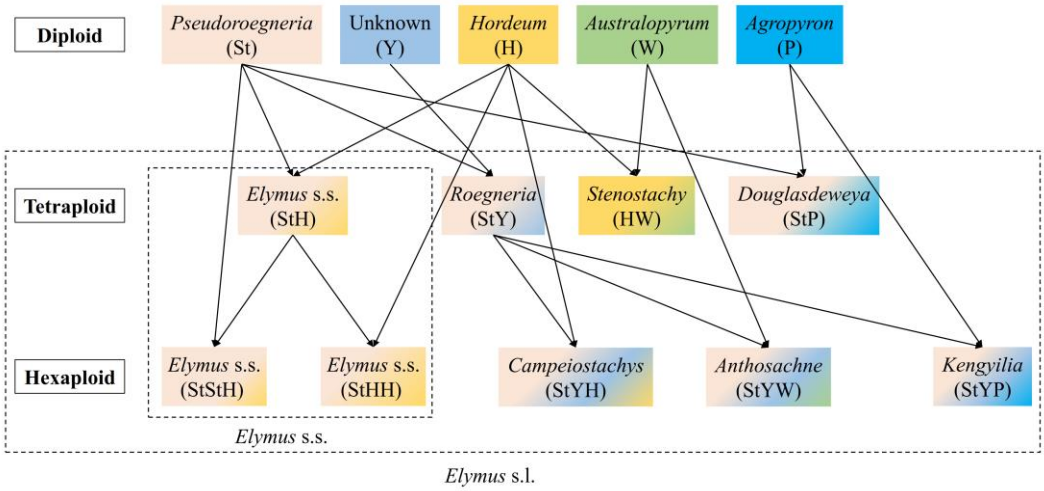
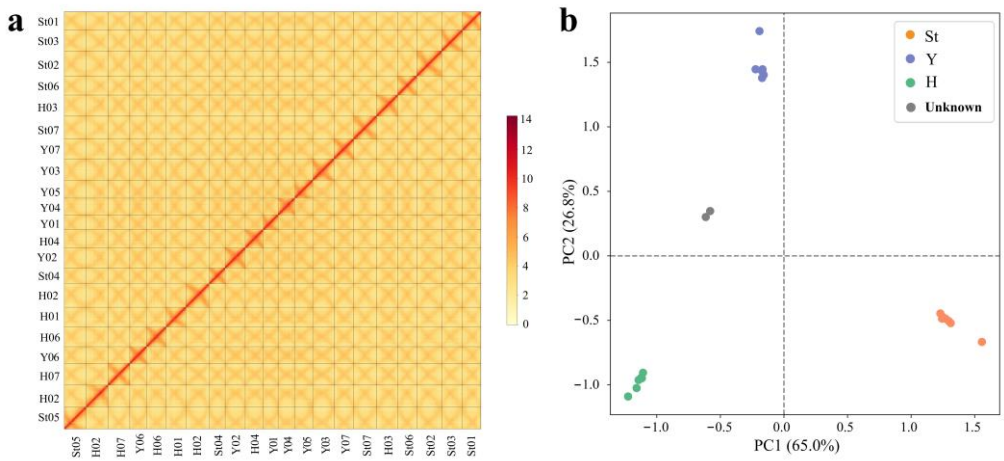
Analysis of allohexaploid wheatgrass genome reveals its Y haplome origin in Triticeae and high-altitude adaptation

Xiong *et al.*

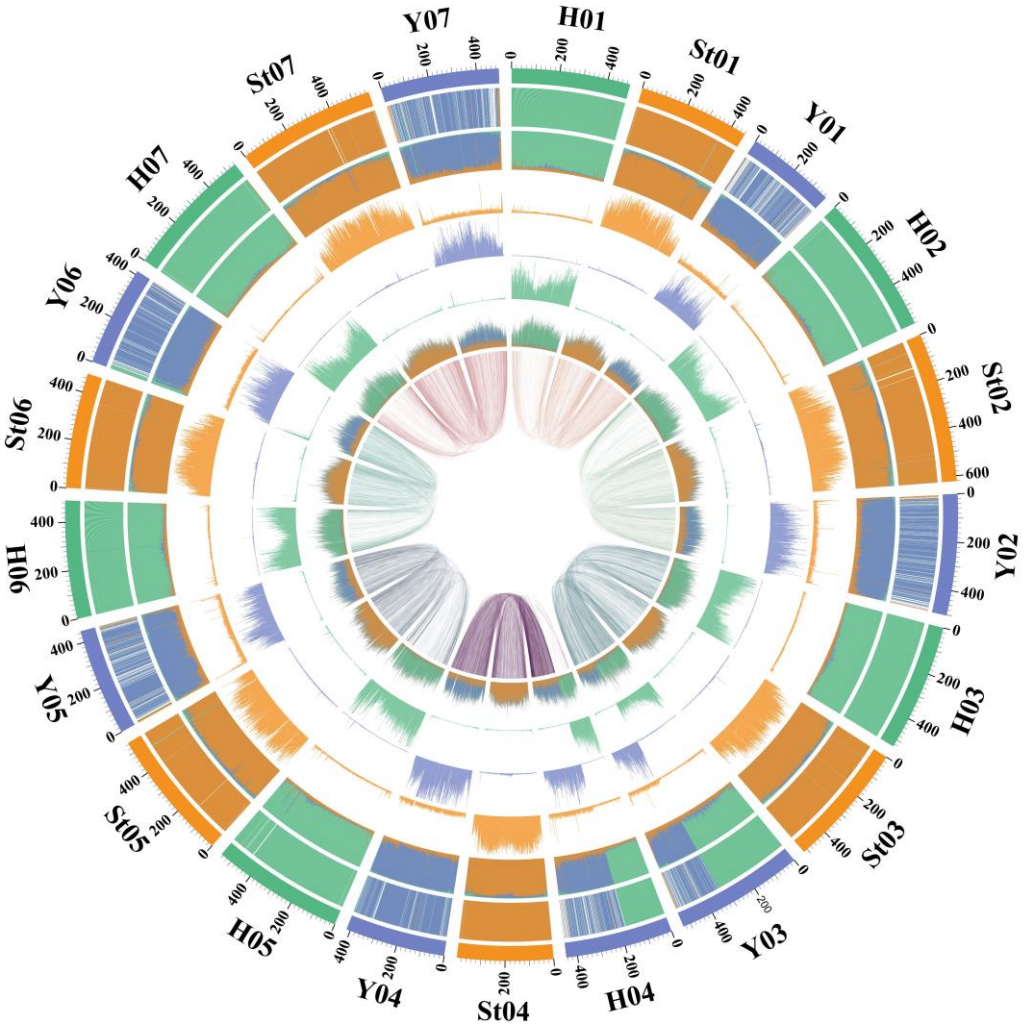


Supplementary Fig. 1. An illustrative diagram showing the potential polyploids speciation process of *Elymus* s.l..

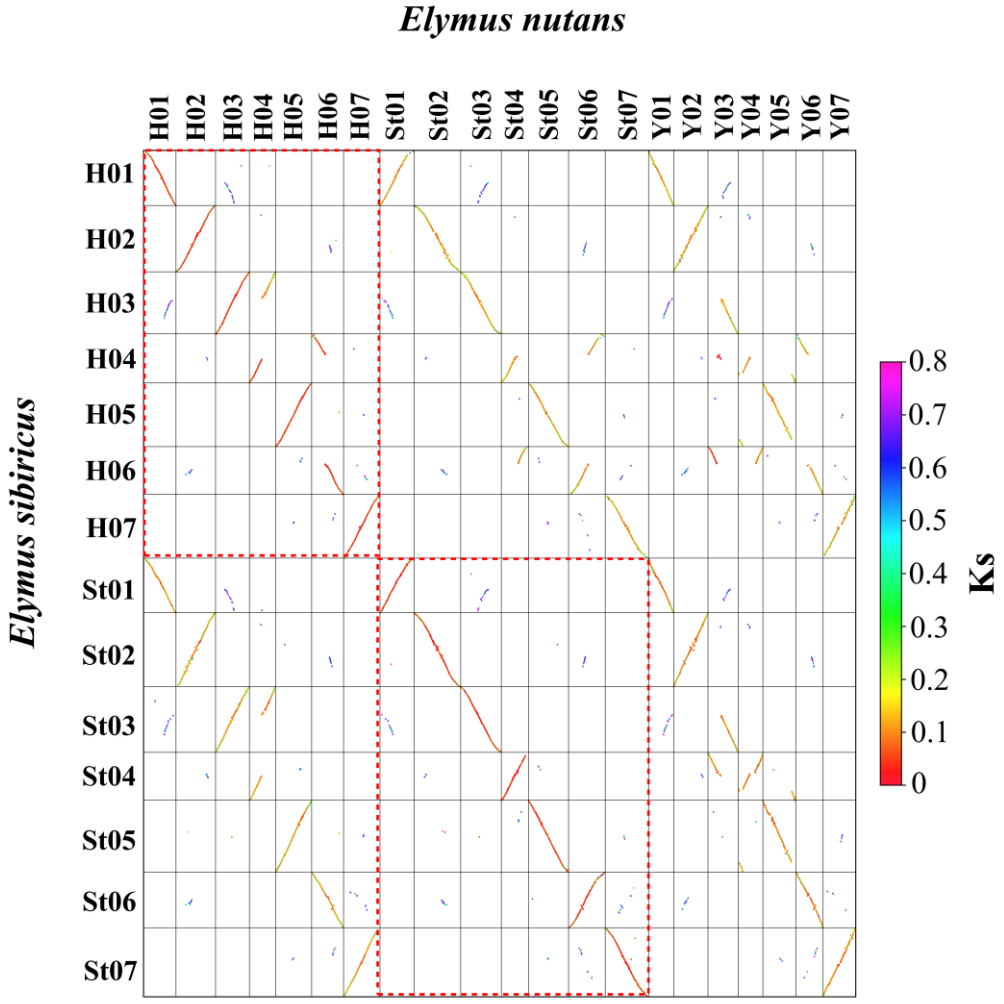
Supplementary Fig. 2. The habitats and spike morphological characteristics of *E. nutans*. a, The growth status of the accession utilized for genome sequencing in the common garden. b, a detailed view of panicle morphology during in-situ collection. c, the natural habitat of the wild *E. nutans* population.



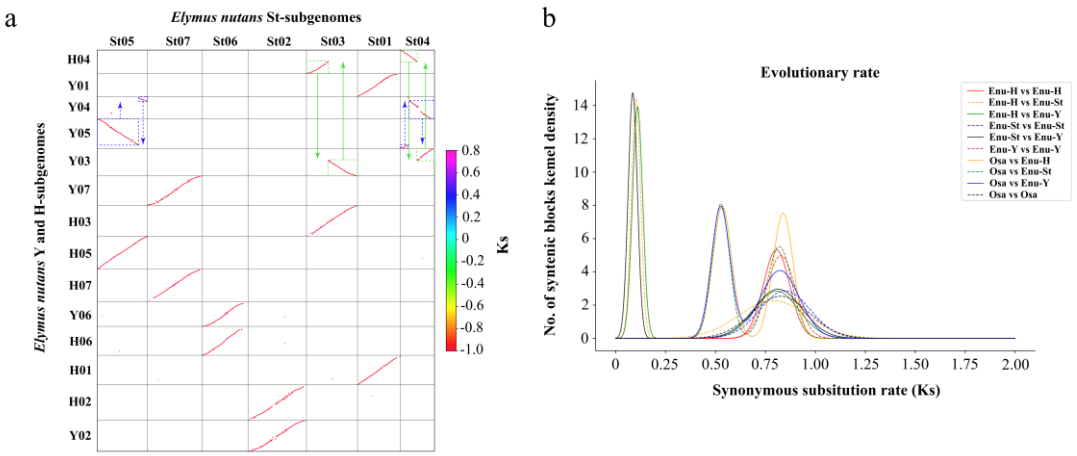
Supplementary Fig. 3. The scaffolding and classification of 21 chromosomes in *E. nutans*. a, The Hi-C interaction map of *E. nutans*, where the color of each dot representing the log value of the interaction intensity between corresponding genomic bin pairs. The colors gradient from white to red indicates the increasing interaction intensity. b, Illustration of chromosome assignment based on PCA analysis.



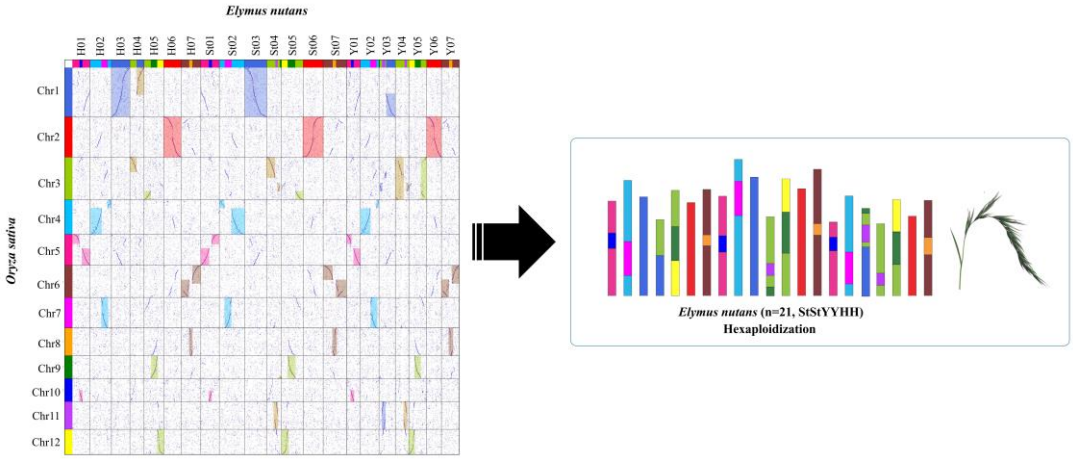
Supplementary Fig. 4. The circos plot illustrates the assembly features of the hexaploid *E. nutans* genome. From the outermost to the innermost circle (1-8): (1) Chromosome assignment based on k-mer analysis; (2) Subgenome-specific k-mer enrichment, with the same color representing the same subgenome; (3) Normalized proportion of subgenome-specific k-mers; (4-6) Absolute counts of subgenome-specific k-mers for each set; (7) LTR density; (8) Homoeologous blocks, with connections of the same color indicating syntenic blocks between different chromosomes.



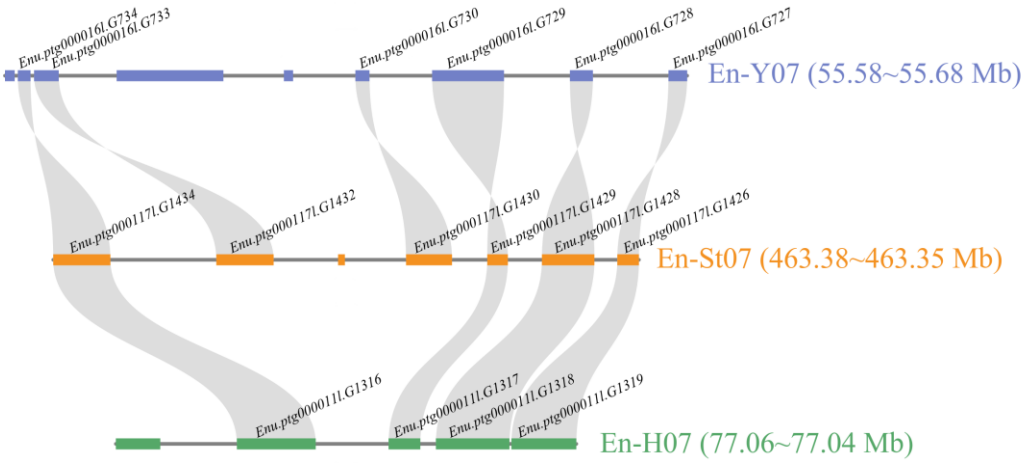
Supplementary Fig. 5. The classification of the three subgenomes (St/Y/H) in *E. nutans*. Analysis of synonymous substitution (Ks) values between *E. nutans* and *E. sibiricus*, where the gradient from deep red to light red indicates an increasing divergence time between *E. nutans* and *E. sibiricus*. The codes starting with S, Y, and H represent the St-, Y-, and H-subgenomes, respectively. From top to bottom, the two red dashed boxes represent the H- and St-subgenomes, respectively.



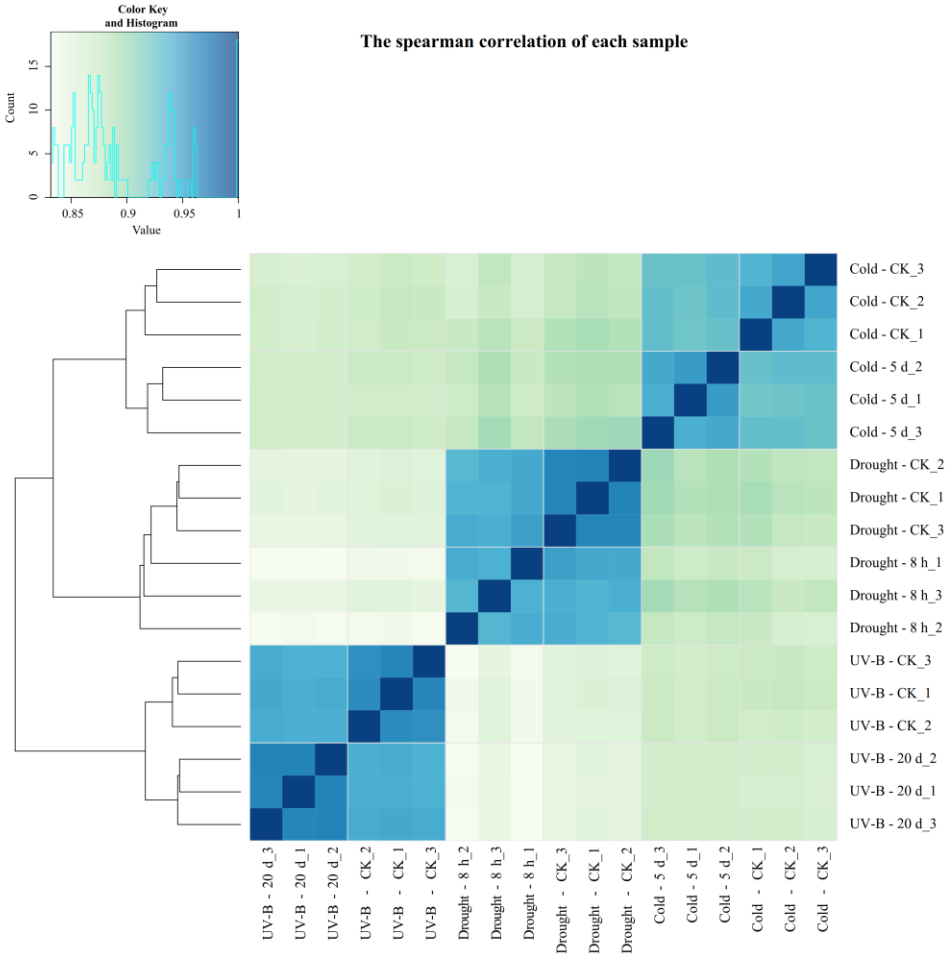
Supplementary Fig. 6. Chromosomal reciprocal translocation events in *E. nutans*. a, The Ks analysis between the En-St and the En-Y / En-H subgenomes, which revealed two significant non-Robertsonian reciprocal translocation (non-RTA) events between En-Y03 vs En-H04 and En-Y04 vs En-Y05. b, Divergence of the synonymous mutation (Ks) between three subgenomes.



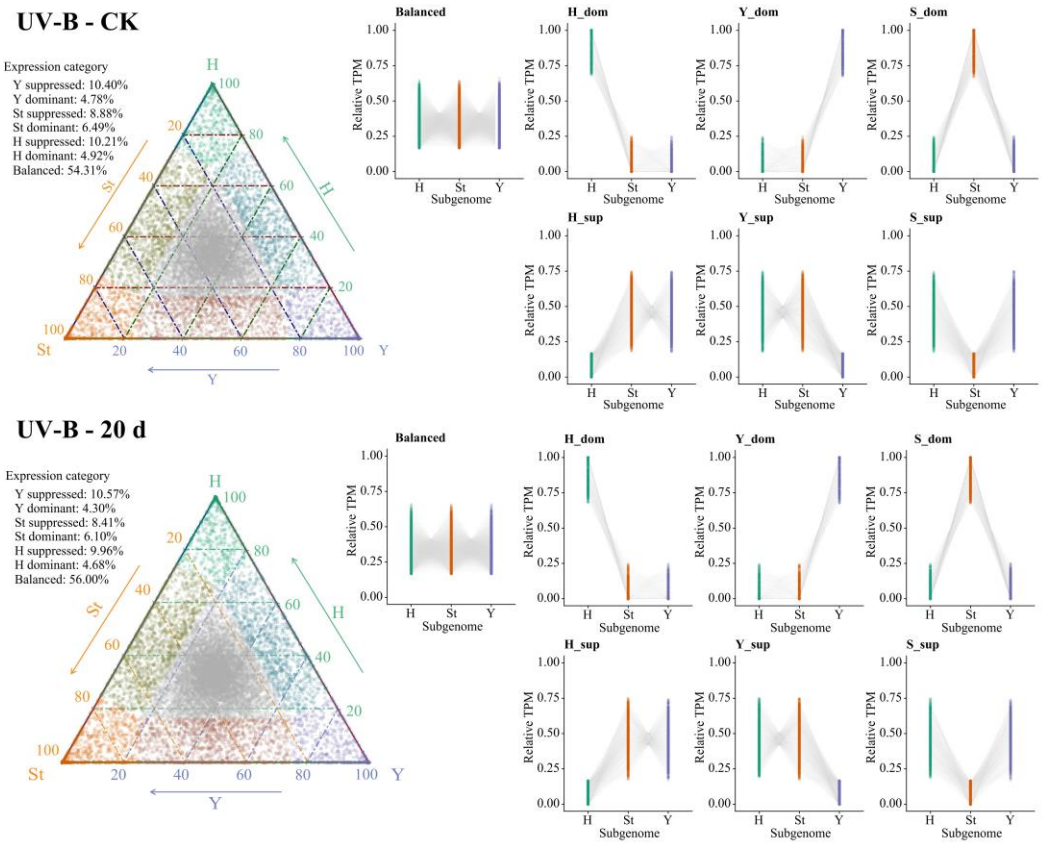
Supplementary Fig. 7. The karyotype history of *E. nutans* predicted with the Whole-Genome Duplication Integrated software with the ancestral species of rice. The chromosome colors of *E. nutans* are correspondingly matched to those of rice chromosomes. The left panel displays the syntenic relationships between chromosomes, while the right panel illustrates the ancestral karyotype of *E. nutans*.



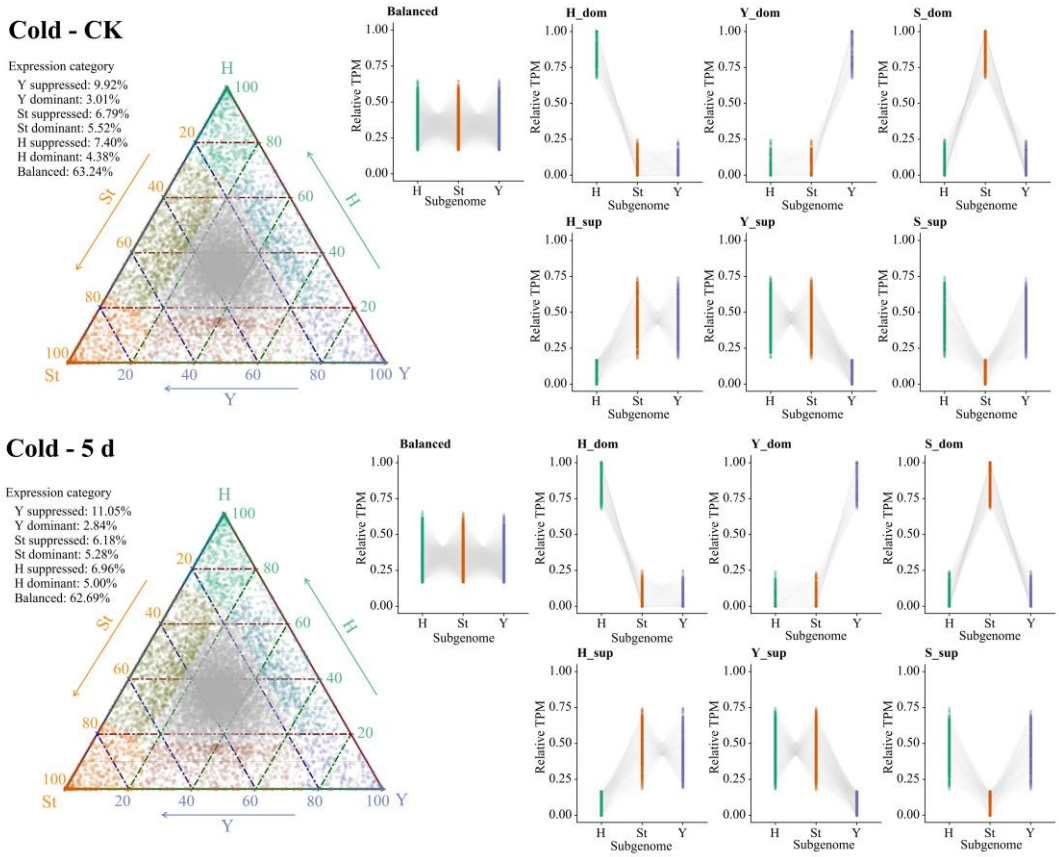
Supplementary Fig. 8. Example of pseudogenization. The syntenic genes in H07 corresponding to those in St07 (*Enu.ptg000117l.G1432* and *Enu.ptg000117l.G1430*) and Y07 (*Enu.ptg000016l.G733* and *Enu.ptg000016l.G730*) are missing.



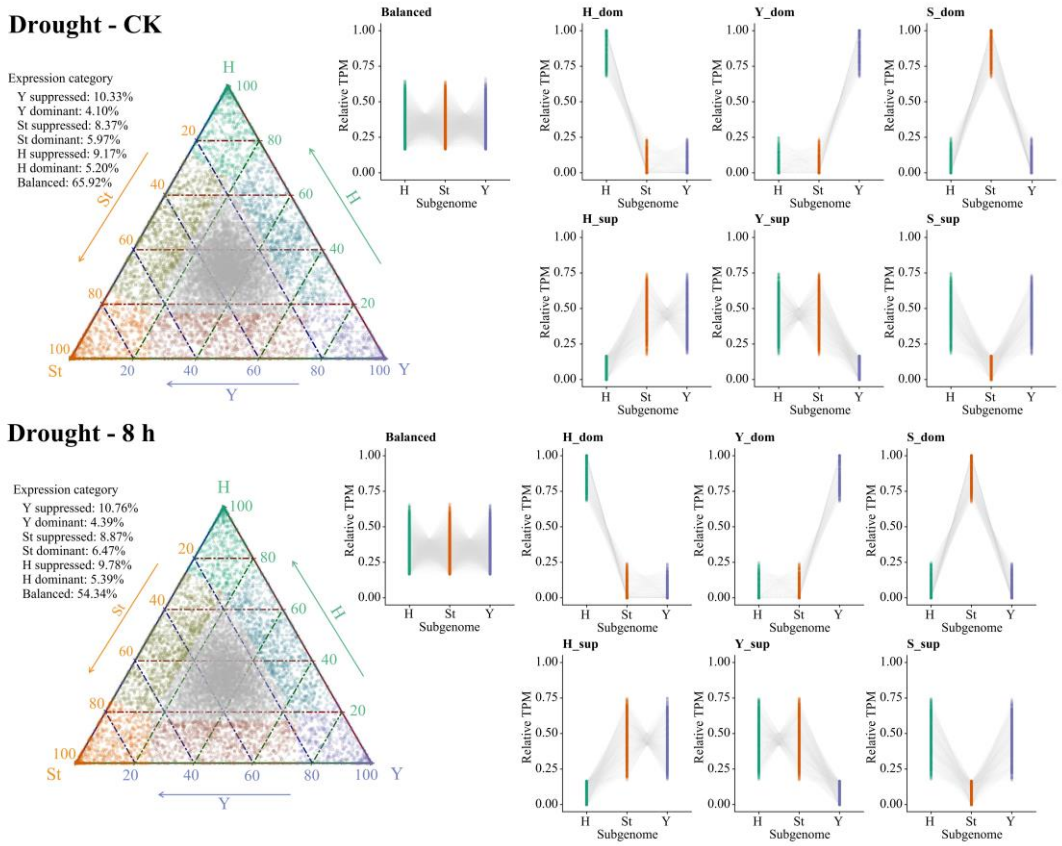
Supplementary Fig. 9. Correlation analysis of *E. nutans* RNA-seq samples under UV-B, drought and cold stresses and their corresponding control.



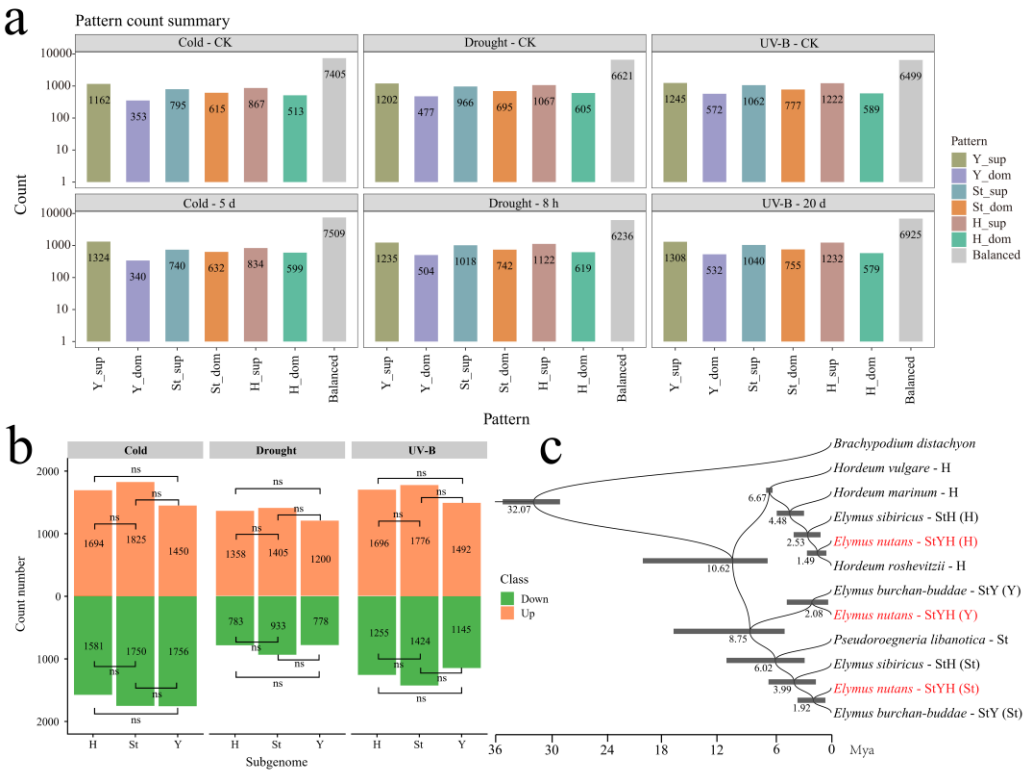
Supplementary Fig. 10. Under UV-B stress, there was no expression bias among the three subgenomes of *E. nutans*. The relative expression abundance of seven homologous expression patterns of 11,996 and 12,971 1:1:1 orthologous gene pair from the three subgenomes of *E. nutans*, under UV-B - CK (above) and UV-B - 20 d (below) respectively. Each point represents a gene triplet. St_sup, St suppressed; Y_sup, Y suppressed; H_sup, H suppressed; St_dom, St dominance; Y_dom, Y dominance; H_dom, H dominance.



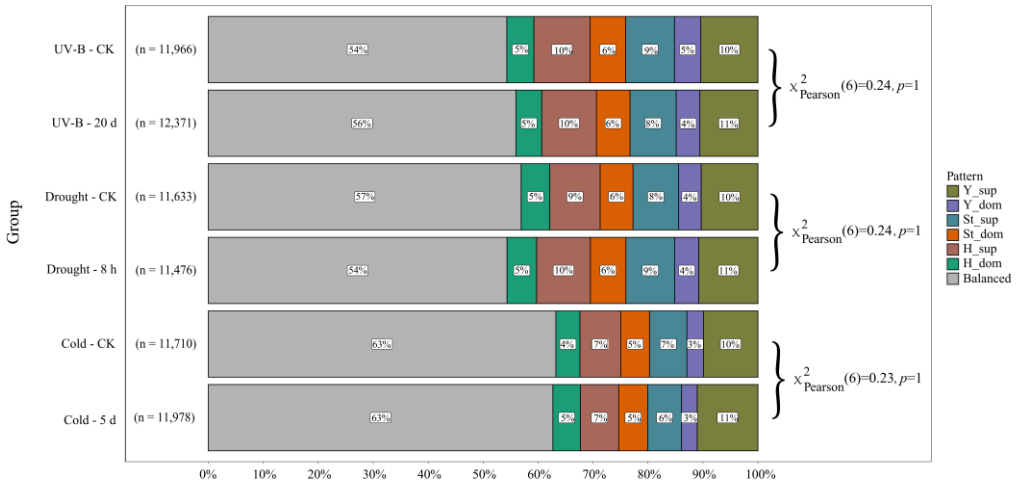
Supplementary Fig. 11. Under cold stress, there was no expression bias among the three subgenomes of *E. nutans*. The Cold - CK (above) and Cold - 5 d (below) treatment revealed the relative expression abundance of seven homologous expression categories of 11,710 and 11,978 1:1:1 orthologous gene pair from the three subgenomes of *E. nutans*, respectively.



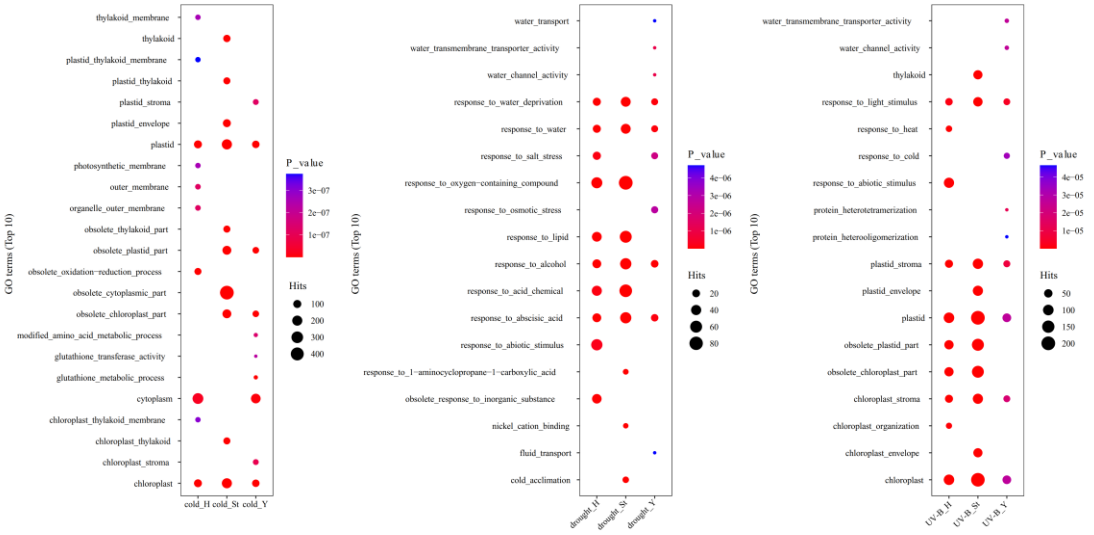
Supplementary Fig. 12. Under drought stress, there was no expression bias among the three subgenomes of *E. nutans*. The Drought - CK (above) and Drought - 8 h treatment (below) revealed the relative expression abundance of seven homologous expression categories of 11,633 and 11,476 1:1:1 orthologous gene pair from the three subgenomes of *E. nutans*, respectively.



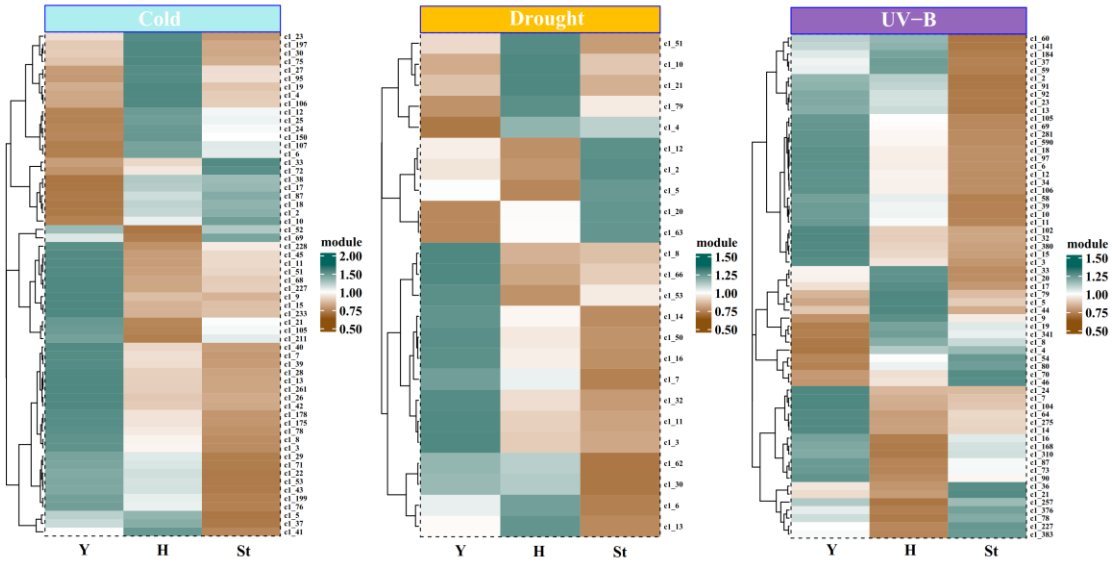
Supplementary Fig. 13. Under the cold, drought, and UV-B stresses, no bias was observed among the three subgenomes (St/Y/H), and a phylogenetic tree of single-copy genes carrying St/Y/H genotypes was constructed. a, The count of 1:1:1 homologous gene pair across seven homologous expression categories. b, The visualization of upregulated and downregulated genes in *E. nutans* under the three types of stresses. Group differences were assessed using chi-square test (df = 2), with significance levels: $*P < 0.05$, $**P < 0.01$, ^{ns} not significant. c, Based on the analysis of 2,046 single-copy genes that exclusively carry the St/Y/H genotypes, a phylogenetic and timescale tree were constructed with *Brachypodium distachyon* as an outgroup. The gray rectangle bars indicating the timescale between nodes.



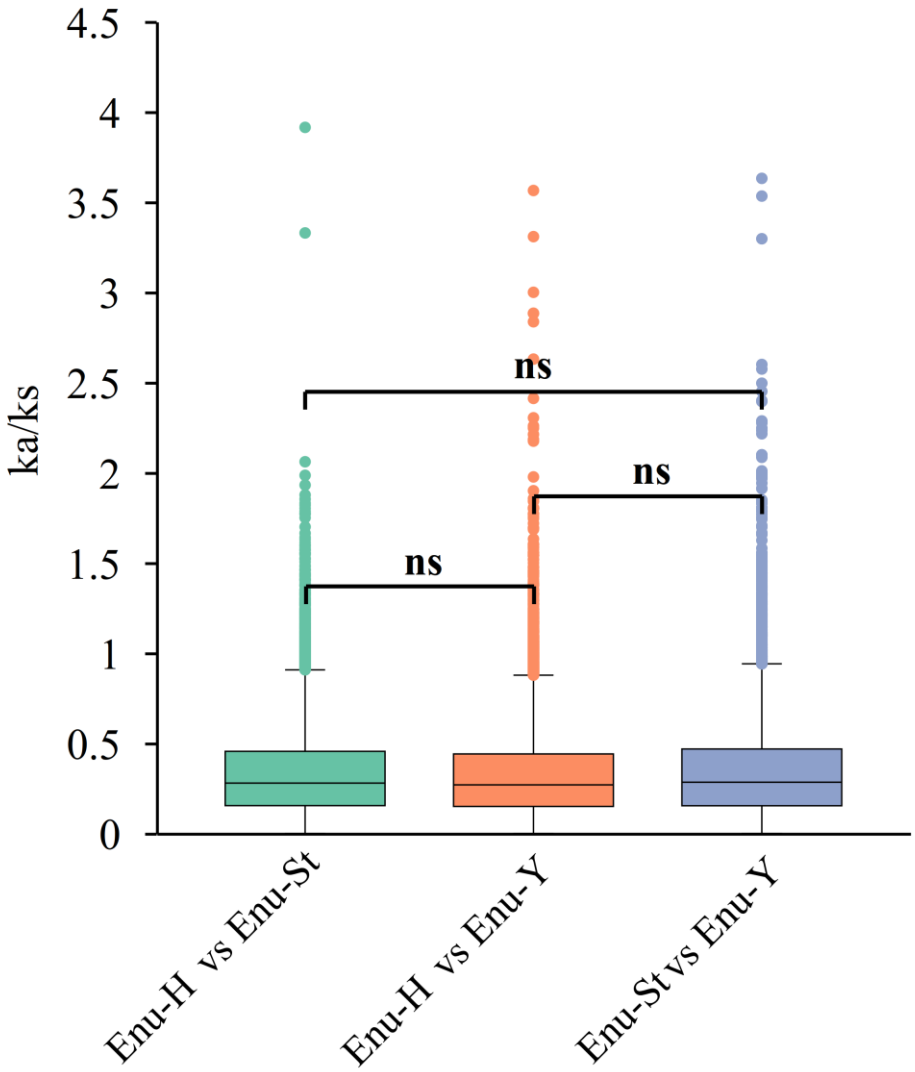
Supplementary Fig. 14. Proportion of seven homologous expression patterns (St/Y/H-suppressed, St/Y/H-dominance, and balanced) under three abiotic stresses (UV-B, cold, and drought). The chi-square test was conducted to compare each stress condition with its corresponding control, with degrees of freedom (df) of 6.



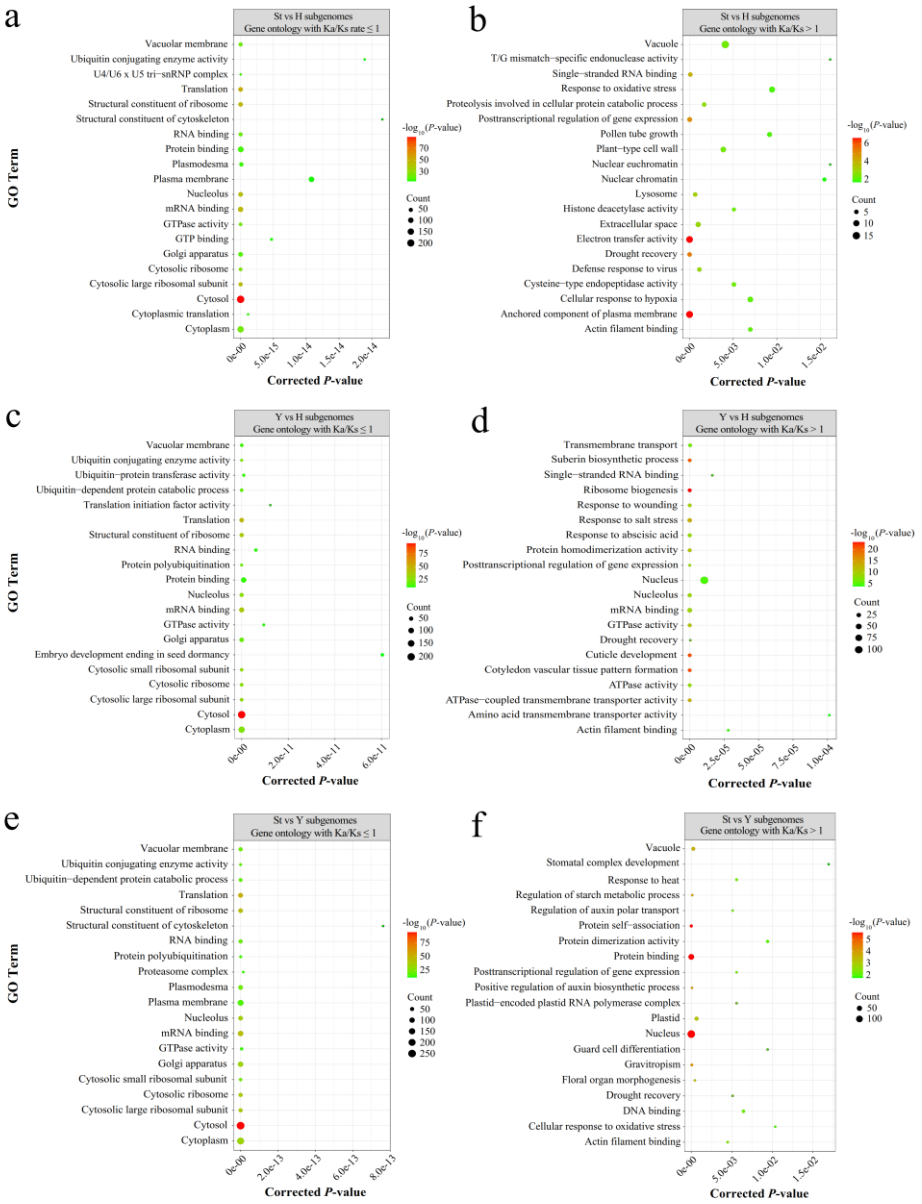
Supplementary Fig. 15. GO enrichment (top 10 terms) of the DEGs responding to cold (left), drought (middle) and UVB (right) stress conditions.



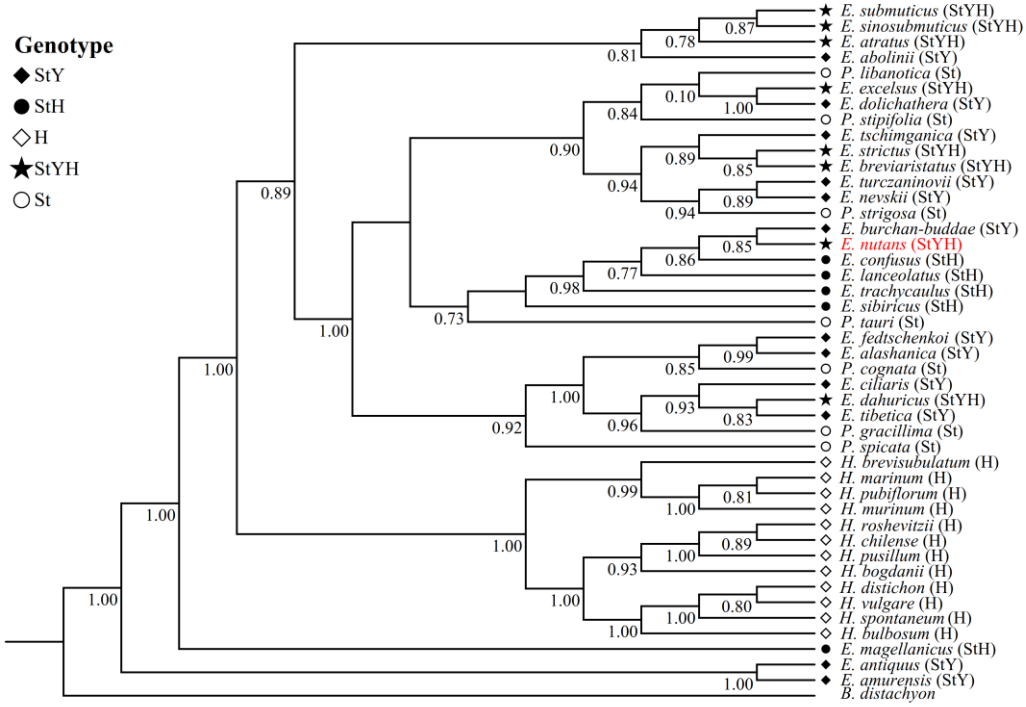
Supplementary Fig. 16. Heatmap representation of MEGENA co-expression modules showing the relation between the expected representation of each subgenome in the module based on the overall number of genes per subgenome and the observed one under cold (left), drought (middle) and UV-B (right) stresses. > 1 indicates higher than expected; < 1 indicates lower than expected; 1 indicates as expected. Source data are provided as a Source Data file.



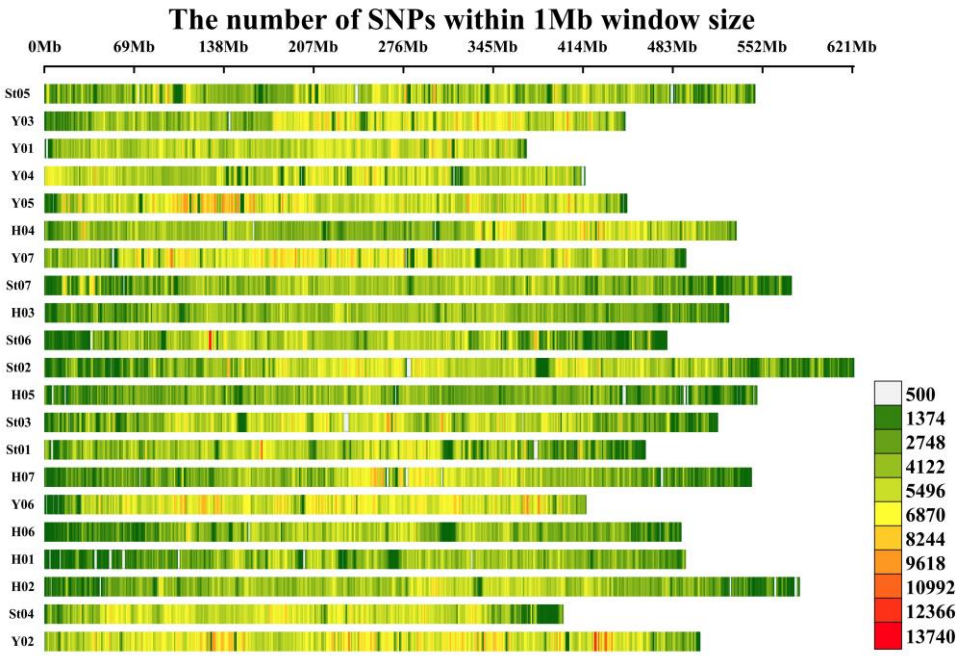
Supplementary Fig. 17. The Ka/Ks ratio between pairwise subgenomes in *E. nutans*. The central box represents the interquartile range, bounded by the first (25th percentile) and third (75th percentile) quartiles. The horizontal line inside the box indicates the median (50th percentile). The whiskers represent the standard deviation of the data. Differences between groups are compared using Welch two-sided *t*-tests, significant differences are indicated as follows: ns, not significant. n = 24133 for Enu-H vs Enu-St, n = 20313 for Enu-H vs Enu-Y, n = 23672 for Enu-St vs Enu-Y. Source data are provided as a Source Data file.



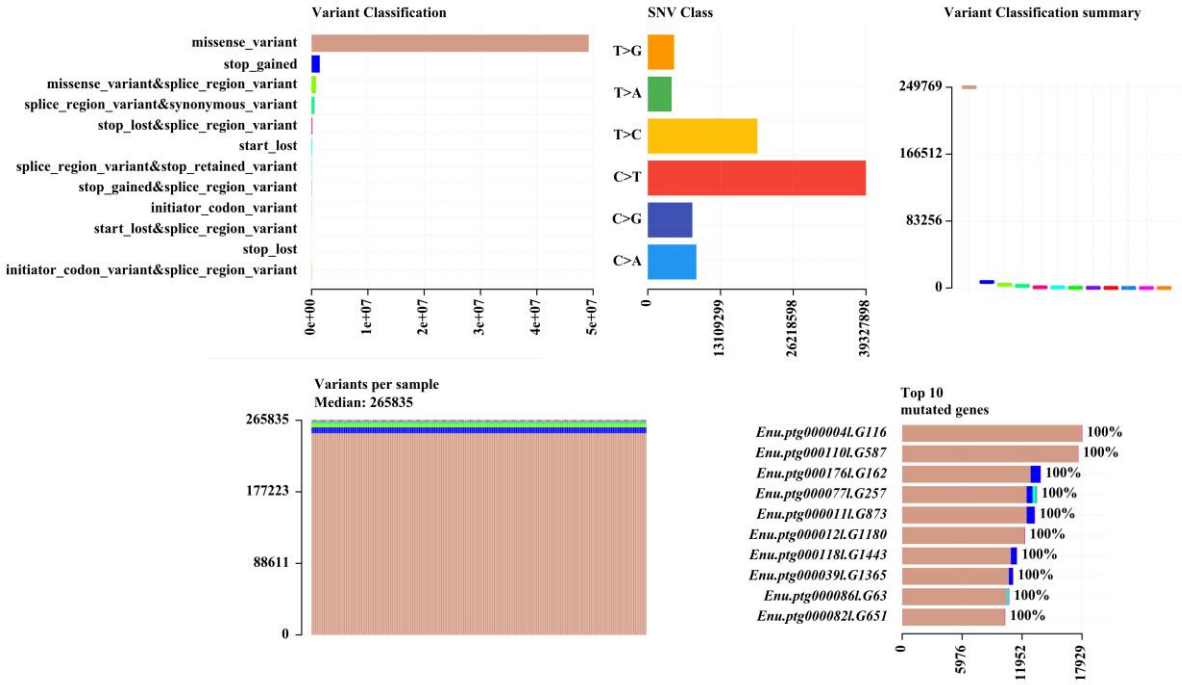
Supplementary Fig. 18. GO enrichment for homologous gene pairs with the Ka/Ks ratio ≤ 1 and > 1, respectively. a-b, The GO enrichment results of gene pairs between St and H subgenome with Ka/Ks ratio ≤ 1 (a) and > 1 (b). c-d, The GO enrichment results of gene pairs between Y and H subgenome with Ka/Ks ratio ≤ 1 (c) and > 1 (d). e-f, The GO enrichment results of gene pairs between Y and St subgenome with Ka/Ks ratio ≤ 1 (e) and > 1 (f).



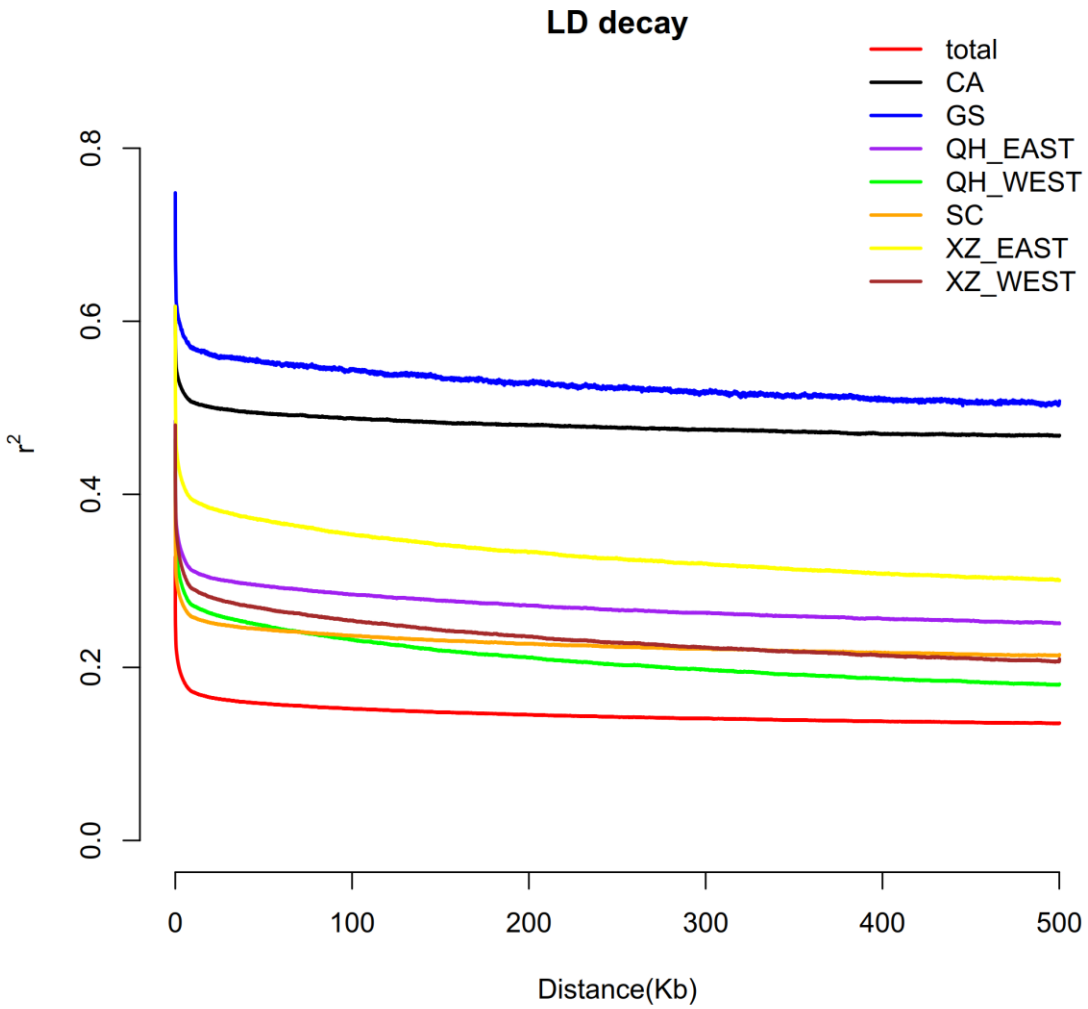
Supplementary Fig. 19. Phylogenetic tree constructed based on the shared chloroplast genes of St, H, StH, StY and StYH species.



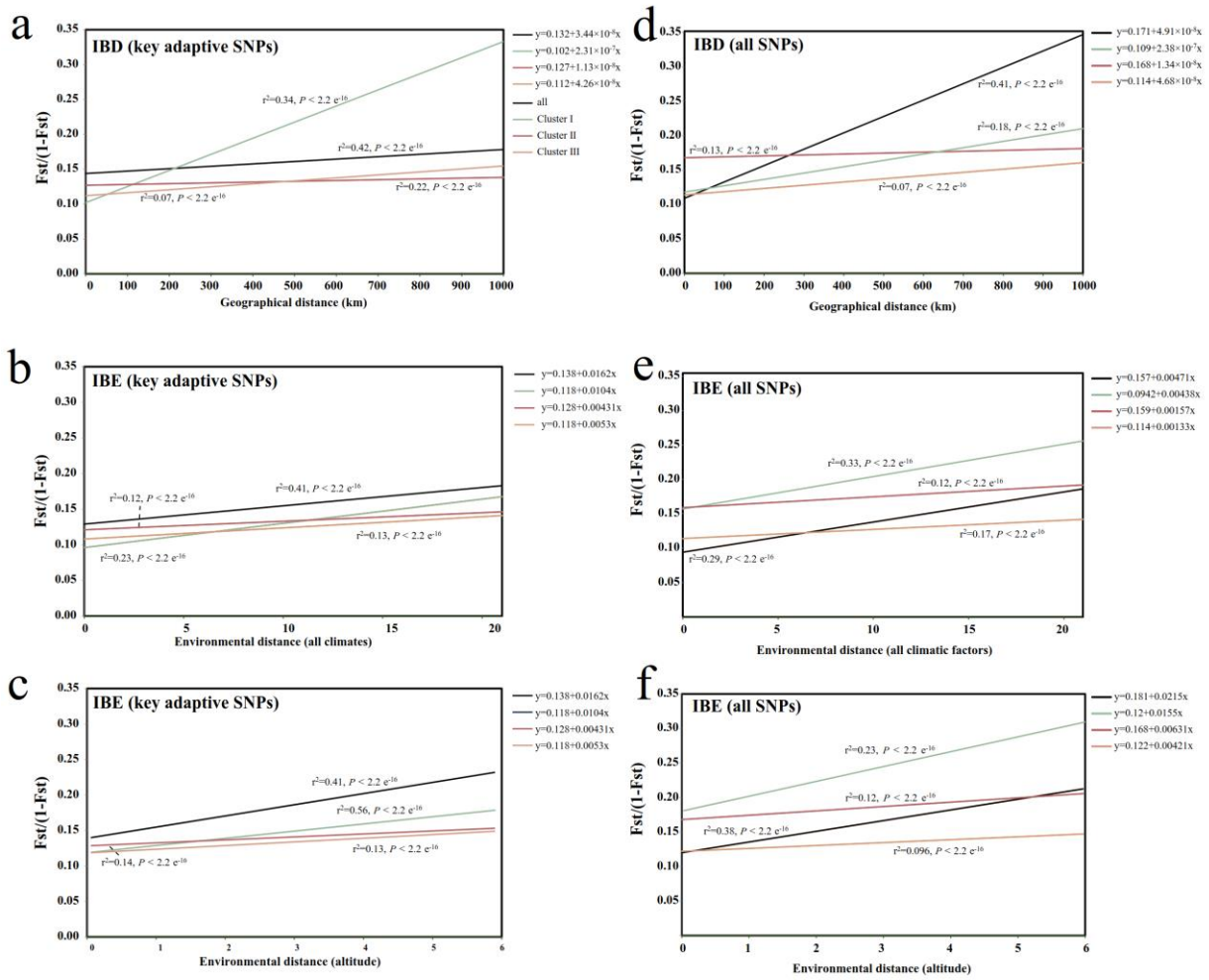
Supplementary Fig. 20. A total of 43,637,347 high quality single nucleotide polymorphisms (SNPs) are evenly distributed across the *E. nutans* genome. The density plot is under a 1 Mb window size.



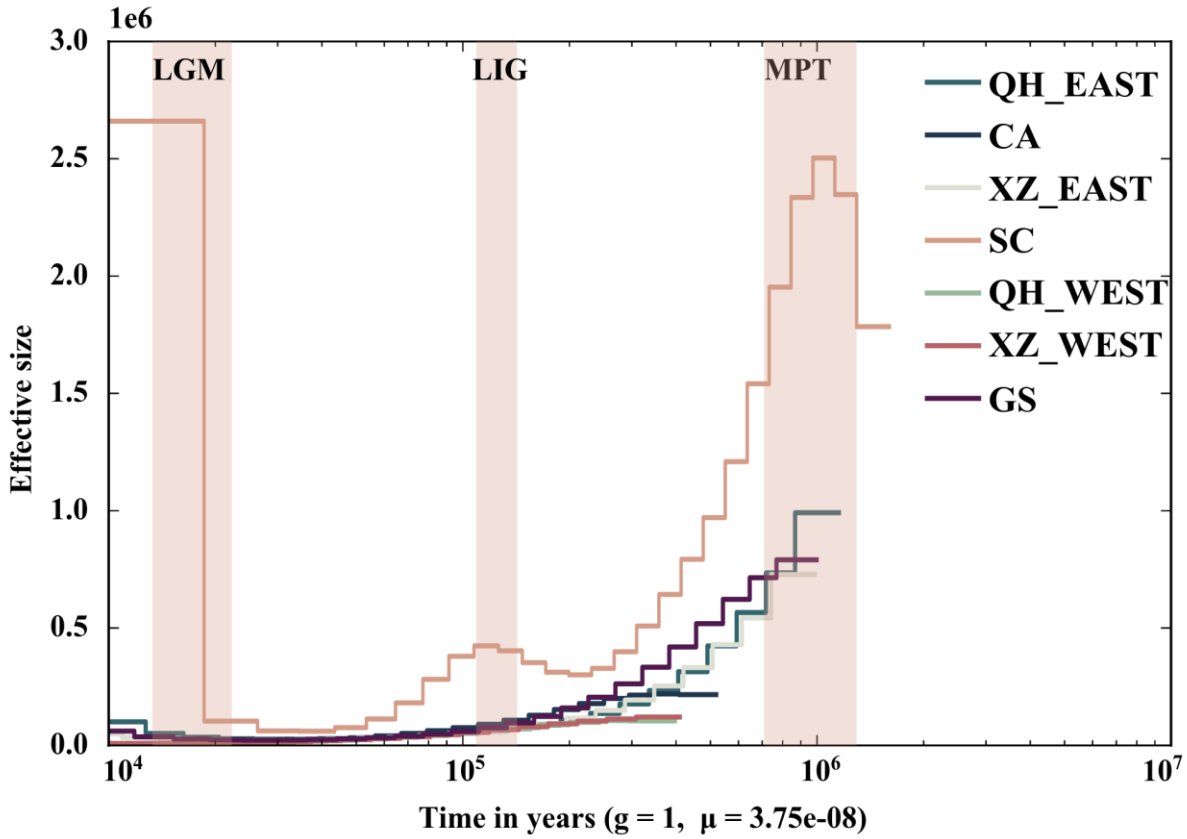
Supplementary Fig. 21. SNPs classification and statistics for variation types, variations per sample and top ten mutated genes. The upper panel presents the quantitative analysis of variant classification (left), SNV classes (middle), and variant classification summary (right). The lower panel displays the mutation profile across samples (left) and the top 10 genes with the highest SNP counts.



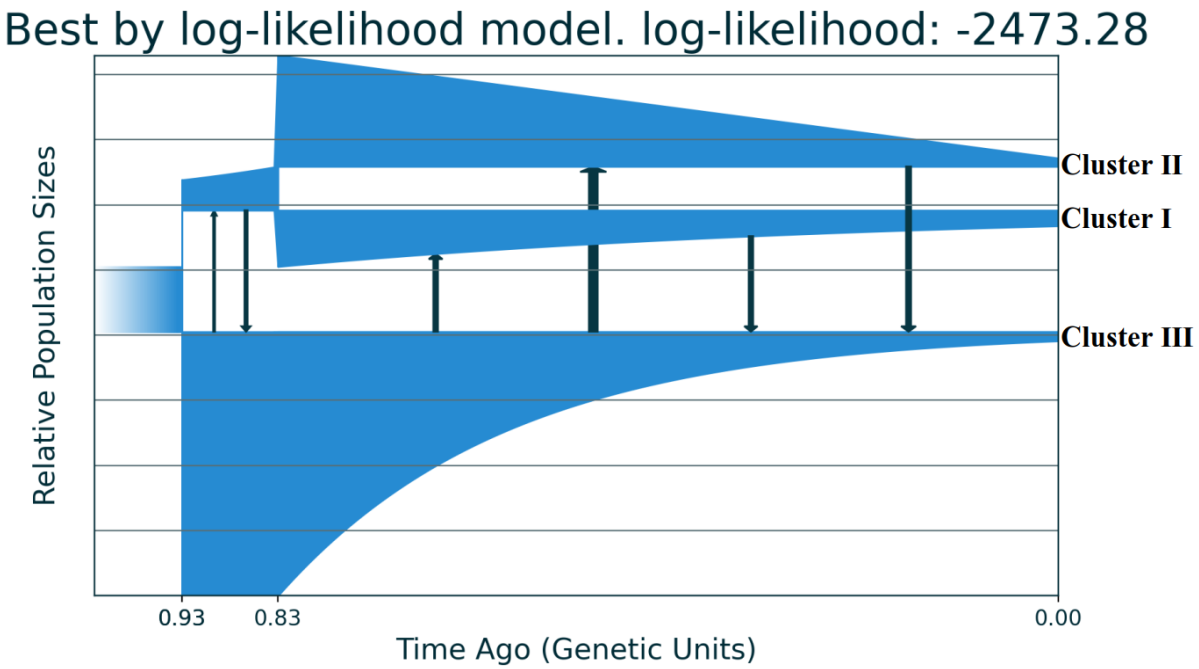
Supplementary Fig. 22. The LD decay of *E. nutans* at the species and geographical group levels



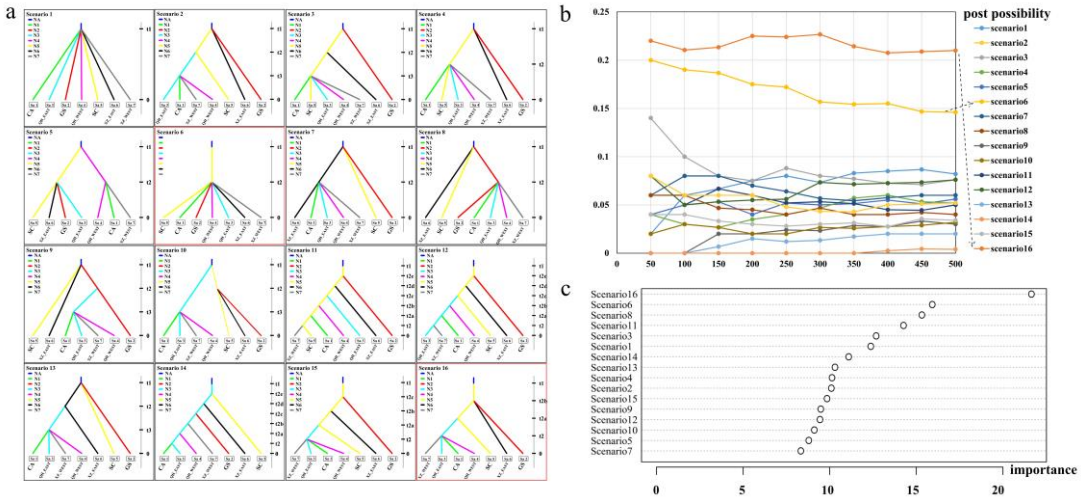
Supplementary Fig. 23. Isolation by distance (IBD) and isolation by environment (IBE) effects are estimated by Mantel test from the four aspects. All groups combined (all), only groups within Cluster I, only groups within Cluster II, and only groups within Cluster III. IBD and IBE patterns are analyzed based on both all SNPs (d, e, f) and key adaptive SNPs that identified in both lfmm and RDA scans (a, b, c).



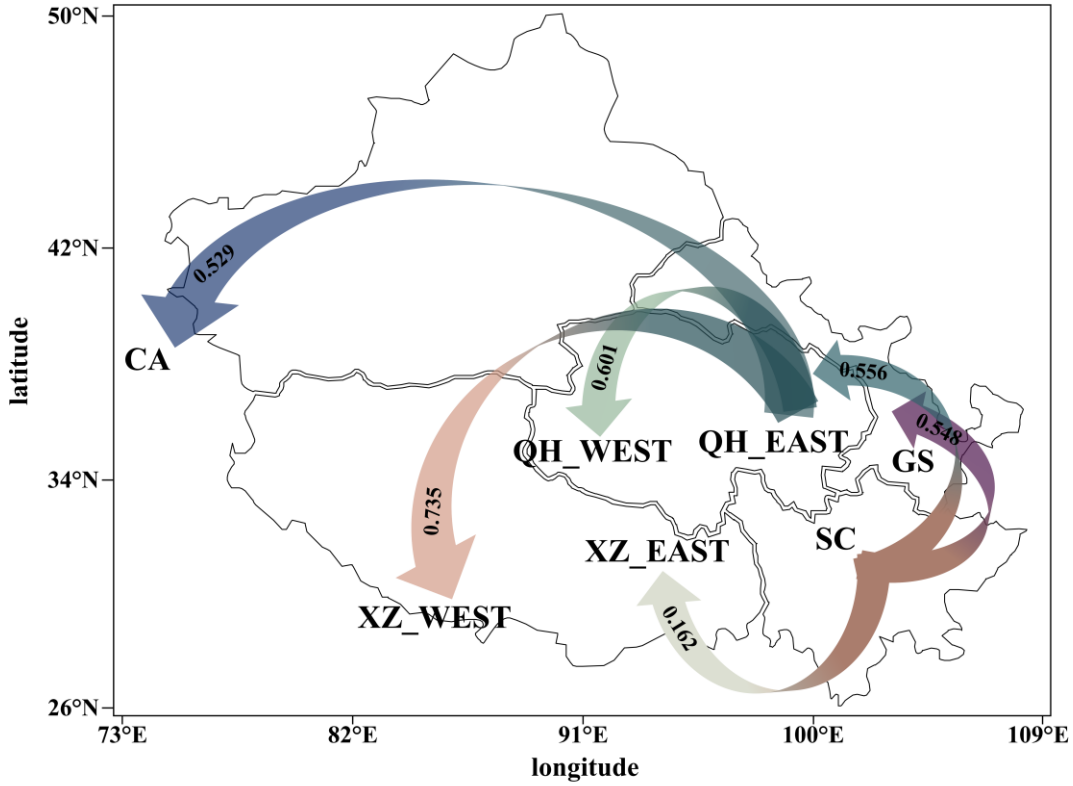
Supplementary Fig. 24. PSMC result of each geographic group. LGM, Last Glacial Maximum; LIG, Last glacial period; MPT, Mid Pleistocene Transition.



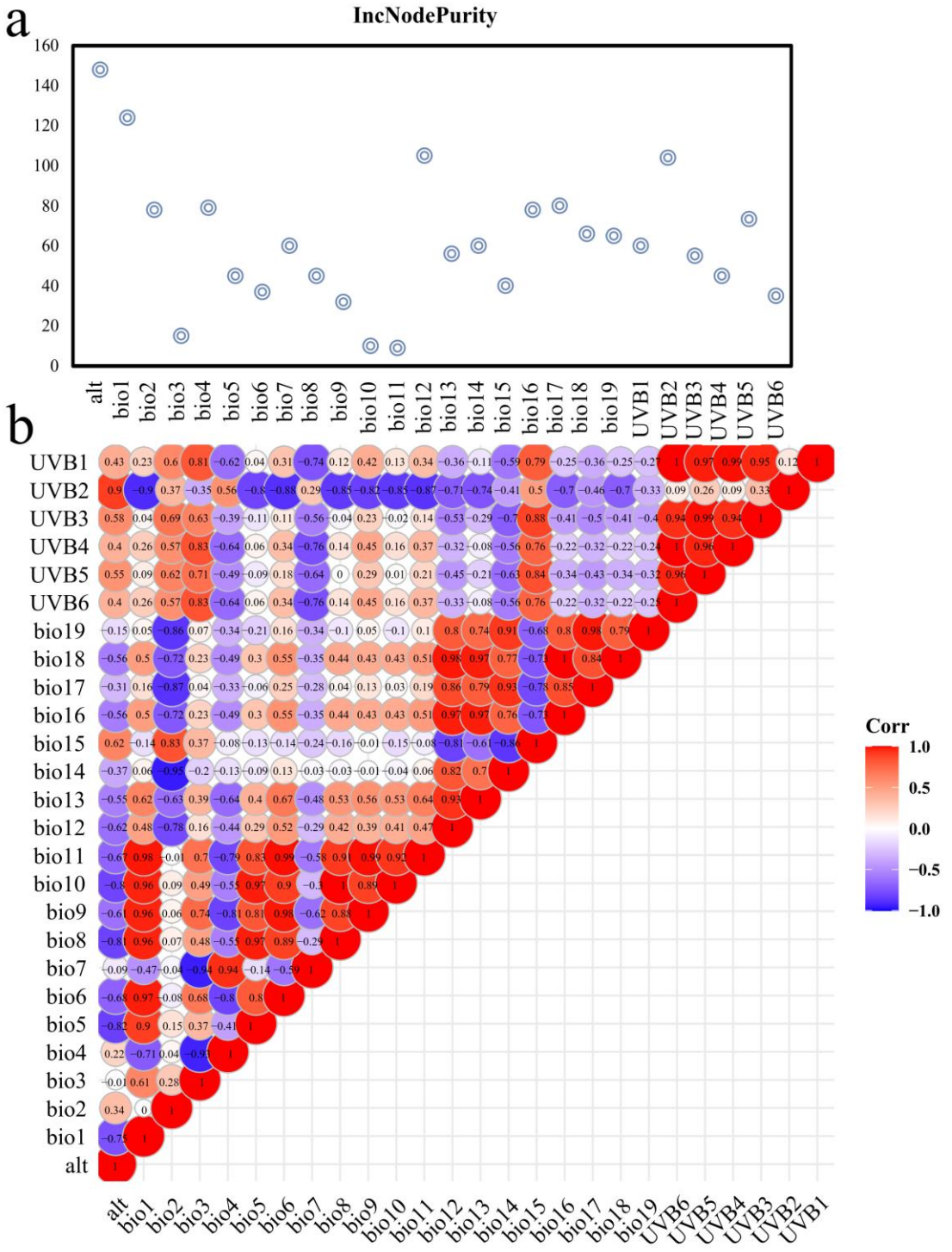
Supplementary Fig. 25. The GADMA result indicate the possible demographical history of three clusters.



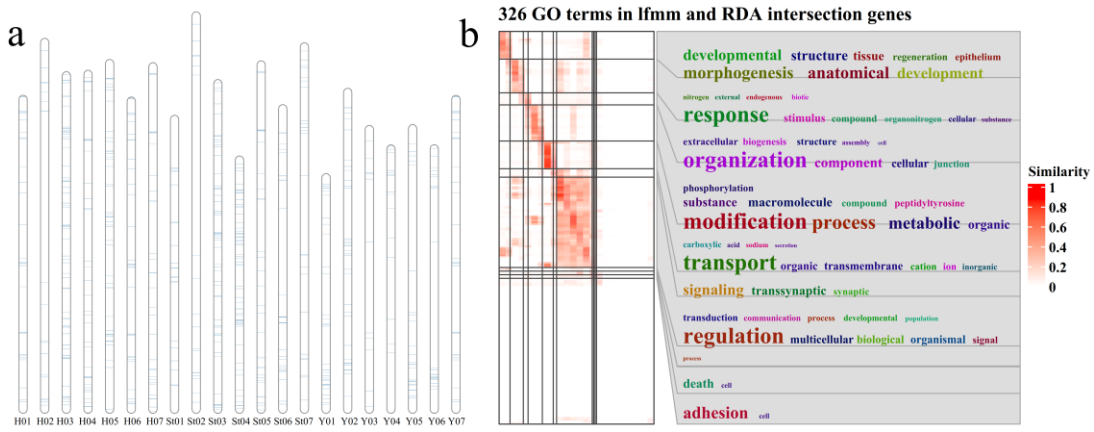
Supplementary Fig. 26. Prediction of the optimal demographic history model from 16 possible scenarios constructed in DIYABC. a, Totally 16 scenarios were assumed based on the basic knowledge from PSMC and GADMA results. The post-possibility (b) and random forest importance (c) values of each scenario are evaluated via DIYABC and DIYABC-RF, respectively. Source data are provided as a Source Data file.



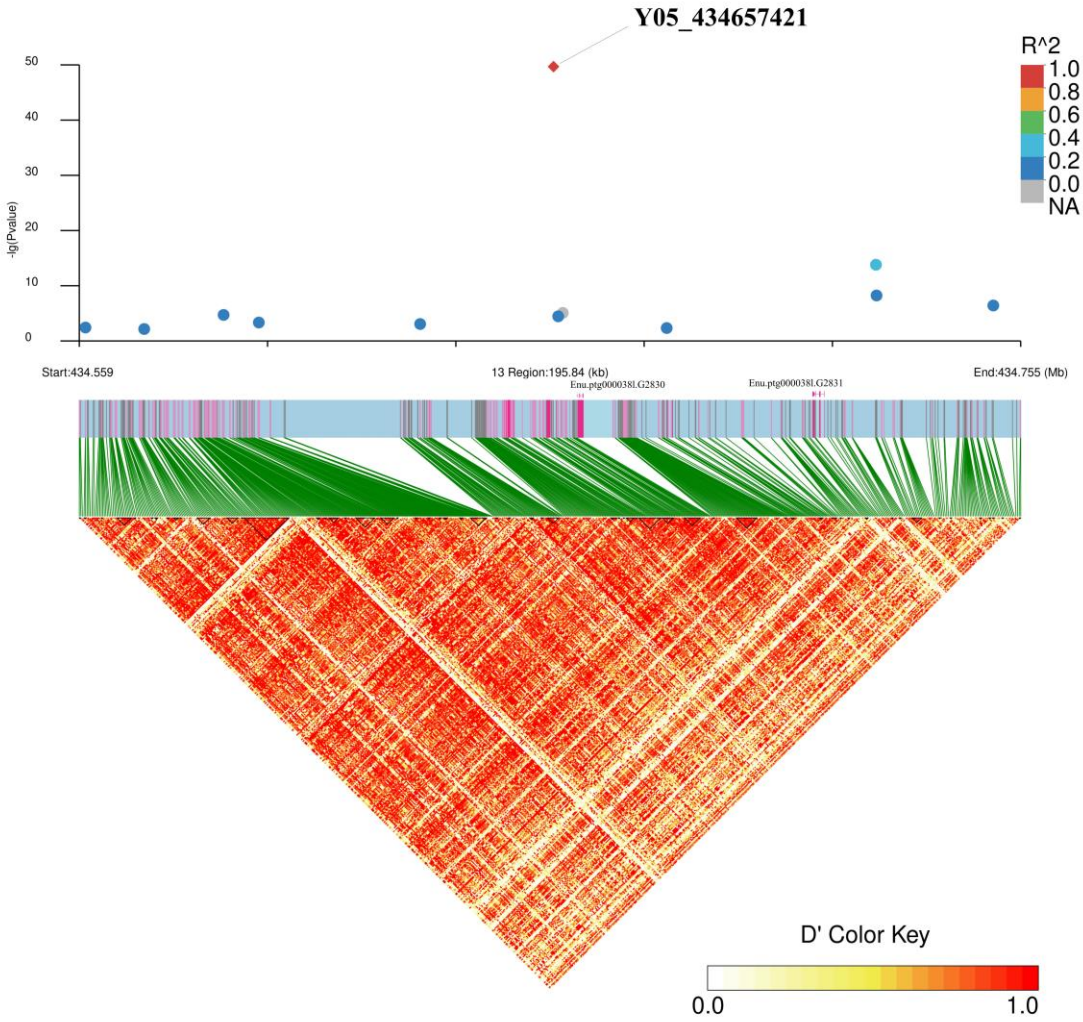
Supplementary Fig. 27. The inference of population origin and spreading rout, as well as the estimated historical gene flow (displayed at the end of the arrows).



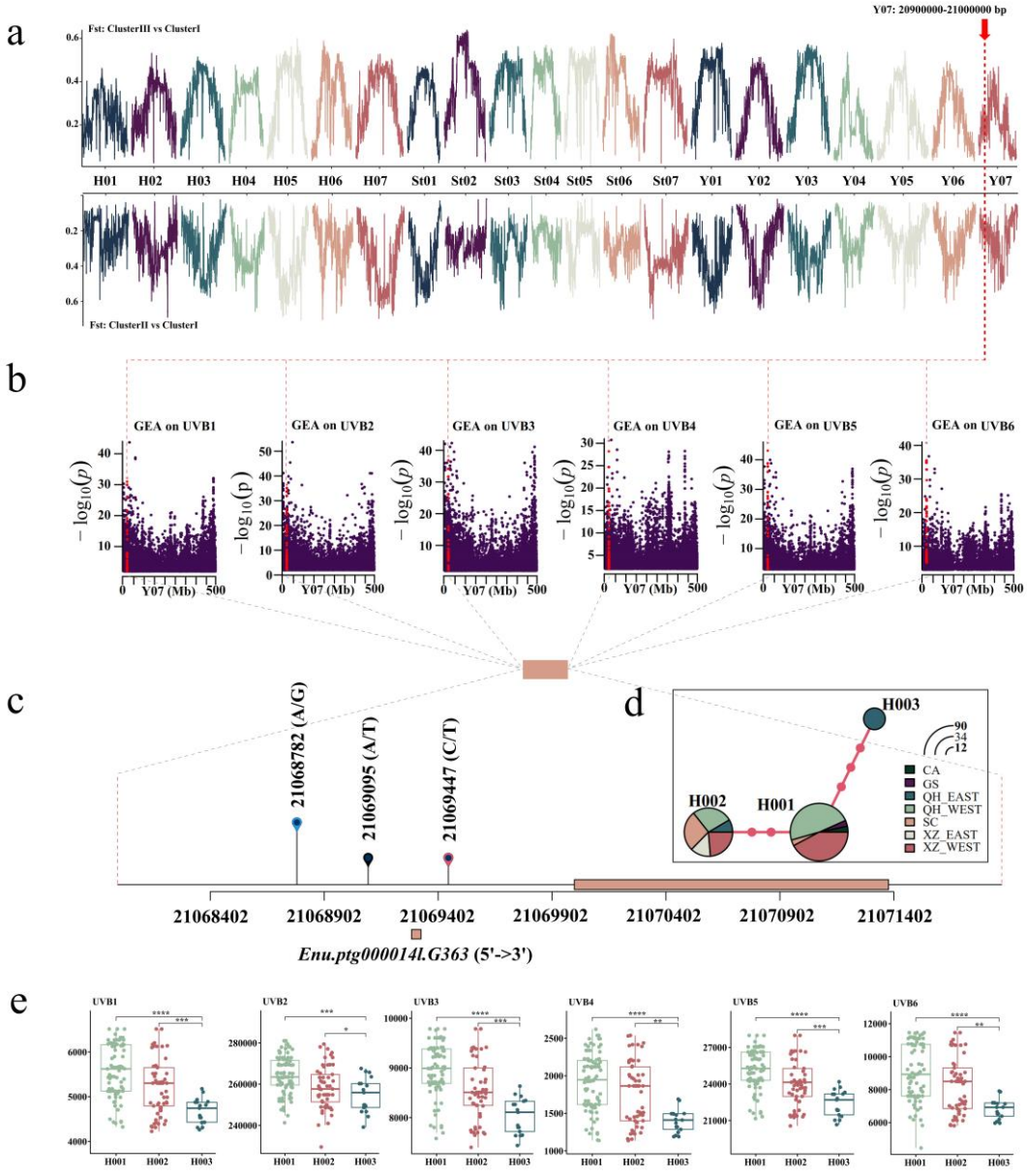
Supplementary Fig. 28. Screening of representative environmental factors using random forest method and correlation analysis. a, The IncNodePurity value of each variable indicates its importance. b, Correlation coefficient between pairwise variables, the color indicates the correlation value and the circle size indicates the degree of significance.



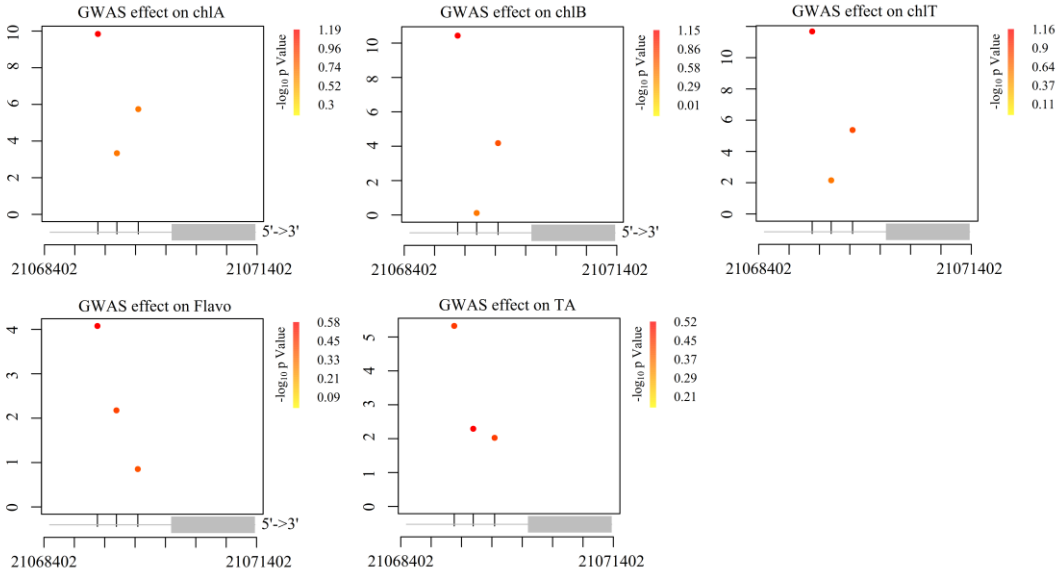
Supplementary Fig. 29. Chromosomal distribution of shared SNPs of lfmm and RDA analysis (a)**, and the genes associated with these SNPs and their GO enrichment classification** (b)**.**



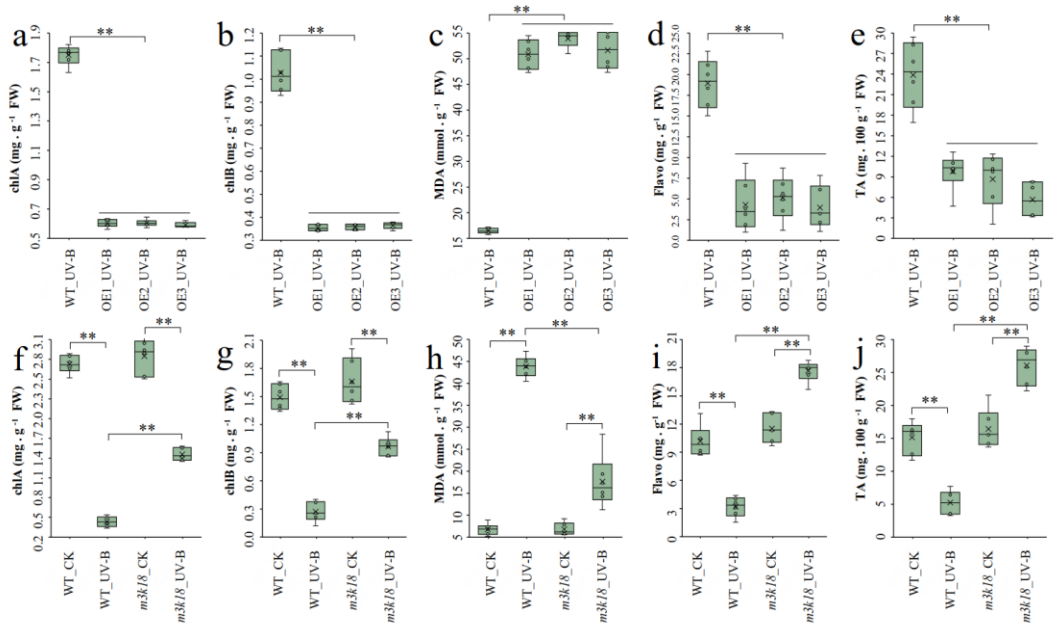
Supplementary Fig. 30. The zoomlocus Manhattan diagram surrounding the locus 'Y05_434657421' that located within the promoter of the gene *Enu.ptg000038l.2830*, and the degree of linkage disequilibrium around this region.



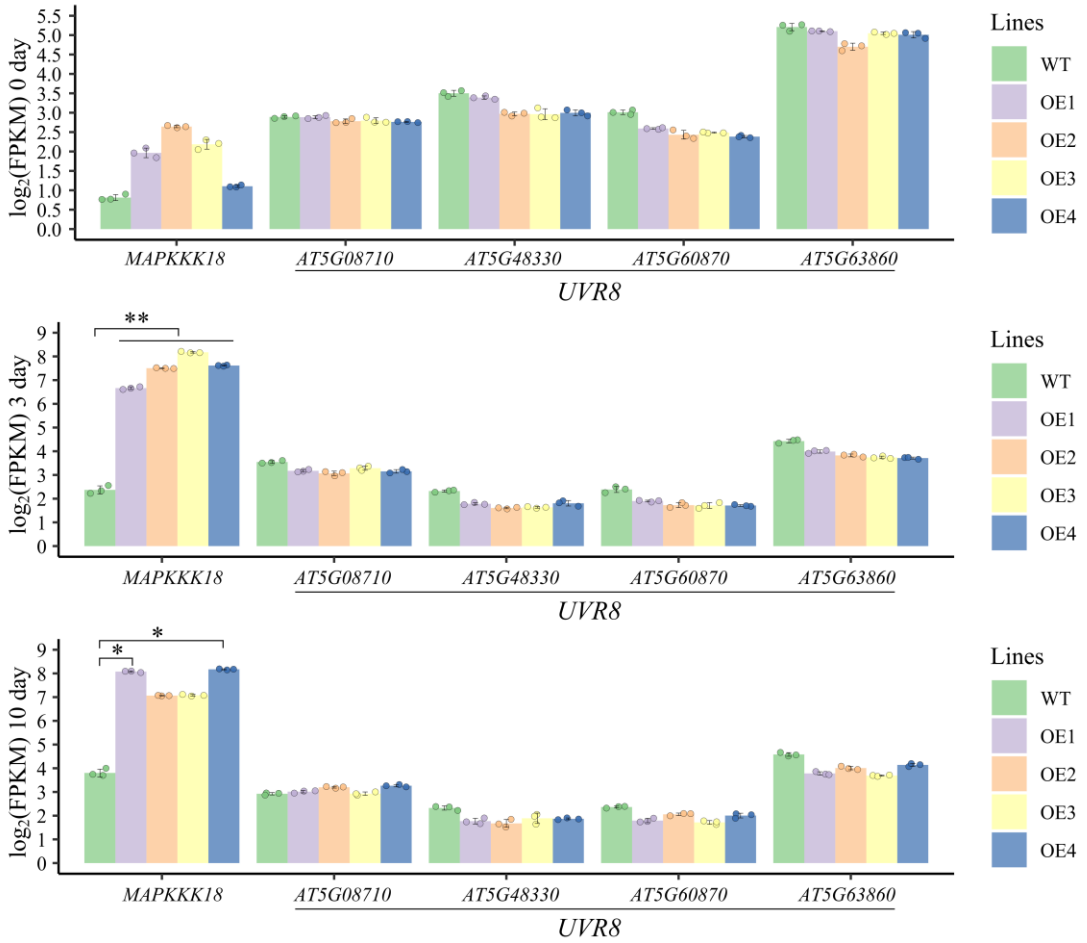
Supplementary Fig. 31. The results of the GEA scans of all six UV-B related environmental factors (UVB1 - UVB6) identified the gene *MAPKKK18* (*Enu.ptg000014l.G363*) in *E. nutans*. a, The Fst comparison with a window size of 100 Kb and a step size of 10 Kb across the whole genome between Cluster I and Cluster III and between Cluster I and Cluster II. b, GEA analysis showed a strong peak on the anterior part of chromosome Y07. c, The GEA locus (Y07: 20900000 - 21000000) selected the gene *MAPKKK18* (*Enu. ptg000014l.G363*) by LD linkage, which contained three SNPs within its promoter. d, The haplotype network constructed based on the three SNPs showed three haplotypes for all accessions, namely H001, H002 and H003, the color within each haplotype representing the origin of the accession from seven geographical groups. e, The differences of *in-situ* UV-B density ($J \cdot m^2 \cdot day^{-1}$) between H001 (n = 76), H002 (n = 52) and H003 (n = 12) were detected by Welch two-sided *t*-tests. The central box represents the interquartile range, bounded by the first (25th percentile) and third (75th percentile) quartiles. The horizontal line inside the box indicates the median (50th percentile), the whiskers represent the standard deviation of the data. Significant differences are indicated as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. Source data are provided as a Source Data file.



Supplementary Fig. 32. The GWAS effect scores of the three SNPs in the promoter region of *Enu.ptg000014l.G363* with traits chlA, chlB, chlT, Flavo and TA.



Supplementary Fig. 33. The chlA, chlB, MDA, Flavo and TA contents of *Arabidopsis thaliana* ecotype Col-0 (wild type, WT), overexpression (a-e) and mutant lines (f-j) of the *MAPKKK18* gene (*Enu.ptg000014l.G363*), with six biological replicates (n = 6). a-e showed the indicators of WT and OE lines after 20 days of UV-B treatment, corresponds to the growth status in the bottom left of Fig. 6h. f-j showed the indicators of WT and mutant lines before and after UV-B treatment, corresponds to the growth status in the right of Fig. 6h. In box plots, the central box represents the interquartile range, bounded by the first (25th percentile) and third (75th percentile) quartiles. The horizontal line inside the box indicates the median (50th percentile). The '×' symbol and whiskers represent the mean and standard deviation of the data. Differences between groups are compared using Welch two-sided *t*-tests, significant differences are indicated as follows: $*P < 0.05$, $**P < 0.01$. Source data are provided as a Source Data file.



Supplementary Fig. 34. The expression profile of *MAPKKK18* and *UVR8* genes in WT and OE *Arabidopsis thaliana* lines across 0 day (top panel), 3 days (middle panel) and 10 days (bottom panel) of UV-B treatment. The data represent the mean and standard deviation of three biological replicates (n = 3). Differences between groups are compared using two-sided Welch *t*-tests. Significant differences between the lines are indicated as follows: **P* < 0.05, ***P* < 0.01. Source data are provided as a Source Data file.

Supplementary Table 1. Summary of the HiFi and Hi-C sequencing data of the *E. nutans* genome.

Data type	Total raw reads number	Total raw bases (Gb)	Q20 (%)	Q30 (%)	Average read length (bp)	Read length N50 (bp)	Sequence coverage (×)	N content (%)
HiFi reads	20,103,316	366.72	-	-	18,255.60	18,316.80	~35.03	-
HI-C reads	3,869,994,364	1,161.00	97.25	94.08	-	-	~110.89	0.35

Supplementary Table 2. Genome assembly and annotation statistics of *E. nutans*.

Assembly	Assembled genome size (bp)	10,469,055,907
	GC content (%)	45.97
	Number of contigs	1,121
	Number of scaffolds	877
	N50 of contig (bp)	76,252,896
	N90 of contig (bp)	20,358,283
	N50 of scaffold (bp)	503,834,971
	N90 of scaffold (bp)	416,155,416
	Longest contig (bp)	288,267,451
	Shortest contig (bp)	15,140
	Anchored rate of Hi-C sequences to the pseudo-chromosomes (%)	95.05
	Complete BUSCOs (%)	99.5
	LAI	13.97
Annotation	Number of protein-coding genes	120,538
	Average gene length (bp)	3,436.38
	Average CDS length (bp)	1,233.23
	Average exon number per gene	5.3
	Average exon length (bp)	232.84
	Average intron length (bp)	512.78
	Content of TEs Interspersed Repeats (%)	83.02
	Number of tRNA	15,600
	Number of rRNA	23,259
	Number of snRNA	3,543
	Complete BUSCOs (%)	99.4

Supplementary Table 3. Statistics of repetitive elements in the *E. nutans* genome.

Classifications			Number of elements	Length (bp)	% of genome
Class I	LTR		3,791,629	6,549,468,585	62.560
		Copia	730,797	1,389,779,400	13.580
		Gypsy	2,235,405	4,172,906,224	39.589
		Other	825,427	986,782,961	9.426
	LINE	-	354,145	210,564,697	2.011
		L1	302,283	198,708,739	1.898
		Other	51,862	11,855,958	0.113
	SINE	-	52,128	10,574,197	0.101
		Penelope	3,765	210,578	0.00002011
		Other	48,363	10,363,619	0.098
Class Ⅱ	DNA	-	2,071,866	1,500,297,164	14.331
		Hat-Ac	128,864	30,670,828	0.293
		Tc1-IS630-Pogo	291,890	68,282,907	0.652
		PIF-Harbinger	197,928	79,549,164	0.760
		Other	1,453,184	1,321,794,265	12.626
Rolling-circles			125,840	32,570,413	0.311
Unclassified			1,380,941	420,640,196	4.018
Total interspersed repeats			-	8,691,544,839	83.021
Small RNA			63,730	36,718,010	0.351
Simple repeats			585,851	45,457,632	0.434
Satellites			86,531	59,058,071	0.564
Low complexity			78,448	4,640,773	0.044

Supplementary Table 4. The annotated genes of the assembled *E. nutans* genome.

Sources	Number	Percent (%)
SwissProt	35,898	29.78
KEGG	4,161	3.45
KOG	38,658	32.07
GO	30,753	25.51
NR	51,741	42.93
Interproscan	42,402	35.18
TrEMBL	51,251	42.52
Annotated genes	110,390	91.58
Total gene models	120,538	-

Supplementary Table 5. Statistics of the differentially expressed and associated genes under UV-B, cold and drought stress. 3/3 triads: the number of 1:1:1 homoeologous gene pairs, 2/3 triads: the number of 1:1 homoeologous gene pairs, 1/3 triads: the number of orthologous genes within one of the St/H/Y subgenomes. The collinear ratio is calculated as (3 * 3/3 triads + 2 * 2/3 triads) / total genes.

Source	Factors	Total genes	3/3 triads	2/3 triads				1/3 triads				Single copy genes				Collinear ratio
				Total	S:H	S:Y	H:Y	Total	S	H	Y	Total	S	H	Y	
Gene expression	UV-B	6731	554	716	239	258	219	1469	543	454	472	2168	833	778	557	45.97%
	Cold	6407	637	641	263	219	159	1341	471	500	370	1873	739	645	489	49.84%
	Drought	2944	213	290	105	98	87	664	240	219	205	1061	438	342	280	41.41%
GWAS	UV-B	3782	0	40	19	19	2	1152	513	214	425	2550	1183	467	902	2.12%
GEA	UV-B	1606	0	4	0	4	0	464	145	16	303	1134	381	54	699	0.50%
	Temperature	2181	0	17	3	10	4	559	220	44	295	1588	689	211	688	1.56%
	Precipitation	1333	0	11	0	10	1	374	107	58	209	937	388	156	393	1.65%

Supplementary Table 6. The SNPs classification based on location and consequence annotated with snpEff.

Type (alphabetical order)	Count	Percent
5_prime_UTR_variant	6	0%
downstream_gene_variant	2,721,400	5.51%
initiator_codon_variant	109	0%
intergenic_region	42,249,710	85.55%
intron_variant	956,623	1.94%
missense_variant	253,763	0.51%
splice_acceptor_variant	1,450	0.00%
splice_donor_variant	1,478	0.00%
splice_region_variant	27,579	0.06%
start_lost	691	0.00%
stop_gained	7,548	0.02%
stop_lost	824	0.00%
stop_retained_variant	272	0.00%
synonymous_variant	167,890	0.34%
upstream_gene_variant	2,999,342	6.07%

Supplementary Table 7. The FISH probes used in this study.

Oligo-probe ID	Probe sequence
Oligo-pTa535	AAAAACTTGACGCACGTCACGTACAAATTGGACAAACTCTTTCGGAGTATCAGGGTTTC
Oligo-pSc119.2	CCGTTTTGTGGACTATTACTCACCGCTTTGGGGTCCCATAGCTAT
Oligo-HvCSR	ACAACGACAACAACGACAATGACGAGA

Supplementary Table 8. The primers used in the qRT-PCR experiment.

Type	Primer	Primer sequence	Tm/℃	Treatment
qRT-PCR	*Enu.ptg000014l.G363* - F	GAAGACAGCAGCGATCCAC	59.9	UVB
qRT-PCR	*Enu.ptg000014l.G363* - R	CTAATAATGGCGAGGCGAGA	58.1	UVB
qRT-PCR	*Enu.ptg000038l.G2830* - F	AGAGGCTTCAAAGCACCTTG	58.1	Cold
qRT-PCR	*Enu.ptg000038l.G2830* - R	GGTTGATAATTGTCGGCAGC	58.1	Cold
qRT-PCR	*Enu.ptg000007l.G716* - F	TGTAATGGAAGAGACCGGGA	58.1	Drought
qRT-PCR	*Enu.ptg000007l.G716* - R	AACTTCGTACAACTTGGAGACC	58.6	Drought
qRT-PCR	*U2AF* - F	ATCGCTGCTCTCGCATCCATAAC	58.2	Internal reference
qRT-PCR	*U2AF* - R	TGCTGCTGCCTGATCTTCCTC	58.5	Internal reference