1 Supplementary Notes

1.1 Supplementary notes of C-Phasing

1.1.1 Supplementary module of Hitig

In this section, we describe some auxiliary modules of Hitig that leverage Pore-C or Hi-C contact data to detect both chimeric and high-confidence regions (HCRs).

Chimeric contigs identification and correction by Pore-C or Hi-C contacts. We also developed a chimeric contig identification and correction submodule that uses paired contact of intra-contig to calculate the depth in 500 bp window size, similar to existing tools, such as SALSA2[1] or HapHiC[2]. We first filter out the contigs with a length of less than 50 kb and an average depth of less than 10. Furthermore, we use the Mean Squared Error (MSE) to estimate the smoothness of the depth distribution and filter out contigs with MSE < 1, which are mostly repeats where it is difficult to identify misjoins by contacts.

$$MSE = \frac{1}{n}\sum_{i=0}^{n-1}(y_{i+1} - y_i)^2 \tag{1}$$

where n is the number of bins in a contig and y is the depth of the bins.

Then, we used the Polynomial Curve Fitting method to fit the depth distribution with the degree=50 parameter. Finding peaks in the signal area was used to identify the breakpoints of a chimeric and the peak points adjacent to the breakpoints. Peak points were used to ensure that the trough points correspond to breakpoints of the chimeric contigs, and we only retained the trough points whose corresponding values are more significant than two peak pairs $1.5E$ times, where E is the value of the distance of the breakpoint to an edge. Meanwhile, to avoid truncating the telomere regions, we filter out the break contigs that enrich the telomere motif, such as ”5’-CCCATT” for Human. Finally, to avoid another round of mapping, we convert the contig name and coordinate to the corrected name and coordinate in the Pore-C table and pairs file.

Identify the HCRs by the depth distribution of Pore-C or Hi-C data. Moreover, we also support identifying the HCRs by contacts based on Pore-C or Hi-C data, in which we calculate the distribution of all contacts (including MAPQ=0) at 10 kb resolution and normalize the contacts by the count of the restriction enzyme (RE) cut site to avoid bias from the RE. Following the normalization of the distribution of contacts, we used the peak value $Peak$ as the primary depth of the contacts and removed the bin that contained contacts greater than $1.5 * Peak$. The remaining bins were merged as the HCRs for the subsequent hypergraph filtering step.

1.1.2 Methalign: Utilize information from allele-specific 5mC sites to improve the mapping accuracy

The Methalign algorithm comprises a pipeline from reads to corrected alignments. Initially, HiFi reads, produced on the PacBio Revio system, are aligned to contigs using pbmm2 (https://github.com/PacificBiosciences/pbmm2, version 1.13.1). These alignments are then sorted by coordinate with SAMtools (version 1.19.2). The commands for these procedures are as follows:

```
$ pbmm2 index --preset CCS contigs.fasta index.mmi
$ pbmm2 align --preset CCS index.mmi <HiFi_reads.bam | samtools view - -b -o
    HiFi.align.bam
$ samtools sort HiFi.align.bam -o HiFi.align.sorted.bam
$ samtools index HiFi.align.sorted.bam
```

Following this, 5mC sites on each contig are identified using pb-CpG-tools (https://github.com/PacificBiosciences/pb-CpG-tools, version 2.3.2) and recorded in a BED file named 'aligned_bam_to _cpg_scores.combined.bed':

```
$ aligned_bam_to_cpg_scores --bam HiFi.align.sorted.bam --ref contigs.fasta --
    model pileup_calling_model.v1.tflite --modsites-mode reference
```

For ONT reads, the dorado basecaller (https://github.com/nanoporetech/dorado, version 0.3.2) identifies methylated CpGs during the base calling process:

```
$ dorado basecaller dna_r10.4.1_e8.2_400bps_sup@v3.5.2 pod5_dir --min-qscore 10
     --modified-bases 5mCG --device cuda:auto
```

These reads are then aligned to contigs using the dorado aligner (version 0.5.2+7969fab):

```
$ dorado aligner contigs.fasta ont_reads.bam --mm2-opts "--secondary=yes" > ont
    .align.bam
```

Finally, the custom script 'filter_bam_methyl.py' generates the final BAM file:

```
$ filter_bam_methyl.py contigs.fasta aligned_bam_to_cpg_scores.combined.bed ont
    .align.bam --penalty penalty
```

Each of the corresponding primary and supplementary alignments is found for each secondary alignment based on their positions on ONT reads. The number of inconsistent 5mC sites between the Pore-C read and the contig in each alignment is counted to adjust the alignment score using a specified penalty with the formula:

$$AS_{adj} = AS_{raw} - N \times P \tag{2}$$

In this formula, AS_{raw} and AS_{adj} represent the raw alignment score calculated by the dorado aligner and the adjusted alignment score, respectively. N denotes the number of inconsistent 5mC modified sites, and P is the specified penalty. If an updated alignment score of a secondary alignment exceeds that of the corresponding primary or supplementary alignment, their alignment types are swapped. The MAPQ of this alignment is then assigned a higher value to prevent it from being filtered out in subsequent analyses.

1.1.3 Rename: Rename chromosome according to a reference genome

To facilitate a comparative analysis of assemblies generated by different scaffolding software, we implemented a scaffold renaming function based on homology alignment with a closely related diploid reference genome. The workflow proceeds as follows:

(1) Contigs alignment: The contigs were aligned to the reference genome using wfmash (v0.17.0-0-g78bff59, https://github.com/waveygang/wfmash) with "-m -s 50k -l 250k -p 90 " parameters.

(2) Chromosomal assignment. For each scaffold, we calculated the cumulative matches of alignments against individual reference chromosomes. The reference chromosome with the maximum aggregate matches was selected to rename the scaffold and reorient it as step (3).

(3) Scaffold reorientation. We calculated the orientation score SO of each scaffold using the formula:

$$SO = \sum_{i,r} O_{ri} M_{ri} O_i \tag{3}$$

where SO quantifies the concordance between the scaffold and the reference chromosome. If $SO < 0$, it indicates an opposite orientation, and the entire orientation of the scaffold is reversed. The O_{ri} represents the direction between contig i and its matched regions. The O_{ri} is set to -1 if contig i is aligned in the opposite direction than the reference and 1 if it is aligned in the same direction. The M_{ri} represents the number of matches between contig i and the reference chromosome. The O_i represents the contig i orientation on the scaffold; it is set to -1 if it is oriented in the opposite direction than the source contig on the scaffold, and 1 if it is oriented in the same direction.

1.2 Simulation of polyploid genomes and corresponding Pore-C, Hi-C and ONT ultra-long data

1.2.1 Simulation of Pseudo-genomes

To benchmark the performance of our chimeric contig correction and haplotype phasing algorithms, we simulated five pseudo-polyploid genomes at different ploidy (2, 4, 6, 8, 12). Supplementary Table 8 shows that we download 12 accessions *Arabidopsis* variants information from 1001 Genome Projects[3] and their reference TAIR10 genome from TAIR (https://www.arabidopsis.org/).

And then, we used “BCFtools consensus” to generate pseudo-genomes:

```
$ bcftools consensus -f tair10.fasta sample.vcf.gz > sample.pseudo.fasta
```

To simulate pseudo-polyploid genomes at different ploidy levels, we merged the pseudo-genome data, as shown in Supplementary Table 1. Finally, we simulated contigs at different contig N50 (50 kb, 100 kb, 500 kb, 1 Mb and 2 Mb).

```
$ python CPhasing/scripts/simulation/simuCTG.py -n n50 -i ploidy.pseudo.fasta -o
    output.contig.fasta
```

1.2.2 Simulation of Pore-C data

To simulate Pore-C data for the pseudo-genomes, we first downloaded the Pore-C data of the *A. thaliana* Col-0 variant from the Genome Sequence Archive (GSA; CRA0051051) and aligned it to the TAIR10 reference genome. Low-quality alignments (mapping quality ≤ 1) were filtered out using the ”cphasing-rs paf2porec”, followed by the removal of inter-chromosomal contacts (Supplementary Figure 7):

```
$ python ~/code/CPhasing/scripts/similation/remove_inter_contact.py sample.porec.gz
```

We then projected the alignments onto different accessions by incorporating their respective variant information, enabling the reconstruction of simulated Pore-C reads that reflect haplotype-specific contacts, with contact fragments grouped and ligated according to read IDs.

```
$ cphasing-rs simulator porec tair10.fasta ${sample}.vcf.gz sample.porec.DpnII.bed -
    o ${sample}.porec.fasta
```

To realistically mimic ONT sequencing characteristics, we introduced sequencing errors using PBSIM3[4] with the ERRHMM-ONT model.

```
$ pbsim --strategy templ --template sample.porec.fasta --method errhmm --errhmm data
    /ERRHMM-ONT.model --prefix output --accuracy-mean 0.95
```

1.2.3 Simulation of Hi-C data

Following a procedure analogous to the Pore-C simulation, we downloaded Hi-C data for the Arabidopsis thaliana Col-0 variant from NCBI BioSample (SRR2626163) and aligned it to the TAIR10 reference genome. After removing low-quality alignments (MAPQ < 1) using samtools view -q 1, we converted the alignments to other accessions using their variant profiles and excluded all inter-chromosomal read pairs.

```
$ cphasing-rs simulator hic tair10.fasta ${sample}.vcf.gz -o ${sample}
```

To ensure consistency, we downsampled the simulated paired-end reads to 50× coverage, matching the conditions of the Pore-C simulation.

1.2.4 Simulation of Ultra-long ONT data

To assess the performance of hitig, we simulated approximately 30x ultra-long data across different ploidy levels using assemblies with a contig N50 of 500 kb.

```
$ pbsim --strategy wgs --genome ploidy.fasta --depth 30 --length-mean 80000 --length
    -sd 80000 --accuracy-mean 0.95 --method errhmm --errhmm ~/software/pbsim3-3.0.0/
    data/ERRHMM-ONT.model --prefix ploidy.ul.fastq
```

1.2.5 Simulation of chimeric contigs

We simulated two types of chimeric errors, namely, intra-chromosomal and inter-chromosomal switches. To identify homologous regions, we performed self-alignment of the contig-level assembly using minimap2 [5] with the parameters ”-DP -k19 -w19 -m200”. Subsequently, approximately 20% of chimeric contigs were introduced using the following script:

```
$ simulate_chimeric.py ploidy-${ploidy_level}.500k.fasta --chimeric_ratio 0.2 >
    chimeric.fasta
```

1.3 Performance of C-Phasing on monoploid genome scaffolding

In this section, we demonstrate the ability of C-Phasing to scaffold monoploids based on the Hyperpartition (basal) algorithm. We collected the public Pore-C data of rice (*Oryza sativa* Azucena, SRR13985193) and its chromosome-level assembly (GCA_009830595.1), breaking the chromosomes to simulate contig-level assembly at different contig N50 (100 kb, 500 kb, 1 Mb, 2 Mb, and 5 Mb).

```
$ cphasing pipeline -f sample.contigs.fasta -pct sample.porec.gz -n group_number --
    mode basal
```

We then utilized these assemblies to benchmark the performance of our algorithm in monoploid scaffolding. As shown in Supplementary Table 6, Hyperpartition in its basal mode consistently achieved the highest contiguity across various contig lengths compared to other tools. Furthermore, it produced fewer interchromosomal misassemblies, demonstrating superior accuracy and robustness in monoploid genome reconstruction.

Furthermore, we assessed the performance of Hyperpartition (basal mode) on real datasets from *Metrosideros polymorpha* (PRJNA670777)[6] and *Gossypium hirsutum* (PRJNA824233)[7]. For this evaluation, the assemblies were fragmented at gap regions, and publicly available Pore-C data were used to guide the scaffolding process. As shown in Supplementary Table 7, Hyperpartition achieved the highest contiguity and anchor rate for M. polymorpha and *G. hirsutum*, demonstrating its effectiveness on real-world genomes. Overall, our Hyperpartition algorithm in basal mode outperforms existing tools, demonstrating robust performance in scaffolding both monoploid and allopolyploid genomes.

1.4 Detail commands of comparison with existing software

To evaluate our algorithm, we compared it with existing tools.

1.4.1 Chimeric contigs identification and correction on simulated data

a. C-Phasing

```
$ hitig pipeline -f chimeric.fasta -i ul.fastq.gz -n ploidy_level -t 20
```

b. CRAQ

```
$ craq -g chimeric.fasta -sms ul.fastq.gz -b T -t 20 -x map-ont
```

c. tigmint

```
$ tigmint-make tigmint-long draft=chimeric reads=ul dist=auto G=genome_sizes t=20
```

d. GAEP

```
$ gaep bkp -r chimeric.fasta -i ul.fastq.gz -t 20 -x ont
```

1.4.2 Polyploid phasing and scaffolding on simulated polyploids

a. C-Phasing

```
$ cphasing pipeline -f sample.contigs.fasta -prs sample.pairs.gz -n
    homo_group_number:ploidy_level -t 10 -x nofilter
```

b. HapHiC

```
$ haphic pipeline sample.contigs.fasta sample.bam group_number --
    remove_allelic_links ploidy_level --processes 10
```

c. ALLHiC

```
$ gmap_build -D . -db DB sample.contigs.fasta
$ gmap -d DB -D . -f 2 -n ploidy_level -t 10 monoploid.cds > gmap.gff3
$ gmap2AlleleTableBED.pl monoploid.bed
$ ALLHiC_prune -i Allele.ctg.table -b sample.bam
$ allhic extract prunning.bam sample.contigs.fasta --RE enzyme_site
$ allhic partition prunning.counts_enzyme_site.txt prunning.pairs.txt
    group_numbers --minREs 25
$ for txt in *.counts_enzyme_site.group_number.g*.txt; do
   echo "allhic optimize $txt prunning.clm"
   done | parallel -j 10
```

1.4.3 Polyploid phasing and scaffolding on real polyploid data

To further evaluate the performance of our pipeline compared to HapHiC, we processed the Hi-C data following the instructions provided in their respective tutorials. For polyploid datasets, C-Phasing was executed with ”-hcr -p AAGCTT -t 10” across all samples, and with ”–disable-merge-in-first –refine” specifically for sweet potato assemblies, and with ”-n 0:0 -x very-sensitive” specifically for cultivated sugarcane ZZ01.

1.5 Performance of C-Phasing pipeline on polyploidy phasing scaffolding

We evaluated the C-Phasing pipeline (default using the Align module for Pore-C) on a ‘synthetic’ dodecaploid genome with contig N50 values of 500 kb and 2 Mb. Using Hi-C data, C-Phasing achieved strong concordance with the reference (median Spearman’s ρ = 1.00 and 1.00, respectively), greater than HapHiC (median ρ = 0.97 and 0.98) and markedly outperforming ALLHiC (median ρ = 0.83 and 0.88) (Supplementary Fig. 12). Notably, with Pore-C data, C-Phasing’s performance further improved, achieving near-perfect phasing and scaffolding (ρ = 1.00 at 500 kb; 1.00 at 2 Mb), despite structural artifacts such as misorientation between the two arms of chromosome 1. These results underscore C-Phasing’s robustness and accuracy in assembling ultra-complex polyploid genomes at haplotype resolution.

References

[1] Jay Ghurye, Arang Rhie, Brian P. Walenz, Anthony Schmitt, Siddarth Selvaraj, Mihai Pop, Adam M. Phillippy, and Sergey Koren. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLOS Computational Biology*, 15(8):e1007273, August 2019.

[2] Xiaofei Zeng, Zili Yi, Xingtan Zhang, Yuhui Du, Yu Li, Zhiqing Zhou, Sijie Chen, Huijie Zhao, Sai Yang, Yibin Wang, and Guoan Chen. Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nature Plants*, pages 1–17, August 2024.

[3] Carlos Alonso-Blanco, Jorge Andrade, Claude Becker, Felix Bemm, Joy Bergelson, Karsten M. Borgwardt, Jun Cao, Eunyoung Chae, Todd M. Dezwaan, Wei Ding, Joseph R. Ecker, Moises Exposito-Alonso, Ashley Farlow, Joffrey Fitz, Xiangchao Gan, Dominik G. Grimm, Angela M. Hancock, Stefan R. Henz, Svante Holm, Matthew Horton, Mike Jarsulic, Randall A. Kerstetter, Arthur Korte, Pamela Korte, Christa Lanz, Cheng-Ruei Lee, Dazhe Meng, Todd P. Michael, Richard Mott, Ni Wayan Muliyati, Thomas Nägele, Matthias Nagler, Viktoria Nizhynska, Magnus Nordborg, Polina Yu Novikova, F. Xavier Picó, Alexander Platzer, Fernando A. Rabanal, Alex Rodriguez, Beth A. Rowan, Patrice A. Salomé, Karl J. Schmid, Robert J. Schmitz, Ümit Seren, Felice Gianluca Sperone, Mitchell Sudkamp, Hannes Svardal, Matt M. Tanzer, Donald Todd,

Samuel L. Volchenboum, Congmao Wang, George Wang, Xi Wang, Wolfram Weckwerth, Detlef Weigel, and Xuefeng Zhou. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*, 166(2):481–491, July 2016. Publisher: Elsevier.

[4] Yukiteru Ono, Michiaki Hamada, and Kiyoshi Asai. PBSIM3: a simulator for all types of PacBio and ONT long reads. *NAR Genomics and Bioinformatics*, 4(4):lqac092, 12 2022.

[5] Heng Li. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100, September 2018.

[6] Jae Young Choi, Xiaoguang Dai, Ornob Alam, Julie Z. Peng, Priyesh Rughani, Scott Hickey, Eoghan Harrington, Sissel Juul, Julien F. Ayroles, Michael D. Purugganan, and Elizabeth A. Stacy. Ancestral polymorphisms shape the adaptive radiation of Metrosideros across the Hawaiian Islands. *Proceedings of the National Academy of Sciences*, 118(37):e2023801118, September 2021.

[7] Xianhui Huang, Xuehan Tian, Liuling Pei, Xuanxuan Luo, Yuqi Zhang, Xingtan Zhang, Xianlong Zhang, Longfu Zhu, and Maojun Wang. Multi-omics mapping of chromatin interaction resolves the fine hierarchy of 3D genome in allotetraploid cotton. *Plant Biotechnology Journal*, 20(9):1639–1641, 2022.

2 Supplementary Tables

Supplementary Table 1 | Statistics of the Hi-C mapping and their coverage

Items			***Sorghum bicolor***		***Saccharum spontaneum***		***Saccharum officinarum* L.**		***Saccharum* hybrid cultivar POJ2878**[1]	
			Number	**Percentage (%)**	**Number**	**Percentage (%)**	**Number**	**Percentage (%)**	**Number**	**Percentage (%)**
Mapping	Sequencing reads		161,155,460	100	639,213,589	100	1,329,647,044	100	4,191,942,972	100
	valid reads	unique	79,842,347	49.54	133,671,824	20.91	110,702,433	8.36	332,133,293	7.92
		deduped	70,861,426	43.97	128,086,634	20.04	109,095,283	8.2	301,956,362	7.2
		trans links	47,398,395	29.41	83,822,929	13.11	75,969,043	5.71	160,565,351	3.83
	invalid reads		81,313,113	50.46	505,541,765	79.09	1,218,944,611	91.67	3,859,809,679	92.07
	# of contigs		4,941		11,009		28,905		27,331	
	contig N50 (kb)		1,179.13		1,781.63		2,943.09		6,418.52	
	Size of contigs (Mb)		698.54		3,192.30		7,322.72		10,109.42	
			Number	**Percentage (%)**	**Number**	**Percentage (%)**	**Number**	**Percentage (%)**	**Number**	**Percentage (%)**
Contigs	Contigs covered by valid RE[2]		2,012	40.72	2,595	23.57	1,897	6.56	1,335	4.88
	Size of contigs covered by valid RE (Mb)		640.3	91.66	2,908.05	91.09	5,055.45	69.03	6,577.37	65.06

[1] The POJ2878 genome is a contig-level assembly generated using HiFi and ultra-long (UL) reads by hifiasm, with an ultra-long read N50 that differs from the ultra-long read length distribution reported in our separate study of the POJ2878 genome.

[2] We define a valid restriction enzyme (RE) site as one with at least three inter-contig links within a ±500 bp window of the site. Only contigs containing at least 25 valid RE sites were valid for subsequent accuracy clustering.

Supplementary Table 2 | Statistics of the assemblies by different assemblers

Sample	**Software**	**# of contigs**	**Length (Gb)**	**N50 (Mb)**	**Chimeric error**	**Erroneous**	**Duplicated**	**Haplotig**	**Collapsed**
alfalfa	hicanu (HiFi-only)	22,876	3.65	1.90	0.77%	7.41%	12.58%	77.34%	2.67%
	verkko (HiFi-UL)	19,040	3.32	1.93	0.26%	1.78%	10.79%	84.54%	2.89%
	hifiasm (HiFi-only)	24,042	3.54	1.24	0.11%	1.38%	17.24%	79.37%	2.01%
	hifiasm (HiFi-UL)	6,312	3.18	4.28	0.13%	1.07%	9.60%	86.49%	2.84%
Sweet potato	hicanu (HiFi-only)	26,703	3.29	2.37	1.20%	8.64%	6.88%	80.80%	3.68%
	verkko (HiFi-UL)	12,100	2.92	5.94	0.28%	1.04%	3.62%	92.29%	3.05%
	hifiasm (HiFi-only)	33,502	3.30	1.13	0.11%	0.82%	13.52%	83.17%	2.49%
	hifiasm (HiFi-UL)	11,041	2.87	3.24	0.14%	0.34%	4.08%	92.73%	2.85%
POJ2878	hicanu (HiFi-only)	31,273	8.79	8.16	2.37%	0.59%	2.14%	80.43%	16.83%
	verkko (HiFi-UL)	49,868	10.79	3.06	0.39%	0.94%	6.00%	88.06%	5.00%
	hifiasm (HiFi-only)	140,205	12.28	0.35	0.23%	0.34%	15.51%	79.86%	4.29%
	hifiasm (HiFi-UL)	27,853	10.13	6.33	0.25%	0.12%	3.25%	90.79%	5.84%

Note: The HiFi reads for sweet potato in this analysis were generated using the PacBio Sequel platform, differing from those used in our downstream contig-level assemblies for benchmarking. Additionally, the ultra-long read N50 for the POJ2878 dataset was approximately 100 kb, which differs from the ultra-long read distributions reported in our separate study of POJ2878.

Supplementary Table 3 | The sequencing output and cost of ePore-C

Source	Accession	Sample	Enzyme	Platform	ONT chip	Read count (M)	Yield (Gb)	Read N50 (kb)	Pairwise contacts (M)	Yield pairwise contacts per Gb (M)	Singleton (%)	Cost* ($)	Cost ($) / Gb	Cost ($) / # of pairwise contacts (M)
Li, Z. et al,. 2022	Col-0	Rep 1	DpnII	MinION	R9.4.1	3.50	7.91	3.01	9.22	1.17	27.93	791	100.00	85.83
		Rep 2				6.45	11.69	2.65	18.41	1.57	25.33	1,169	100.00	63.51
Huang, X, et al,. 2022	TM-1	TM-1_part1		PromethION		36.08	61.25	3.05	63.98	1.04	43.67	1,268	20.70	19.82
		TM-1_part2				36.08	51.99	2.24	42.46	0.82	51.71	1268	24.39	29.86
Serra Mari, R, et al,. 2024	Altus	Altus-202304			R10.4.1	14.72	24.54	2.35	2.65	0.11	77.13	1268	51.67	478.28
		Altus-202308				59.28	33.17	0.63	14.96	0.45	58.14	1,268	38.23	84.73
Data generated in this study (ePore-C)	Col-0	AT-1	DpnII			5.55	14.45	3.45	49.80	3.45	11.38	566	20.41	6.44
		AT-2				5.22	13.29	3.30	38.14	2.87	15.74			
	1393	SP-1	HindIII			25.80	110.97	6.05	43.24	0.39	43.98	3,165	9.04	22.93
		SP-2				26.53	142.03	7.22	61.91	0.44	36.09			
		SP-3				29.61	96.95	4.66	32.88	0.34	52.73			
	82-114	82-114-1				29.36	108.25	4.60	31.06	0.29	49.64	3,165	10.63	32.12
		82-114-2				23.85	95.04	5.27	34.27	0.36	45.50			
		82-114-3				23.96	94.59	5.21	33.22	0.35	45.69			
	Zhong Zhe No. 1 (ZZ01)	ZZ01-1				22.15	93.22	5.80	29.00	0.31	48.35	8,530	10.46	28.26
		ZZ01-2				24.68	90.25	4.43	20.01	0.22	56.25			
		ZZ01-3				28.48	107.04	4.43	34.39	0.32	43.42			
		ZZ01-4				26.23	98.16	4.21	35.97	0.37	39.47			
		ZZ01-5				13.19	84.77	8.58	29.74	0.35	33.79			
		ZZ01-6				24.20	102.98	5.73	30.65	0.30	44.19			
		ZZ01-7				46.31	138.45	5.32	63.46	0.46	53.00			
		ZZ01-8				20.75	102.55	9.67	58.57	0.57	43.55			
		ZZ01-DpnII-1	DpnII			30.10	87.06	3.86	313.63	3.60	9.18	2,232	13.35	3.71
		ZZ01-DpnII-2				28.67	80.13	3.83	288.42	3.60	10.50			

Note(*): The cost of Pore-C in *Arabidopsis* was referenced from a previous study (Li et al,. 2022). All other costs were estimated using a standardized approach, with library preparation ranging from $272.11 to (HindIII) to $303.67 (DpnII), and sequencing costs totaling $964.19. In this study, cost estimates were based on preparing and sequencing 2-3 cells per run.

Supplementary Table 4 | Pseudocode of hypergraph construction

```
Algorithm 1: C-Phasing Hypergraph Construction
Input: Pore-C Table
Output: Incidence matrix $H$
for each concatemer $j$ do
    if fragment locus in contig $i$ then
        if number of contigs $2 \leq n_j < 50$ and alignment length $l_i \geq 150$ then
            $H(i, j) = 1$
        else
            $H(i, j) = 0$
        end
    else
        $H(i, j) = 0$
    end
end
```

Supplementary Table 5 | Pseudocode of HyperPartition (basal)

```
Algorithm 2: Basal mode of HyperPartition
Input: Hypergraph incidence matrix $H$, Group number $n$
Output: Cluster assignments $C$
// Compute adjacency matrix
$A = H(D_e - I)^{-1}H^{\tau}$;
$A = zero\_diag(A)$;
$C = LOUVAIN\_ALGORITHM(A)$;
$c = length(C)$;
if $c < n$ then
    $C = MERGE\_ALGORITHM(C)$;
    // Merge groups iteratively according to the edge weights
else
    $C$
end
```

Supplementary Table 6 | Performance of HyperPartition (basal) on simulation contigs of *Oryza sativa* Azucena by Pore-C.

Contig N50	Software	% AR	#G / #C	% Contiguity	% IE	Precision	Recall	F1 score
100 kb	C-Phasing	99.44	12/12	99.94	0.05	1.00	1.00	1.00
	ALLHiC	99.73	12/12	99.30	0.83	0.99	0.99	0.99
	YAHS	94.06	791/12	94.87	11.12	0.98	0.74	0.77
	HapHiC	94.32	12/12	93.31	7.31	0.95	0.88	0.90
500 kb	C-Phasing	99.49	12/12	100.00	0.00	1.00	1.00	1.00
	ALLHiC	99.70	12/12	99.45	0.68	0.99	0.99	0.99
	YAHS	99.70	40/12	97.69	20.81	0.89	0.77	0.71
	HapHiC	88.39	12/12	89.98	13.24	0.92	0.71	0.78
1 Mb	C-Phasing	99.49	12/12	99.95	0.04	1.00	1.00	1.00
	ALLHiC	99.63	12/12	96.92	4.20	0.98	0.96	0.97
	YAHS	99.81	20/12	99.96	13.86	0.84	0.83	0.79
	HapHiC	92.31	12/12	90.64	9.62	0.96	0.83	0.87
2 Mb	C-Phasing	99.49	12/12	100.00	0.00	1.00	1.00	1.00
	ALLHiC	100.00	12/12	93.24	7.47	0.96	0.93	0.94
	YAHS	99.92	11/12	100.00	19.96	0.63	0.67	0.64
	HapHiC	91.53	12/12	87.86	15.36	0.89	0.71	0.77
5 Mb	C-Phasing	100.00	12/12	100.00	0.00	1.00	1.00	1.00
	ALLHiC	100.00	12/12	100.00	0.00	1.00	1.00	1.00
	YAHS	99.95	10/12	100.00	19.95	0.58	0.58	0.58
	HapHiC	96.40	12/12	89.77	9.84	0.96	0.87	0.89

Note: AR, Anchor rate; G, Group; C, Chromosome; IE, Interchromosomal error. The scaffolds or contigs from the initial assembly were excluded for evaluation.

Supplementary Table 7 | Performance of HyperPartition (basal) on public haploid assemblies and Pore-C data.

Species	Software	% AR	#G/#C	% Contiguity	% IE	Precision	Recall	F_1
M. polymorpha	C-Phasing	**100.00**	11/11	**99.44**	0.53	0.99	**0.99**	**0.99**
	ALLHiC	99.63	11/11	98.10	8.31	0.99	0.98	0.98
	YAHS	97.24	217/11	98.95	**0.36**	**1.00**	0.96	0.98
	HapHiC	81.92	11/11	95.94	14.62	0.90	0.67	0.75
G. hirsutum	C-Phasing	**99.99**	26/26	**99.94**	**0.06**	**1.00**	**1.00**	**1.00**
	ALLHiC	99.80	26/26	92.38	8.88	0.95	0.87	0.90
	YAHS	95.70	46/26	99.16	17.47	0.95	0.76	0.75
	HapHiC	99.44	26/26	98.62	1.76	0.98	0.98	0.98

Note: AR, Anchor rate; G, Group; C, Chromosome; IE, Interchromosomal error. The scaffolds or contigs from the initial assembly were excluded for evaluation. Furthermore, the %IE of YAHS in the *M. polymorpha* genome was the lowest, which was 0.36. Because the group numbers of YAHS are higher than those of other software, which will significantly decrease the % IE calculation.

Supplementary Table 8 | Synthetic polyploid simulation based on different ecotypes of *Arabidopsis*.

Assession	Ecotype ID	# SNPs	# Indels	ploidy				
				2	4	6	8	12
An-1	6898	421,863	29,793		A	A	A	A
Cvi-0	6911	684,310	62,458		B	B	B	B
Kyoto	7207	439,345	30,844			C	C	C
Ler-1	6932	597,706	60,884			D	D	D
Altenb-2	9970	164,735	3,907	E	E	E	E	E
TAL-07	6180	150,846	13,745	F	F	F	F	F
Got-22	6920	541,987	47,266				G	G
Ms-0	6938	474,252	33,848				H	H
Bor-1	5837	418,323	30,809					I
Hs-0	7162	431,195	28,549					J
Cdm-0	9943	422,633	12,397					K
LL-0	6933	512,355	34,970					L

Supplementary Table 9 | Average normalized h-*trans* error between pseudo genomes

	An-1	Cvi-0	Kyoto	Ler-1	Altenb-2	TAL-07	Got-22	Ms-0	Bor-1	Hs-0	Cdm-0	LL-0
An-1												
Cvi-0	0.3940											
Kyoto	0.5282	0.3877										
Ler-1	0.5033	0.3751	0.5143									
Altenb-2	0.6215	0.4474	0.6051	0.5688								
TAL-07	0.6526	0.4664	0.6313	0.6159	0.8231							
Got-22	0.5181	0.3822	0.4970	0.4919	0.5755	0.6041						
Ms-0	0.4984	0.3697	0.5090	0.4849	0.5659	0.5889	0.4640					
Bor-1	0.5331	0.3894	0.5616	0.5159	0.6174	0.6420	0.4997	0.5093				
Hs-0	0.5635	0.3941	0.5317	0.5060	0.6222	0.6503	0.5292	0.4990	0.5348			
Cdm-0	0.5137	0.3737	0.4927	0.4641	0.5923	0.6067	0.4760	0.4589	0.4980	0.5059		
LL-0	0.5038	0.3774	0.4934	0.4710	0.5683	0.5931	0.4852	0.4621	0.4958	0.4987	0.4885	

Supplementary Table 10 | Assembly statistics for simulated polyploid genomes at contig-level

Ploidy	# of contigs	Genome size (bp)	Contig N50 (bp)	Minimum length (bp)	Average length (bp)	Maximum length (bp)	Q1 (bp)	Q2 (bp)	Q3 (bp)
2	4,471	238,298,347	48,616	6,794	53,298.70	4,933,106	21,633.50	32,176	43,746
2	2,466	238,298,347	96,363	10,002	96,633.60	4,957,432	36,375	62,459.50	88,199
2	563	238,298,347	483,495	10,084	423,265.30	4,915,373	144,612.50	297,993	455,831
2	293	238,298,347	993,289	11,168	813,304.90	4,881,612	289,070	585,017	929,864
2	162	238,298,347	1,945,088	13,293	1,470,977.50	4,909,547	594,949	1,302,699	1,943,540
4	9,084	476,605,983	48,360	6,121	52,466.50	4,959,454	21,028.50	32,840.50	44,410
4	4,865	476,605,983	97,099	898	97,966.30	4,989,449	36,784	62,247	88,211
4	1,099	476,605,983	484,653	12,158	433,672.40	4,917,970	156,546.50	294,649	444,035.50
4	593	476,605,983	976,713	14,846	803,720	4,956,264	324,286	633,401	934,641
4	322	476,605,983	1,948,762	15,599	1,480,142.80	4,936,277	646,564	1,206,224	1,936,073
6	13,518	714,925,126	48,381	393	52,886.90	4,985,818	21,745	33,197	44,750
6	7,423	714,925,126	96,813	556	96,312.20	4,986,816	36,462.50	62,285	88,297
6	1,658	714,925,126	482,513	10,209	431,197.30	4,977,688	161,780	304,543.50	449,900
6	865	714,925,126	973,746	10,222	826,503	4,957,760	344,694	670,828	932,460
6	461	714,925,126	1,932,880	22,186	1,550,813.70	4,997,430	709,540	1,469,641	1,943,279
8	17,949	953,235,842	48,501	58	53,108	4,858,648	21,175	32,723	44,078
8	9,913	953,235,842	96,592	5,129	96,160.20	4,886,758	36,296	62,248	88,289
8	2,175	953,235,842	487,126	1,678	438,269.40	4,993,863	158,269	306,642	450,715.50
8	1,188	953,235,842	978,786	11,178	802,387.10	4,935,403	293,158	594,134	919,943.50
8	623	953,235,842	1,928,734	15,445	1,530,073.60	4,983,117	689,125.50	1,355,870	1,926,144.50
12	27,576	1,429,847,861	48,297	231	51,851.20	4,998,911	21,506.50	33,093	44,807.50
12	14,867	1,429,847,861	96,501	1,364	96,176	4,995,445	36,410.50	62,171	88,338
12	3,293	1,429,847,861	482,280	10,091	434,208.30	4,993,829	166,057	309,654	452,939
12	1,706	1,429,847,861	975,919	3,377	838,128.90	4,995,301	335,196	641,677.50	928,530
12	956	1,429,847,861	1,944,548	10,266	1,495,656.80	4,977,643	668,236	1,330,420	1,947,805.50

Supplementary Table 11 | Partitioning performance on simulated dodecaploid genomes

Ploidy level	Contig N50	Software	Whole genome			Haplotype E and F		
			Correct rate (%)	Mis-assignment rate (%)	Un-anchored rate (%)	Correct rate (%)	Mis-assignment rate (%)	Un-anchored rate (%)
12	500 kb	C-Phasing (Pore-C)	99.82	0.09	0.09	99.50	0.26	0.24
12	500 kb	C-Phasing (Hi-C)	98.75	0.18	1.06	94.73	1.10	4.17
12	500 kb	HapHiC	73.09	8.93	17.98	18.56	36.70	44.75
12	500 kb	ALLHiC	81.01	9.65	9.33	44.74	15.77	39.50
12	2 Mb	C-Phasing (Pore-C)	99.96	0.00	0.03	99.99	0.01	0.00
12	2 Mb	C-Phasing (Hi-C)	99.87	0.00	0.13	99.42	0.00	0.58
12	2 Mb	HapHiC	90.46	4.94	4.60	48.31	29.64	22.05
12	2 Mb	ALLHiC	91.22	2.68	6.09	81.75	0.00	18.25

Supplementary Table 12 | Statistics of 700-bp ONT read alignments in sweet potato

Type of alignments	Number
Total mapped	50,238,766
Mapped to a homologous chromosome	4,551,285
Having a correct secondary alignment	4,272,417
Same MAPQ as the secondary alignment	3,639,682
Same alignment score as the secondary alignment	3,456,352
Same identity as the secondary alignment	3,547,768
Same edit distance as the secondary alignment	3,455,516
Same inconsistent 5mC sites as the secondary alignment	2,260,454
Primary alignment MAPQ = 0	3,810,406
Secondary alignment MAPQ = 0	4,272,069

Supplementary Table 13 | Summary statistics of sequencing datasets used for benchmarking.

Sample	Data type	Size (Gb)	Coverage	Platform	Source
alfalfa Zhongmu-4	HiFi	136.06	43.89	PacBio Sequel	CNCB:PRJCA004062
	Ultra-long	74.66	24.08	Oxford Nanopore (R9.4)	CNCB:PRJCA031790
	Hi-C	284.82	91.88	Illumina NovaSeq	CNCB:PRJCA004062
	Pore-C	139.34	44.95	Oxford Nanopore (R10.4)	CNCB:PRJCA041059
sweet poatao 1393	HiFi	100.22	36.05	PacBio Revio	This study
	Ultra-long	115.03	41.38	Oxford Nanopore (R10.4)	This study
	Hi-C	451.73	162.49	Illumina HiSeq 1500	CNGBdb:CNP0004414
	Pore-C	349.95	125.88	Oxford Nanopore (R10.4)	This study
wild sugarcane 82-114	HiFi	230.84	33.07	PacBio Revio	This study
	Ultra-long	97.9	14.03	Oxford Nanopore (R10.4)	This study
	Hi-C	302.7	43.37	MGI DNBSEQ T7	This study
	Pore-C	297.88	42.68	Oxford Nanopore (R10.4)	This study

Supplementary Table 14 | Benchmarking of C-Phasing on real datasets

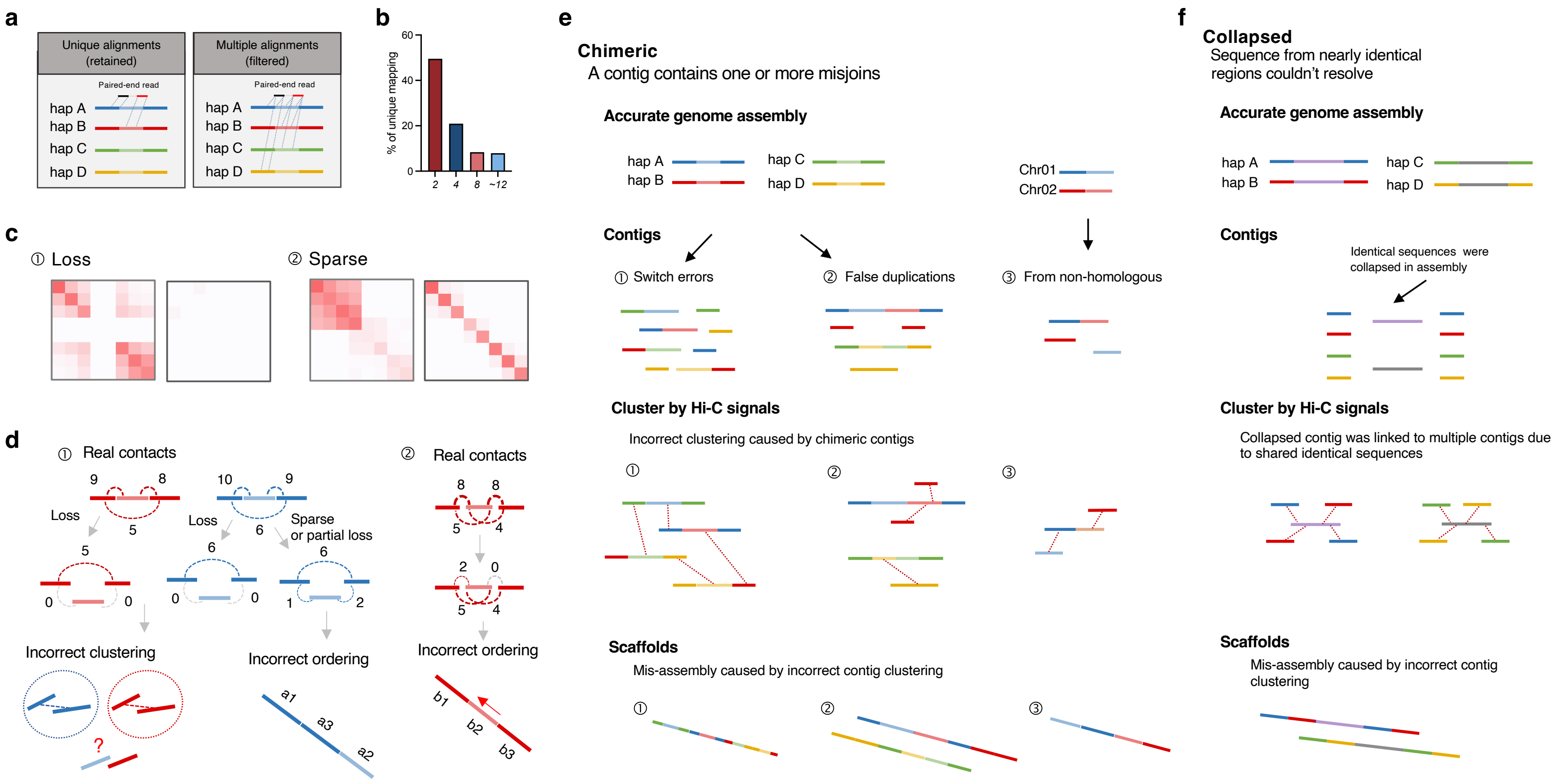
Species	Assemble strategy	Data	Software	Assembly size (Gb) / Estimate size (Gb)	# of contigs	Contig NG50 (Mb)	Anchored size (Mb)	Anchor rate (%)	Adjusted anchor rate (%)	# of Misjoined	% of Misjoined	Wall time (s)	Peak memory (Gb)
Alfalfa Zhongmu-4 (2n=4x=32)	HiFi-only	Pore-C	C-Phasing	3.54 / 3.10	24,042	1.54	2,847.73	80.55	91.86	107	0.47	471.04	10.08
		Hi-C	C-Phasing				2,754.58	77.92	88.86	49	0.53	1033.55	29.35
			HapHiC				2,745.03	77.65	88.55	127	0.72	1030.73	10.60
	HiFi-UL	Pore-C	C-Phasing	3.18 / 3.10	6,312	4.37	2,989.89	94.05	96.45	0	0.00	465.68	10.69
		Hi-C	C-Phasing				2,928.30	92.11	94.46	39	0.44	554.43	25.45
			HapHiC				2,903.29	91.33	93.65	88	1.00	925.46	9.25
Sweet potato 1393 (2n=6x=90)	HiFi-only	Pore-C	C-Phasing	3.26 / 2.78	27,220	1.62	2,505.69	76.86	90.13	524	1.38	915.73	20.46
		Hi-C	C-Phasing				2,366.98	72.60	85.14	179	0.98	1116.20	35.63
			HapHiC				2,318.44	71.11	83.40	404	2.25	1192.92	9.44
	HiFi-UL	Pore-C	C-Phasing	2.85 / 2.78	7,529	5.52	2,620.83	92.04	94.27	306	1.15	762.93	19.81
		Hi-C	C-Phasing				2,503.80	87.93	90.06	150	0.75	760.74	31.68
			HapHiC				2,438.27	85.63	87.71	156	1.39	1091.74	6.54
Wild sugarcane 82-114 (2n=10x=80)	HiFi-only	Pore-C	C-Phasing	7.45 / 6.98	20,332	5.52	6,811.67	91.39	97.59	153	0.44	765.52	17.64
		Hi-C	C-Phasing				6,681.72	89.64	95.73	65	0.52	639.22	13.90
			HapHiC				6,631.42	88.97	95.01	61	0.94	567.64	15.64
	HiFi-UL	Pore-C	C-Phasing	7.09 / 6.98	3,796	25.75	6,927.14	97.72	99.24	52	1.33	729.41	17.99
		Hi-C	C-Phasing				6,834.05	96.41	97.91	31	0.47	588.17	13.25
			HapHiC				6,817.21	96.17	97.67	50	2.15	514.66	14.68

Note: The contig NG50 was computed using the estimated genome size as the reference. The adjusted anchor rate is calculated by dividing the total length of anchored contigs by the estimated genome size. For instance, alfalfa (HiFi-only) anchored 2,847.73 Mb with Pore-C data out of an estimated 3,100 Mb genome, yielding an adjusted anchor rate of 91.86%.

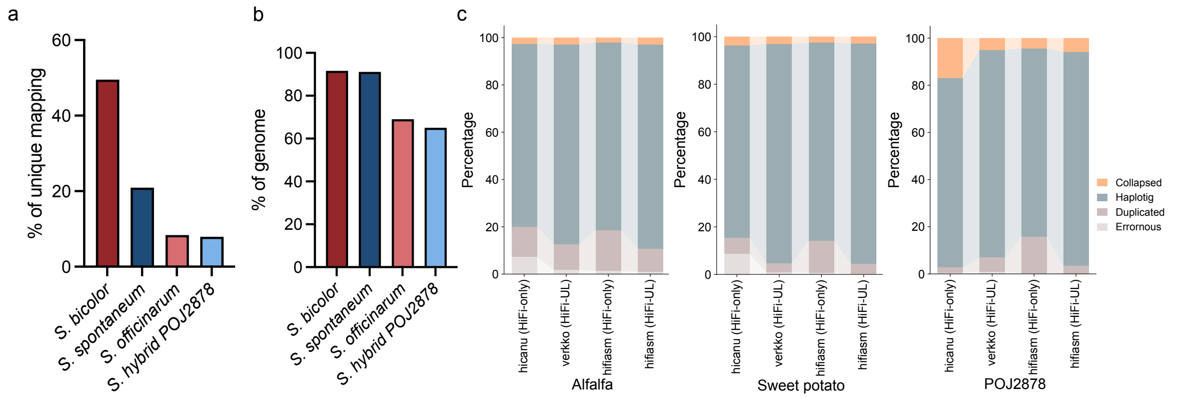
Supplementary Table 15 | Statistics of alignments and contacts before and after Methalign.

Items	HiFi-only			HiFi-UL		
	Align	HiFi-Methalign	ONT-Methalign	Align	HiFi-Methalign	ONT-Methalign
# of valid alignments (M)	41.13	47.21	46.57	42.18	48.16	47.72
% of valid alignments	65.34	75.01	73.98	67.00	76.50	75.81
# of valid reads (M)	10.30	12.11	11.91	10.58	12.40	12.27
% of valid reads	40.52	47.67	46.88	41.64	48.79	48.27
# of valid contacts (M)	37.77	46.14	45.53	39.30	48.08	47.65
Total length of weakly connected contigs (Mb)	423.40	63.73	61.82	61.65	12.08	12.26
Anchored length (Mb)	2,393.93	2,857.71	2,842.17	2,534.75	2,725.55	2,724.41
Adjusted anchor rate (%)	86.11	102.80	102.24	91.18	98.04	98.00

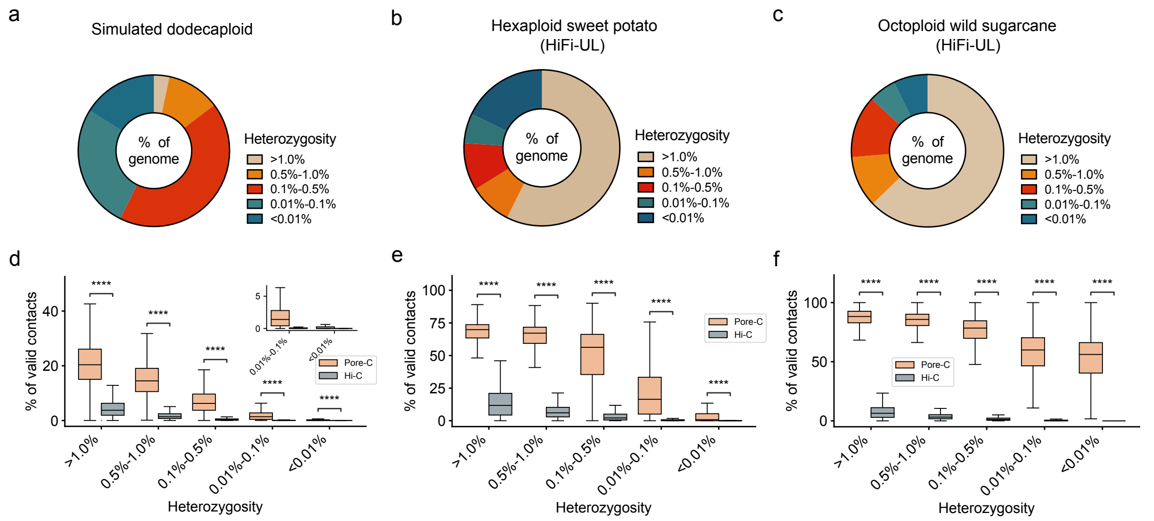
3 Supplementary Figures



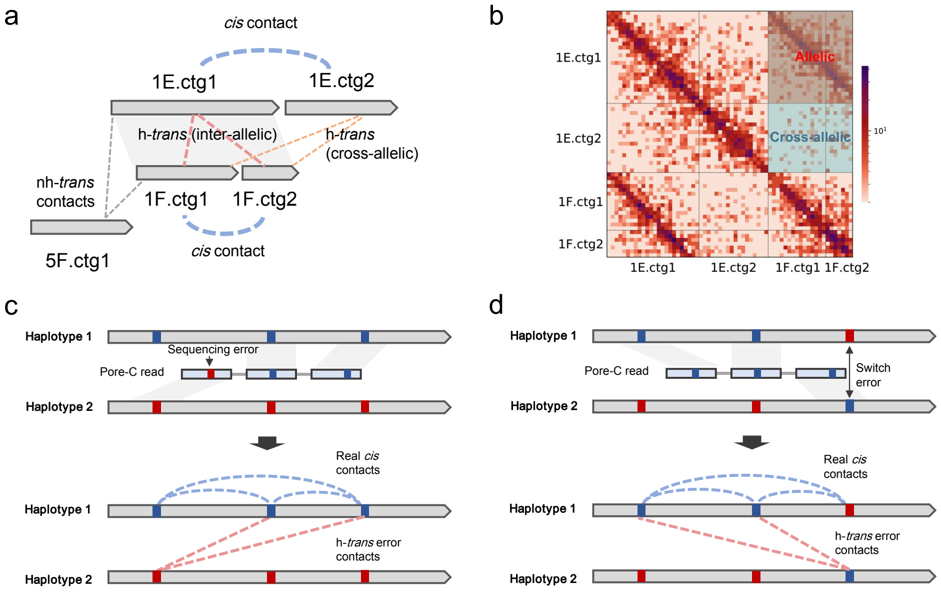
Supplementary Fig. 1 | Schematic illustrating the challenges of phasing and scaffolding using Hi-C data in complex polyploids. **a**, Schematic illustrating unique and multiple alignments. **b**, Schematic showing the percentage of unique alignments across different ploidy levels. **c**, Schematic illustrating two types of complex contig regions: (1) ”loss”, where a genomic region or entire contig loses all contacts; and (2) ”sparse”, where the loss of several contacts results in sparse signals. **d**, Schematic showing how loss or sparse contacts affect contig clustering and scaffolding. **e, f**, Schematics illustrating how chimeric or collapsed contigs affect contig clustering and scaffolding.



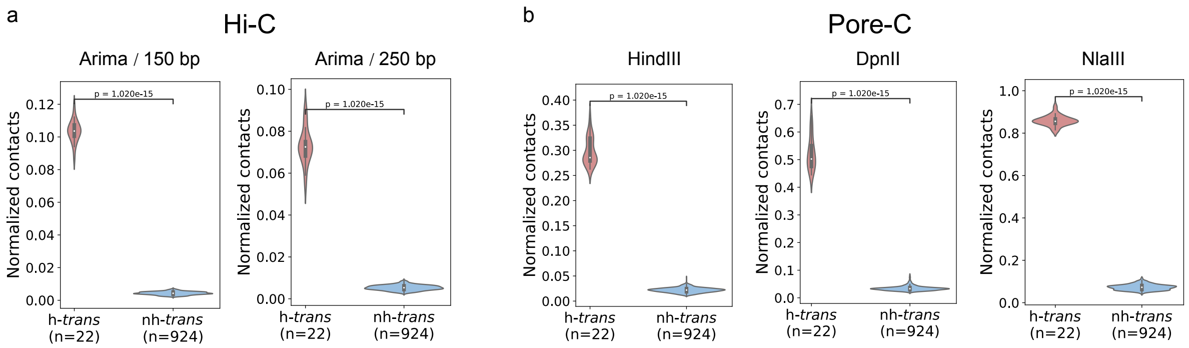
Supplementary Fig. 2 | Problem of polyploid assembly. **a**, Statistics of the unique mapped read pairs of different genomes. **b**, Statistics of the contigs covered by at least one valid restriction site indicate that the up- and downstream 500 bp of each site covered enough contacts (at least 25). **c**, Statistics of genome components (including erroneous, duplicated, haplotig, and collapsed regions) for each assembly, based on comparisons among different assembly results.



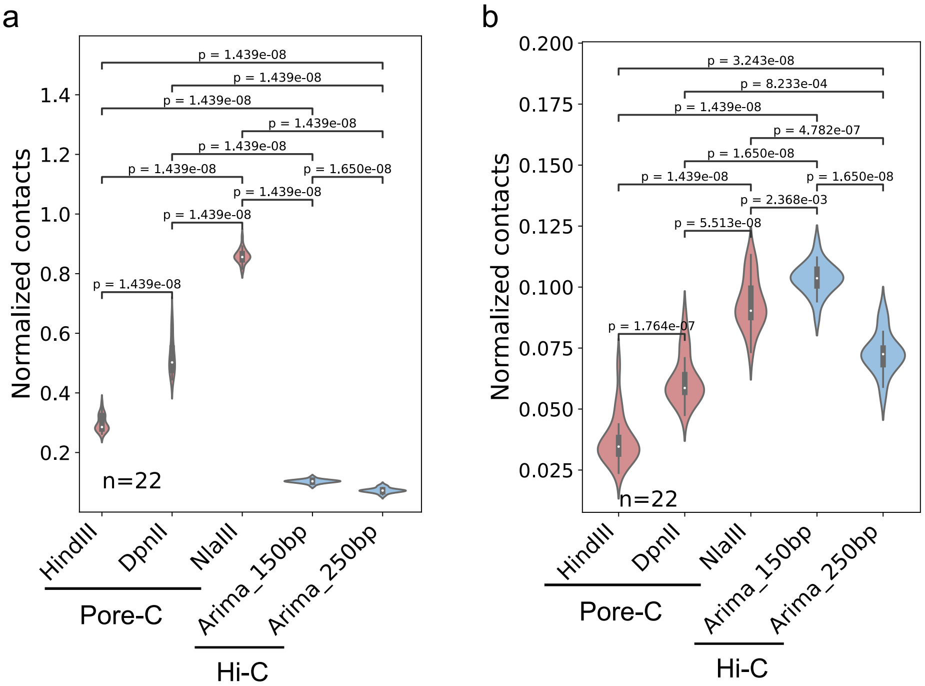
Supplementary Fig. 3 | Comparison of valid contacts between Pore-C and Hi-C at different levels of heterozygosity. These cases are shown: simulated dodecaploid (**a, d**), hexaploid sweet potato (**b, e**), and octoploid wild sugarcane (**c, f**). We categorize sequences with heterozygosity between 0.01% and 1.0% as highly similar, and those with heterozygosity below 0.01% as nearly or completely identical. Heterozygosity was estimated by self-mapping using 5-kb sliding windows. For sequences with multiple mappings, the minimum heterozygosity value was retained. **d-f**, Comparison of valid contacts between Pore-C and Hi-C under varying levels of heterozygosity. Valid contacts for Pore-C were derived from virtual pairwise contacts with MAPQ ≥ 2, whereas Hi-C contacts were filtered using MAPQ ≥ 1. The proportion of valid contacts was defined as the ratio of valid to total contacts. P-values were calculated using two-sided Wilcoxon rank-sum tests. **** indicates P < 1.00e–04.



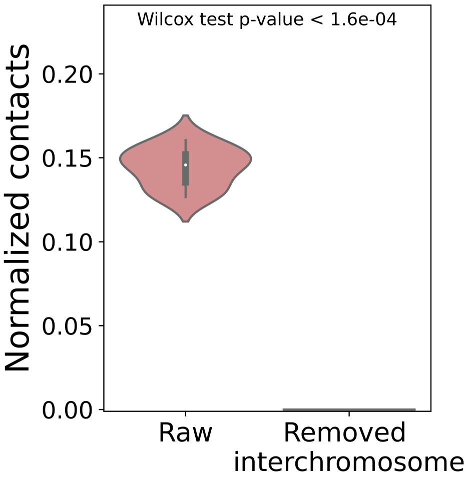
Supplementary Fig. 4 | Schematic illustration of h-*trans* contacts and its associated error sources. **a,** Diagram showing *cis* and h-*trans* contacts. *Cis* contacts refer to interactions between two contigs from the same chromosome (e.g., 1E.ctg1 and 1E.ctg2, 1F.ctg1 and 1F.ctg2), while *trans* contacts occur between contigs from different chromosomes. In diploid or polyploid genomes, trans contacts can be further classified as h-*trans* (between homologous chromosomes; red and orange dashed lines) and nh-*trans* (between non-homologous chromosomes; grey dashed lines). An inter-allelic contig pair representing different alleles at the same genomic position (e.g., 1E.ctg1 and 1F.ctg1 or 1F.ctg2), whereas a cross-allelic contig pair refers to two contigs from homologous chromosomes located at different genomic positions (e.g., 1E.ctg2 and 1F.ctg1 or 1F.ctg2). **b,** Heatmap of Pore-C contacts at 100 kb resolution, highlighting *cis* signals from inter-allelic and cross-allelic contig pairs. **c,d,** Schematic representations of two major sources of h-*trans* errors: sequencing errors in Pore-C reads (**c**) and switch errors between homologous regions (**d**).



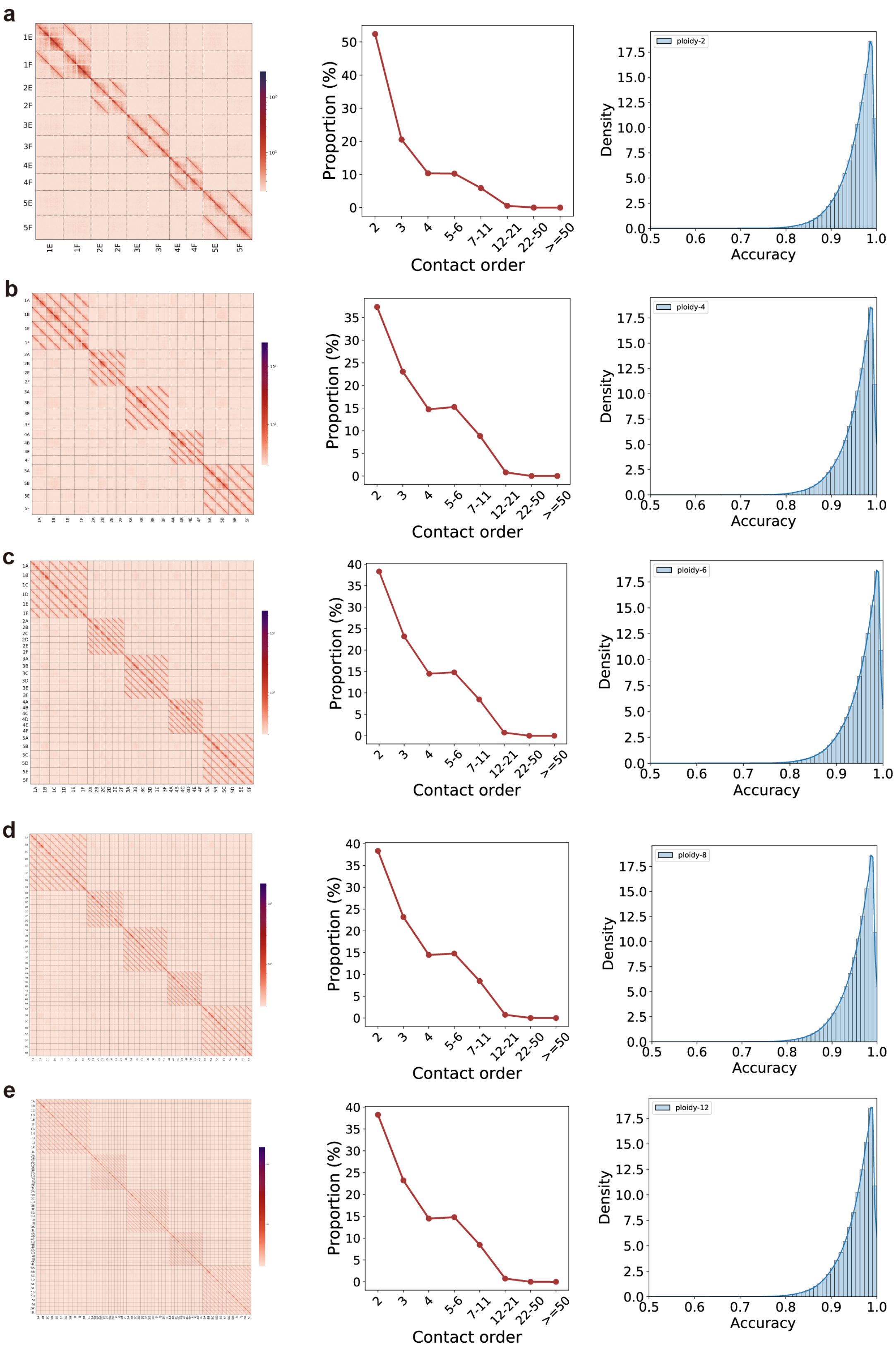
Supplementary Fig. 5 | Comparison of h-*trans* and nh-*trans* contacts. Comparison of h-*trans* and nh-*trans* contacts was performed using the haplotype-resolved HG002 human genome with corresponding Hi-C (**a**) and Pore-C (**b**) data. Pore-C and Hi-C contacts were filtered with MAPQ ≥ 1. h-*trans* contacts refer to interactions between homologous chromosomes (e.g., chr1_maternal and chr1_paternal), while nh-*trans* contacts represent interactions between non-homologous chromosomes (e.g., chr1_maternal and chr2_maternal or chr2_paternal). P-values were calculated using two-sided Wilcoxon rank-sum tests.



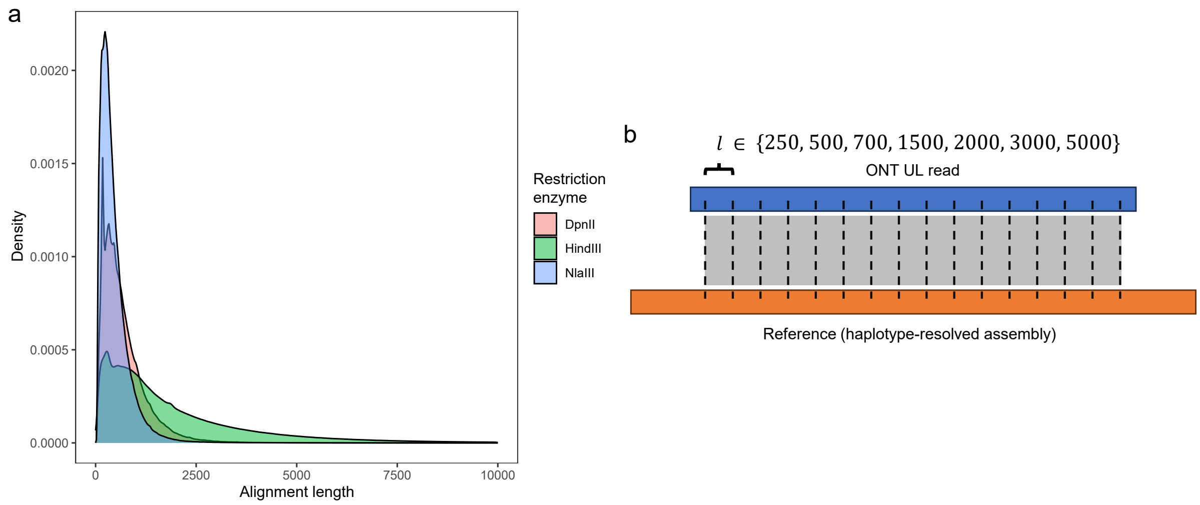
Supplementary Fig. 6 | Comparison of h-*trans* contacts across different datasets. Comparison of h-*trans* contacts was performed between Hi-C and Pore-C data from the HG002 genome, including libraries constructed using different restriction enzymes. **a**, Contacts filtered with a minimum mapping quality (MAPQ) of 1. **b**, Contacts filtered with a minimum mapping quality (MAPQ) of 2. P-values were calculated using two-sided Wilcoxon rank-sum tests.



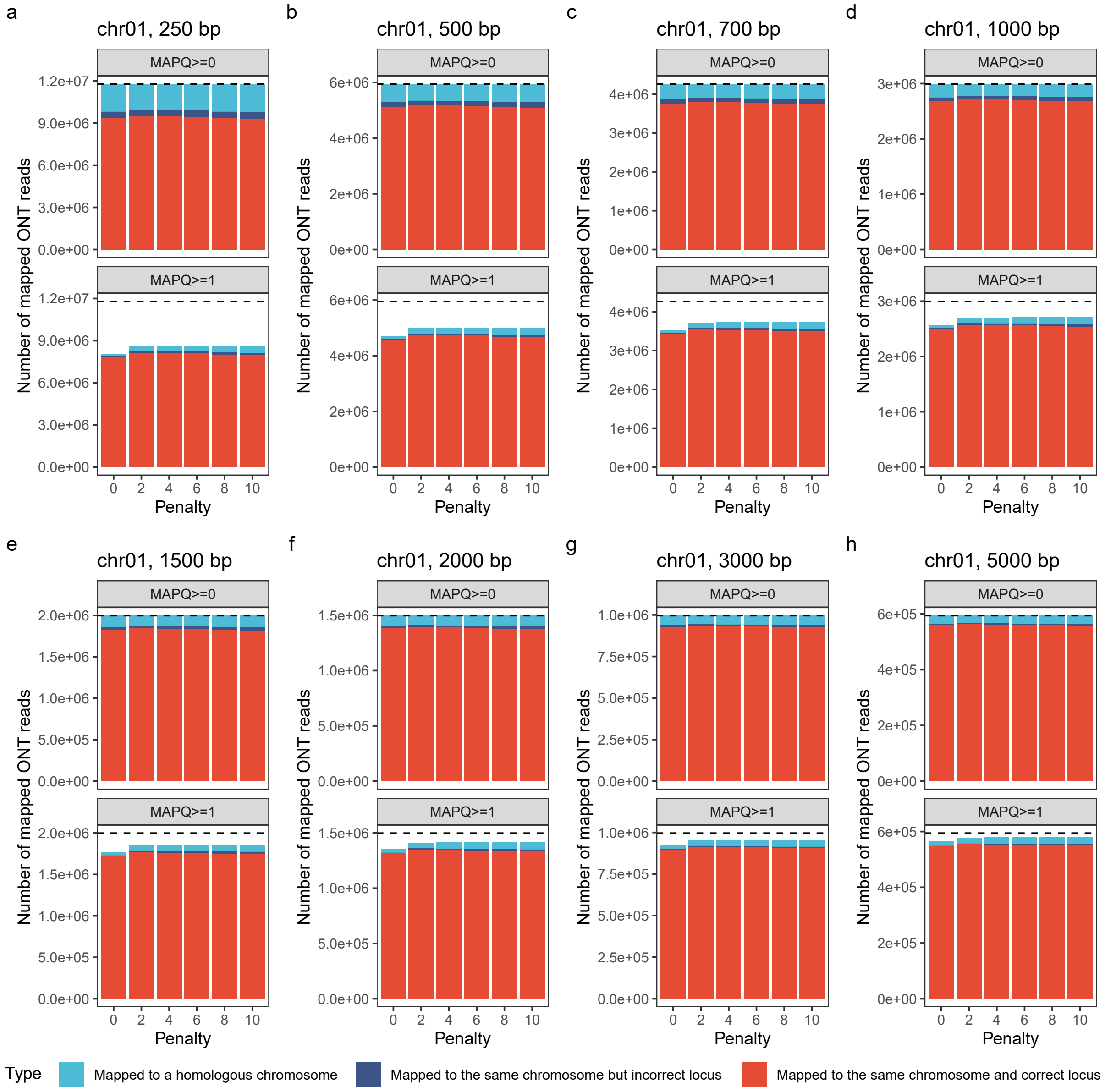
Supplementary Fig. 7 | Comparison of interchromosomal contacts between raw and filtered Pore-C contact maps. To simulate Pore-C data from different ecotypes, interchromosomal interactions were initially removed to assess their impact on downstream polyploid simulations. Violin plots show the distribution of interchromosomal contacts in the raw Pore-C maps and after filtering.



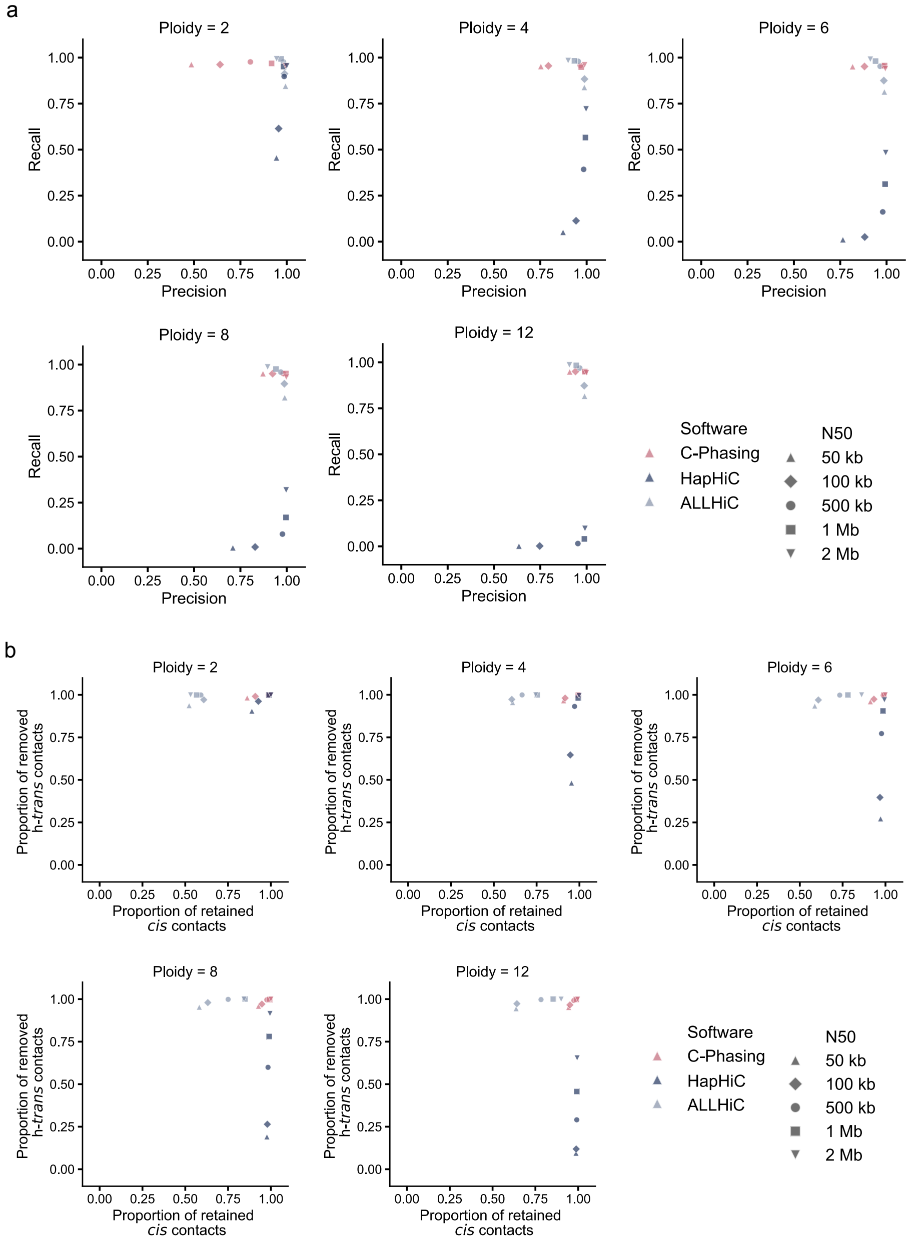
Supplementary Fig. 8 | Summary statistics of the synthetic Pore-C data. Approximately 5× coverage of synthetic Pore-C data was generated for evaluating polyploid genomes. The left panel shows contact heatmaps at 100-kb resolution; the middle panel displays the distribution of concatemer orders; and the right panel shows the accuracy distribution of simulated Pore-C reads. Panels **a** to **e** correspond to ploidy levels of 2, 4, 6, 8, and 12, respectively.



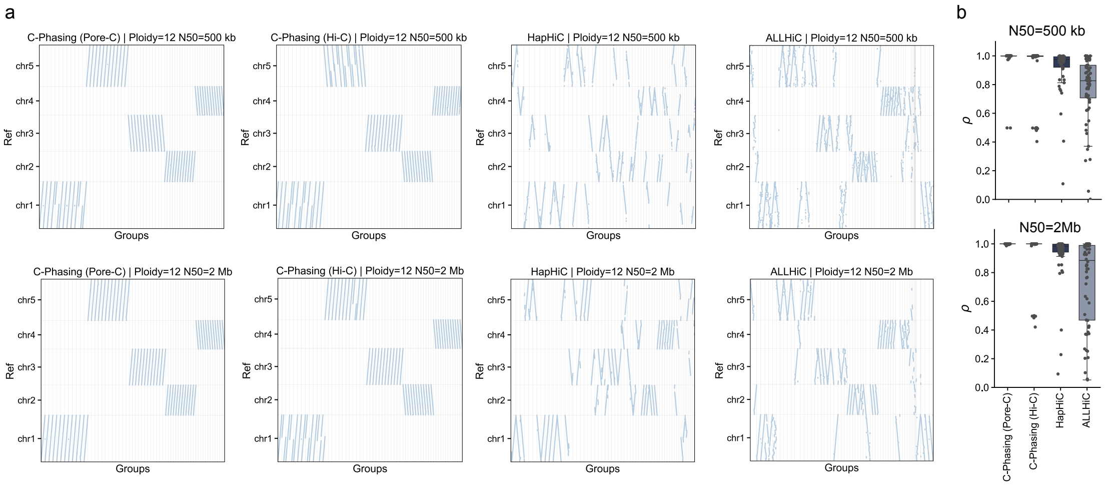
Supplementary Fig. 9 | Simulation of short ONT reads. a, Distribution of alignment lengths of Pore-C reads generated using different restriction enzymes. These Pore-C reads are obtained from the publicly available Human HG002 datasets. **b**, Artificially splitting of ONT ultra-long reads exceeding 100 kb in length into shorter reads ranging from 250 bp to 5 kb.



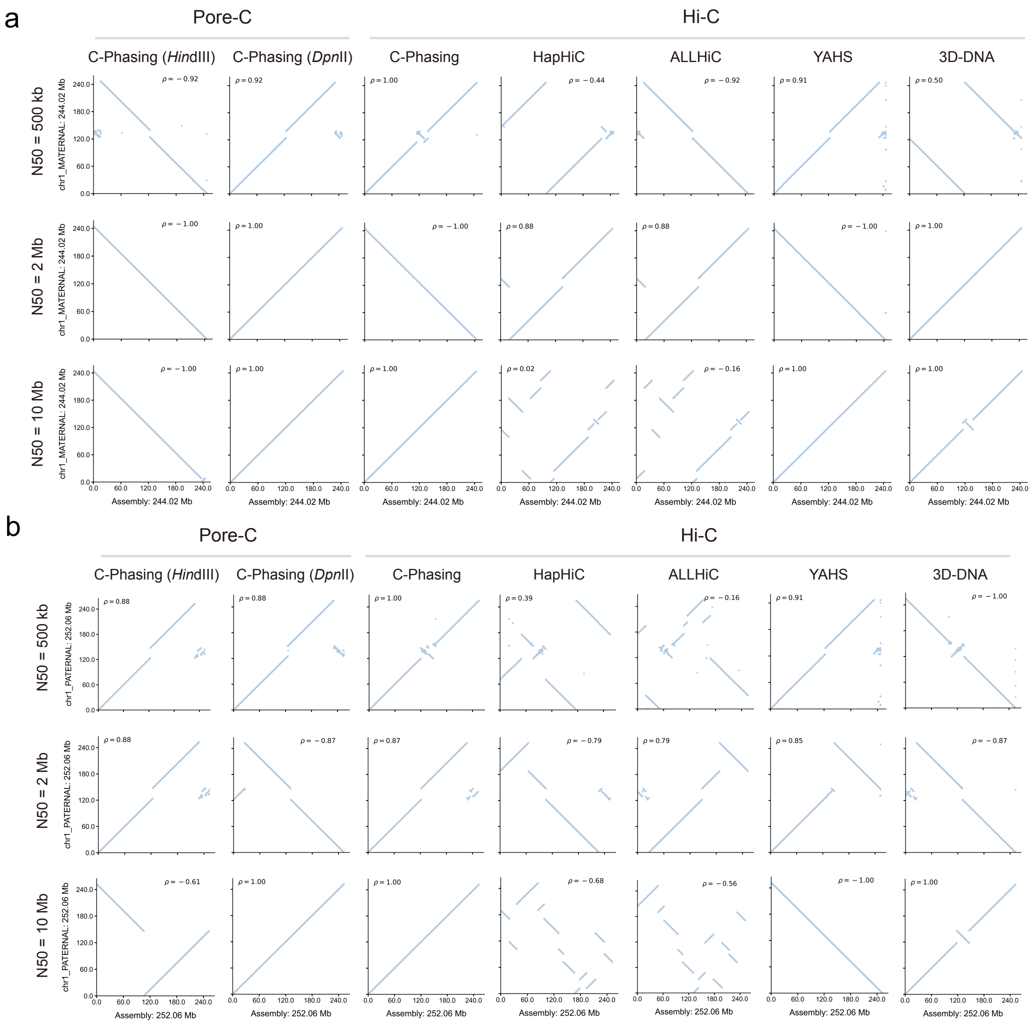
Supplementary Fig. 10 | Mapping accuracy statistics of simulated short ONT reads. The analysis was conducted on homologous chromosome group 1 of sweet potato using simulated ONT reads of varying lengths: 250 bp (**a**), 500 bp (**b**), 700 bp (**c**), 1,000 bp (**d**), 1,500 bp (**e**), 2,000 bp (**f**), 3,000 bp (**g**), and 5,000 bp (**h**).



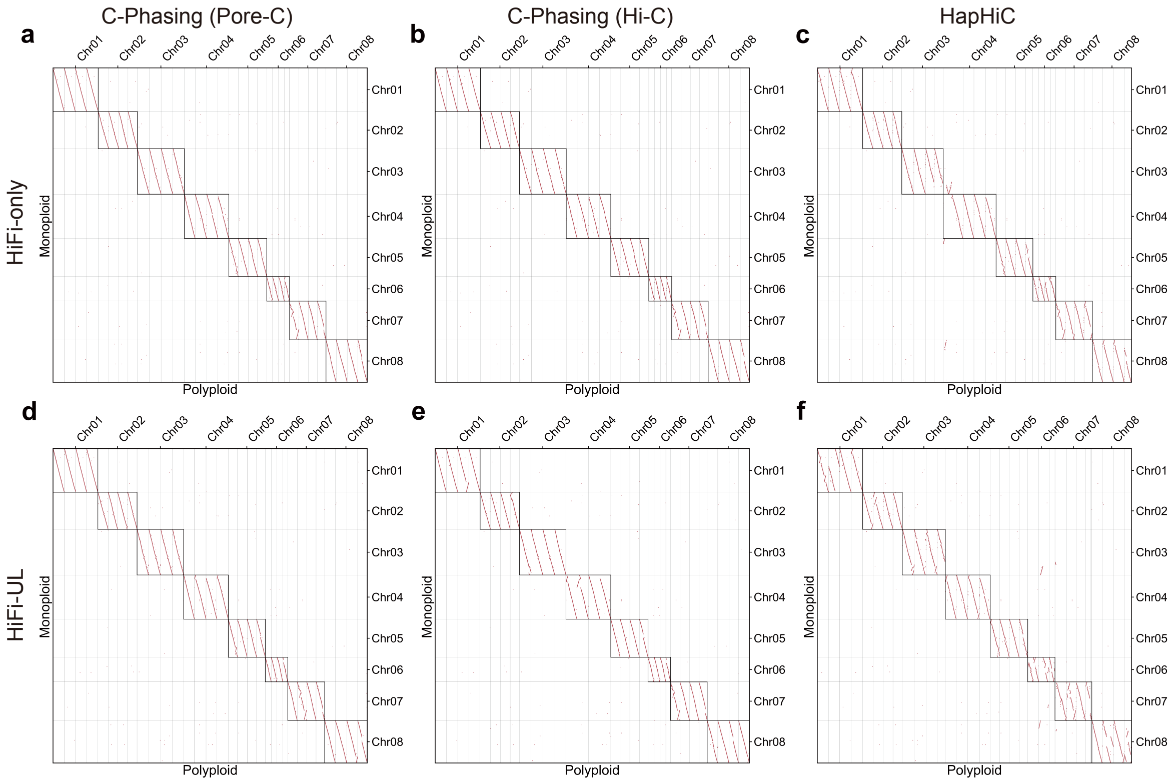
Supplementary Fig. 11 | Evaluation of Kprune's performance on synthetic polyploid datasets across varying contiguity and ploidy levels. **a**, Performance of inter-allelic contig pair identification. A dot located in the upper right corner indicates the best performance, reflecting both high precision and recall. Notably, a higher recall rate facilitates comprehensive detection of cross-allelic contig pairs. **b**, Performance in removing h-*trans* contacts while retaining *cis* contacts. A dot in the upper right corner represents an optimal trade-off, ensuring effective removal of erroneous h-*trans* signals while preserving informative *cis* interactions, which is critical for accurate separation of homologous chromosomes. In both panels, each dot represents an individual sample, with colors denoting different software tools and shapes indicating contig N50 values. Red dots: C-Phasing-Kprune; dark blue: HapHiC-Remove; light blue: ALLHiC-Prune. Shapes: upward triangle (50 kb), diamond (100 kb), circle (500 kb), rectangle (1 Mb), and downward triangle (2 Mb).



Supplementary Fig. 12 | Benchmarking of phasing and scaffolding using simulation data. The results were obtained from phasing and scaffolding analyses performed on simulated dodecaploid contig-level genomes, using either Pore-C or Hi-C data. The top and bottom panels correspond to assemblies with contig N50 values of 500 kb and 2 Mb, respectively. **a**, Dot plots under different experimental conditions. **b**, Box plots showing the absolute Spearman correlation coefficients between the ground truth and the results produced by different software tools. Each box represents the distribution across all chromosomes, with the central line indicating the median, box edges denoting the interquartile range (IQR), and whiskers extending to 1.5× IQR. Outliers are shown as individual points. The number of chromosomes (n = 60) was used for each tool under both N50 settings.



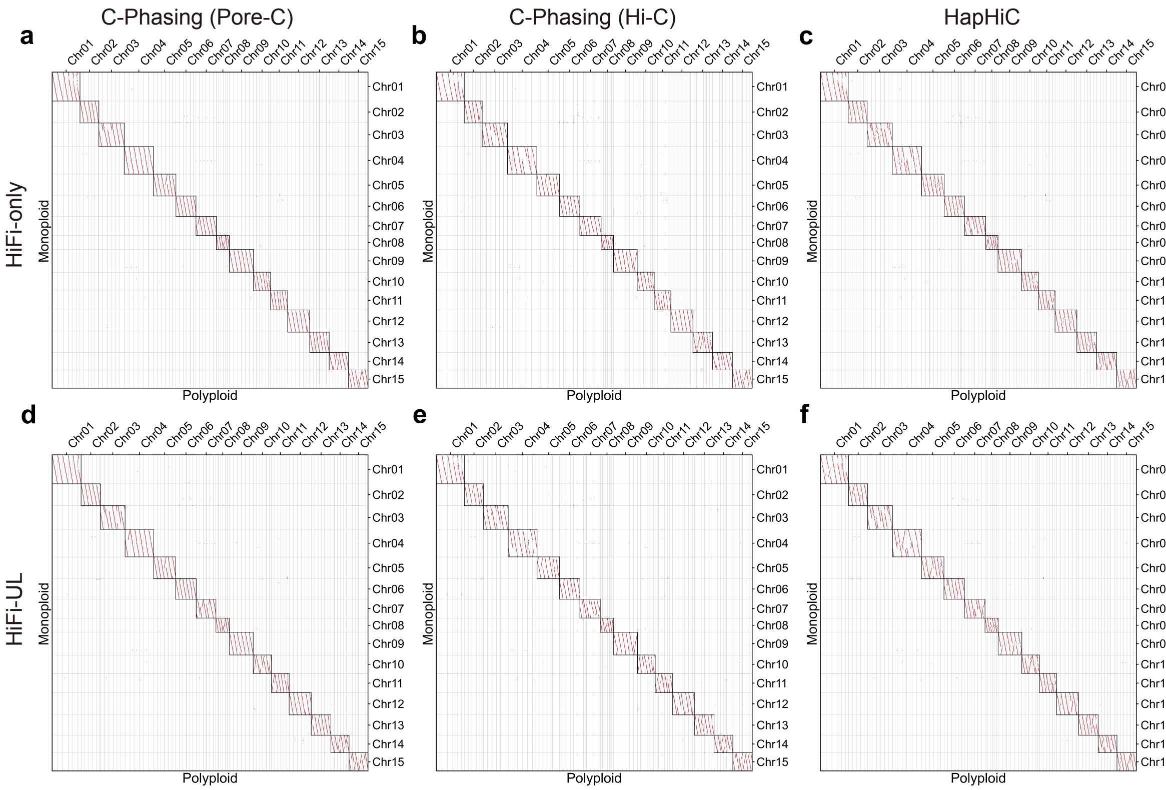
Supplementary Fig. 13 | Evaluation of scaffolding using simulation data. The haplotype-resolved genome of HG002 was downloaded and its chromosomes were fragmented into contigs with varying N50 values (500 kb, 2 Mb, and 10 Mb). Dot plots and absolute Spearman correlation coefficients were generated to compare the scaffolding accuracy of different software tools. Our method, C-Phasing, was evaluated using two Pore-C datasets of HG002, digested with HindIII and DpnII, respectively. In addition, the performance of C-Phasing using Hi-C data was compared with several state-of-the-art tools, including HapHiC, ALLHiC, YAHS, and 3D-DNA. Results for chromosome 1 of the maternal (**a**) and paternal (**b**) haplotypes are shown.



Supplementary Fig. 14 | Synteny dot plots of alfalfa assemblies generated by different scaffolding strategies, used to benchmark chromosome-scale assembly accuracy on real data.

Supplementary Fig. 15 | Heatmaps of alfalfa Pore-C or Hi-C contact maps at 500 kb resolution for benchmarking chromosome-scale assembly performance using real data.

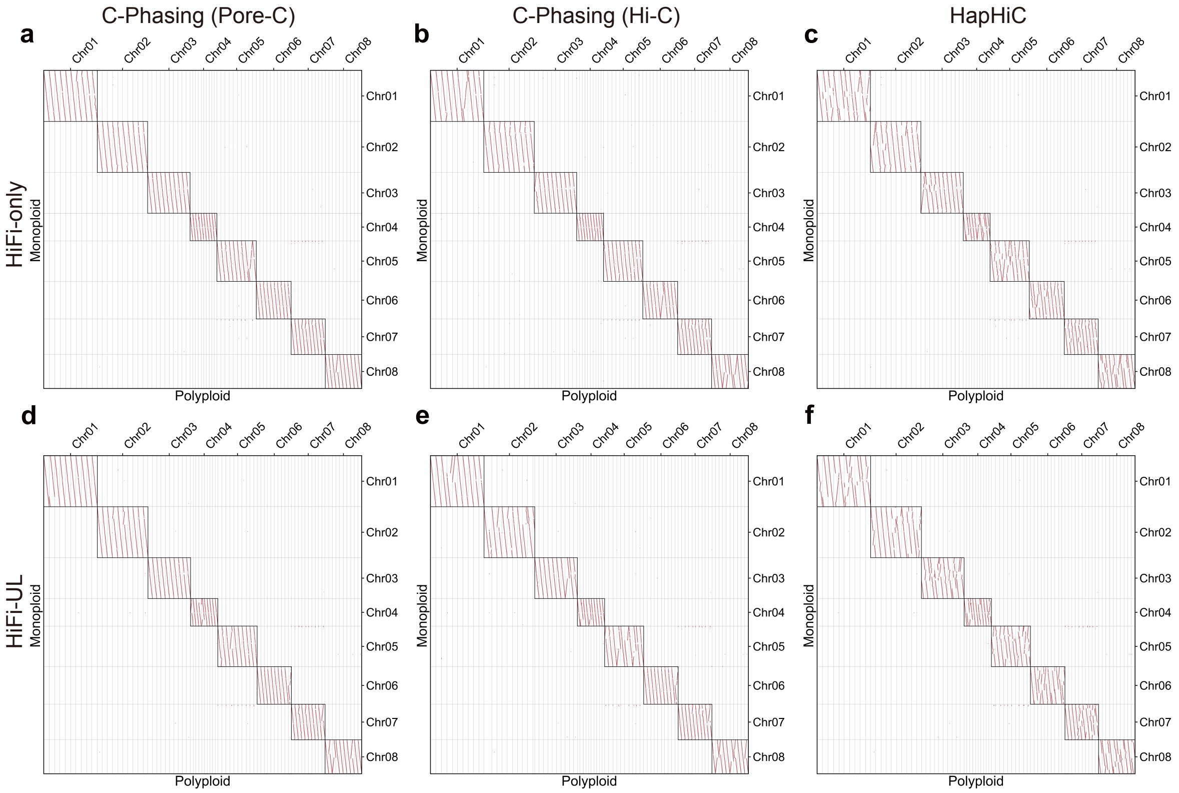
(*included in separate PDF file*)



Supplementary Fig. 16 | Synteny dot plots of sweet potato assemblies generated by different scaffolding strategies, used to benchmark chromosome-scale assembly accuracy on real data.

Supplementary Fig. 17 | Heatmaps of sweet potato Pore-C or Hi-C contact maps at 500 kb resolution for benchmarking chromosome-scale assembly performance using real data.

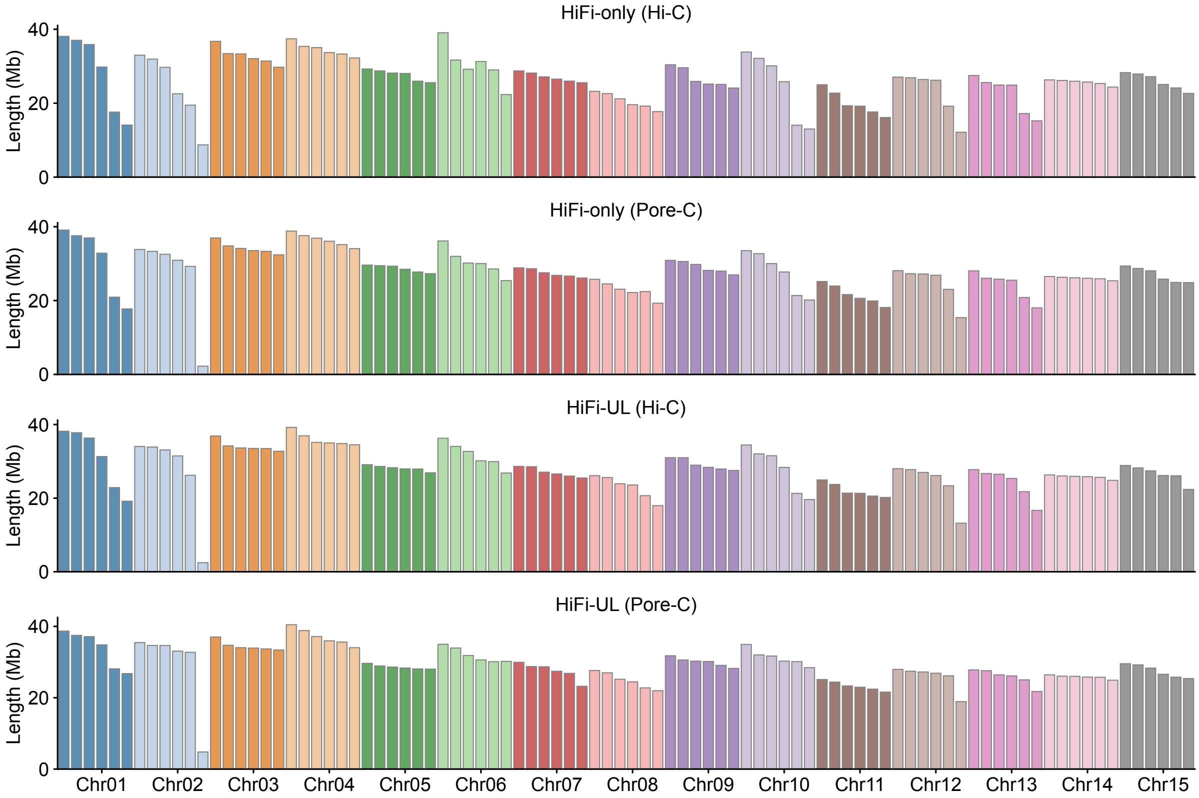
(*included in separate PDF file*)



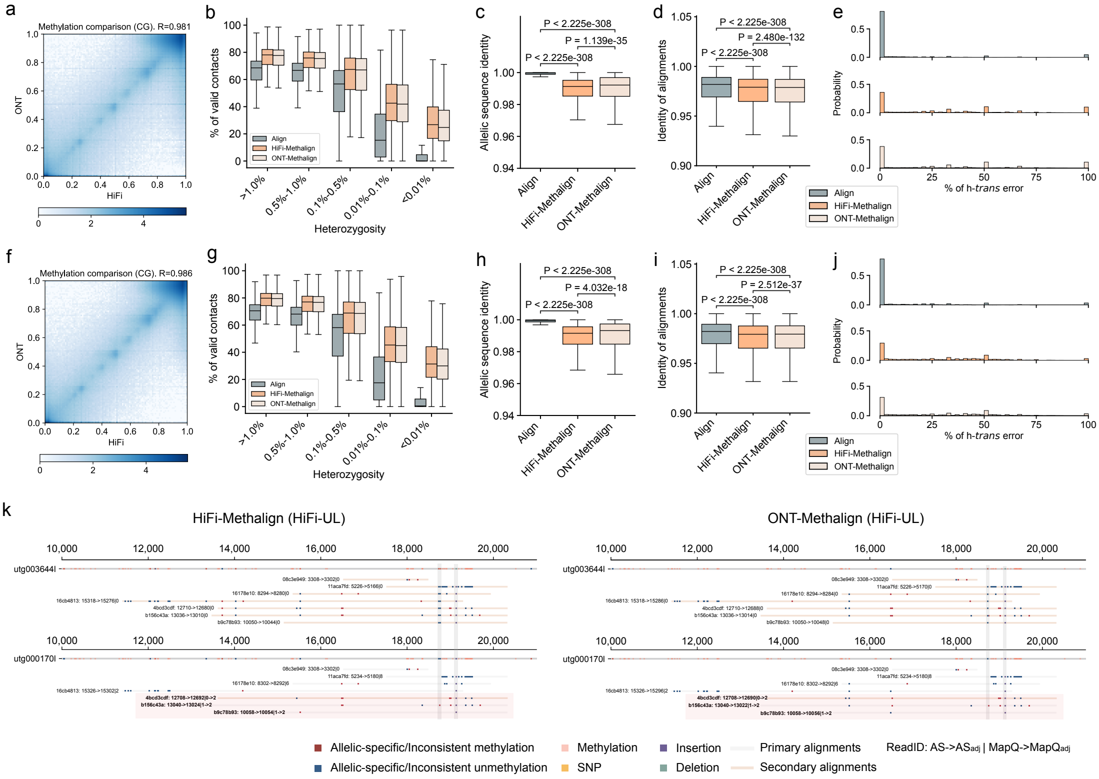
Supplementary Fig. 18 | Synteny dot plots of wild sugarcane assemblies generated by different scaffolding strategies, used to benchmark chromosome-scale assembly accuracy on real data.

Supplementary Fig. 19 | Heatmaps of wild sugarcane Pore-C or Hi-C contact maps at 1 Mb resolution for benchmarking chromosome-scale assembly performance using real data.

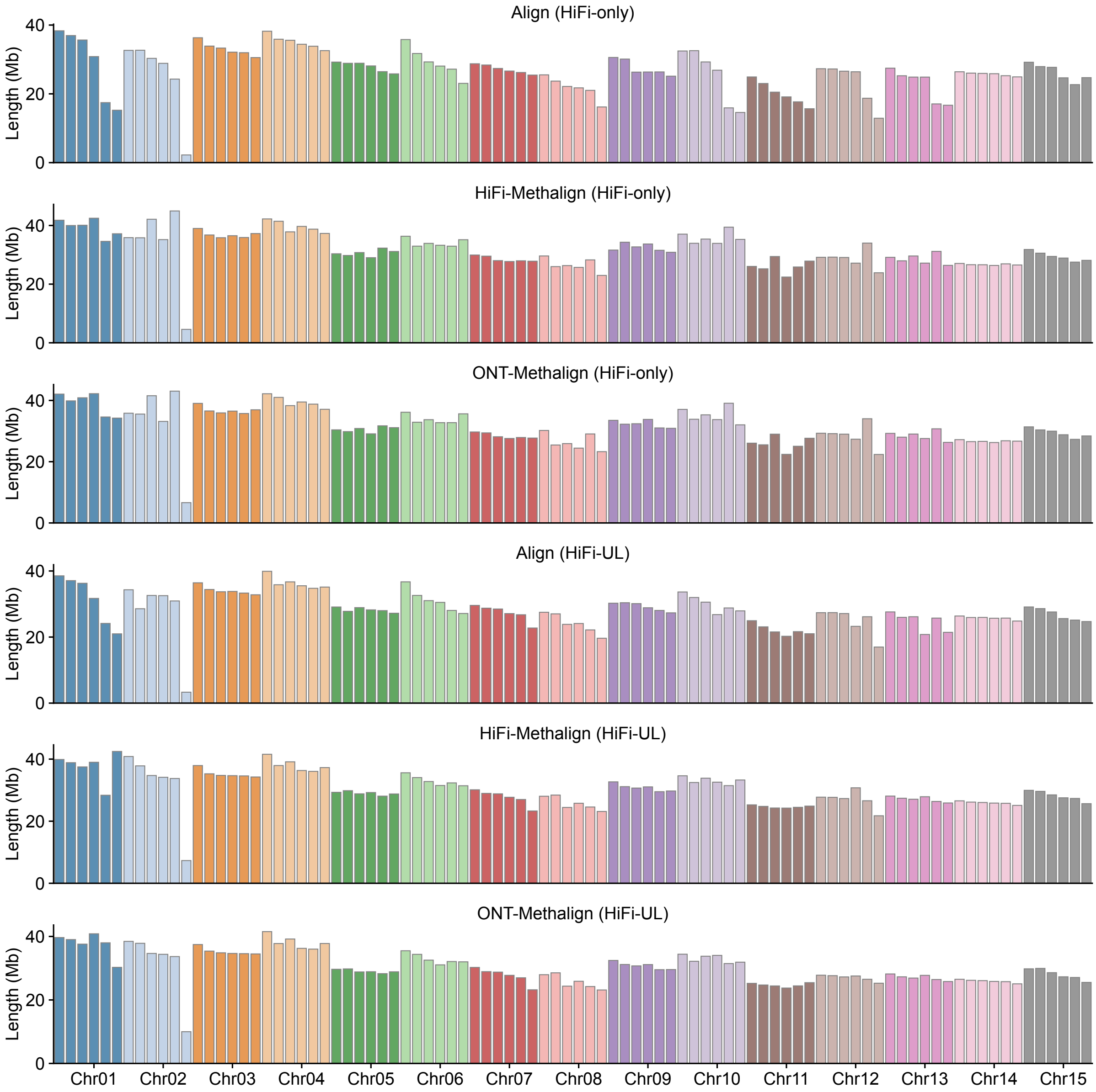
(*included in separate PDF file*)



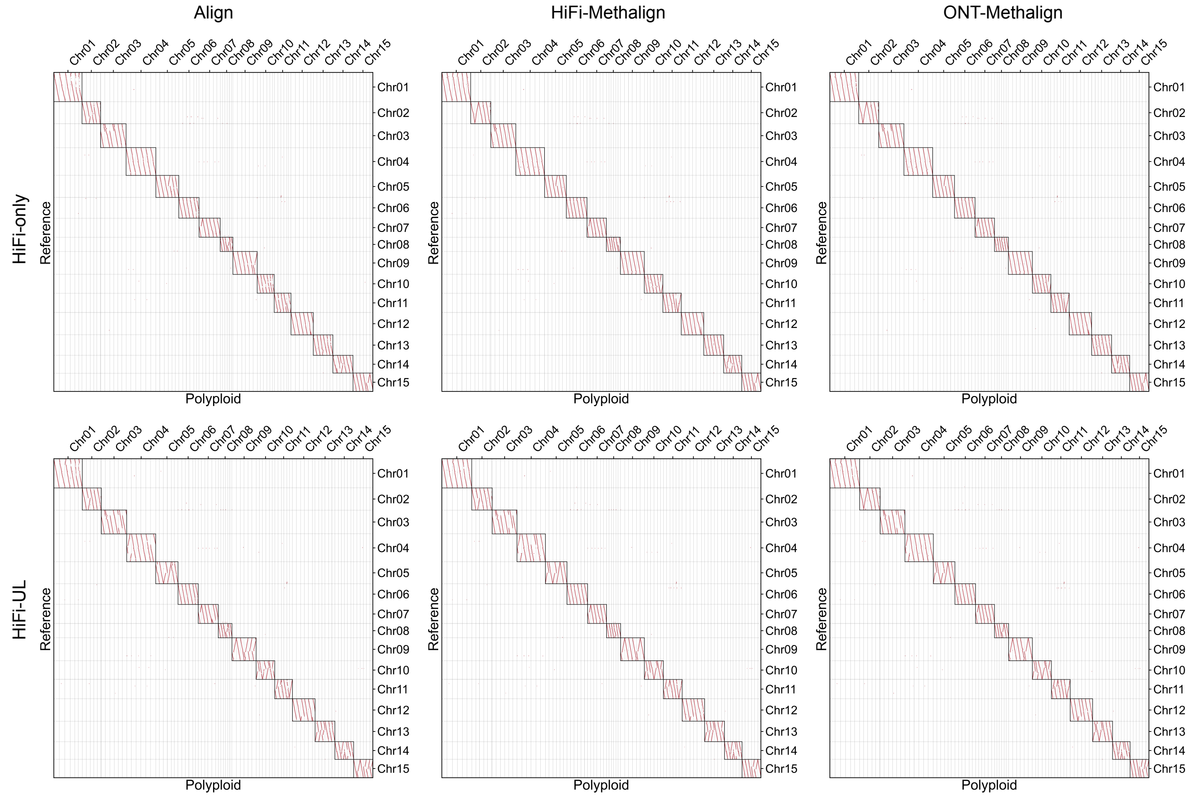
Supplementary Fig. 20 | Comparison of the scaffold length between C-Phasing (Hi-C) and C-Phasing (Pore-C). Note: these assemblies were generated by C-Phasing on all Hi-C data or 125x Pore-C data.



Supplementary Fig. 21 | Comparative analysis of Pore-C alignments processed using Align and Methalign. The analyses were performed on two assemblies of sweet potato: HiFi-only (**a–e**) and HiFi-UL (**f–j**). **a, f**, Comparison of 5mCG methylation detection between PacBio HiFi and ONT data across the genome. **b, g**, Proportion of valid contacts derived from Align and Methalign across genomic regions with varying levels of heterozygosity. **c, h**, Sequence identity between allelic contigs as evaluated using Align and Methalign. **d, i**, Read-to-contig alignment identity comparison between Align and Methalign. **e, j**, Statistics of h-*trans* contact errors. **k**, Read pileup plot of an allelic contig pair, illustrating how Methalign refines ambiguous alignments.



Supplementary Fig. 22 | Comparison of the scaffold length between Align and Methalign. Note: these assemblies were generated by C-Phasing on 50x Pore-C data.



Supplementary Fig. 23 | Synteny dot plots of the sweet potato for using Align or Methalign results. Note: these assemblies were generated by C-Phasing on 50x Pore-C data.

Supplementary Fig. 24 | Heatmaps of cultivated sugarcane ZZ01 Pore-C or Hi-C contact maps at 1 Mb resolution. (*included in separate PDF file*)