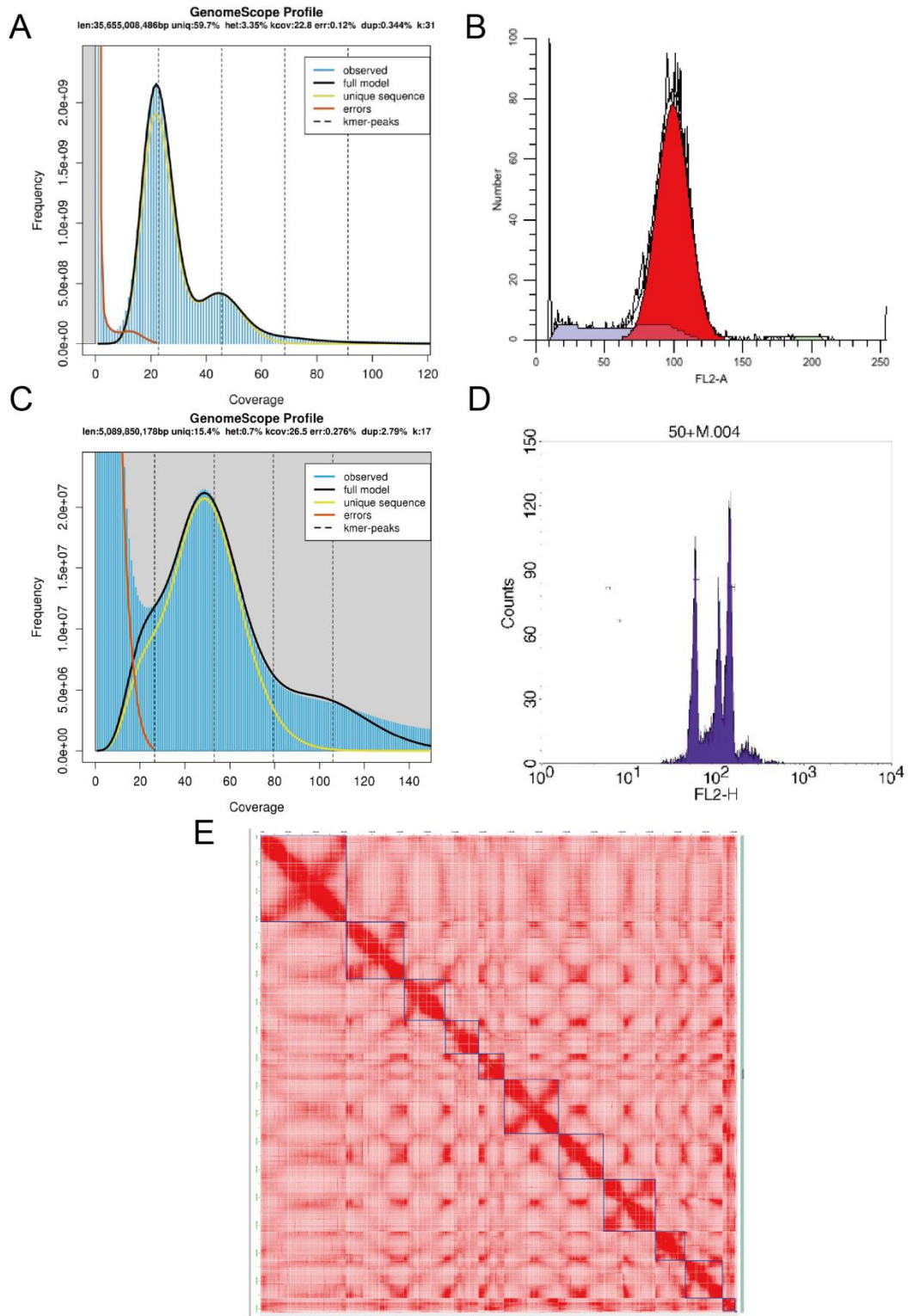
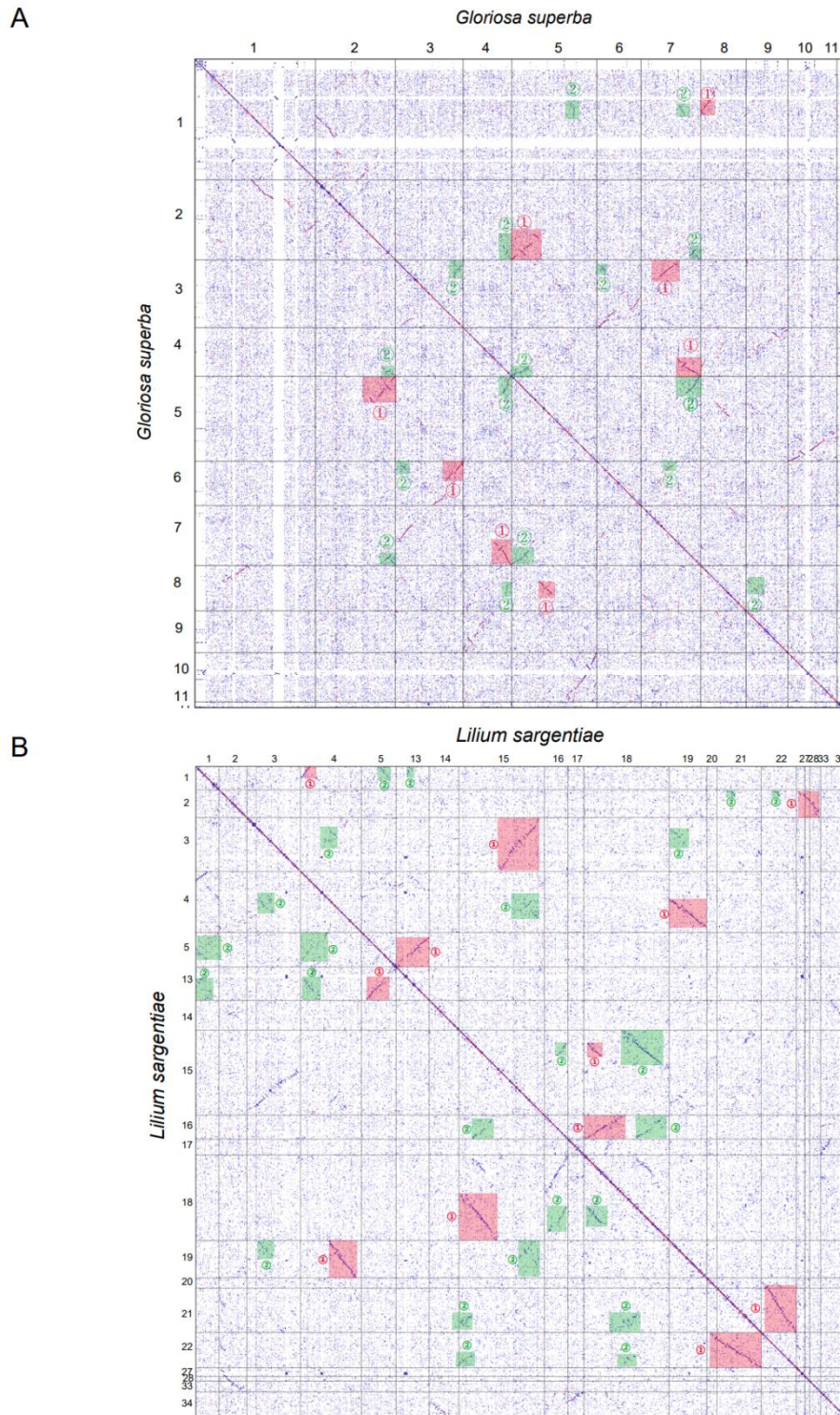


**The giant genome of lily provides insights into the hybridization of  
cultivated lilies**

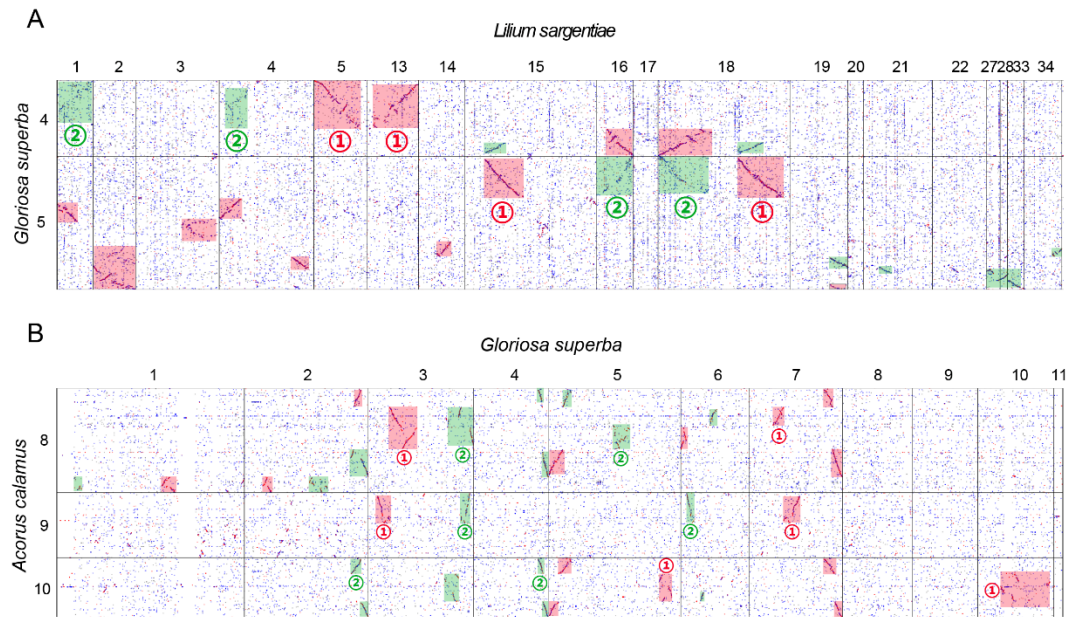
Liang *et al.*



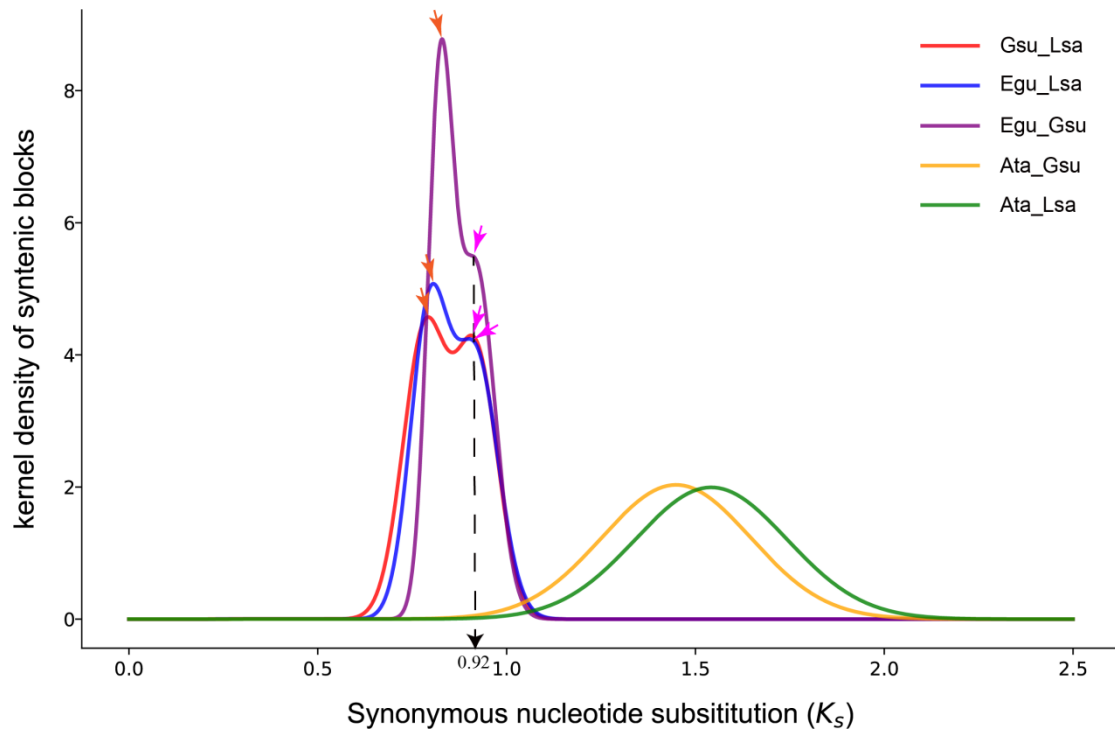
**Supplementary Figure 1. Genome assemblies of *Lilium sargentiae* and *Gloriosa superba*.** Estimation of *L. sargentiae* genome size by K-mer (A) and flow cytometry analysis (B). Estimation of *G. superba* genome size by K-mer (C) and flow cytometry analysis (D). (E) Hi-C heatmap of the *G. superba* genome.



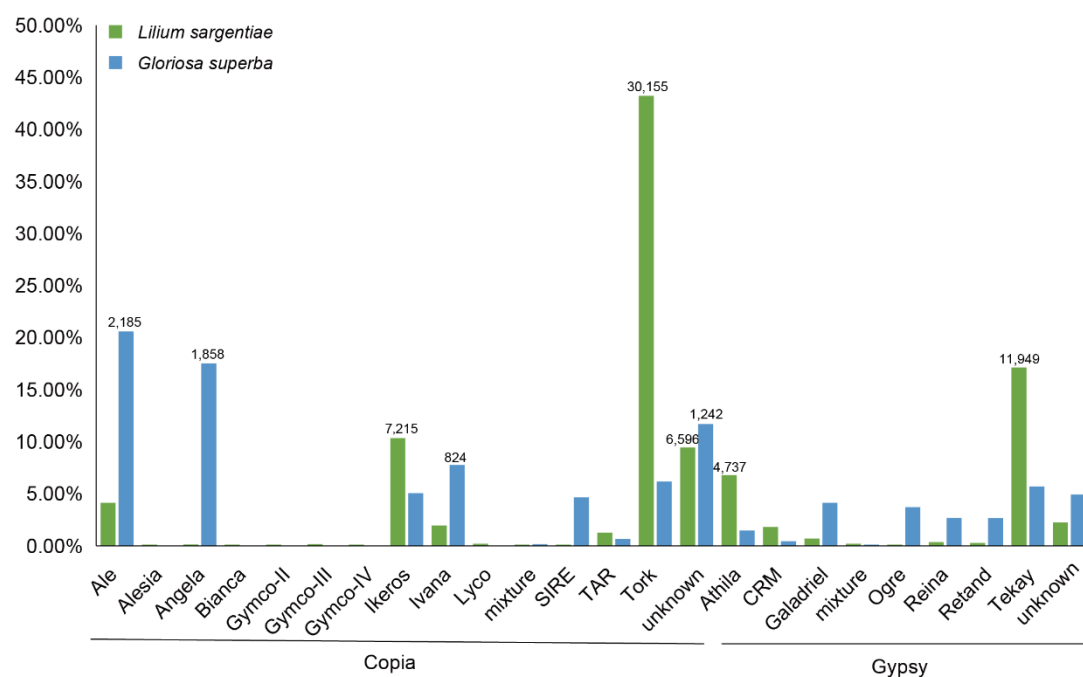
**Supplementary Figure 2. Synteny analysis within genomes.** (A) Dotplot of homologies within *Gloriosa superba* genome. (B) Dotplot of homologies within *Lilium sargentiae* genome. The ancient monocot-shared WGD event ( $\tau$ ) is blocked by green triangles and numbered as 2. The recent species-specific WGD events in both *G. superba* and *L. sargentiae* are blocked by red triangles and numbered as 1.



**Supplementary Figure 3. Synteny analysis between genomes.** Dotplot of homologies between *Gloriosa superba* and *Lilium sargentiae* (A) and between *Acorus calamus* and *G. superba* (B). The ancient monocot-shared WGD event ( $\tau$ ) is blocked by green triangles and numbered as 2. The recent species-specific WGD events in *L. sargentiae* and *G. superba* are blocked by red triangles and numbered as 1.



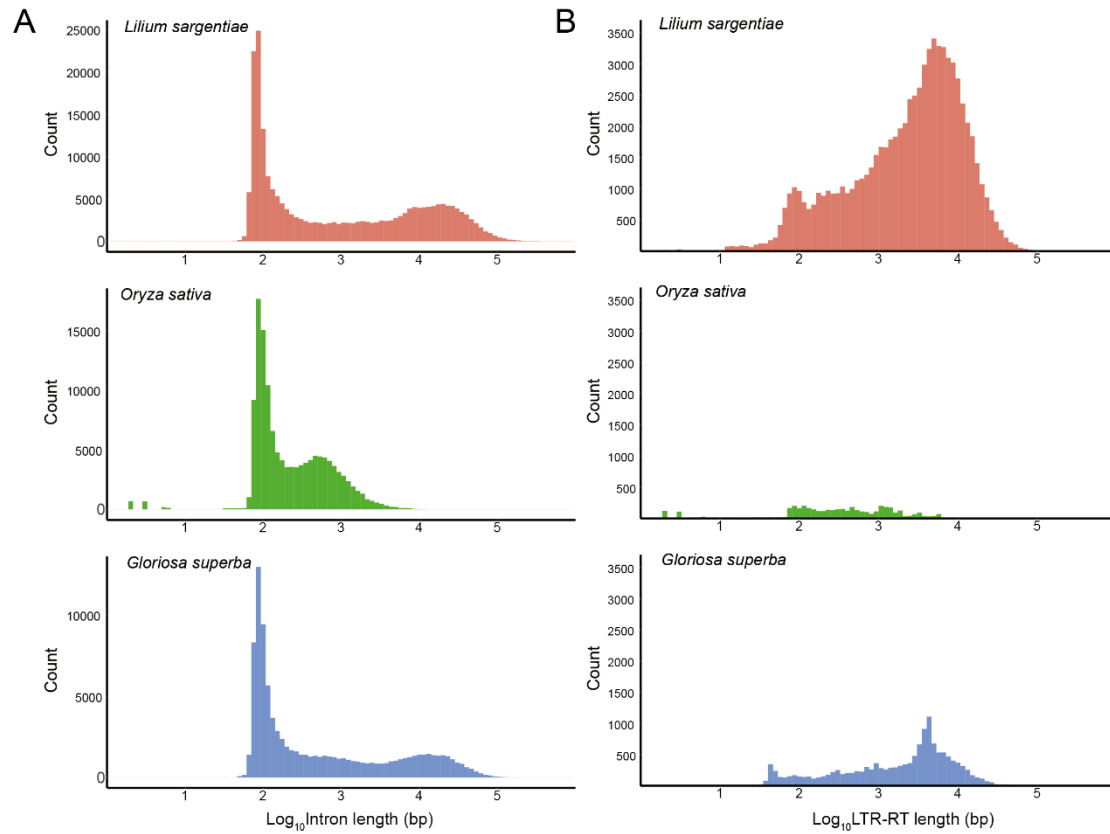
**Supplementary Figure 4.  $K_s$  distributions of orthologous gene pairs between species after correlation.** Gsu, *Gloriosa superba*; Lsa, *Lilium sargentiae*; Egu, *Elaeis guineensis*; Ata, *Acorus tatarinowii*.  $K_s$  distribution results show a bimodal pattern between *G. superba*, *L. sargentiae*, and *Elaeis guineensis*, indicating that these three species experienced a common  $\tau$  WGD event before divergence, as represented by the right-hand peaks pointed by pink arrows. The left-hand peaks pointed by orange arrows represented the divergence between species. In contrast, only a single peak representing species divergence was identified between *A. tatarinowii* and *G. superba*, as well as between *A. tatarinowii* and *L. sargentiae*, respectively. This suggests that *A. tatarinowii* did not undergo the  $\tau$  WGD event.



**Supplementary Figure 5. Statistics of the proportion of the number of long terminal repeats retroelements (LTR-RTs) belonging to different subfamilies of *Ty1/Copia* and *Ty3/Gypsy* LTR-RTs accounting for the total number of LTR-RTs in *Lilium sargentiae* and *Gloriosa superba*.** The numbers of main LTR-RTs are marked above the corresponding columns. Source data are provided as a Source Data file.

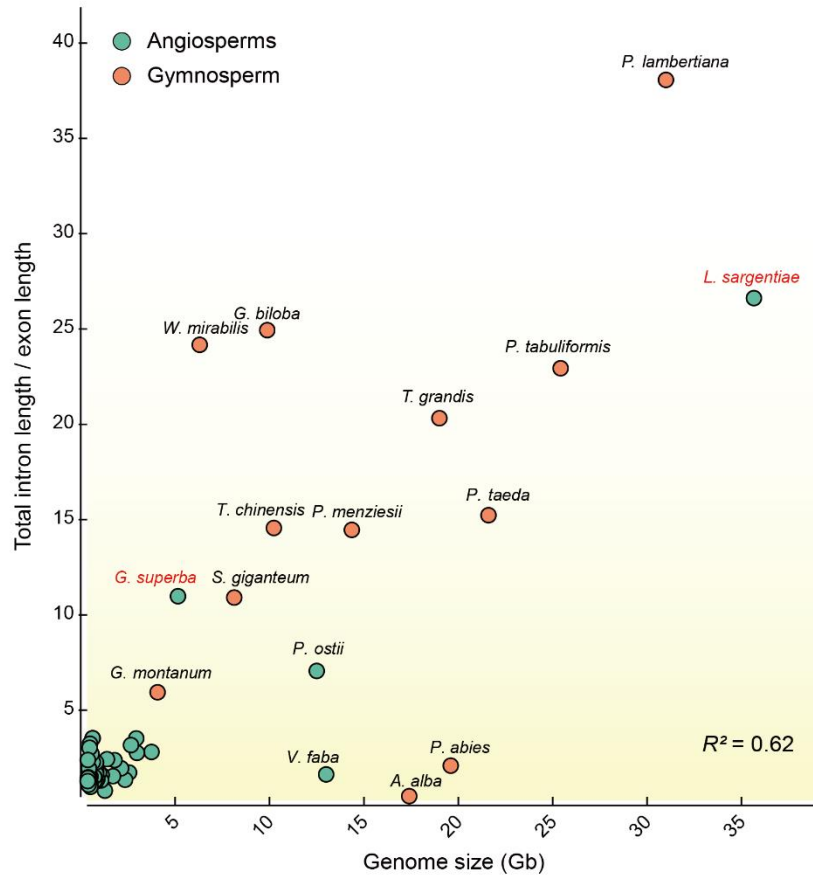


**Supplementary Figure 6. GO enrichment analysis.** GO enrichment analysis of genes 2 Kb upstream and downstream of Tork and Tekay long terminal repeats retroelements (LTR-RTs) (A) as well as that of genes containing these two types of LTRs in introns (B). GO terms related to unsexual organ propagation are in red fonts. GO terms related to underground organ sensing changes in temperature and light are in bold fonts. Source data are provided as a Source Data file.

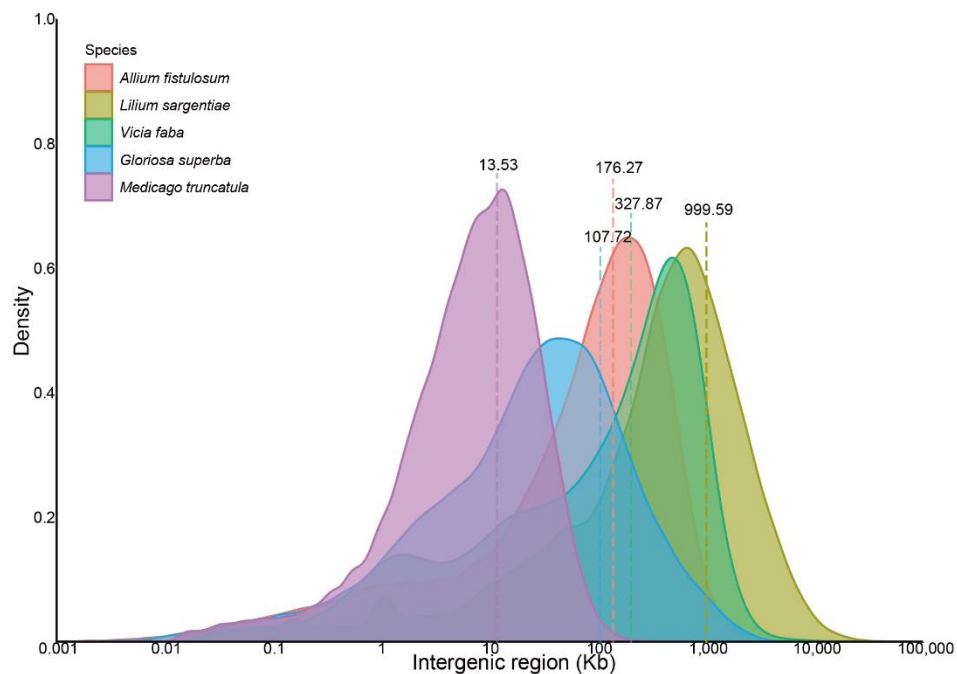


**Supplementary Figure 7. Statistics of lengths of introns and long terminal repeats retroelements (LTR-RTs) in introns in different genomes.** (A) Distribution of intron lengths within the *Lilium sargentiae* genome, the *Oryza sativa* genome, and the *Gloriosa superba* genome. (B) Distribution of lengths of long terminal repeats retroelements (LTR-RTs) in introns within the *L. sargentiae* genome, the *O. sativa* genome, and the *G. superba* genome.

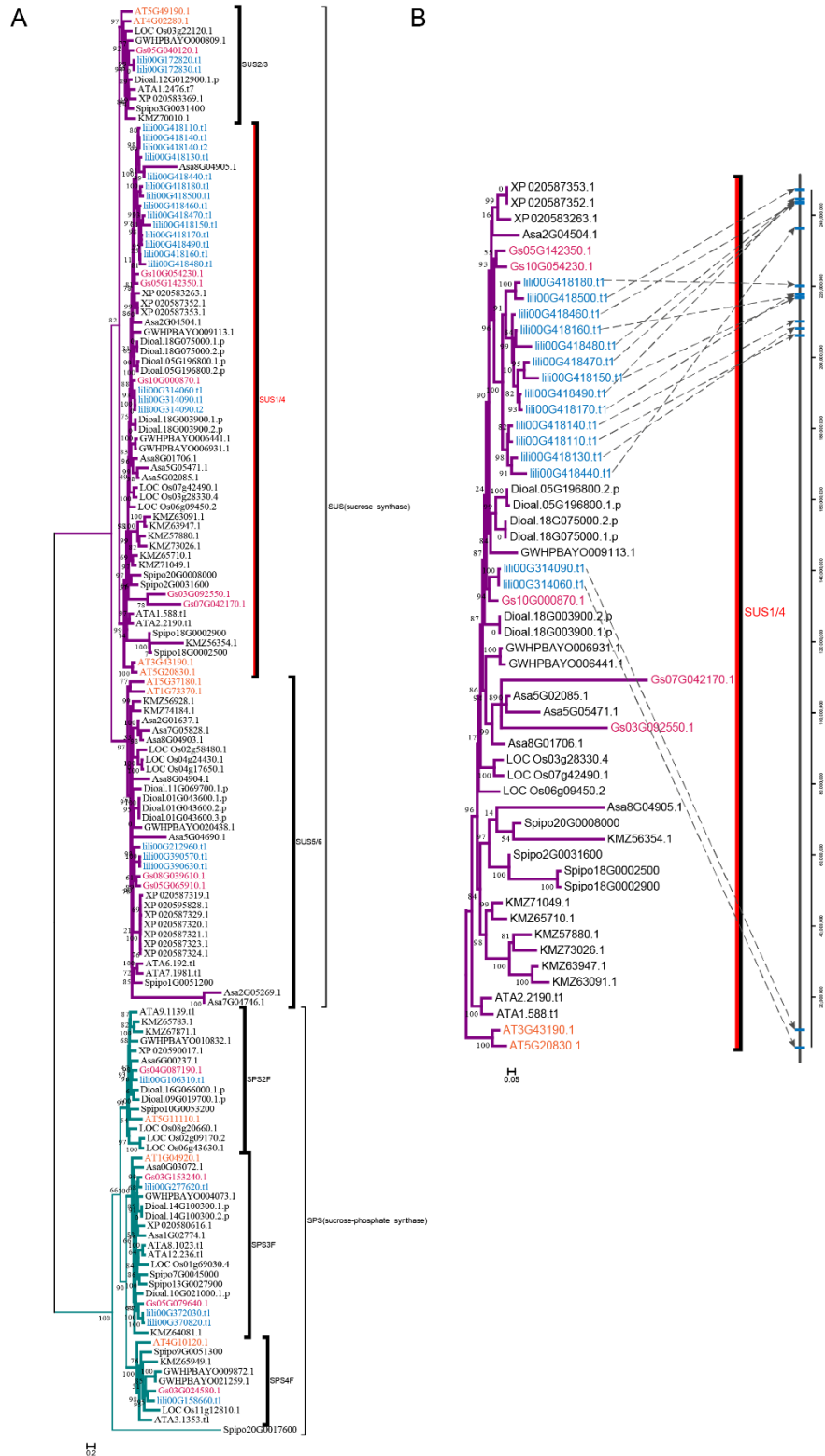
A



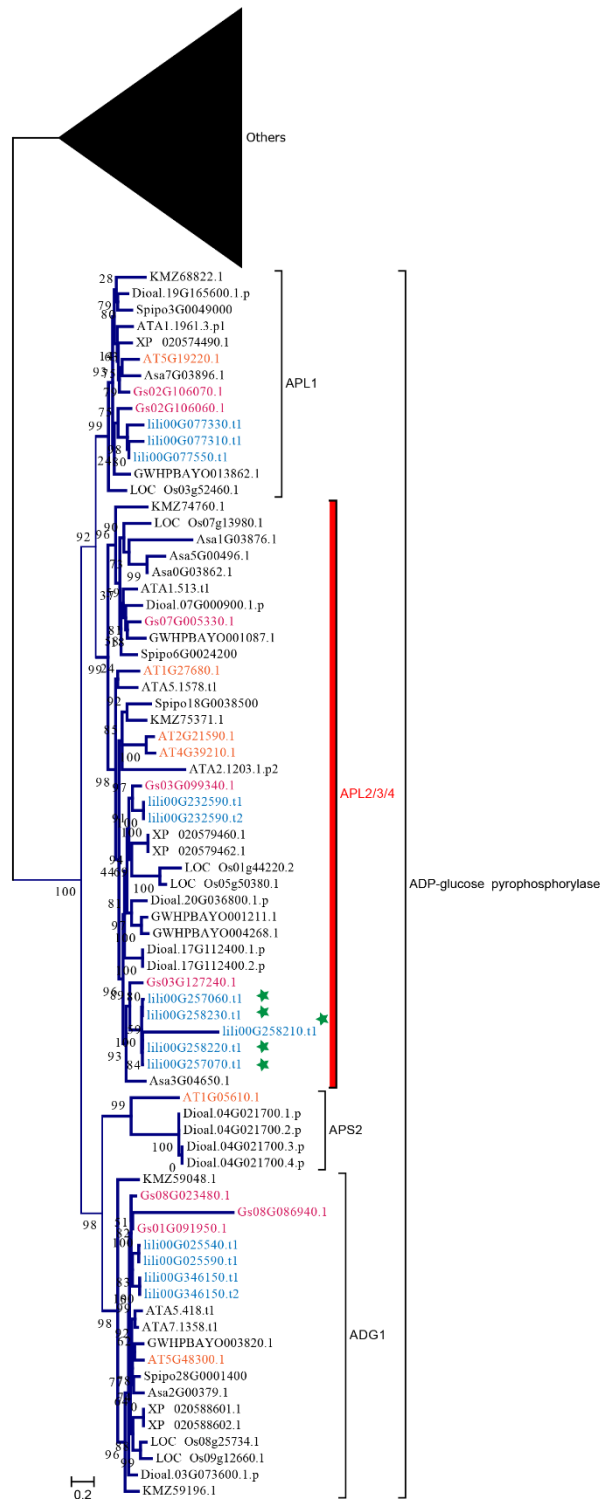
B



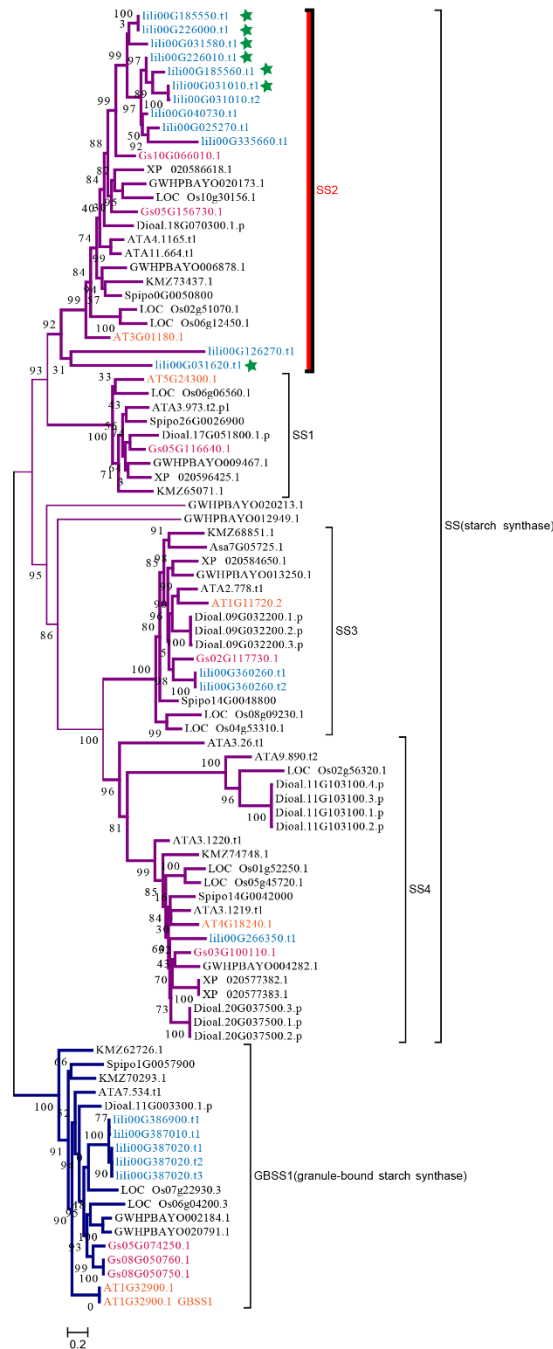
**Supplementary Figure 8. Statistics of genomic features in different genomes.** (A) Statistics of the ratio of total intron length to total exon length. All the data were used to calculate the Pearson correlation coefficient ( $R^2$ ). The ratio for other species was obtained from previous study<sup>23</sup>. (B) Statistics of lengths of intergenic regions. Source data are provided as a Source Data file.



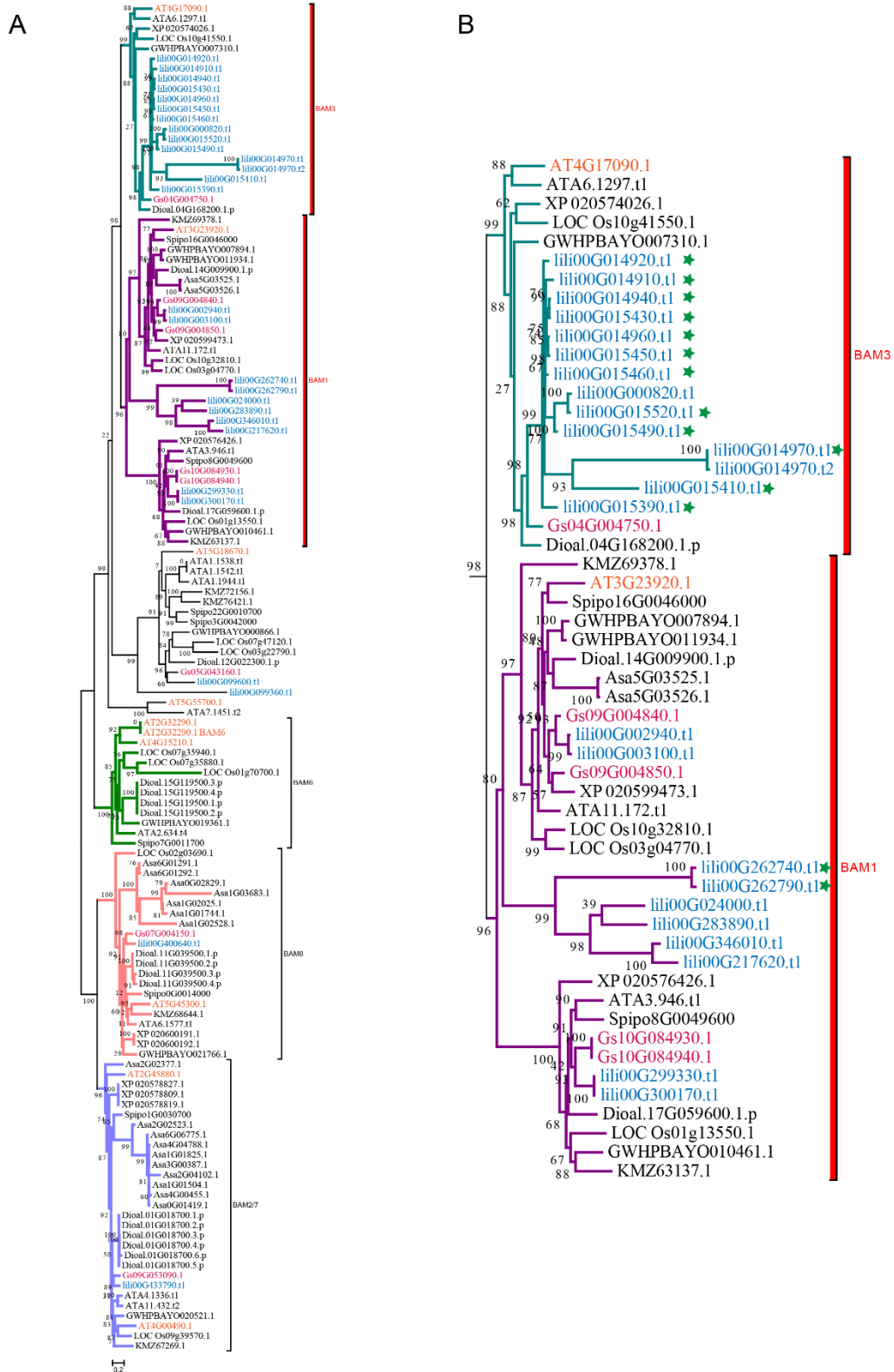
**Supplementary Figure 9. Identification of SUS (SUCROSE SYNTHASE) and SPS (SUCROSE-PHOSPHATE SYNTHASE) genes.** (A) Phylogenetic tree of SUS and SPS genes. Genes in *Lilium sargentiae*, *Gloriosa superba*, and *Arabidopsis* are in blue, pink, and yellow fonts. (B) Phylogenetic tree of SUS1/4 homologs.



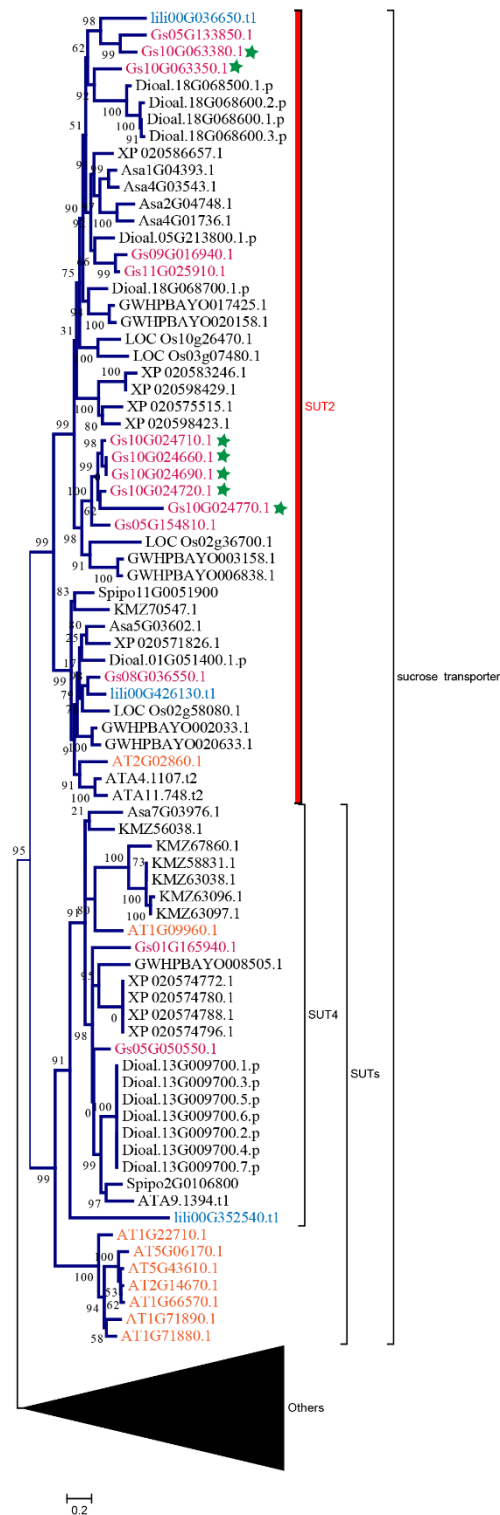
**Supplementary Figure 10. Identification of ADP GLUCOSE PYROPHOSPHORYLASE genes.** Genes in *Lilium sargentiae*, *Gloriosa superba*, and *Arabidopsis* are in blue, pink, and yellow fonts. Genes generated by tandem duplication are represented by green stars.



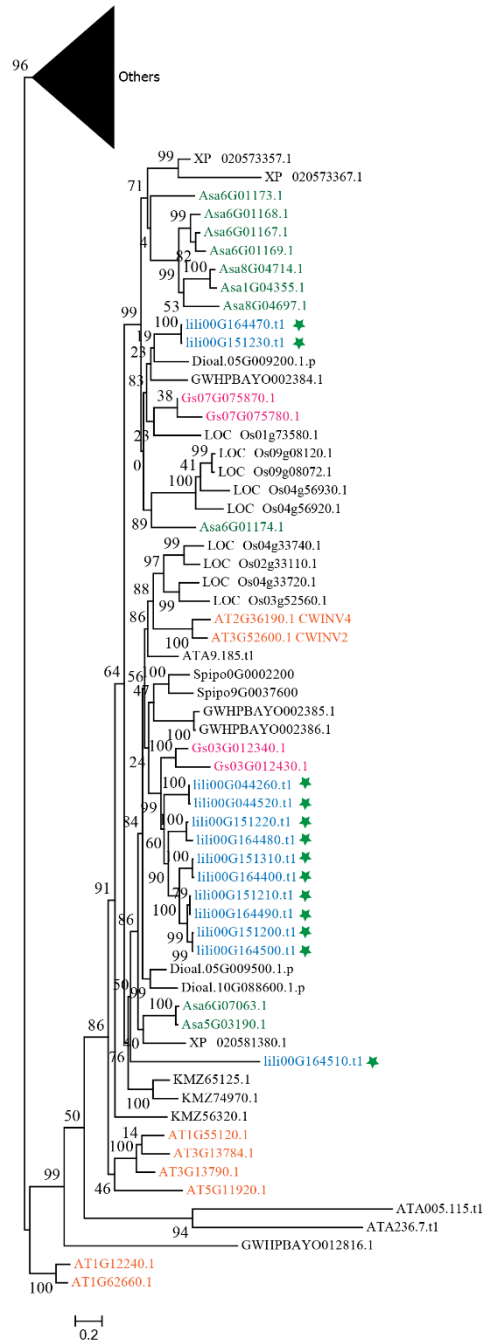
**Supplementary Figure 11. Phylogenetic tree of SS (STARCH SYNTHASE) and GBSS (GRANULE-BOUND STARCH SYNTHASE) genes.** Genes in *Lilium sargentiae*, *Gloriosa superba*, and *Arabidopsis* are in blue, pink, and yellow fonts. Genes generated by tandem duplication are represented by green stars.



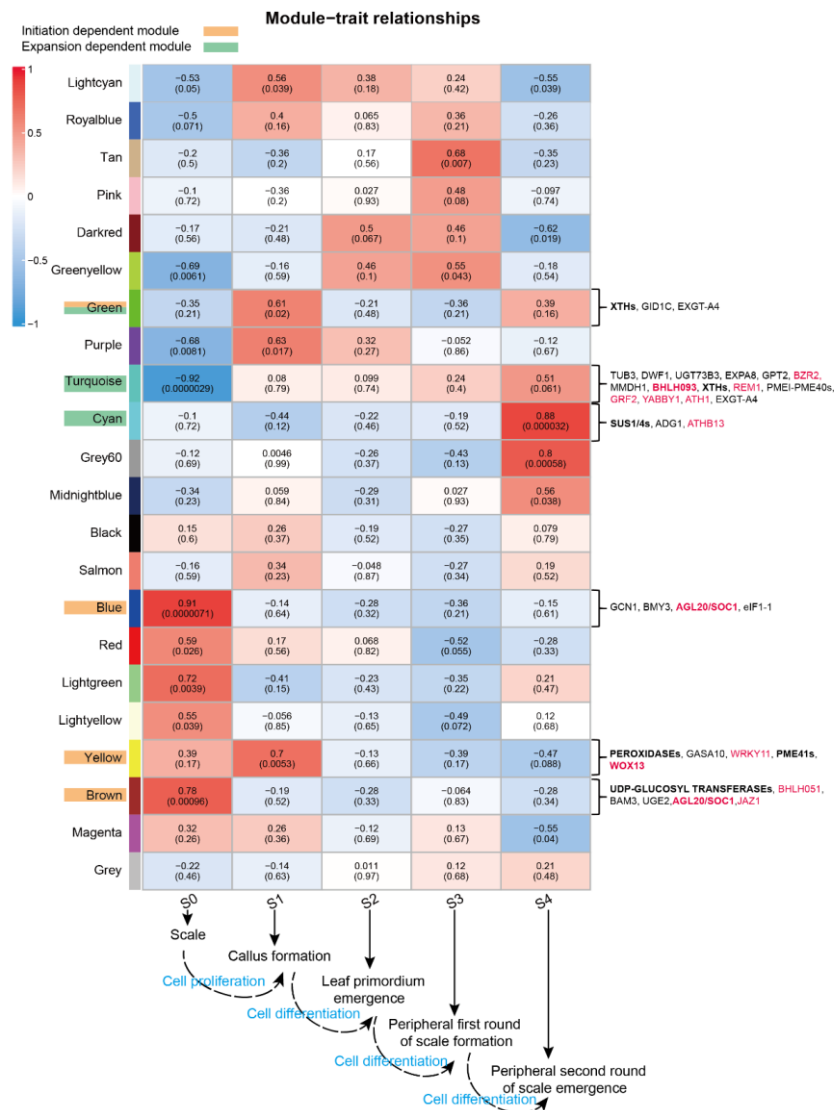
**Supplementary Figure 12. Identification of BETA-AMYLASE genes (BAMs).** (A) Phylogenetic tree of BAMs. (B) The detailed tree of *BAM1* and *BAM3* genes. Genes in *Lilium sargentiae*, *Gloriosa superba*, and *Arabidopsis* are in blue, pink, and yellow fonts.



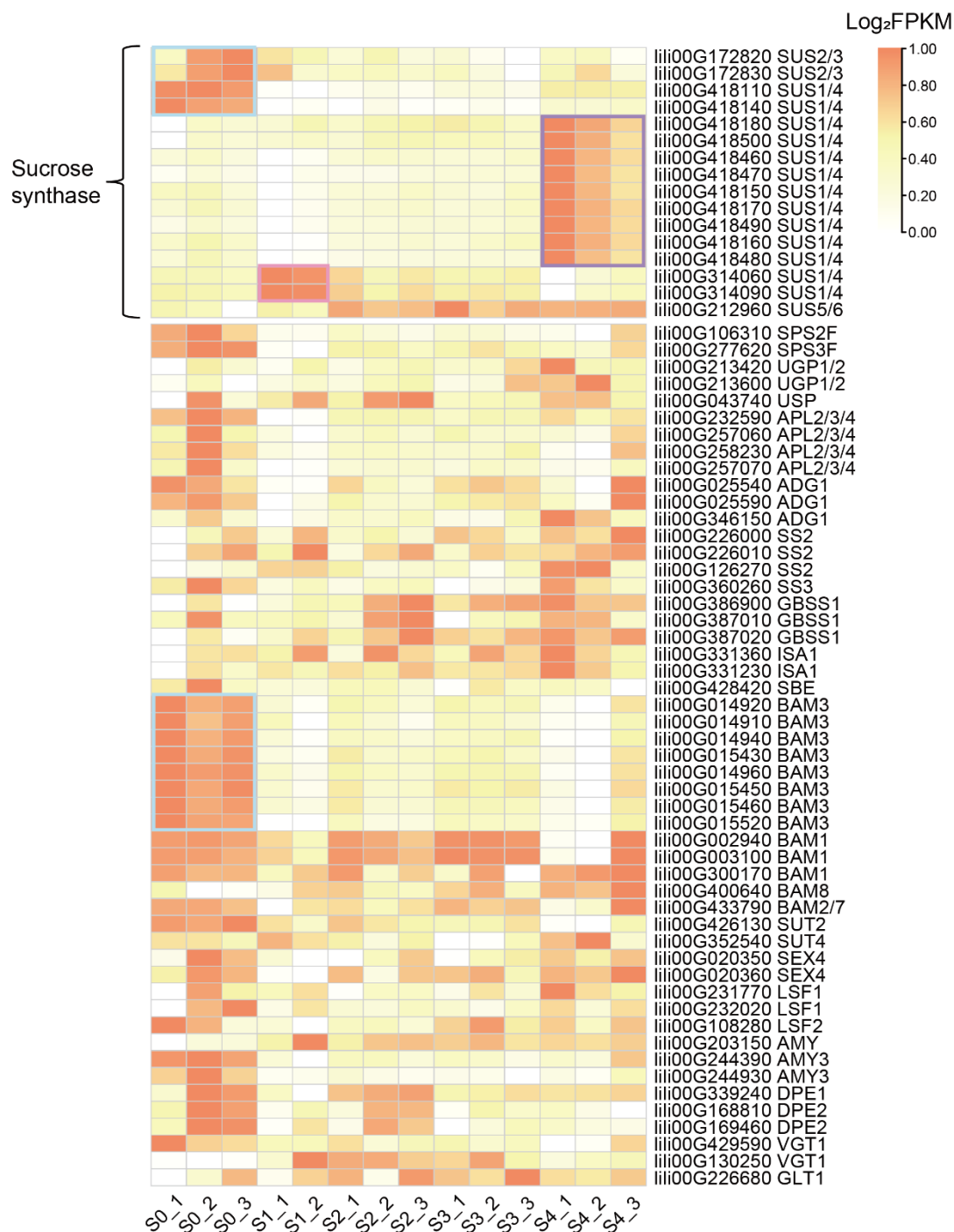
**Supplementary Figure 13. Identification of SUT (SUCROSE TRANSPORTER) genes.** Genes in *Lilium sargentiae*, *Gloriosa superba*, and *Arabidopsis* are in blue, pink, and yellow fonts. Genes generated by tandem duplication are represented by green stars.



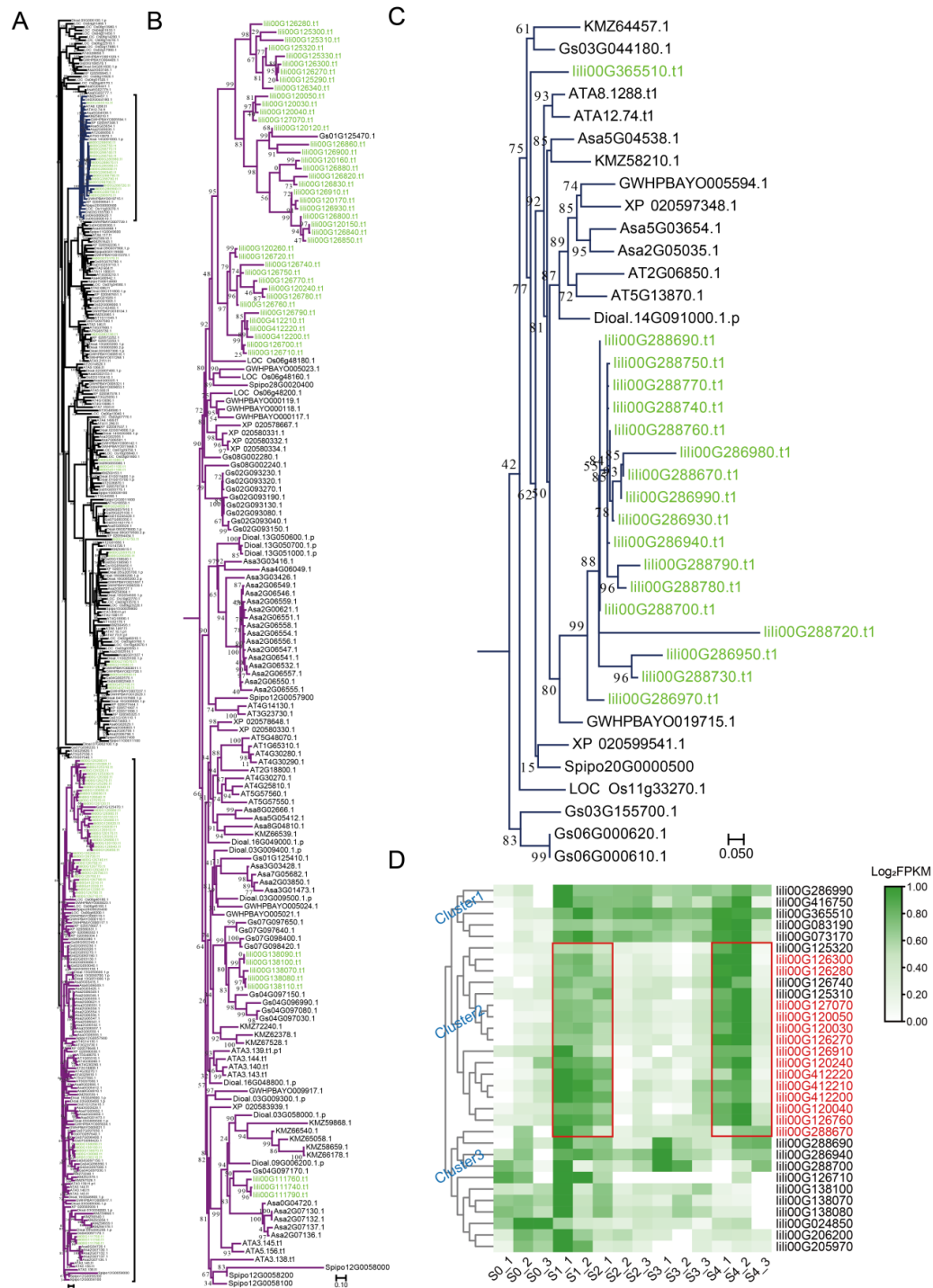
**Supplementary Figure 14. Identification of INVERTASE genes.** Genes in *Lilium sargentiae*, *Gloriosa superba*, and *Arabidopsis* are in blue, pink, and yellow fonts. Genes generated by tandem duplication are represented by green stars.



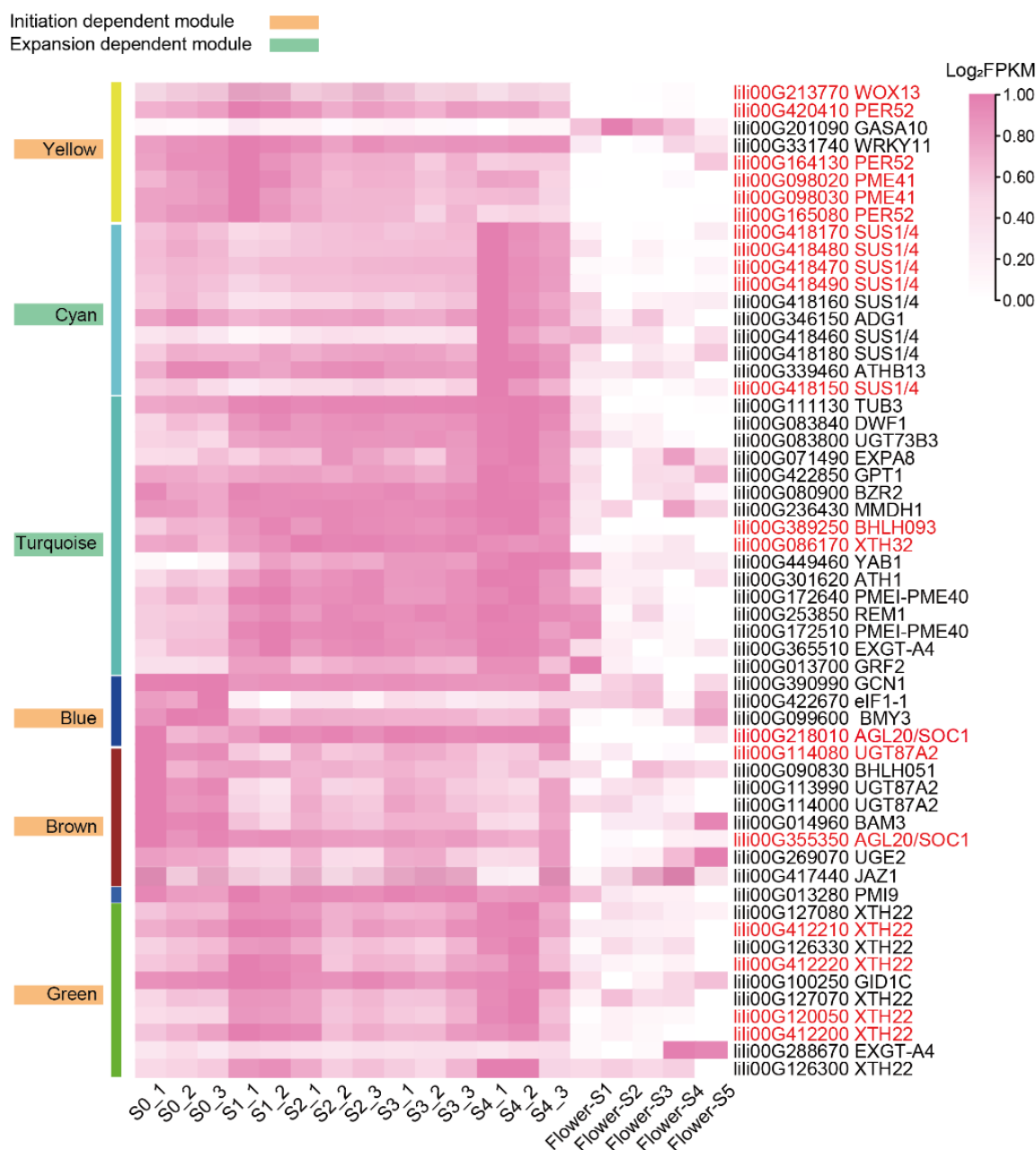
**Supplementary Figure 15. Weighted gene co-expression network analysis (WGCNA) of bulb development in lily.** (A) Relationship between modules and time-series (S0-S4, four development stages from bulblet initiation to bulblet expansion) lily scales (the base of scale as well as the new tissue). Before S3, the lily bulblet is during the initiation period. From S3, the lily bulblet enters the expansion period. Each column corresponds to a tissue, and each row corresponds to a gene module. Each cell contains the corresponding Pearson's correlation efficient  $r$  and  $P$ -value. The cell is color-coded by correlation according to the color legend on the right of the heatmap. The characteristics of each stage are shown below the picture. Out of the 22 modules, 18 were analyzed in detail, excluding the black, salmon, magenta, and grey modules due to their low correlation with our traits. From these, we selected six modules of particular interest (highlighted by rectangular boxes) based on the expression patterns and predicted functions of the genes. The key genes in these modules are indicated next to the corresponding modules, with genes encoding transcription factors highlighted in red. Potential bulb-specific regulators in lily are in bold fonts (genes with the maximum FPKM in flower samples less than 10 are deemed as not expressed in flowers). Source data are provided as a Source Data file.



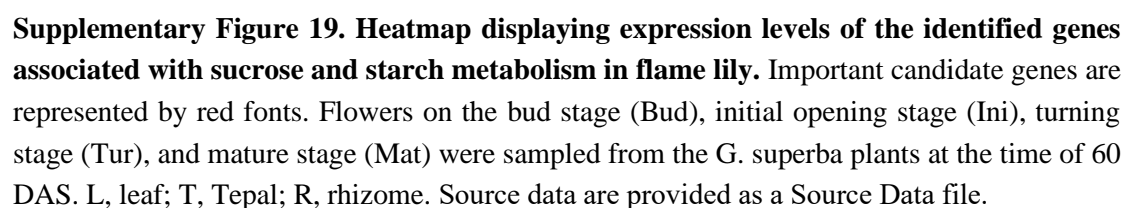
**Supplementary Figure 16. Heatmap of expression patterns of gene members involved in sucrose and starch metabolism in lily.** Genes with a maximum FPKM value of less than 5 were filtered. S0-S4 represents four development stages from bulblet initiation to bulblet expansion. Source data are provided as a Source Data file.

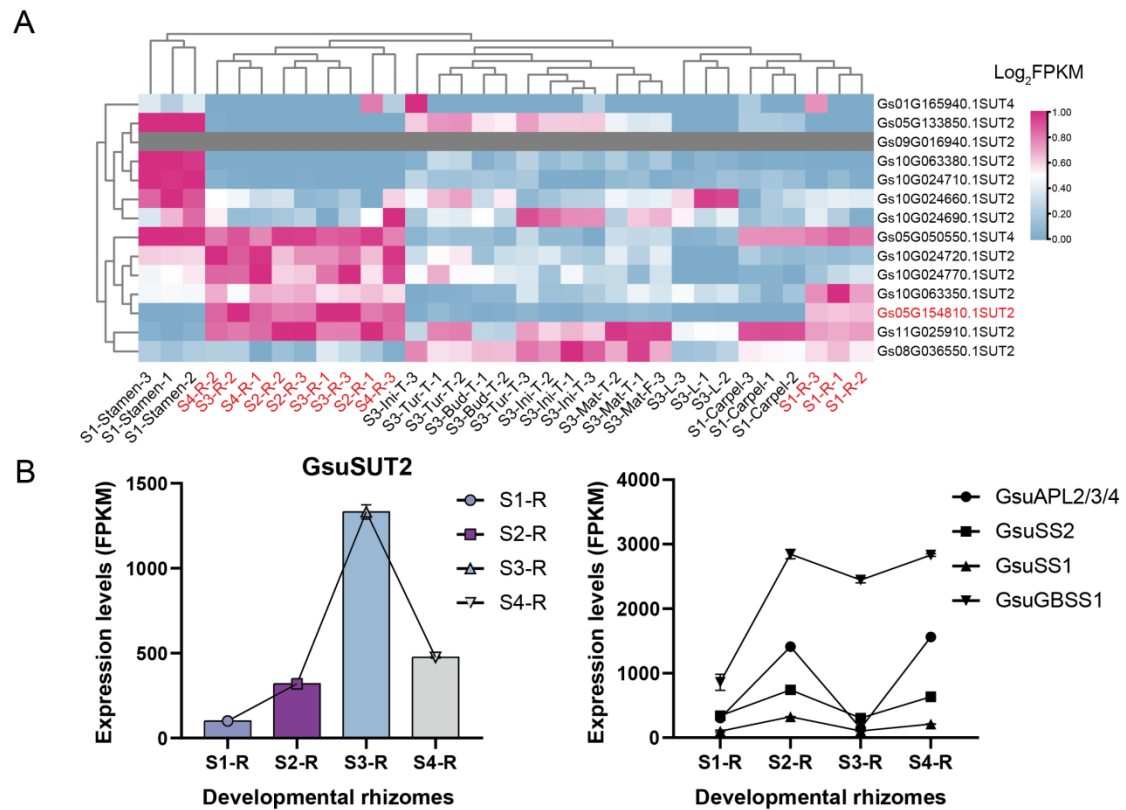


**Supplementary Figure 17. Phylogenetic and expression analysis of the XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE (XTH) family.** (A) Phylogenetic tree of gene members of the XTH family. (B) and (C) showed the subtrees of (A). Lily genes are in green fonts. (D) Heatmap of expression patterns of gene members of the XTH family. Members that were co-expressed in the green module generated by WGCNA are in red fonts. In total, the expression profiles could be divided into three clusters. The genes in cluster 2 are mainly expressed in the samples of S1 and S4. S0-S4 represents four development stages from bulblet initiation to bulblet expansion. Genes with a maximum FPKM value of less than 5 are filtered. Source data are provided as a Source Data file.

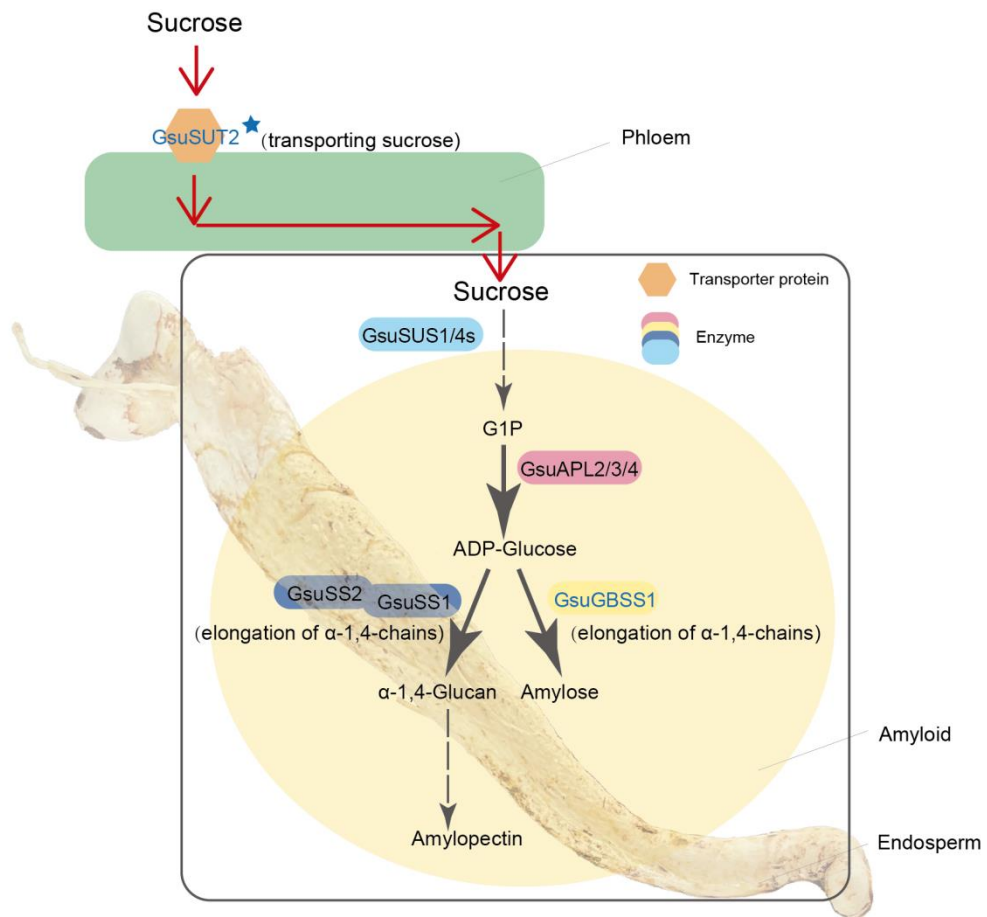


**Supplementary Figure 18. Heatmap of expression patterns of candidate gene members obtained by weighted gene co-expression network analysis (WGCNA) of time-series RNA-seq lily samples.** Potential bulb-specific regulators in lily are in red fonts (genes with a maximum FPKM in flower samples less than 10 are deemed as not expressed in flowers). Flower-S1 to Flower-S5 are the lily flowers from the bud stages to the mature stages. S0-S4 represents four development stages from bulblet initiation to bulblet expansion. The colors of lines on the left of the heatmap show the modules where the genes are co-expressed.

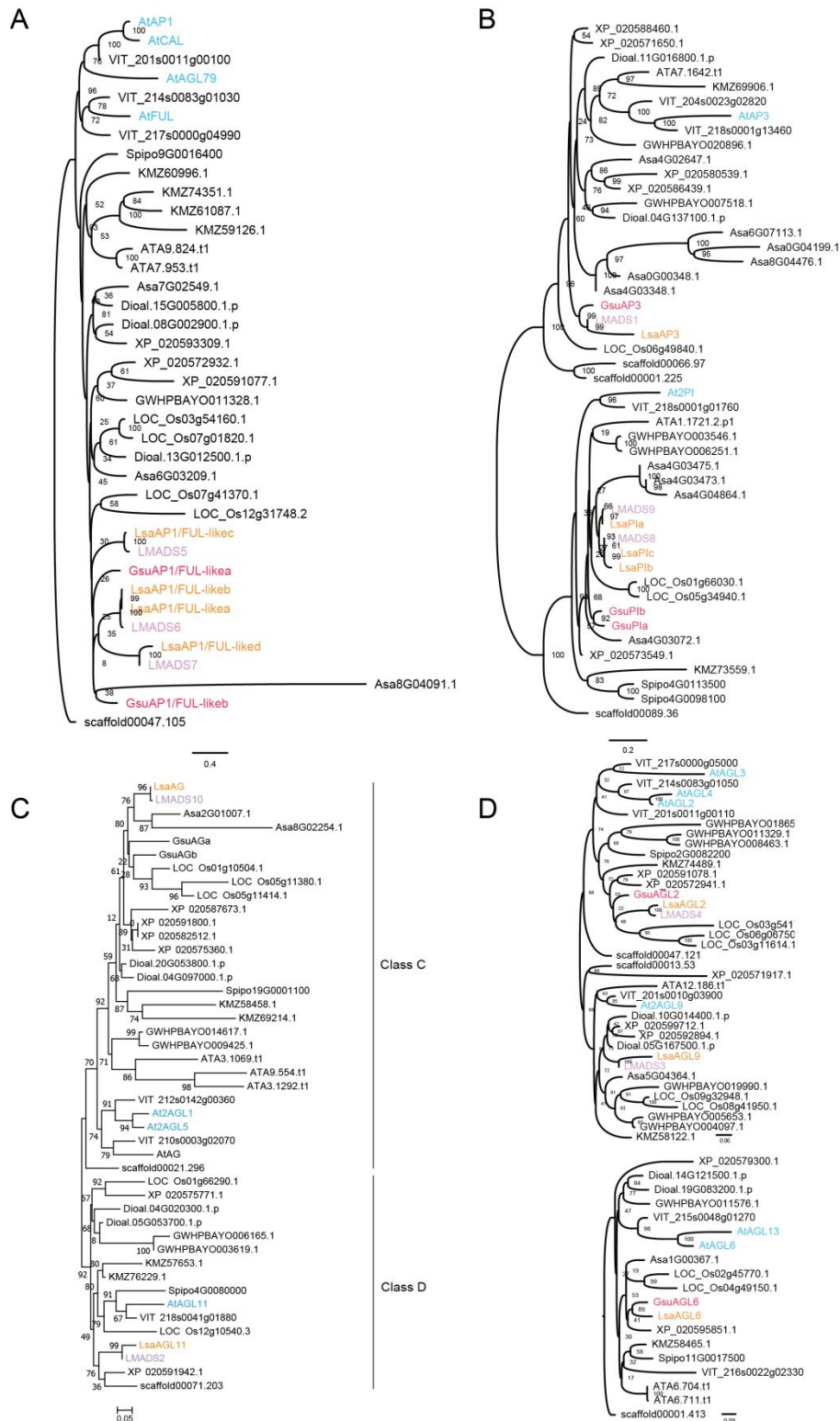




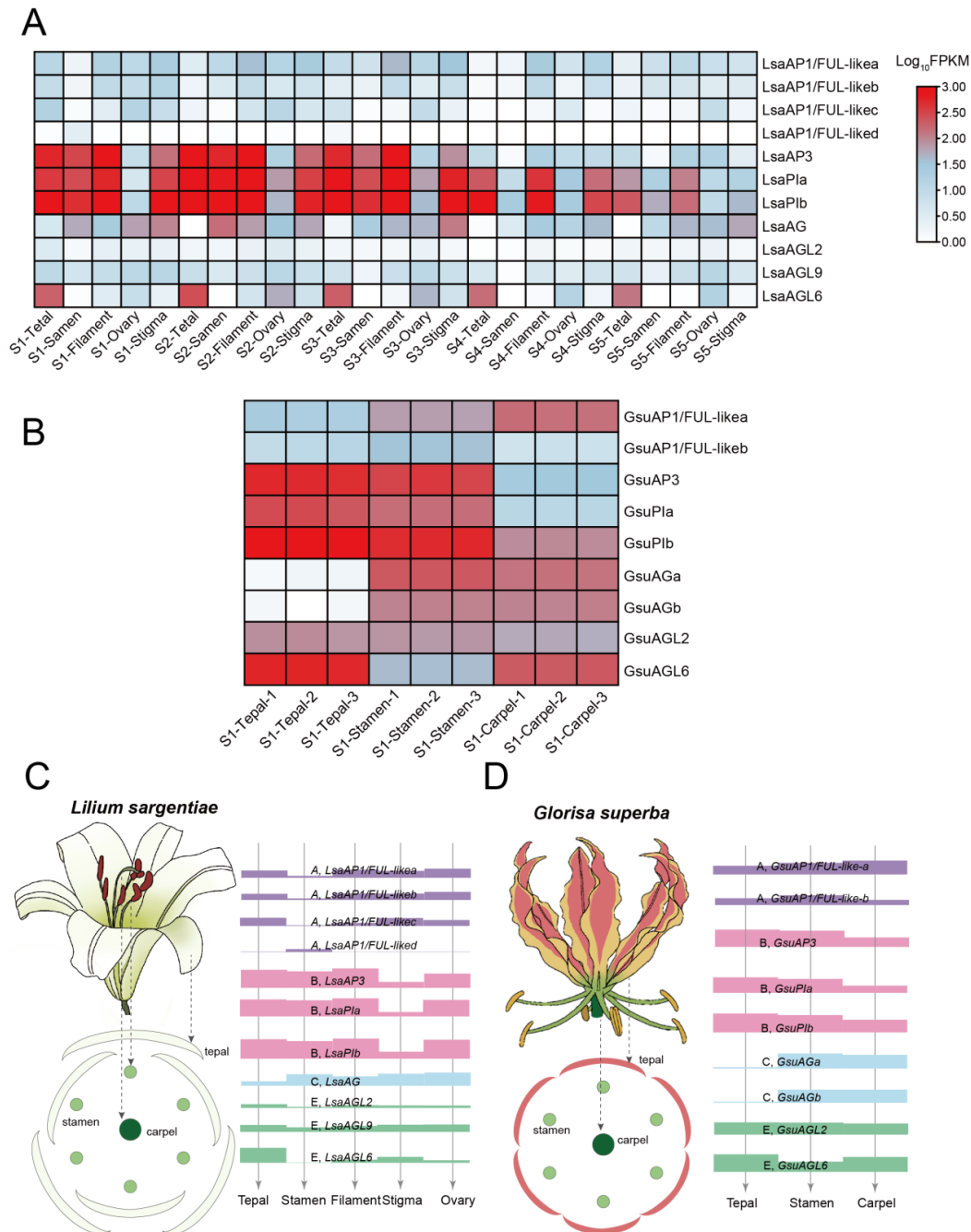
**Supplementary Figure 20. Expression of important genes involved in sucrose and starch metabolism during flame lily rhizome development.** (A) Heatmap illustrating expression levels of *SUT2* homologs. Key genes involved in rhizome development are highlighted in red. (B) Histograms depicting expressions of significant regulators of rhizomes. Sampling time points of 20, 40, 60, and 80 days after sprouting (DAS) in *G. superba* were named S1, S2, S3, and S4, respectively. Flowers on the bud stage (Bud), initial opening stage (Ini), turning stage (Tur), and mature stage (Mat) were sampled from the *G. superba* plants at the time of 60 DAS. L, leaf; T, Tepal; R, rhizome. Source data are provided as a Source Data file.



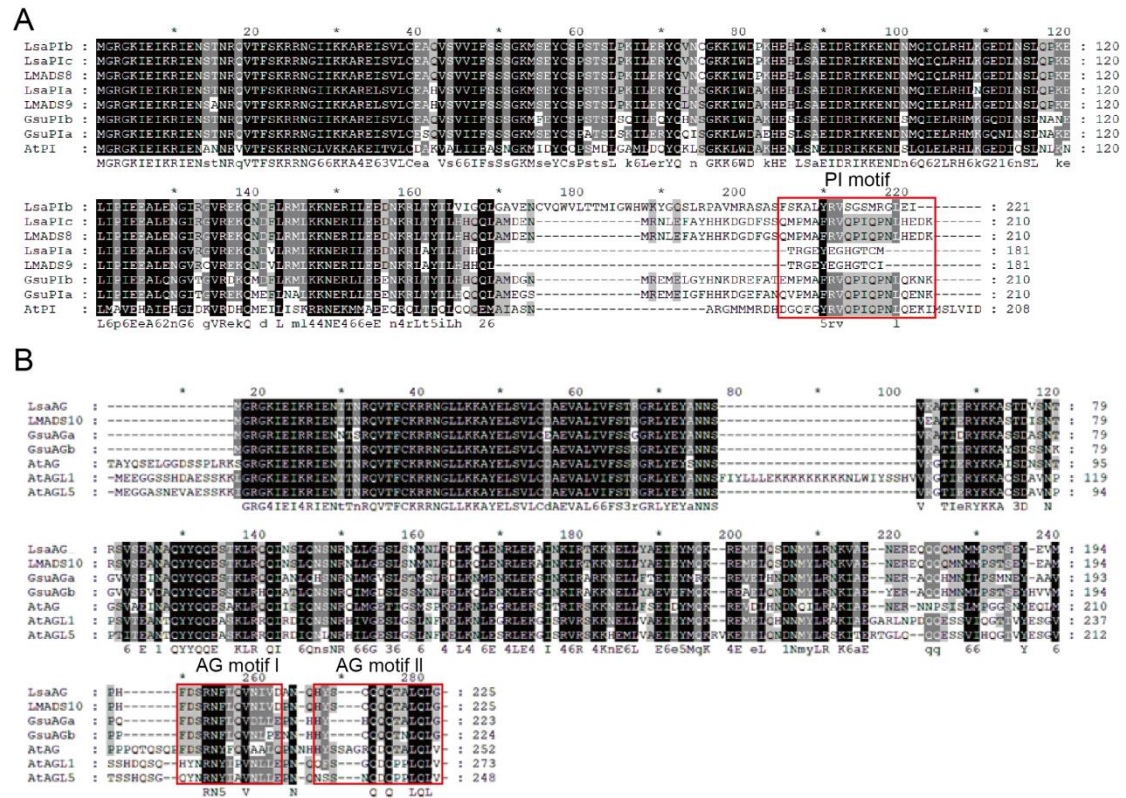
**Supplementary Figure 21. A proposed model of rhizome evolution and formation in *G. superba*.** The homologs of *SUT2* in *G. superba* have expanded by tandem duplication. The potential rhizome-specific regulators are in blue fonts. The role of *GsSUT2* in the sucrose transport process through apoplasmic unloading or symplasmic unloading remains unclear. Here, we assume it operates via symplasmic unloading to propose this model.



**Supplementary Figure 22. Phylogenetic analysis of floral identity genes in *Lilium sargentiae*, *Gloriosa superba*, and other species.** (A) Class A, (B) class B, (C) class C and D, and (D) class E genes. Genes in *L. sargentiae*, *G. superba*, and *Arabidopsis thaliana* are presented by orange, red, and blue fonts, respectively. And the previously cloned MADS genes from *L. longiflorum* are in light purple fonts.

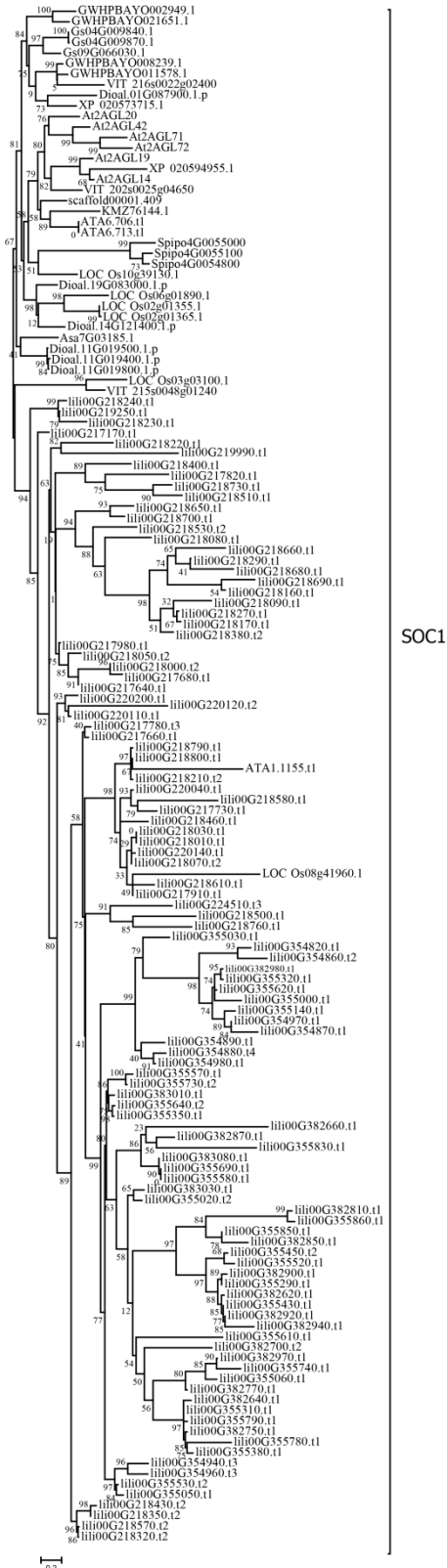


**Supplementary Figure 23. Expression profiles of ABCE MADS-box genes.** (A) and (B) showed gene expression patterns of ABCE MADS-box genes from various organs of *Lilium sargentiae* and *Gloriosa superba*, respectively. Expression values were scaled by  $\log_{10}$ FPKM, in which FPKM was fragments per kilobase of exon per million mapped reads. For *L. sargentiae*, the expression profiles of flower tissues, including tepals, stamens, filaments, ovaries, and stigmas at five consecutive stages (S1-S5) from the bud stage to the mature stage were analyzed. For *G. superba*, tepals, stamens, and carpels at the bud stage (S1) were analyzed. (C) and (D) are the proposed flowering ABCE models that specified floral organs in *L. sargentiae* and *G. superba*, respectively, which were based on the  $\log_{10}(\text{FPKM}+1)$  format of FPKMs. FPKM, fragments per kilobase of exon per million mapped reads. Source data are provided as a Source Data file.

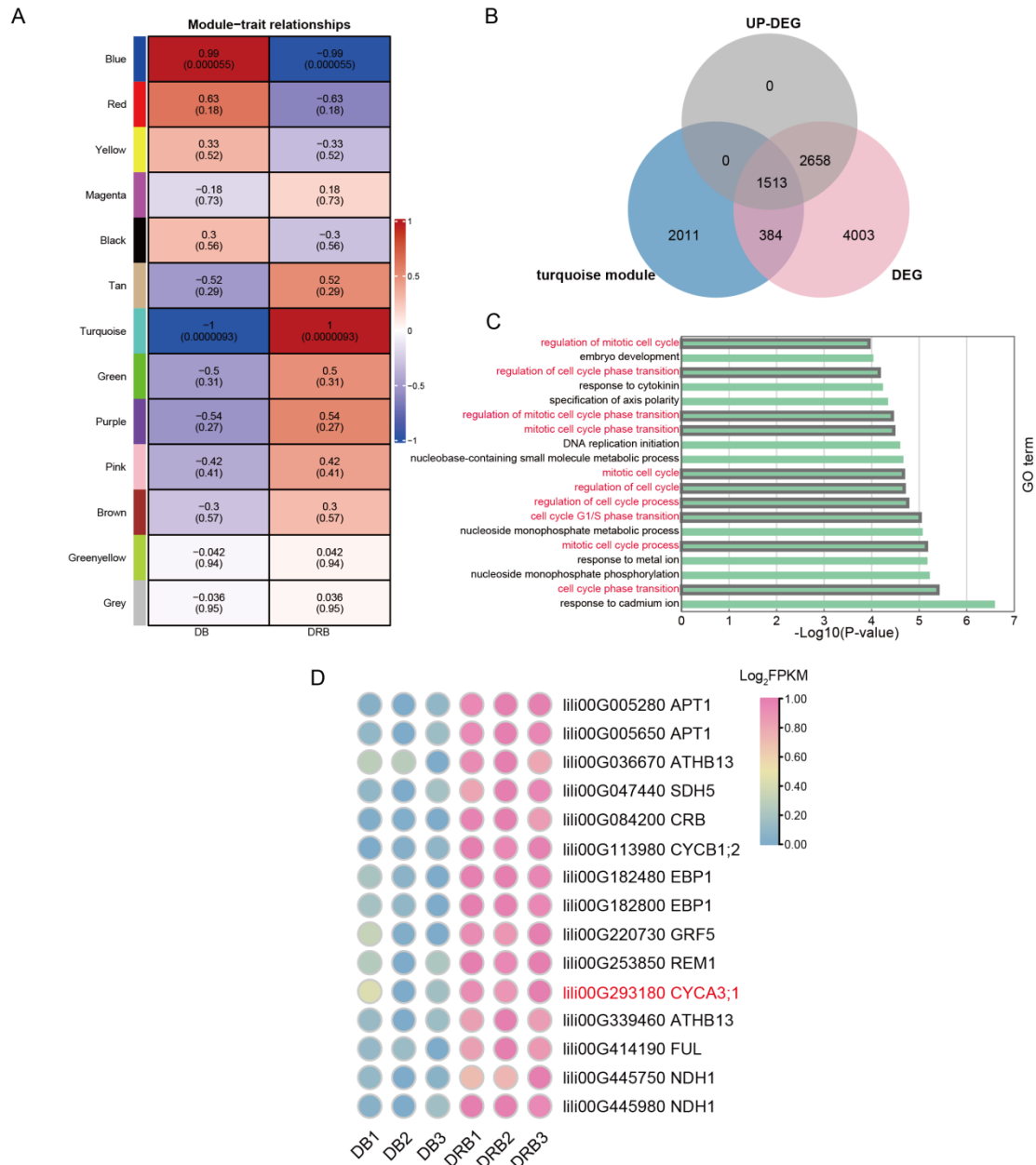


**Supplementary Figure 24. Comparison of protein sequences of MADS-box genes.**

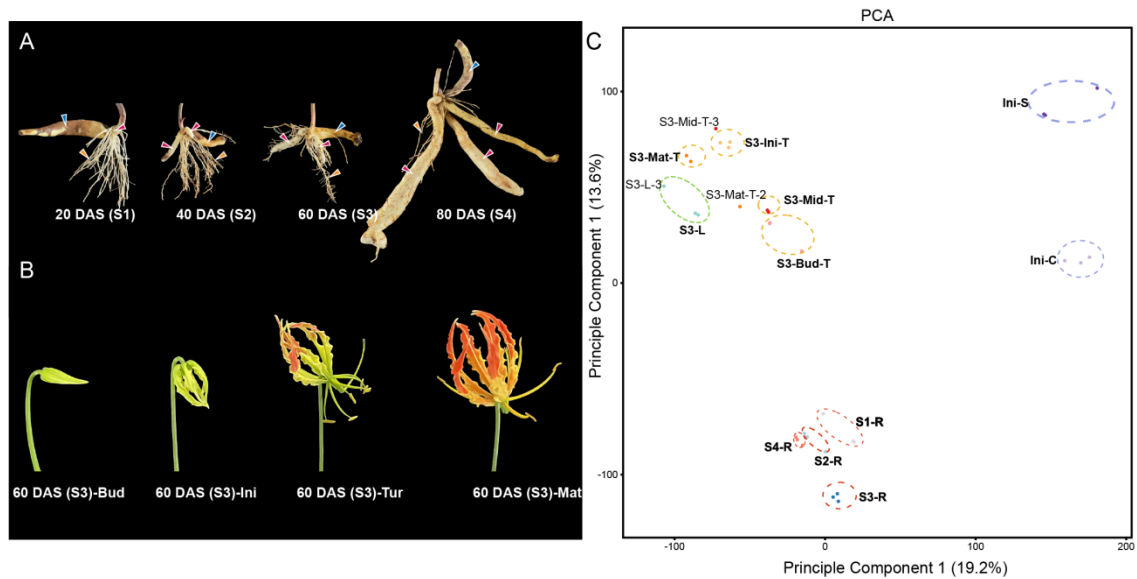
Comparison of C-terminal regions of *PI*-like MADS-box proteins (A) and AG-like MADS-box proteins (B).



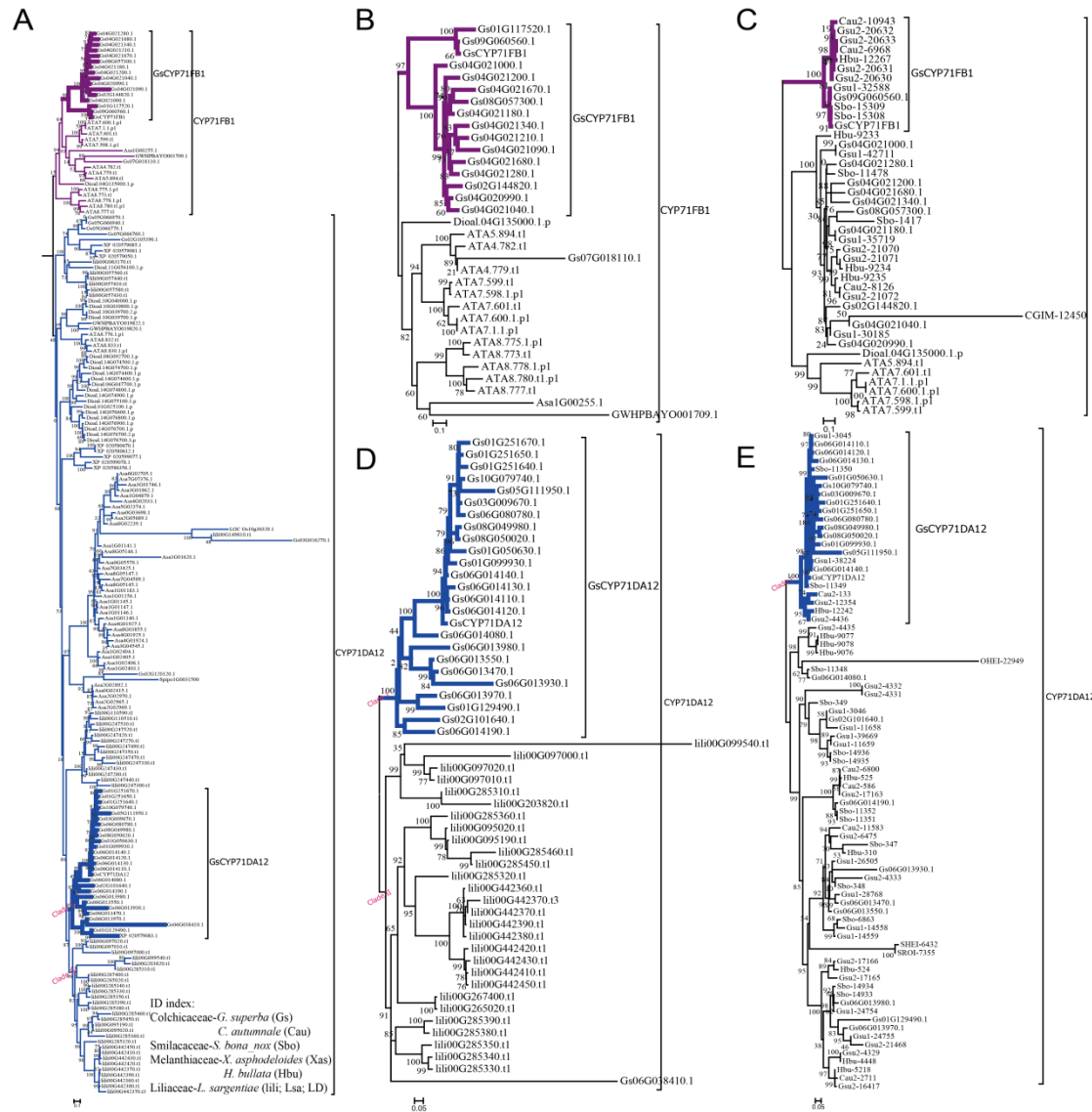
**Supplementary Figure 25. Phylogenetic tree of the SOC1 group of MADS-box genes.**



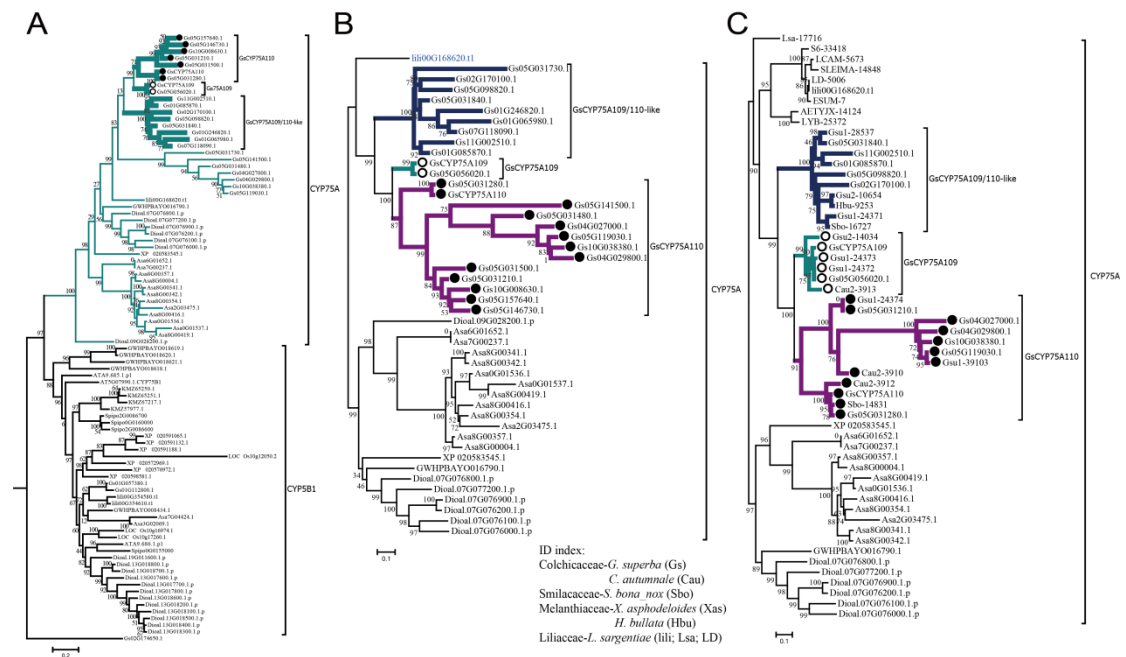
**Supplementary Figure 26. Weighted gene co-expression network analysis (WGCNA) of bulb dormancy in lily.** (A) Relationship between modules and traits, including dormancy bulb (DB) and dormancy-released bulb (DRB). Each column corresponds to a tissue, and each row corresponds to a gene module. Each cell contains the corresponding Pearson's correlation efficient  $r$  and  $P$ -value. The cell is color-coded by correlation efficient according to the color legend on the right of the heatmap. (B) Venn diagram of genes of the turquoise module generated by WGCNA and differentially expressed genes (DEGs) between DB and DRB. (C) GO enrichment of genes of the turquoise module. GO terms related to cell cycle are in red fonts. The turquoise module was considered to be dormancy-related because it showed a significant ( $P < 0.01$ ) positive correlation ( $r = 1$ ) with DRB. (D) Heatmap of expression patterns of candidate genes involved in lily bulb dormancy. Source data are provided as a Source Data file.



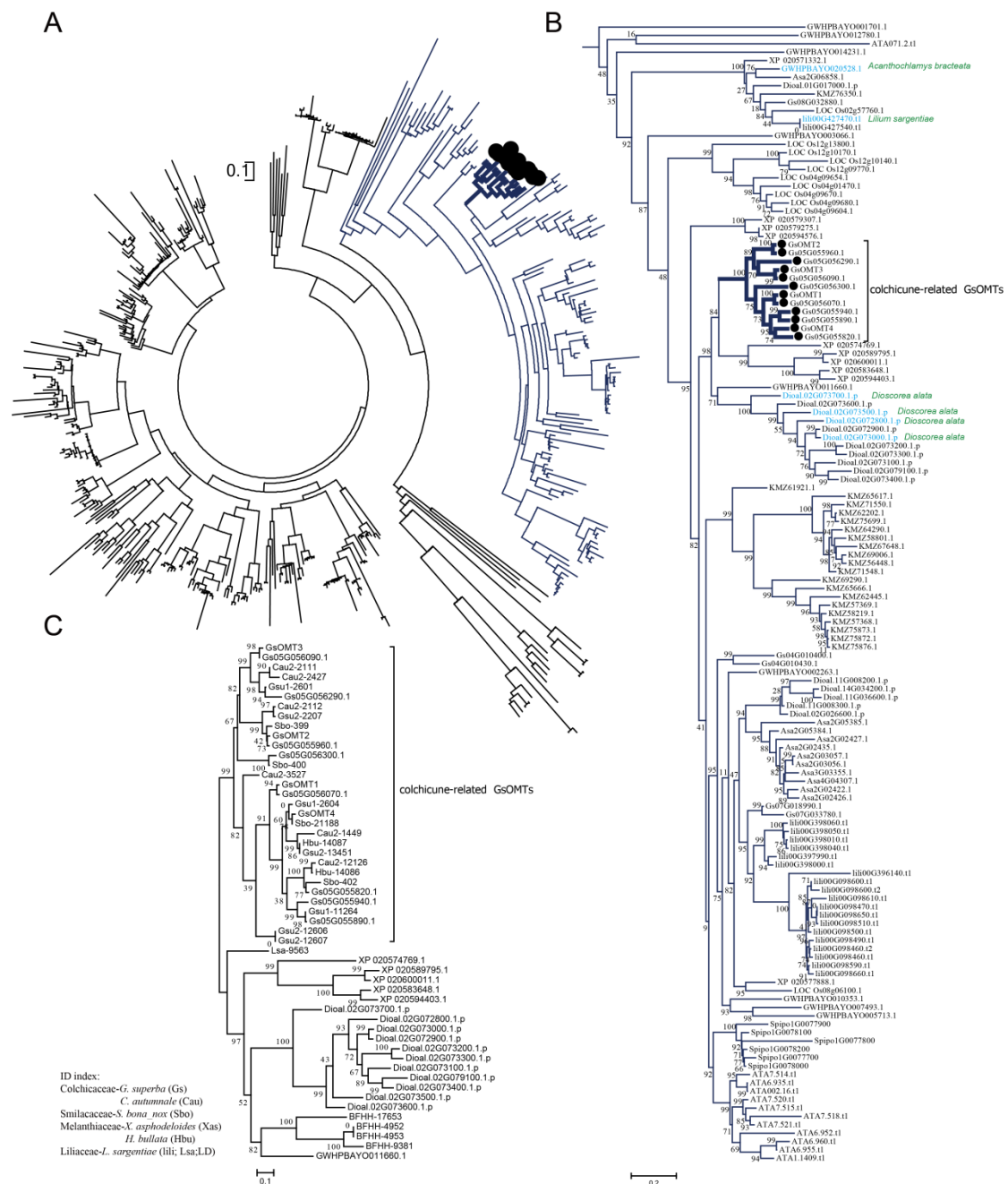
**Supplementary Figure 27. RNA-seq samples of flame lily.** (A) Characteristics of roots at sampling time points of 20, 40, 60, and 80 days after sprouting (DAS) in *Gloriosa superba*. Pink arrows indicate developing rhizomes (DRs), blue arrows indicate senescent rhizomes (SRs), and yellow arrows indicate fibrous roots (FRs). (B) Characteristics of flowers of different development stages (C) Principal component analysis of samples used in RNA sequencing of *G. superba*. Sampling time points of 20, 40, 60, and 80 days after sprouting (DAS) in *G. superba* were named S1, S2, S3, and S4, respectively. Flowers on the bud stage (Bud), initial opening stage (Ini), turning stage (Tur), and mature stage (Mat) were sampled from the *G. superba* plants at the time of 60 DAS. L, leaf; T, Tepal; R, rhizome.



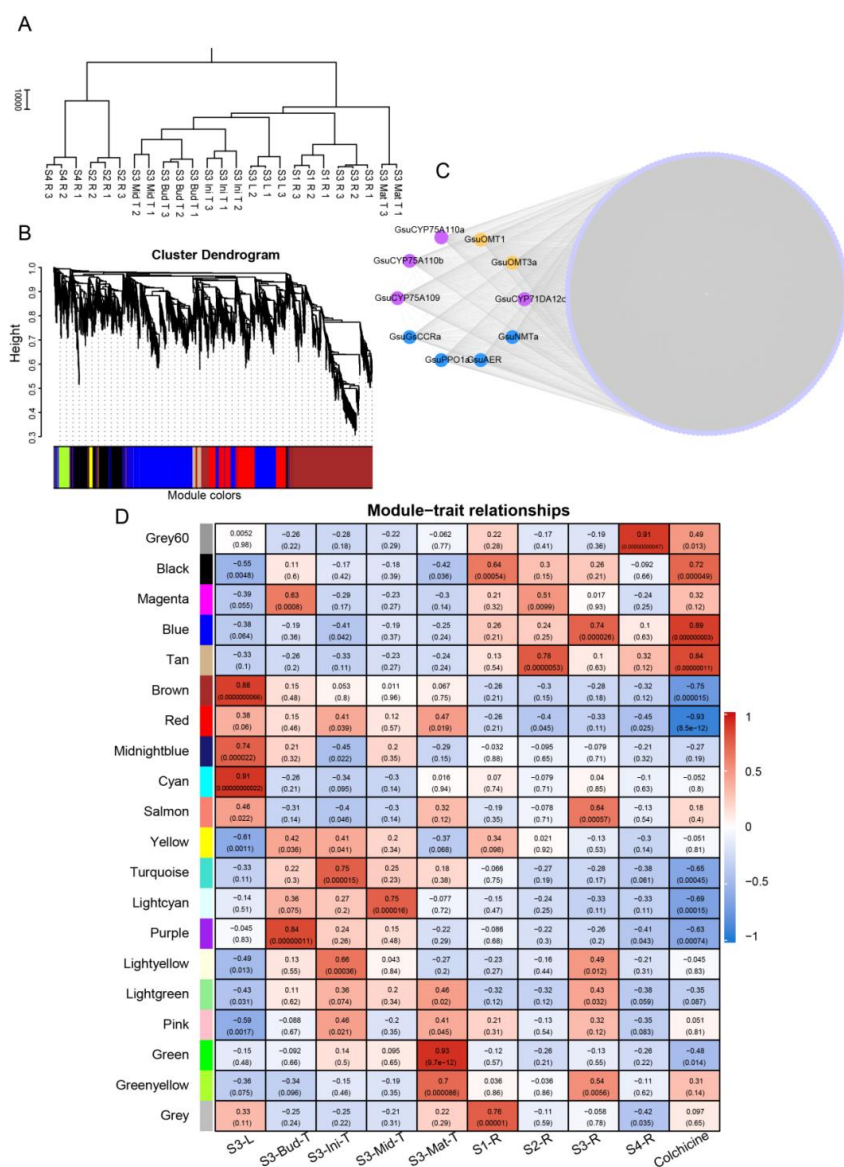
**Supplementary Figure 28. Identification of colchicine-related members of the cytochrome P450 (P450) gene family.** Identification of homologs of *GsCYP71FB1* (A, B, C) and *GsCYP71DA12* (A, D, E). (A) is the subtree extracted from the phylogenetic tree of P450 gene family. The RNA-seq sequences of a total of 57 *Lilium* plants and other 5 species from Colchicaceae, Smilacaceae, and Melanthiaceae were used in (C) and (E).



**Supplementary Figure 29. Identification of homologs of *GsCYP75A110* and *GsCYP75A109*.** (A) was the subtree extracted from the phylogenetic tree of P450 gene family. The lily CYP75A gene identified within the syntenic genomic region of the colchicine gene cluster is in blue font in (B). The RNA-seq sequences of a total of 71 transcriptome assemblies of wild Liliales species were used in (C).



**Supplementary Figure 30. Identification of the O-methyltransferase (OMT) gene family.** (A) Phylogenetic analysis of the OMT gene family. (B) The clade related to GsOMTs, extracted from (A), highlights colchicine-related OMT genes, including homologs of *GsOMT1*, *GsOMT2*, *GsOMT3*, and *GsOMT4*. The OMT genes identified in other monocot plant species except for flame lily within the syntenic genomic region of the CYP75A109-OMTs gene cluster are in blue fonts. (C) Identification of colchicine-related OMT genes across multiple Liliaceae species. The RNA-seq sequences of a total of 57 *Lilium* plants and other 5 species from Colchicaceae, Smilacaceae, and Melanthiaceae were used in C.

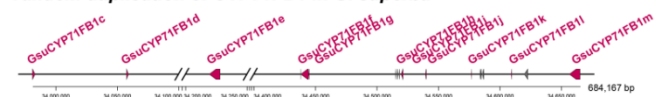


**Supplementary Figure 31. Weighted gene co-expression network analysis (WGCNA) of colchicine biosynthesis in flame lily.** (A) Clustering of RNA-seq samples used in WGCNA. (B) Hierarchical cluster dendrogram using dynamic tree cut, revealing 20 modules generated by WGCNA. The different colors under the dendrogram show co-expressed modules identified using WGCNA. (C) The co-expression network of the magenta module generated by WGCNA where multiple colchicine-related genes are co-expressed, which shows the correlation of colchicine-related genes. (D) Relationship between modules and traits. The traits includes different flame lily tissues across development as well as the relative colchicine content in the rhizome of flame lily of 80 days after sprouting (DAS). Each column corresponds to a tissue, and each row corresponds to a gene module. Each cell contains the corresponding Pearson's correlation efficient  $r$  and  $P$ -value. The cell is color-coded by correlation according to the color legend on the right of the heatmap. The used tissues (from left to right of the horizontal axis of the heatmap) include the leaves from flame lilies of DAS, flowers on the bud stage (Bud), initial opening stage (Ini), turning stage (Tur), and mature stage (Mat) from flame lilies of 60 DAS, and rhizomes of 20, 40, 60, and 80 DAS. Source data are provided as a Source Data file.

**Tandem duplication of CYP71DA12 in *G. superba***



**Tandem duplication of CYP71FB1 in *G. superba***



**Gene cluster of CYP75A109 and OMTs in *G. superba***

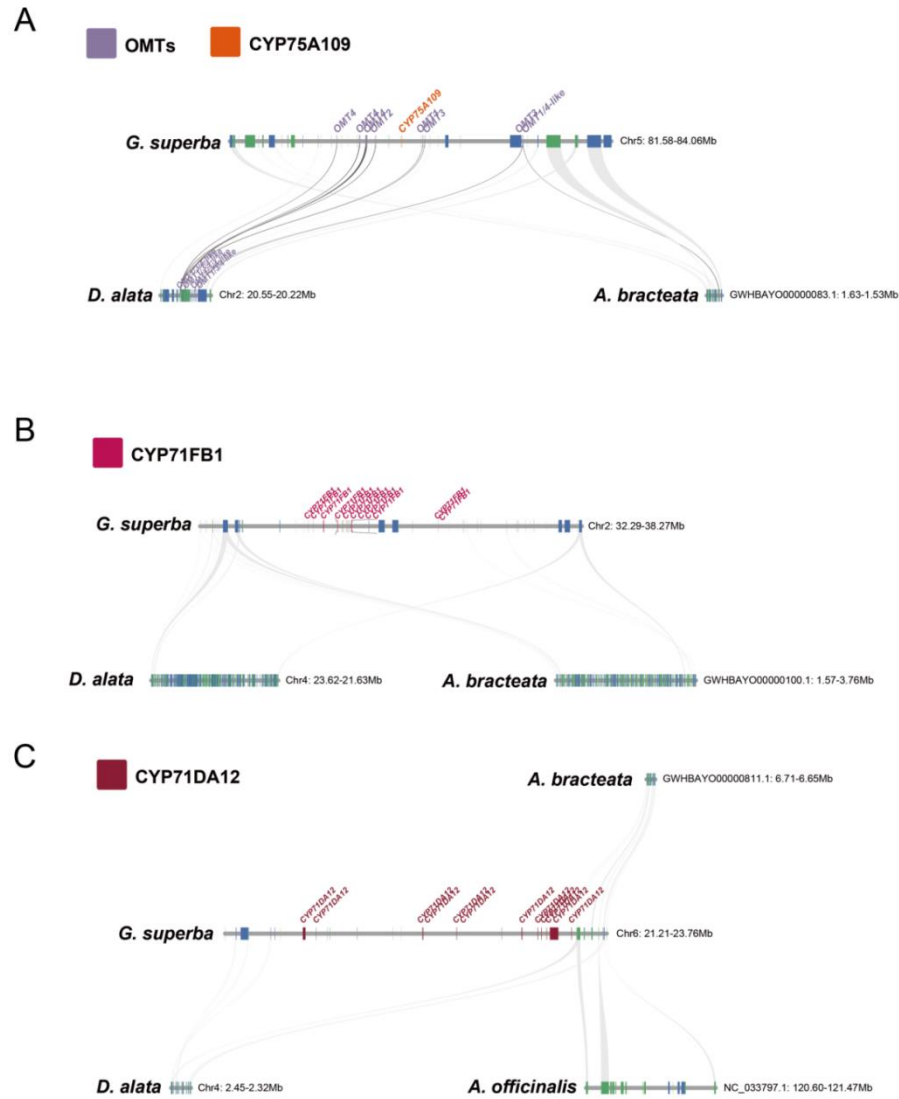


**Type**

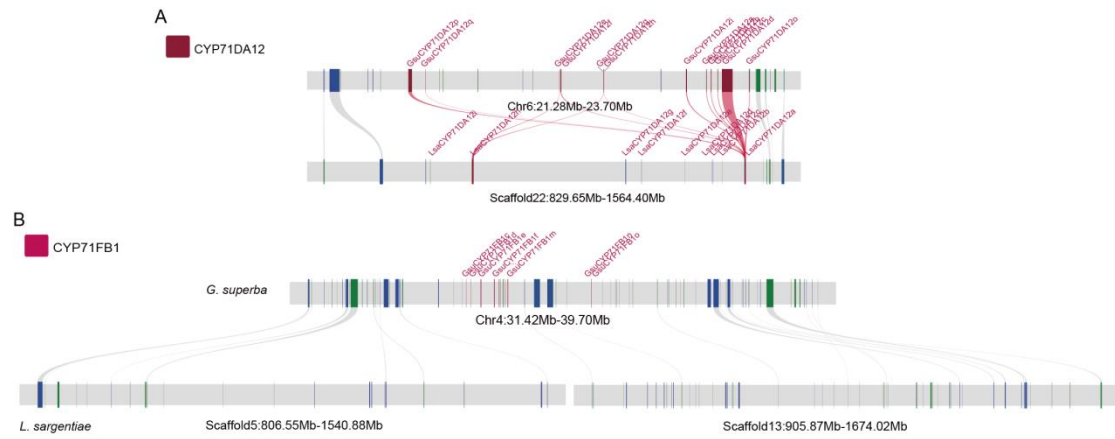
- CYP71DA12
- Others
- CYP71FB1
- OMT
- CYP75A109

20kb

**Supplementary Figure 32. Gene clusters of colchicine biosynthesis in *Gloriosa superba*.**

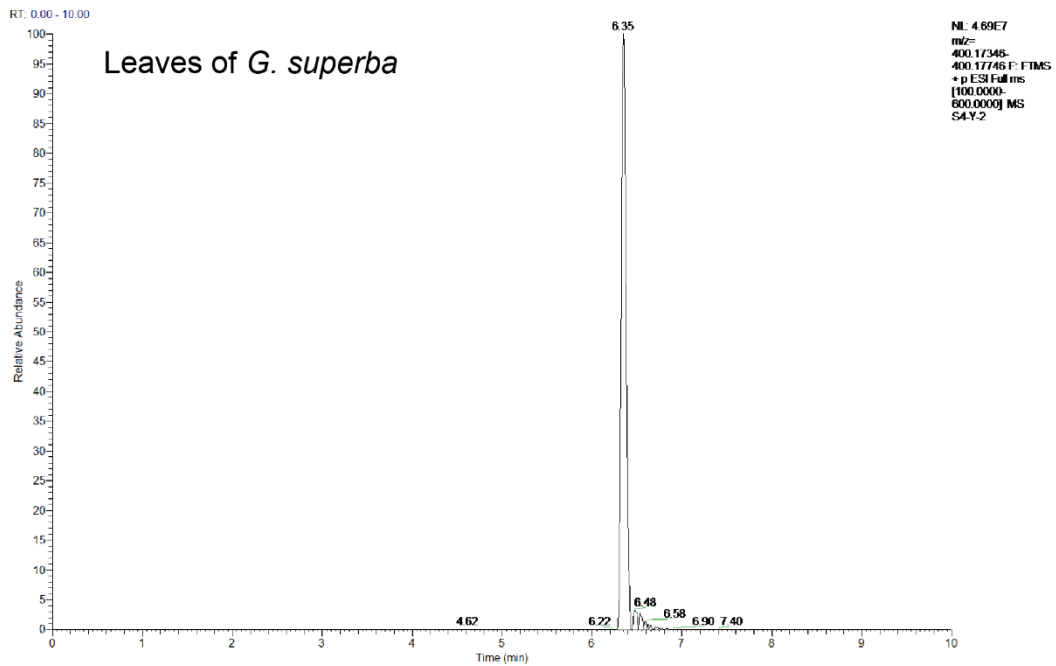


**Supplementary Figure 33. Microsynteny analysis of colchicine-related gene clusters in flame lily.** Microsynteny analysis of the colchicine-related OMT genes and CYP75A109 (A), CYP71FB1 (B), and CYP71DA12 (C) genes among *Gloriosa superba*, *Lilium sargentiae*, *Dioscorea alata*, and *Acanthochlamys bracteata*.

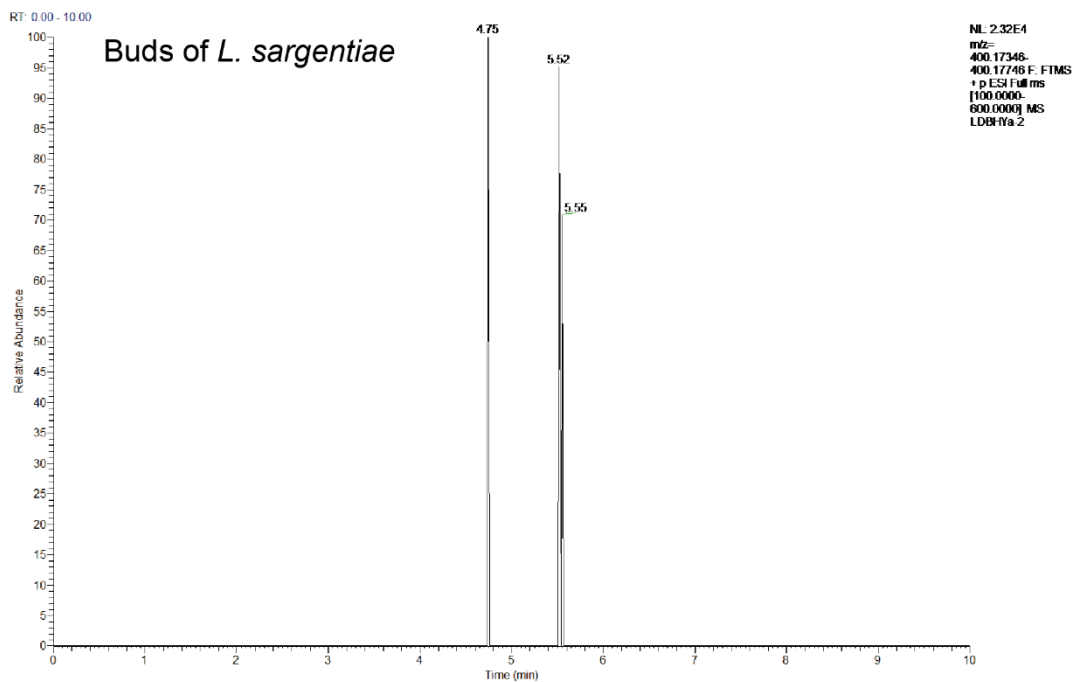


**Supplementary Figure 34. Microsynteny analysis of colchicine-related tandem arrays in flame lily.** Microsynteny analysis of CYP71DA12 (A) and CYP71FB1 (B) and genes between *Gloriosa superba* and *Lilium sargentiae*.

A



B



**Supplementary Figure 35. Colchicine detection in flame lily and lily.** Detection and identification of colchicine in the leaves of *Gloriosa superba* (A) and the buds of *Lilium sargentiae* (B). Mass tolerance is within 5.0 ppm. The calculated molecular weight of colchicine in the positive ion mode is 400.17546.

**Supplementary Table 1. Genome assembly statistics of *Lilium sargentiae*.**

	Contig		Scaffold	
	Length (bp)	Number	Length (bp)	Number
Longest reads length	60,480,349		5,309,664,549	
N10	2,481,798	1,236	5,309,664,549	1
N20	1,787,646	3,548	5,205,347,602	2
N30	1,411,064	6,581	3,792,423,142	3
N40	1,138,171	10,364	2,970,388,937	5
N50	929,352	15,019	2,845,141,673	6
N60	753,141	20,731	2,327,064,200	8
N70	593,946	27,867	2,119,186,429	10
N80	440,331	37,175	1,933,104,692	13
N90	279,343	50,670	1,294,474,416	16
Total length	47,833,891,648		47,092,191,210	
Number $\geq$ 100 bp		85,041		35
Number $\geq$ 2,000 bp		85,041		35
GC rate	0.322		0.317	

**Supplementary Table 2. Scaffold length statistics of *Lilium sargentiae*.**

<b>Scaffold</b>	<b>Length (bp)</b>	<b>Scaffold</b>	<b>Length (bp)</b>	<b>Scaffold</b>	<b>Length (bp)</b>
Scaffold1	2,032,120,186	Scaffold13	2,156,195,688	Scaffold25	11,342,490
Scaffold2	2,845,141,673	Scaffold14	1,731,449,773	Scaffold26	174,617,907
Scaffold3	3,551,658,967	Scaffold15	3,792,423,142	Scaffold27	465,339,060
Scaffold4	5,205,347,602	Scaffold16	1,933,104,692	Scaffold28	419,303,571
Scaffold5	2,790,071,271	Scaffold17	2,058,904,771	Scaffold29	122,730,325
Scaffold6	175,874,487	Scaffold18	5,309,664,549	Scaffold30	47,210,273
Scaffold7	95,418,508	Scaffold19	1,487,956,767	Scaffold31	146,586,096
Scaffold8	209,350,961	Scaffold20	631,479,043	Scaffold32	153,717,186
Scaffold9	75,178,950	Scaffold21	2,970,388,937	Scaffold33	1,294,474,416
Scaffold10	86,796,608	Scaffold22	2,327,064,200	Scaffold34	2,119,186,429
Scaffold11	59,754,209	Scaffold23	224,691,177	Scaffold35	79,479,384
Scaffold12	49,876,872	Scaffold24	258,291,040		

**Supplementary Table 3. Genome assembly statistics of *Gloriosa superba*.**

	Contig		Scaffold	
	Length (bp)	Number	Length (bp)	Number
Longest reads length	4,643,141		933,321,873	
N10	1,128,891	344	933,321,873	1
N20	845,313	876	616,235,262	2
N30	686,692	1,554	616,235,262	2
N40	570,690	2,377	582,368,806	3
N50	481,812	3,357	561,520,468	4
N60	407,741	4,514	490,290,024	5
N70	336,083	5,900	449,022,805	6
N80	266,255	7,613	361,845,312	8
N90	196,835	9,838	314,643,379	9
Total length	5,134,916,349		5,145,710,378	
Number>=100 bp		14,810		64
Number>=2,000 bp		14,810		64
GC rate	0.424		0.423	

**Supplementary Table 4. Chromosome length statistics of *Gloriosa superba*.**

<b>Chromosome</b>	<b>Length (bp)</b>
Chr1	933,321,873
Chr2	616,235,262
Chr3	582,368,806
Chr4	561,520,468
Chr5	490,290,024
Chr6	449,022,805
Chr7	405,278,612
Chr8	361,845,312
Chr9	314,643,379
Chr10	277,140,901
Chr11	133,886,167

**Supplementary Table 5. BUSCO evaluation of the genomic completeness of *Lilium sargentiae*.**

<b>Type</b>	<b>Genome</b>	<b>%</b>	<b>Protein</b>	<b>%</b>
Complete BUSCOs (C)	1465	90.77%	1425	88.30%
Complete and single-copy BUSCOs (S)	865	53.59%	682	42.30%
Complete and duplicated BUSCOs (D)	600	37.17%	743	46.00%
Fragmented BUSCOs (F)	41	2.54%	56	3.50%
Missing BUSCOs (M)	108	6.69%	133	8.20%
Total BUSCO groups searched	1614		1614	

# BUSCO version is: 5.4.0

# The lineage dataset is: embryophyta\_odb10 (Creation date: 2024-01-08, number of genomes: 50, number of BUSCOs: 1614)

**Supplementary Table 6. BUSCO evaluation of the genomic completeness of *Gloriosa superba*.**

Type	Genome	%	Protein	%
Complete BUSCOs (C)	1481	91.76%	1387	85.90%
Complete and single-copy BUSCOs (S)	1411	87.42%	1309	81.10%
Complete and duplicated BUSCOs (D)	70	4.34%	78	4.80%
Fragmented BUSCOs (F)	52	3.22%	107	6.60%
Missing BUSCOs (M)	81	5.02%	120	7.50%
Total BUSCO groups searched	1614		1614	

# BUSCO version is: 5.4.0

# The lineage dataset is: embryophyta\_odb10 (Creation date: 2024-01-08, number of genomes: 50, number of BUSCOs: 1614)

**Supplementary Table 7. Statistics of genomic sequence features of *Lilium sargentiae*.**

Type	Count
Gene count	47,139
Maximum gene length	439,555 bp
Minimum gene length	261 bp
Median/Average of gene length	23945/42804 bp
mRNA count	51,523
Maximum mRNA length	14,580 bp
Minimum mRNA length	201 bp
Median/Average of mRNA length	1137/1445 bp
Median/Average of mRNA count	1/1.09
CDS count	51,523
Maximum CDS length	14,580 bp
Minimum CDS length	201 bp
Median/Average of CDS length	1137/1445 bp
Median/Average of protein length	379/482 bp
Median/Average of intron length	498/9070 bp
Median/Average of intron count	3/4.51

**Supplementary Table 8. Statistics of genomic sequence features of *Gloriosa superba*.**

Type	Count
Gene count	56,639
Maximum gene length	435,182 bp
Minimum gene length	51
Median/Average of gene length	636/9,984 bp
mRNA count	56639
Maximum mRNA length	14316 bp
Minimum mRNA length	51 bp
Median/Average of mRNA length	534/802 bp
Median/Average of mRNA count	1/1.00
CDS count	56639
Maximum CDS length	14316 bp
Minimum CDS length	51 bp
Median/Average of CDS length	534/802 bp
Median/Average of protein length	178/267 bp
Median/Average of intron length	213/4979 bp
Median/Average of intron count	0/1.84

**Supplementary Table 9. Statistics of genomic sequence features of representative species.**

Type	<i>Oryza sativa</i>	<i>Zea mays</i>	<i>Triticum aestivum</i>	<i>Glycine max</i>	<i>Arabidopsis thaliana</i>	Human	<i>Lilium sargentiae</i>	<i>Gloriosa superba</i>
Gene count	55,986	39,756	120,744	57,147	37,513	62,703	45,419	54,617
Maximum gene length (bp)	57,094	751,401	124,945	94,722	58,519	2,473,539	439,555	434,638
Minimum gene length (bp)	84	213	42	42	22	8	272	147
Median/Average of gene length (bp)	2393/2965.12	2710/447.6.79	2031/3133.43	2860/3775.02	1850/2161.83	3605/32370.78	22197/40117.12	611/8874.39
mRNA count	66860	452083	762086	525918	41671	117725	49,803	54612
Maximum mRNA length (bp)	16311	10979	46252	28210	16347	347561	14,580	14316
Minimum mRNA length (bp)	3	1	1	1	22	8	201	147
Median/Average of mRNA length (bp)	1464/1694.83	148/291.33	152/300.74	147/341.64	1361/1556.65	1505/2147.04	1128/1440.39	519/782.32
Median/Average of mRNA count	1/1.18	1/1.00	1/1.00	1/1.00	1/1.00	1/1.88	1/1.10	1/1.00
CDS count	66,860	72,539	133,346	88,412	35,386	64,538	49,803	54,617
Maximum CDS length (bp)	16,311	16,278	16,080	16,308	16,182	107,976	14,580	14,316
Minimum CDS length (bp)	3	87	54	87	22	8	201	147
Median/Average of CDS length (bp)	999/1331.53	1014/1206.40	1131/1332.42	1056/1275.48	1050/1230.62	1215/1607.49	1128/1440.39	519/782.29
Median/Average of protein length (bp)	333/443.84	338/402.13	377/444.14	352/425.16	350/410.21	405/535.83	376/480.13	173/260.76
Median/Average of exon length (bp)	178/362.62	148/291.33	152/300.74	147/341.64	150/298.67	133/296.13	136/268.37	171/285.57
Median/Average of exon count	3/4.67	1/1.00	1/1.00	1/1.00	3/5.21	4/7.25	4/5.37	1/2.74

**Supplementary Table 10. Statistics of repeat sequences of *Lilium sargentiae*.**

Type	Length (bp)	Rate (%)
<b>Class I: Retrotransposon</b>	<b>23,422,933,505</b>	<b>49.74%</b>
<b>LTR-Retrotransposon</b>	<b>22,319,951,664</b>	<b>47.40%</b>
LTR/Copia	6,864,245,374	14.58%
LTR/Gypsy	15,161,340,188	32.20%
LTR-other	294,366,102	0.63%
<b>Non-LTR Retrotransposon</b>	<b>1,102,981,841</b>	<b>2.34%</b>
SINE	127,134,377	0.27%
LINE	975,847,464	2.07%
<b>Class II: DNA Transposon</b>	<b>2,758,560,438</b>	<b>5.86%</b>
EnSpm/CACTA	1,763,161,088	3.74%
Harbinger	263,017,356	0.56%
Helitron	221,268,853	0.47%
MuDR	301,971,260	0.64%
Tcl/Mariner	52,529,547	0.11%
hAT	124,057,818	0.26%
DNA-other	32,554,516	0.07%
<b>Low Complexity</b>	<b>2,483,693</b>	<b>0.01%</b>
<b>Tandem repeat</b>	<b>21,655,675</b>	<b>0.05%</b>
<b>Unclassified</b>	<b>12,559,879,578</b>	<b>26.67%</b>
<b>Total content</b>	<b>38,765,512,889</b>	<b>82.32%</b>

**Supplementary Table 11. Statistics of repeat sequences of *Gloriosa superba*.**

<b>Type</b>	<b>Length (bp)</b>	<b>Rate (%)</b>
<b>Class I: Retrotransposon</b>	3,033,094,138	58.94%
LTR-Retrotransposon	2,978,566,345	57.88%
LTR/Copia	774,177,976	15.05%
LTR/Gypsy	2,185,123,440	42.46%
LTR-other	19,264,929	0.37%
Non-LTR Retrotransposon	54,527,793	1.06%
SINE	0	0.00%
LINE	54,527,793	1.06%
<b>Class II: DNA Transposon</b>	232,433,893	4.52%
EnSpm/CACTA	46,236,668	0.90%
Harbinger	8,072,857	0.16%
Helitron	5,192,145	0.10%
MuDR	97,339,147	1.89%
Tcl/Mariner	0	0.00%
hAT	72,112,927	1.40%
DNA-other	3,480,149	0.07%
<b>Low Complexity</b>	36,499,938	0.71%
<b>Tandem repeat</b>	1,494,736	0.03%
<b>Unclassified</b>	1,146,591,507	22.28%
<b>Total content</b>	<b>4,450,114,212</b>	<b>86.48%</b>

**Supplementary Table 12. Statistics of LTR-RT distribution of *Lilium sargentia* and *Gloriosa superba*.**

<b>Species</b>	<b><i>Lilium sargentia</i></b>	<b><i>Gloriosa superba</i></b>
Number of LTR-RTs	23,719,431	1,395,216
Number of LTR-RTs in intergenic regions	23,160,279	1,325,468
Number of LTR-RTs in intergenic regions/Number of LTR-RTs	97.64%	95.00%
Number of LTR-RTs in genic regions	559,152	69,748
Number of LTR-RTs in genic regions/Number of LTR-RTs	2.36%	5.00%
Number of genes	45,419	54,617p
Number of genes with LTR-RTs	33,508	18,987
Number of genes with LTR-RTs/Number of genes	73.78%	34.76%
Number of genes with LTR-RTs in introns	33,222	8,452
Number of genes with LTR-RTs in introns/Number of genes	73.15%	15.48%

**Supplementary Table 13. Statistics of gene function annotation results of *Lilium sargentiae*.**

	<b>Number</b>	<b>Percent (%)</b>
NR	45113	99.85
Swiss-Prot	38030	84.17
TAIR	41731	92.36
Rice	42596	94.27
GO	15976	35.36
KEGG	20607	45.61
eggNOG	41542	91.94
Pfam	36406	80.57
All database	45183	95.85
Total	47139	

Note: NR, Non-Redundant Protein Sequence Database; Rice, Rice Genome Annotation Database; TAIR, Arabidopsis Information Resource; GO, Gene Ontology Database; KEGG, Kyoto Encyclopedia of Genes and Genomes Database; eggnog, Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups Database.

**Supplementary Table 14. Statistics of gene function annotation results of *Gloriosa superba*.**

	<b>Number</b>	<b>Percent (%)</b>
NR	54,214	99.26
Swiss-Prot	41,399	75.8
TAIR	43,779	80.15
Rice	50,075	91.68
GO	17,431	31.91
KEGG	22,784	41.71
eggNOG	49,647	90.9
Pfam	35,694	65.35
All database	54,380	99.57
Total	54,617	

Note: NR, Non-Redundant Protein Sequence Database; Rice, Rice Genome Annotation Database; TAIR, Arabidopsis Information Resource; GO, Gene Ontology Database; KEGG, Kyoto Encyclopedia of Genes and Genomes Database; eggnog, Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups Database.

**Supplementary Table 15. Primers used in qRT-PCR.**

<b>Symbol</b>	<b>Primer name</b>	<b>Primer sequence (5'-3')</b>
LaXTH	XTH-F	CCTTTGGAACGCCGACGACT
	XTH-R	GCAGTCTCTTGTGGCTCGCA
LoFUL	FUL-F	GCTGAAGAAGGCGCACGAGA
	FUL-R	AAGACGATGGCAGCGACCTC
FP	FP-F	TCGCCTACATCGCTAACC
	FP-R	TTCCCAATAATCGCAAGACC

Note: Forward primer, F; downward primer, R.