

Supplementary Information

Genomic and epigenomic insight into giga-chromosome architecture and adaptive evolution of Royal Lily (*Lilium regale*)

Jie Sun^{1,#}, Xiangfeng Wang^{1,#}, Ke Wang^{1,2}, Dian Meng¹, Yu Mu¹, Lili Zhang¹, Jingxuan Wang¹, Gang Yao¹, Li Guo^{1,§}

¹Shandong Key Laboratory of Precision Molecular Crop Design and Breeding, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Weifang, Shandong 261325, China.

²College of Life Sciences, Shandong Agricultural University, Tai'an, Shandong 271018, China.

[#]These authors contributed equally.

§ Correspondence: li.guo@pku-iaas.edu.cn

Supplementary Table 1: Hi-C interaction heatmaps for pseudomolecules Chr1 to Chr6. **a-f**, of *Lilium regale*.

Species	Libraries	Read length (bp)	Total data (Gb)	Depth (x)
<i>Lilium regale</i>	NGS	150	4397	124
	PacBio HiFi	/	806	23
	Hi-C	150	5460	154

Supplementary Table 2: Statistics of large plant genomes (~>10G) published in recent years.

Species	ContigN50 (bp)	Genome Size (Gb)	Year	Source
<i>Paeonia ostii</i>	228000	12.28	2022	Yuan et al.,2022
<i>Allium africanus</i>	7320000	20.2	2023	Hao et al., 2023
<i>Allium cepa</i> L.	81660000	15.9	2023	Hao et al., 2023
<i>Allium fistulosum</i> L.	507270000	11.7	2023	Hao et al., 2023
<i>Allium sativum</i> L.	109820000	16.5	2023	Hao et al., 2023
<i>Cycas panzhihuaensi</i>	12000000	10.5	2022	Liu et al., 2022
<i>Lilium davidii</i>	7720000	36.68	2024	Xu et al., 2024
<i>Lilium regale</i>	32225586	35.57	2024	This Study
<i>Lilium sargentiae</i>	929352	47.09	2025	Liang et al., 2025
<i>Pinus lambertiana</i> Dougl.	4250	31	2016	Stevens KA et al., 2016
<i>Pinus tabuliformis</i>	2601000	25.4	2020	Niu et al., 2022
<i>Sequoia sempervirens</i>	96800	26.4	2022	Neale DB et al., 2022
<i>Triticum aestivum</i>	30220000	14.68	2022	Athiyannan N et al., 2022

Supplementary Table 3: Chromosome lengths of the assembled lily genome.

Chromosome ID	GC content	Contig number	Length (bp)
Chr01	31.89%	208	4,300,821,314
Chr02	31.98%	210	3,979,246,691
Chr03	32.59%	172	3,190,401,390
Chr04	32.66%	151	3,144,698,111
Chr05	32.35%	126	2,789,839,680
Chr06	32.20%	166	2,714,556,596
Chr07	32.36%	173	2,685,516,968
Chr08	32.02%	120	2,534,867,411
Chr09	32.42%	121	2,498,420,967
Chr10	32.37%	145	2,351,978,224
Chr11	32.44%	113	2,225,902,421
Chr12	32.61%	113	2,017,080,394
TOTAL	32.29%	1818	34,433,330,167

Supplementary Table 4: Survey results of *Allium* type telomeric sequence (CTCGGTTATGGG) in lily genome assembly.

Chromosome ID	Telomere-start (bp)	Telomere-end (bp)	Telomere-length (bp)	Chromosome size (bp)
Chr04	891	1,819	928	3,144,698,111
Chr05	2,789,824,082	2,789,839,680 bp	15,598	2,789,839,680
Chr07	1	7,949	7,948	2,685,516,968
Chr08	9	5,282	5,273	2,534,867,411
Chr09	2,498,409,164	2,498,420,967	11,803	2,498,420,967
Chr10	1	10,109	10,108	2,351,978,224
Chr12	1	12,268	12,267	2,017,080,394

Supplementary Table 5: The quality of lily genome assembly assessed by Merqury.

<i>Lilium regale</i>	
QV (consensus quality value)	63.8
Completeness	72.06%
Heterozygosity	3.32%
Base accuracy	99.99996%

Supplementary Table 6: RNA sequencing (RNA-Seq) data summary.

	Transcriptome sequencing (NGS)	Transcriptome sequencing (PacBio)
Leaf	65.8 G	44.3 G
Stem	68.8 G	36.0 G
Bulb	74.4 G	46.9 G
Flower	70.0 G	36.8 G

Supplementary Table 7: The analysis of transposable elements in the lily genome.

	Number of elements*	length occupied(bp)	percentage of sequence ⁺
Retroelements	25,770,839	20,922,007,440	59.95%
SINEs:	343,476	49,919,961	0.14%
LINEs:	1,380,605	1,013,184,807	2.90%
CRE/SLACS	1,380,605	1,013,184,807	2.90%
L2/CR1/Rex	64,913	55,577,396	0.16%
RTE/Bov-B	515,921	261,359,698	0.75%
R1/LOA/Jockey	7,578	2,006,384	0.01%
R2/R4/NeSL	4,623	1,497,670	0.00%
RTE/Bov-B	196,218	112,814,868	0.32%
L1/CIN4	423,092	498,518,051	1.43%
LTR elements:	24,046,758	19,858,902,672	56.90%
Ty1/Copia	6,727,998	4,927,427,156	14.12%
Gypsy/DIRS1	16,730,972	14,620,930,226	41.89%
Retroviral	149,710	128,577,816	0.37%
DNA transposons	10,370,428	4,007,422,675	11.48%
hobo-Activator	3,513,285	954,781,914	2.74%
Tc1-IS630-Pogo	1,750,068	432,327,322	1.24%
MULE-MuDR	113,150	49,004,747	0.14%
Tourist/Harbinger	1,019,765	172,382,905	0.49%
Other (Mirage, P-element, Transib)	26,725	19,651,782	0.06%
Rolling-circles	56,896	23,098,690	0.07%
Unclassified:	8,587,739	2,541,854,912	7.28%
Total interspersed repeats:		27,471,285,027	78.71%
Small RNA:	109,371	32,598,315	0.09%
Satellites:	161,466	73,812,740	0.21%
Simple repeats:	6,294,702	301,943,555	0.87%
Low complexity:	858,634	50,228,821	0.14%

*most repeats fragmented by insertions or deletions have been counted as one element.

⁺percentage means one element takes part in of what proportion of whole genome.

Supplementary Table 8: The percentage of repeat elements in different species.

	DNA	LTR/Gypsy	LTR/Copia	other LTR	LINE	SINE	Other	Total
<i>Lilium regale</i>	11.48	41.89	14.12	0.89	2.90	0.14	8.64	80.06
<i>Arabidopsis thaliana</i>	6.45	5.14	1.45	0.22	1.06	0.02	12.22	26.56
<i>Oryza Sativa</i> L.	17.67	23.71	3.33	0.15	0.42	0.72	0.22	46.22
<i>Zea mays</i>	9.78	45.51	25.27	4.03	0.70	0.01	3.07	88.37
<i>Allium fistulosum</i>	17.02	35.98	8.08	0.69	2.86	0.06	16.03	80.72
<i>Lilium davidii</i>	4.53		64.40		1.70	0.07	17.61	88.31
<i>Lilium sargentiae</i>	5.86	32.20	14.58	0.63	2.07	0.27	26.71	82.32
<i>Gloriosa superba</i>	4.52	42.46	15.05	0.37	1.06	0.00	23.02	86.48
<i>Homo sapiens</i>	3.58		8.84		20.68	12.78	8.07	53.94

Supplementary Table 9: The stats of LTR elements in the lily genome.

Type	Number of elements*	length occupied(bp)	percentage of sequence
intact LTRs	79,821	650,877,132	3.28%
truncated LTRs	4,249,484	7,129,800,897	35.90%
solo LTRs	19,230,536	9,111,917,037	45.88%

*most repeats fragmented by insertions or deletions have been counted as one element.

Supplementary Table 10: Statistics for gene annotation of the assembled genomes.

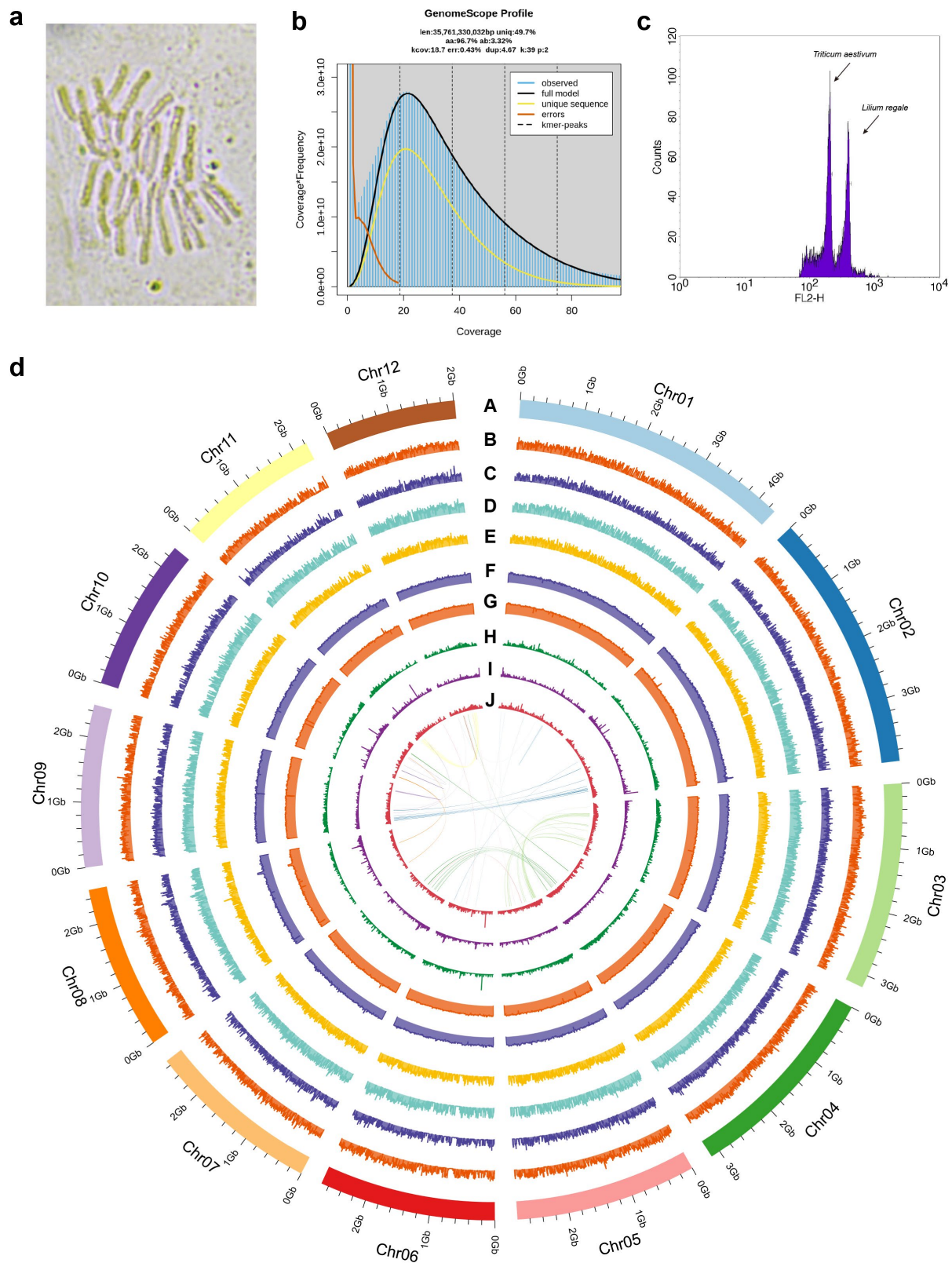
Species	<i>Lilium regale</i>	<i>Arabidopsis thaliana</i>	<i>Oryza Sativa L.</i>	<i>Zea mays</i>	<i>Triticum monococcum</i>
Gene number	67,862	27,561	60,935	42,580	37,313
Average gene length (bp)	19,940	2,376	2,959	4,409	5,809
Average exons per gene	5.00	5.16	3.94	4.83	5.16
Average exon length (bp)	232	331	411	265	318
Average intron length (bp)	4,680	71	439	815	1,001

Supplementary Table 11: Statistics for alternative splicing events of the different species.

Species	<i>Arabidopsis thaliana</i>	<i>Oryza sativa L.</i>	<i>Triticum</i>	<i>Lilium regale</i>	<i>Homo sapiens</i>
RI	10,622 (48.46%)	659 (48.46%)	16.00%	707 (27.40%)	19,703 (17.69%)
A3	6,784 (30.95%)	223 (48.46%)	33.00%	843 (32.67%)	23,026 (20.67%)
A5	3,314 (15.12%)	138 (48.46%)	19.00%	391 (15.16%)	17,443 (15.66%)
SE	1200 (5.47%)	89 (48.46%)	13.00%	364 (14.11%)	34,762 (31.21%)
AF	-	-	13.00%	143 (5.54%)	6,699 (6.01%)
AL	-	-	5.00%	121 (4.69%)	3,237 (2.91%)
MX	-	5 (0.45%)	0.60%	11 (0.43%)	6,512 (5.85%)
Total	21,920	1,104	100.00%	2,580	111,382

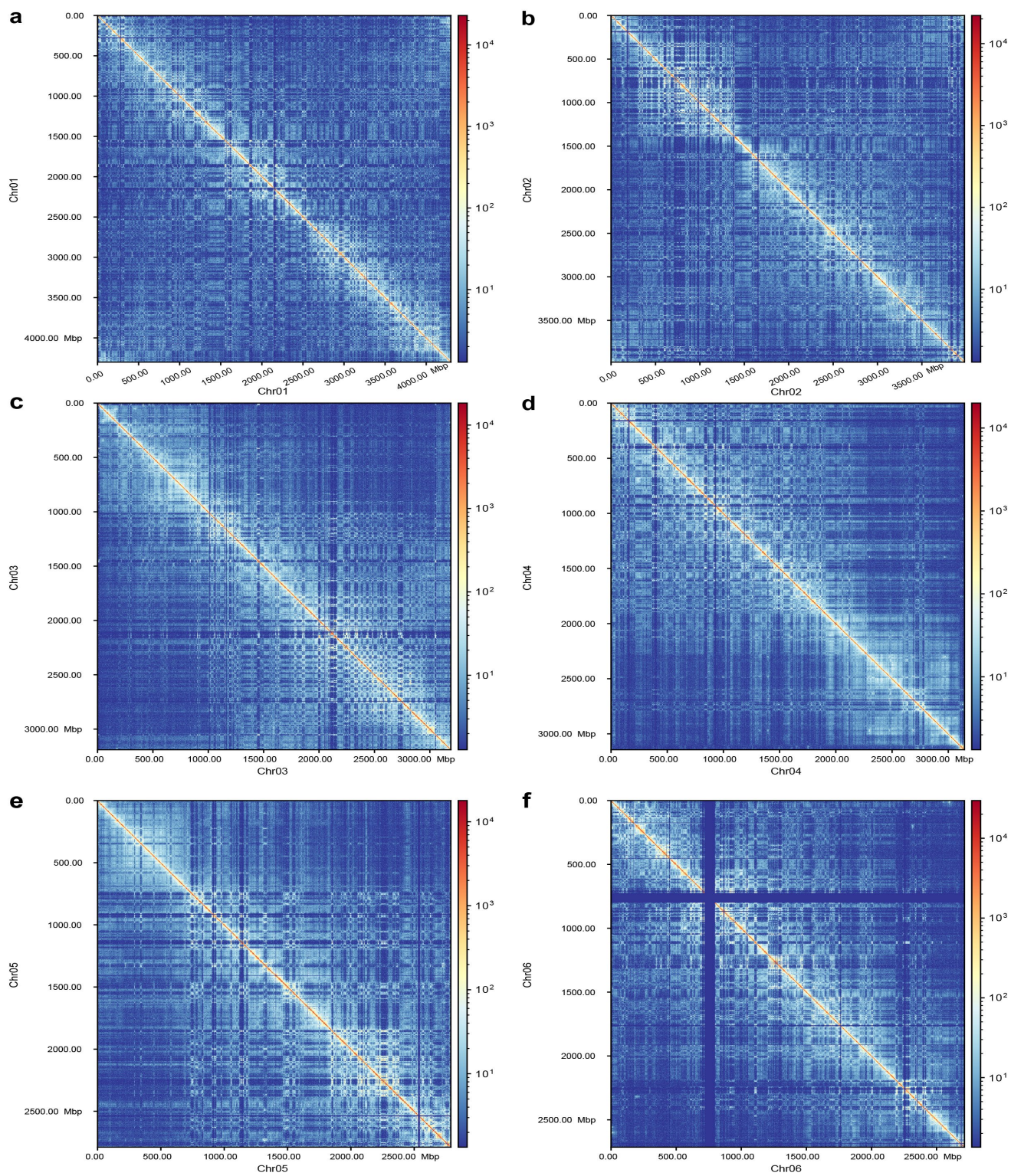
Supplementary Table 12: Species for Phylogenetic analysis..

Species	Reference
<i>Zea mays</i>	Chen, J. <i>et al.</i> Nat Genet (2023)
<i>Ananas comosus</i>	Ming, R. <i>et al.</i> Nat Genet (2015)
<i>Musa acuminata</i>	Huang, H.R. <i>et al.</i> Horticulture Res (2023)
<i>Apostasia shenzhenica</i>	Zhang, G.Q. <i>et al.</i> Nature (2017)
<i>Cymbidium ensifolium</i>	Ai, Y. <i>et al.</i> Horticulture Res (2021)
<i>Phalaenopsis aphrodite</i>	Chao, Y.T. <i>et al.</i> Plant Biotechnology J (2018)
<i>Asparagus kiusianus</i>	Shirasawa, K. <i>et al.</i> DNA Res (2022)
<i>Allium sativum</i>	Sun, X. <i>et al.</i> Mol Plant (2020)
<i>Allium cepa</i>	Finkers, R. <i>et al.</i> G3 (2021)
<i>Allium fistulosum</i>	Liao, N. <i>et al.</i> Nat Commun (2022).
<i>Dioscorea alata</i>	Bredeson, J.V. <i>et al.</i> Nat Commun (2022)
<i>Gloriosa superba</i>	Liang, Y. <i>et al.</i> Nat Commun (2025)
<i>Lilium davidii</i>	Xu, S.J. <i>et al.</i> Innovation (Camb) (2024)
<i>Lilium sargentiae</i>	Liang, Y. <i>et al.</i> Nat Commun (2025)
<i>Lilium regale</i>	This study
<i>Papaver somniferum</i>	Guo, L. <i>et al.</i> Science (2018).

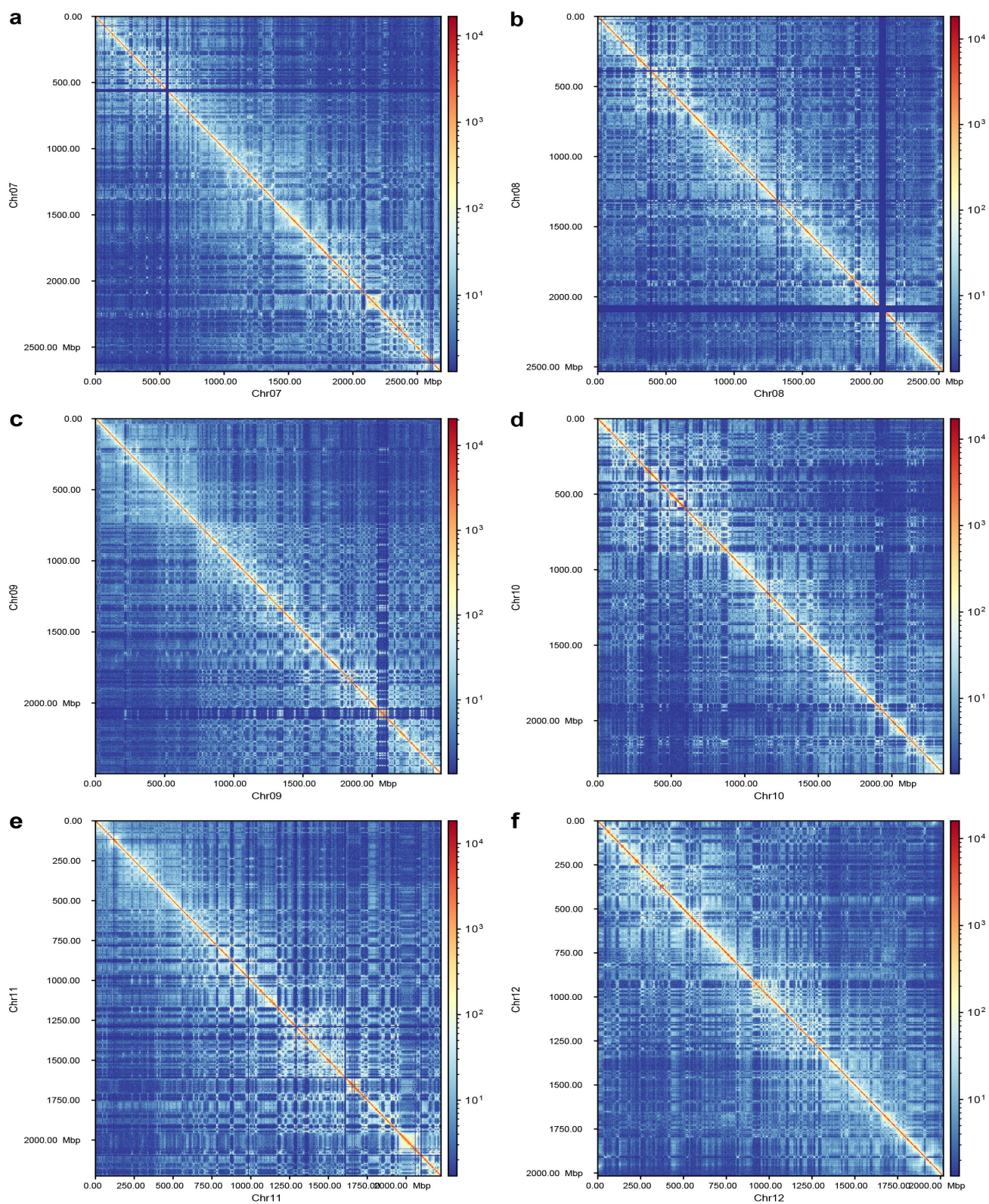


Supplementary Fig. 1: Additional genomic characteristics of *Lilium regale*. **a**, The microscopic identification of karyotype ($2n=24$) of *L. regale*. **b-c**, Estimation of *L. regale* genome size by K-mer (**b**) and flow cytometry analysis (**c**). **d**, Circos plot showing the transcriptomic and epigenomic characteristics for *L. regale*. Track A to H: chromosomes, flower RNA-seq, leaf RNA-seq, stem

RNA-seq, bulb RNA-seq, leaf DNA methylation, flower DNA methylation, H3K4Me3, H3K9Me2 and H3K27Me3, and color ribbons representing genome-wide syntenic blocks.

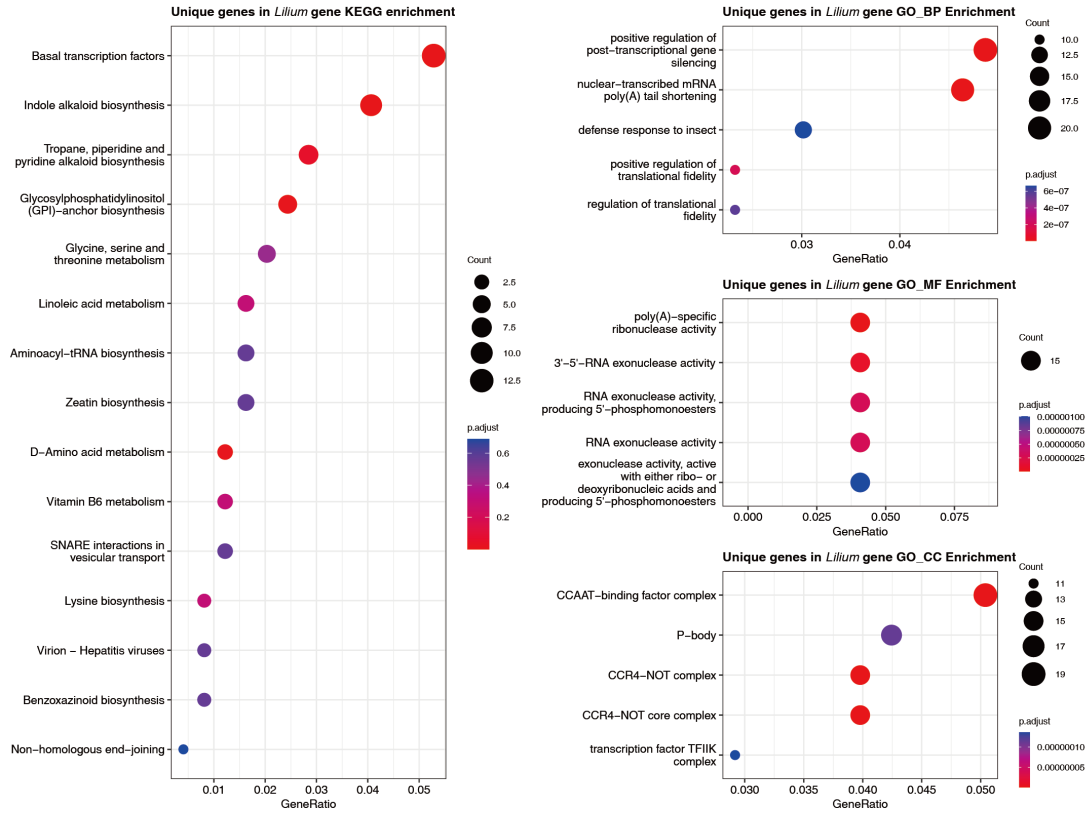


Supplementary Fig. 2: Hi-C interaction heatmaps for pseudomolecules Chr1 to Chr6. a-f, of *Lilium regale*.

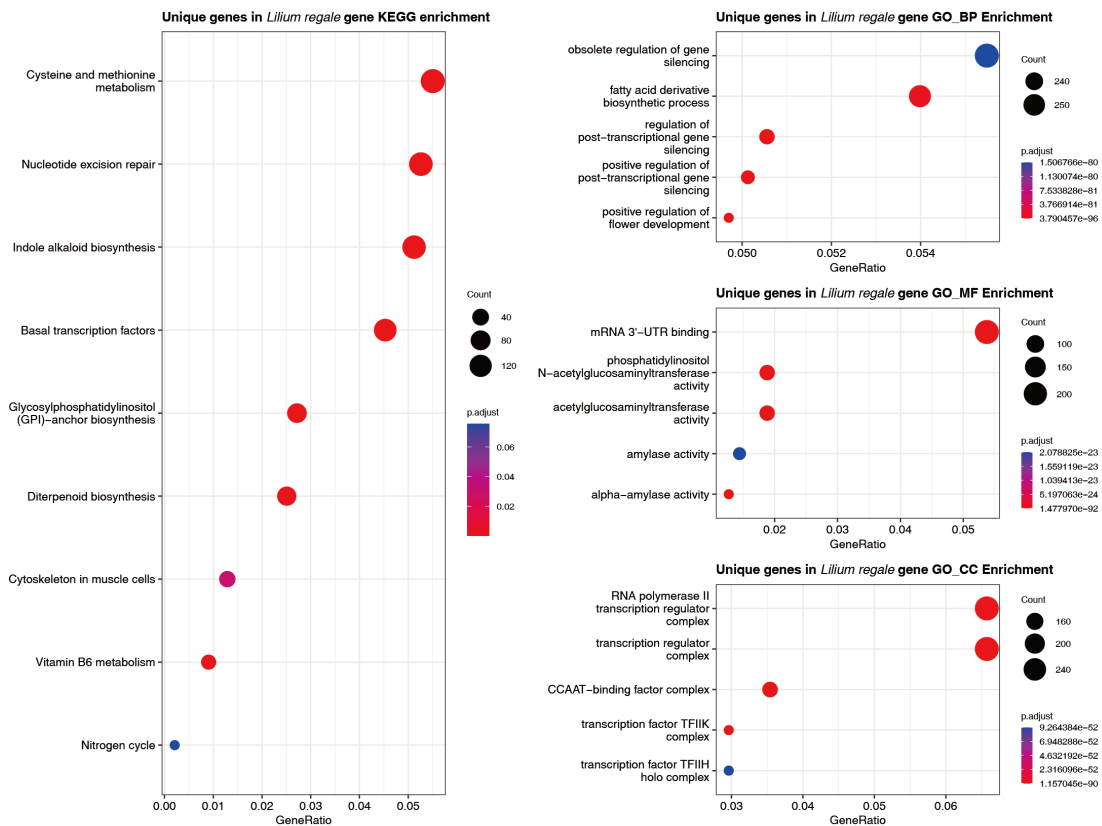


Supplementary Fig. 3: Hi-C interaction heatmaps for pseudomolecules Chr7 to Chr12. a-f, of *L. regale*.

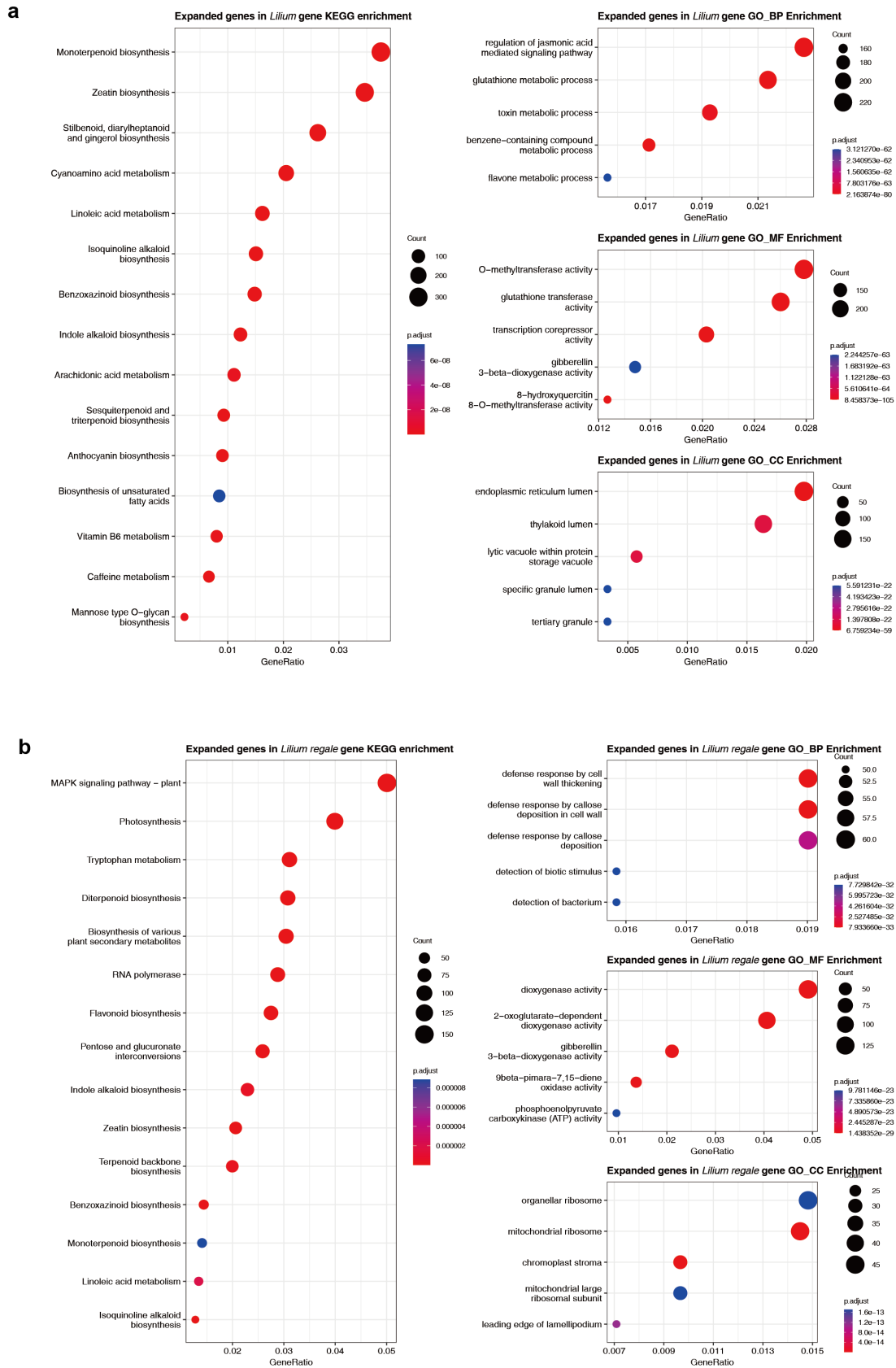
a



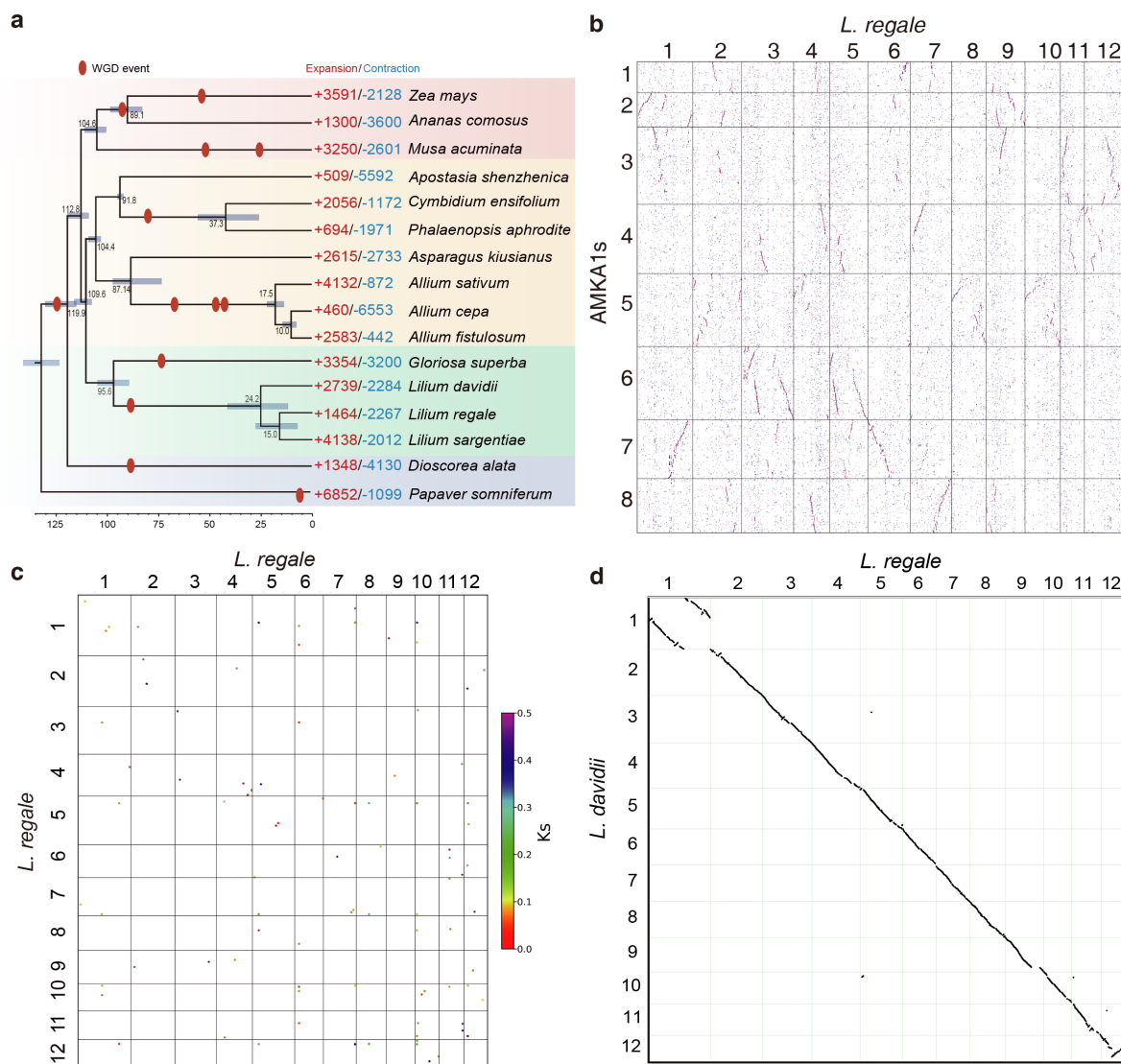
b



Supplementary Fig. 4: KEGG and GO term enrichment of unique genes in *Lilium*(a) and *L. regale* (b) identified in phylogenomic analysis.

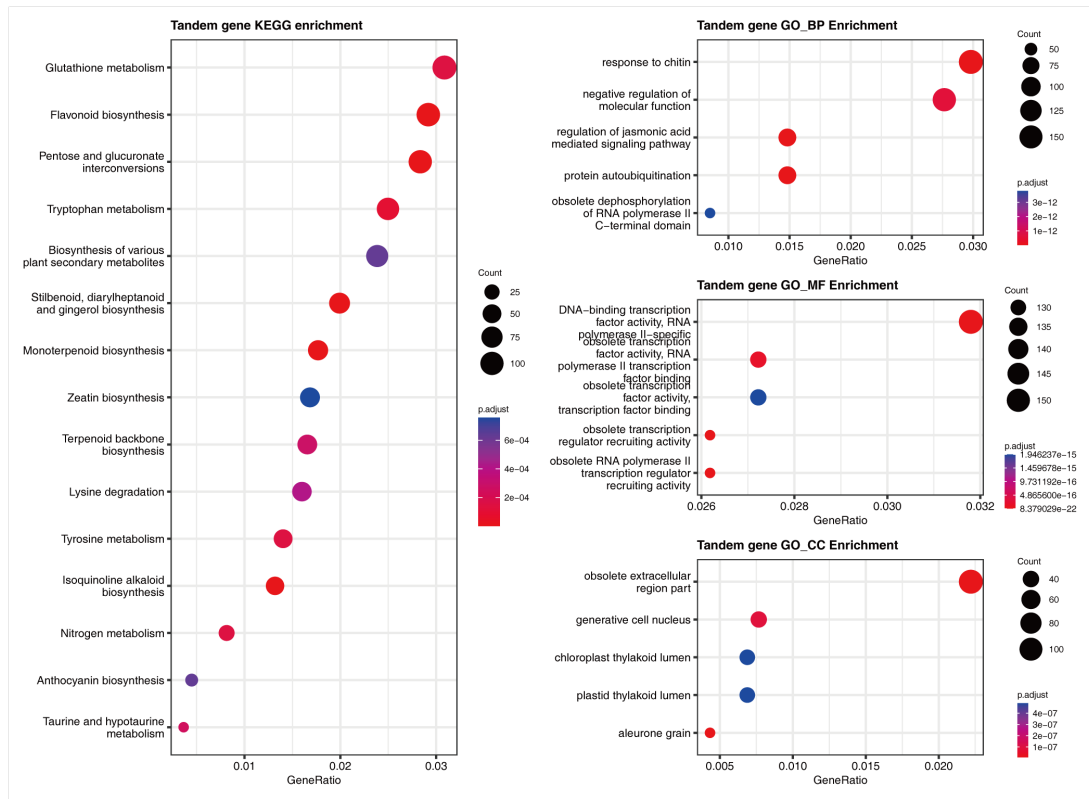


Supplementary Fig. 5: KEGG and GO enrichment of significant expanded gene in *Lilium*(a) and *L. regale* (b) identified in phylogenomic analysis.

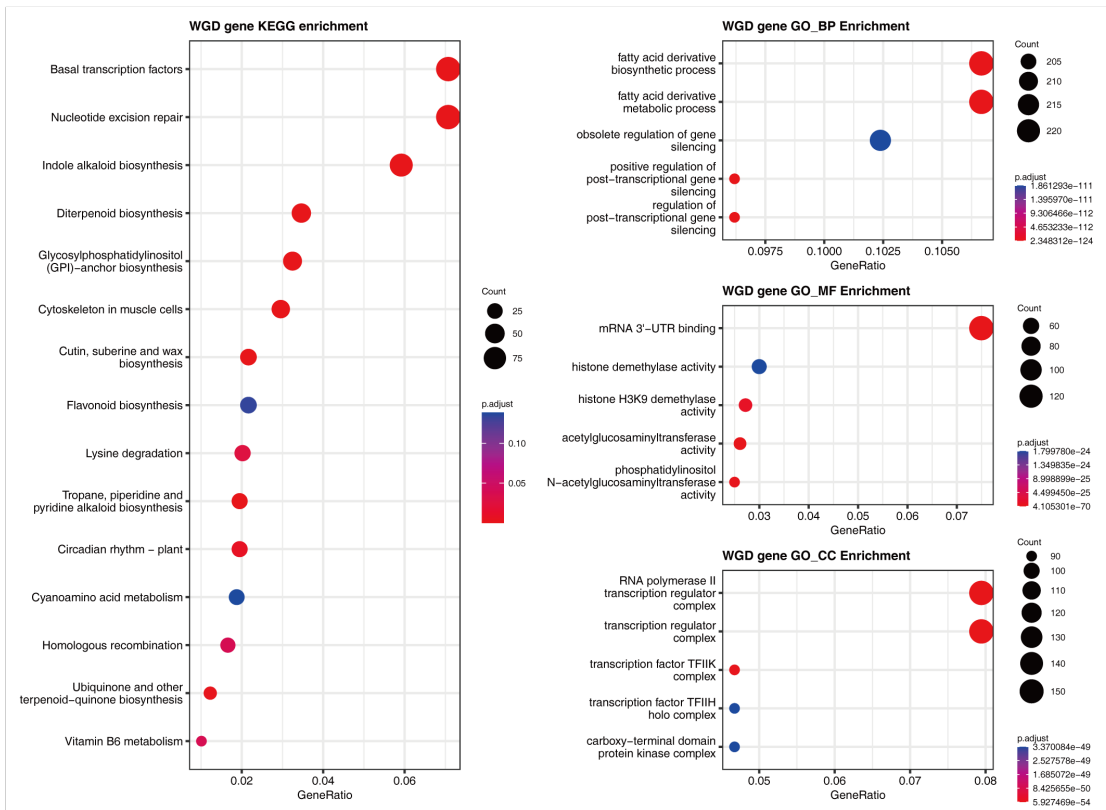


Supplementary Fig. 6: Whole-genome alignment of *Lilium regale* genome. **a**, A maximum likelihood phylogenetic tree of the genomes of 16 species. The numbers near the internal nodes represent the estimated divergence time, while the blue bars indicate the 95% confidence interval for the estimated divergence time. The number of gene-family expansion and contraction events ($P \leq 0.01$) is depicted in red and blue, respectively. **b**, Genomic collinearity among *L. regale* to Monocotyledon ancestral karyotype (AMKA). **c**, Dot plot showing the syntenic relationship of homoeologous gene pairs which K_s less than 0.5. Each point represents a pair of homologous blocks. **d**, Inter-genomic comparison: *L. regale* and *L. davidii*. Each point represents a pair of homologous blocks.

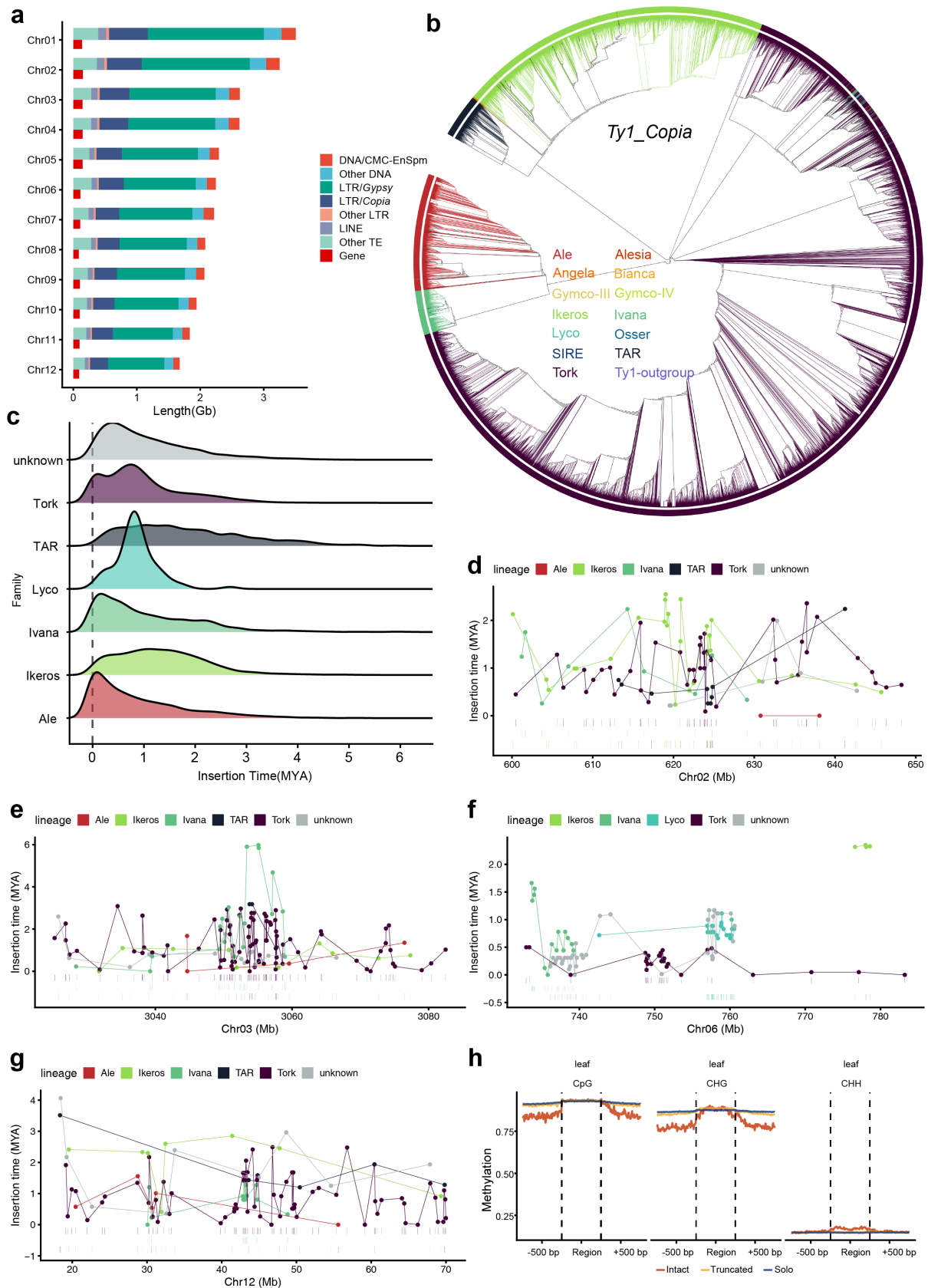
a



b



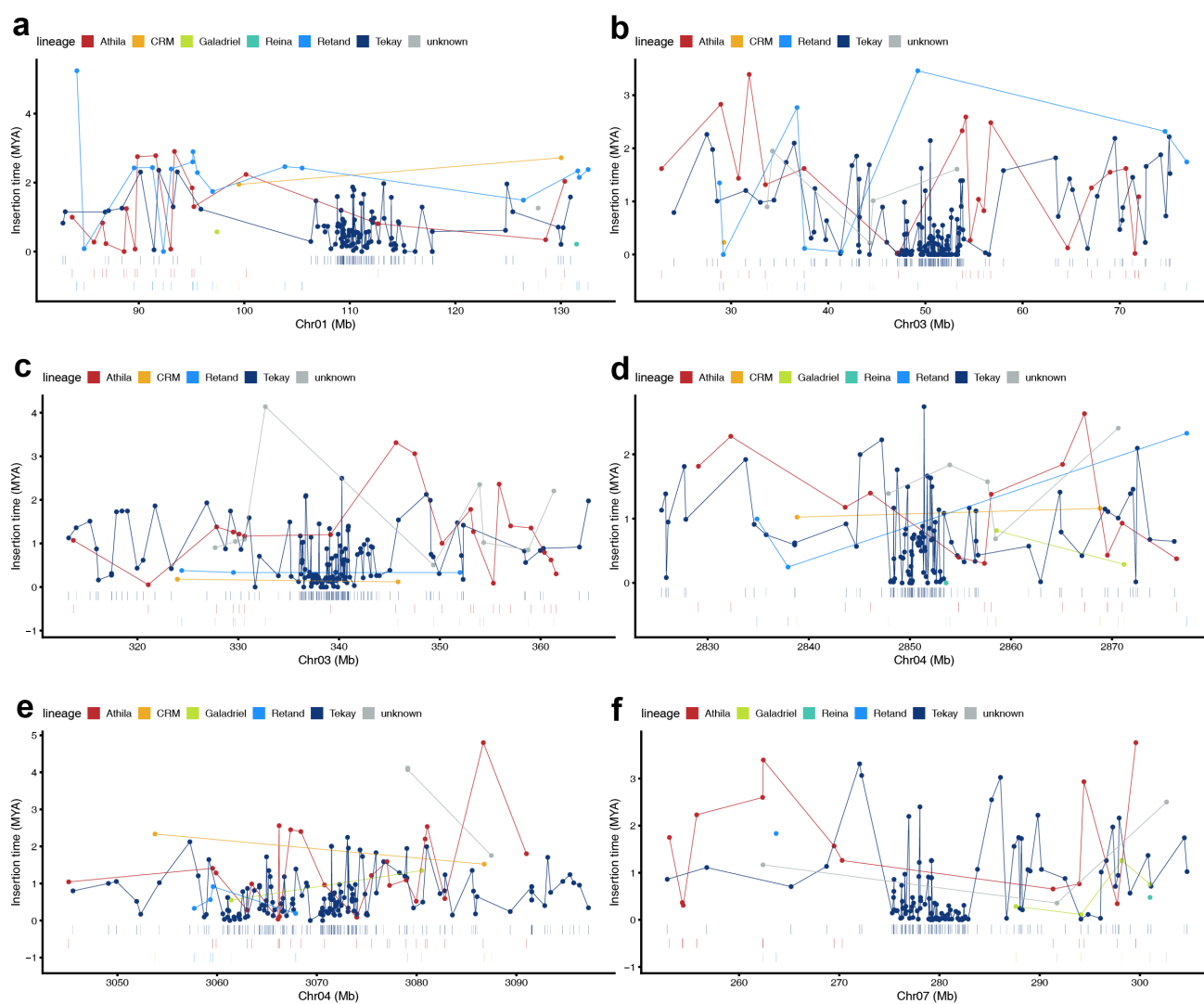
Supplementary Fig. 7: KEGG and GO term enrichment of tandem duplication and WGD genes. a, KEGG and GO terms enrichment of tandem duplication genes. b, KEGG and GO terms enrichment of WGD genes.



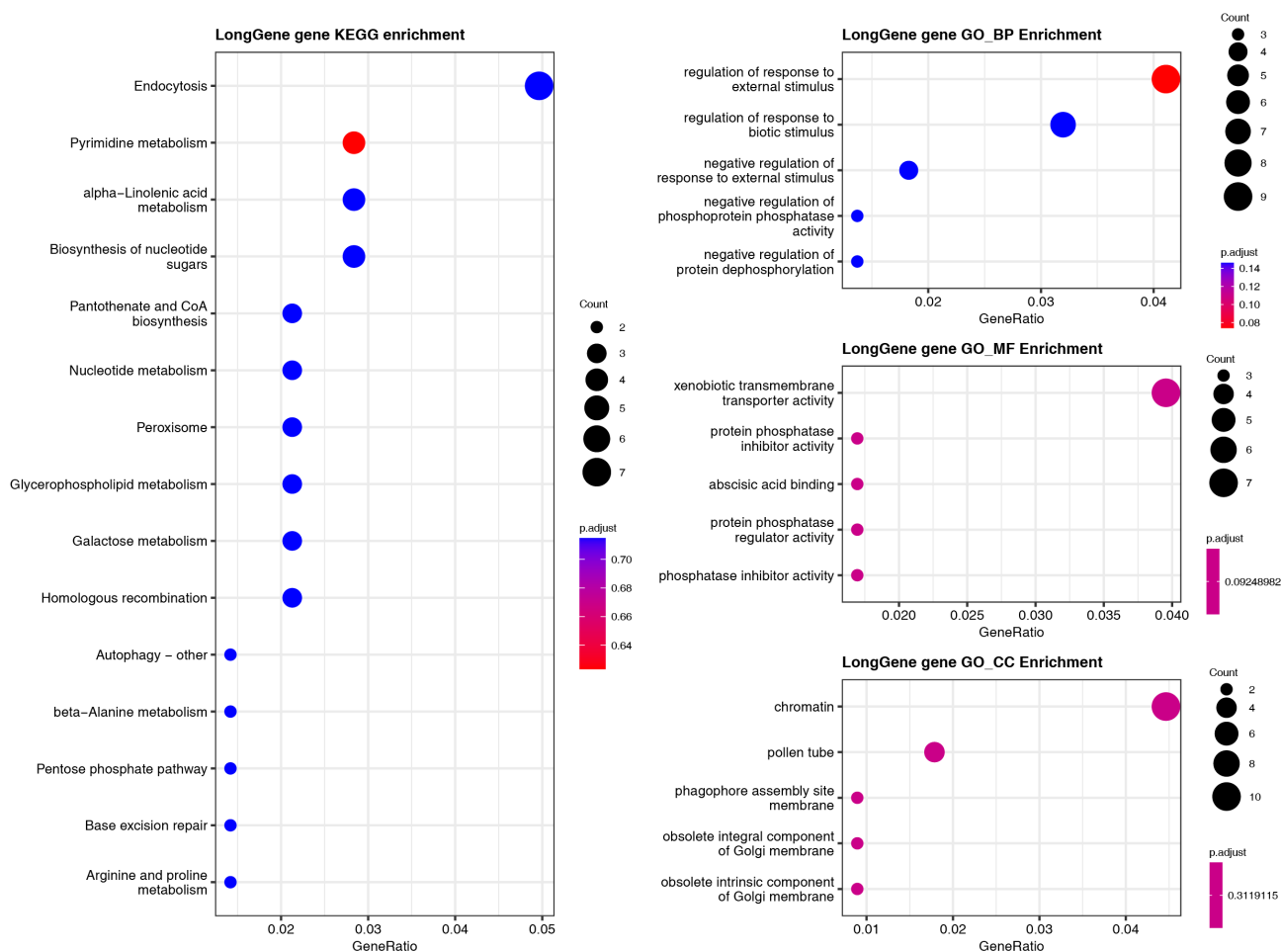
Supplementary Fig. 8: Genome expansion due to extensive insertion of *Copia* retrotransposons.

a, The distribution of different transposable elements across 12 chromosomes in *L. regale*. LTR: long terminal repeats. DNA: DNA transposons. LINE: long interspersed nuclear elements. SINE: short

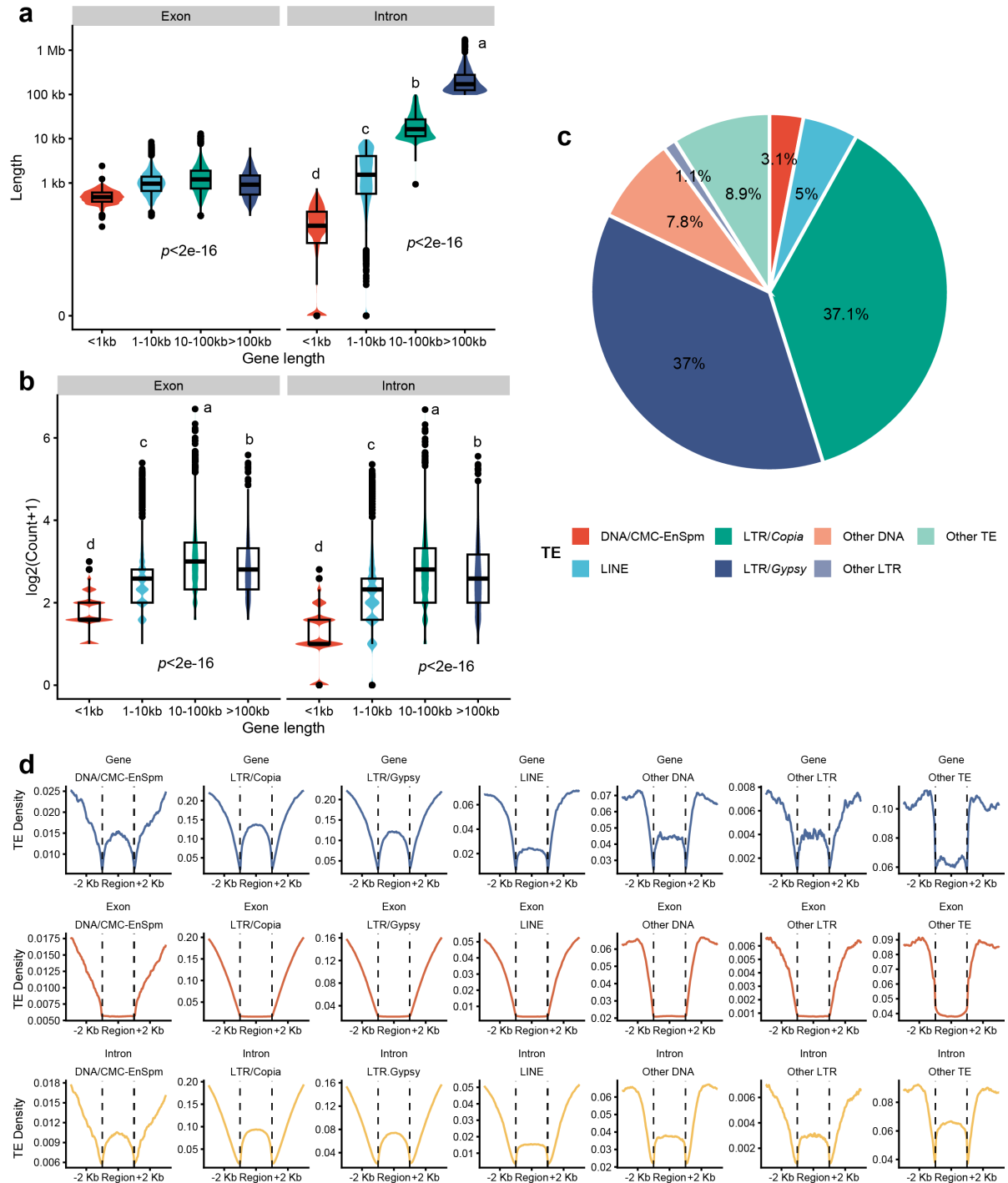
interspersed nuclear elements. **b**, Phylogenetic analyses of intact LTR/Gypsy retrotransposons. **c**, Distribution of insertion times of different families of *Copia* retrotransposons in *L. regale*. **d-g**, *Copia* retrotransposons bursts on Chr2(**d**), Chr3(**e**), Chr6(**f**) and Chr12(**g**). Different colored dots represent individual intact LTR insertions of different lineages along the chromosomes, with the colored bars below showing the distribution of different lineages on the chromosomes. **h**, Overview of DNA methylation level in intact, truncated and solo LTRs of *L. regale* leaves. Source data are provided as a Source Data file.



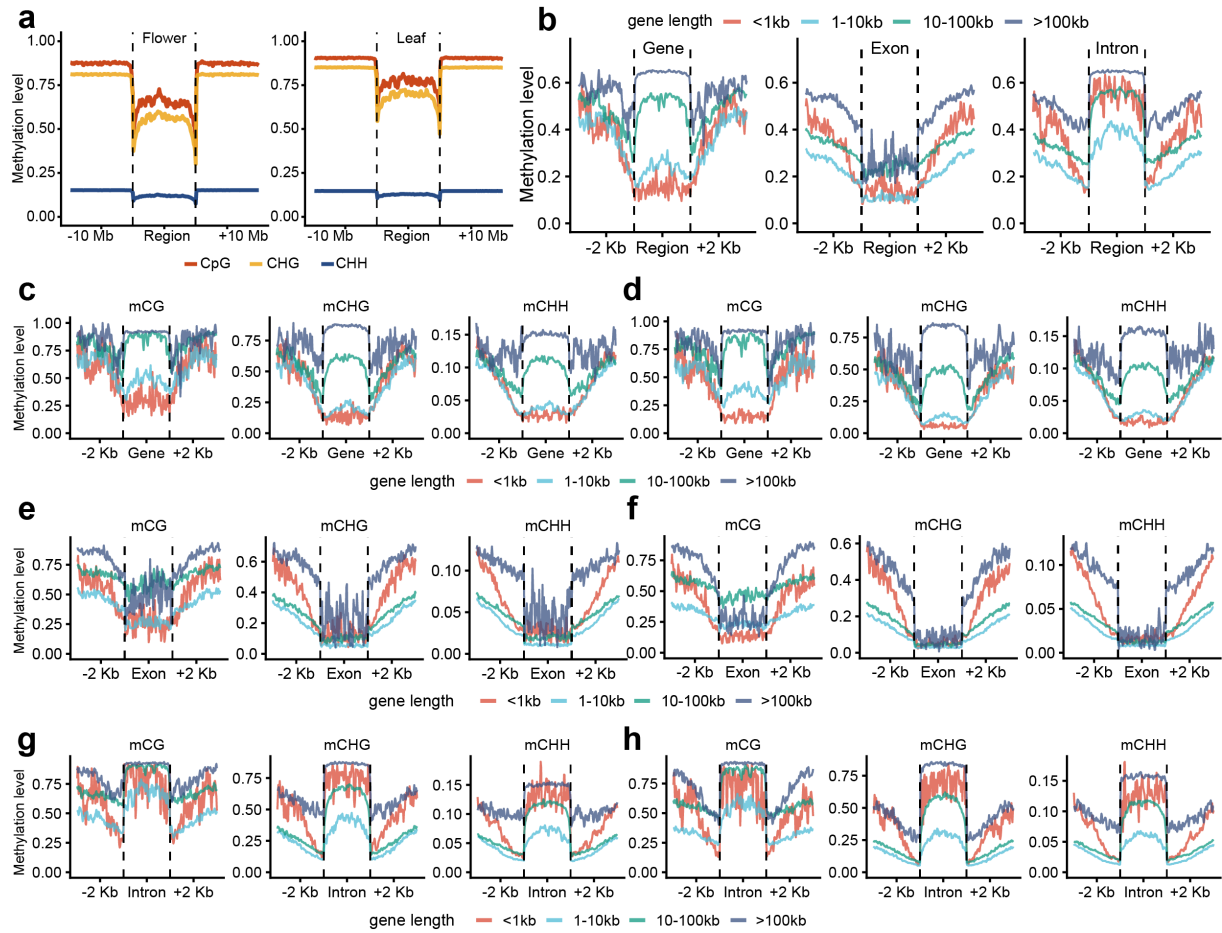
Supplementary Fig. 9: *Gypsy* retrotransposons bursts on Chr01 (a), Chr03 (b,c), Chr04 (d,e) and Chr07 (f). *Gypsy* transposons of different subfamilies are colored differently.



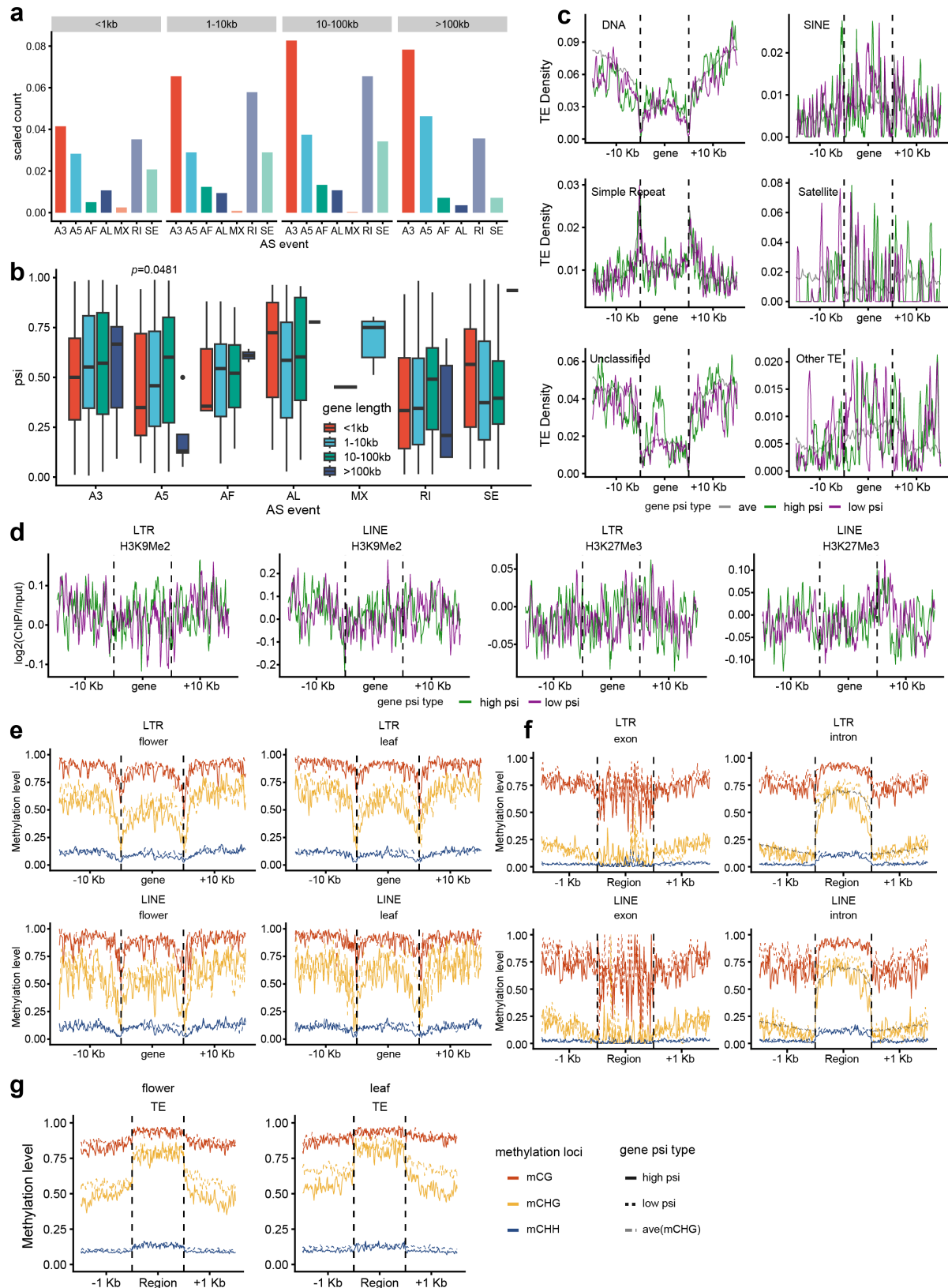
Supplementary Fig. 10: KEGG and GO term enrichment of ultra-long genes (>100kb).



Supplementary Fig. 11: The long gene and its DNA methylation landscape. a,b, The length (**a**) and count (**b**) of exons and introns of each gene ($n = 15667(<1\text{kb})$, $34295(1-10\text{kb})$, $11963(10-100\text{kb})$, $952(>100\text{kb})$ genes) in *Lilium regale*. **c,** The total percentage of different transposable elements (TE) in introns of *L. regale*. **d,** Overview of TE density in genes, exons and introns and their 2kb flanking regions in *L. regale*. In **a**, and **b**, groups labeled with different lowercase letters indicate significant differences ($p < 0.05$) based on LSD comparisons following one-way ANOVA, box plots denote the 25th percentile, the median and the 75th percentile, with minimum to maximum whiskers. Source data are provided as a Source Data file.

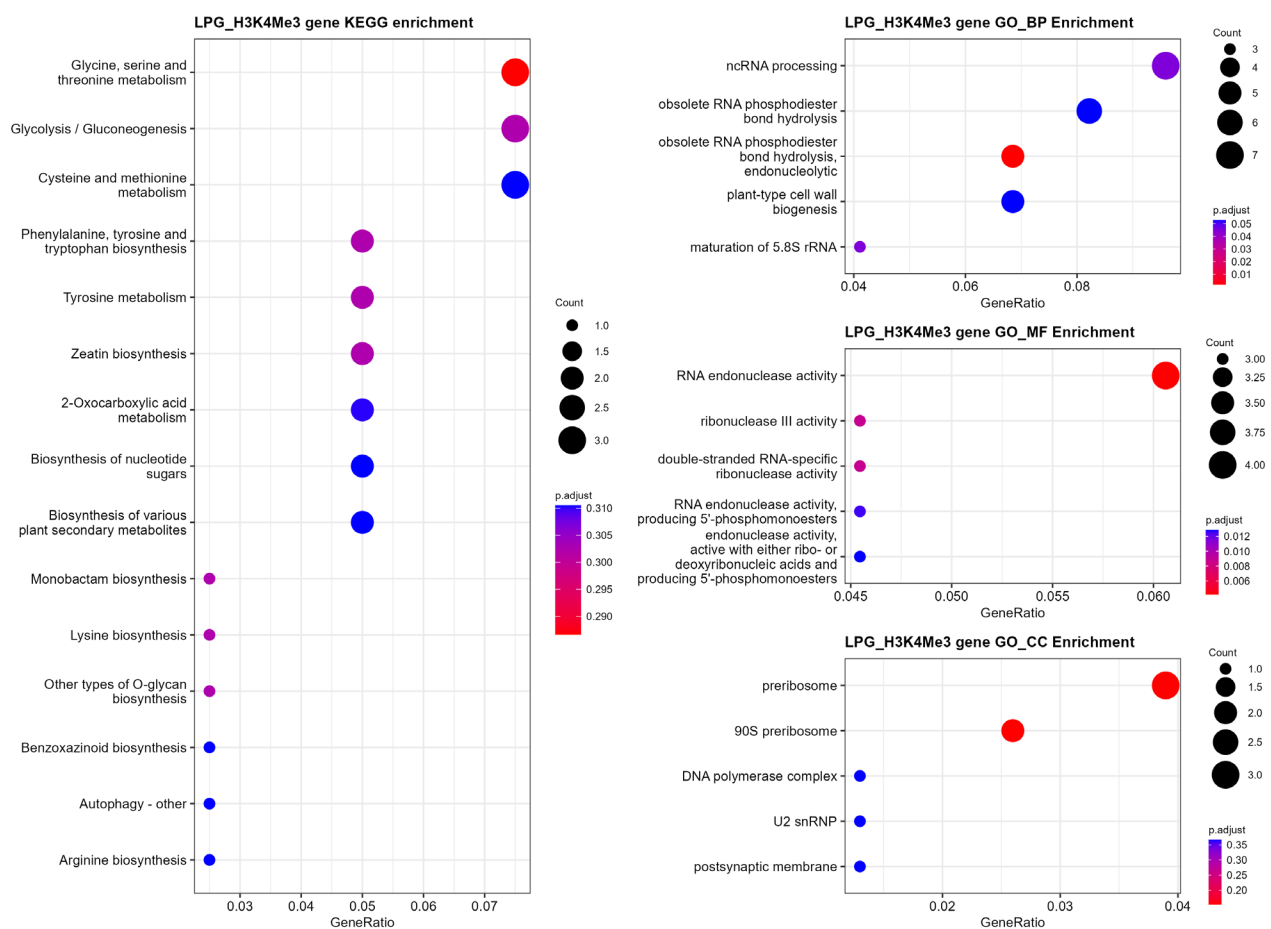


Supplementary Fig. 12: DNA methylation landscape of *Lilium regale* genome. **a**, Overview of DNA methylation level gene and 10Mb flanking regions in *L. regale* leaf and flower. **b**, Overview of DNA methylation level in genes, exons and introns of *L. regale* leaf with lines colored by different gene-size ranges. **c-h**, DNA methylation patterns of genes (**c,d**), exons(**e,f**), introns (**g,h**) and their 2kb flanking regions in leaves (**c,e,g**) and flower (**d,f,h**) tissue of *L. regale* with lines colored by different length categories of full-length genes.

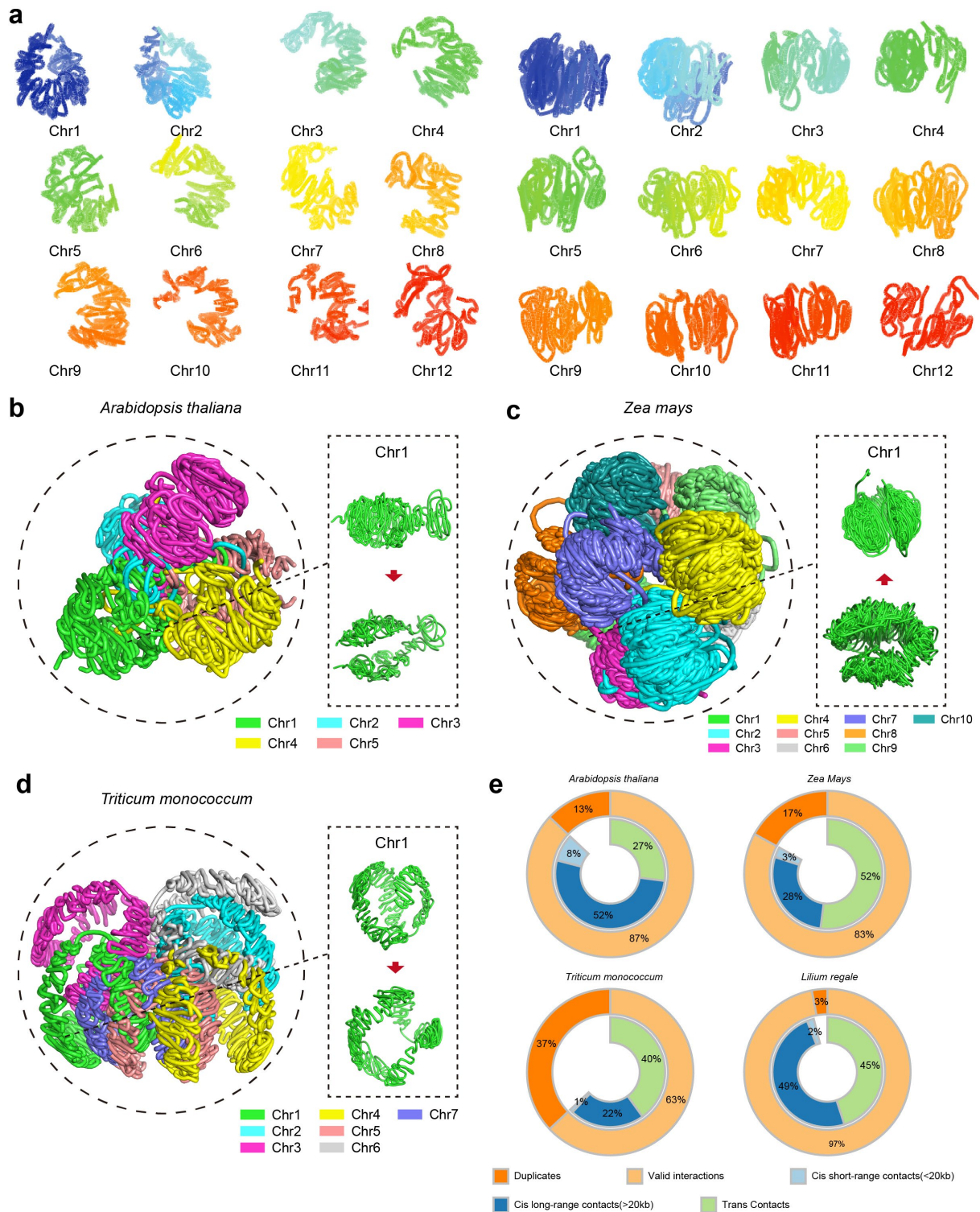


Supplementary Fig. 13: The landscape of alternative splicing events in *Lilium regale*. **a**, The scaled count of alternative splicing (AS) events in different gene length. The scaled count means the ratio of (AS counts)/(identified gene number in respective length). **b**, The percent spliced in index (PSI) of AS events in different gene length (n = 229(<1kb), 1431(1-10kb) 842(10-100kb), 50

(>100kb) genes). **c**, The density of different transposon (TE) insertions in gene bodies, upstream and downstream of high PSI genes (HPG) and low PSI genes (LPG). **d**, The enrichment of different histones between upstream and downstream of HPGs and LPGs. **e**, The methylation levels of different PSI genes with LTR insertions in leaves and flower. **f**, The methylation levels of exons and introns for different PSI genes with LTR and LINE insertions in leaves. **g**, The methylation levels of TEs on introns for different PSI genes with LTR and LINE insertions in leaves and flower. In **e,f,g**, the figure legends are same. Source data are provided as a Source Data file.

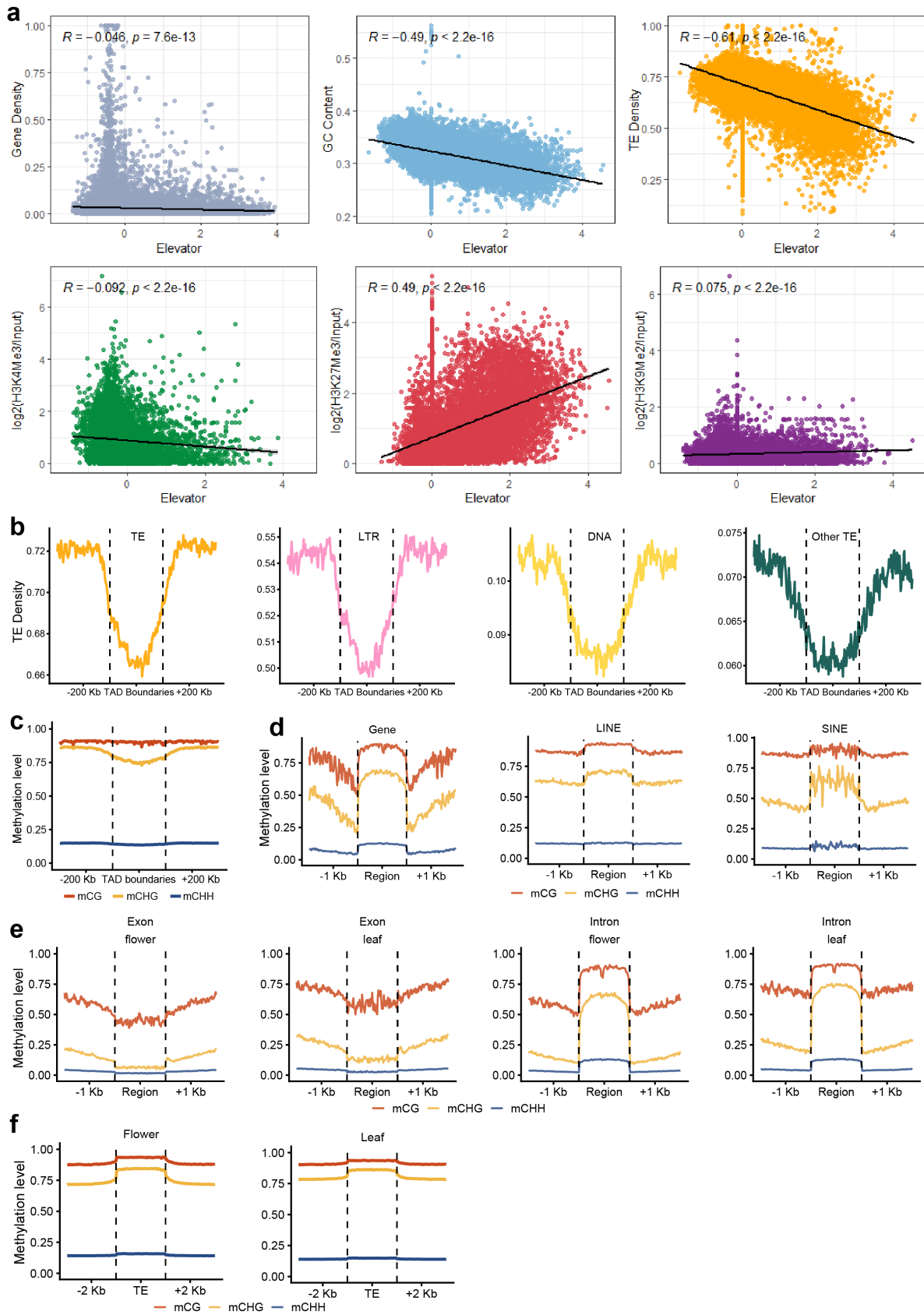


Supplementary Fig. 14: KEGG and GO term enrichment of low PSI (percent spliced in index) genes with H3K4Me3 enrichment. LPG: low PSI genes.



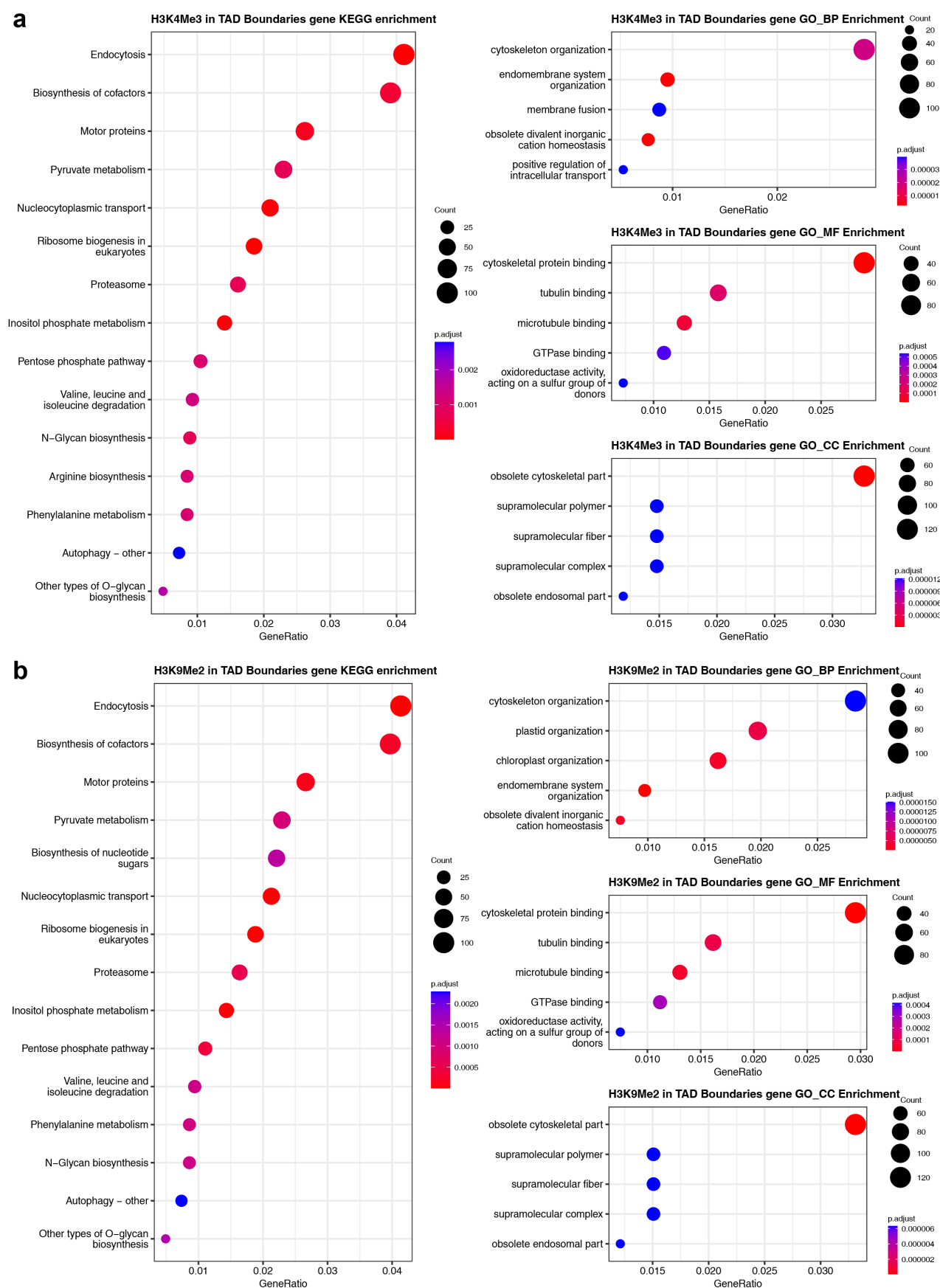
Supplementary Fig. 15: The 3D model of chromatin folding in *Lilium regale* and other species.

a, 3D structure plots of *L. regale* twelve chromosomes (haplotype) in two views based on Hi-C data. **b-d**, Chromosome (haplotype) 3D structure plots of *Arabidopsis thaliana* (**b**), *Zea mays* (**c**), and *Triticum monococcum* (**d**) based on Hi-C data. **e**, The percentage of cis-contacts and trans-contacts in valid interactions on Hi-C data.

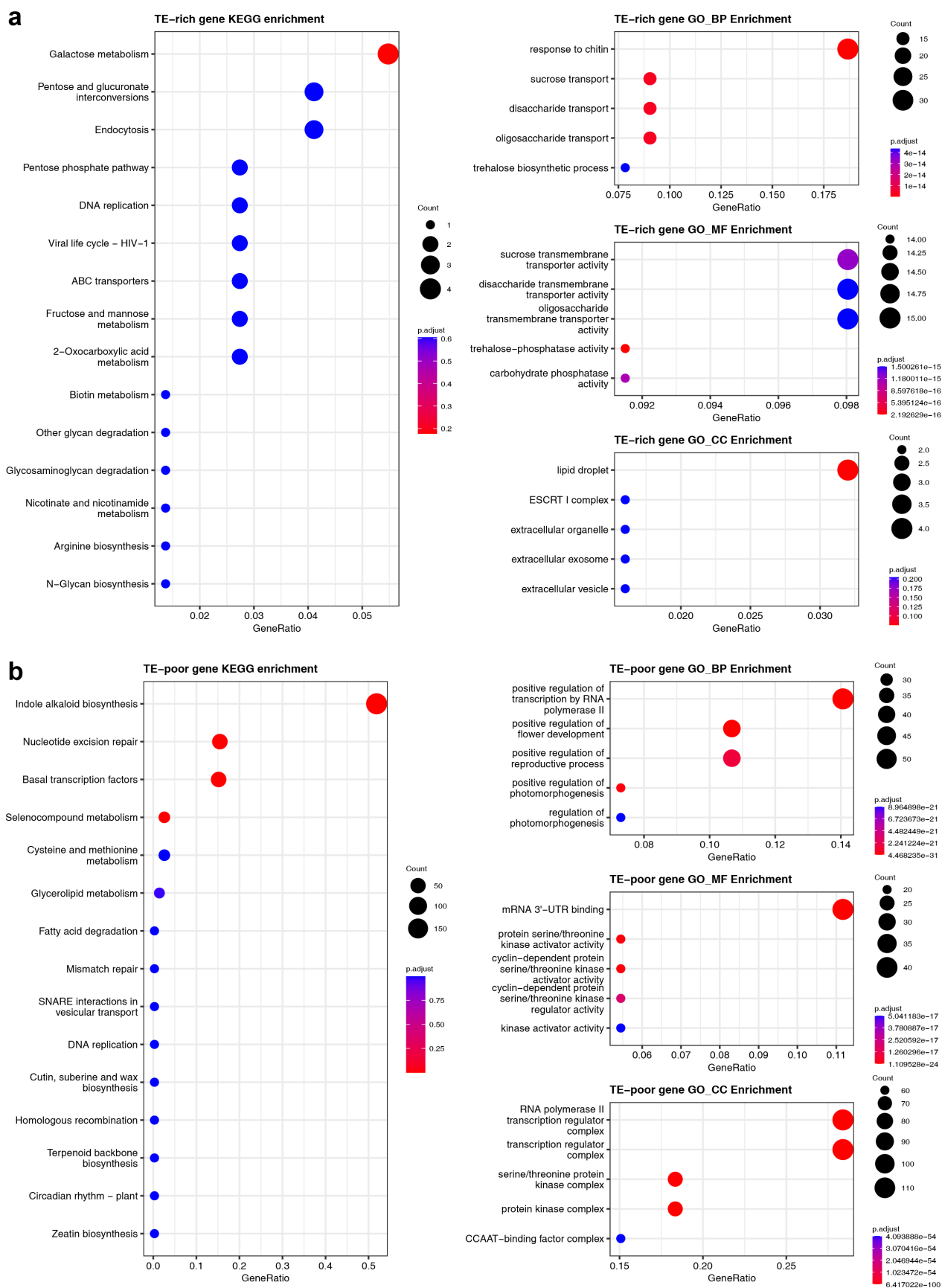


Supplementary Fig. 16: Three-dimensional genomic landscape of *Lilium regale* genome. **a**, The linear correlation (Pearson) between gene density, GC content, intron density, H3K4Me3 coverage, H3K27Me3 coverage, H3K9Me2 coverage, with Hi-C data PCA elevator 1 in *L. regale* genome. The

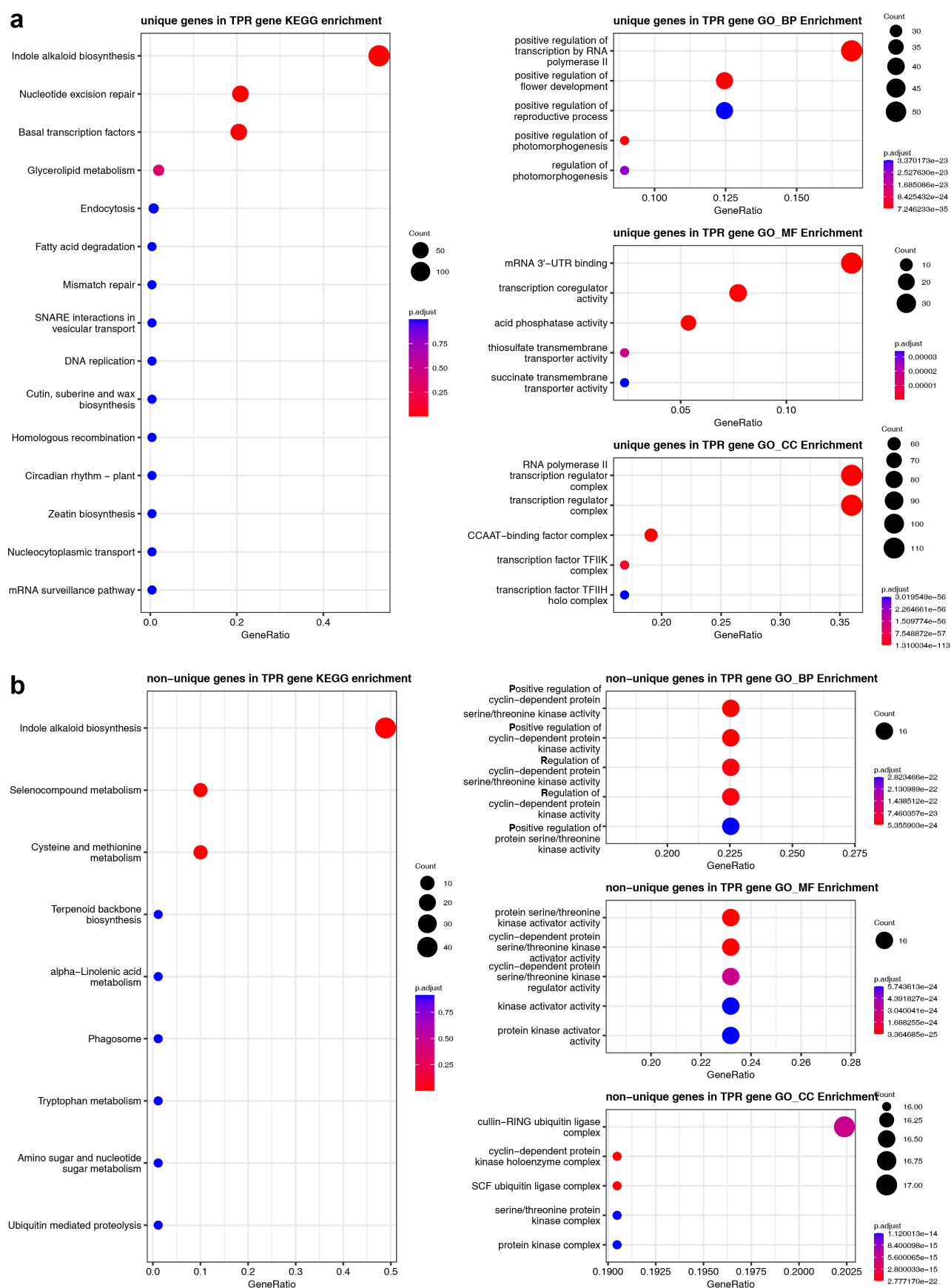
black lines represent the linear regression fitted to all data points. **b**, Enrichment status of different transposable elements (TEs) on TAD boundaries. **c**, The methylation level of TAD boundaries upstream and downstream in leaves. **d**, The methylation level of genes, LINEs, and SINEs between their upstream and downstream on TAD boundaries in leaves. **e**, The methylation level of exons and introns of genes on TAD boundaries. **f**, The methylation level of all TEs upstream and downstream on TAD boundaries in leaves. Source data are provided as a Source Data file.



Supplementary Fig. 17: KEGG and GO term enrichment of genes related to H3K4Me3 (a) and H3K9Me2 (b) on TAD (topologically associated domain) boundaries.



Supplementary Fig. 18: KEGG and GO term enrichment of genes in Top100 TPRs (TE-poor region) (a) and Top100 TRRs (TE-rich region) (b).



Supplementary Fig. 19: KEGG and GO term enrichment of unique genes (a) and non-unique genes (b) of *L. regale* in Top100 TPR (TE-poor region).