04G02958.1
ML1h07G00642.1	ML1h05G05100.1
ML1h02G00112.1	ML1h07G01555.1
ML1h01G02504.1	ML1h06G00231.1
ML1h04G02467.1	ML1h12G00074.1
ML1h02G00897.1	ML1h03G00168.1
ML1h08G02909.1	ML1h06G00004.1
ML1h05G00616.1	ML1h237G00003.1
ML1h02G02679.1	ML1h18G00065.1
ML1h09G00002.1	ML1h18G00163.1
ML1h03G02070.1	ML1h08G02372.1
ML1h09G02708.1	ML1h212G00003.1
ML1h07G02305.1	ML1h11G00024.1
ML1h02G02459.1	ML1h15G00194.1
ML1h03G01538.1	ML1h06G00288.1
ML1h02G01132.1	ML1h185G00004.1
ML1h01G00338.1	ML1h07G01542.1
ML1h06G01489.1	ML1h12G00008.1
ML1h08G02170.1	ML1h18G00078.1
ML1h02G00658.1	ML1h05G04871.1
ML1h01G01941.1	ML1h14G00002.1
ML1h03G01081.1	ML1h122G00003.1
ML1h06G02398.1	ML1h09G01691.1
ML1h03G02450.1	ML1h06G02547.1
ML1h04G02262.1	ML1h15G00106.1
ML1h06G00447.1	ML1h06G00155.1
ML1h09G02472.1	ML1h05G05068.1
ML1h04G02556.1	ML1h03G03902.1
ML1h05G00951.1	ML1h05G04934.1
ML1h08G02782.1	ML1h18G00043.1
ML1h07G02527.1	ML1h12G00061.1
ML1h05G01119.1	ML1h12G00128.1
ML1h05G02575.1	ML1h15G00008.1
ML1h04G03741.1	ML1h06G00253.1
ML1h05G02463.1	ML1h12G00096.1
ML1h01G01298.1	ML1h186G00014.1
ML1h05G01196.1	ML1h114G00001.1
ML1h06G00685.1	ML1h53G00004.1
ML1h05G01418.1	ML1h18G00141.1
ML1h08G03698.1	ML1h61G00055.1
ML1h02G00185.1	ML1h07G01479.1
ML1h03G00417.1	ML1h28G00031.1
ML1h01G02468.1	ML1h61G00033.1
ML1h08G02978.1	ML1h09G01407.1
ML1h01G01525.1	ML1h04G02593.1
ML1h01G02258.1	ML1h38G00001.1
ML1h06G02652.1	ML1h11G00002.1
ML1h05G03672.1	ML1h06G00120.1
ML1h09G02614.1	ML1h06G00373.1
ML1h09G00767.1	ML1h06G00017.1
ML1h01G02805.1	ML1h08G01770.1
ML1h05G04842.1	ML1h18G00154.1
ML1h02G00528.1	ML1h18G00056.1
	ML1h14G00024.1
	ML1h17G00024.1

Further lines deleted in pdf, please see xls file for all data

Supplementary Table S12. KEGG and GO enrichment results of MLA varieties under cold treatment.

A,B: KEGG Terms and Classes
C,D: GO Terms

Table S12A. KEGG of 38 genes among 88 up-regulated DEGs (Fig. 2D, bottom left).

Term Name	MainClass	GeneHitsInSelectedSet	AllGeneInSelectedSet	GeneHitsInBackground	AllGeneInBackground	p-value	enrichFactor	GeneListInSelectedSets	corrected p-value(BH method)
03400 DNA repair and recombination proteins	A09180 Brite Hierarchies	5	33	407	14311	0.00223	5.327600328	[ML1h02G00119.1, ML1h08G03395.1, ML1h09G01055.1, ML1h09G02708.1, ML1h08G02782.1]	0.06701056
A09150 Organismal Systems	A09150 Organismal Systems	5	33	4.41E+02	14311	0.00316	4.916855631	[ML1h01G01941.1, ML1h09G02614.1, ML1h03G02777.1, ML1h01G02258.1, ML1h03G02107.1]	0.06322914
B 09159 Environmental adaptation	A09150 Organismal Systems	5	33	4.41E+02	14311	0.00316	4.916855631	[ML1h01G01941.1, ML1h09G02614.1, ML1h03G02777.1, ML1h01G02258.1, ML1h03G02107.1]	0.06322914
B 09182 Protein famil	A09180 Brite Hierarchie	23	33	7310	14311	0.023508	1.364477884	[ML1h02G02459.1, ML1h06G01645.1, ML1h09G01055.1, ML1h01G02258.1, ML1h03G02107.	0.23507519

Table S12B. KEGG of 410 genes among 1,470 down-regulated DEGs (Fig. 2D, bottom right).

Term Name	MainClass	GeneHitsInSelectedSet	AllGeneInSelectedSet	GeneHitsInBackground	AllGeneInBackground	p-value	enrichFactor	GeneListInSelectedSets	corrected p-value(BH method)
00194 Photosynthesis proteins	A09180 Brite Hierarchies	49	184	185	14311	0	20.60044066	[ML1h08G01679.1, ML1h84G00001.1, ML1h05G02619.1, ML1h09G02622.1, ML1h57G00004.1, ML1h103G00003.1, ML1h57G00002.1, ML1h122G00005.1, ML1h02G00705.1, ML1h170G00004.1, ML1h08G03573.1, ML1h02G02129.1, ML1h64G00001.1, ML1h217G00001.1, ML1h50G00003.1, ML1h08G02615.1, ML1h101G00002.1, ML1h99G00003.1, ML1h53G00004.1, ML1h53G00006.1, ML1h05G02621.1, ML1h67G00002.1, ML1h19G00002.1, ML1h02G01825.1, ML1h131G00001.1, ML1h07G03107.1, ML1h57G00005.1, ML1h03G01583.1, ML1h50G00002.1, ML1h239G00003.1, ML1h122G00006.1, ML1h170G00003.1, ML1h05G02658.1, ML1h05G02656.1, ML1h122G00002.1, ML1h187G00001.1, ML1h60G00001.1, ML1h46G00003.1, ML1h07G01168.1, ML1h50G00004.1, ML1h03G02240.1, ML1h08G02907.1, ML1h35G00002.1, ML1h02G00392.1,	0
B 09102 Energy metabolism	A09100 Metabolism	72	184	6.42E+02	14311	0	8.72267371	[ML1h08G01679.1, ML1h84G00001.1, ML1h54G00001.1, ML1h05G02619.1, ML1h09G02622.1, ML1h57G00004.1, ML1h103G00003.1, ML1h57G00002.1, ML1h50G00001.1, ML1h07G01565.1, ML1h170G00002.1, ML1h122G00005.1, ML1h122G00001.1, ML1h02G00705.1, ML1h76G00002.1, ML1h170G00004.1, ML1h122G00003.1, ML1h06G02547.1, ML1h08G03573.1, ML1h46G00002.1, ML1h02G02129.1, ML1h232G00001.1, ML1h64G00001.1, ML1h217G00001.1, ML1h50G00003.1, ML1h08G02615.1, ML1h109G00001.1, ML1h101G00002.1, ML1h99G00003.1, ML1h53G00004.1, ML1h02G01821.1, ML1h53G00006.1, ML1h05G02621.1, ML1h67G00002.1, ML1h19G00002.1, ML1h02G01825.1, ML1h131G00001.1, ML1h07G03107.1, ML1h57G00005.1, ML1h57G00003.1, ML1h03G01583.1, ML1h57G00001.1, ML1h50G00002.1, ML1h239G00003.1, ML1h76G00003.1, ML1h122G00006.1, ML1h170G00003.1, ML1h05G02658.1, ML1h05G02656.1, ML1h05G02861.1, ML1h106G00001.1, ML1h122G00002.1, ML1h06G03265.1, ML1h06G02546.1, ML1h187G00001.1, ML1h60G00001.1, ML1h46G00003.1, ML1h07G01168.1, ML1h50G00004.1, ML1h03G02240.1, ML1h08G02907.1, ML1h35G00002.1, ML1h07G00253.1, ML1h02G00392.1,	0
00195 Photosynthesis	A09100 Metabolism	46	184	1.56E+02	14311	1.11E-16	22.93429487	[ML1h08G01679.1, ML1h84G00001.1, ML1h57G00005.1, ML1h05G02619.1, ML1h09G02622.1, ML1h57G00004.1, ML1h103G00003.1, ML1h03G01583.1, ML1h57G00002.1, ML1h50G00002.1, ML1h239G00003.1, ML1h122G00005.1, ML1h122G00006.1, ML1h170G00003.1, ML1h05G02658.1, ML1h05G02656.1, ML1h02G00705.1, ML1h122G00002.1, ML1h170G00004.1, ML1h187G00001.1, ML1h60G00001.1, ML1h46G00003.1, ML1h02G02129.1, ML1h64G00001.1, ML1h217G00001.1, ML1h07G01168.1, ML1h50G00003.1, ML1h50G00004.1, ML1h08G02615.1, ML1h03G02240.1, ML1h08G02907.1, ML1h35G00002.1, ML1h101G00002.1, ML1h99G00003.1, ML1h53G00004.1, ML1h19G00001.1, ML1h53G00006.1, ML1h05G04513.1, ML1h05G02621.1, ML1h67G00002.1,	3.33E-15
A09100 Metabolism	A09100 Metabolism	97	184	3.94E+03	14311	3.65E-13	1.913847253	[ML1h08G01679.1, ML1h54G00001.1, ML1h06G00647.1, ML1h57G00002.1, ML1h50G00001.1, ML1h122G00005.1, ML1h122G00001.1, ML1h76G00002.1, ML1h170G00004.1, ML1h05G02482.1, ML1h08G03573.1, ML1h09G00139.1, ML1h64G00001.1, ML1h07G01017.1, ML1h53G00006.1, ML1h05G02621.1, ML1h67G00002.1, ML1h19G00002.1, ML1h131G00001.1, ML1h08G02747.1, ML1h03G00225.1, ML1h07G03107.1, ML1h08G00774.1, ML1h57G00005.1, ML1h57G00001.1, ML1h50G00002.1, ML1h08G02369.1, ML1h03G02969.1, ML1h76G00003.1, ML1h122G00006.1, ML1h170G00003.1, ML1h05G02656.1, ML1h05G02861.1, ML1h08G02743.1, ML1h122G00002.1, ML1h06G03265.1, ML1h187G00001.1, ML1h46G00003.1, ML1h53G00005.1, ML1h19G00001.1, ML1h08G02372.1, ML1h09G00340.1, ML1h84G00001.1, ML1h05G02619.1, ML1h09G02622.1, ML1h57G00004.1, ML1h103G00003.1, ML1h02G01581.1, ML1h07G01565.1, ML1h170G00002.1, ML1h07G01964.1, ML1h02G00705.1, ML1h122G00003.1, ML1h07G00617.1, ML1h06G02547.1, ML1h46G00002.1, ML1h02G02129.1, ML1h232G00001.1, ML1h217G00001.1, ML1h01G00152.1, ML1h01G02201.1, ML1h50G00003.1, ML1h06G02927.1, ML1h08G02615.1, ML1h109G00001.1, ML1h101G00002.1, ML1h99G00003.1, ML1h53G00004.1, ML1h02G01821.1, ML1h04G01557.1, ML1h08G02371.1, ML1h02G01825.1, ML1h57G00003.1, ML1h03G01583.1, ML1h239G00003.1, ML1h04G04188.1, ML1h05G02658.1, ML1h106G00001.1, ML1h09G01117.1, ML1h06G02546.1, ML1h07G00616.1, ML1h60G00001.1, ML1h07G01168.1, ML1h50G00004.1, ML1h03G02240.1, ML1h08G02907.1, ML1h35G00002.1, ML1h07G00253.1,	8.22E-12
00190 Oxidative phosphorylation	A09100 Metabolism	22	184	2.55E+02	14311	1.66E-12	6.710187553	[ML1h06G02546.1, ML1h06G02547.1, ML1h54G00001.1, ML1h60G00001.1, ML1h46G00003.1, ML1h57G00003.1, ML1h57G00002.1, ML1h07G01168.1, ML1h50G00002.1, ML1h08G02615.1, ML1h109G00001.1, ML1h07G01565.1, ML1h99G00003.1, ML1h02G01820.1, ML1h76G00003.1, ML1h19G00001.1, ML1h02G01821.1, ML1h05G02861.1, ML1h122G00002.1, ML1h76G00002.1, ML1h05G02865.1, ML1h122G00003.1]	2.98E-11
03020 RNA polymerase	A09120 Genetic Information Processing	9	184	6.70E+01	14311	1.78E-07	10.44768008	[ML1h222G00001.1, ML1h57G00006.1, ML1h151G00001.1, ML1h181G00001.1, ML1h21G00001.1, ML1h225G00001.1, ML1h218G00002.1, ML1h53G00003.1, ML1h207G00004.1]	2.67E-06

B 09181 Protein families: metabolism	A09180 Brite Hierarchies	64	184	2.72E+03	14311	2.79E-07	1.829378585		[ML1h08G01679.1, ML1h01G00013.1, ML1h06G02034.1, ML1h84G00001.1, ML1h05G04460.1, ML1h04G04235.1, ML1h06G01948.1, ML1h06G02241.1, ML1h05G02619.1, ML1h09G02622.1, ML1h57G00004.1, ML1h103G00003.1, ML1h57G00002.1, ML1h01G00199.1, ML1h122G00005.1, ML1h02G00705.1, ML1h05G02655.1, ML1h170G00004.1, ML1h08G03573.1, ML1h02G02129.1, ML1h64G00001.1, ML1h217G00001.1, ML1h01G02201.1, ML1h08G01353.1, ML1h50G00003.1, ML1h161G00002.1, ML1h08G02615.1, ML1h101G00002.1, ML1h99G00003.1, ML1h03G02816.1, ML1h53G00004.1, ML1h53G00006.1, ML1h05G02621.1, ML1h07G03630.1, ML1h67G00002.1, ML1h19G00002.1, ML1h02G01825.1, ML1h131G00001.1, ML1h07G03107.1, ML1h08G00774.1, ML1h57G00005.1, ML1h03G01583.1, ML1h50G00002.1, ML1h239G00003.1, ML1h122G00006.1, ML1h170G00003.1, ML1h05G02658.1, ML1h05G02656.1, ML1h122G00002.1, ML1h187G00001.1, ML1h60G00001.1, ML1h46G00003.1, ML1h07G01168.1, ML1h50G00004.1, ML1h03G02240.1, ML1h05G04804.1]	3.59E-06
00630 Glyoxylate and dicarboxylate metabolism	A09100 Metabolism	9	184	1.24E+02	14311	3.15E-05	5.645117461		[ML1h08G02369.1, ML1h46G00002.1, ML1h170G00002.1, ML1h53G00005.1, ML1h232G00001.1, ML1h57G00001.1, ML1h122G00001.1, ML1h106G00001.1, ML1h50G00001.1]	3.54E-04
00710 Carbon fixation in photosynthetic	A09100 Metabolism	8	184	1.25E+02	14311	2.10E-04	4.97773913		[ML1h46G00002.1, ML1h170G00002.1, ML1h53G00005.1, ML1h232G00001.1, ML1h57G00001.1, ML1h122G00001.1, ML1h106G00001.1, ML1h50G00001.1]	0.00209648
03021 Transcription machinery	A09180 Brite Hierarchies	9	184	3.36E+02	14311	0.02984	2.083317158		[ML1h222G00001.1, ML1h57G00006.1, ML1h151G00001.1, ML1h181G00001.1, ML1h21G00001.1, ML1h225G00001.1, ML1h218G00002.1, ML1h53G00003.1, ML1h207G00004.1]	0.2441493
03011 Ribosome	A09180 Brite Hierarchies	13	184	5.63E+02	14311	0.03002	1.795920534		[ML1h136G00002.1, ML1h26G00002.1, ML1h225G00001.1, ML1h118G00001.1, ML1h110G00003.1, ML1h07G01567.1, ML1h01G01682.1, ML1h36G00001.1, ML1h03G02693.1, ML1h101G00001.1, ML1h09G01610.1, ML1h230G00001.1, ML1h122G00004.1]	0.22515778
03010 Ribosome	A09120 Genetic Information Processing	12	184	5.33E+02	14311	0.04268	1.751080839		[ML1h07G01567.1, ML1h01G01682.1, ML1h36G00001.1, ML1h136G00002.1, ML1h101G00001.1, ML1h26G00002.1, ML1h225G00001.1, ML1h118G00001.1, ML1h110G00003.1, ML1h09G01610.1, ML1h230G00001.1, ML1h122G00004.1]	0.27439648

Table S12C. GO-terms of up-regulated DEGs (Fig. S5C).

Class	GO_Name	GO_ID	GO_Level	P_value	EnrichmentScore	HitsGenesCountsInSelectedSet	HitsGenesCountsInBackground	AllGenesCountsInSelectedSet	AllGenesCountsInBackground	GenesOfSelectedSetInGOterm	corrected p-value(BH method)
Biological process	root morphogenesis	GO:0010015	5	4.38E-03	4.55292	5	426	33	12801	ML1h01G01298,ML1h09G01055,ML1h01G02100,ML1h05G01119,ML1h01G01525,	2.59E-01
Biological process	root development	GO:0048364	5	1.31E-02	3.00705	6	774	33	12801	ML1h01G01298,ML1h07G00779,ML1h09G01055,ML1h01G02100,ML1h05G01119,ML1h01G01525,	2.77E-01
Biological process	root system development	GO:0022622	5	1.32E-02	2.9993	6	776	33	12801	ML1h01G01298,ML1h07G00779,ML1h09G01055,ML1h01G02100,ML1h05G01119,ML1h01G01525,	2.70E-01
Biological process	plant organ morphogenesis	GO:1905392	4	2.80E-02	2.8734	5	675	33	12801	ML1h01G01298,ML1h09G01055,ML1h01G02100,ML1h05G01119,ML1h01G01525,	3.10E-01

Table S12D. GO-terms of down-regulated DEGs (Fig. S5B bottom right).

Class	GO_Name	GO_ID	GO_Level	P_value	EnrichmentScore	HitsGenesCountsInSelectedSet	HitsGenesCountsInBackground	AllGenesCountsInSelectedSet	AllGenesCountsInBackground	GenesOfSelectedSetInGOterm	corrected p-value(BH method)
Molecular function	structural constituent of ribosome	GO:0003735	3	3.05E-03	3.27278	8	292	92	10990	ML1h122G00004,ML1h101G00001,ML1h136G00002,ML1h09G01610,ML1h110G00003,ML1h225G00001,ML1h36G00001,ML1h07G01567,	0.739839
Molecular function	calcium ion binding	GO:0005509	7	4.49E-03	4.6301	5	129	92	10990	ML1h07G01479,ML1h02G01902,ML1h01G00199,ML1h04G03146,ML1h03G02816,	0.363662
Molecular function	structural molecule activity	GO:0005198	2	4.84E-03	2.79249	9	385	92	10990	ML1h122G00004,ML1h02G00392,ML1h101G00001,ML1h136G00002,ML1h09G01610,ML1h110G00003,ML1h225G00001,ML1h36G00001,ML1h07G01567,	0.294012
Molecular function	oxidoreductase activity	GO:0016491	3	7.33E-03	2.0596	14	812	92	10990	ML1h01G02201,ML1h08G02743,ML1h08G02747,ML1h02G01825,ML1h109G00001,ML1h09G00139,ML1h08G00774,ML1h03G00225,ML1h06G02546,ML1h09G01117,ML1h01G01706,ML1h04G03734,ML1h57G00003,ML1h02G01152,	0.296988
Molecular function	metal ion binding	GO:0046872	6	3.99E-02	1.9583	9	549	92	10990	ML1h122G00002,ML1h07G01479,ML1h99G00003,ML1h46G00003,ML1h02G01902,ML1h01G00199,ML1h04G03146,ML1h50G00002,ML1h03G02816,	0.387766
Cellular component	chloroplast thylakoid	GO:0009534	8	0.00E+00	6.34684	31	413	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G026	0
Cellular component	plastid	GO:0009536	6	0.00E+00	3.04774	82	2275	151	12768	ML1h02G00705,ML1h01G02201,ML1h04G04188,ML1h08G01679,ML1h02G01825,ML1h35G00002,ML1h26G00002,ML1h21G00001,ML1h109G00001,ML1h230G00001,ML1h19G00003,ML1h118G00001,ML1h122G00001,ML1h122G00002,ML1h122G00003,ML1h122G00004,ML1h02G00392,ML1h181G00001,ML1h101G00001,ML1h07G01481,ML1h02G02129,ML1h218G00002,ML1h02G02585,ML1h05G04513,ML1h136G00002,ML1h103G00003,ML1h09G01610,ML1h08G02907,ML1h01G01682,ML1h151G00002,ML1h151G00001,ML1h99G00003,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02693,ML1h03G02240,ML1h110G00003,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h67G00002,ML1h225G00001,ML1h06G03265,ML1h02G02190,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h05G02032,ML1h237G00003,ML1h04G01557,ML1h05G03347,ML1h04G03734,ML1h02G01581,ML1h84G00001,ML1h57G00006,ML1h57G00004,ML1h57G00003,ML1h57G00001,ML1h222G00001,ML1h09G02622,ML1h239G00003,ML1h36G00001,ML1h05G04635,ML1h07G01567,ML1h07G01565,ML1h02G01152,ML1h53G00001,ML1h53G00003,ML1h53G00004,ML1h08G03914,ML1h0	0
Cellular component	chloroplast thylakoid membrane	GO:0009535	7	1.11E-16	7.38594	29	332	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02	7.48E-15
Cellular component	thylakoid membrane	GO:0042651	5	2.22E-16	6.96629	29	352	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02	1.12E-14

Cellular component	obsolete thylakoid part	GO:0044436	0	2.22E-16	7.06476	32	383	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h05G04513,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02656,ML1h05G02620,ML1h05G02621	1.12E-14
Cellular component	photosynthetic membrane	GO:0034357	4	3.33E-16	7.18609	30	353	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h05G04513,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02656,ML1h05G02620,ML1h05G02621	1.12E-14
Cellular component	plastid thylakoid membrane	GO:0055035	6	3.33E-16	7.25483	29	338	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02656,ML1h05G02620,ML1h05G02621	1.12E-14
Cellular component	plastid thylakoid	GO:0031976	7	4.44E-16	6.31625	31	415	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02656,ML1h05G02620,ML1h05G02621	1.12E-14
Cellular component	thylakoid	GO:0009579	6	4.44E-16	5.60313	33	498	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h05G04513,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h170G00003,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02656,ML1h05G02620,ML1h05G02621	1.12E-14
Cellular component	chloroplast	GO:0009507	7	7.77E-16	3.05201	78	2161	151	12768	ML1h02G00705,ML1h01G02201,ML1h04G04188,ML1h08G01679,ML1h02G01825,ML1h26G00002,ML1h21G00001,ML1h109G00001,ML1h230G00001,ML1h19G00003,ML1h118G00001,ML1h122G00001,ML1h122G00002,ML1h122G00003,ML1h122G00004,ML1h02G00392,ML1h181G00001,ML1h101G00001,ML1h07G01481,ML1h02G02129,ML1h218G00002,ML1h02G02585,ML1h05G04513,ML1h136G00002,ML1h09G01610,ML1h08G02907,ML1h01G01682,ML1h151G00002,ML1h151G00001,ML1h99G00003,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02693,ML1h03G02240,ML1h110G00003,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h225G00001,ML1h06G03265,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h05G02032,ML1h237G00003,ML1h04G01557,ML1h05G03347,ML1h04G03734,ML1h02G01581,ML1h84G00001,ML1h57G00006,ML1h57G00004,ML1h57G00003,ML1h57G00001,ML1h222G00001,ML1h09G02622,ML1h239G00003,ML1h36G00001,ML1h05G04635,ML1h07G01567,ML1h07G01565,ML1h02G01152,ML1h53G00001,ML1h53G00003,ML1h53G00004,M	1.57E-14
Cellular component	plastoglobule	GO:0010287	3	5.77E-15	15.2011	16	89	151	12768	ML1h02G00705,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h46G00003,ML1h217G00001,ML1h03G02240,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h02G01581,ML1h57G00004,ML1h239G00003,ML1h53G00004,ML1h08G02615,ML1h05G02656.	1.06E-13
Cellular component	organelle outer membrane	GO:0031968	6	4.27E-14	5.51041	29	445	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02	7.20E-13
Cellular component	outer membrane	GO:0019867	4	4.27E-14	5.51041	29	445	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02	7.20E-13
Cellular component	plastid membrane	GO:0042170	5	4.65E-13	5.01459	29	489	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02	6.71E-12
Cellular component	obsolete chloroplast part	GO:0044434	0	4.14E-10	2.78757	42	1274	151	12768	ML1h02G00705,ML1h01G02201,ML1h04G04188,ML1h08G01679,ML1h02G01825,ML1h19G00003,ML1h122G00002,ML1h122G00004,ML1h02G00392,ML1h07G01481,ML1h02G02129,ML1h136G00002,ML1h08G02907,ML1h01G01682,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h110G00003,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h02G01581,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h36G00001,ML1h05G04635,ML1h53G00001,ML1h53G00004,ML1h08G03914,ML1h0	5.57E-09
Cellular component	plastid envelope	GO:0009526	4	6.49E-10	3.32185	33	840	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h136G00002,ML1h08G02907,ML1h01G01682,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h36G00001,ML1h05G04635,ML1h53G00004,ML1h08G03914,ML1h0	8.19E-09
Cellular component	obsolete plastid part	GO:0044435	0	7.02E-10	2.74025	42	1296	151	12768	ML1h02G00705,ML1h01G02201,ML1h04G04188,ML1h08G01679,ML1h02G01825,ML1h19G00003,ML1h122G00002,ML1h122G00004,ML1h02G00392,ML1h07G01481,ML1h02G02129,ML1h136G00002,ML1h08G02907,ML1h01G01682,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h110G00003,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h02G01581,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h36G00001,ML1h05G04635,ML1h53G00001,ML1h53G00004,ML1h08G03914,ML1h0	8.34E-09
Cellular component	chloroplast stroma	GO:0009570	4	3.48E-09	3.35985	30	755	151	12768	ML1h02G00705,ML1h04G04188,ML1h08G01679,ML1h02G01825,ML1h19G00003,ML1h122G00002,ML1h122G00004,ML1h02G00392,ML1h07G01481,ML1h136G00002,ML1h01G01682,ML1h46G00003,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h110G00003,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h02G01581,ML1h57G00004,ML1h239G00003,ML1h36G00001,ML1h53G00001,ML1h53G	3.91E-08
Cellular component	plastid stroma	GO:0009532	3	4.45E-09	3.32463	30	763	151	12768	ML1h02G00705,ML1h04G04188,ML1h08G01679,ML1h02G01825,ML1h19G00003,ML1h122G00002,ML1h122G00004,ML1h02G00392,ML1h07G01481,ML1h136G00002,ML1h01G01682,ML1h46G00003,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h110G00003,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h02G01581,ML1h57G00004,ML1h239G00003,ML1h36G00001,ML1h53G00001,ML1h53G	4.73E-08

Cellular component	obsolete cytoplasmic part	GO:0044444	0	2.04E-08	1.40115	115	6940	151	12768	ML1h02G00705,ML1h05G01969,ML1h01G02201,ML1h08G02324,ML1h06G00021,ML1h04G04188,ML1h08G01679,ML1h02G01825,ML1h35G00002,ML1h01G02256,ML1h03G01048,ML1h01G02085,ML1h26G00002,ML1h21G00001,ML1h109G00001,ML1h07G02271,ML1h230G00001,ML1h19G00003,ML1h118G00001,ML1h122G00001,ML1h122G00002,ML1h122G00003,ML1h122G00004,ML1h02G00392,ML1h181G00001,ML1h101G00001,ML1h07G01481,ML1h09G00139,ML1h02G02129,ML1h218G00002,ML1h02G02585,ML1h05G04513,ML1h136G00002,ML1h08G00774,ML1h103G00003,ML1h09G01610,ML1h07G00549,ML1h07G02959,ML1h08G02907,ML1h01G01682,ML1h151G00002,ML1h151G00001,ML1h99G00001,ML1h99G00003,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h36G00001,ML1h05G04635,ML1h07G01567,ML1h07G01565,ML1h02G01152,ML1h53G00001,ML1h06G01553,ML1h53G00003,ML1h53G00004,ML1h08G03914,ML1h03G01719,ML1h06G02546,ML1h06G02547,ML1h03G02240,ML1h06G00123,ML1h110G00003,ML1h04G02220,ML1h08G02029,ML1h131G00001,ML1h182G00008,ML1h05G05114,ML1h09G01117,ML1h67G00002,ML1h225G00001,ML1h06G03265,ML1h02G02190,ML1h08G00064,ML1h04G03146,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h05G02032,ML1h01G01706,ML1h237G00003,ML1h07G02680,ML1h04G01557,ML1h05G03347,ML1h04G03734,ML1h02G01581,ML1h84G00001,ML1h07G00275,ML1h06G01948,ML1h57G00006,ML1h09G02845,ML1h57G00004,ML1h57G00003,ML1h57G00001,ML1h222G00001,ML1h09G02622,ML1h239G00003,ML1h36G00001,ML1h05G04635,ML1h07G01567,ML1h07G01565,ML1h02G01152,ML1h53G00001,ML1h53G00003,ML1h53G00004,ML1h08G03914,ML1h03G01719,ML1h0	2.20E-07
Cellular component	organelle envelope	GO:0031967	3	2.24E-08	2.60807	38	1232	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h109G00001,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h136G00002,ML1h08G02907,ML1h01G01682,ML1h99G00003,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h06G02546,ML1h06G02547,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h36G00001,ML1h05G04635,ML1h06G01553,ML1h53G00001,ML1h53G00003,ML1h53G00004,ML1h08G03914,ML1h03G01719,ML1h0	2.16E-07
Cellular component	organelle subcompartment	GO:0031984	3	6.80E-08	2.64238	35	1120	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G01679,ML1h01G02256,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h09G00139,ML1h02G02129,ML1h07G02959,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02656,ML1	6.24E-07
Cellular component	cytoplasm	GO:0005737	3	3.64E-06	1.26637	122	8146	151	12768	ML1h02G00705,ML1h05G01969,ML1h01G02201,ML1h08G02324,ML1h06G00021,ML1h04G04188,ML1h08G01679,ML1h02G01825,ML1h35G00002,ML1h01G02256,ML1h03G01048,ML1h01G02085,ML1h26G00002,ML1h21G00001,ML1h109G00001,ML1h07G02271,ML1h230G00001,ML1h19G00003,ML1h118G00001,ML1h122G00001,ML1h122G00002,ML1h122G00003,ML1h122G00004,ML1h02G00392,ML1h181G00001,ML1h101G00001,ML1h07G01481,ML1h09G00139,ML1h02G02129,ML1h218G00002,ML1h02G02585,ML1h05G04513,ML1h136G00002,ML1h08G00774,ML1h103G00003,ML1h09G01610,ML1h07G00549,ML1h07G02959,ML1h08G02907,ML1h01G01682,ML1h151G00002,ML1h151G00001,ML1h99G00001,ML1h99G00003,ML1h46G00003,ML1h03G00225,ML1h07G00616,ML1h07G00617,ML1h217G00001,ML1h03G01583,ML1h02G01902,ML1h03G02693,ML1h06G00157,ML1h06G02546,ML1h06G02547,ML1h03G02240,ML1h06G00123,ML1h110G00003,ML1h01G00199,ML1h04G02220,ML1h08G02029,ML1h131G00001,ML1h182G00008,ML1h05G05114,ML1h09G01117,ML1h67G00002,ML1h225G00001,ML1h06G03265,ML1h02G02190,ML1h08G00064,ML1h04G03146,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h05G02032,ML1h01G01706,ML1h237G00003,ML1h07G02680,ML1h04G01557,ML1h05G03347,ML1h04G03734,ML1h02G01581,ML1h84G00001,ML1h07G00275,ML1h06G01948,ML1h57G00006,ML1h09G02845,ML1h57G00004,ML1h57G00003,ML1h57G00001,ML1h222G00001,ML1h09G02622,ML1h239G00003,ML1h36G00001,ML1h05G04635,ML1h07G01567,ML1h07G01565,ML1h02G01152,ML1h03G02816,ML1h53G00001,ML1h06G01553,ML1h53G00003,ML1h53G00004,ML1h08G03914,ML1h03G01719,ML1h08G02615,ML1h03G028	3.20E-05
Cellular component	bounding membrane of organelle	GO:0098588	5	9.67E-06	2.17139	34	1324	151	12768	ML1h02G00705,ML1h01G02201,ML1h08G02324,ML1h08G01679,ML1h07G02271,ML1h19G00003,ML1h122G00002,ML1h02G00392,ML1h02G02129,ML1h08G02907,ML1h46G00003,ML1h03G00225,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h04G02220,ML1h131G00001,ML1h09G01117,ML1h50G00001,ML1h50G00003,ML1h50G00002,ML1h07G00275,ML1h57G00004,ML1h09G02622,ML1h239G00003,ML1h05G04635,ML1h53G00004,ML1h08G03914,ML1h08G02615,ML1h05G02658,ML1h05G02656,ML1h05G02620,ML1	8.14E-05
Cellular component	small ribosomal subunit	GO:0015935	5	1.31E-05	6.28931	9	121	151	12768	ML1h230G00001,ML1h118G00001,ML1h122G00004,ML1h101G00001,ML1h136G00002,ML1h09G01610,ML1h110G00003,ML1h225G00001,ML1h36G00001,	1.05E-04
Biological process	generation of precursor metabolites and energy	GO:0006091	4	1.25E-14	6.961	25	362	127	12801	ML1h06G00021,ML1h08G01679,ML1h02G01825,ML1h109G00001,ML1h02G00392,ML1h76G00003,ML1h76G00002,ML1h07G01481,ML1h05G04513,ML1h217G00001,ML1h03G01583,ML1h06G00157,ML1h06G02546,ML1h06G02547,ML1h03G02240,ML1h06G00123,ML1h182G00008,ML1h05G05114,ML1h01G01706,ML1h57G00003,ML1h239G00003,ML1h0	1.06E-11
Biological process	regulation of cellular respiration	GO:0043457	7	2.83E-13	78.3963	7	9	127	12801	ML1h06G00021,ML1h06G00157,ML1h06G00123,ML1h182G00008,ML1h05G05114,ML1h06G00208,ML1h06G00216,	1.20E-10
Biological process	photosynthesis	GO:0015979	4	1.23E-12	9.84781	17	174	127	12801	ML1h08G01679,ML1h02G01825,ML1h109G00001,ML1h19G00003,ML1h02G00392,ML1h07G01481,ML1h05G04513,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h131G00001,ML1h50G00003,ML1h57G00004,ML1h239G00003,ML1h53G00004,ML1h08G02615,ML1h05G02656.	3.45E-10
Biological process	cellular respiration	GO:0045333	6	1.62E-12	13.5686	14	104	127	12801	ML1h06G00021,ML1h109G00001,ML1h76G00003,ML1h76G00002,ML1h06G00157,ML1h06G02546,ML1h06G02547,ML1h06G00123,ML1h182G00008,ML1h05G05114,ML1h01G01706,ML1h57G00003,ML1h06G00208,ML1h06G00216,	3.41E-10
Biological process	energy derivation by oxidation of organic compounds	GO:0015980	5	1.19E-10	9.93756	14	142	127	12801	ML1h06G00021,ML1h109G00001,ML1h76G00003,ML1h76G00002,ML1h06G00157,ML1h06G02546,ML1h06G02547,ML1h06G00123,ML1h182G00008,ML1h05G05114,ML1h01G01706,ML1h57G00003,ML1h06G00208,ML1h06G00216,	2.00E-08
Biological process	regulation of generation of precursor metabolites and	GO:0043467	6	8.15E-09	18.7526	8	43	127	12801	ML1h06G00021,ML1h03G01583,ML1h06G00157,ML1h06G00123,ML1h182G00008,ML1h05G05114,ML1h06G00208,ML1h06G00216,	1.15E-06
Biological process	photosynthesis, light reaction	GO:0019684	5	4.98E-08	8.7303	11	127	127	12801	ML1h08G01679,ML1h02G01825,ML1h02G00392,ML1h07G01481,ML1h05G04513,ML1h217G00001,ML1h03G01583,ML1h03G02240,ML1h239G00003,ML1h08G02615,ML1h05G02656	6.00E-06
Biological process	electron transport chain	GO:0022900	5	2.21E-05	6.83358	8	118	127	12801	ML1h02G01825,ML1h76G00003,ML1h76G00002,ML1h05G04513,ML1h03G01583,ML1h06G02546,ML1h06G02547,ML1h08G02615,	0.00233
Biological process	aerobic respiration	GO:0009060	7	1.25E-03	6.22193	5	81	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h01G01706,	0.075124
Biological process	obsolete oxidation-reduction process	GO:0055114	0	3.11E-03	1.98047	19	967	127	12801	ML1h01G02201,ML1h08G02743,ML1h08G02747,ML1h02G01825,ML1h109G00001,ML1h76G00003,ML1h76G00002,ML1h09G00139,ML1h05G04513,ML1h08G00774,ML1h03G00225,ML1h06G02546,ML1h06G02547,ML1h09G01117,ML1h01G01706,ML1h04G03734,ML1h57G00003,ML1h02G01152,ML1h08G02615.	0.137972
Biological process	purine nucleoside	GO:0009126	9	8.91E-03	3.93732	5	128	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h04G01557,	0.35792
Biological process	purine ribonucleoside	GO:0009167	10	8.91E-03	3.93732	5	128	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h04G01557,	0.35792
Biological process	purine ribonucleotide metabolic process	GO:0009150	10	9.53E-03	3.32292	6	182	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h04G01557,ML1h03G02613,	0.349865
Biological process	ribonucleoside	GO:0009161	9	1.28E-02	3.59983	5	140	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h04G01557,	0.372074

Biological process	ribonucleotide metabolic process	GO:0009259	9	1.64E-02	2.95011	6	205	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h04G01557,ML1h03G02613.	0.445868
Biological process	nucleoside monophosphate metabolic	GO:0009123	8	2.00E-02	3.21004	5	157	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h04G01557,	0.433519
Biological process	ribose phosphate metabolic process	GO:0019693	6	2.23E-02	2.74896	6	220	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h04G01557,ML1h03G02613,	0.471214
Biological process	purine nucleotide metabolic process	GO:0006163	9	2.41E-02	2.69987	6	224	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h04G01557,ML1h03G02613,	0.485073
Biological process	response to toxic substance	GO:0009636	4	2.54E-02	2.26506	8	356	127	12801	ML1h08G02743,ML1h08G02747,ML1h08G02371,ML1h08G02372,ML1h02G00392,ML1h05G03347.ML1h04G03734.ML1h07G01964.	0.49874
Biological process	amide metabolic process	GO:0043603	3	2.93E-02	1.84663	12	655	127	12801	ML1h08G02371,ML1h08G02372,ML1h122G00004,ML1h101G00001,ML1h136G00002,ML1h09G01610,ML1h110G00003,ML1h08G02029,ML1h225G00001,ML1h36G00001,ML1h07G01567.ML1h07G01964.	0.538156
Biological process	amide biosynthetic process	GO:0043604	4	3.42E-02	2.02491	9	448	127	12801	ML1h08G02371,ML1h08G02372,ML1h101G00001,ML1h136G00002,ML1h09G01610,ML1h110G00003,ML1h225G00001,ML1h36G00001,ML1h07G01567,	0.53482
Biological process	defense response	GO:0006952	4	3.73E-02	1.59101	17	1077	127	12801	ML1h01G02201,ML1h08G01679,ML1h122G00002,ML1h07G01479,ML1h08G00774,ML1h08G00783,ML1h46G00003,ML1h08G02029,ML1h08G00064,ML1h50G00002,ML1h05G03347,ML1h09G02625,ML1h02G01152,ML1h08G02615,ML1h03G02613,ML1h06G00429,ML1h06G00647	0.551803
Biological process	purine-containing compound metabolic	GO:0072521	5	4.02E-02	2.3904	6	253	127	12801	ML1h76G00003,ML1h76G00002,ML1h06G02546,ML1h06G02547,ML1h04G01557,ML1h03G02613,	0.530161
Biological process	translation	GO:0006412	6	4.32E-02	1.93837	9	468	127	12801	ML1h122G00004,ML1h101G00001,ML1h136G00002,ML1h09G01610,ML1h110G00003,ML1h08G02029,ML1h225G00001,ML1h36G00001,ML1h07G01567,	0.561282

Liu et al. Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa* .

Supplementary Table S13. Expression levels before row scale normalization for upregulated and down-regulated DEGs of MLA varieties under cold treatment.

Expression levels of 14 up-regulated transcription factors in heatmap of Figure 2E right

	R51115cold-1	R51115cold-2	R51115cold-3	R511Normal-1	R511Normal-2	R511Normal-3	Y51015cold-1	Y51015cold-2	Y51015cold-3	Y510Normal-1	Y510Normal-2	Y510Normal-3
ML1h01G01941_bZIP	390.8466832	419.6009192	344.4168926	59.02615217	64.5497132	127.2508725	181.7200238	193.050853	363.9149879	59.60177322	39.9822674	37.90063644
ML1h07G00779_GRF	0.475114367	0.297665979	0.263544543	0.031887822	0.095719607	0	0.277114612	0.376853103	0.225453766	0	0	0.031413381
ML1h03G01081_HD-ZIP	69.84231253	94.38401826	71.43505065	13.7396065	12.71141316	25.9776371	61.66731102	80.30398709	73.53640577	8.857636074	13.3740848	14.52141437
ML1h07G00642_bHLH	1.203962431	0.768749951	0.725406356	0.078018871	0.039032328	0.250895128	0.150668241	0.499435042	0.078801459	0	0	0
ML1h06G00685_ERF	24.45347687	20.58357913	9.365575014	2.076840372	7.157759389	2.296927224	10.75898566	8.278825731	31.39849684	0.670840102	1.487646999	7.209503596
ML1h01G02258_WRKY	0.846784166	0.275085714	0.308246327	0.033152495	0.099515852	0.142150214	1.664606628	1.991645517	0.904096059	0.087996639	0	0.130636948
ML1h04G02467_Dof	76.95903613	58.76138038	35.29727103	3.390759791	4.28805862	5.906384291	74.80359852	64.65368666	71.77750602	5.500055492	8.741429806	28.85657166
ML1h05G01119_MYB	24.27894558	11.79245012	12.35792473	1.333233873	2.038080442	2.699504537	3.361427528	6.419223525	6.77044181	0.688100297	0.504528001	0.474282402
ML1h08G01223_G2-like	14.48906512	11.04572299	8.463074149	2.067500084	2.02968108	2.718030549	3.879707205	3.841808019	5.003892647	0.138056948	0.23250332	0.960725892
ML1h09G00614_bHLH	2.197903018	0.74045403	1.712106504	0.254963631	0.076533977	0.792588094	0.590855847	0.903954828	0.103008443	0.112791625	0.06512698	0.125585084
ML1h07G02527_bHLH	3.402138084	7.150268332	2.010231389	0.972819414	1.279444496	0.059595042	2.147290845	2.409114767	0.140382647	0.14756681	0.071005425	0
ML1h01G02805_TALE	0.225937121	0.495434985	0.740210567	0.079611093	0.026552604	0.056892319	1.255568675	1.82943239	0.321638608	0	0	0.052284402
ML1h09G02614_bHLH	3.64747239	6.791587917	4.692112538	0.975235889	1.115209385	1.284344106	5.838394338	13.11703024	8.921450895	0.616325663	0.948993144	0.960725892
ML1h08G02733_bHLH	3.092417838	6.387048511	5.618181342	0.533158572	0.453450741	1.88599982	1.261288805	2.336589387	0.323103932	0.23586039	0.068094041	0.2888743

Expression levels of 34 down-regulated transcription factors in Figure 2E right

	R51115cold-1	R51115cold-2	R51115cold-3	R511Normal-1	R511Normal-2	R511Normal-3	Y51015cold-1	Y51015cold-2	Y51015cold-3	Y510Normal-1	Y510Normal-2	Y510Normal-3
ML1h04G03139_WRKY	0	0.098684204	0.131058059	0.856304683	0.85680721	4.045598941	0.520601646	0.8120895	5.029198807	2.132587009	4.077584363	22.90120582
ML1h01G02930_C2H2	0.227484635	0.2217015	0.607265594	3.420005309	1.657537236	0.458255941	0.361190989	1.210432664	0.161920806	7.422924969	1.04649244	0.368497602
ML1h07G02834_ERF	0.049228394	0	0	3.700499815	0.13885018	0	1.563257441	1.002210788	2.1491307	15.42909177	8.743490193	6.470659763
ML1h05G00785_bHLH	0.300115272	0.146242857	0.194218569	3.712946276	2.022156776	0.151141643	0.090764001	0.370294749	0.332285309	3.24350662	9.004001216	9.769309071
ML1h04G02593_bZIP	0.151425942	0.184470473	0.257236296	1.031556502	0.925386511	0.457559503	0.240428044	0.140126736	0.143710867	0.849742616	0.938897472	0.105125022
ML1h07G01922_MYB	0	0	0	0.797920273	0.684333032	0.203648643	0.623707816	0.56130312	0.19188667	4.134984901	6.373351684	1.122926367
ML1h03G00080_ERF	2.707561689	2.902602791	1.471838982	98.86347819	14.81531424	0.818136286	4.372654385	8.117907379	5.293402355	88.64156463	11.00419441	3.107739407
ML1h04G01712_NAC	1.323035639	0.311234798	0.310002716	5.787114066	1.629921408	1.562350611	0.372531365	0.464394376	0.043297505	1.744675722	1.642488134	2.069255768
ML1h04G01545_HD-ZIP	1.150282577	0.927757233	1.30911975	9.61570163	8.912796022	2.476990113	7.01758925	3.486991355	5.157229244	26.41328401	16.53333438	17.62350681
ML1h03G00680_ERF	23.43465894	14.4806085	13.94604324	189.5037317	60.50010912	11.00417226	16.55368174	34.22242143	22.08514575	251.181902	57.90731192	33.87822883
ML1h04G04188_ERF	0.401767218	0.326294546	0.585005125	12.26909667	1.322062738	0.337224634	0	0.185893936	0	3.228751214	0.428577549	0
ML1h03G03234_bHLH	0.138386486	0.269736825	0.13433451	7.606839933	0.780646569	0.209079273	1.381125542	4.354049088	7.486138606	27.32377105	8.912627278	20.23929213
ML1h06G01412_ERF	0	0	0	3.343665905	0.107923489	0	0.052074277	0	0	4.103536026	0.505109254	0
ML1h08G00293_CO-like	0.5874898	0.057255456	0.342172809	30.91313761	35.54335382	1.213054273	1.119351319	0.108730416	0.41816812	10.37869159	10.22293312	0.435045687
ML1h03G03392_ZF-HD	0.928003497	0.238003081	0.426709621	5.461320977	3.444028982	0.442756108	0.664712828	0.180790966	0.602599392	1.096334591	1.641199908	2.577005922
ML1h06G00732_ERF	4.129031576	4.551808923	3.285354871	214.4670924	39.01111524	5.113351785	4.913094815	12.96610206	7.901559341	164.4303275	16.46296802	10.71418223
ML1h06G01520_MYB	0.041654795	0.202979216	0.080870274	1.213336959	0.509117328	0.881069846	0.302344296	0.192732844	0.118597514	1.870002145	2.599416003	0.462690731
ML1h08G02035_MYB	0.334806015	0.065258909	0.390003417	2.516737778	5.162340216	0.20233478	0.037666149	1.982868655	4.766217279	26.44235908	3.589336972	2.231363362
ML1h09G00769_MYB	0.308068856	0.643365219	0.341769779	5.748033437	13.1992715	0.354622088	3.633606872	5.783074968	4.636198913	23.63554471	19.40070613	23.46479635
ML1h03G02358_ERF	0.233527196	0.37931741	0.226689486	11.33711721	10.17280061	0.078404727	1.695017711	1.584745808	2.068538299	22.52053972	2.927050729	2.305742141
ML1h01G02076_MYB	0.286316868	0.046506349	0	1.300314518	0.807565417	0.72096301	0.56284113	1.148126534	0.679322922	6.347445906	2.786985613	10.91031243
ML1h07G00856_ERF	0.769677649	0.545535152	1.358438868	106.2459233	22.69576402	1.339047029	0.507870475	5.115216295	1.062491582	35.71641841	3.806635421	1.94304113
ML1h08G00320_ERF	0	0	0	8.753717338	0.515226736	0	0	0	0	1.739517551	0	0
ML1h06G00429_WRKY	0.437009957	0.307594625	0.353511869	3.262192564	2.282592308	1.956297291	0.396495371	0.224667136	0.368661797	0.807350576	0.912920305	2.472043231
ML1h03G00168_MYB	0.038799949	0.037813574	0	0.364574164	0.43774574	0.468962855	0.10560858	0	0	0.548357038	0.310418318	0.035914987
ML1h01G01776_TALE	0	0	0	1.905111969	1.463712317	0.07293463	1.511062301	0.268033118	1.443166255	2.919663663	3.678727784	7.607608518
ML1h08G02789_ERF	6.145091552	7.111783692	4.314090661	80.32163074	13.36040494	3.150004906	5.625168909	9.545285122	2.551500985	45.24862099	9.569362585	4.774003552
ML1h07G00626_ZF-HD	0.097302998	0.474146763	0.850085573	9.417121551	7.958935726	2.646159549	6.532880761	1.800847509	2.308636494	10.59694936	6.850544259	17.56327021
ML1h05G03347_WRKY	3.885311555	1.8755753	1.797654234	25.99870836	3.652879719	3.364774306	1.218963174	2.755349191	1.068299663	7.909093403	5.46155238	4.940876016
ML1h06G00772_WOX	1.334441118	0.303453928	0.129536849	2.758524371	5.896669622	0.582435118	1.614302582	1.276029092	2.617334174	4.178687994	6.583639937	19.63998217
ML1h05G00166_CO-like	1.652767514	0.690321732	1.060032984	22.79556115	20.97756421	4.964394112	1.981537293	1.584063022	2.128759793	24.29082565	28.40570355	16.30957415
ML1h01G02176_WRKY	0	0.133876733	0	12.8214836	1.463712317	0.092240856	3.282019598	2.330508541	9.038900886	50.89581069	12.67533858	107.2339635
ML1h01G02045_G2-like	0.712900081	0.326953727	0.529196556	2.718839448	1.774196748	1.858482427	0.34242782	0.23283685	0.994967917	4.776212611	3.153724691	1.591505518
ML1h03G02929_ERF	1.494574052	1.553684112	1.982777372	50.79028509	10.30453471	1.605728817	1.491615586	6.546440864	2.36404377	43.03235084	3.387905524	10.92921775

Supplementary Table S14. Statistics of the structural variations detected in MLA haplotypes by SyRI.

Parameters to count as an inversion, translocation or duplication: 100bp									Recombination rate (nb_recomb/Mb)		
		Count	Length_MLAh1 (bp)	Length_MLAh2 (bp)	MLAh1 (total bp)	%MLAh1	MLAh2 (total bp)	%MLAh2	All arranger	MLAh1 (	MLAh2
MLA01	Syntenic regions	324	40,126,562	39,402,025	51,603,189	77.76%	49,567,952	79.49%	276		
	Inversions	15	658,764	589,952		1.28%		1.19%		0.29	0.30
	Translocations	116	1,239,843	1,134,594		**2.40%**		**2.29%**		2.25	2.34
	Duplications	145	1,963,948	2,016,080		**3.81%**		**4.07%**		2.81	2.93
	All rearrangments	**276**									
	Not aligned (Hap1)	465	9,142,269	-							
	Not aligned (Hap2)	726	-	5,927,665							
MLA02	Syntenic regions	316	48,899,063	47,633,913	58,414,654	83.71%	54,496,339	87.41%	157		
	Inversions	12	315,091	119,029		0.54%		0.22%		0.21	0.22
	Translocations	39	1,151,665	1,132,508		1.97%		2.08%		0.67	0.72
	Duplications	106	1,343,894	1,429,661		2.30%		2.62%		1.81	1.95
	All rearrangments	**157**									
	Not aligned (Hap1)	426	7,944,806								
	Not aligned (Hap2)	549	-	4,730,025							
MLA03	Syntenic regions	278	40,226,354	39,394,875	56,565,741	71.11%	54,952,116	71.69%	168		
	Inversions	16	8,045,525	8,310,779		**14.22%**		**15.12%**		0.28	0.29
	Translocations	19	334,337	283,917		0.59%		0.52%		0.28	0.35
	Duplications	133	1,764,936	1,651,135		3.12%		3.00%		2.35	2.42
	All rearrangments	**168**									
	Not aligned (Hap1)	339	7,743,164	-							
	Not aligned (Hap2)	637	-	5,020,471							
MLA04	Syntenic regions	268	39,008,395	38,076,183	49,080,785	79.48%	50,361,345	75.61%	144		
	Inversions	19	4,943,401	4,703,625		10.07%		9.34%		0.39	0.38
	Translocations	32	824,796	796,587		1.68%		1.58%		0.65	0.64
	Duplications	93	1,828,871	2,184,168		3.73%		4.34%		1.89	1.85
	All rearrangments	**144**									
	Not aligned (Hap1)	317	4,723,539	-							
	Not aligned (Hap2)	527	-	4,938,988							
MLA05	Syntenic regions	306	38,927,251	38,234,622	50,726,621	76.74%	49,514,395	77.22%	280		
	Inversions	20	2,830,932	2,980,962		5.58%		6.02%		0.39	0.40
	Translocations	60	772,470	732,045		1.52%		1.48%		1.18	1.21
	Duplications	200	1,185,373	1,231,473		2.34%		2.49%		3.94	4.04
	Not aligned (Hap1)	413	7,563,825	-							
	Not aligned (Hap2)	655	-	5,284,941							
MLA06	Syntenic regions	205	33,731,496	32,930,696	45,995,787	73.34%	40,909,305	80.50%	141		
	Inversions	13	2,574,998	2,550,331		5.60%		6.23%		0.28	0.32
	Translocations	35	744,565	737,743		1.62%		1.80%		0.76	0.86
	Duplications	93	1,688,651	1,557,113		3.67%		3.81%		2.02	2.27
	All rearrangments	**141**									
	Not aligned (Hap1)	341	7,859,039	-							
	Not aligned (Hap2)	425	-	3,499,234							
MLA07	Syntenic regions	374	49,113,873	47,305,082	59,458,675	82.60%	56,090,541	84.34%	239		
	Inversions	13	383,182	368,841		0.64%		0.66%		0.22	0.23
	Translocations	79	1,046,266	953,794		1.76%		1.70%		1.33	1.41
	Duplications	147	1,561,685	1,656,873		2.63%		2.95%		2.47	2.62
	All rearrangments	**239**									
	Not aligned (Hap1)	501	8,490,131	-							
	Not aligned (Hap2)	781	-	5,028,965							
MLA08	Syntenic regions	257	43,010,704	42,149,693	51,979,136	82.75%	53,045,455	79.46%	180		
	Inversions	20	3,435,468	3,557,873		6.61%		6.71%		0.38	0.38
	Translocations	52	194,504	199,850		0.37%		0.38%		1.00	0.98
	Duplications	108	864,538	806,643		1.66%		1.52%		2.08	2.04
	All rearrangments	**437**									
	Not aligned (Hap1)	327	5,010,742	-							
	Not aligned (Hap2)	533	-	4,229,182							
MLA09	Syntenic regions	280	44,239,560	43,478,921	50,644,119	87.35%	53,318,599	81.55%	122		
	Inversions	12	129,069	96,619		0.25%		0.18%		0.24	0.23
	Translocations	27	1,103,879	1,078,237		2.18%		2.02%		0.53	0.51
	Duplications	83	1,330,723	1,351,551		2.63%		2.53%		1.64	1.56
	All rearrangments	**122**									
	Not aligned (Hap1)	323	5,461,536	-							

	Not aligned (Hap2)	635	-	6,787,066							
Total	Syntenic regions	2,608	377,283,258	368,606,010	474,468,707	79.52%	462,256,047	79.74%	1,707		
	Inversions	140	23,316,430	23,278,011		4.91%		5.04%		0.30	0.30
	Translocations	459	7,412,325	7,049,275		1.56%		1.52%		0.97	0.99
	Duplications	1,108	13,532,619	13,884,697		2.85%		3.00%		2.34	2.40
Average		569				3.11%		3.19%		1.20	1.23

Liu et al. Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa*.

Supplementary Table S15. Repetitive DNA composition of MLA haplotypes.

Table S15A. Repetitive DNA proportion of MLA haplotypes.

MLAh1					MLAh2				
Class		**Count**	**Masked base pair**	**%masked**	**Class**		**Count**	**Masked base pairs**	**%masked**
DNA		--		--	DNA		--	--	
	DTA	33077	7775821	1.56%		DTA	34533	19900372	3.99%
	DTC	28667	6207894	1.24%		DTC	26154	6675021	1.34%
	DTH	4718	1044242	0.21%		DTH	4682	1294918	0.26%
	DTM	63659	16667694	3.33%		DTM	57803	13392454	2.69%
	DTT	5318	1233437	0.25%		DTT	6054	723776	0.15%
	Helitron	28541	9080566	1.82%		Helitron	26731	5876820	1.18%
LINE		1754	2371677	0.47%	LINE		1902	2171524	0.44%
LTR	--	--	--		LTR	--	--	--	
	Copia	44483	21845750	4.37%		Copia	50518	31417387	6.31%
	Gypsy	39706	23149617	4.63%		Gypsy	41326	18669128	3.75%
	mixture	4518	1019359	0.20%		unknown	65622	35530364	7.13%
	unknown	71322	42977680	8.60%	LTR/Copia	--	--	--	
LTR/Copia		--		--		Copia/Ale	4111	4687625	0.94%
	Copia/Ale	4531	5078242	1.02%		Copia/Alesia	108	55745	0.01%
	Copia/Alesia	67	6857	0.00%		Copia/Angela	7219	11482809	2.30%
	Copia/Angela	12767	15042410	3.01%		Copia/Gymco-	78	6438	0.00%
	Copia/Ikeros	4941	7402171	1.48%		Copia/Ikeros	4858	7705641	1.55%
	Copia/Ivana	3427	3360557	0.67%		Copia/Ivana	1929	2696250	0.54%
	Copia/SIRE	14892	13367978	2.67%		Copia/SIRE	18539	15405108	3.09%
	Copia/TAR	95	193131	0.04%		Copia/TAR	303	267971	0.05%
	Copia/Tork	1270	2303444	0.46%		Copia/Tork	1210	2321117	0.47%
	Copia/mixture	555	227865	0.05%	LTR/Gypsy	--	--	--	
LTR/Gypsy		--		--		Gypsy/Athila	351	104410	0.02%
	Gypsy/CRM	4251	3537511	0.71%		Gypsy/CRM	6284	4274750	0.86%
	Gypsy/Galadriel	6299	4481931	0.90%		Gypsy/Galadri	2652	4433139	0.89%

Table S15B. Statistics of repetitive DNA proportion of MLA haplotypes.

Class		MLAh1	MLAh2	
Transposable elements		43.85%%	43.40%	43.40%
	Class I Retrotransposons	35.91%	33.79%	34.85%
	LTR retroelements	35.44%	33.35%	34.40%
	Copia	13.77%	15.26%	14.52%
	Gypsy	12.87%	10.96%	11.92%
	unknown and mixture	8.80%	7.13%	8.07%
	LINEs	0.47%	0.44%	0.46%
	Class II DNA transposons	8.41%	9.61%	9.01%
Simple repeats and low complexity		1.40%	1.17%	1.29%
MITE		3.80%	3.58%	3.69%
Unknown interspersed repeats		4.71%	3.24%	3.98%
Total repeats (RepeatMasker)		**54.23%**	**51.39%**	**52.81%**

	Gypsy/Reina	23651	21443921	4.29%		Gypsy/Reina	15935	16264874	3.26%
	Gypsy/Retand	9097	9319113	1.86%		Gypsy/Retand	6823	8477480	1.70%
	Gypsy/Tekay	3890	2341439	0.47%		Gypsy/Tekay	4016	2355591	0.47%
	Gypsy/mixture	287	57895	0.01%		Gypsy/chromc	123	50722	0.01%
MITE		--		--	MITE	--	--	--	
	DTA	16415	16073665	3.21%		DTA	14693	15037851	3.02%
	DTC	2572	55925	0.01%		DTC	489	59495	0.01%
	DTH	222	60374	0.01%		DTH	157	28826	0.01%
	DTM	7329	2858863	0.57%		DTM	7385	2667682	0.54%
	DTT	11	3036	0.00%		DTT	12	3107	0.00%
Unknown		65444	23552725	4.71%	Unknown		54416	16134576	3.24%
total interspersed		507776	264142790	52.83%	total interspersed		467016	250172971	50.22%
Low_complexity		14621	749387	0.15%	Low_complexity		14369	730303	0.15%
Simple_repeat		109761	6231014	1.25%	Simple_repeat		110384	5059645	1.02%
Total		632158	271123191	54.23%	Total		591769	255962919	51.38%

Total average MLAh1 & MLAh2 263543055

Liu et al. Variation between haplotype-resolved genome assemblies of *Musella lasiocarpa*.

Supplementary Table S16. Fluorescence *in situ* hybridization (FISH) probes used in FISH experiments.

Slide code	Green	Red
22R-12_MLA_CenG_TeloR-011	EgCEN-FAM-green EGLG2640L: (6FAM)GGCCTATATTTTGAAATTCCGAGACGGTGCATGAAAAACCGATCGAAACGAAACATTGCG	Telo-Cy3-red Plant Telo Red (Cyanine3)TTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTT
22R-12_MLA_CenG_TeloR-017		
D4-8MLA_45Sg_TELOr_23042	45S rDNA-FAM-green 18SmiddleFAM (6FAM)GAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAGCAGG 45S_5rDNApolygon5 (6FAM)ACGAGCTCCAGCTATCCTGAGGGAAACTTCGGAGGGAACCAG 45S_6rDNApolygon6 (6FAM)GTCAACGCGAGCTGATGACTCGCGCTTACTAGGAATTCCTCG	Telo-Cy3-red Plant Telo Red (Cyanine3)TTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGG
21L-6_MLA_cenG_5Sr_298	EgCEN-FAM-green EGLG2640L: (6FAM)GGCCTATATTTTGAAATTCCGAGACGGTGCATGAAAAACCGATCGAAACGAAACATTGCG	5SrDNA-Cy3-red 5S_3R_Cy3: (Cyanine 3)AGTACTAGGATGGGTGACCCCCTGGGAAGTCCTCGTGTTGC 5S-4M_Cy3: (Cyanine3)TCAGAACTCCGAAGTTAAGCGTGCTTGGGCGAGAGTAGTAC
21L-6_MLA_cenG_5Sr_ 288		
21L-6_MLA_cenG_5Sr_293		
22R-11MLA_5Sg_45Sr_L091	5S rDNA-FAM-green 5S_3R_FAM: (6FAM)AGTACTAGGATGGGTGACCCCCTGGGAAGTCCTCGTGTTGC 5S_6L_FAM: (6FAM)GCGATCATACCAGCACTAAAGCACCGGATCCCATCAGAACTCC	45SrDNA Cy3-red 18Smidright Cy3: (Cyanine3)CAAGAACGAAAGTTGGGGGCTCGAAGACGATCAGATACCGTCC 45S_7rDNApolygon7: (Cyanine3)GGCATCACAGACCTGTTATTGCCTCAAACTTCCGTGGCCTAG 5.8ScentreRight: (Cyanine3)CCGTGAACCATCGAGTCTTTGAACGCAAGTTGCGCCCGAGGCC
20A-4; Mus4_MuTRr 5Sg_007	5S rDNA-dig-green 5S dig; PCR of insert of pTa794 (Gerlach and Dyer 1980, 410bp insert from Triticum aestivum) using M13 sequencing primers, cut band from gel and purify, random priming with digoxygenin-11- dUTP.	MuTR-bio-red MuTR bio, PCR product (~, using MuTR primers and using Ensete ventriocosum total genomic DNA as template, 200bp band cut from gel and purify and use for random priming with biotin-16-dUTP. MuTrFor TGCTTCGCTCKCTGTCCGC MuTrRev ASGAGCRGCACCTTCCCTGC