

RAmbler resolves complex repeats in human Chromosomes 8, 19 and X (Supplemental Material)

Sakshar Chakravarty, Glennis Logsdon, Stefano Lonardi

Additional experimental results on *Homo sapiens*

Supplemental Figure 2 shows the NucFreq plots for the mapping coverage of PacBio HiFi reads across the chromosomes for GRCh38.p13 and T2T-CHM-13.v2.0. For both assemblies of the human genome, observe that centromeres are likely not well assembled yet. They appear to contain a few gaps and a significant number of collapsed repeats. There are also some long repeats spread across the non-centromeric regions of several chromosomes.

Supplemental Table 7 reports the QUAST results obtained for human Chromosomes 8, 19, and X from GRCh38.p13 (HG38) and RAmbler (RA.HG38.HG002, RA.HG38.HG00733 and RA.HG38.HG01346) using T2T-CHM13.v2.0 (T2T) as the reference where the lengths of the corresponding chromosomes are 146,259,331, 61,707,364, and 154,259,566 respectively. Observe that RAmbler achieves higher genome fraction, normalized NG50 (η), and smaller number of misassemblies for all three of the chromosomes than HG38. Supplemental Figure 7 shows a chromosome-level alignment with SyRI to determine syntenic regions, inversions, translocations and duplications of RAmbler's assemblies (RA.HG38.HG002, RA.HG38.HG00733 and RA.HG38.HG01346) against HG38 and T2T.

Supplemental Figure 8A shows the NucFreq plots for the mapping coverage of ultra-long ONT reads across the Chromosomes 19, 8, and X for T2T-CHM-13.v2.0 (T2T) and RAmbler (RA.T2T.CHM13) where the RAmbler assemblies are reconstructed using the long PacBio HiFi reads (SRA accession SRR11292120-SRR11292123) for the five corresponding blocks. Supplemental Figure 8B shows the Circos plots obtained from the CRAQ assessment of the same three chromosomes indicating comparable performance between the T2T and RAmbler assemblies with no misjoins.

Supplemental Figure 9A shows the NucFreq plots for the mapping coverage of PacBio HiFi reads across the Chromosomes 19, 8, and X for the maternal HG002 assembly (MAT002) and RAmbler assembly (RA.MAT002.HG002). The RAmbler assemblies used the PacBio HiFi reads ([link](#)) for the five corresponding blocks. Supplemental Figure 9B shows the Circos plots obtained from the CRAQ assessment of the same three chromosomes indicating comparable performance between the MAT002 and RAmbler assemblies with no misjoins.

The plots in Supplemental Figure 11A show a perfect synteny between the RA.T2T.CHM13 assembly (blue) and the T2T assembly (orange) across Chromosomes 19, 8 and X. Supplemental Figure 11B demonstrates perfect synteny between RA.MAT002.HG002 (blue) and MAT002 (orange) assemblies across Chromosomes 19 and 8, and a near perfect synteny on Chromosome X with a small inversion.

Supplemental Tables 8, and 9 report the results obtained from CRAQ assessment of human Chromosomes 19, 8, and X from T2T-CHM13.v2.0 (T2T), and RAmbler (RA.T2T.CHM13) and Supplemental Table 10 reports the results obtained from CRAQ assessment of human Chromosomes 19, 8, and X from the maternal HG002 assembly (MAT002) and RAmbler (RA.MAT002.HG002). Observe that RAmbler achieves the best scores with respect to coverage rate, R-AQI, and S-AQI for most of the chromosomes.

Command-line listing

All the commands used to execute tools, assemblers, and RAMbler are listed below.

Commands for *k*-mer analysis

```
jellyfish count -m $k -C -o $count_output -c 3 -s 10000000 --disk -t 16 $reads
jellyfish histo -o $histo_output -v $count_output
jellyfish dump -c -t -L $lower -U $upper $count_output | awk '{print $1}' > $dump_output
```

Commands for assemblers

```
# hifiasm
hifiasm -o $output/prefix -t 20 $reads

# LJA
lja -t 20 -o $output --reads $reads

# LJA (for simulated diploid genomes)
lja --diploid -t 20 -o $output --reads $reads

# HiCANU
canu -p prefix -d $output genomeSize=$ASM_SIZE -pacbio-hifi $reads

# Verkko
verkko -d $output --hifi $reads --no-nano
```

Commands for RAMbler

```
bash run.sh -r read_file -u unikmer_file -o output_directory -s assembly_length \
  [-k kmer_length] [-l tolerance] [-h threshold]
required:
-r path to the fasta file containing the HiFi reads extracted in step A
-u path to the file containing the list of unikmers extracted in step B
-o path to the directory where RAMbler will put all the generated outputs
-s estimated length of the repeat resolved part of the genome in bp
optional:
-k length of kmers used (default: 21)
-l tolerance parameter (default: 15)
-h threshold parameter (default: 15)

bash stitch.sh -g draft_genome_file -n contig_id \
  -b start_position -e end_position -o output_directory -s contig_length
required:
-g path to the fasta file containing the draft/reference genome
-n id of the contig that is being re-assembled
-b approximate start position of the repeat region in the contig sequence \
  from the draft genome
-e approximate end position of the repeat region in the contig sequence \
  from the draft genome
-o path to the directory where RAMbler will put all the generated outputs \
  (always put the same one as run.sh)
-s estimated length of the repeat resolved contig in bp
```


Commands for evaluation tools

```
# CRAQ
perl craq -g reference.fasta -ngs ngs.bam -sms hifi.bam \
  -pl T -D output_directory

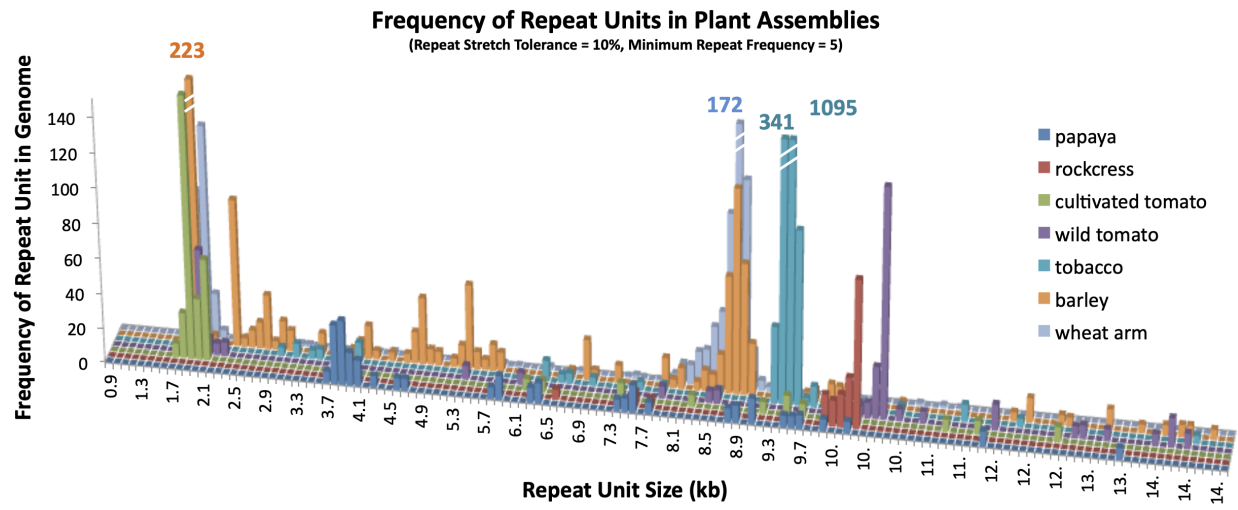
# QUAST
quast <assembly1> <assembly2> ... <assemblyN> -r reference.fasta

# BLAST
makeblastdb -in assembly.fasta -dbtype nucl -out prefix_db
blastn -query repeat_unit.fasta -db prefix_db \
  -outfmt "6 qseqid sseqid pident length evalue bitscore" -out out_file.tsv
```

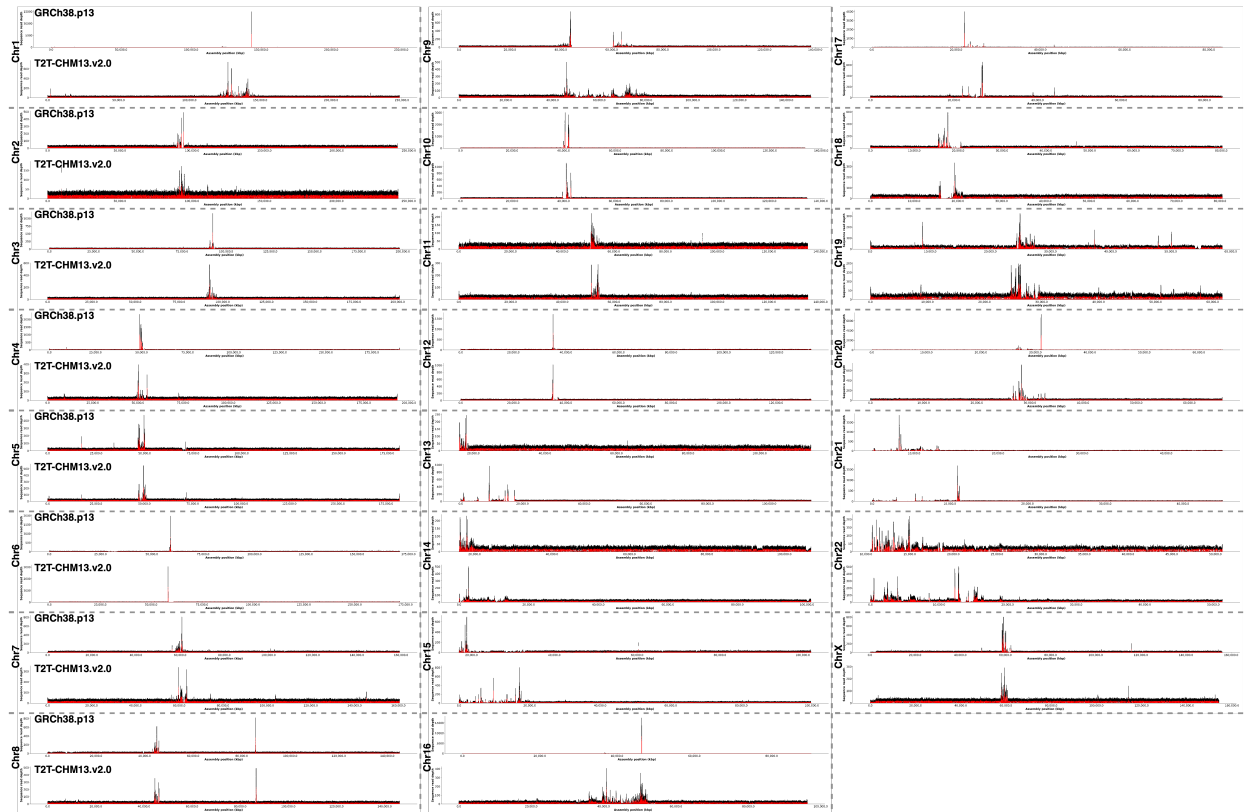
Supplemental Code

Detailed instructions for executing RAMbler are provided in the GitHub repository (<https://github.com/ucrbioinfo/rambler>).

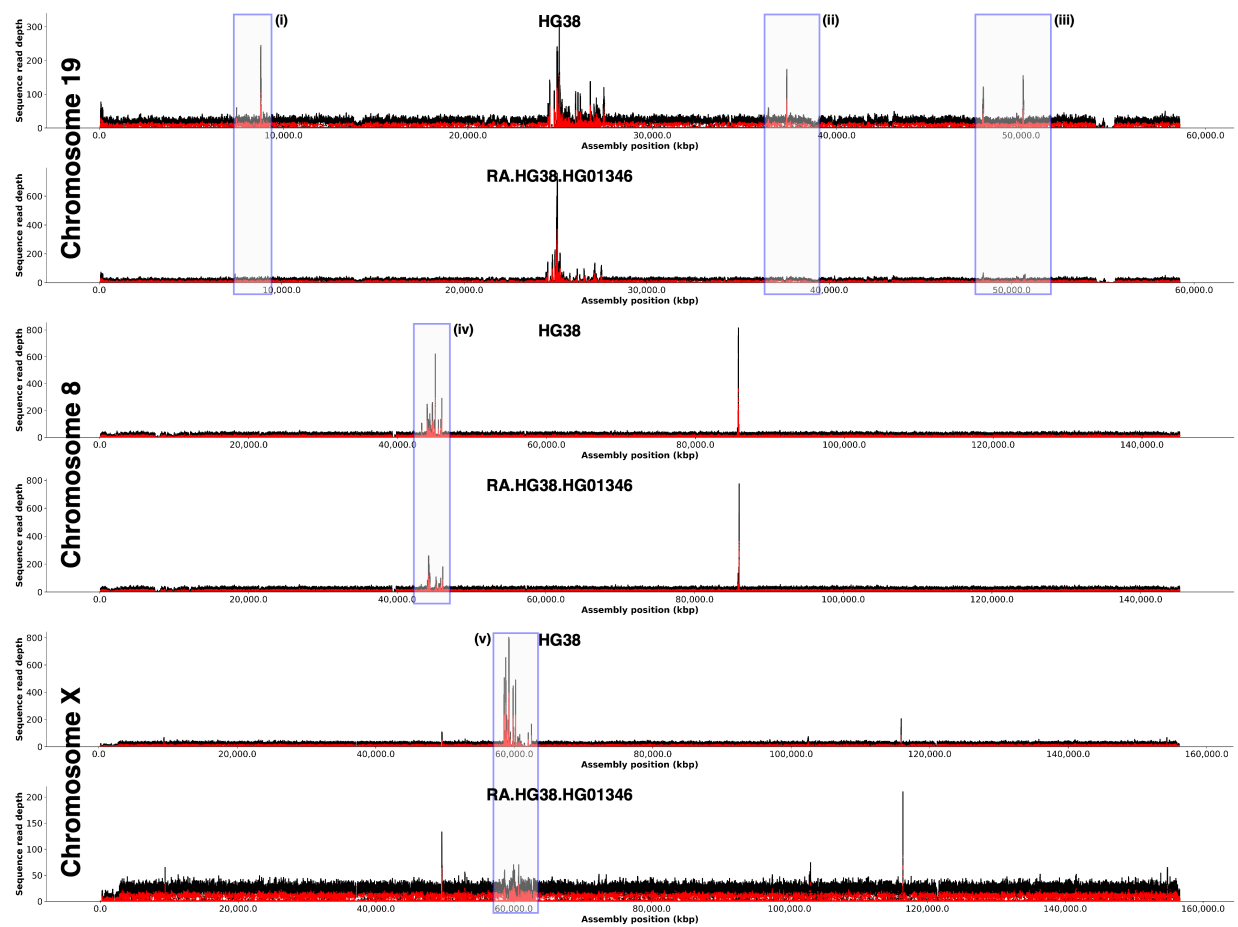
Supplemental Figures



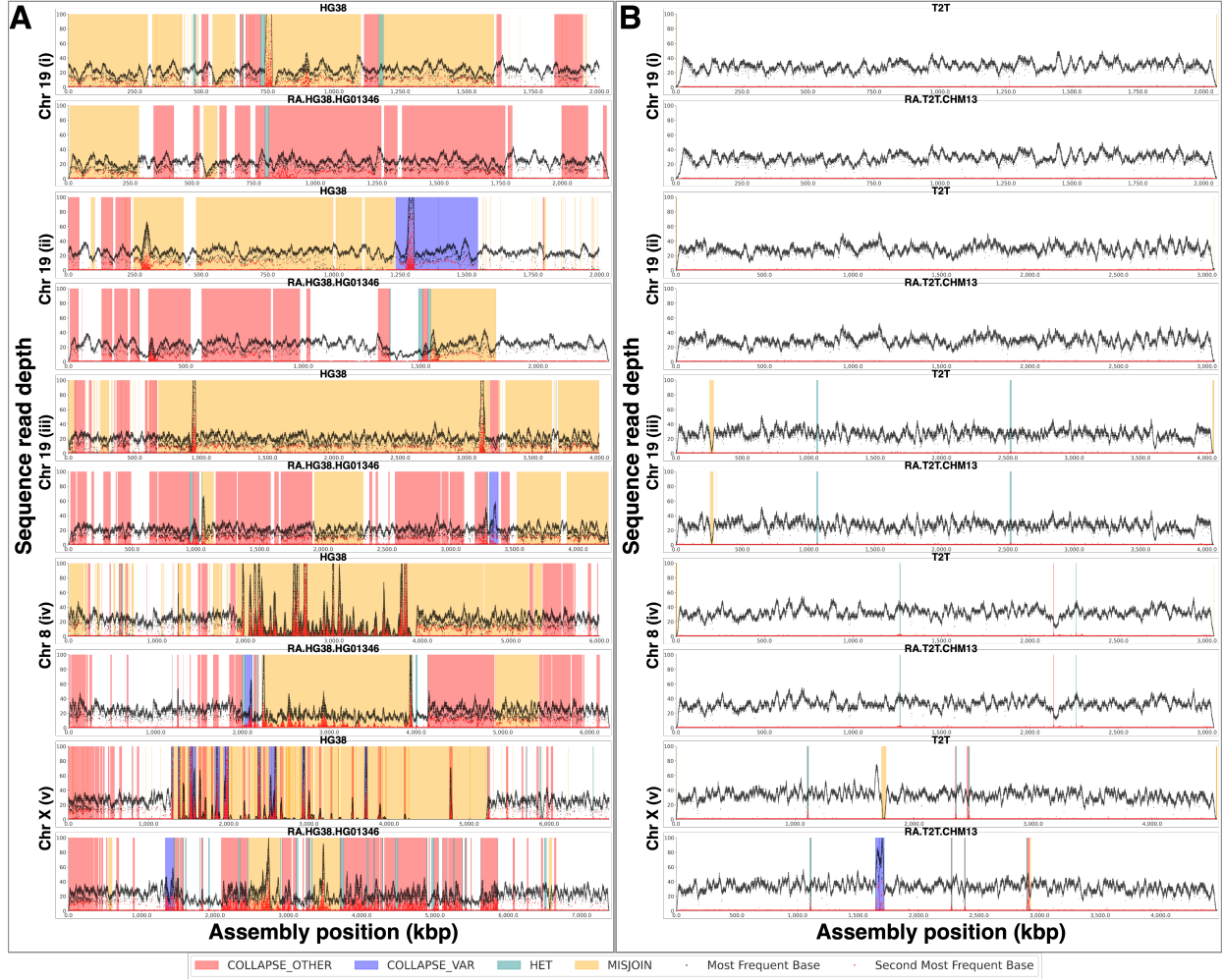
Supplemental Figure 1: A frequency histogram of unresolved repeats in seven plant genomes (2015); the x-axis represents the repeat size (in kb); the y-axis represents the number of copies in each genome of those repeats' sizes (Figure reproduced with permission (Chan et al. 2015)).



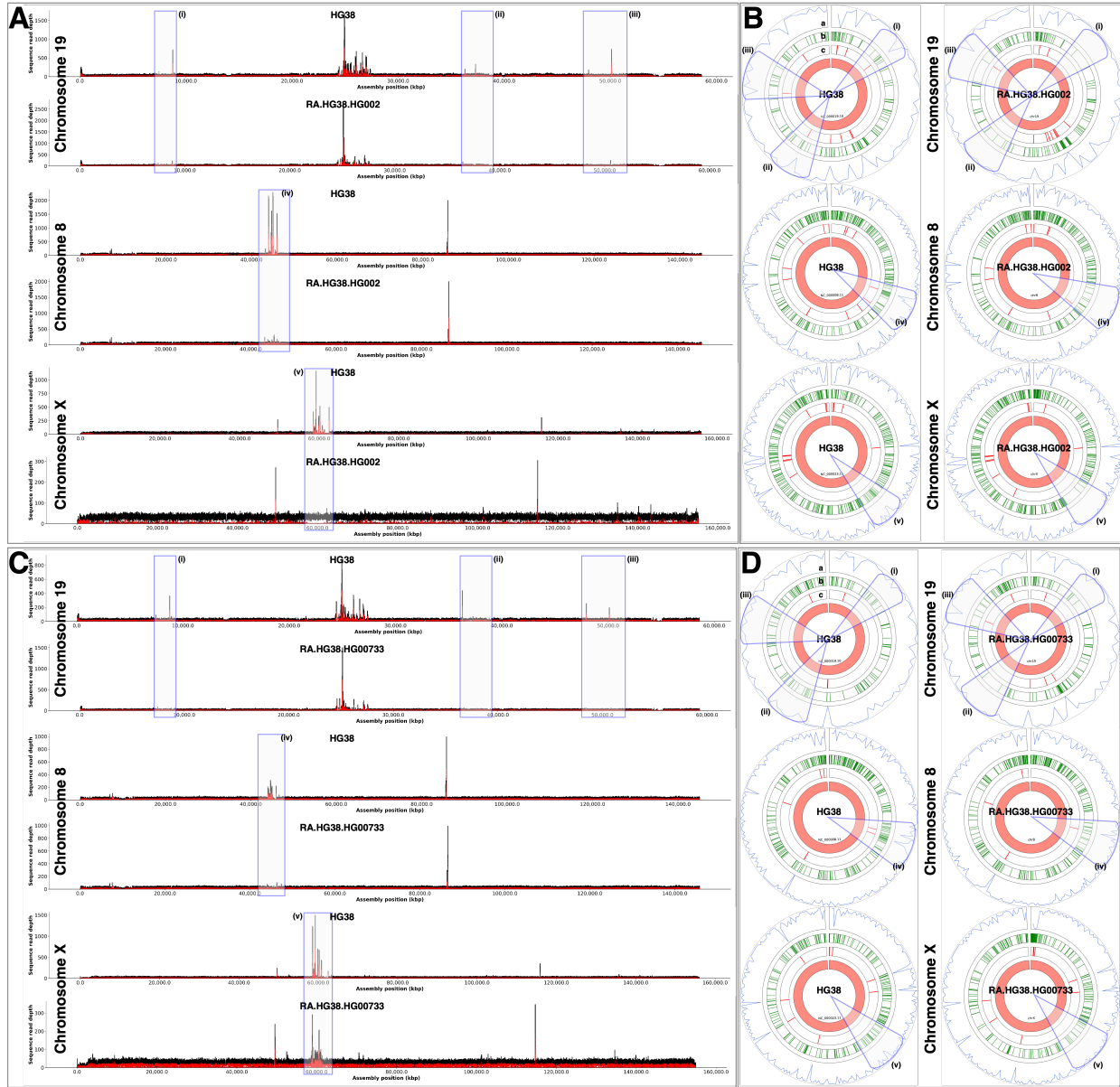
Supplemental Figure 2: Mapping coverage for HiFi HG01346 reads ($\sim 26\times$ coverage) on all chromosomes in GRCh38.p13 and T2T-CHM13.v2.0.



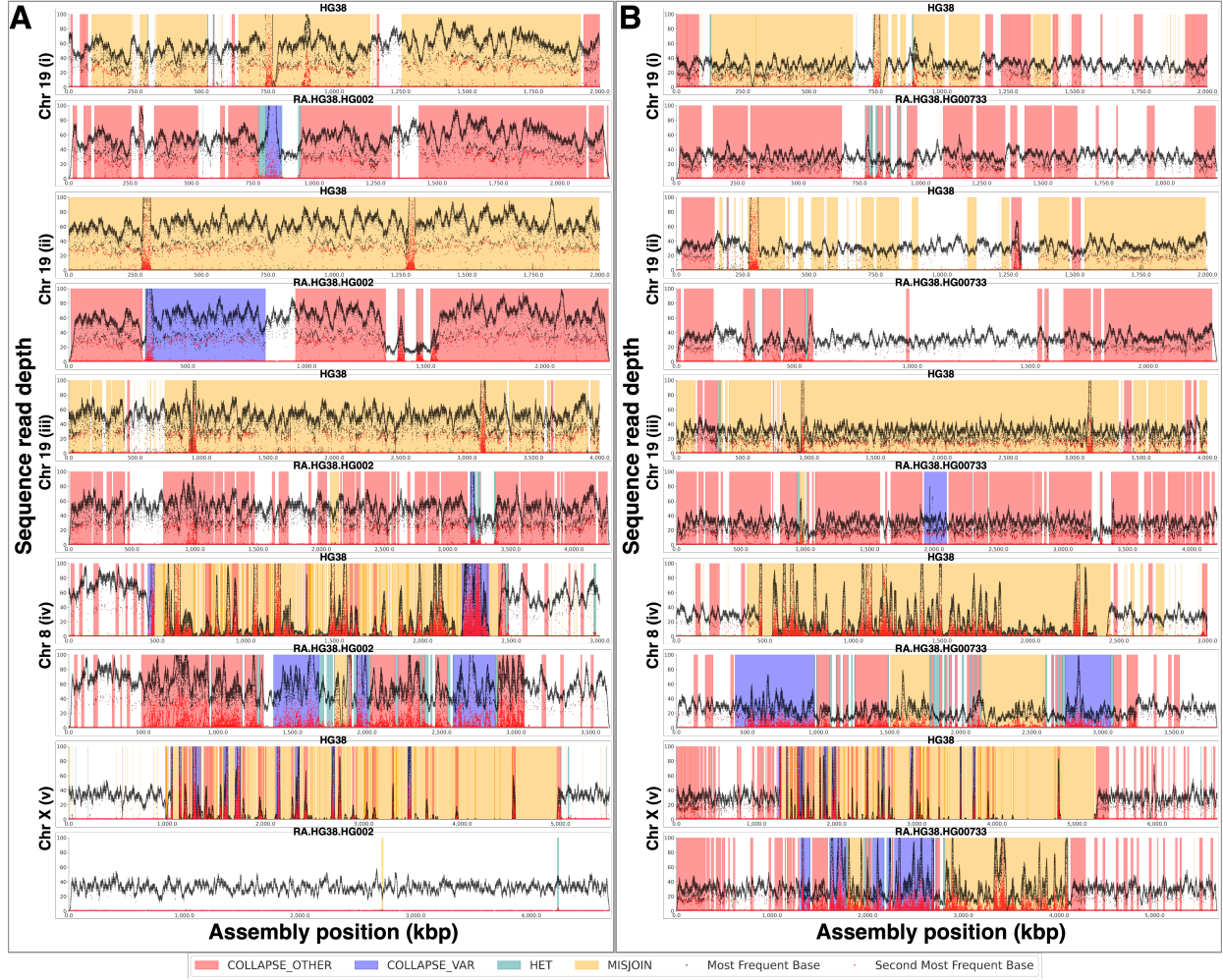
Supplemental Figure 3: NucFreq plots illustrating HiFi read mapping coverage for the comparison of RA.HG38.HG01346 assembly against GRCh38.p13 (HG38) of human Chromosomes 8, 19, and X; in all plots, blocks (i), (ii), and (iii) are non-centromeric repeats in Chromosome 19; blocks (iv) and (v) are the centromeric regions of Chromosomes 8 and X, respectively.



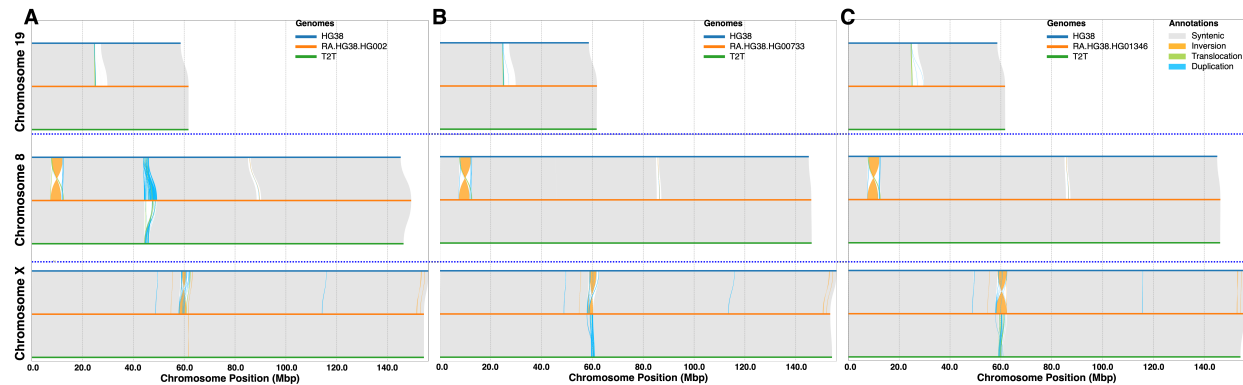
Supplemental Figure 4: NucFlag plots illustrating HiFi read mapping coverage for the comparison of (A) RA.HG38.HG01346 assembly against GRCh38.p13 (HG38) and (B) RA.T2T.CHM13 assembly against T2T-CHM13.v2.0 (T2T) of human Chromosomes 8, 19, and X; in all plots, blocks (i), (ii), and (iii) are non-centromeric repeats in Chromosome 19; blocks (iv) and (v) are the centromeric regions of Chromosomes 8 and X, respectively.



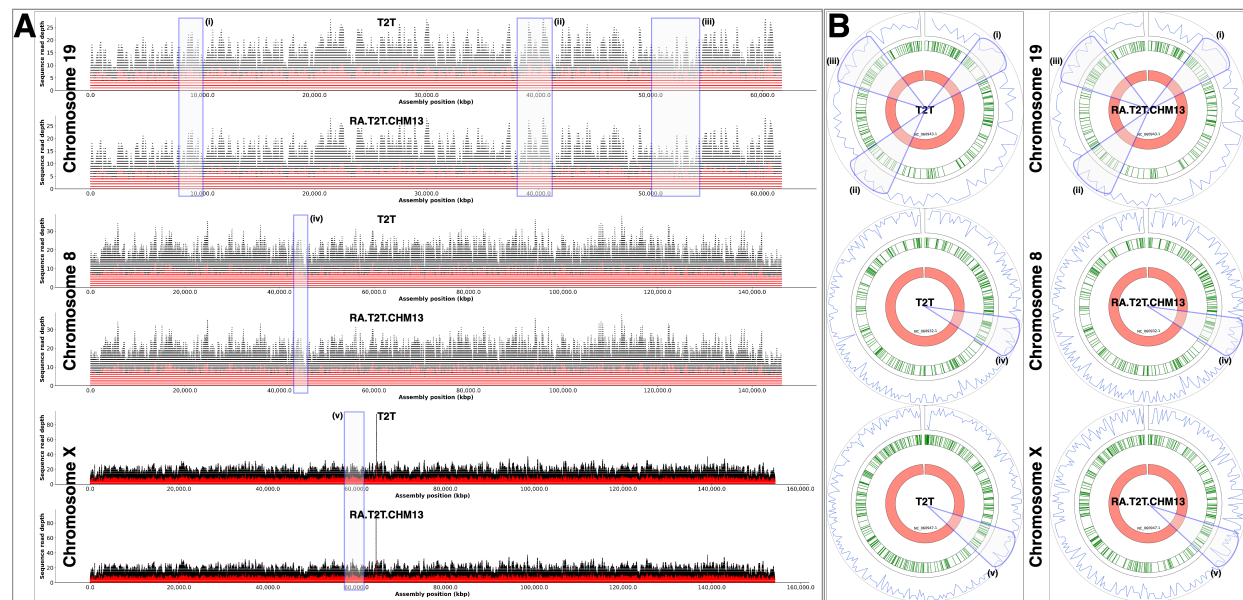
Supplemental Figure 5: Comparing assemblies RA.HG38.HG002 and RA.HG38.HG00733 against GRCh38.p13 (HG38) of human Chromosomes 8, 19, and X; in all plots, blocks (i), (ii), and (iii) are non-centromeric repeats in Chromosome 19; blocks (iv) and (v) are the centromeric regions of Chromosomes 8 and X, respectively; (A, C) NucFreq plots illustrating HiFi read mapping coverage; (B, D) CRAQ Circos plots illustrating (a) assembly quality index (AQI) score (higher is better), (b) base errors (lower is better), and (c) misjoins (lower is better).



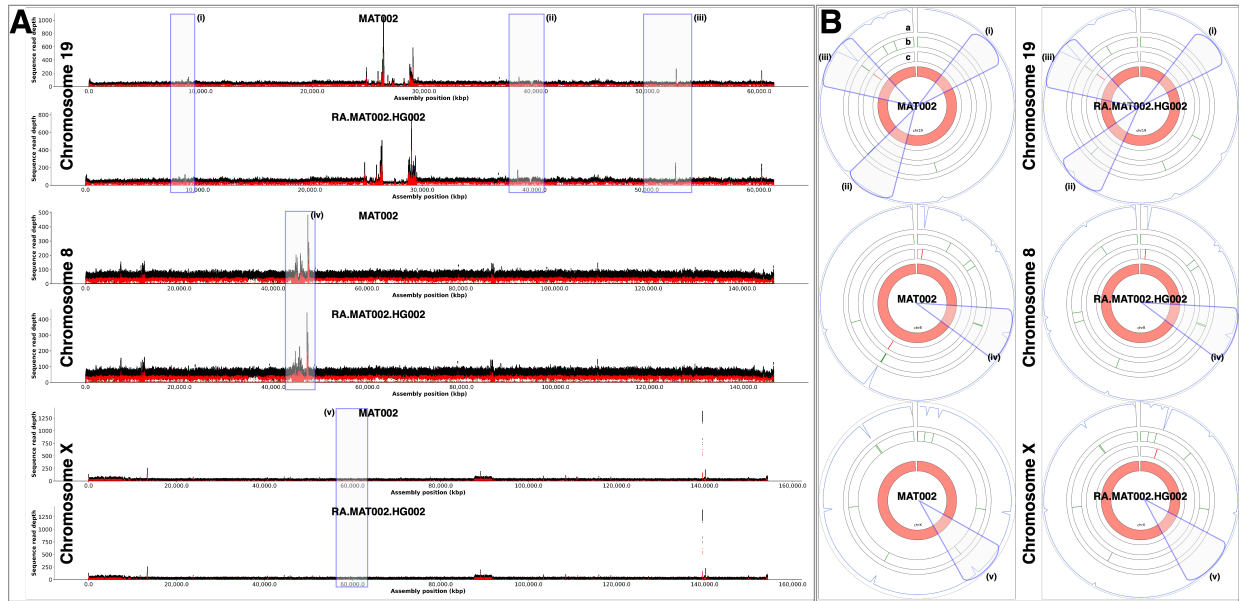
Supplemental Figure 6: NucFlag plots illustrating HiFi read mapping coverage for the comparison of (A) RA.HG38.HG002 assembly and (B) RA.HG38.HG00733 assembly against GRCh38.p13 (HG38) of human Chromosomes 8, 19, and X; in all plots, blocks (i), (ii), and (iii) are non-centromeric repeats in Chromosome 19; blocks (iv) and (v) are the centromeric regions of Chromosomes 8 and X, respectively.



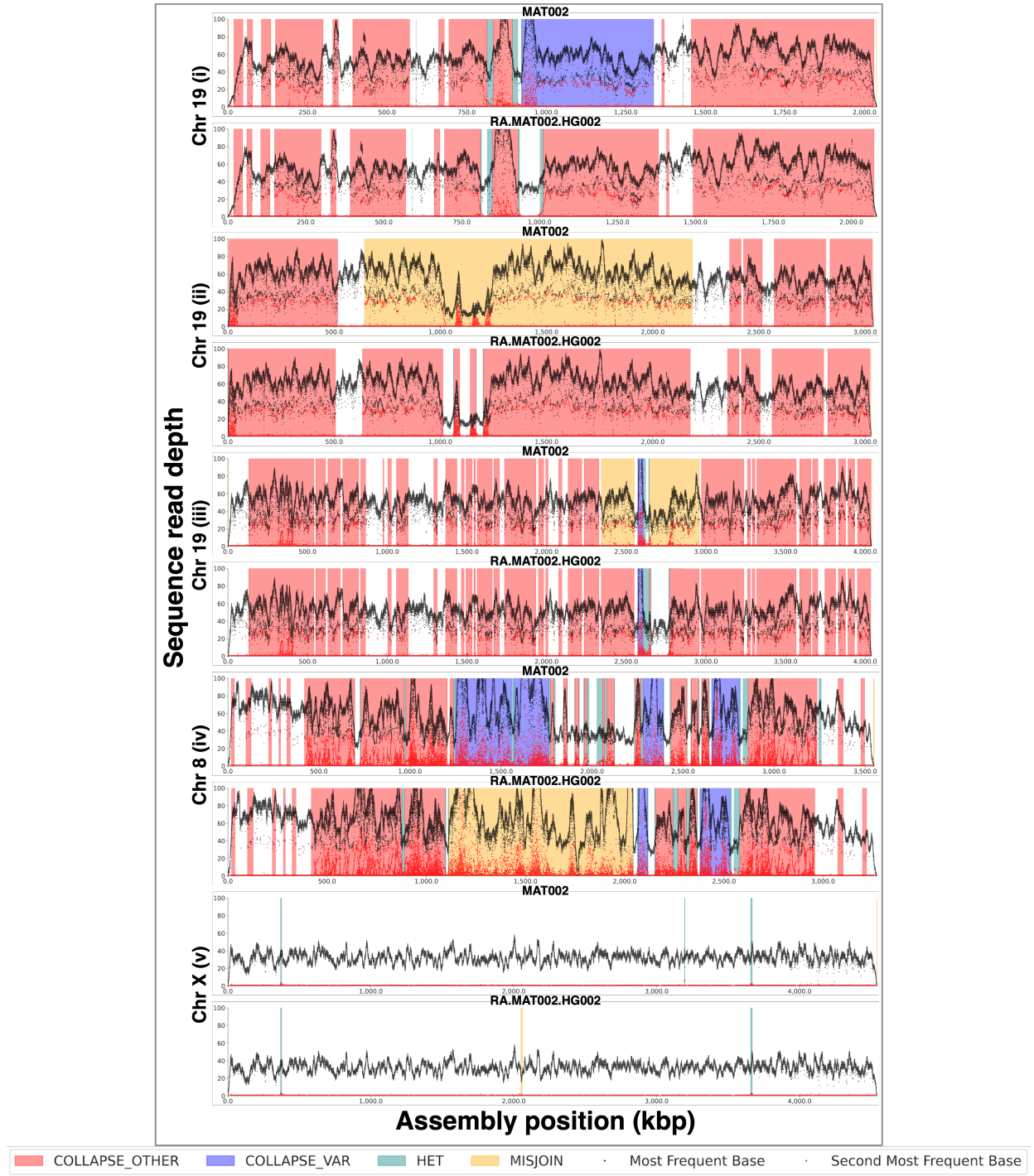
Supplemental Figure 7: Synteny plots of RAMbler's assemblies against GRCh38.p13 assembly (HG38) and the T2T-CHM13.v2.0 assembly (T2T) on human Chromosomes 19, 8, and X using SyRI (A) Synteny plots between HG38 (blue), RA.HG38.HG002 (orange) and T2T (green); (B) Synteny plots between HG38 (blue), RA.HG38.HG00733 (orange) and T2T (green); (C) Synteny plots between HG38 (blue), RA.HG38.HG01346 (orange) and T2T (green).



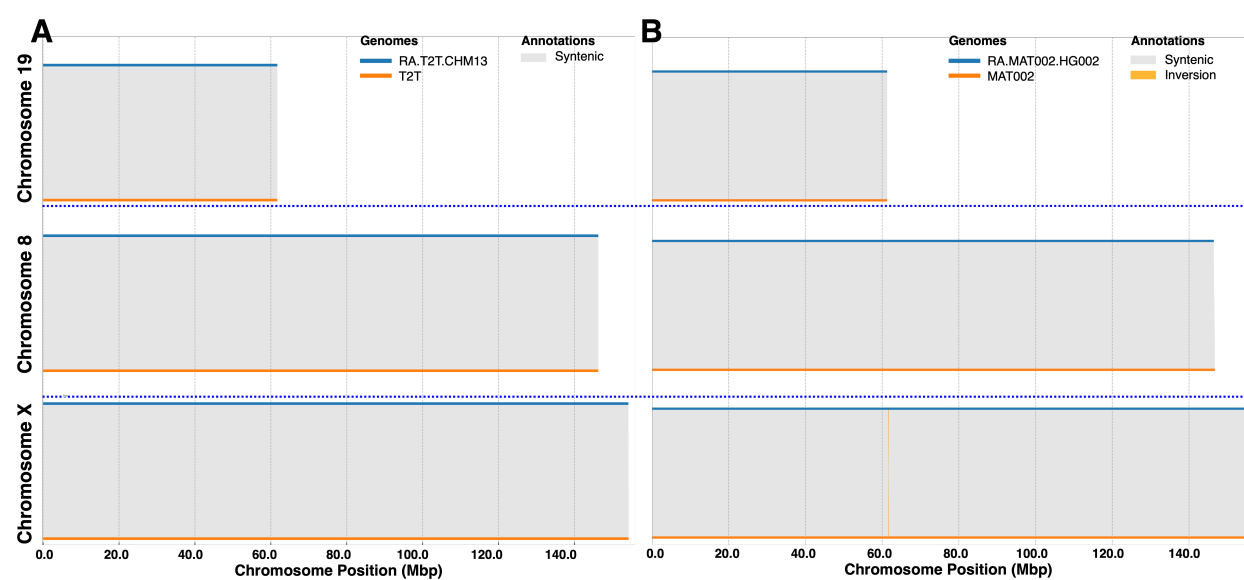
Supplemental Figure 8: Comparing RAMbler's assembly (RA.T2T.CHM13) quality against the T2T-CHM13.v2.0 assembly (T2T) on human Chromosomes 19, 8, and X. Blocks (i), (ii), and (iii) are non-centromeric repeats in Chromosome 19; blocks (iv) and (v) are the centromeric regions of Chromosomes 8 and X (A) NucFreq plots illustrating ultra-long ONT (<https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=T2T/CHM13/nanopore/rel2/>, ~17 $\times$ coverage (≥ 100 kb)) read mapping coverage for T2T (top), and RAMbler (bottom); (B) CRAQ Circos plots illustrating regional AQI scores, base errors, and misjoins.



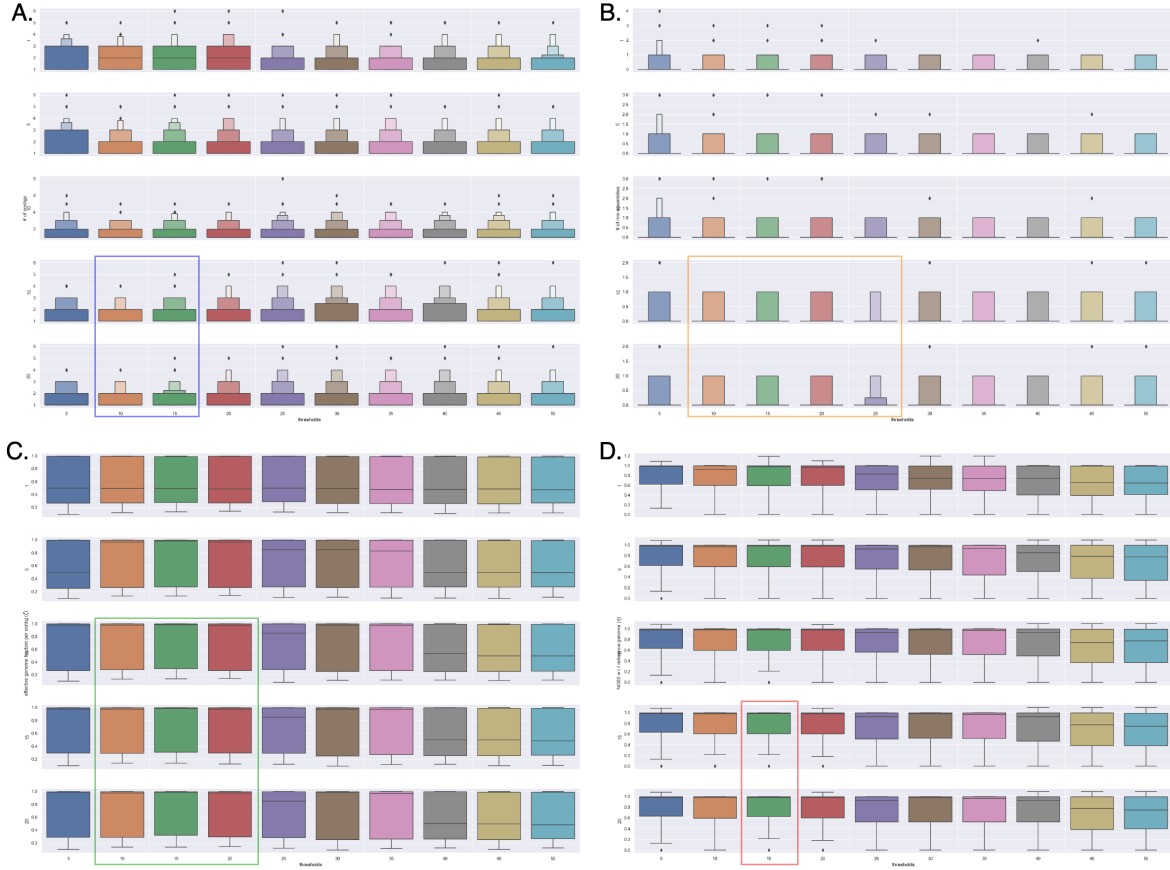
Supplemental Figure 9: Comparing RA.MAT002.HG002 against the maternal HG002 assembly (MAT002) of human Chromosomes 8, 19, and X; in all plots, blocks (i), (ii), and (iii) are non-centromeric repeats in Chromosome 19; blocks (iv) and (v) are the centromeric regions of Chromosomes 8 and X, respectively; **(A)** NucFreq plots illustrating HiFi read mapping coverage; **(B)** CRAQ Circos plots illustrating (a) assembly quality index (AQI) score (higher is better), (b) base errors (fewer is better), and (c) misjoins (fewer is better).



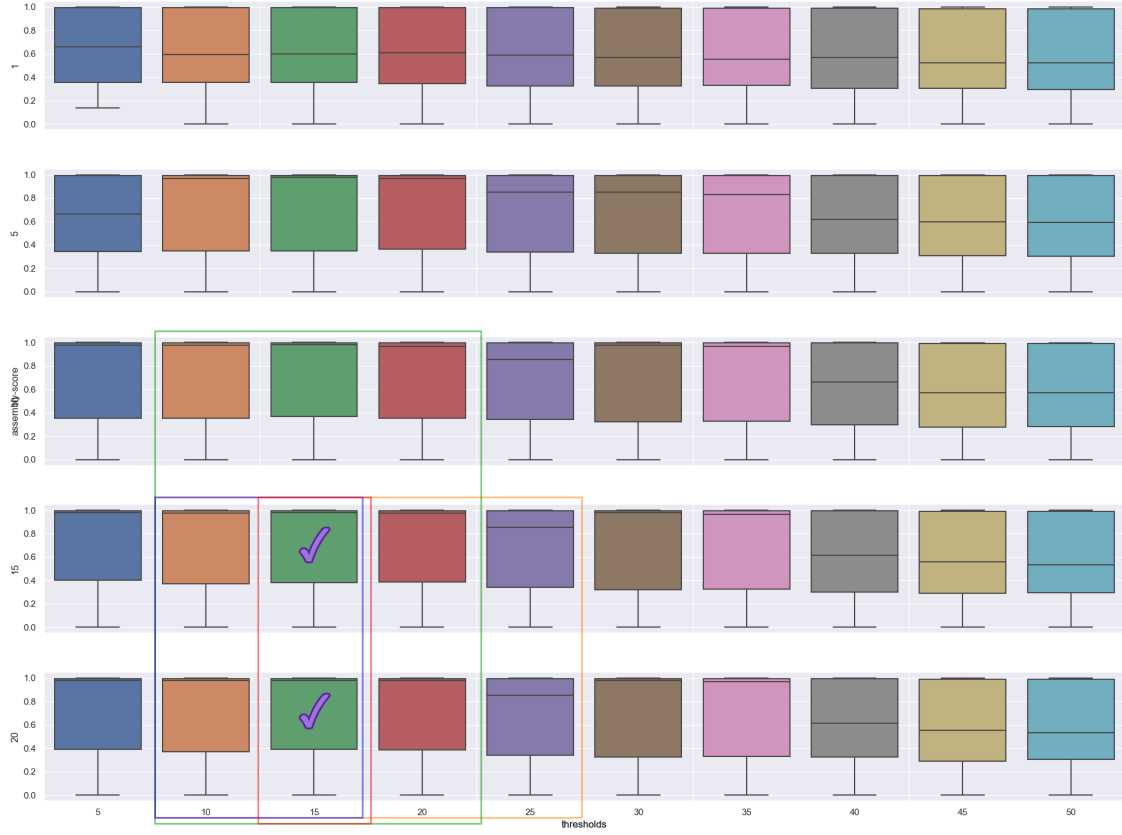
Supplemental Figure 10: NucFlag plots illustrating HiFi read mapping coverage for the comparison of RA.MAT002.HG002 assembly against the maternal HG002 assembly (MAT002) of human Chromosomes 8, 19, and X; in all plots, blocks (i), (ii), and (iii) are non-centromeric repeats in Chromosome 19; blocks (iv) and (v) are the centromeric regions of Chromosomes 8 and X, respectively.



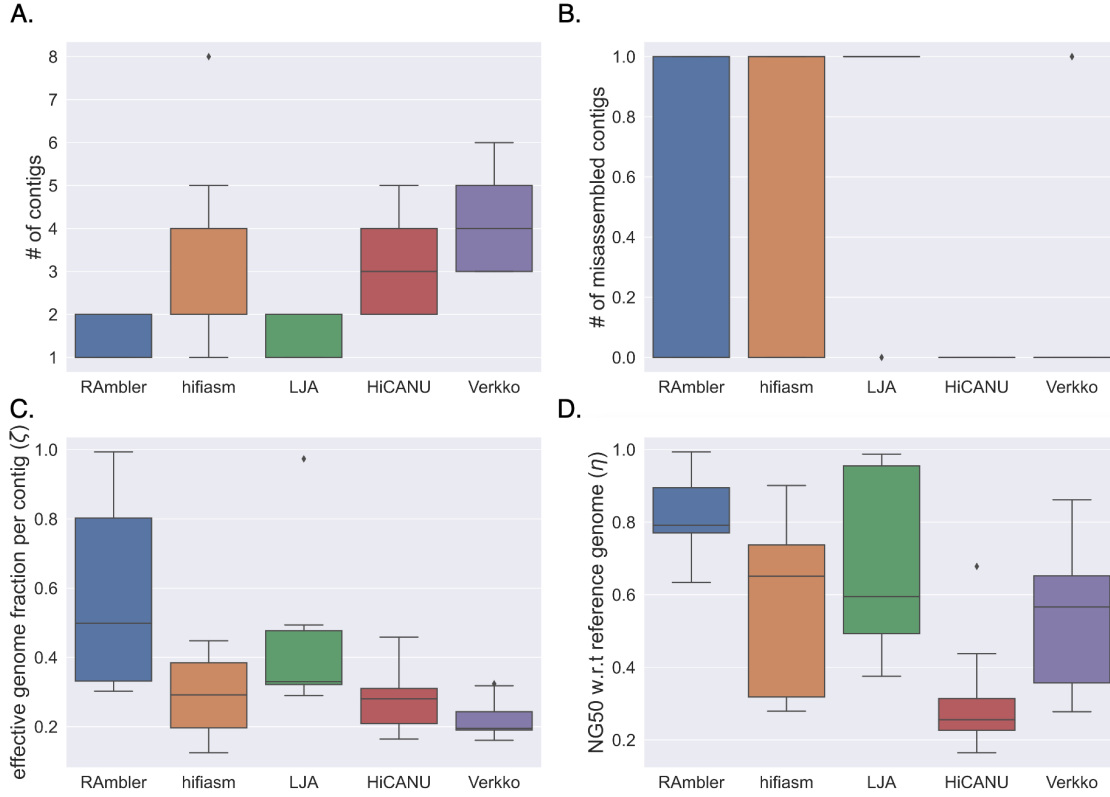
Supplemental Figure 11: Synteny plots of RAMbler's assembly against T2T-CHM13.v2.0 (T2T) and the maternal HG002 assembly (MAT002) on human Chromosomes 19, 8, and X using SyRI (A) Synteny plots between RA.T2T.CHM13 (blue), and T2T (orange); (B) Synteny plots between RA.MAT002.HG002 (blue) and MAT002 (orange).



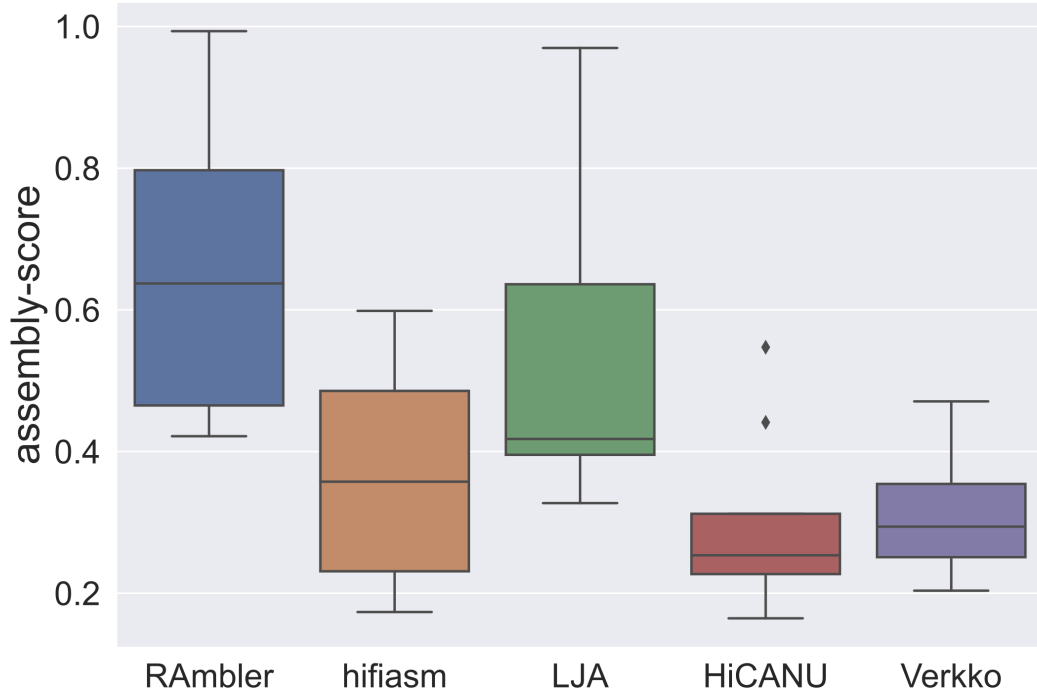
Supplemental Figure 12: Performance of RAMbler for difference combinations of tolerance $to = \{1, 5, 10, 15, 20\}$, threshold, $th = \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$, repeat size $= \{10, 15, 20\}$ kb, number of copies $= \{2, 5, 10\}$, mutation rate $p = \{\frac{1}{100}, \frac{1}{250}, \frac{1}{500}, \frac{1}{1000}, \frac{1}{2000}\}$, on synthetic reads with coverage depth $= \{20\times, 30\times, 40\times\}$; the rows correspond to different values of to , the columns correspond to different values of th ; (A) number of contigs; the best performance is achieved for $to = \{15, 20\}$ and $th = \{10, 15\}$, (B) number of misassembled contigs; the best performance is achieved for $to = \{15, 20\}$ and $th = \{10, 15, 20, 25\}$, (C) effective genome fraction per contig ζ ; the best performance is achieved for $to = \{10, 15, 20\}$ and $th = \{10, 15, 20\}$, and (D) normalized NG50 η ; the best performance is achieved for $to = \{15, 20\}$ and $th = 15$.



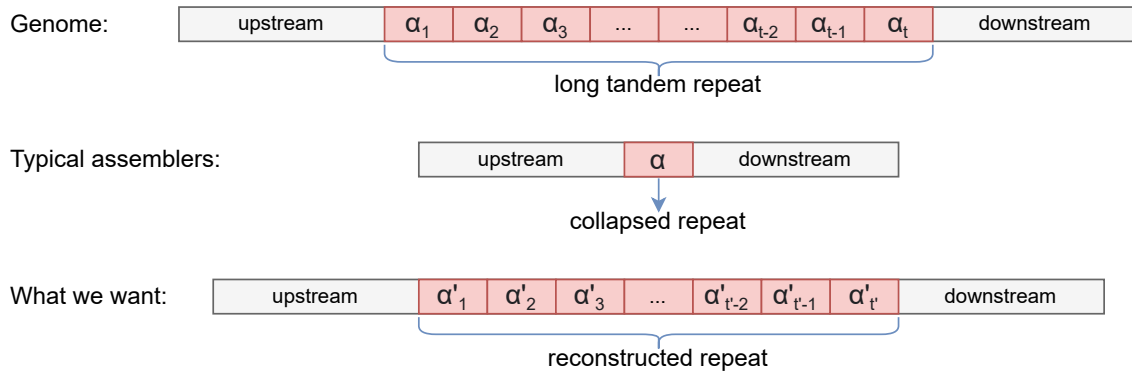
Supplemental Figure 13: Assembly scores achieved by RAMbler for different combinations of tolerance $to = \{1, 5, 10, 15, 20\}$, threshold $th = \{5, 10, 15, 20, 25, 30, 35, 40, 45, 50\}$, repeat size = $\{10, 15, 20\}$ kb, number of copies = $\{2, 5, 10\}$, mutation rate $p = \{\frac{1}{100}, \frac{1}{250}, \frac{1}{500}, \frac{1}{1000}, \frac{1}{2000}\}$, on synthetic reads with coverage depth = $\{20\times, 30\times, 40\times\}$; the rows correspond to different values of to , the columns correspond to different values of th ; the optimal values are indicated by tick marks ($to = \{15, 20\}$ and $th = 15$).



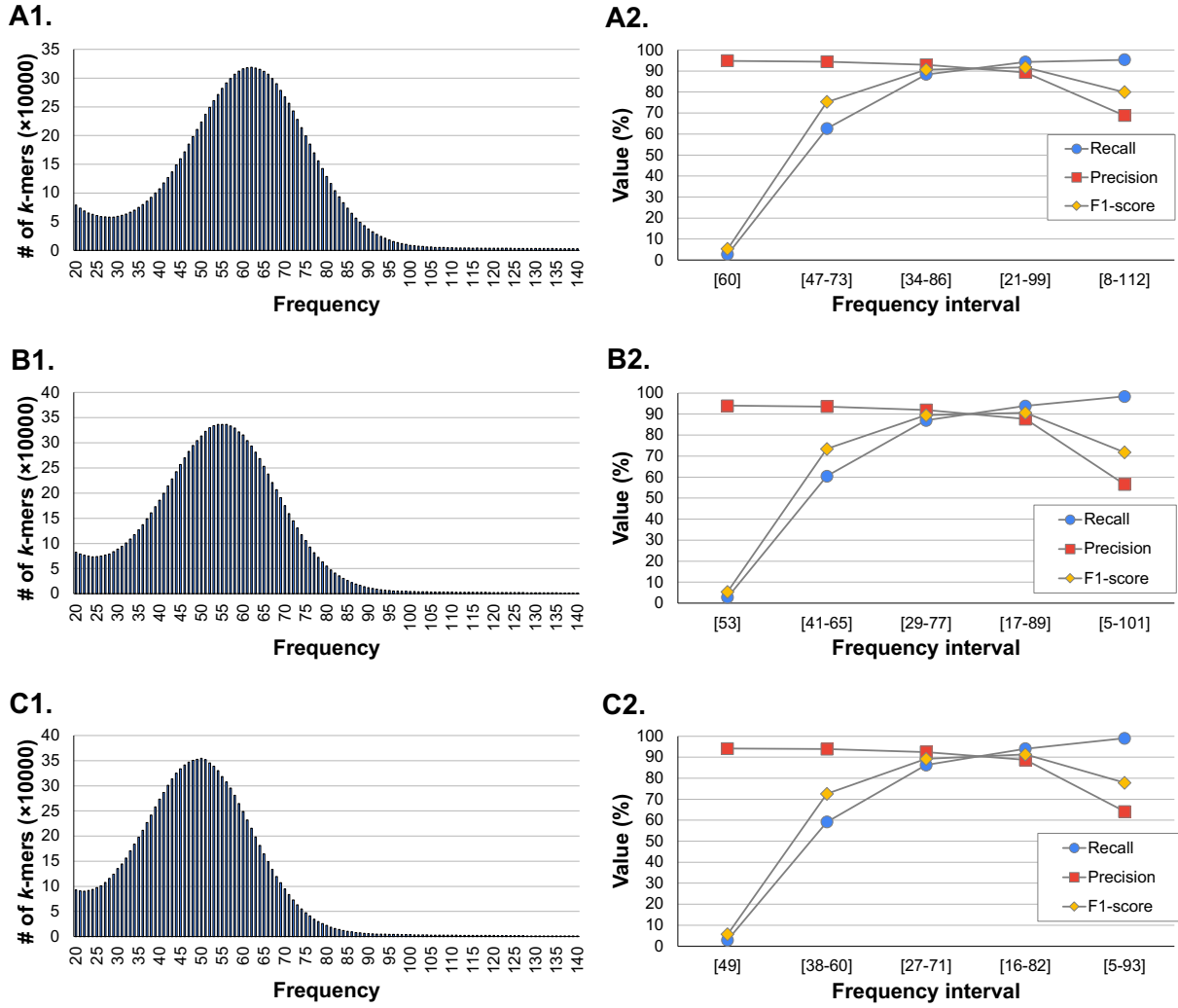
Supplemental Figure 14: Assembly statistics for RAmbler, hifiasm, LJA, HiCANU and Verkko on synthetic data with $30\times$ coverage depth for nine different combinations of repetitive regions containing five copies of a repeat size = $\{15, 20, 25\}$ kb and variable-length (up to $\pm 5\%$ per copy), mutation rate $p = \{\frac{1}{250}, \frac{1}{500}, \frac{1}{1000}\}$; (A) number of contigs, (B) number of misassembled contigs, (C) effective genome fraction per contig ζ , and (D) normalized NG50 η .



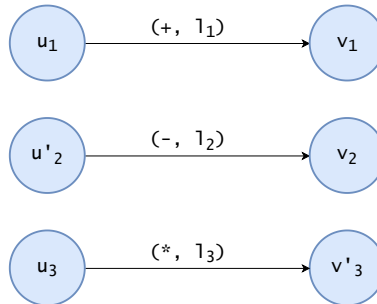
Supplemental Figure 15: Assembly score for RAMbler, hifiasm, LJA, HiCANU and Verkko on synthetic data with $30\times$ coverage depth for nine different combinations of repetitive regions containing five copies of a repeat size = $\{15, 20, 25\}$ kb and variable-length (up to $\pm 5\%$ per copy), mutation rate $p = \{\frac{1}{250}, \frac{1}{500}, \frac{1}{1000}\}$.



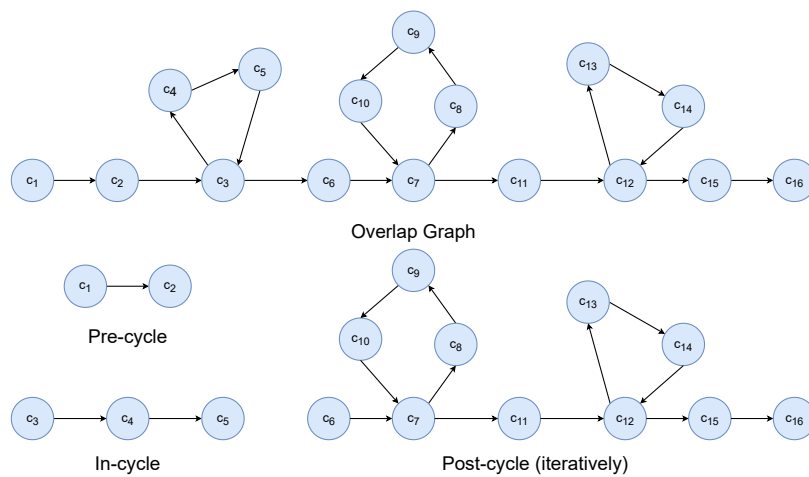
Supplemental Figure 16: The problem formulation: tandem repeats $\alpha_1, \alpha_2, \dots, \alpha_t$ are over-collapsed by traditional assemblers; we want to reconstruct all the copies of the repetitive regions.



Supplemental Figure 17: Oxford Nanopore k -mer distribution for $k = 17$ (A1), $k = 21$ (B1) and $k = 25$ (C1) for *Saccharomyces cerevisiae* dataset (SRA accession SRR18365585, $\sim 50\times$ coverage); precision, recall, and F1-score for unikmers when $k = 17$ (A2), $k = 21$ (B2) and $k = 25$ (C2) for larger and larger intervals centered at the average sequencing coverage.



Supplemental Figure 18: The three types of edges in our overlap graph.



Supplemental Figure 19: Resolving cycles in an overlap graph.

Supplemental Tables

Reference Genome	Sample	Illumina (NGS)		Oxford Nanopore (ONT)		PacBio HiFi (HiFi)	
		Accession/Link	Coverage	Accession/Link	Coverage	Accession/Link	Coverage
T2T-CHM13v2.0	CHM13	SRR10035392	~17×	T2T_rel2 (link)	~50×	SRR11292120-23	~32.4×
HG002v1.0.1	HG002	runs: 1-3, 17 (link)	~13.5×	T2T_rel2 (≥100Kb)	~17×		
GRCh38.p13	HG00733	28 runs (1 2 3 4)	~12×	-	-	link	~66×
	HG01346	SRR769453-54	~10×	-	-	link	~33×
						SRR25618906	~26×

Supplemental Table 1: Data sets for *H. sapiens*.

Organism	Reference Genome	Illumina (NGS)		Oxford Nanopore (ONT)		PacBio HiFi (HiFi)	
		Accession/Link	Coverage	Accession/Link	Coverage	Accession/Link	Coverage
Yeast (<i>S. cerevisiae</i>)	R64 (S288C)	-	-	SRR15597407	~140×	SRR13577847	~85.5×
				SRR18365585	~50×	SRR18210286	~188×

Supplemental Table 2: Data sets for *S. cerevisiae*.

Chromosome statistics			HiFi reads: HG002			HiFi reads: HG00733			HiFi reads: HG01346		
Chr	Length (bp)	Length of SDs (bp)	Coverage	Coverage (SDs masked)	Change (%)	Coverage	Coverage (SDs masked)	Change (%)	Coverage	Coverage (SDs masked)	Change (%)
1	248,956,422	14,814,748	65.75	65.30	0.68	34.07	32.88	3.49	25.83	25.28	2.13
2	242,193,529	10,341,536	69.02	70.67	2.39	33.97	34.75	2.30	26.72	27.30	2.17
3	198,295,559	3,642,738	69.89	71.52	2.33	34.06	34.83	2.26	26.66	27.24	2.18
4	190,214,555	5,059,972	72.27	72.82	0.76	34.49	34.17	0.93	27.66	27.54	0.43
5	181,538,259	5,948,025	70.77	72.84	2.92	33.94	35.01	3.15	26.8	27.61	3.02
6	170,805,979	4,071,190	69.6	71.19	2.28	33.11	33.90	2.39	26.20	26.81	2.33
7	159,345,973	13,686,825	67.68	71.05	4.98	33.6	35.18	4.70	26.37	27.65	4.85
8	145,138,636	3,143,789	67.61	68.10	0.72	32.78	33.07	0.88	25.94	26.18	0.93
9	138,394,717	13,547,270	59.75	60.90	1.92	30.15	30.57	1.39	23.66	24.00	1.44
10	133,797,422	7,574,751	69.74	70.85	1.59	35.86	35.52	0.95	27.60	27.58	0.07
11	135,086,622	5,719,550	67.50	69.48	2.93	34.01	35.05	3.06	26.39	27.18	2.99
12	133,275,309	3,027,056	68.23	70.01	2.61	33.89	34.78	2.63	26.21	26.81	2.29
13	114,364,328	3,363,038	59.80	61.53	2.89	28.27	29.03	2.69	22.60	23.25	2.88
14	107,043,718	2,952,467	57.97	58.58	1.05	28.58	28.95	1.29	22.24	22.62	1.71
15	101,991,189	8,669,292	52.15	54.87	5.22	26.70	27.82	4.19	20.68	21.73	5.08
16	90,338,345	9,086,009	62.23	64.38	3.45	34.45	35.55	3.19	26.25	27.37	4.27
17	83,257,441	7,711,895	61.83	67.13	8.57	33.34	35.81	7.41	24.94	26.58	6.58
18	80,373,285	1,938,507	67.91	69.06	1.69	32.71	33.28	1.74	25.57	26.00	1.68
19	58,617,616	3,985,208	61.11	62.98	3.06	33.61	34.70	3.24	24.35	25.23	3.61
20	64,444,167	2,894,651	71.59	74.02	3.39	37.99	36.60	3.66	28.73	27.61	3.90
21	46,709,983	6,163,779	62.94	61.22	2.73	31.45	29.44	6.39	24.88	23.52	5.47
22	50,818,468	5,996,890	48.68	47.69	2.03	26.12	26.22	0.38	20.20	19.93	1.34
X	156,040,895	10,321,878	34.20	35.21	2.95	31.75	32.66	2.87	25.52	26.29	3.02
Total	3,031,042,417	153,661,064	64.73	66.19	2.26	32.56	33.03	1.44	25.3	25.71	1.62

Supplemental Table 3: Changes in HiFi reads' coverage for each Chromosome when segmental duplications (SD) are masked.

Input assembly	HiFi reads	Coverage	Repetitive region		# Selected reads	Selected reads sum (Mb)	Expected size (Mb)		Assembler				
			start (Mb)	end (Mb)					Rambler	hifiasm	LJA	HiCANU	Verkko
T2T	CHM13	30.5 ×	8	10	3,428	61.473	2.016	# contigs	1	1	1	8	1
								Total (bp)	2,050,034	2,050,034	2,040,669	2,294,767	2,047,043
								Longest contig	2,050,034	2,050,034	2,040,669	2,047,040	2,047,043
			38.5	41.5	5,154	92.475	3.032	# contigs	1	1	1	10	4
								Total (bp)	3,042,992	3,042,992	3,042,941	3,309,519	3,125,757
								Longest contig	3,042,992	3,042,992	3,042,941	3,037,943	2,793,414
MAT002	HG002	60 ×	50	54	6,422	114.784	3.763	# contigs	1	1	1	7	1
								Total (bp)	4,051,224	4,051,222	4,051,225	4,251,698	4,051,206
								Longest contig	4,051,224	4,051,222	4,051,225	4,044,012	4,051,206
			8	10	7,709	123.020	2.050	# contigs	1	3	14	15	23
								Total (bp)	2,082,651	2,149,253	4,034,730	4,275,016	4,213,623
								Longest contig	2,082,651	2,082,652	734,364	2,002,029	751,170
			38.5	41.5	11,249	180.077	3.001	# contigs	1	2	21	14	21
								Total (bp)	3,029,470	3,077,117	5,880,214	6,065,751	5,925,770
								Longest contig	3,029,470	3,041,587	1,395,837	3,023,701	1,581,545
			50	54	13,854	220.561	3.676	# contigs	1	1	23	17	41
								Total (bp)	4,040,445	4,040,444	7,849,165	8,056,324	8,090,995
								Longest contig	4,040,445	4,040,444	1,721,763	4,035,820	1,602,573

Supplemental Table 4: Assembly statistics for Rambler, hifiasm, LJA, HiCANU and Verkko on three non-centromeric repeats of Chromosome 19 from the T2T-CHM13.v2.0 assembly (T2T) and the maternal HG002 assembly (MAT002) using the corresponding HiFi reads as inputs.

Input assembly	HiFi reads	Coverage	Repetitive region		# Selected reads	Selected reads sum (Mb)	Expected size (Mb)		Assembler				
			start (Mb)	end (Mb)					Rambler	hifiasm	LJA	HiCANU	Verkko
HG38	HG002	61 ×	8	10	8,183	130.625	2.141	# contigs	1	1	14	12	23
								Total (bp)	2,203,257	2,225,655	4,344,711	4,496,850	4,509,590
								Longest contig	2,203,257	2,225,655	770,032	2,223,194	780,635
			36	38	8,657	138.825	2.276	# contigs	1	1	9	5	14
								Total (bp)	2,272,627	2,272,627	4,386,424	4,502,066	4,586,729
								Longest contig	2,272,627	2,272,627	1,169,520	1,355,295	1,346,483
			47	51	14,573	232.556	3.812	# contigs	1	2	31	22	53
								Total (bp)	4,283,959	4,313,161	8,347,481	8,639,631	8,658,198
								Longest contig	4,283,959	4,294,647	1,651,643	3,958,242	1,565,100
	HG00733	33 ×	8	10	5,236	70.550	2.138	# contigs	1	1	52	33	72
								Total (bp)	2,214,533	2,214,533	4,221,660	4,370,351	4,541,272
								Longest contig	2,214,533	2,214,533	672,674	2,104,790	538,893
			36	38	5,529	74.748	2.265	# contigs	1	1	27	48	76
								Total (bp)	2,292,968	2,292,968	3,552,638	4,493,447	4,362,704
								Longest contig	2,292,968	2,292,968	697,275	2,287,243	320,798
			47	51	10,283	137.920	4.179	# contigs	2	12	32	31	62
								Total (bp)	4,197,958	7,150,118	8,335,398	8,595,284	8,670,120
								Longest contig	3,217,913	4,215,911	843,864	4,279,895	783,679
	HG01346	24.5 ×	8	10	2,926	53.411	2.180	# contigs	1	1	25	18	31
								Total (bp)	2,190,205	2,198,110	4,198,112	4,460,687	4,395,359
								Longest contig	2,190,205	2,198,110	705,013	1,538,054	580,887
			36	38	2,992	55.055	2.247	# contigs	1	1	21	20	42
								Total (bp)	2,303,078	2,303,078	3,744,185	4,229,499	4,120,267
								Longest contig	2,303,078	2,303,078	531,893	2,296,104	451,056
			47	51	5,216	95.285	3.889	# contigs	1	2	36	28	64
								Total (bp)	4,244,920	4,329,037	8,412,138	8,695,951	8,722,578
								Longest contig	4,244,920	4,271,958	793,309	4,069,987	656,797

Supplemental Table 5: Assembly statistics for Rambler, hifiasm, LJA, HiCANU and Verkko on three non-centromeric repeats of Chromosome 19 using HiFi reads HG002, HG00733, HG01346 and GRCh38.p13 (HG38) as input assembly.

Chr	Input Assembly	Region	# Selected reads	# Common reads	Common (%)
8	HG38	Centromere	15,620	14,392	92.14
	MAT002		14,393		99.99
X	HG38		9,665	8,217	85.02
	MAT002		9,362		87.77
19	HG38	8 Mb - 10 Mb	8,183	7,548	92.24
	MAT002		7,709		97.91
	HG38	36 Mb - 38 Mb	8,657	7,453	86.09
	MAT002	38.5 Mb - 40.5 Mb	7,775		95.86
	HG38	47 Mb - 51 Mb	14,573	12,454	85.46
	MAT002	50 Mb - 54 Mb	13,854		89.89

Supplemental Table 6: Comparing the number of selected HiFi HG002 reads when the GRCh38.p13 assembly (HG38) or the maternal HG002 assembly (MAT002) were used as the input reference assembly to Rambler.

Chromosome	Reference	Query	Metric		
			Genome fraction	Normalized NG50 (η)	# of misassemblies
8	T2T	RA.T2T.CHM13	100.000	1.000	0
		HG38	97.971	0.992	371
		RA.HG38.HG002	97.779	0.996	239
		RA.HG38.HG00733	98.303	0.998	278
		RA.HG38.HG01346	97.813	0.994	239
	MAT002	RA.MAT002.HG002	99.482	0.998	49
		HG38	97.532	0.989	382
		RA.HG38.HG002	98.379	0.993	205
19	T2T	RA.T2T.CHM13	100.000	1.000	0
		HG38	90.706	0.950	101
		RA.HG38.HG002	91.480	0.960	92
		RA.HG38.HG00733	91.458	0.961	98
		RA.HG38.HG01346	91.561	0.960	100
	MAT002	RA.MAT002.HG002	99.940	1.000	3
		HG38	91.233	0.956	119
		RA.HG38.HG002	92.275	0.966	102
X	T2T	RA.T2T.CHM13	99.950	1.000	2
		HG38	98.611	1.012	374
		RA.HG38.HG002	99.126	1.006	205
		RA.HG38.HG00733	98.849	1.004	213
		RA.HG38.HG01346	98.960	1.015	250
	MAT002	RA.MAT002.HG002	100.000	1.000	2
		HG38	98.576	1.011	382
		RA.HG38.HG002	99.595	1.005	169

Supplemental Table 7: QUAST scores for human Chromosomes 8, 19, and X from GRCh38.p13 (HG38) and RAmbler (RA.T2T.CHM13, RA.HG38.HG002, RA.HG38.HG00733 and RA.HG38.HG01346) using T2T-CHM13.v2.0 (T2T) as the ground truth; the length of the chromosomes are 146,259,331 bp, 61,707,364 bp, and 154,259,566 bp, respectively, and from GRCh38.p13 (HG38) and RAmbler (RA.MAT002.HG002 and RA.HG38.HG002) using the maternal strand of HG002 (MAT002) as the ground truth; the length of the chromosomes are 146,740,790 bp, 61,317,360 bp, and 154,341,406 bp respectively.

Chromosome	Assembly	Coverage Rate (%)	R-AQI	S-AQI
19	T2T	99.73	97.91	100
	RA.T2T.CHM13	99.68	98.23	
8	T2T	99.85	99.69	
	RA.T2T.CHM13	99.77	99.49	
X	T2T	99.56	98.83	
	RA.T2T.CHM13	99.64	99.25	

Supplemental Table 8: CRAQ report on assembly coverage and errors obtained using short Illumina reads (SRA accession SRR10035392, $\sim 17\times$ coverage) and PacBio HiFi reads (SRA accession SRR11292120-SRR11292123, $\sim 32.4\times$ coverage) on human Chromosomes 19, 8, and X from T2T-CHM13.v2.0 (T2T) and RAmbler (RA.T2T.CHM13); numbers in bold indicate the best scores.

Chromosome	Assembly	Coverage Rate (%)	R-AQI	S-AQI
19	T2T	99.98	68.60	100
	RA.T2T.CHM13	99.98	72.14	
8	T2T	99.41	84.61	
	RA.T2T.CHM13	99.41	85.57	
X	T2T	99.99	82.03	
	RA.T2T.CHM13	99.99	84.07	

Supplemental Table 9: CRAQ report on assembly coverage and errors obtained using short Illumina reads (SRA accession SRR10035392, $\sim 17\times$ coverage) and ultra-long Oxford Nanopore reads (<https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=T2T/CHM13/nanopore/rel2/>, $\sim 17\times$ coverage (≥ 100 kb)) on human Chromosomes 19, 8, and X from T2T-CHM13.v2.0 (T2T), and RAmbler (RA.T2T.CHM13); numbers in bold indicate the best scores.

Chromosome	Assembly	Coverage Rate (%)	R-AQI	S-AQI
19	MAT002	99.07	98.69	98.37
	RA.MAT002.HG002	99.20	98.86	98.37
8	MAT002	98.97	99.07	98.63
	RA.MAT002.HG002	99.00	99.24	99.31
X	MAT002	99.40	99.29	100
	RA.MAT002.HG002	99.25	99.35	99.35

Supplemental Table 10: CRAQ report on assembly coverage and errors obtained using short Illumina reads ([illuminaHG002](#), $\sim 13.5\times$ coverage) and PacBio HiFi reads ([hifiHG002](#), $\sim 66\times$ coverage) on human Chromosomes 19, 8, and X from the maternal HG002 assembly (MAT002) and RAmbler (RA.MAT002.HG002); numbers in bold indicate the best scores.

Centromere	Input Assembly	HiFi Reads Sample	Coverage	Time (hours)	Memory (GB)
8	T2T	CHM13	$32.4\times$	0.84	31.71
	MAT002	HG002	$66\times$	2.42	45.66
	HG38	HG002	$67.5\times$	4.65	57.80
		HG00733	$32.8\times$	1.18	33.61
		HG01346	$26\times$	1.25	43.02
X	T2T	CHM13	$34\times$	2.45	39.67
	MAT002	HG002	$33\times$	2.28	40.47
	HG38	HG002	$32.2\times$	1.88	44.32
		HG00733	$31.8\times$	6.52	58.38
		HG01346	$25.5\times$	4.69	48.46

Supplemental Table 11: Time and memory usage for RAmbler for the reconstruction of the centromeres of Chromosome 8 and X.

Parameter	Value	Parameter	Value
Mean read length	13,500	Mean read accuracy	0.998
Standard deviation of read length	1,500	Standard deviation of read accuracy	0.01
Minimum read length	7,500	Minimum read accuracy	0.95
Maximum read length	20,000	Maximum read accuracy	1.0

Supplemental Table 12: PBSim's parameters and their values.

	Parameter	Value
k	length of $unik$ mers	21
t	k -mers in the interval $[\mu - t\sigma, \mu + t\sigma]$ of the k -mer distribution considered $unik$ mers	3
th	minimum number of $unik$ mers shared between two reads to declare them overlapping	15
to	maximum allowed tolerance between the position of two consecutive shared $unik$ mers	15
mo	minimum overlap between contigs that is allowed in the overlap graph	1,000 bp

Supplemental Table 13: RAMbler’s main parameters.

Dataset				Assembly score				
Repeat unit size	# of copies	p	Coverage	RAmbler	hifiasm	LJA	HiCANU	Verkko
15 kb	5	1/250	20	0	0.992421	0.362516	0	0.103149
			30	0.993685	0.993682	0.66135	0.279093	0.336791
			40	0.992072	0.996525	0.664114	0.596943	0.472819
		1/500	20	0.442281	0.986855	0.64146	0.236397	0.273222
			30	0.680773	0.548473	0.658083	0.388264	0
			40	0.999324	0.684225	0.661653	0.960054	0.982176
		1/1000	20	0.979912	0.996561	0	0	0.141093
			30	0.999289	0.999281	0.921949	0.402803	0.170338
			40	0.996749	0.996706	0.404907	0.57493	0.212441
	10	1/250	20	0.378988	0.085941	0.213113	0.164032	0.111124
			30	0.381808	0.194383	0.326858	0.566358	0.464188
			40	0.697521	0.164275	0.651712	0.436001	0.645386
		1/500	20	0.503772	0.18682	0	0	0.092472
			30	0.397242	0.214671	0.655464	0.13711	0.170362
			40	0.208417	0.290759	0.64724	0.645192	0
		1/1000	20	0.27603	0.325317	0.395507	0.117697	0.105373
			30	0.246113	0.42346	0.445131	0.166698	0.264898
			40	0.23118	0.423723	0.434755	0.969895	0.984059
20 kb	5	1/250	20	0.993377	0.993362	0.613642	0	0.158968
			30	0.995045	0.995115	0.651152	0.143198	0.320305
			40	0.995057	0.995112	0.414397	0.939415	0
		1/500	20	0.989612	0.99032	0.290029	0	0.228378
			30	0.99039	0.310043	0.406863	0.257214	0.551989
			40	0.998585	0.998585	0.659891	0.962357	0
		1/1000	20	0.994817	0.596527	0.211835	0	0.125832
			30	0.61509	0.255791	0.662906	0.228406	0
			40	0.265227	0.305306	0.661373	0.411638	0.97693
	10	1/250	20	0.321816	0.16656	0	0	0.218633
			30	0.475847	0.399262	0.658377	0.373103	0.139977
			40	0.73109	0.86521	0.457416	0.264454	0.661972
		1/500	20	0.362629	0.077912	0.258614	0	0.090594
			30	0.299438	0.310596	0.414516	0.110915	0.146964
			40	0.336683	0	0.660191	0.374036	0.970193
		1/1000	20	0.187074	0.177739	0.389218	0.087065	0
			30	0.270521	0	0.456605	0.116092	0
			40	0.194154	0	0.559774	0.95271	0

Supplemental Table 14: Assembly score of RAMbler, hifiasm, LJA, HiCANU, and Verkko on 36 synthetic data sets; numbers in bold indicate the best assembly score on each row.

Repeat unit size (bp)	Haplotype 1 genome size (bp)	Haplotype 2		<i>p</i>	Number of contigs					Assembly size (bp)						
		Number of copies	Genome size (bp)		RAmbler	hifiasm		LJA	HiCANU	RAmbler	hifiasm		LJA	HiCANU		
						primary	alternate				primary	alternate				
5,000	125,000	3	115,000	1/250	1	1	1	2	2	124,014	124,044	114,044	236,342	153,146		
				1/500	1	1	1	2	2	123,792	124,733	114,733	238,326	158,925		
				1/1000	1	1	1	2	1	124,516	124,506	114,506	234,154	118,297		
		4	120,000	1/250	1	1	1	1	2	124,772	124,772	119,772	124,610	146,101		
				1/500	1	1	1	2	1	122,954	124,190	119,190	237,137	120,875		
				1/1000	1	1	1	2	2	119,684	124,678	119,678	242,786	155,645		
		5	125,000	1/250	1	1	1	2	3	124,816	124,845	124,845	248,430	167,028		
				1/500	1	1	1	2	2	124,459	124,453	124,453	245,850	172,225		
				1/1000	1	1	1	2	2	124,516	124,509	124,509	247,492	137,479		
		10,000	150,000	3	130,000	1/250	1	1	1	2	2	149,704	149,768	129,768	275,648	196,186
						1/500	1	1	1	2	1	149,460	149,522	129,522	277,752	124,994
						1/1000	1	1	1	2	2	149,938	149,894	129,894	279,752	178,305
4	140,000			1/250	1	1	1	2	2	149,740	149,734	139,734	281,318	199,970		
				1/500	1	1	1	2	2	149,272	149,277	139,277	286,966	203,767		
				1/1000	1	1	1	2	2	149,606	149,606	139,606	287,786	183,662		
5	150,000			1/250	1	1	1	1	2	149,824	149,827	149,827	149,599	170,422		
				1/500	1	1	1	2	4	149,686	149,747	149,747	246,773	203,411		
				1/1000	1	1	1	6	4	148,757	149,564	149,564	259,079	211,558		
15,000	175,000			3	145,000	1/250	1	1	1	2	4	174,718	174,719	144,719	314,955	244,935
						1/500	1	1	1	2	8	174,012	174,557	144,557	303,079	255,277
						1/1000	1	1	1	2	5	174,244	174,304	144,304	318,018	231,616
		4	160,000	1/250	1	1	1	3	4	174,683	174,649	159,649	333,299	243,259		
				1/500	1	1	1	1	5	174,163	174,166	159,166	173,611	246,750		
				1/1000	1	1	1	3	3	174,696	174,683	159,683	275,311	248,097		
		5	175,000	1/250	1	1	1	2	6	173,791	173,792	173,792	234,321	253,340		
				1/500	1	1	1	5	3	174,765	174,847	174,847	301,644	224,413		
				1/1000	1	1	1	4	3	219,085	174,650	174,650	242,524	249,085		
		20,000	200,000	3	160,000	1/250	1	1	1	2	5	199,897	199,884	159,884	358,247	282,097
						1/500	1	1	1	2	3	199,853	199,854	159,854	358,946	283,088
						1/1000	1	1	1	5	9	238,526	238,525	72,170	289,383	352,156
4	180,000			1/250	1	1	1	2	3	199,378	199,490	179,490	334,210	247,370		
				1/500	1	1	1	2	7	198,949	199,456	179,456	367,218	295,081		
				1/1000	1	1	1	4	8	199,560	199,562	179,562	311,928	358,247		
5	200,000			1/250	1	1	1	3	3	199,555	199,555	199,555	397,504	268,243		
				1/500	1	1	1	2	4	199,349	199,344	199,344	397,630	233,761		
				1/1000	1	1	2	4	6	199,723	219,716	202,723	376,751	353,151		
50,000	350,000			3	250,000	1/250	1	1	1	4	5	349,544	349,540	249,541	619,973	433,210
						1/500	1	1	1	3	6	399,665	349,654	249,654	550,810	389,677
						1/1000	1	1	1	12	14	448,922	349,542	249,542	556,744	616,291
		4	300,000	1/250	1	1	1	2	13	349,296	349,335	299,335	599,897	550,346		
				1/500	1	1	1	3	7	349,649	349,679	299,679	661,998	534,924		
				1/1000	1	2	1	6	8	349,634	364,122	299,685	572,956	533,915		
		5	350,000	1/250	1	1	1	4	19	349,651	349,646	349,646	704,151	667,578		
				1/500	1	1	1	4	14	449,784	449,965	183,554	655,052	457,076		
				1/1000	1	1	1	5	16	349,220	349,219	308,939	611,751	656,526		
		100,000	600,000	3	400,000	1/250	1	1	1	6	15	598,560	599,235	399,237	1,035,574	803,380
						1/500	1	1	2	5	18	599,336	699,758	310,632	983,101	802,113
						1/1000	1	2	1	17	24	599,644	615,325	399,668	963,902	867,330
4	500,000			1/250	1	1	1	4	15	599,834	599,831	499,831	1,111,210	829,144		
				1/500	1	1	2	6	16	599,322	599,323	522,346	1,131,188	832,552		
				1/1000	1	2	2	10	22	599,015	410,687	679,805	1,010,779	987,162		
5	600,000			1/250	1	1	1	3	14	599,784	599,779	599,779	1,143,983	930,005		
				1/500	1	1	1	5	23	599,418	699,418	322,057	1,225,788	1,032,964		
				1/1000	1	3	2	10	28	699,614	565,106	640,252	1,215,171	1,114,022		

Supplemental Table 15: Number of contigs and assembly size for the assemblies produced by RAmbler, hifiasm, LJA, and HiCANU; assemblers were given in input synthetic HiFi reads with coverage 30× generated from synthetic diploid genomes; in each synthetic genome the primary haplotype (hap1) had five copies of the repeat unit, while the secondary haplotype (hap2) had 3-5 copies; Verkko was omitted because it failed to produce any output in most cases.

Repeat unit size (bp)	Haplotype 1 genome size (bp)	Haplotype 2		<i>p</i>	Number of contigs					Assembly size (bp)				
		Number of copies	Genome size (bp)		RAmbler	hifiasm		LJA	HiCANU	RAmbler	hifiasm		LJA	HiCANU
						primary	alternate				primary	alternate		
5,000	150,000	6	130,000	1/250	1	1	1	2	2	149,798	149,807	129,807	278,735	204,153
				1/500	1	1	1	2	2	149,600	149,584	129,584	273,803	175,265
				1/1000	1	1	1	3	3	149,821	149,816	129,816	157,944	185,932
		8	140,000	1/250	1	1	1	2	4	149,814	149,814	139,814	281,722	204,717
				1/500	1	1	1	2	3	149,598	149,700	139,700	287,414	160,116
				1/1000	1	1	1	4	3	149,146	149,698	139,699	213,196	154,213
		10	150,000	1/250	1	1	1	2	3	149,520	149,522	149,522	294,735	209,336
				1/500	2	1	1	2	3	157,601	149,596	149,596	299,977	177,134
				1/1000	1	1	1	2	4	149,533	149,534	149,534	251,604	218,802
10,000	200,000	6	160,000	1/250	1	1	1	3	3	198,907	198,943	158,943	356,387	284,989
				1/500	1	1	1	3	3	208,918	198,966	158,966	313,973	270,487
				1/1000	1	1	1	6	4	198,803	198,799	158,799	367,402	277,901
		8	180,000	1/250	1	1	1	2	3	199,909	199,903	179,903	259,132	313,321
				1/500	1	1	1	2	3	259,751	199,808	179,808	334,697	250,747
				1/1000	1	1	1	3	5	159,892	199,802	179,802	280,797	338,852
		10	200,000	1/250	1	1	1	2	8	199,143	199,138	199,138	286,264	311,887
				1/500	1	1	1	3	6	199,387	199,386	199,386	342,217	329,246
				1/1000	1	1	1	2	12	199,921	199,912	199,912	343,594	406,570
15,000	250,000	6	190,000	1/250	1	1	1	5	5	249,666	249,666	189,666	453,741	336,368
				1/500	1	1	1	3	8	249,774	249,774	189,783	454,197	364,820
				1/1000	2	1	2	3	8	218,440	264,806	204,702	318,102	374,178
		8	220,000	1/250	1	1	1	2	5	249,801	249,801	219,801	464,700	404,715
				1/500	1	1	1	2	8	219,761	309,760	115,902	461,678	371,045
				1/1000	2	1	1	10	7	393,419	249,891	219,891	463,866	399,737
		10	250,000	1/250	1	1	1	4	3	263,324	369,706	83,795	479,210	394,700
				1/500	1	1	1	2	3	219,559	249,554	249,554	495,509	387,150
				1/1000	3	1	0	7	10	235,122	324,922	0	437,369	420,817
20,000	300,000	6	220,000	1/250	1	1	1	4	4	398,624	298,623	218,623	534,817	373,343
				1/500	1	1	1	3	6	299,828	299,831	219,831	445,980	364,339
				1/1000	1	1	1	3	4	298,587	263,652	219,618	389,514	388,469
		8	260,000	1/250	1	1	1	2	10	398,674	299,601	259,601	559,192	468,951
				1/500	1	1	1	6	7	299,403	299,466	259,466	504,472	467,505
				1/1000	3	1	1	5	5	296,036	299,811	259,811	502,147	392,303
		10	300,000	1/250	1	1	1	5	12	299,429	299,473	299,479	549,615	495,926
				1/500	1	1	1	2	10	319,347	299,345	299,345	594,195	451,130
				1/1000	1	1	2	13	19	239,745	419,694	199,766	466,960	556,297
50,000	600,000	6	400,000	1/250	1	1	2	6	10	749,565	749,564	261,223	994,336	715,594
				1/500	1	1	1	7	13	598,937	598,942	398,942	925,192	726,548
				1/1000	2	1	1	15	26	679,365	614,317	399,952	976,037	1,077,795
		8	500,000	1/250	1	1	1	10	18	598,481	599,775	499,775	1,098,032	817,291
				1/500	1	1	1	2	16	599,795	599,791	499,791	1,097,101	838,820
				1/1000	1	1	1	19	35	599,763	599,764	499,764	1,018,018	1,246,016
		10	600,000	1/250	3	1	1	8	18	845,259	598,997	598,997	1,230,662	998,718
				1/500	1	1	1	8	22	599,707	599,707	599,707	1,207,967	1,165,785
				1/1000	1	3	1	32	43	749,327	614,349	560,593	1,104,150	1,344,584
100,000	1,100,000	6	700,000	1/250	1	1	1	7	31	1,399,769	1,099,766	699,766	1,661,520	1,384,759
				1/500	1	1	1	4	29	1,098,917	1,099,745	699,742	1,761,361	1,298,079
				1/1000	3	2	3	55	42	1,021,424	971,330	946,443	1,910,184	1,912,471
		8	900,000	1/250	1	1	1	10	23	1,598,393	1,099,455	899,455	1,927,926	1,546,332
				1/500	2	1	1	11	26	1,146,069	1,098,945	898,945	1,978,251	1,432,989
				1/1000	1	1	1	38	61	1,399,349	1,599,474	899,474	1,971,769	2,237,345
		10	1,100,000	1/250	1	1	1	13	33	1,099,462	1,099,460	1,099,460	2,199,749	1,822,173
				1/500	1	1	1	7	37	1,099,255	1,199,265	934,535	2,256,628	1,942,282
				1/1000	2	1	1	26	55	376,999	1,299,672	1,099,672	2,352,688	2,362,493

Supplemental Table 16: Number of contigs and assembly size for the assemblies produced by RAmbler, hifiasm, LJA, and HiCANU; assemblers were given in input synthetic HiFi reads with coverage 30× generated from synthetic diploid genomes; in each synthetic genome the primary haplotype (hap1) had ten copies of the repeat unit, while the secondary haplotype (hap2) had 6,8 or 10 copies; Verkko was omitted because it failed to produce any output in most cases.

Repeat unit size (bp)	# of copies in hap2	p	RAmbler				hifiasm (primary)				hifiasm (alternate)				LJA				HiCANU			
			number of copies resolved			# haplotype switches	number of copies resolved			# haplotype switches	number of copies resolved			# haplotype switches	number of copies resolved			# haplotype switches	number of copies resolved			# haplotype switches
			hap1	hap2	total		hap1	hap2	total		hap1	hap2	Total		hap1	hap2	total		hap1	hap2	total	
5,000	3	1/250	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	5	3	8	0
		1/500	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	5	3	8	0
		1/1000	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	2	2	4	1
		1/250	5	0	5	0	5	0	5	0	0	4	4	0	5	0	5	0	5	1	6	0
	4	1/500	5	0	5	0	5	0	5	0	0	4	4	0	5	3	8	0	5	0	5	0
		1/1000	0	4	4	0	5	0	5	0	0	4	4	0	4	3	7	0	5	4	9	0
		1/250	0	5	5	0	0	5	5	0	5	0	5	0	5	5	10	0	4	5	9	1
		1/500	0	5	5	0	0	5	5	0	5	0	5	0	5	4	9	0	4	4	8	0
	5	1/1000	0	5	5	0	0	5	5	0	5	0	5	0	5	5	10	0	1	5	6	0
		1/250	5	0	5	0	5	0	5	0	0	3	3	0	4	3	7	0	5	3	8	0
		1/500	5	0	5	0	5	0	5	0	0	3	3	0	3	3	6	0	0	3	3	0
		1/1000	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	1	3	4	0
10,000	3	1/250	5	0	5	0	5	0	5	0	0	4	4	0	5	4	9	0	5	4	9	0
		1/500	5	0	5	0	5	0	5	0	0	4	4	0	5	4	9	0	5	2	7	0
		1/1000	5	0	5	0	5	0	5	0	0	4	4	0	4	3	7	0	3	3	6	0
		1/250	0	5	5	0	0	5	5	0	5	0	5	0	5	0	5	0	5	1	6	0
	4	1/500	5	0	5	0	0	5	5	0	5	0	5	0	4	4	8	0	4	2	6	1
		1/1000	0	5	5	0	0	5	5	0	5	0	5	0	5	5	10	1	3	4	7	0
		1/250	5	0	5	0	5	0	5	0	0	3	3	0	4	3	7	0	5	3	8	1
		1/500	5	0	5	0	5	0	5	0	0	3	3	0	3	1	4	0	1	1	2	0
	5	1/1000	5	0	5	0	5	0	5	0	0	3	3	0	4	3	7	0	1	0	1	0
		1/250	5	0	5	0	5	0	5	0	0	4	4	0	5	4	9	0	4	1	5	0
		1/500	5	0	5	0	5	0	5	0	0	4	4	0	4	0	4	0	2	2	4	0
		1/1000	5	0	5	0	5	0	5	0	0	4	4	0	5	4	9	0	4	2	6	0
15,000	3	1/250	0	5	5	0	5	0	5	0	0	5	5	0	2	4	6	0	2	3	5	0
		1/500	5	0	5	0	5	0	5	0	0	5	5	0	5	5	10	2	1	1	2	1
		1/1000	4	4	8	1	0	5	5	0	5	0	5	0	5	5	10	2	4	4	8	1
		1/250	5	0	5	0	5	0	5	0	0	3	3	0	4	2	6	0	5	1	6	1
	4	1/500	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	5	3	8	0
		1/1000	4	3	7	1	4	2	6	1	0	2	2	0	5	3	8	0	1	2	3	1
		1/250	5	0	5	0	5	0	5	0	0	4	4	0	5	3	8	0	3	4	7	0
		1/500	3	2	5	1	5	0	5	0	0	4	4	0	5	4	9	0	4	2	6	0
	5	1/1000	5	0	5	0	5	0	5	0	0	4	4	0	5	4	9	0	5	0	5	0
		1/250	0	5	5	0	0	5	5	0	5	0	5	0	5	3	8	0	3	2	5	0
		1/500	0	5	5	0	0	5	5	0	5	0	5	0	3	5	8	0	2	2	4	0
		1/1000	5	0	5	0	3	2	5	1	4	4	8	0	4	4	8	1	5	4	9	1
20,000	3	1/250	5	0	5	0	5	0	5	0	0	3	3	0	4	2	6	0	5	1	6	1
		1/500	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	5	3	8	0
		1/1000	4	3	7	1	4	2	6	1	0	2	2	0	5	3	8	0	1	2	3	1
		1/250	5	0	5	0	5	0	5	0	0	4	4	0	5	3	8	0	3	4	7	0
	4	1/500	3	2	5	1	5	0	5	0	0	4	4	0	5	4	9	0	4	2	6	0
		1/1000	5	0	5	0	5	0	5	0	0	4	4	0	5	4	9	0	5	0	5	0
		1/250	0	5	5	0	0	5	5	0	5	0	5	0	5	3	8	0	3	2	5	0
		1/500	0	5	5	0	0	5	5	0	5	0	5	0	3	5	8	0	2	2	4	0
	5	1/1000	5	0	5	0	3	2	5	1	4	4	8	0	4	4	8	1	5	4	9	1
		1/250	5	0	5	0	5	0	5	0	0	3	3	0	2	2	4	0	3	2	5	0
		1/500	4	2	6	1	5	0	5	0	0	3	3	0	4	2	6	0	1	1	2	0
		1/1000	5	2	7	1	5	0	5	0	0	3	3	0	4	3	7	0	3	2	5	0
50,000	3	1/250	5	0	5	0	5	0	5	0	0	4	4	0	4	3	7	0	1	2	3	0
		1/500	5	0	5	0	5	0	5	0	0	4	4	0	5	3	8	0	4	1	5	0
		1/1000	5	0	5	0	5	0	5	0	0	4	4	0	5	3	8	0	1	2	3	0
		1/250	5	0	5	0	5	0	5	0	0	4	4	0	5	3	8	0	4	1	5	0
	4	1/500	5	0	5	0	5	0	5	0	0	4	4	0	5	3	8	0	1	2	3	0
		1/1000	5	0	5	0	5	0	5	0	0	4	4	0	5	3	8	0	1	2	3	0
		1/250	5	0	5	0	5	0	5	0	0	5	5	0	4	5	9	0	2	1	3	0
		1/500	4	3	7	1	2	4	6	1	3	0	3	0	5	5	10	0	5	5	10	1
	5	1/1000	5	0	5	0	0	5	5	0	5	0	5	0	1	2	3	0	1	3	4	0
		1/250	5	0	5	0	5	0	5	0	0	3	3	0	4	2	6	0	5	3	8	1
		1/500	5	0	5	0	5	1	6	1	5	3	8	1	5	3	8	0	1	1	2	0
		1/1000	5	0	5	0	5	0	5	0	0	3	3	0	4	3	7	0	1	2	4	0
100,000	3	1/250	5	0	5	0	5	0	5	0	0	4	4	0	2	4	6	0	2	3	5	0
		1/500	5	0	5	0	5	0	5	0	0	4	4	0	4	3	7	0	1	1	2	0
		1/1000	5	0	5	0	5	0	5	0	0	4	4	0	4	4	8	0	0	1	1	0
		1/250	0	5	5	0	0	5	5	0	5	0	5	0	3	4	7	0	3	3	6	0
	4	1/500	5	0	5	0	3	2	5	1	0	3	3	0	3	4	7	0	1	0	1	0
		1/1000	4	2	6	1	5	5	10	1	5	2	7	0	4	5	9	0	5	5	10	1
		1/250	5	0	5	0	5	0	5	0	0	3	3	0	4	3	7	0	5	3	8	0
		1/500	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	5	3	8	0
	5	1/1000	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	5	3	8	0
		1/250	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	5	3	8	0
		1/500	5	0	5	0	5	0	5	0	0	3	3	0	5	3	8	0	5	3	8	0
		1/1000	4	2	6	1	5	5	10	1	5	2	7	0	4	5	9	0	5	5	10	1

Supplemental Table 17: Number of repeat copies resolved and number of haplotype switching in the assemblies produced by RAmbler, hifiasm, LJA, and HiCANU; assemblers were given in input synthetic HiFi reads with coverage 30× generated from synthetic diploid genomes; in each synthetic genome the primary haplotype (hap1) had five copies of the repeat unit, while the secondary haplotype (hap2) had 3-5 copies; red numbers indicate incorrect repeat copy counts (neither matching either haplotype copy counts nor their sum); blue numbers indicate copy counts that match the sum of copies for both haplotypes (hap1+hap2); grey cells indicate assemblies with haplotype switches; Verkko was omitted because it failed to produce any output in most cases.

Repeat unit size (bp)	# of copies in hap2	<i>p</i>	RAmbler				hifiasm (primary)				hifiasm (alternate)				LJA				HiCANU			
			number of copies resolved			# haplotype switches	number of copies resolved			# haplotype switches	number of copies resolved			# haplotype switches	number of copies resolved			# haplotype switches	number of copies resolved			# haplotype switches
			hap1	hap2	total		hap1	hap2	total		hap1	hap2	Total		hap1	hap2	total		hap1	hap2	total	
5,000	6	1/250	10	0	10	0	10	0	10	0	0	6	6	0	9	6	15	0	10	6	16	0
		1/500	10	0	10	0	10	0	10	0	0	6	6	0	9	4	13	0	10	2	12	0
		1/1000	10	0	10	0	10	0	10	0	0	6	6	0	10	6	16	1	9	3	12	1
		1/250	10	0	10	0	10	0	10	0	0	8	8	0	9	6	15	0	10	3	13	0
	8	1/500	10	0	10	0	10	0	10	0	0	8	8	0	10	7	17	0	3	3	6	1
		1/1000	6	4	10	1	8	7	15	1	6	8	14	1	10	8	18	4	7	7	14	2
		1/250	10	0	10	0	10	0	10	0	0	10	10	0	10	9	19	0	10	7	17	0
		1/500	7	6	13	0	10	0	10	0	0	10	10	0	9	10	19	0	1	10	11	1
	10	1/1000	0	10	10	0	0	10	10	0	10	0	10	0	10	10	20	0	5	7	12	0
		1/250	10	0	10	0	10	0	10	0	0	6	6	0	7	6	13	0	8	6	14	0
10,000	6	1/500	10	1	11	1	10	0	10	0	0	6	6	0	6	3	9	0	7	4	11	0
		1/1000	10	0	10	0	10	0	10	0	0	6	6	0	10	6	16	4	7	3	10	0
		1/250	10	0	10	0	10	0	10	0	0	8	8	0	6	5	11	0	10	6	16	0
		1/500	10	6	16	1	10	0	10	0	0	8	8	0	8	8	16	0	4	6	10	0
	8	1/1000	2	3	5	1	8	2	10	1	3	5	8	1	10	8	18	1	7	6	13	1
		1/250	10	0	10	0	0	10	10	0	10	0	10	0	7	8	15	0	6	8	14	0
		1/500	0	10	10	0	10	0	10	0	0	10	10	0	6	8	14	0	8	8	16	0
		1/1000	10	0	10	0	10	0	10	0	0	10	10	0	9	10	19	0	6	8	14	1
	10	1/250	10	0	10	0	10	0	10	0	0	6	6	0	10	6	16	0	6	6	12	0
		1/500	10	0	10	0	10	0	10	0	0	6	6	0	9	6	15	0	6	4	10	0
15,000	6	1/1000	3	6	9	0	7	4	11	1	5	2	7	1	10	4	14	2	10	6	16	1
		1/250	10	0	10	0	10	0	10	0	0	8	8	0	8	4	12	0	10	5	15	0
		1/500	0	8	8	0	9	4	13	1	0	5	5	0	10	7	17	0	10	8	18	2
		1/1000	10	7	17	0	10	0	10	0	0	8	8	0	10	7	17	4	8	5	13	1
	8	1/250	10	1	11	1	9	8	17	1	0	3	3	0	9	6	15	0	9	6	15	0
		1/500	0	8	8	0	0	9	9	0	8	0	8	0	9	9	18	0	7	8	15	0
		1/1000	2	4	6	0	8	6	14	1	0	0	0	0	10	10	20	2	10	10	20	2
		1/250	10	5	15	1	9	0	9	0	0	6	6	0	7	6	13	0	7	3	10	0
	10	1/500	10	0	10	0	6	4	10	1	4	2	6	1	9	4	13	0	4	5	9	0
		1/1000	10	0	10	0	7	3	10	1	3	3	6	1	9	6	15	0	6	5	11	1
20,000	6	1/250	10	5	15	1	10	0	10	0	0	8	8	0	8	6	14	2	4	4	8	0
		1/500	10	0	10	0	9	2	11	1	0	6	6	0	9	8	17	0	6	7	13	0
		1/1000	3	3	6	0	8	0	8	0	2	8	10	1	10	8	18	1	5	5	10	0
		1/250	10	0	10	0	0	10	10	0	10	0	10	0	7	7	14	0	10	10	20	3
	8	1/500	2	9	11	1	10	0	10	0	0	10	10	0	10	10	20	0	5	2	7	1
		1/1000	7	0	7	0	7	8	15	1	10	10	20	1	10	10	20	2	10	10	20	3
		1/250	10	3	13	1	9	3	12	1	1	3	4	1	6	4	10	0	5	2	7	1
		1/500	10	0	10	0	10	0	10	0	0	6	6	0	10	4	14	1	4	2	6	0
	10	1/1000	9	1	10	1	10	0	10	0	0	6	6	0	10	6	16	0	7	3	10	0
		1/250	10	0	10	0	10	0	10	0	0	8	8	0	10	8	18	1	2	4	6	0
50,000	6	1/500	10	0	10	0	6	4	10	1	4	4	8	1	9	8	17	0	3	5	8	1
		1/1000	10	0	10	0	10	0	10	0	0	8	8	0	10	8	18	5	10	8	18	4
		1/250	8	7	15	0	7	3	10	1	3	7	10	1	7	7	14	0	4	3	7	0
		1/500	0	10	10	0	10	0	10	0	0	10	10	0	7	10	17	0	10	10	20	1
	8	1/1000	8	5	13	1	10	4	14	1	0	10	10	0	10	10	20	9	5	6	11	0
		1/250	10	3	13	1	10	0	10	0	0	6	6	0	10	6	16	1	4	2	6	0
		1/500	10	0	10	0	10	0	10	0	0	6	6	0	8	5	13	0	10	6	16	1
		1/1000	9	1	10	0	3	6	9	0	8	1	9	1	10	6	16	5	10	6	16	1
	10	1/250	10	5	15	1	10	0	10	0	0	8	8	0	4	5	9	0	3	5	8	0
		1/500	4	6	10	0	5	5	10	1	5	3	8	1	8	6	14	0	5	3	8	0
		1/1000	10	3	13	1	2	8	10	1	8	0	8	0	10	8	18	3	2	2	4	0
		1/250	10	0	10	0	0	10	10	0	10	0	10	0	8	9	17	0	6	1	7	0
100,000	6	1/500	10	0	10	0	9	2	11	1	0	8	8	0	6	10	16	0	6	1	7	0
		1/1000	1	2	3	0	0	10	10	0	10	0	10	0	10	10	20	4	6	6	12	2

Supplemental Table 18: Number of repeat copies resolved and number of haplotype switching in the assemblies produced by RAmbler, hifiasm, LJA, and HiCANU; assemblers were given in input synthetic HiFi reads with coverage 30× generated from synthetic diploid genomes; in each synthetic genome the primary haplotype (hap1) had ten copies of the repeat unit, while the secondary haplotype (hap2) had 6,8 or 10 copies; red numbers indicate incorrect repeat copy counts (neither matching either haplotype copy counts nor their sum); blue numbers indicate copy counts that match the sum of copies for both haplotypes (hap1+hap2); grey cells indicate assemblies with haplotype switches; Verkko was omitted because it failed to produce any output in most cases.

References

Chan S, Wang W, Hallers B ten, Peters S, Gaiero P, Jong H de, Perez G, Hastie A, and Cao H. 2015. Detection, Characterization, and Biological Analysis of Long Tandem Repeats Using Nanochannel Technology. In *poster at Plant and Animal Genome conference*.