

Supplementary Note

Identification of tandem duplication gene in *S. spontaneum* Np-X

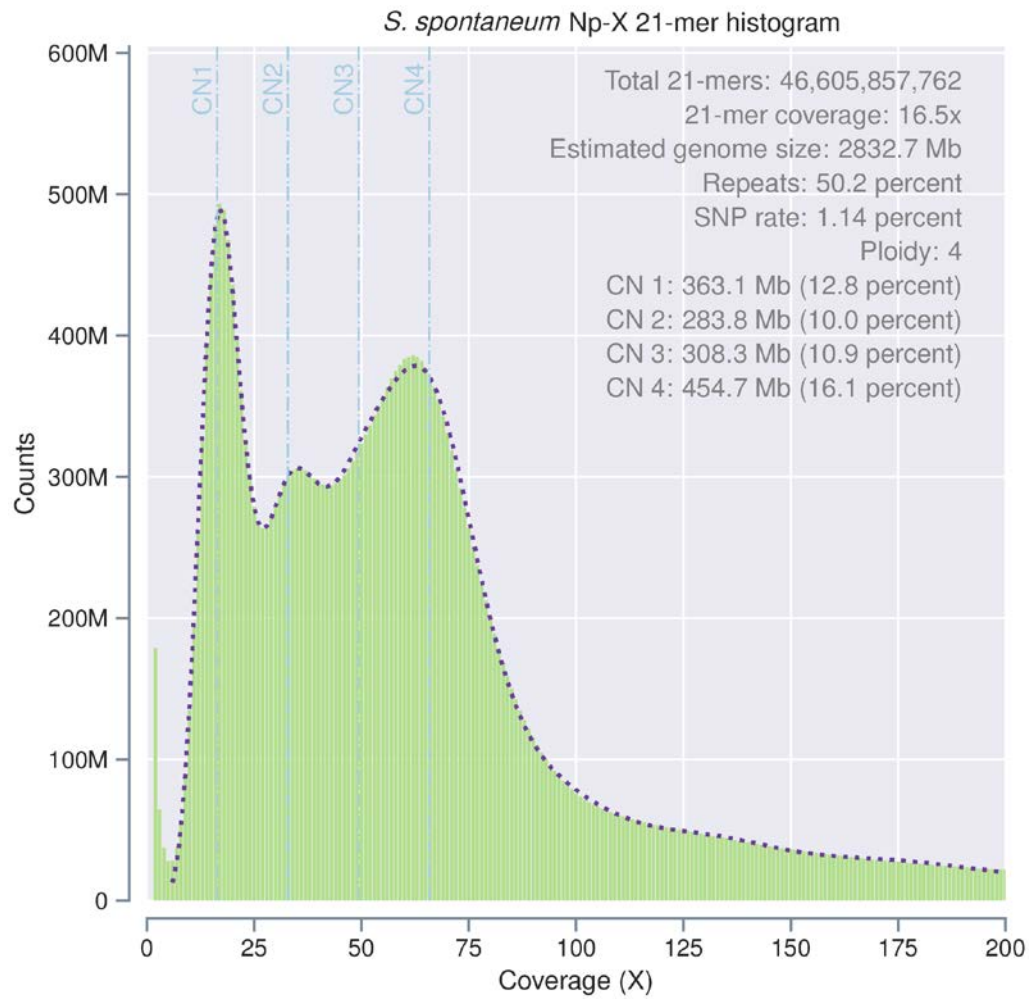
We identified 4,316 tandem arrays of two or more genes in the *S. spontaneum* Np-X genome using MCScanX¹ (in BLASTP E = 1e -10 and num_alignment = 5). A total of 9,060 genes were involved in such tandem arrays, accounting for about 7.37% of all genes. Among these, the tandem arrays of two genes were found more generally on chromosomes in *S. spontaneum* Np-X (3,939 of this type). The number of tandem arrays tended to decrease with increased average array size. The largest type of tandem array (2 in total) we identified is a single array of six genes. A total of 4,276 tandem arrays carrying 2–3 genes were found on all chromosomes. We identified 40 tandem arrays carrying 4–6 genes (Supplementary Fig. S34).

The downstream program *duplicate_gene_classifier* of MCScanX package¹ was used for gene origin assignment of the entire set of *S. spontaneum* Np-X genes, which resulted in the gain of a total of 3,652 (approximately 2.97% of the total set of genes) tandem duplicated genes located on all chromosomes. The number of tandem duplicates on the four sets of homologous chromosomes was evenly distributed, and the number of tandem duplicated genes on Chromosome 5 (Chr5A/B/C/D) accounted for the highest percentage of all genes on corresponding chromosomes (4.51–5.72%) (Supplementary Fig. S35).

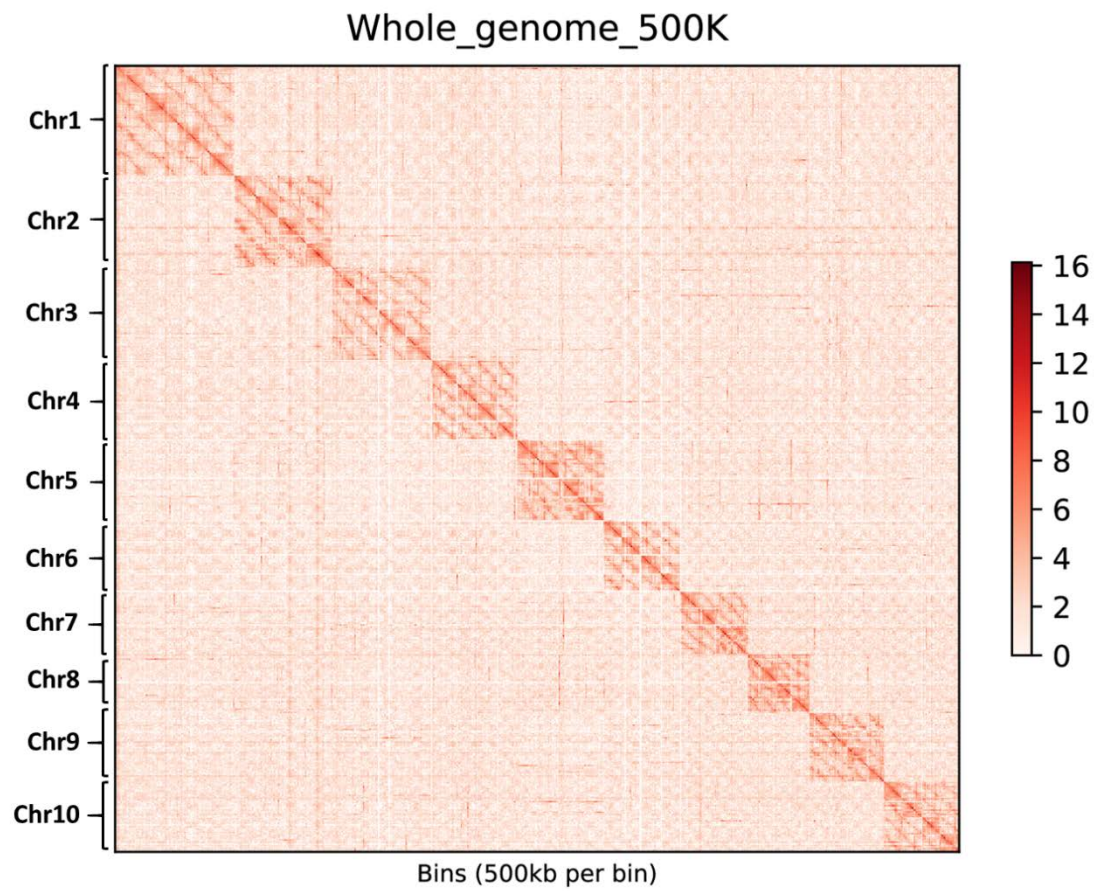
A total of 84 genes in tandem arrays, including 45 tandem duplicates were found in the recombination regions of Chromosome 5 and Chromosome 8 in the *S. spontaneum* Np-X genome (Table 1). Functional annotations of genes in tandem arrays were obtained using eggNOG-mapper² (<http://egglog-mapper.embl.de/>). We then compared the functional annotations of genes in tandem arrays between the non-recombinant and recombinant regions of these chromosomes. We annotated two tandem arrays carrying four genes in the recombination regions, as members of the Subtilase-encoding and Alginate lyase-encoding gene families, respectively. Plant subtilases, a type of serine peptidase, target cell walls regulate the properties of the cell wall and the activity of extracellular signaling molecules that play roles in plant growth and development, including embryogenesis, cuticle formation, programmed cell death, organ abscission and senescence, and responses to their biotic and abiotic environment³. Some proteases within the large subtilase (SBT) family have been named phytaspases due to their strict Asp specificity for peptide and protein targets. Further, many such tandemly arrayed genes have been found on chromosomes 1, 4, and 8 in the tomato genome⁴. Genes encoding putative alginate lyase occupies a relatively high proportion of the tandemly arrayed gene annotations in the recombination regions (7.89%). These enzymes convert alginate into unsaturated monosaccharides via saccharification, and could be useful biocatalysts for future applications in biochemistry and biofuels fields⁵. Interestingly, we annotated the functions of the largest tandem arrays of six genes in the non-recombinant regions as Disease resistance protein and RNA-mediated transposition, respectively. Among the tandemly arrayed genes in both the recombinant and non-recombinant regions of these chromosomes are some unique annotations and also some common annotations. The unique annotations of tandemly arrayed genes in the recombinant regions include those encoding aaa-type ATPase family protein, members of the disease-resistance NB-LRR gene family, and genes encoding PsaA and PsaB, which bind P700 (Supplementary Fig. S36). The unique annotations of tandemly arrayed genes in the non-recombinant regions includes those encoding members of the cytochrome P450 family, Ser-Thr protein kinase family, the UDP-glycosyltransferase family, and others (Supplementary Fig. S36), in relatively high

proportions. The common annotations of tandemly arrayed genes in both the non-recombinant and recombinant regions included those encoding NB-ARC domain proteins, a Disease resistance protein, and a Protein kinase domain (Supplementary Fig. S36).

Supplemental figures



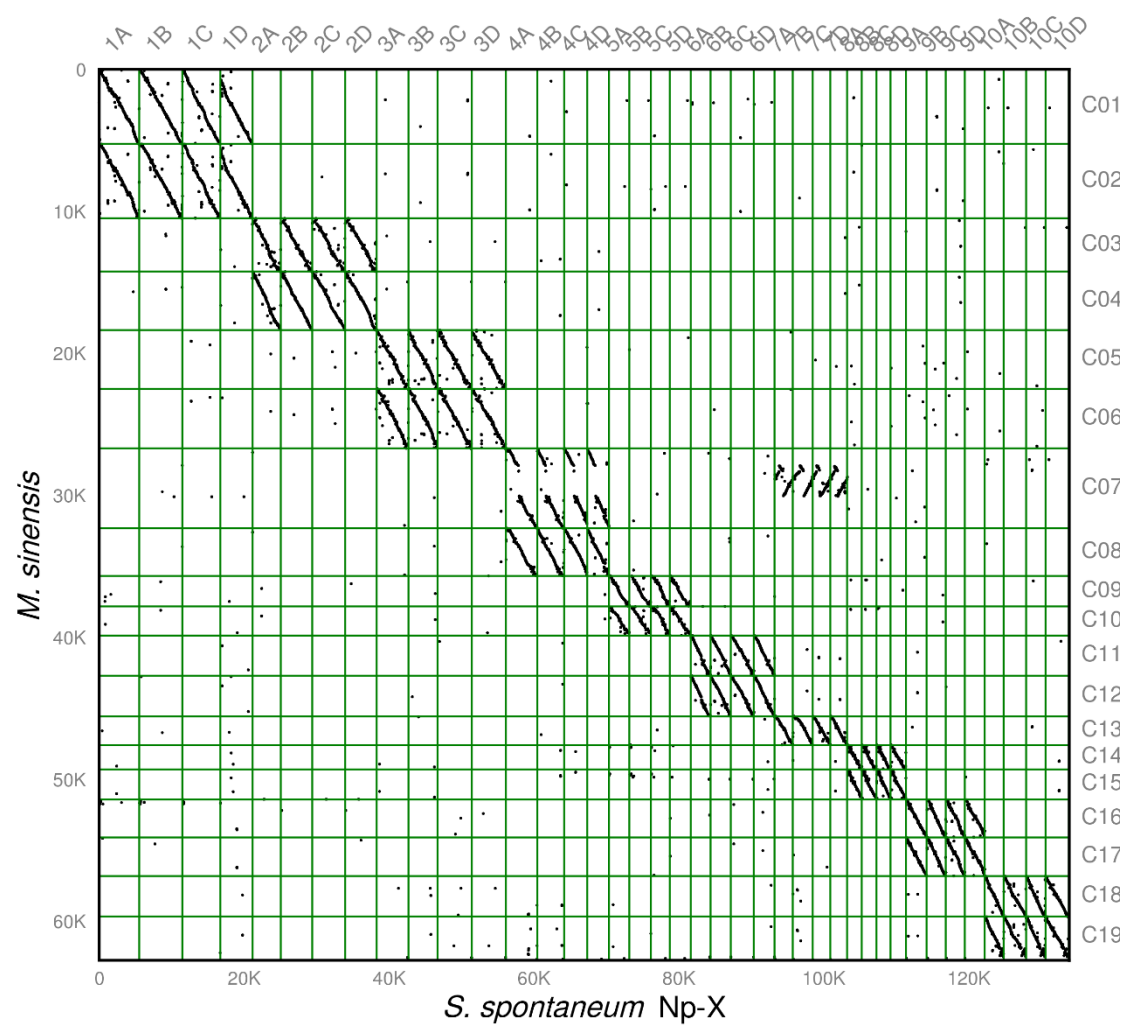
Supplementary Fig. S1. Estimation of *S. spontaneum* Np-X genome size based on a genomic survey of 21-Kmer frequency distribution.



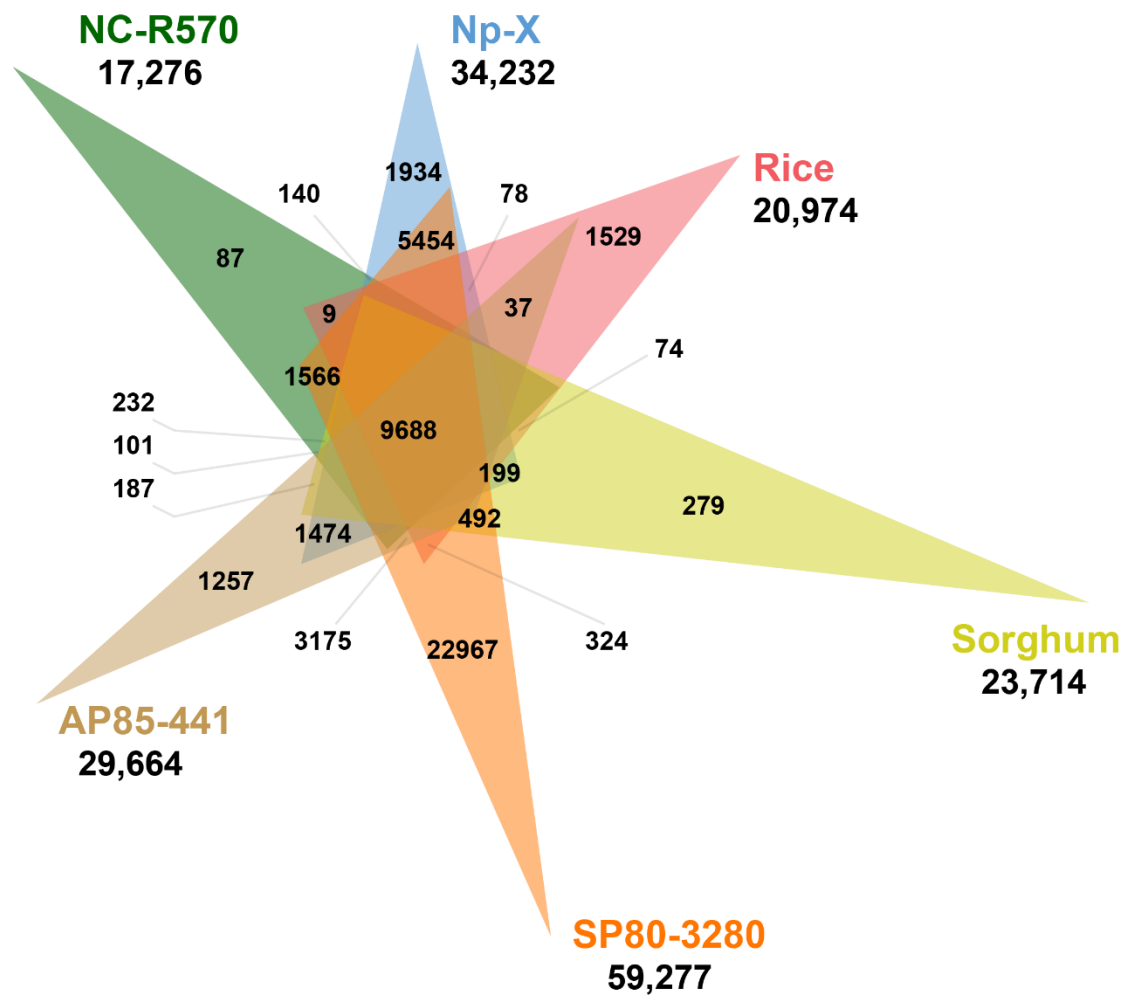
Supplementary Fig. S2. Genome-wide chromatin interactions at 500-kb resolution in the *S. spontaneum* Np-X genome.



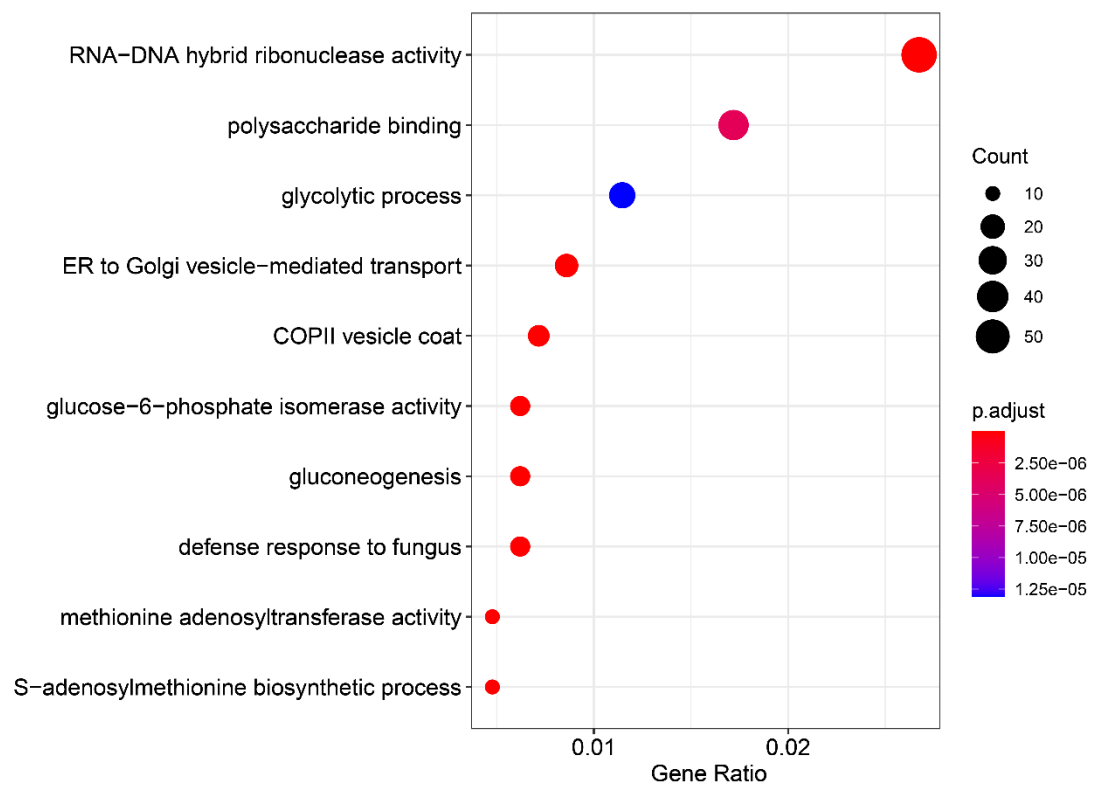
Supplementary Fig. S3. Genomic features distributed along the monoploid genome of *S. spontaneum* Np-X. The tracks indicate: (a) monoploid genome in Mbp; (b) SNP densities in four haplotypes; (c) Gene density.



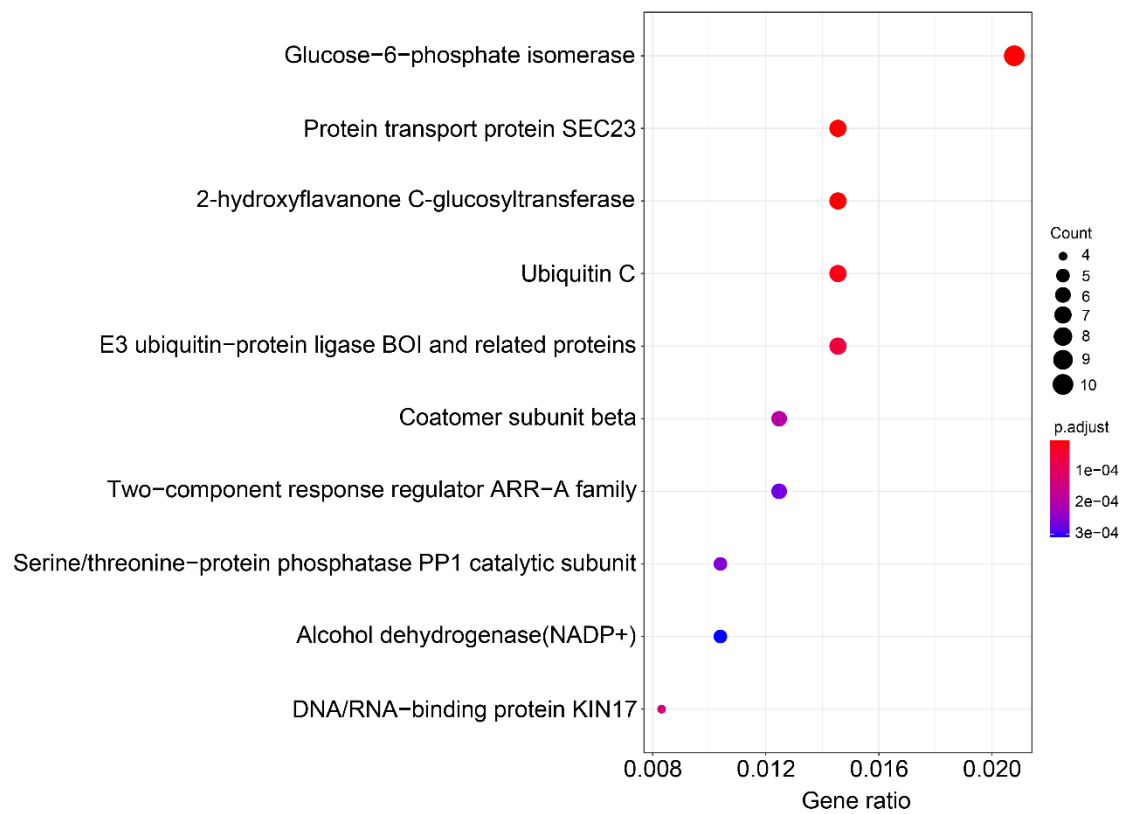
Supplementary Fig. S4. Alignment of the *S. spontaneum* Np-X and *Miscanthus sinensis* genomes shown as a dot plot.



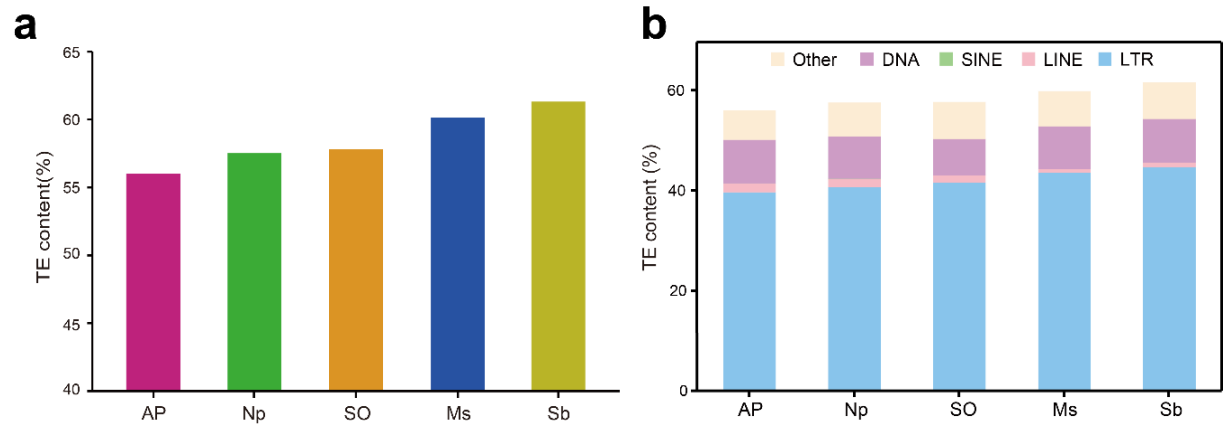
Supplementary Fig. S5. Orthologous gene families among *S. spontaneum* Np-X, *S. spontaneum* AP85-441, *Sorghum*, *Saccharum* hybrid SP80-3280, *Saccharum* hybrid R570, and rice. The numbers of gene families are shown for each species and for those shared among two or more species.



Supplementary Fig. S6. GO term enrichment of *S. spontaneum*-specific gene families.

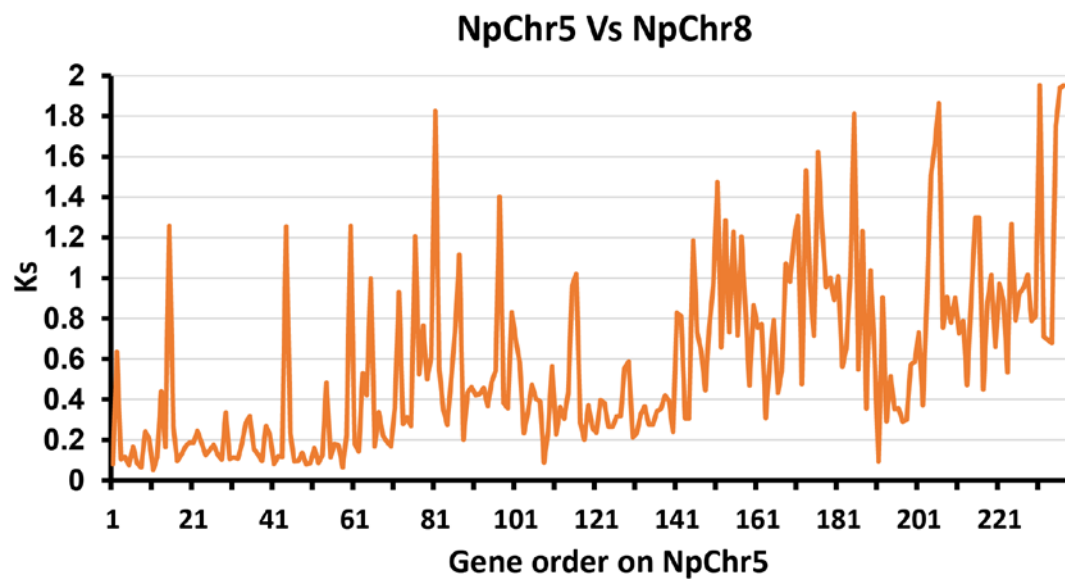


Supplementary Fig. S7. KEGG term enrichment for *S. spontaneum*-specific gene families.

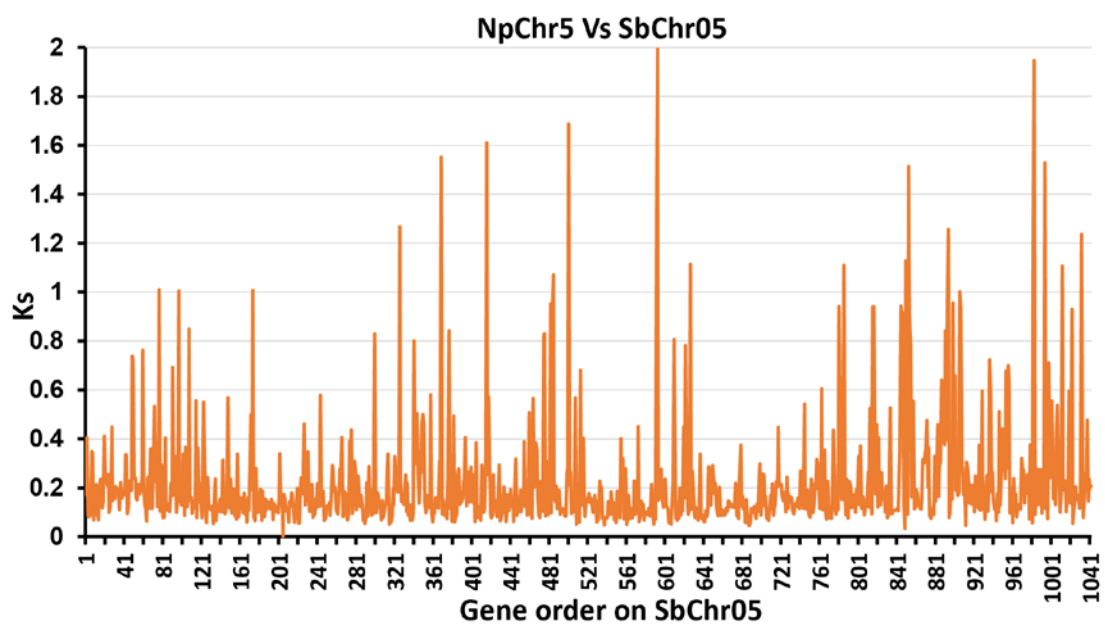


Supplementary Fig. S8. Transposable elements contents of the genomes of three species in the Andropogoneae and two *S. spontaneum* accessions. (a) TE content in each genome; (b) TE composition of each genome.

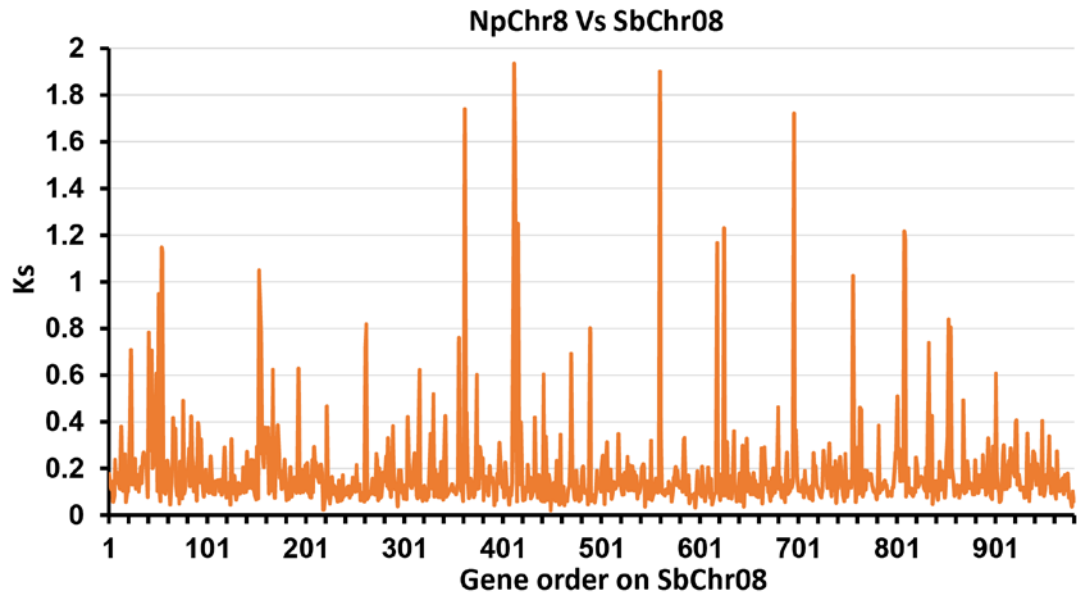
a



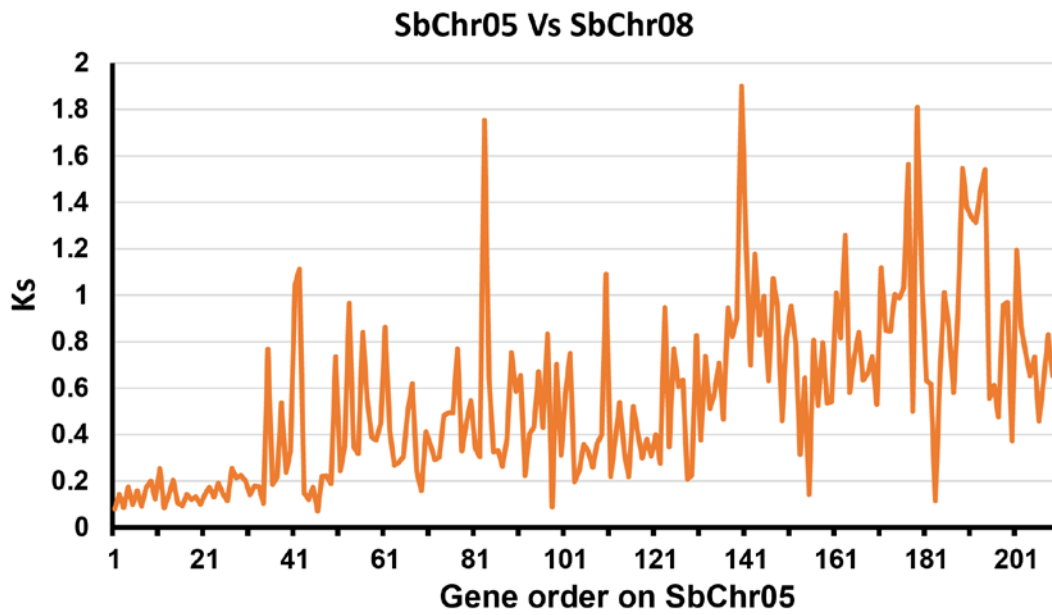
b



c

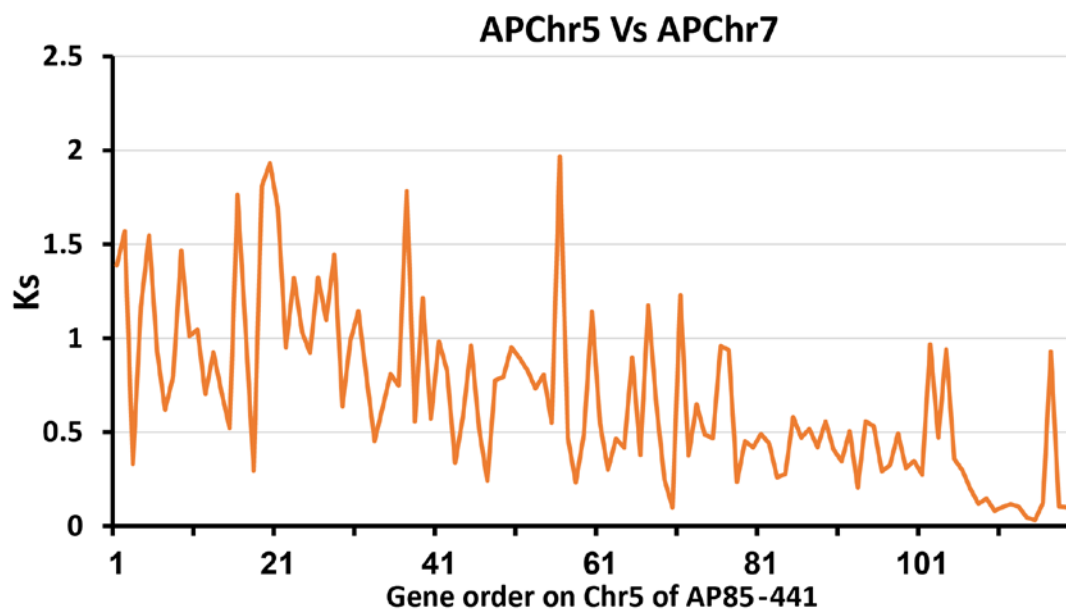


d

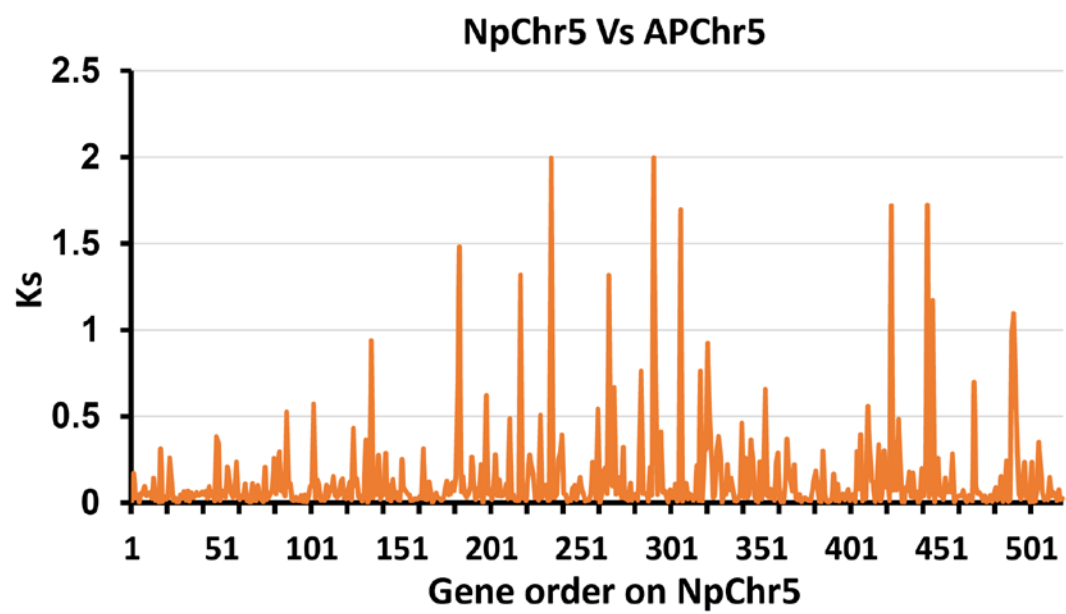


Supplementary Fig. S9. Ks distribution between syntenic gene pairs along chromosomes of *S. spontaneum* Np-X and *Sorghum bicolor*. Pairs with Ks >2 are not shown. Np represents *S. spontaneum* Np-X and Sb represents *Sorghum bicolor*. (a) NpChr5-NpChr8 paralogs; (b) NpChr5-SbChr05 orthologs; (c) NpChr8-SbChr08 orthologs; (d) SbChr05-SbChr08 paralogs.

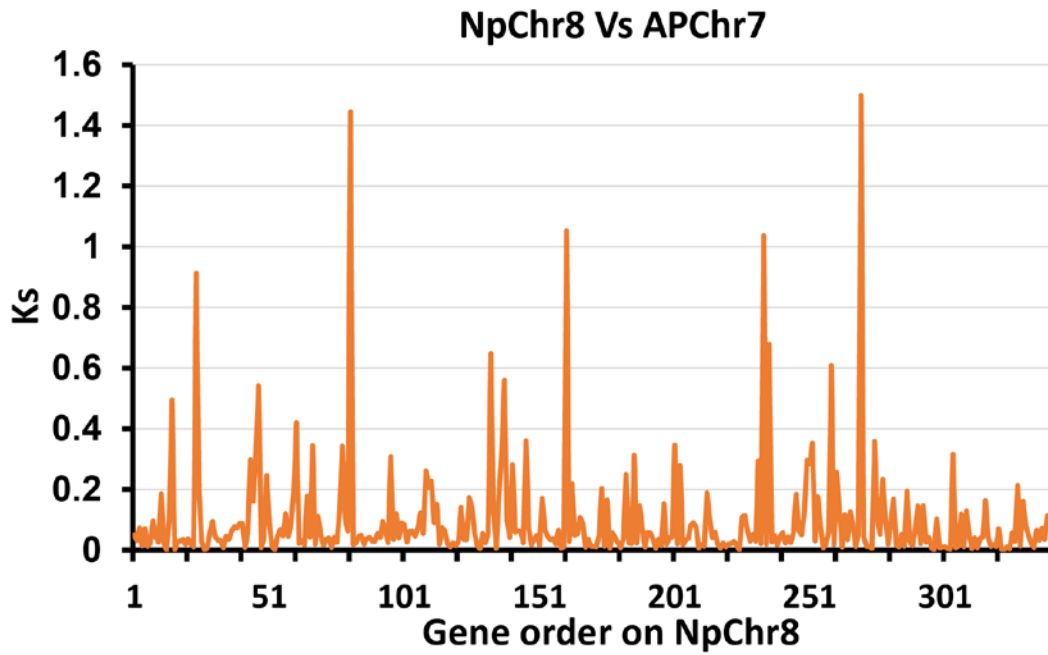
a



b



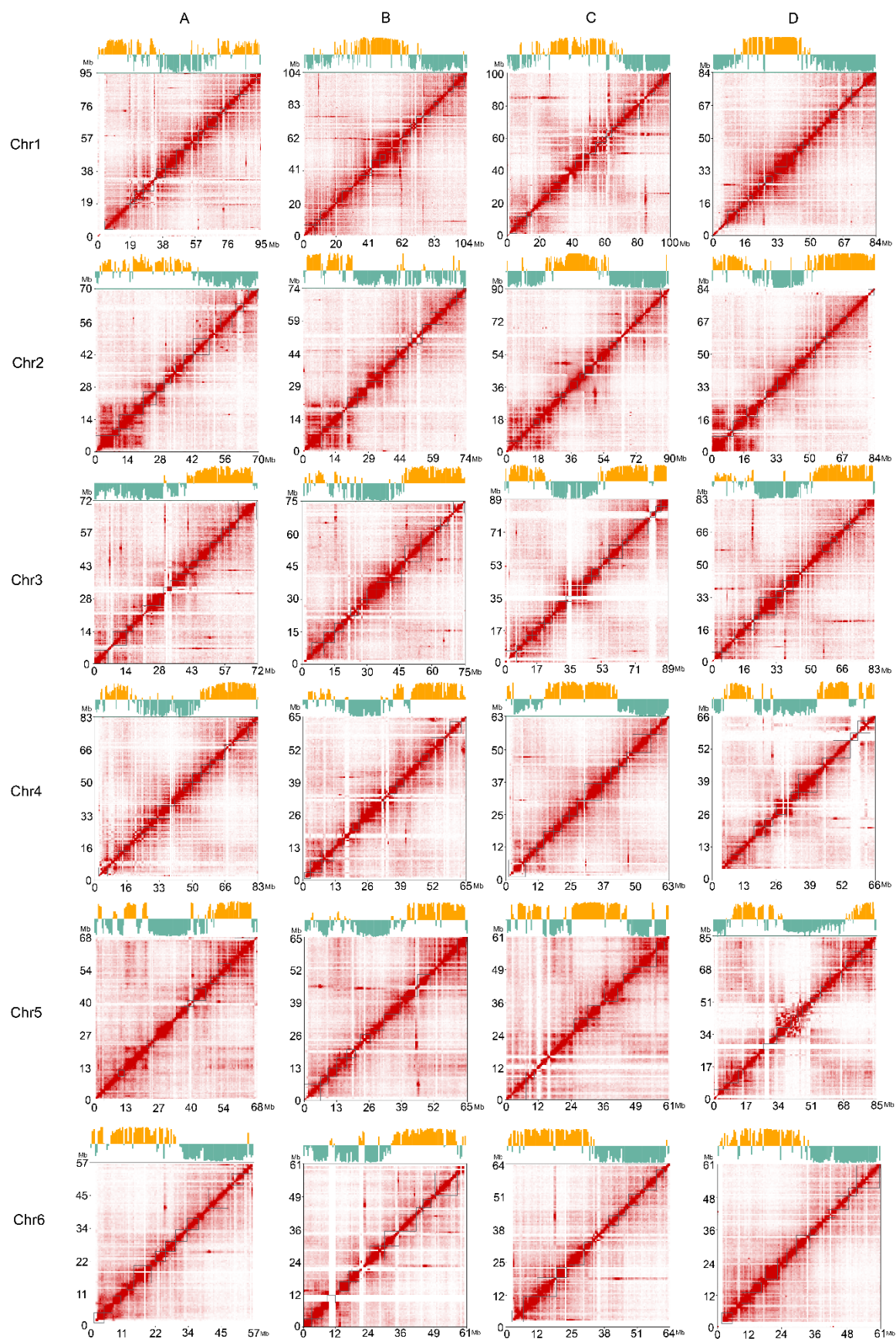
C

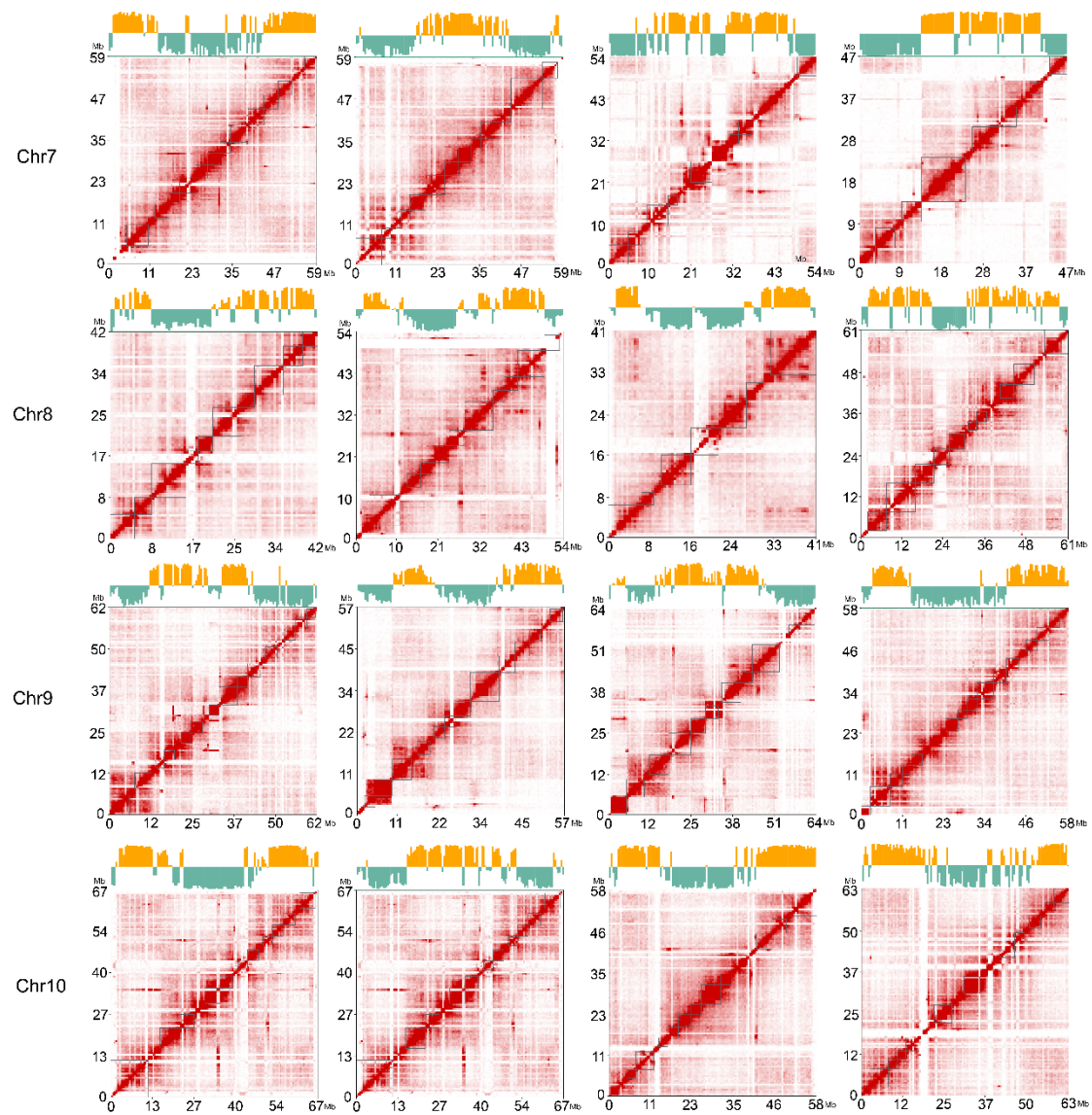


Supplementary Fig. S10. Ks distribution between syntenic gene pairs along chromosomes of *S. spontaneum* Np-X and *S. spontaneum* AP85-441. Pairs with Ks >2 are not shown. APChr7 represents the short Chromosome arm of *S. spontaneum* AP85-441 Chr7 and Np represents *S. spontaneum* Np-X. (a) APChr5-APChr7 paralogs; (b) NpChr5-APChr05 orthologs; (c) NpChr8-APChr7 orthologs.

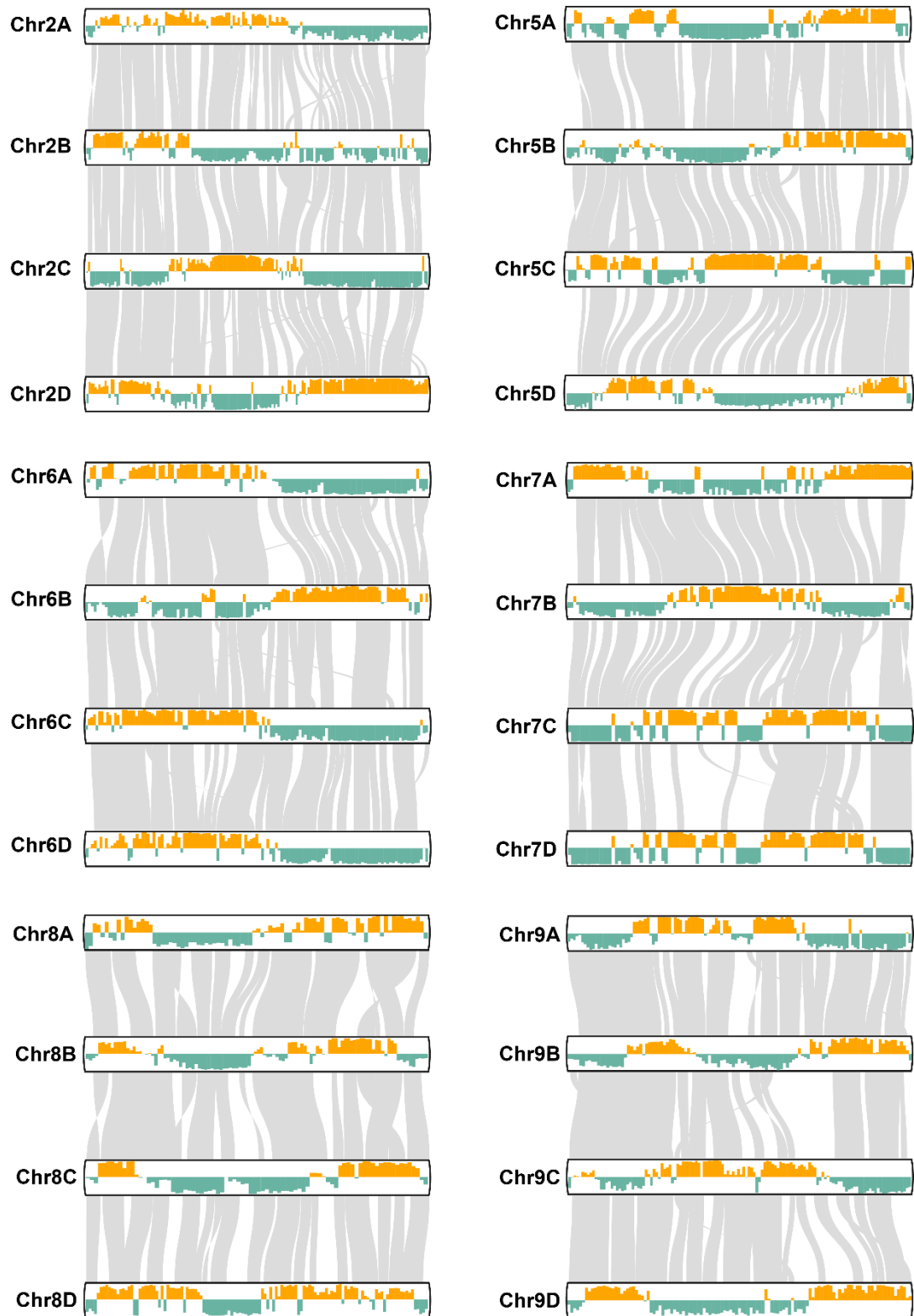


Supplementary Fig. S11. Total allelic expression of genes on the chromosomes of *S. officinarum* in mature leaves and stems. A–H represent the four homologous chromosomes in each group.

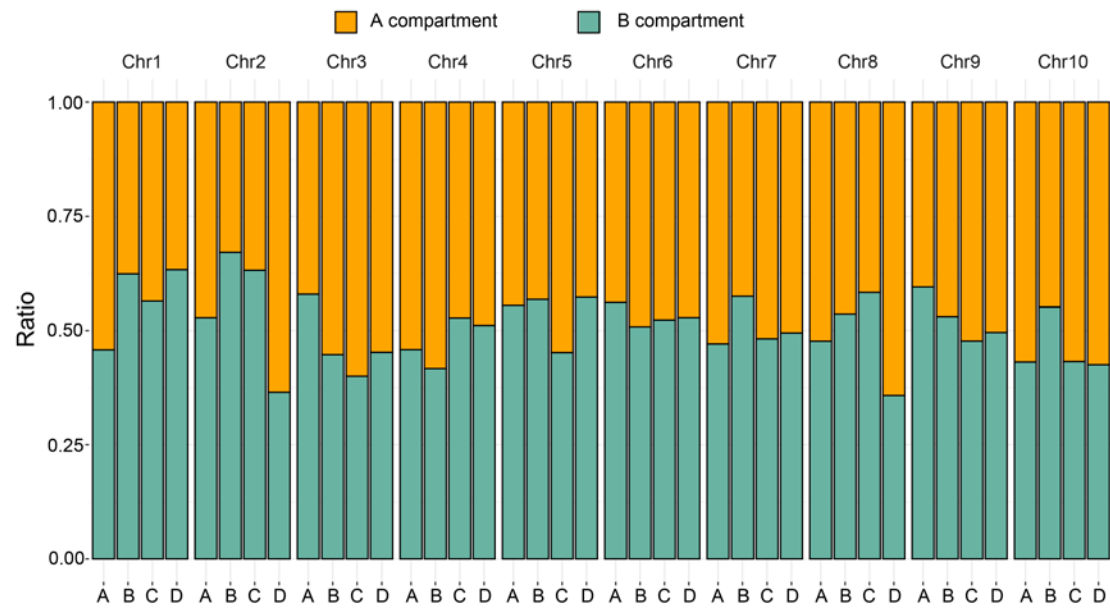




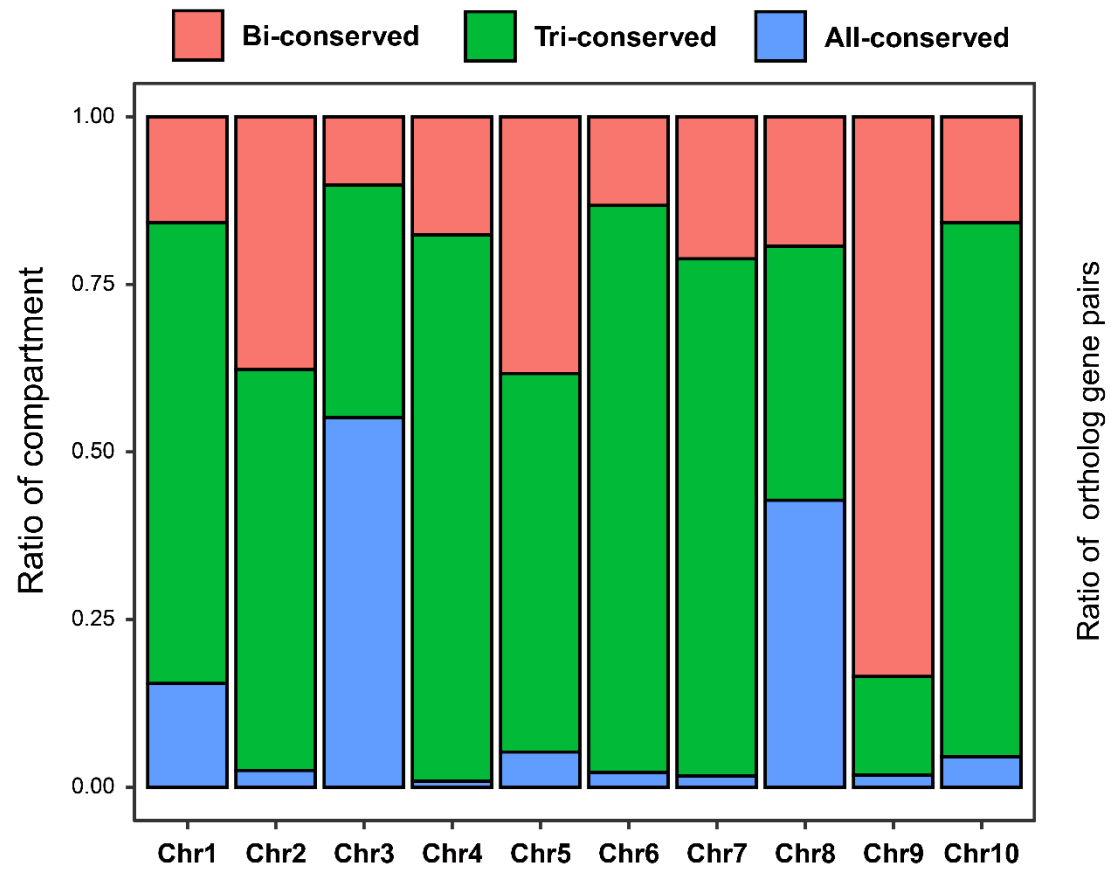
Supplementary Fig. S12. Global patterns of chromatin interaction in the *S. spontaneum* Np-X genome. The upper track shows the partition of each set of four alleles into A (yellow histogram) and B (blue histogram) compartments. The lower track shows chromatin interactions for each allele at 500-kb resolution. In each heat map, strong interactions are indicated in red and weak interactions are indicated in white.



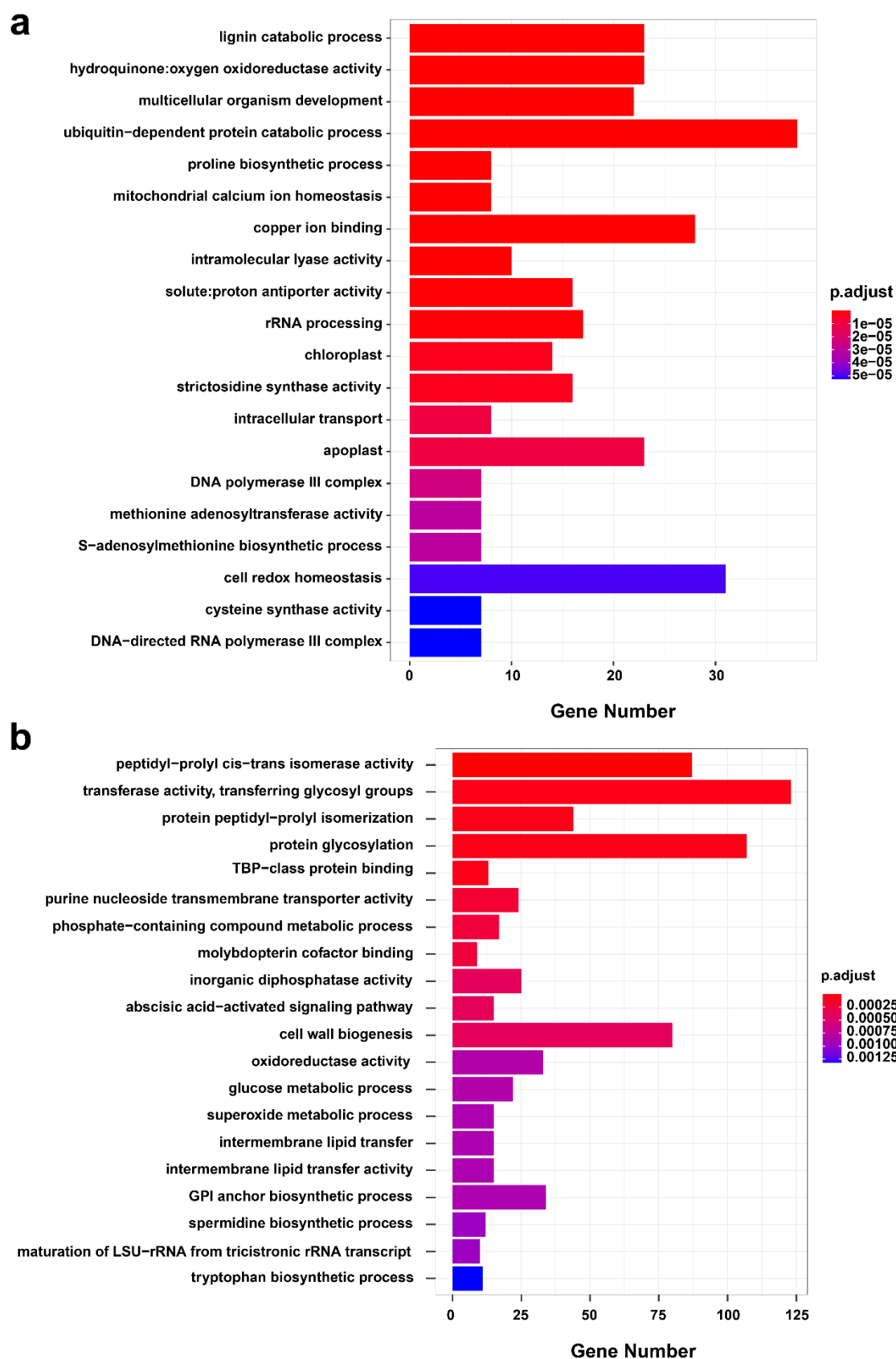
Supplementary Fig. S13. Conservation of higher order structure of homologous chromosome sets in the *S. spontaneum* Np-X genome. Note: The four tracks represent first principal component values showing A or B compartment status for four alleles on chromosome 2,5,6,7,8 and 9. In the analysis for each set of alleles, the grey line indicates the syntenic regions between alleles of the chromosome.



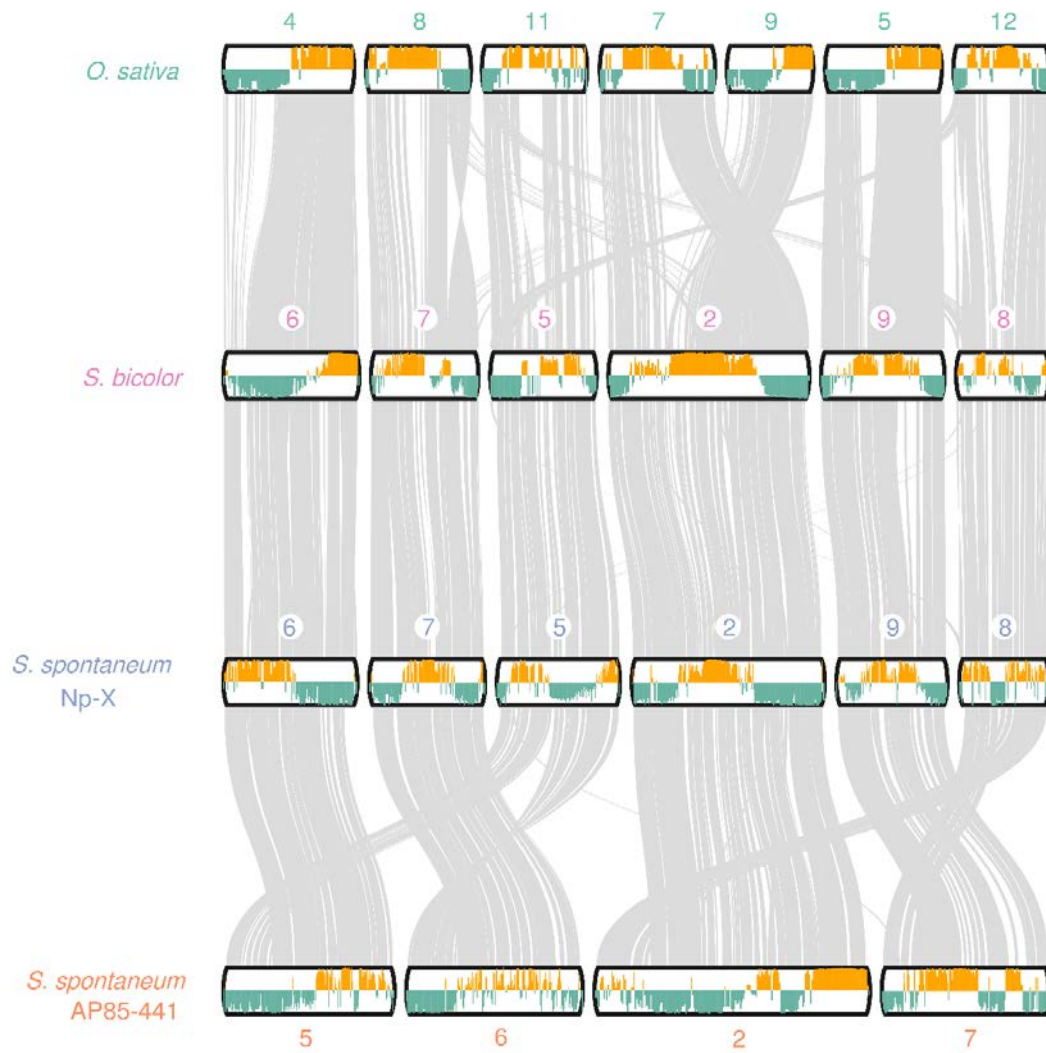
Supplementary Fig. S14. A/B compartment distribution on each chromosome of the *S. spontaneum* Np-X genome. A–D represent the four homologous groups.



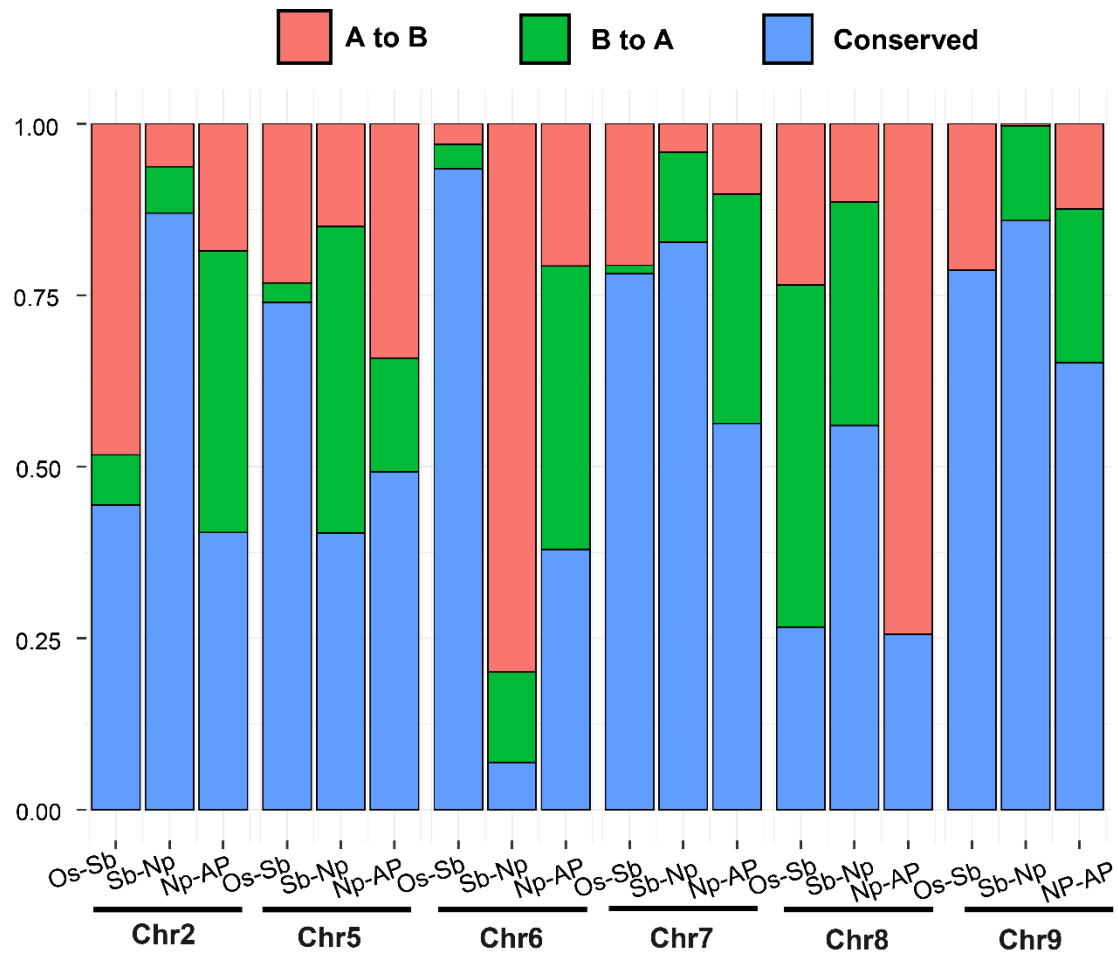
Supplementary Fig. S15. Analysis of variation in the 3D structure of allelic chromosomes in the *S. spontaneum* Np-X genome. Note: The red, green, and blue bars show the proportion of genes with two, three, or four conserved alleles, respectively.



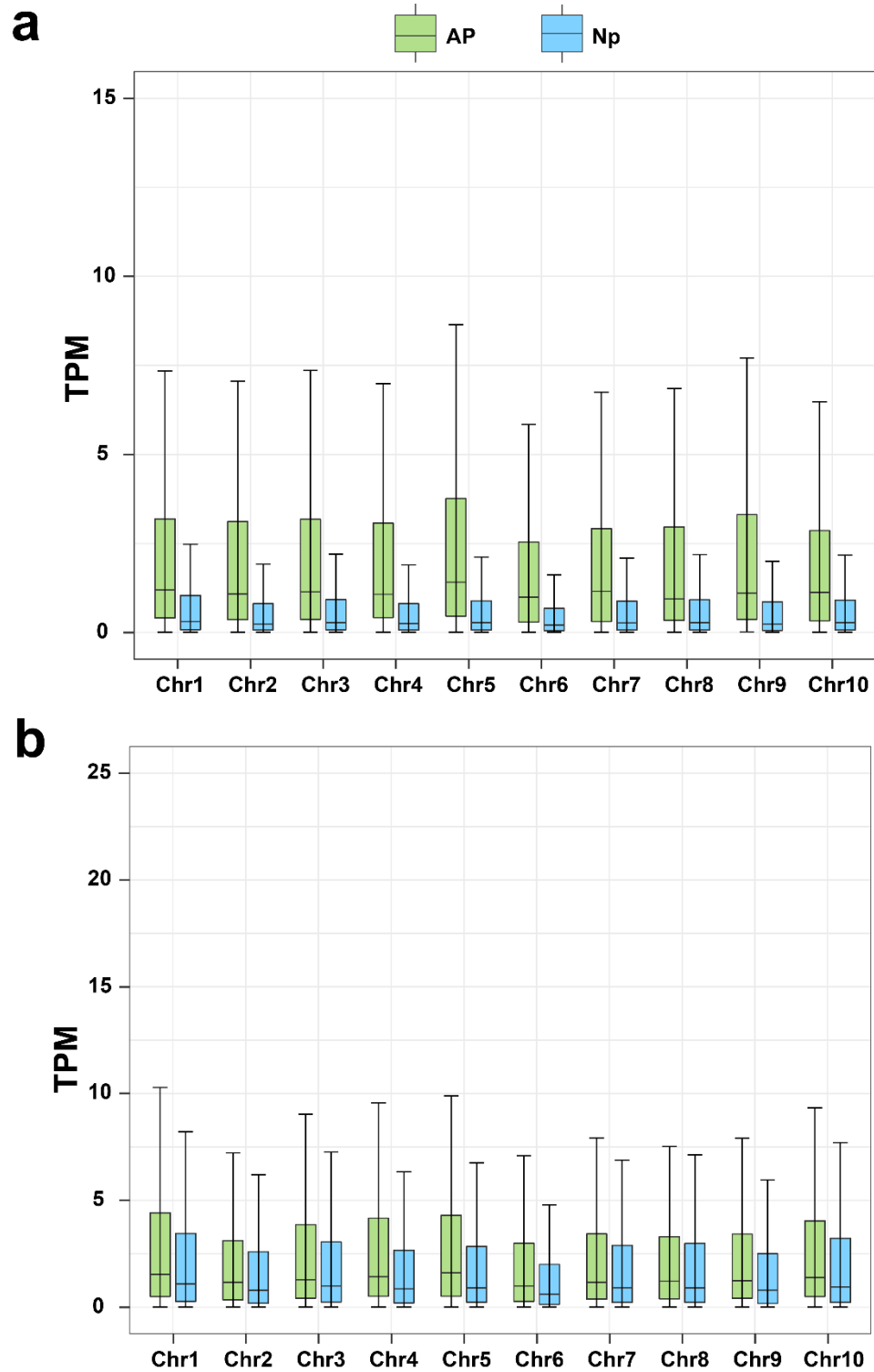
Supplementary Fig. S16. GO term enrichment of the All-conserved and non-All-conserved genes. The GO term enrichment of All-conserved genes(a) and non-All-conserved genes(b) including two and three conserved.



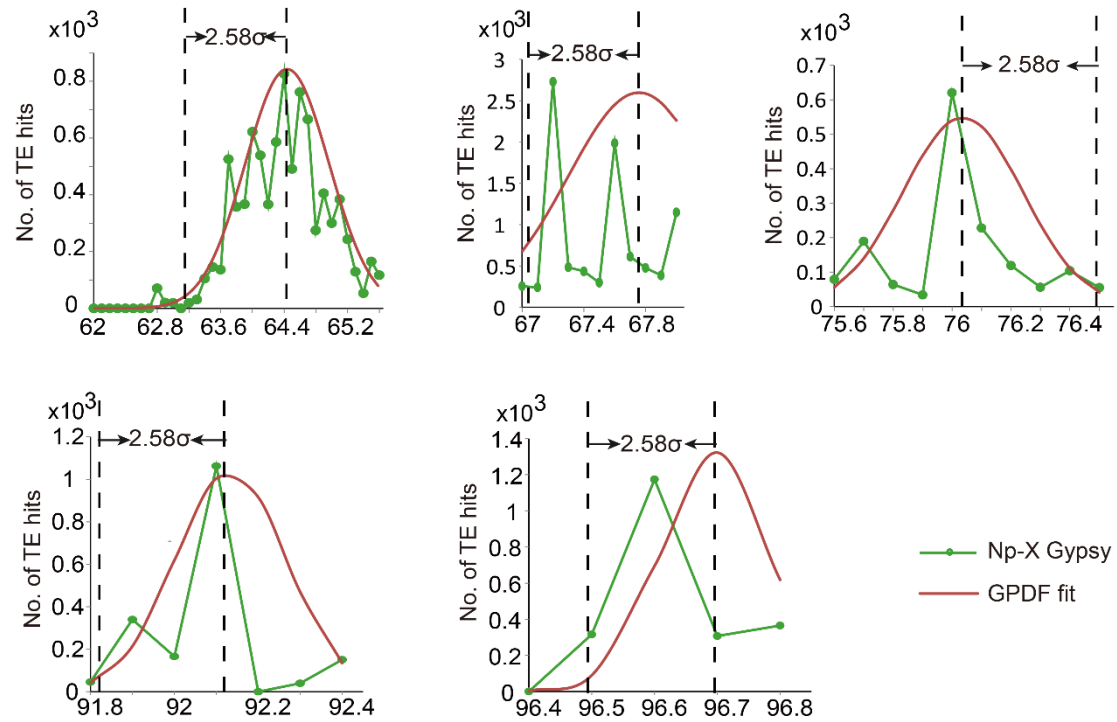
Supplementary Fig. S17. Conservation of higher order structure of reconstructed chromosomes between *S. spontaneum* Np-X and its relatives. Note: The four tracks represent first principal component values showing A/B compartment status for *O. sativa*, *S. bicolor*, *S. spontaneum* Np-X, and *S. spontaneum* AP85-441, respectively. In the analysis of each set of species, the grey line indicates the syntenic regions of the chromosome between species.



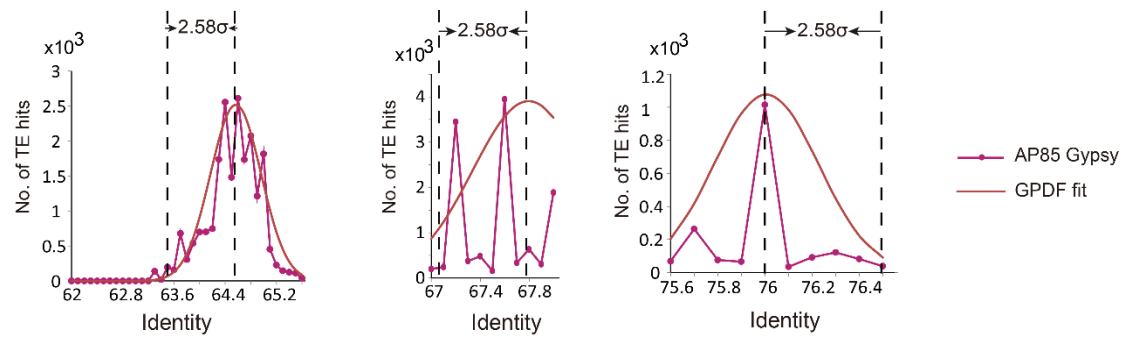
Supplementary Fig. S18. Recombination of higher chromatin structure in the *S. spontaneum* Np-X and *S. spontaneum* AP85-441 genomes and related species. Note: The red, green, and blue bars indicated A-to-B compartment switching, B-to-A switching, and conserved compartments during species evolution, respectively.



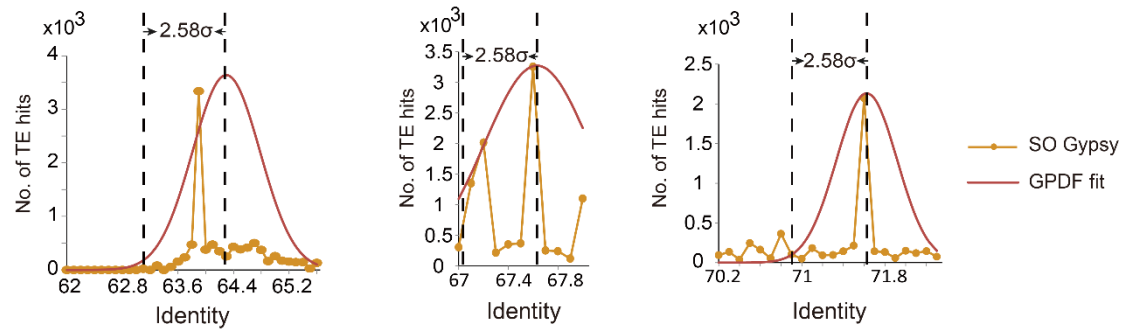
Supplementary Fig. S19. The expression level of genes within B to A compartment switching in each chromosome of *S. spontaneum* Np-X and corresponding in *S. spontaneum* AP85-441. Note: The genes expression in leaf are indicating in **a and that in stem are in **b**.**



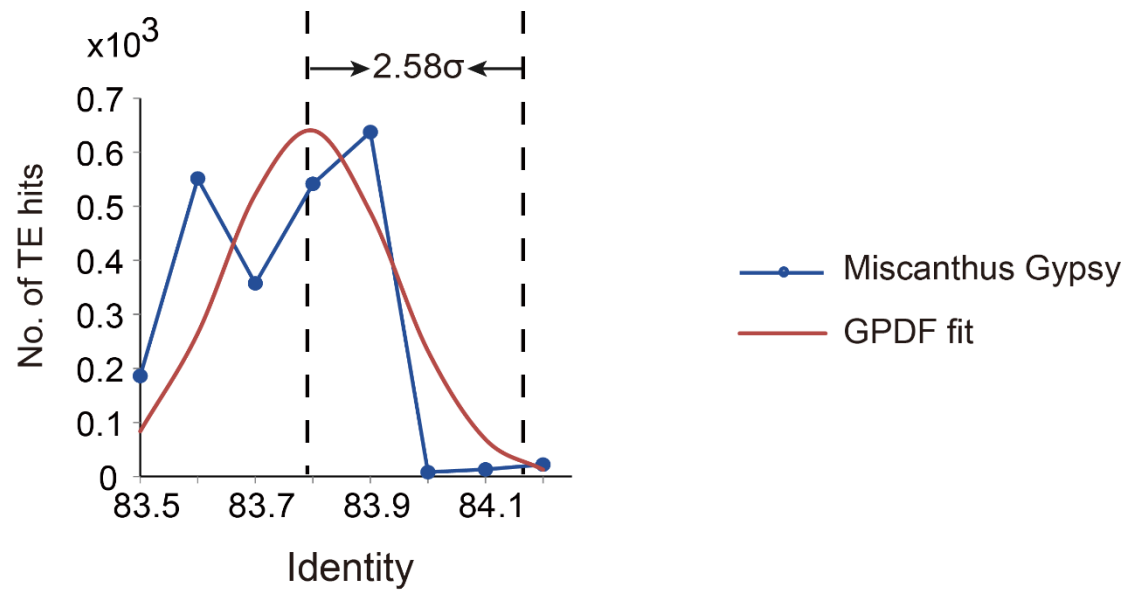
Supplementary Fig. S20. Burst of LTR/Gypsy amplification in *S. spontaneum* Np-X. This LTR/Gypsy amplification burst fit a Gaussian distribution, which allows time estimation of burst events. The peak width was defined as 2.58σ covering >99.5% of the nucleotide substitution events in TEs.



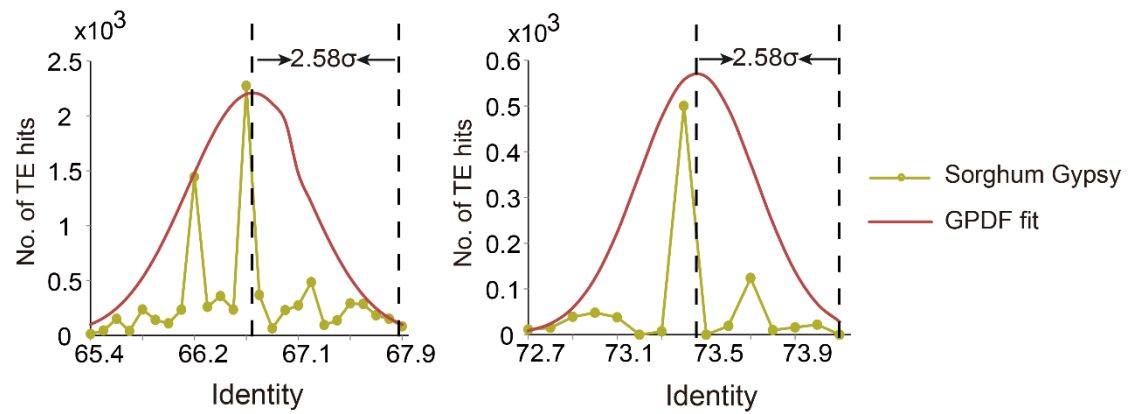
Supplementary Fig. S21. Burst of LTR/Gypsy amplification in *S. spontaneum* AP85-441. This LTR/Gypsy amplification burst fit a Gaussian distribution, which allows time estimation of burst events. The peak width was defined as 2.58σ covering >99.5% of the nucleotide substitution events in TEs.



Supplementary Fig. S22. Burst of LTR/Gypsy amplification in *S. officinarum*. This LTR/Gypsy amplification burst fit a Gaussian distribution, which allows time estimation of burst events. The peak width was defined as 2.58σ covering >99.5% of the nucleotide substitution events in TEs.

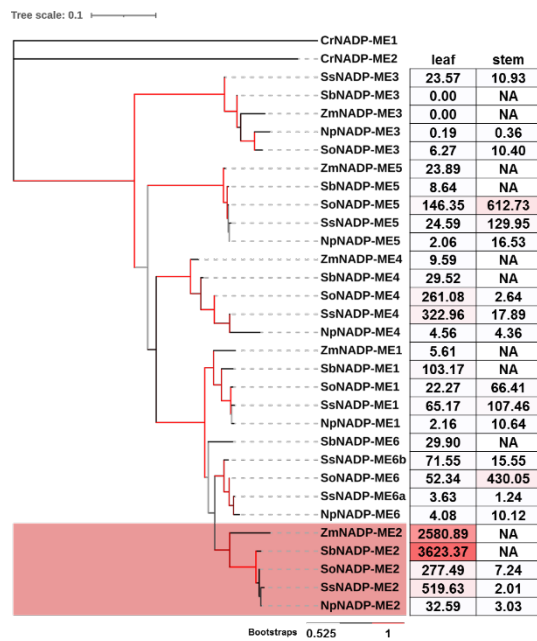


Supplementary Fig. S23. Burst of LTR/Gypsy amplification in *Miscanthus*. This LTR/Gypsy amplification burst fit a Gaussian distribution, which allows time estimation of burst events. The peak width was defined as 2.58σ covering >99.5% of the nucleotide substitution events in TEs.

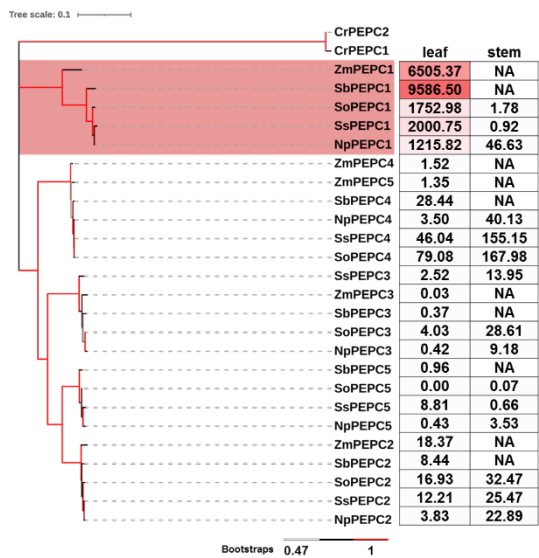


Supplementary Fig. S24. Burst of LTR/Gypsy amplification of *Sorghum bicolor* This LTR/Gypsy amplification burst fits a Gaussian distribution, which allows time estimation of burst events. The peak width was defined as 2.58σ covering >99.5% of the nucleotide substitution events in TEs.

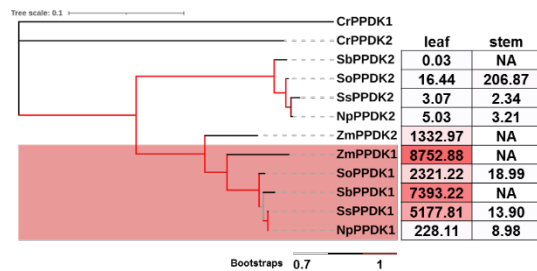
NADP-ME



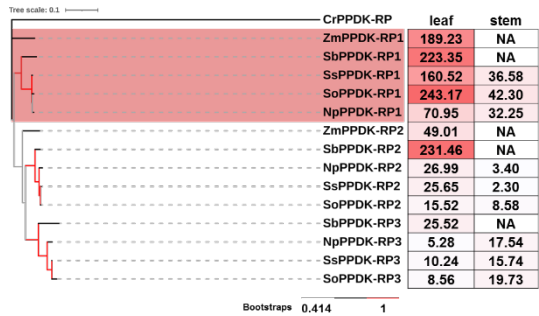
PEPC



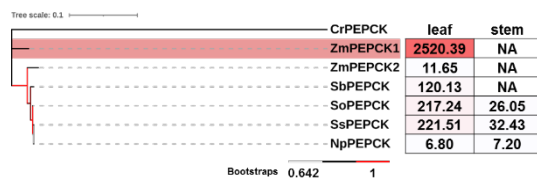
PPDK



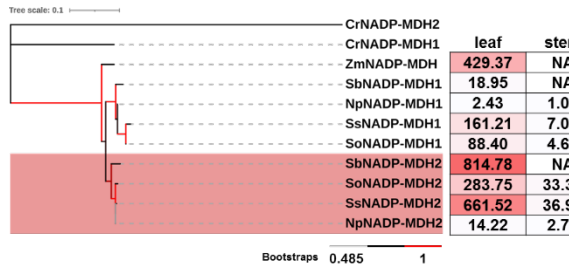
PPDK-RP



PEPCK

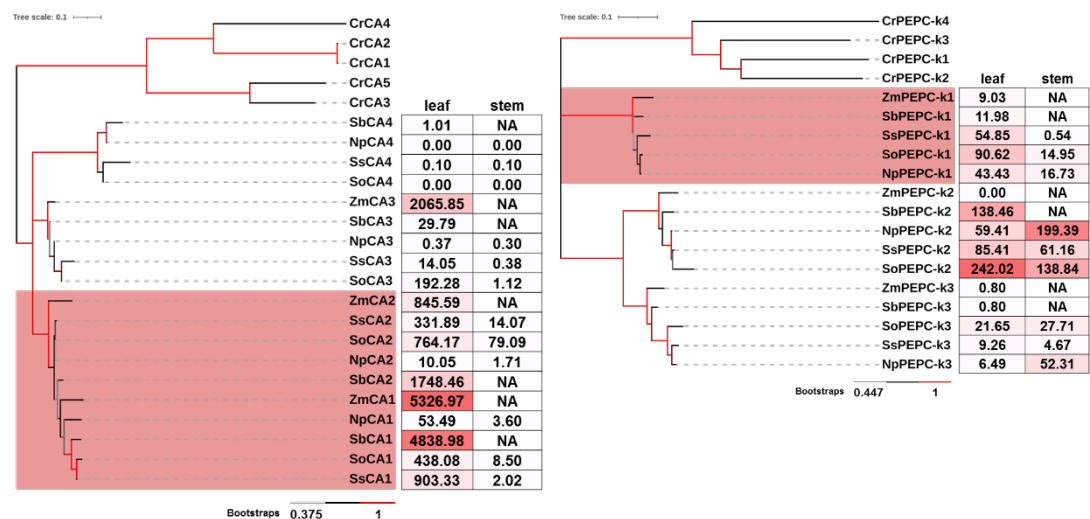


NADP-MDH

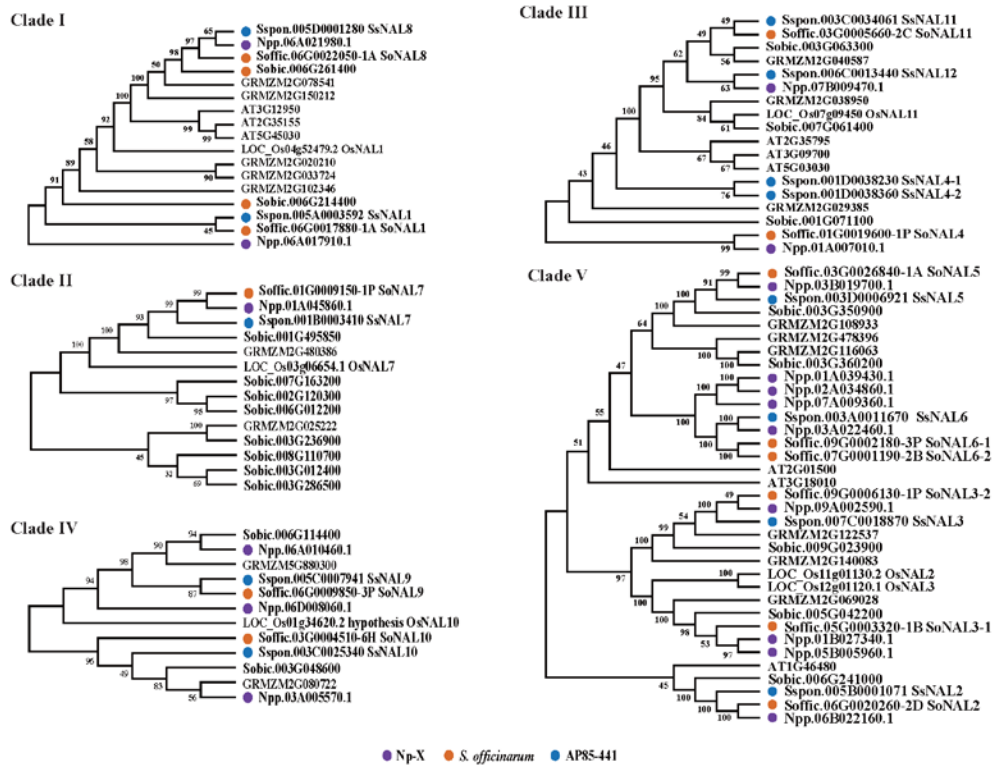


CA

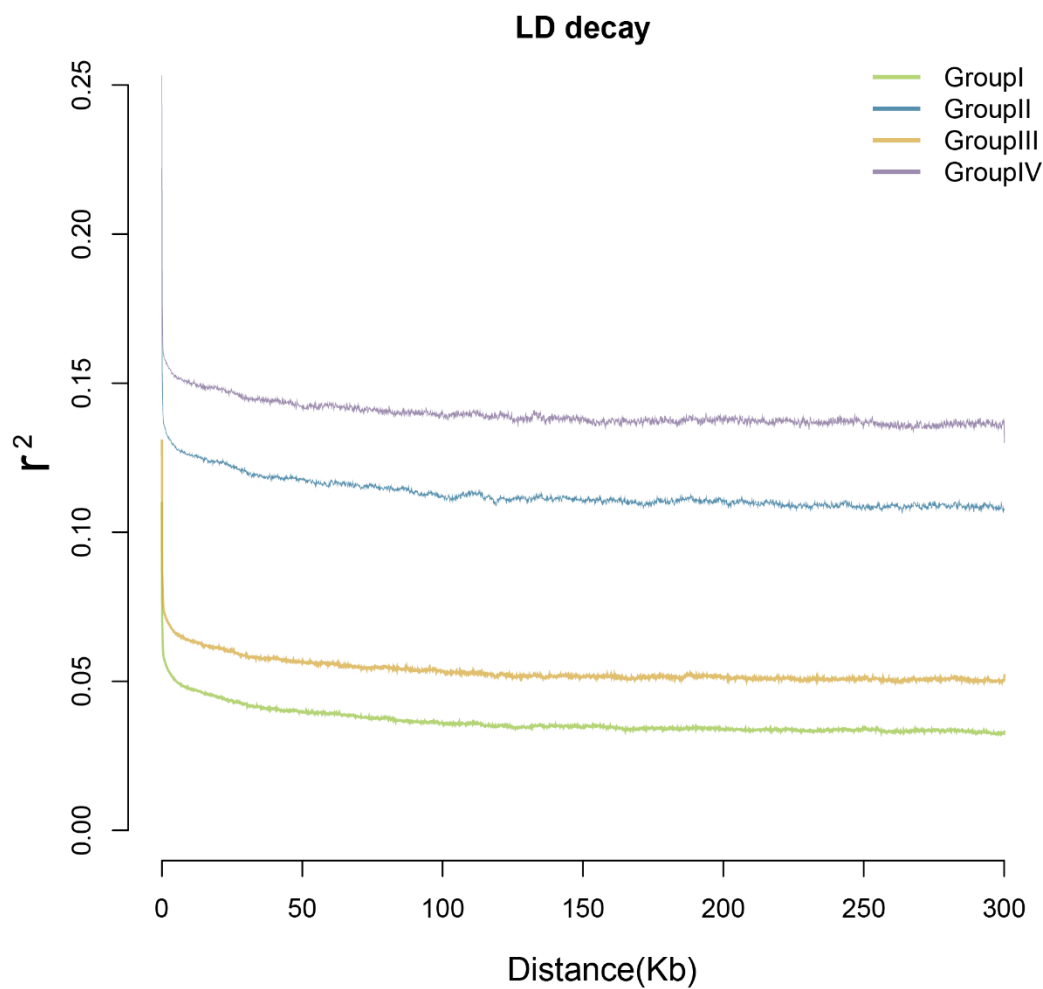
PEPC-k



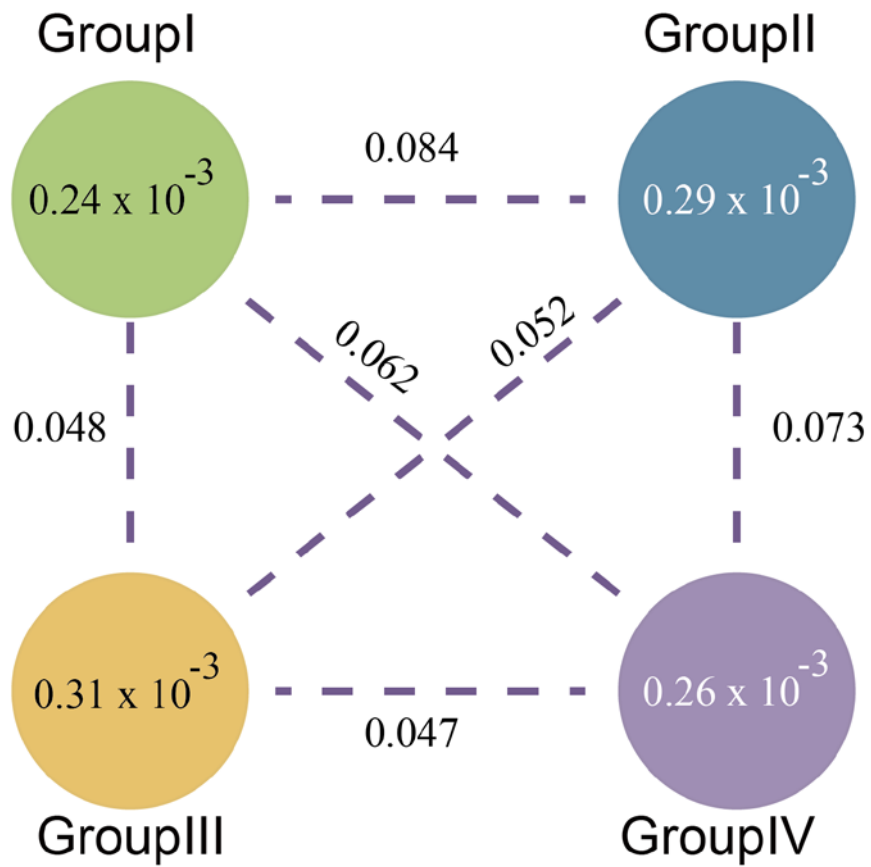
Supplementary Fig. S25. Phylogenetic analysis of genes encoding C4 enzymes and their non-C4 isoforms. XSs indicates *Saccharum spontaneum* Np-X, Ss indicates *Saccharum spontaneum* AP85-441, Sb indicates *Sorghum bicolor*, Zm indicates *Zea mays*, Os indicates *Oryza sativa*, and Cr indicates *Chlamydomonas reinhardtii*. Grey circles indicate bootstraps. Genes in the red clade encodes proteins in the C4 pathway. Tree scale: 0.1.



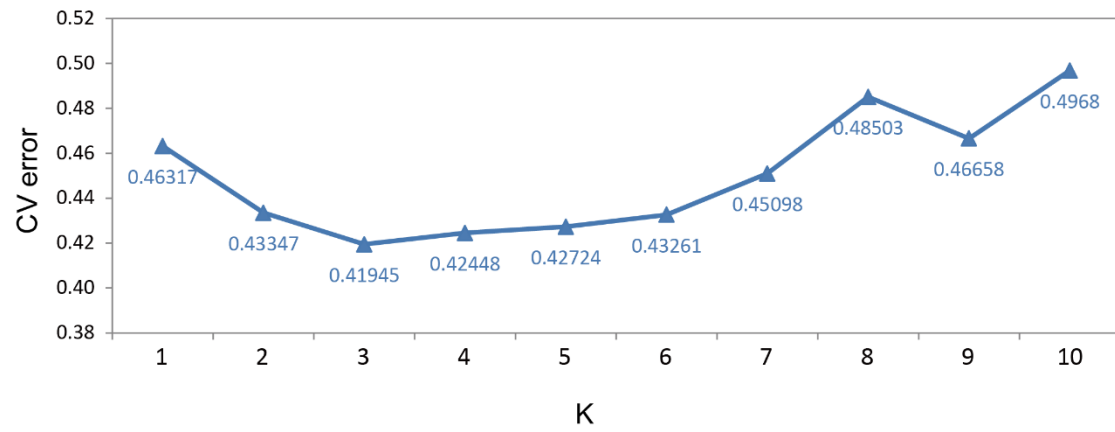
Supplementary Fig. S26. Phylogenetic analysis of Narrow Leaf (NAL)-related genes in *S. spontaneum* Np-X, *S. spontaneum* AP85-441, *S. officinarum*, *Z. mays*, rice, *Sorghum*, and *Arabidopsis*.



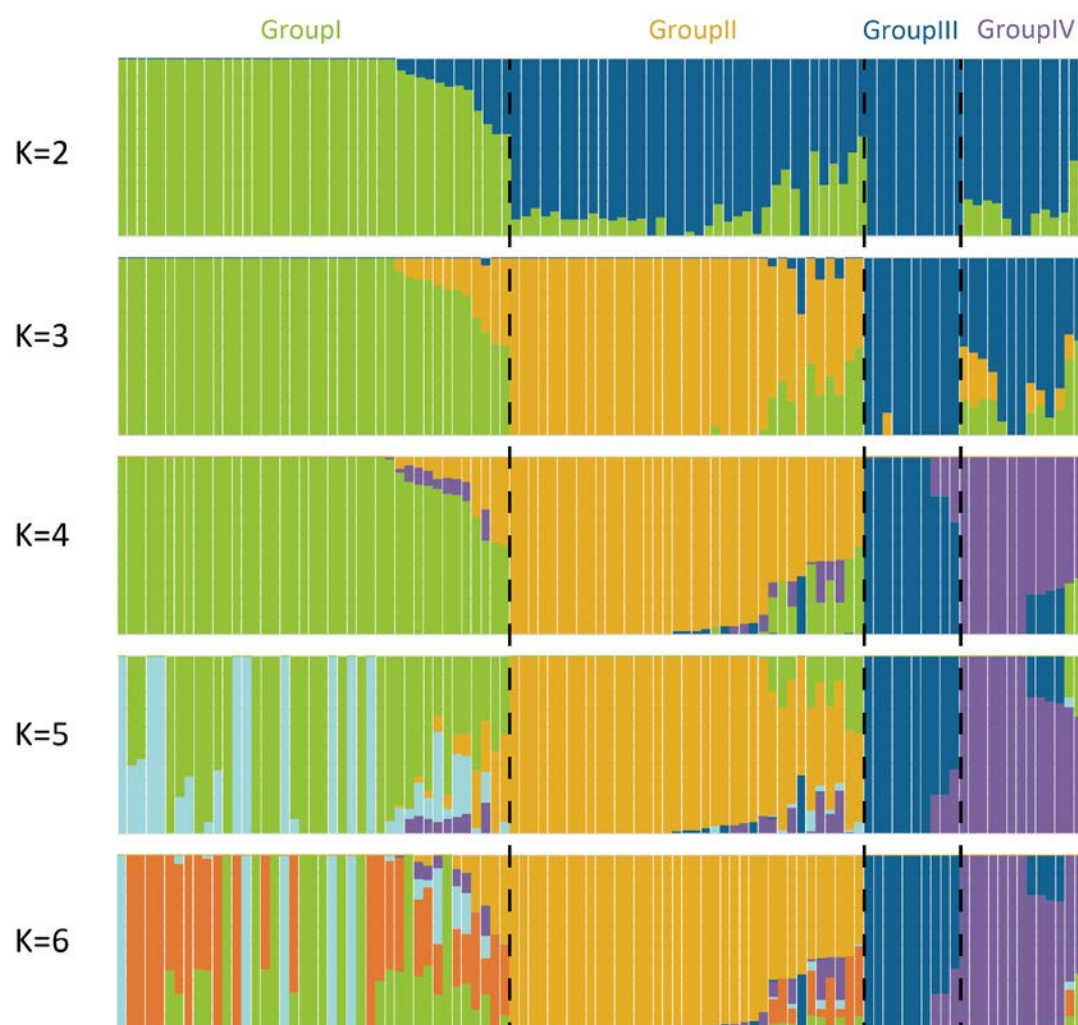
Supplementary Fig. S27. Linkage disequilibrium decay in different subpopulations of *S. spontaneum*.



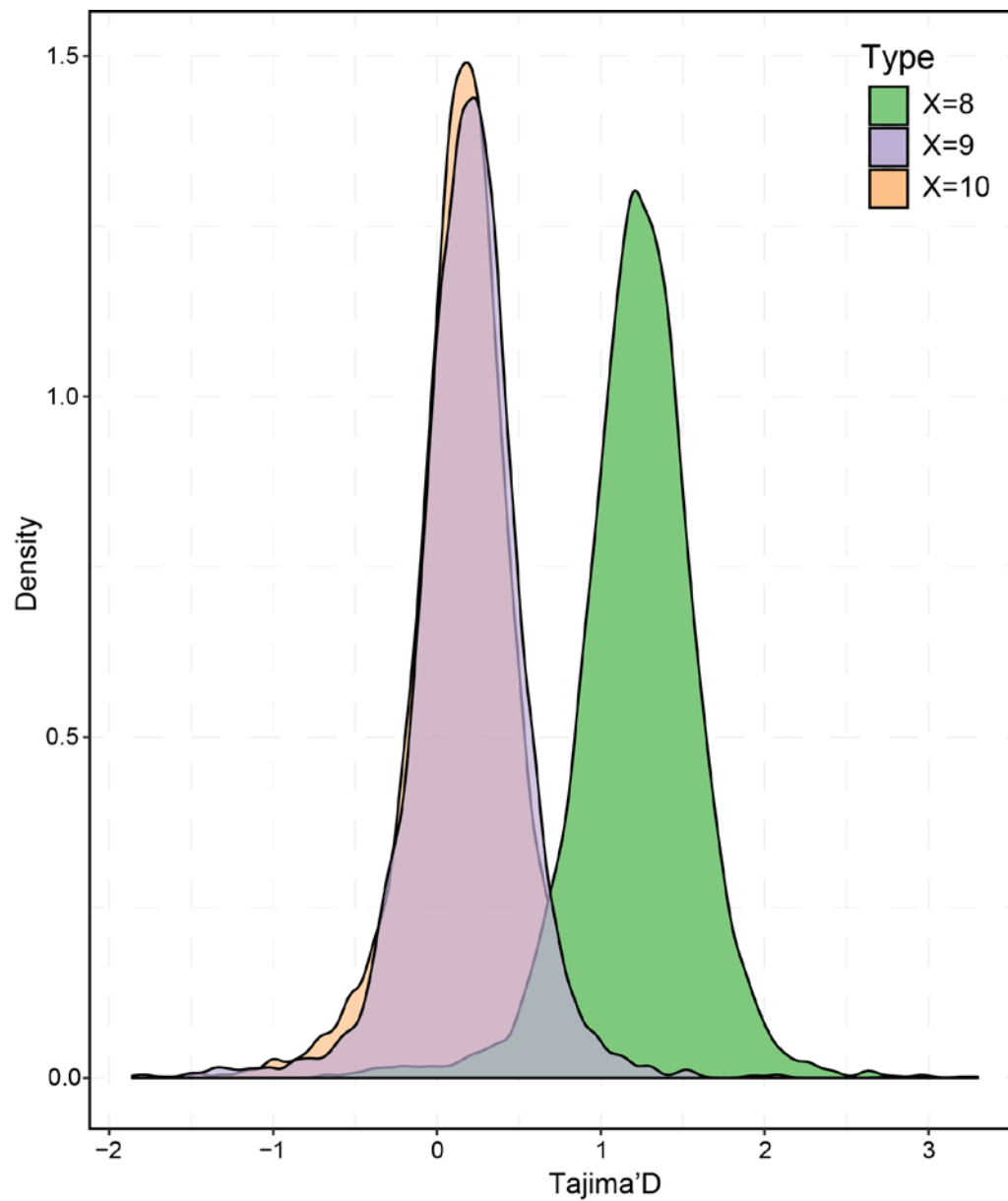
Supplementary Fig. S28. Nucleotide diversity (π) and population divergence (F_{ST}) between different subpopulation of *S. spontaneum*. The value in each circle represents a measure of π for this group, and the value on each line indicates F_{ST} between two groups.



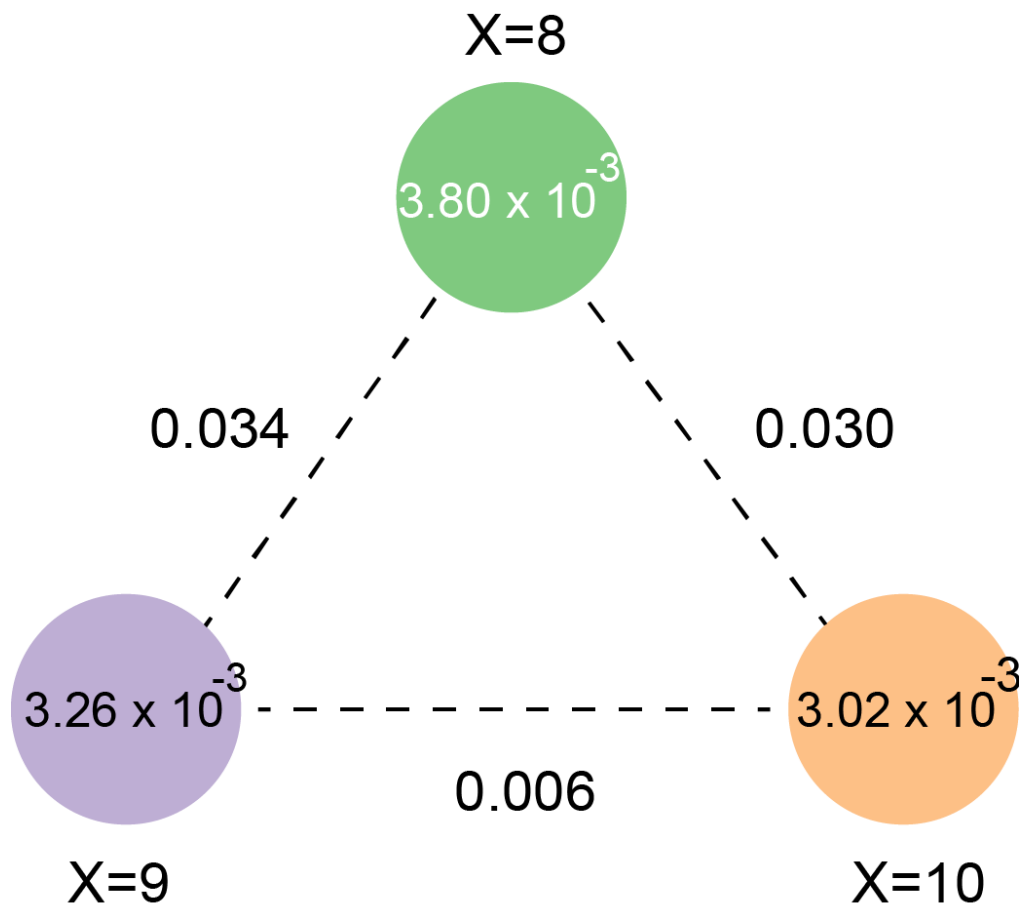
Supplementary Fig. S29. Plot of ADMIXTURE cross-validation errors for K = 1 through K = 10.



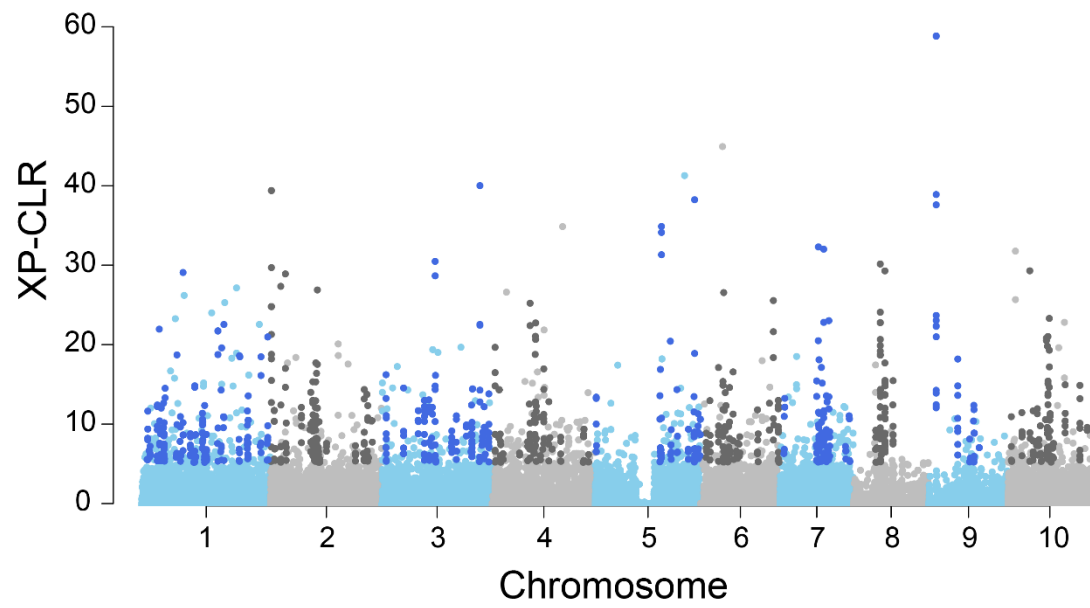
Supplementary Fig. S30. STRUCTURE analysis of the resequenced populations with K values from 2 to 6.



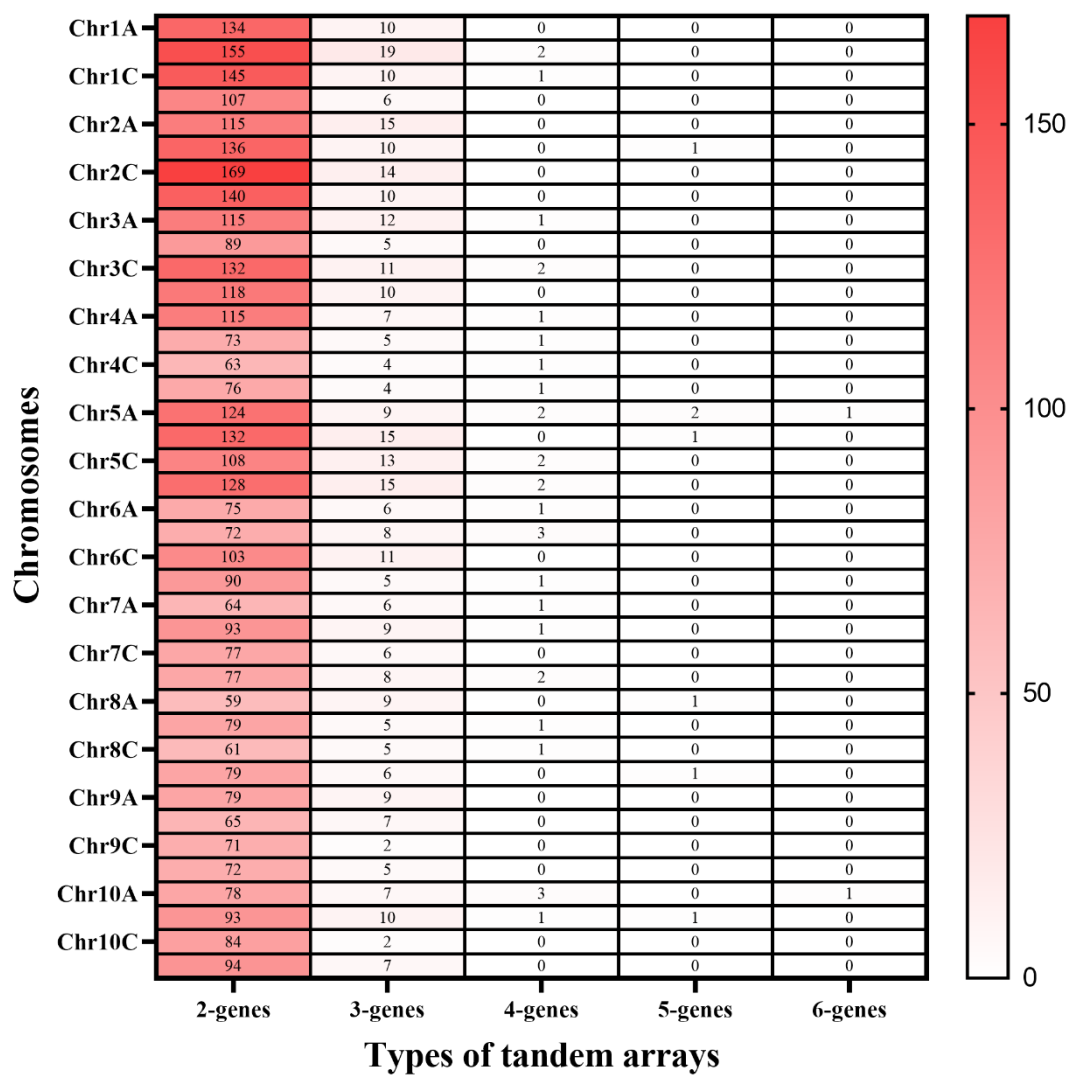
Supplementary Fig. S31. Tajima's D of *S. spontaneum* subpopulations with $x = 8$, $x = 9$, or $x = 10$.



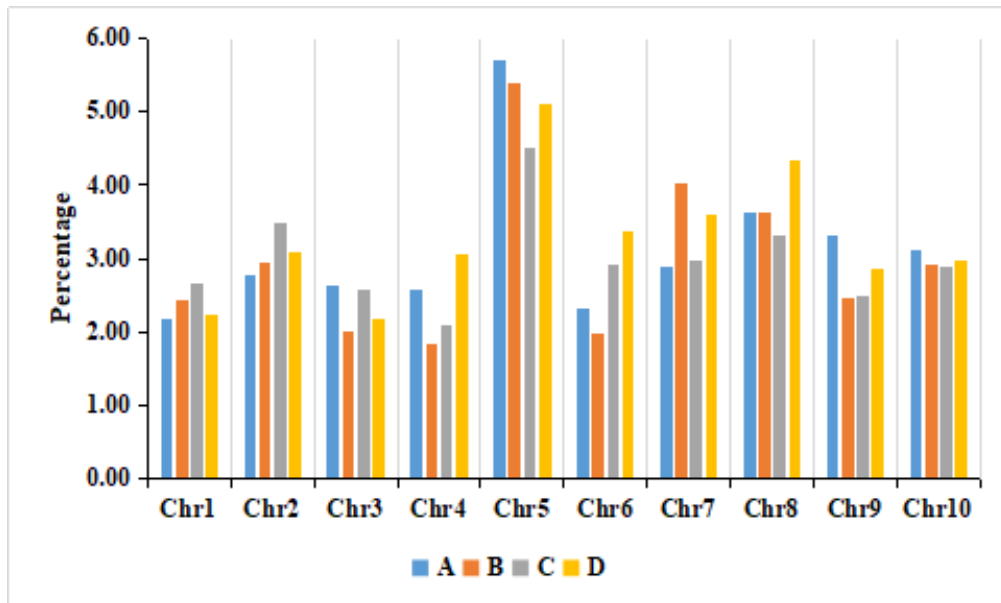
Supplementary Fig. S32. Nucleotide diversity (π) and population divergence (F_{ST}) across *S. spontaneum* population with $x = 8$, $x = 9$, or $x = 10$. The value in each circle represents a measure of nucleotide diversity for this group, and the value on each line indicates population divergence between two groups.



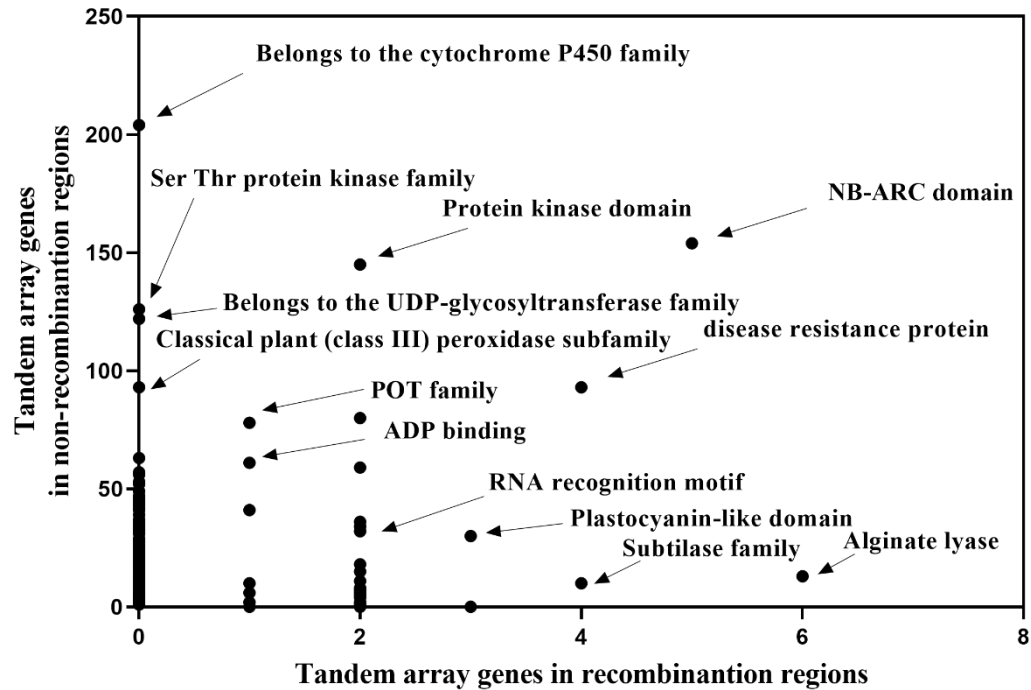
Supplementary Fig. S33. Genome-wide screening for selective sweeps due to natural selection in *S. spontaneum* populations with $x = 8$, and $x = 9$, or $x = 10$.



Supplementary Fig. S34. The distribution of tandem array length types on each chromosome of the *S. spontaneum* Np-X genome. The x-axis shows types of single tandem arrays containing 2 to 6 genes. The y-axis indicates chromosomes of *S. spontaneum* Np-X. The colors in the legend indicate the proportions of each tandem array length type distributed on the chromosome.



Supplementary Fig. S35. The distribution of tandemly duplicated genes in the genome of *S. spontaneum* Np-X. Different colors represent homologous chromosomes A, B, C, and D, respectively. The x-axis refers to chromosomes of *S. spontaneum* Np-X, while the y-axis refers to the percentage of tandemly duplicated genes on this chromosome out of the total number of genes on the chromosome.



Supplementary Fig. S36. Comparison of the functional annotations of tandemly arrayed genes in the non-recombinant and recombinant regions of chromosomes in the genome of *S. spontaneum* Np-X. The x-axis shows the counts of tandemly arrayed genes in the recombinant region with functional annotations and the y-axis shows the counts of tandemly arrayed genes in the non-recombinant regions with functional annotations.

Supplemental Tables

Supplementary Table S1. Statistics for CCS reads used for *S. spontaneum* Np-X genome assembly

Items	Pacbio CCS	Illumina X10	Hi-C
Total Num. of sequencing Cell(Library)	5	-	2
Total Num. of reads	5,540,670	2,782,017,034	1,937,852,694
Total Num. of sequenced Base(bp)	52,190,133,430	417,302,555,100	290,037,320,954
Mean reads length(bp)	9,644	150	150
Coverage(X)	18	148	105

Supplementary Table S2. Assessment of the consistency of *S. spontaneum* Np-X genome.

Items	Statistics
Number of reads	404,929,026
Data size (Gb)	60.74
Mapped bases (Gb)	59.97
Map rate (%)	98.73
Genome Length (Mbp)	2,759,716,890
Mean Depth(x)	21.7
Coverage Rate (%)	99.05

Supplementary Table S3. Statistics for Hi-C sequencing and mapping in *S. spontaneum* Np-X.

Statistics of mapping	
Clean Paired-end Reads	1,937,852,694
Mapped Reads	1,119,749,423
Mapped Ratio (%)	57.78
Unique Mapped Paired-end Reads	167,615,280
Unique Mapped Ratio (%)	8.65
Statistics of valid reads	
Unique Mapped Paired-end Reads	167,615,280
Valid Interaction Pairs	48,858,596
Valid Interaction Rate (%)	58.30
Dangling End Pairs	9,986,977
Dangling End Rate (%)	11.92
Re-ligation Pairs	1,531,655
Re-ligation Rate (%)	1.83
Self Circle Paired-end Reads	304,359
Self Circle Rate (%)	0.36
Dumped Paired-end Reads	85,558
Dumped Rate (%)	0.11

Supplementary Table S4. Statistics for chromosome-level scaffolding of the genome of *S. spontaneum* Np-X using Hi-C data.

Chromosome	Haplotype	No. of Anchored	Total length	No. of genes
Chr1	A	524	93,044,720	5,029
	B	580	104,759,444	5,416
	C	700	100,769,705	4,792
	D	403	84,453,026	4,094
Chr2	A	411	70,752,275	3,613
	B	415	74,057,031	3,982
	C	416	90,787,606	4,182
	D	389	84,572,824	4,021
Chr3	A	428	72,154,782	3,990
	B	407	75,508,239	3,642
	C	431	89,648,162	4,335
	D	489	83,912,231	4,271
Chr4	A	407	83,891,557	3,938
	B	309	65,715,293	3,371
	C	325	63,242,035	2,926
	D	392	66,330,368	2,710
Chr5	A	457	72,028,841	2,652
	B	321	69,201,893	2,789
	C	336	66,068,211	2,509
	D	377	73,606,151	2,740
Chr6	A	287	56,802,555	2,519
	B	341	61,627,438	2,537
	C	353	64,567,053	2,902
	D	279	61,065,058	2,615
Chr7	A	268	59,240,971	2,375
	B	278	59,761,120	2,513
	C	235	54,872,887	2,277
	D	221	47,449,267	2,158
Chr8	A	146	42,784,806	1,844
	B	320	54,634,649	1,901
	C	181	41,941,943	1,803
	D	360	61,380,726	1,940
Chr9	A	413	62,824,460	2,622
	B	318	57,226,191	2,444
	C	366	63,676,737	2,404
	D	290	58,305,488	2,526
Chr10	A	234	54,170,167	2,439
	B	253	67,739,541	2,831
	C	295	58,708,265	2,456
	D	240	63,487,932	2,792
Oriented	-	14,195(95.44%)	2,736,771,648(98.96%)	122,900(99.81%)
Unanchored	-	680(4.56%)	24,418,322(1.04%)	228(0.19%)
In total		14,872	2,765,452,713	123,128

Supplementary Table S5. Identification of genetic variation in **homolog group compared to the monoploid *S. spontaneum* Np-X genome.**

Haplotype	Variation	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10
A	SNPs	352741	222861	237667	252152	160584	165734	150672	92451	216652	147559
	No. of Indels(1-10bp)	31016	19158	22174	14231	12218	14203	12102	6834	11021	11677
	No. of large Indels(>10bp)	2404	1438	1589	282	794	1011	777	426	176	785
	Size of Indels(1-10bp)	62196	38222	43813	15512	23181	28329	23622	13099	12111	23112
	No. of Repeat expansion/contraction	325	189	208	6	59	132	78	32	8	87
	Size of Repeat expansion/contraction	117838	63367	75993	2463	14590	48884	27077	12466	2574	30250
B	SNPs	370516	243388	249163	211432	148243	195997	164237	152965	178479	190296
	No. of Indels(1-10bp)	22040	21191	21222	17406	11394	15926	8850	11099	17670	11080
	No. of large Indels(>10bp)	498	1490	1492	1318	664	1060	130	737	919	40
	Size of Indels(1-10bp)	23724	41642	42209	35428	21836	31147	9624	20512	31102	11919
	No. of Repeat expansion/contraction	22	208	162	176	74	137	3	60	90	0
	Size of Repeat expansion/contraction	10125	67638	61843	61284	20434	47915	1326	19704	37413	0
C	SNPs	351549	279737	292665	221412	140139	211015	133408	109060	190834	154578
	No. of Indels(1-10bp)	30622	16194	17519	17062	10344	12152	10578	8806	17782	12772
	No. of large Indels(>10bp)	2499	369	385	1177	747	234	693	539	1039	823
	Size of Indels(1-10bp)	61426	17357	18989	34232	20151	13031	21046	16845	32064	25290
	No. of Repeat expansion/contraction	322	16	15	144	62	12	66	51	103	67
	Size of Repeat expansion/contraction	116580	7130	6180	48812	26342	5288	20808	21354	37462	25322
D	SNPs	284697	255938	290473	193803	193096	177198	121946	151240	170096	173414
	No. of Indels(1-10bp)	23137	21278	25238	15254	9499	15876	9546	6823	14555	14238
	No. of large Indels(>10bp)	1750	1533	1882	1029	151	1031	646	97	836	973
	Size of Indels(1-10bp)	46468	42588	50050	30557	10105	31441	18817	7150	27304	28581
	No. of Repeat expansion/contraction	209	172	219	130	3	124	61	3	94	104
	Size of Repeat expansion/contraction	71780	57152	79980	46192	1240	42540	21556	1148	33122	39141

Supplementary Table S6. BUSCO analysis of the *S. spontaneum* Np-X genome assembly.

Description	Number	Percentage (%)
Complete BUSCOs (C)	1389	96.46
Complete and single-copy BUSCOs (S)	154	10.69
Complete and duplicated BUSCOs (D)	1235	85.76
Fragmented BUSCOs (F)	6	0.42
Missing BUSCOs (M)	45	3.12
Total BUSCO groups searched	1440	100

Supplementary Table S7. CEGMA analysis of the completeness of the *S. spontaneum* Np-X genome.

Description	Fully mapped CEGs	Fully+partially mapped CEGs
Number of CEGs present in the assembly	226	236
Completeness of the genome (%)	91.13	95.16
Average number of orthologs per CEG	4.15	4.44
CEGs with more than one ortholog (%)	96.02	97.46

Supplementary Table S8. Distribution of centromeric regions in each *S. spontaneum* Np-X chromosome.

Chromosome	Start	End	Length (Kb)
Chr1B	48,544,356	51,938,914	3,394.56
Chr1C	42,167,324	50,797,552	8,630.23
Chr1D	32,358,756	35,736,704	3,377.95
Chr2C	35,126,811	39,583,066	4,456.26
Chr2D	33,281,957	37,479,202	4,197.25
Chr3B	31,142,464	34,458,593	3,316.13
Chr3C	36,679,178	42,933,051	6,253.87
Chr3D	33,486,849	36,221,715	2,734.87
Chr4A	35,552,432	39,395,515	3,843.08
Chr4B	27,663,126	31,272,689	3,609.56
Chr4D	26,098,526	32,023,448	5,924.92
Chr5A	33,050,133	39,134,772	6,084.64
Chr5B	31,573,653	35,492,974	3,919.32
Chr5C	35,028,151	37,922,080	2,893.93
Chr5D	32,741,207	36,826,796	4,085.59
Chr6A	15,603,276	20,044,078	4,440.80
Chr7A	27,604,959	31,647,111	4,042.15
Chr7B	29,688,848	33,493,217	3,804.37
Chr7C	19,537,064	25,426,544	5,889.48
Chr7D	22,763,279	23,810,362	1,047.08
Chr8B	20,517,727	24,556,859	4,039.13
Chr8D	25,058,818	31,557,234	6,498.42
Chr9A	26,718,432	29,809,884	3,091.45
Chr9B	31,432,527	32,024,319	591.79
Chr9C	30,029,768	35,391,925	5,362.16
Chr9D	26,789,645	27,694,285	904.64
Chr10A	27,464,928	30,430,936	2,966.01
Chr10B	31,018,907	36,242,171	5,223.26
Chr10C	26,693,743	31,123,951	4,430.21
Chr10D	32,606,994	35,820,300	3,213.31

Supplementary Table S9. Characteristics of the telomere in each *S. spontaneum* Np-X chromosomes.

Chromosome	Start	End	Copy Number	Monomer
Chr1A	3,850,279	3,852,474	317.4	TAAACCC
Chr1A	58,791,550	58,792,631	155.3	AGGGTTT
Chr1B	782,689	783,956	183	ACCCTAA
Chr1B	104,135,301	104,137,868	370.4	AACCCTA
Chr1C	99,882,266	99,883,543	175.3	TTTAGGG
Chr1D	84,194,024	84,201,062	955.4	AACCCTA
Chr2B	18,094,098	18,095,477	195	TTAGGGT
Chr2B	73,779,502	73,781,800	336.3	TTTAGGG
Chr2C	754,038	755,255	175.6	AGGGTTT
Chr2D	9,811,260	9,812,036	104	GGGTTTA
Chr3A	20,886,528	20,887,561	138.4	AACCCTA
Chr3A	71,111,714	71,115,159	492.9	TTAGGGT
Chr3C	52,842,002	52,845,060	452.1	TAAACCC
Chr3C	89,625,503	89,630,644	736.9	TAGGGTT
Chr4A	83,494,887	83,496,823	279	AGGGTTT
Chr4B	63,454,736	63,456,470	250.3	TTAGGGT
Chr4C	2,946,418	2,954,821	1218.3	GGTTTAG
Chr5A	1,062,308	1,064,769	351.6	AACCCTA
Chr5A	66,861,228	66,862,180	136	TTAGGGT
Chr5A	67,397,233	67,399,798	368.3	TTTAGGG
Chr5B	580,934	583,665	389	CCTAAAC
Chr5B	43,948,316	43,951,395	397.9	CCTAAAC
Chr6A	88,191	92,611	609.6	TAAACCC
Chr6A	96,778	102,755	827.6	TAAACCC
Chr6A	112,012	114,113	292.6	TAAACCC
Chr6A	123,117	129,729	918.6	TAAACCC
Chr6A	136,207	137,357	161.6	TAAACCC
Chr6A	217,866	220,768	403.6	TAAACCC
Chr6A	231,745	233,163	197.6	TAAACCC
Chr6A	55,172,508	55,175,338	394	CCCTAAA
Chr7B	522,226	530,115	1127.6	TAAACCC
Chr7B	56,463,173	56,468,813	805.4	TAAACCC
Chr7D	47,429,334	47,434,973	809.9	CCTAAAC
Chr8A	995,742	996,458	105.9	TTAGGGT
Chr8C	1,395,874	1,400,807	727	CCTAAAC
Chr8D	49,262,491	49,265,661	449.9	TAGGGTT
Chr8D	60,068,247	60,070,328	299.9	CCCTAAA
Chr9B	9,399,632	9,407,086	1065.3	AAACCTT
Chr9C	5,663,423	5,665,990	371	CCCTAAA
Chr9D	2,318,989	2,320,539	231.3	AACCCTA

Chr10A	170,746	174,035	480	AAACCCT
Chr10B	1,884,298	1,886,290	287	CCTAAAC
Chr10B	23,443,490	23,444,544	149.1	CCCTAAA

Supplementary Table S10. General statistics for annotation of gene functions in the *S. spontaneum* Np-X genome

	Type	Number	Percent (%)
Annotated	InterPro	95,988	78.1
	GO	68,379	55.6
	COG	6,146	5.0
	KOG	28,944	23.5
	Swissprot	76,073	61.9
	uniprot	76,073	61.9
	Nr	112,431	91.4
	Annotated	115,167	93.67
	Unannotated	7,778	6.33
	Total	122,945	100.0

Supplementary Table S11. Statistics for annotations of alleles in the *S. spontaneum* Np-X genome.

	Total no. of genes	Total no. of alleles	No. of genes with 4 alleles	No. of genes with 3 alleles	No. of genes with 2 alleles	No. of genes with 1 allele	No. of dispersely duplicated genes	No. of tandem duplicated genes
Chr01	5779	19331	1258	2321	1692	508	965	602
Chr02	4514	15798	1203	1600	1311	400	613	581
Chr03	4750	16238	1198	1881	1263	408	570	552
Chr04	3862	12945	840	1501	1164	357	471	447
Chr05	2931	10690	614	860	1070	387	342	581
Chr06	3049	10573	868	1031	861	289	241	396
Chr07	2720	9323	625	930	852	313	278	371
Chr08	2204	7488	427	765	749	263	202	364
Chr09	2951	9996	795	942	910	304	298	384
Chr10	3061	10518	817	1029	907	308	302	418
Unanchored	9	228	0	1	2	6	-	-
Gene with annotated alleles	35830	123128	8645	12861	10781	3543	-	-

Supplementary Table S12. Statistics for transposable elements in the *S. spontaneum* Np-X genome.

	Length(Mb)	% of Repeat	% of genome
Total Repeat Fractions	1590.66	100	57.52
Class I :Retrotransposon	1170.64	73.60	42.33
LTR Retrotransposon			
Ty3/Gypsy	798.49	50.20	28.87
Ty1/Copia	317.12	19.94	11.47
Other	8.19	0.51	0.30
Non-LTR Retrotransposon			
LINE	42.75	2.69	1.55
SINE	4.10	0.26	0.15
Class II :DNA transposon	233.64	14.69	8.45
CACTA	71.76	4.51	2.59
Mutator	39.01	2.45	1.41
hAT	23.51	1.48	0.85
PIF/Harbinger	57.94	3.64	2.10
Helitron	11.69	0.74	0.42
Other	29.73	1.87	1.07
Tandem Repeat	88.82	5.58	3.21
Unknown	97.56	6.13	3.53

Supplementary Table S13. Statistics for transposable elements in the genomes of some species in the Poaceae.

	AP85-441	Np-X	<i>S. officinarum</i>	Sorghum	Miscanthus
Total Repeat Fractions (%)	56.02	57.52	57.79	61.30	60.14
Class I :Retrotransposon (%)	41.39	42.33	43.01	45.59	44.24
LTR Retrotransposon (%)					
Ty3/Gypsy (%)	25.33	28.87	25.45	38.78	31.52
Ty1/Copia (%)	14.19	11.47	15.96	5.70	11.93
Other (%)	0.12	0.30	0.18	0.18	0.10
Non-LTR Retrotransposon (%)					
LINE (%)	1.72	1.55	1.39	0.93	0.68
SINE (%)	0.03	0.15	0.03	0	0.02
Class II :DNA transposon (%)	8.71	8.45	7.28	8.69	8.56
CACTA (%)	2.89	2.59	2.47	3.56	4.46
Mutator (%)	1.51	1.41	1.15	0.68	1.62
hAT (%)	0.99	0.85	0.71	0.51	0.55
PIF/Harbinger (%)	1.73	2.10	1.86	2.79	0.95
Helitron (%)	0.51	0.42	0.38	0.28	0.34
Other (%)	1.08	1.07	0.71	0.86	0.64
Tandem Repeat (%)	1.39	3.21	1.49	2.56	0.75
Unknown (%)	4.53	3.53	5.90	4.46	6.59
Repeat total content (Mb)	1759.72	1590.66	3932.30	434.52	1250.62

Supplementary Table S14. Genes absent near the recombination breakpoint of NpChr5 in *S. spontaneum* Np-X compared to *S. spontaneum* AP85-441.

Gene ID	Ortholog genes	The ortholog gene in Species	Function
Npp.05D013180	XP_021320472.1	Sorghum bicolor	probable Ufm1-specific protease isoform X2
Npp.05D013190	XP_002489311.1	Sorghum bicolor	Probable beta-D-xylosidase 7
Npp.05D013200	NP_001132410.1	Zea mays	uncharacterized protein LOC100193857
Npp.05D013210	OQU87166.1	Sorghum bicolor	hypothetical protein SORBI_3003G220100
Npp.05D013220	AAX96588.1	Oryza sativa	Transposable element protein, putative
Npp.05D013230	AAX96588.1	Oryza sativa	Transposable element protein, putative
Npp.05D013240	XP_002444762.1	Sorghum bicolor	plant intracellular Ras-group-related LRR protein 4
Npp.05D013250	AGT16190.1	Saccharum hybrid cultivar R570	translation initiation factor
Npp.05D013260	XP_021317279.1	Sorghum bicolor	uncharacterized protein LOC8059223 isoform X1
Npp.05D013270	AGT16190.1	Saccharum hybrid cultivar R570	translation initiation factor
Npp.05D013280	ACE86399.1	Sorghum bicolor	amino acid selective channel protein
Npp.05D013290	WP_188249283.1	Serratia marcescens	DDE-type integrase/transposase/recombinase
Npp.05D013300	XP_021305712.1	Sorghum bicolor	uncharacterized protein LOC110431233
Npp.05D013310	NP_001132772.1	Zea mays	eukaryotic initiation factor 5C CG2922-PF
Npp.05D013320	PWZ28846.1	Zea mays	Flavin mononucleotide hydrolase 1
Npp.05D013330	XP_021310236.1	Sorghum bicolor	DNA polymerase epsilon subunit 3
Npp.05D013340	ABA98082.1	Oryza sativa	retrotransposon protein, putative, Ty3-gypsy subclass
Npp.05D013350	XP_039787907.1	Panicum virgatum	uncharacterized protein LOC120654436
Npp.05D013360	XP_002450710.1	Zea mays	uncharacterized protein LOC8071849
Npp.05D013370	XP_021316769.1	Sorghum bicolor	protein NRT1/ PTR FAMILY 1.2
Npp.05D013380	XP_021316769.1	Sorghum bicolor	protein NRT1/ PTR FAMILY 1.2
Npp.05D013390	XP_004979230.1	Setaria italica	40S ribosomal protein S5
Npp.05D013400	XP_021316769.1	Sorghum bicolor	protein NRT1/ PTR FAMILY 1.2
Npp.05D013410	XP_021316769.1	Sorghum bicolor	protein NRT1/ PTR FAMILY 1.2
Npp.05D013420	XP_004979230.1	Setaria italica	40S ribosomal protein S5

Npp.05D013430	XP_002489340.1	<i>Sorghum bicolor</i>	protein transport protein SEC23
Npp.05D013440	XP_002439601.2	<i>Sorghum bicolor</i>	uncharacterized protein LOC8061323
Npp.05D013450	XP_021317888.1	<i>Sorghum bicolor</i>	embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein
Npp.05D013460	XP_002439601.2	<i>Sorghum bicolor</i>	uncharacterized protein LOC8061323
Npp.05D013470	XP_021316769.1	<i>Sorghum bicolor</i>	protein NRT1/ PTR FAMILY 1.2
Npp.05D013480	XP_004979230.1	<i>Setaria italica</i>	40S ribosomal protein S5
Npp.05D013490	XP_021316769.1	<i>Sorghum bicolor</i>	protein NRT1/ PTR FAMILY 1.2
Npp.05D013500	XP_020112239.1	<i>Ananas comosus</i>	40S ribosomal protein S5
Npp.05D013510	XP_021316769.1	<i>Sorghum bicolor</i>	protein NRT1/ PTR FAMILY 1.2

Supplementary Table S15. Genes absent near the recombination breakpoint of NpChr8 in *S. spontaneum* Np-X compared to *S. spontaneum* AP85-441.

Gene ID	Ortholog genes	The ortholog gene in Species	Function
Npp.08D008970	XP_002442723.1	Sorghum bicolor	cucumber peeling cupredoxin
Npp.08D008980	XP_002442723.1	Sorghum bicolor	cucumber peeling cupredoxin
Npp.08D008990	XP_002442723.1	Sorghum bicolor	cucumber peeling cupredoxin
Npp.08D009000	XP_002489290.2	Sorghum bicolor	nuclear pore complex protein NUP1
Npp.08D009010	XP_002489290.2	Sorghum bicolor	nuclear pore complex protein NUP1
Npp.08D008920	KAF5797026	Helianthus annuus	putative photosystem I PsaA/PsaB
Npp.08D008870	PVH34726.1	Panicum hallii	Retrotransposon gag protein; pfam03732
Npp.08D008930	KAF5797026.1	Helianthus annuus	putative photosystem I PsaA/PsaB
Npp.08D008940	PWZ53246.1	Zea mays	60S ribosomal protein L10a
Npp.08D008910	KAF5797026.1	Helianthus annuus	putative photosystem I PsaA/PsaB
Npp.08D008900	BAH80072.1	Oryza sativa	putative unclassified retrotransposon protein
Npp.08D008860	XP_002442112.2	Sorghum bicolor	ATPase family AAA domain-containing protein FIGL1
Npp.08D008950	XP_025801448.1	Panicum hallii	uncharacterized protein LOC112880939
Npp.08D008880	KAF5797026.1	Helianthus annuus	putative photosystem I PsaA/PsaB
Npp.08D008890	YP_009134927.1	Chusquea spectabilis	photosystem II protein D1 (plastid)
Npp.08D008960	PWZ53246.1	Zea mays	60S ribosomal protein L10a

Supplementary Table S16. A/B compartment variation between homologous chromosome groups in *S. spontaneum* Np-X.

Chr	Bi-conserved	Tri-conserved	All-conserved
Chr1	192(15.8%)	838(68.8%)	188(15.4%)
Chr2	438(37.7%)	696(59.8%)	29(2.5%)
Chr3	119(10.2%)	406(34.7%)	645(55.1%)
Chr4	146(17.6%)	677(81.6%)	7(0.8%)
Chr5	234(38.3%)	345(56.5%)	32(5.2%)
Chr6	114(13.2%)	729(84.6%)	19(2.2%)
Chr7	129(21.2%)	470(77.2%)	10(1.6%)
Chr8	83(19.3%)	163(37.9%)	184(42.8%)
Chr9	659(83.5%)	116(14.7%)	14(1.8%)
Chr10	126(15.8%)	635(79.7%)	36(4.5%)
Intotal	2,240(26.4%)	5,075(59.9%)	1,164(13.7%)

Supplementary Table S17. The distribution of A/B compartments in four species of Poaceae.

	Np-X	AP85-441	<i>O. sativa</i>	Sorghum
Size of A compartment (Mb)	367.54	367.42	180.52	306.57
Ratio of A compartment(%)	48	47	47	45
No. of genes in A compartment	14,067	9,627	19,092	16,735
Size of B compartment (Mb)	405.48	417.89	201.28	377.07
Ratio of B compartment(%)	52	53	53	55
No. of genes in B compartment	19,069	11,685	22,912	17,292

Supplementary Table S18. The number of genes likely involved in sugar metabolism in *S. spontaneum* Np-X, *S. spontaneum* AP85-441, Sorghum, Rice and Arabidopsis.

Gene families		Gene numbers				
		Arabidopsis	Rice	Sorghum	<i>S. spontaneum</i> AP85-441	<i>S. spontaneum</i> Np-X
Sugar metabolism	Sucrose phosphate synthase (SPS)	4	5	5	7	6
	Sucrose synthase (SUS)	6	6	5	4	8
	Invertase (INV)	17	17	19	20	26
	Fructokinase (FRK)	13	7	7	9	11
Total		40	35	36	40	51
Sugar transporter	SWEET	17	23	23	22	32
	SUT	9	5	6	9	8
	PLT	6	11	17	31	31
	STP/HXT	14	21	22	35	25
	INT	4	2	4	4	4
	VGT	3	2	2	4	3
	pGlcT	1	1	1	3	3
	SFP	17	4	7	14	13
	TMT	3	4	3	5	5
Total		75	68	86	127	124

Supplementary Table S19. The C4 pathway genes in *S. spontaneum* Np-X, *S. spontaneum* AP85-441 and Sorghum.

<i>S. spontaneum</i> Np-X	Representative allele	<i>S. bicolor</i>	<i>S. spontaneum</i> AP85-441	Representative allele	% identity XSs vs Sb	% identity XSs vs Ss
<i>XSsCA1*</i>	<i>Npp.03C022280.1</i>	<i>SbCA1*</i>	<i>SsCA1*</i>	<i>Sspon.003C0020330</i>	90.57	87.02
<i>XSsCA2*</i>	<i>Npp.03A021921.0</i>	<i>SbCA2*</i>	<i>SsCA2*</i>	<i>Sspon.003C0020320</i>	96.10	99.51
<i>XSsCA3</i>	<i>Npp.03C021922.1</i>	<i>SbCA3</i>	<i>SsCA3</i>	<i>Sspon.003B0015790</i>	98.03	96.99
<i>XSsCA4</i>	<i>Npp.03A021930.1</i>	<i>SbCA4</i>	<i>SsCA4</i>	<i>Sspon.003B0015800</i>	96.15	95.85
<i>XSsPEPC1*</i>	<i>Npp.10A014060.1</i>	<i>SbPEPC1*</i>	<i>SsPEPC1*</i>	<i>Sspon.008A0009990</i>	96.67	97.19
<i>XSsPEPC2</i>	<i>Npp.02C017520.1</i>	<i>SbPEPC2</i>	<i>SsPEPC2</i>	<i>Sspon.002C0022221</i>	99.17	99.90
<i>XSsPEPC3</i>	<i>Npp.03C028320.1</i>	<i>SbPEPC3</i>	<i>SsPEPC3</i>	<i>Sspon.003C0015360</i>	98.97	95.54
<i>XSsPEPC4</i>	<i>Npp.04A013350.1</i>	<i>SbPEPC4</i>	<i>SsPEPC4</i>	<i>Sspon.004D0019501</i>	99.48	99.90
<i>XSsPEPC5</i>	<i>Npp.07B014690.1</i>	<i>SbPEPC5</i>	<i>SsPEPC5</i>	<i>Sspon.006C0010170</i>	98.85	98.85
<i>XSsPEPC-k1*</i>	<i>Npp.04B030200.1</i>	<i>SbPEPC-k1*</i>	<i>SsPEPC-k1*</i>	<i>Sspon.004D0002690</i>	96.79	99.25
<i>XSsPEPC-k2</i>	<i>Npp.04B019480.1</i>	<i>SbPEPC-k2</i>	<i>SsPEPC-k2</i>	<i>Sspon.004B0009111</i>	93.18	99.01
<i>XSsPEPC-k3</i>	<i>Npp.06B014490.1</i>	<i>SbPEPC-k3</i>	<i>SsPEPC-k3</i>	<i>Sspon.005C0005400</i>	96.77	97.81
<i>XSsNADP-MDH1</i>	<i>Npp.07C016531.0</i>	<i>SbNADP-MDH1</i>	<i>SsNADP-MDH1</i>	<i>Sspon.006D0005110</i>	96.82	81.28
<i>XSsNADP-MDH2*</i>	<i>Npp.07D014800.1</i>	<i>SbNADP-MDH2*</i>	<i>SsNADP-MDH2*</i>	<i>Sspon.006D0004990</i>	96.77	96.00
<i>XSsNADP-ME1</i>	<i>Npp.03D002980.1</i>	<i>SbNADP-ME1</i>	<i>SsNADP-ME1</i>	<i>Sspon.003C0036192</i>	83.64	77.74
<i>XSsNADP-ME2*</i>	<i>Npp.03D002990.1</i>	<i>SbNADP-ME2*</i>	<i>SsNADP-ME2*</i>	<i>Sspon.003B0032970</i>	98.43	83.57
<i>XSsNADP-ME3</i>	<i>Npp.03C027450.1</i>	<i>SbNADP-ME3</i>	<i>SsNADP-ME3</i>	<i>Sspon.003A0008210</i>	98.65	96.68
<i>XSsNADP-ME4</i>	<i>Npp.09D007600.1</i>	<i>SbNADP-ME4</i>	<i>SsNADP-ME4</i>	<i>Sspon.007D0012940</i>	97.20	99.47
<i>XSsNADP-ME5</i>	<i>Npp.03C026470.1</i>	<i>SbNADP-ME5</i>	<i>SsNADP-ME5</i>	<i>Sspon.003A0008980</i>	99.66	96.30
<i>XSsNADP-ME6</i>	<i>Npp.09B011130.1</i>	<i>SbNADP-ME6</i>	<i>SsNADP-ME6</i>	<i>Sspon.007C0010990</i>	95.71	88.25
<i>XSsPPDK1*</i>	<i>Npp.09A013470.1</i>	<i>SbPPDK1*</i>	<i>SsPPDK1*</i>	<i>Sspon.007A0010811</i>	98.52	99.27
<i>XSsPPDK2</i>	<i>Npp.01B032990.1</i>	<i>SbPPDK2</i>	<i>SsPPDK2</i>	<i>Sspon.001C0014951</i>	97.36	99.86
<i>XSsPPDK-RP1*</i>	<i>Npp.02C031010.1</i>	<i>SbPPDK-RP1*</i>	<i>SsPPDK-RP1*</i>	<i>Sspon.002A0008810</i>	91.00	98.64
<i>XSsPPDK-RP2</i>	<i>Npp.02C031020.1</i>	<i>SbPPDK-RP2</i>	<i>SsPPDK-RP2</i>	<i>Sspon.002B0007451</i>	95.95	98.80
<i>XSsPPDK-RP3</i>	<i>Npp.04B028840.1</i>	<i>SbPPDK-RP3</i>	<i>SsPPDK-RP3</i>	<i>Sspon.002D0005590</i>	91.07	100.00
NA	NA	NA	<i>SsPEPCK1</i>	<i>Sspon.001B0013950</i>	NA	NA
<i>XSsPEPCK</i>	<i>Npp.01C037370.1</i>	<i>SbPEPCK</i>	<i>SsPEPCK2</i>	<i>Sspon.001A0020841</i>	99.09	95.64
<i>XSsNAD-ME1</i>	<i>Npp.02C029720.1</i>	<i>SbNAD-ME1</i>	<i>SsNAD-ME1</i>	<i>Sspon.002C0010890</i>	99.53	100.00
<i>XSsNAD-ME2</i>	<i>Npp.01C015670.1</i>	<i>SbNAD-ME2</i>	<i>SsNAD-ME2</i>	<i>Sspon.001C0025371</i>	99.02	93.49

Supplementary Table S20. Detailed information for all resequenced accessions.

Accessions	Species	location	Ploidy	Basic chromosome number	Data size(Gb)
AP85-441	<i>S. spontaneum</i>	-	4	8	241.28
AP85-68	<i>S. spontaneum</i>	-	4	8	34.83
FJ-89-1-1	<i>S. spontaneum</i>	Fujian, China	11	8	79.17
FU-89-1-16	<i>S. spontaneum</i>	Fujian, China	11	8	61.97
IN84-021	<i>S. spontaneum</i>	Raha, Indonesia	10	8	46.90
IN84-089	<i>S. spontaneum</i>	Salodik, Indonesia	6	8	32.55
GZ-78-1-11	<i>S. spontaneum</i>	Guangzhou, China	9	8	121.33
Gugu	<i>S. spontaneum</i>	Kenya	8	8	36.71
HN-2	<i>S. spontaneum</i>	Hainan, China	10	8	124.46
Holes1	<i>S. spontaneum</i>	Coimbatore, India	6	9	53.04
IK76-067	<i>S. spontaneum</i>	Panajam, Indonesia	10	8	47.48
IN76-086	<i>S. spontaneum</i>	Indonesia	10	8	43.99
IND81-03	<i>S. spontaneum</i>	India	8	9	41.01
IND81-05	<i>S. spontaneum</i>	India	8	9	59.90
IND81-08	<i>S. spontaneum</i>	India	10	8	48.22
IND81-13	<i>S. spontaneum</i>	India	8	10	46.65
IND81-14	<i>S. spontaneum</i>	India	8	8	40.95
IND81-15	<i>S. spontaneum</i>	India	8	8	38.66
IND81-17	<i>S. spontaneum</i>	India	10	8	47.61
IND81-18	<i>S. spontaneum</i>	India	8	10	49.96
Iranspon	<i>S. spontaneum</i>	Iran	14	8	67.88
PCANOR84	<i>S. spontaneum</i>	Philippines	8	8	37.20
PPGN84-0	<i>S. spontaneum</i>	Philippines	6	8	33.40
PTAR84-0	<i>S. spontaneum</i>	Philippines	9	8	46.51
s15-95	<i>S. spontaneum</i>	-	8	8	44.44
SC-79-2-11	<i>S. spontaneum</i>	Sichuan,China	11	8	118.72
SES004A	<i>S. spontaneum</i>	India	10	8	46.26
SaudiAra	<i>S. spontaneum</i>	Coimbatore, India	12	9	60.34
SES014	<i>S. spontaneum</i>	India	12	8	55.83
SES072	<i>S. spontaneum</i>	India	6	8	30.50
SES113A	<i>S. spontaneum</i>	India	6	8	38.68
SES184B	<i>S. spontaneum</i>	India	15	9	55.62
SES196	<i>S. spontaneum</i>	India	6	9	42.13
SES197A	<i>S. spontaneum</i>	India	8	8	69.28
SES208	<i>S. spontaneum</i>	India	8	8	53.67
SES234	<i>S. spontaneum</i>	Malaysia	8	8	36.57
SES239-43	<i>S. spontaneum</i>	-	9	8	38.42
SES264	<i>S. spontaneum</i>	India	8	9	45.01

SES275	<i>S. spontaneum</i>	India	6	8	29.81
SES289	<i>S. spontaneum</i>	-	10	8	56.18
SES294	<i>S. spontaneum</i>	India	8	8	38.57
SES297B	<i>S. spontaneum</i>	India	10	8	49.83
SES341	<i>S. spontaneum</i>	India	6	8	32.14
SES365	<i>S. spontaneum</i>	India	5	8	35.69
SES517	<i>S. spontaneum</i>	India	8	8	36.45
SES519	<i>S. spontaneum</i>	India	7	8	44.95
SES561	<i>S. spontaneum</i>	Sudan	16	8	37.92
SES602	<i>S. spontaneum</i>	India	10	8	45.68
SLC92-81	<i>S. spontaneum</i>	sri lanka	8	8	40.14
SLC92-94	<i>S. spontaneum</i>	sri lanka	8	8	34.90
Shoaguan	<i>S. spontaneum</i>	Sichuan, China	8	8	33.12
S-spont-jing	<i>S. spontaneum</i>	-	8	8	32.85
S-spontI	<i>S. spontaneum</i>	-	10	8	31.64
SM7916	<i>S. spontaneum</i>	-	6	8	33.24
Taiwansp	<i>S. spontaneum</i>	Taiwan, China	12	8	62.27
Tongza	<i>S. spontaneum</i>	China	8	8	38.31
US48-60	<i>S. spontaneum</i>	-	13	8	42.56
US56-14-4	<i>S. spontaneum</i>	Thailand	10	8	90.61
US60-004	<i>S. spontaneum</i>	Afghanistan	6	9	29.70
US78-500	<i>S. spontaneum</i>	Pakistan	8	8	37.41
YN-76-1-20	<i>S. spontaneum</i>	Yunnan, China	14	8	153.94
YNMZ	<i>S. spontaneum</i>	Yunnan, China	10	8	84.14
YNXD	<i>S. spontaneum</i>	Yunnan, China	11	8	115.53
Yacheng-jing	<i>S. spontaneum</i>	China	10	8	41.77
Yunnan82-25	<i>S. spontaneum</i>	Yunnan, China	8	8	24.75
Yunnan83-160	<i>S. spontaneum</i>	Yunnan, China	8	8	24.14
Yunnan75- II -35	<i>S. spontaneum</i>	Yunnan, China	10	8	31.98
Yunnan76- I -024	<i>S. spontaneum</i>	Yunnan, China	10	8	31.77
Yunnan82-33	<i>S. spontaneum</i>	Yunnan, China	10	8	28.59
Yunnan84-260	<i>S. spontaneum</i>	Guizhou, China	10	8	29.58
Guangdong-No.1	<i>S. spontaneum</i>	Guangdong, China	10	8	29.79
Fujiandongshan	<i>S. spontaneum</i>	Fujian, China	10	8	30.82
Yunnan82-67	<i>S. spontaneum</i>	Yunnan, China	8	8	22.01
Yunnan76- I -016	<i>S. spontaneum</i>	Yunnan, China	9	8	24.65
Sichuan92-42	<i>S. spontaneum</i>	Sichuan, China	9	8	27.13
Guizhou78- II -28	<i>S. spontaneum</i>	Guizhou, China	12	8	30.01
Guangdong30	<i>S. spontaneum</i>	Guangdong, China	12	8	33.57
Fujianhuian	<i>S. spontaneum</i>	Fujian, China	12	8	58.37
Sichuan79- II -11	<i>S. spontaneum</i>	Sichuan, China	12	8	31.09
Fujian89- I -19	<i>S. spontaneum</i>	Fujian, China	13	8	33.35
Yunnan76- I -003	<i>S. spontaneum</i>	Yunnan, China	9	8	27.82

Yunnan82-70	<i>S. spontaneum</i>	Yunnan, China	9	8	24.16
Guangdong70	<i>S. spontaneum</i>	Guangdong, China	9	8	23.08
Yunnan82-91	<i>S. spontaneum</i>	Yunnan, China	9	8	31.93
Fujian79- II -9	<i>S. spontaneum</i>	Fujian, China	9	8	25.51
Sichuan79- I -31	<i>S. spontaneum</i>	Sichuan, China	9	8	26.54
Yunnan83-189	<i>S. spontaneum</i>	Yunnan, China	10	8	32.11
Sichuan79- II -18	<i>S. spontaneum</i>	Sichuan, China	10	8	31.15
Fujian87- I -11	<i>S. spontaneum</i>	Fujian, China	12	8	34.48
Sichuan79- I -1	<i>S. spontaneum</i>	Sichuan, China	12	8	40.61
2015-25	<i>S. spontaneum</i>	Yangon; Myanmar	8	8	25.37
2015-68	<i>S. spontaneum</i>	Myanmar	10	9	35.32
2015-95	<i>S. spontaneum</i>	Mandalay; Myanmar	6	8	15.91
2015-31	<i>S. spontaneum</i>	Irrawaddy; Myanmar	12	8	33.94
2015-61	<i>S. spontaneum</i>	Pegu river; Myanmar	7	8	23.26
2015-77	<i>S. spontaneum</i>	Myanmar	12	8	32.94
2012-43	<i>S. spontaneum</i>	Tibet; China	6	9	17.68
Vietnam3	<i>S. spontaneum</i>	Vietnam	10	8	28.13
Laos2	<i>S. spontaneum</i>	Laos	10	8	27.17
Guizhou78- I -11	<i>S. spontaneum</i>	Guizhou, China	9	8	25.91
2013-8	<i>S. spontaneum</i>	Nepal	4	10	11.42
Np-X	<i>S. spontaneum</i>	Tibet; China	4	10	53.24

Note: Data for accessions highlighted in red were downloaded from Zhang *et al.*⁶

Reference:

- 1 Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**, e49, doi:10.1093/nar/gkr1293 (2012).
- 2 Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular biology and evolution* **34**, 2115-2122, doi:10.1093/molbev/msx148 (2017).
- 3 Schaller, A. *et al.* From structure to function - a family portrait of plant subtilases. *The New phytologist* **218**, 901-915, doi:10.1111/nph.14582 (2018).
- 4 Reichardt, S. *et al.* The tomato subtilase family includes several cell death-related proteinases with caspase specificity. *Scientific reports* **8**, 10531, doi:10.1038/s41598-018-28769-0 (2018).
- 5 Kim, H. S., Lee, C. G. & Lee, E. Y. Alginate Lyase: Structure, Property, and Application. *Biotechnology & Bioprocess Engineering* **16**, p.843-851.
- 6 Zhang, J. *et al.* Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics* **50**, 1565-1573, doi:10.1038/s41588-018-0237-2 (2018).