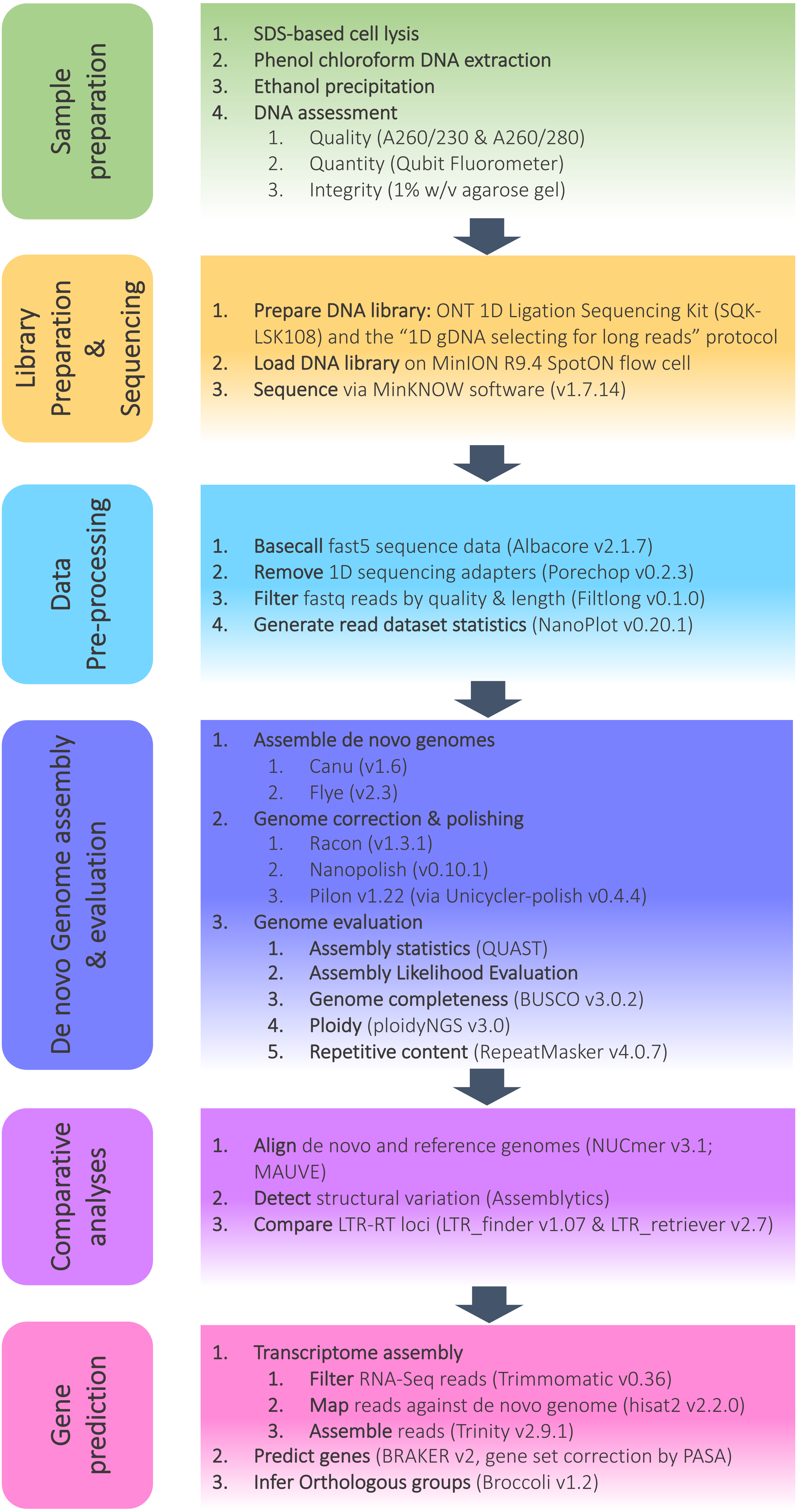


Sample preparation

1. SDS-based cell lysis
2. Phenol chloroform DNA extraction
3. Ethanol precipitation
4. DNA assessment
    1. Quality (A260/230 & A260/280)
    2. Quantity (Qubit Fluorometer)
    3. Integrity (1% w/v agarose gel)

Library Preparation & Sequencing

1. **Prepare DNA library:** ONT 1D Ligation Sequencing Kit (SQK-LSK108) and the "1D gDNA selecting for long reads" protocol
2. **Load DNA library** on MinION R9.4 SpotON flow cell
3. **Sequence** via MinKNOW software (v1.7.14)

Data Pre-processing

1. **Basecall** fast5 sequence data (Albacore v2.1.7)
2. **Remove** 1D sequencing adapters (Porechop v0.2.3)
3. **Filter** fastq reads by quality & length (Filtlong v0.1.0)
4. **Generate read dataset statistics** (NanoPlot v0.20.1)

De novo Genome assembly & evaluation

1. Assemble de novo genomes
    1. Canu (v1.6)
    2. Flye (v2.3)
2. Genome correction & polishing
    1. Racon (v1.3.1)
    2. Nanopolish (v0.10.1)
    3. Pilon v1.22 (via Unicycler-polish v0.4.4)
3. Genome evaluation
    1. **Assembly statistics** (QUAST)
    2. **Assembly Likelihood Evaluation**
    3. **Genome completeness** (BUSCO v3.0.2)
    4. **Ploidy** (ploidyNGS v3.0)
    5. **Repetitive content** (RepeatMasker v4.0.7)

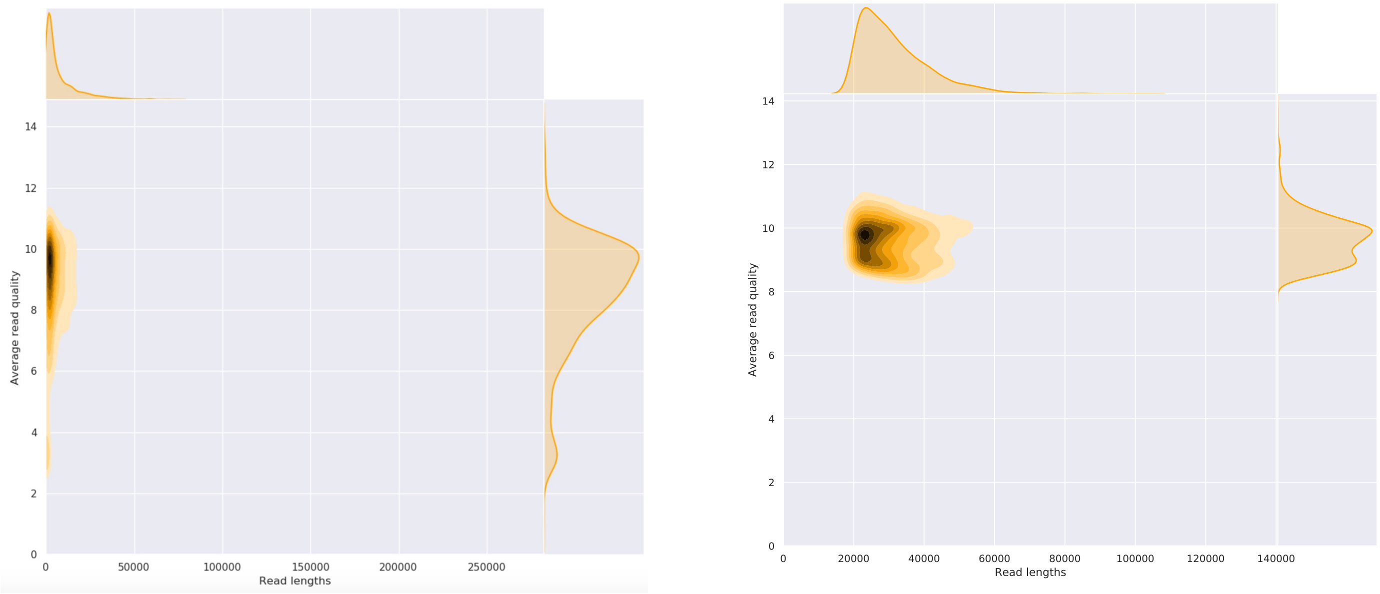
Comparative analyses

1. **Align** de novo and reference genomes (NUCmer v3.1; MAUVE)
2. **Detect** structural variation (Assemblytics)
3. **Compare** LTR-RT loci (LTR_finder v1.07 & LTR_retriever v2.7)

Gene prediction

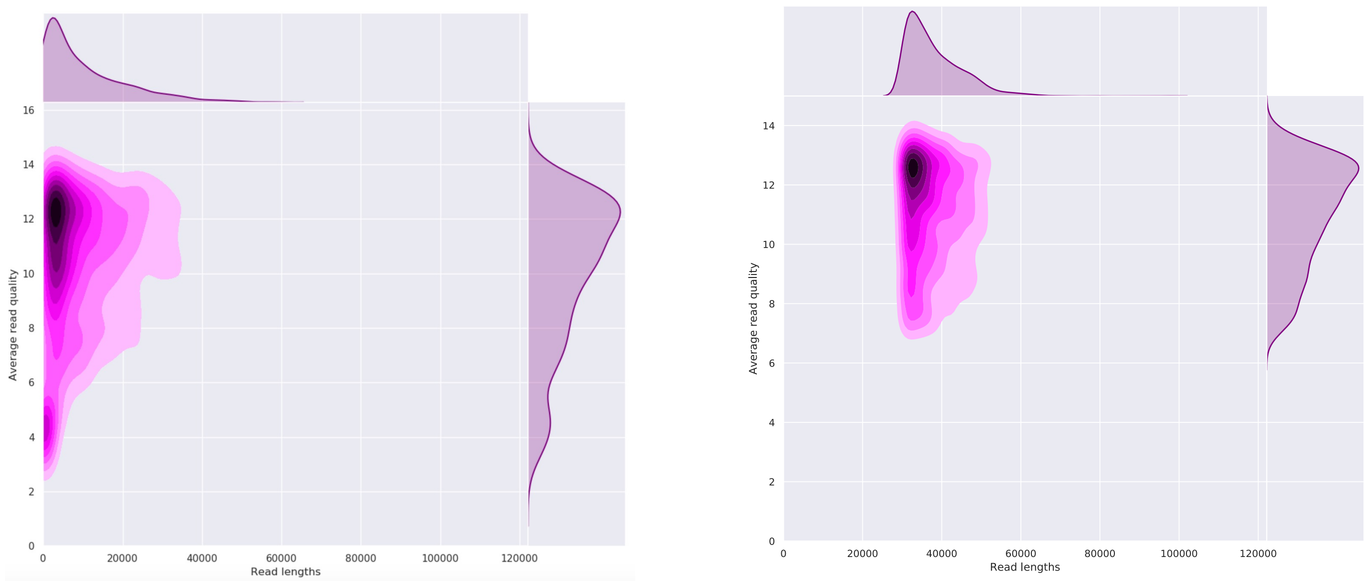
1. **Transcriptome assembly**
    1. **Filter** RNA-Seq reads (Trimmomatic v0.36)
    2. **Map** reads against de novo genome (hisat2 v2.2.0)
    3. **Assemble** reads (Trinity v2.9.1)
2. **Predict genes** (BRAKER v2, gene set correction by PASA)
3. **Infer Orthologous groups** (Broccoli v1.2)

Supplementary Figure 1



A Unfiltered reads

B Filtered reads > 20kb, Q7

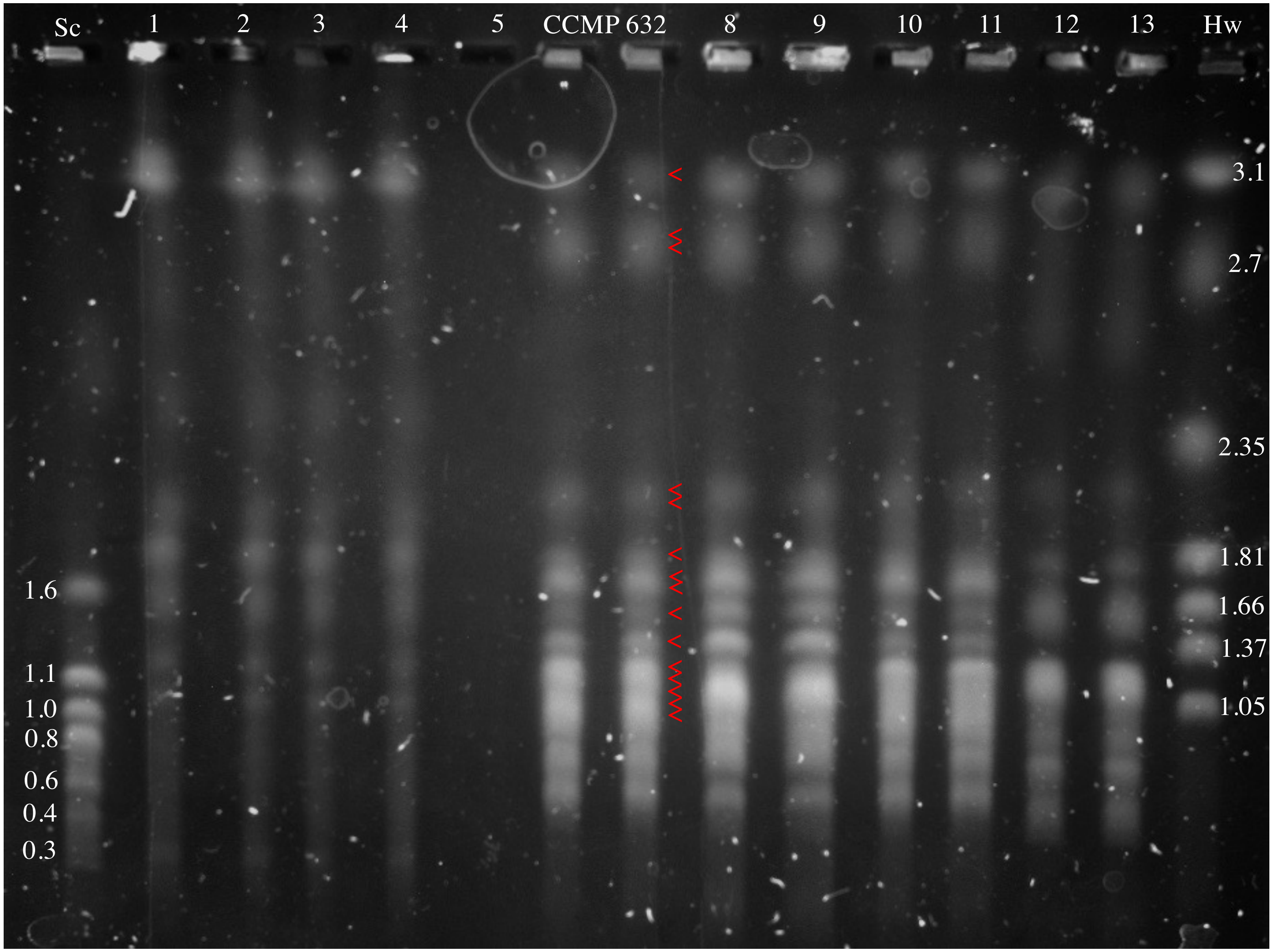


C Unfiltered reads

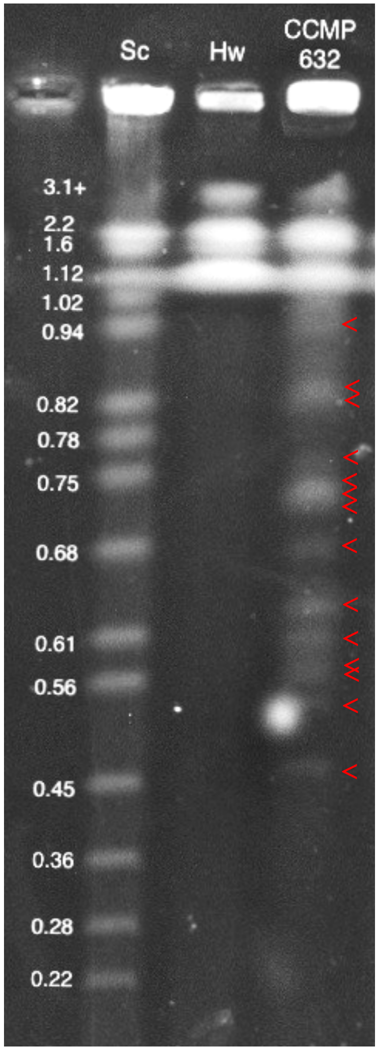
D Filtered reads > 30 kb, Q7

Supplementary Figure 2

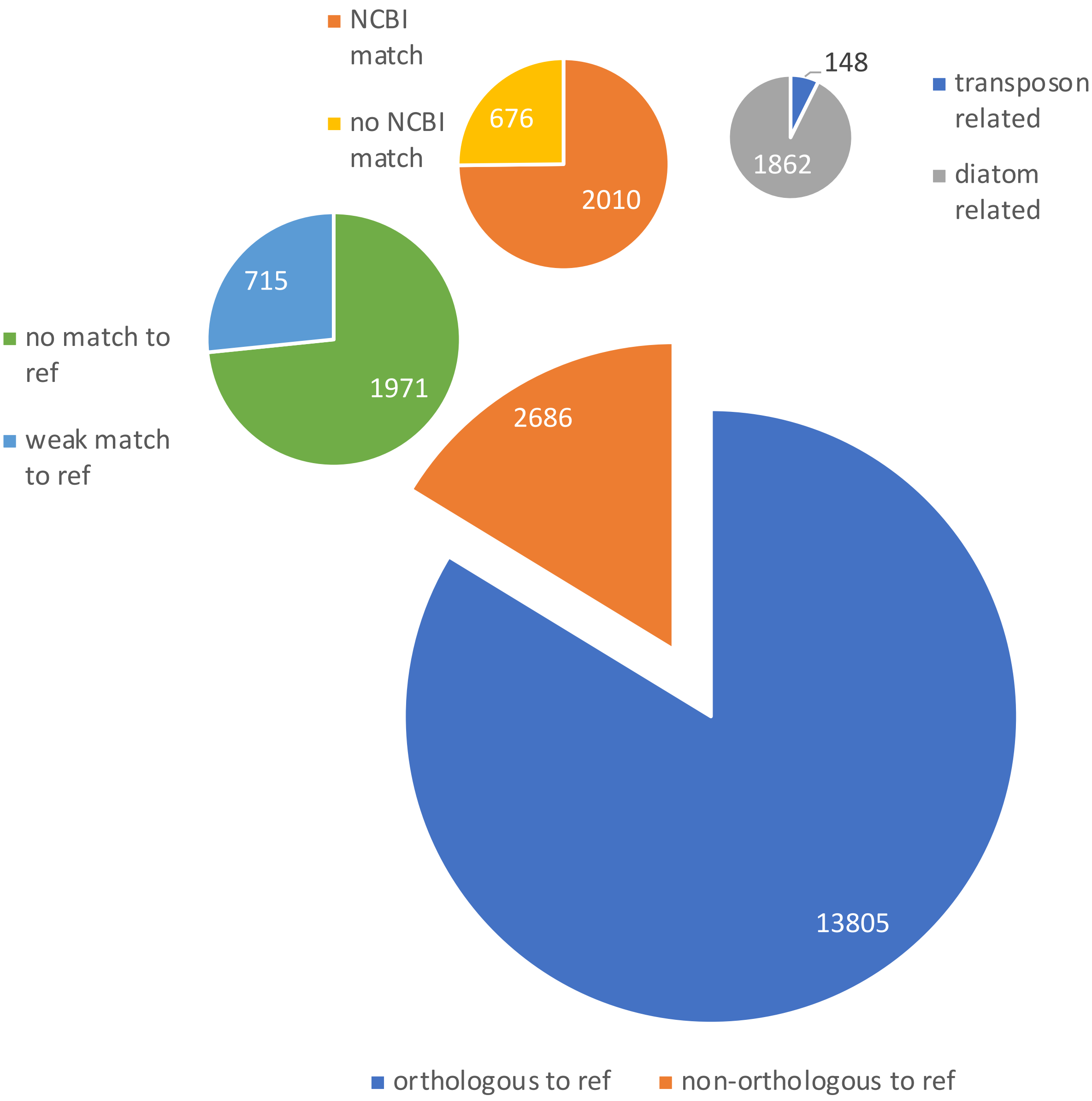
A



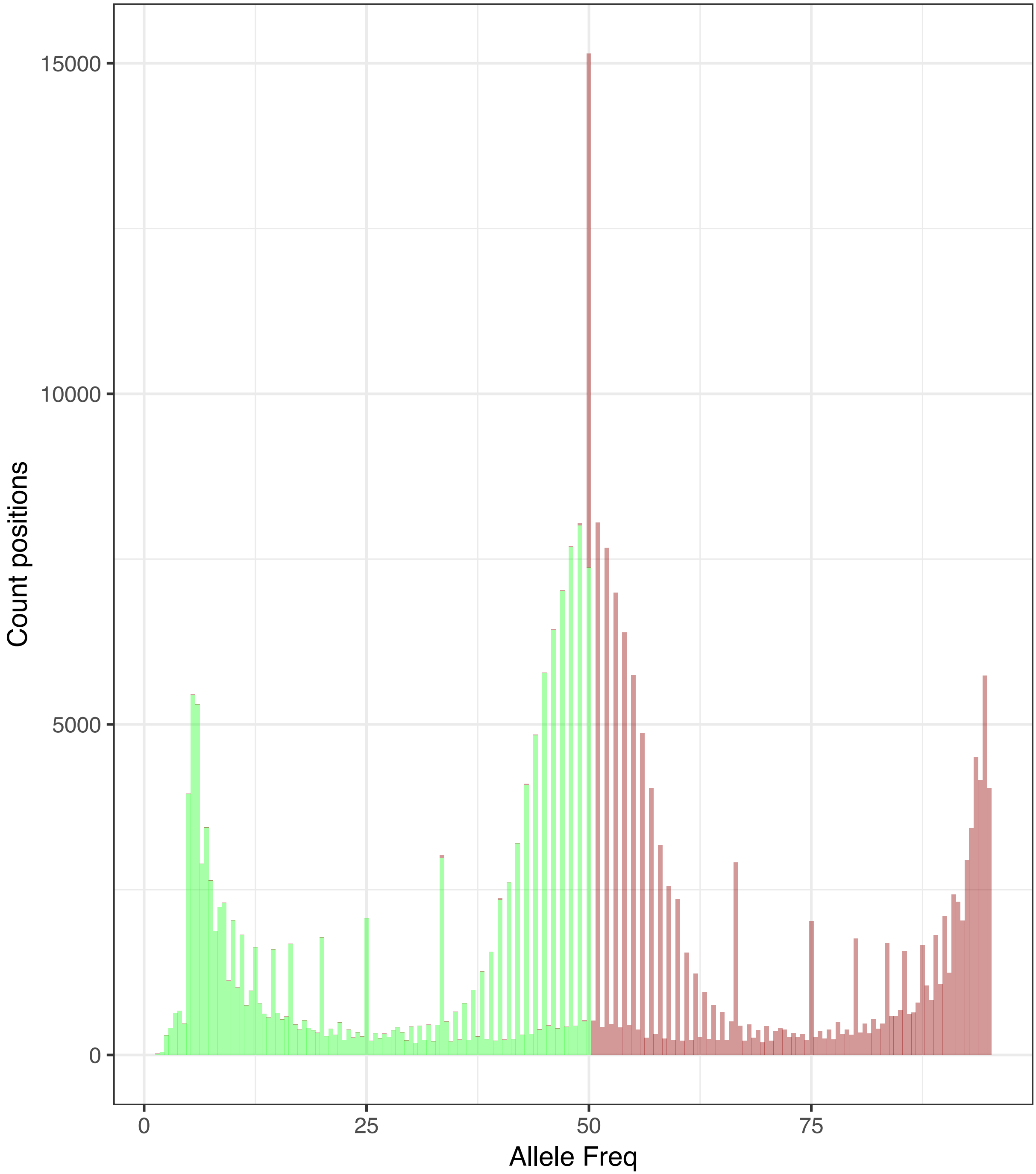
B



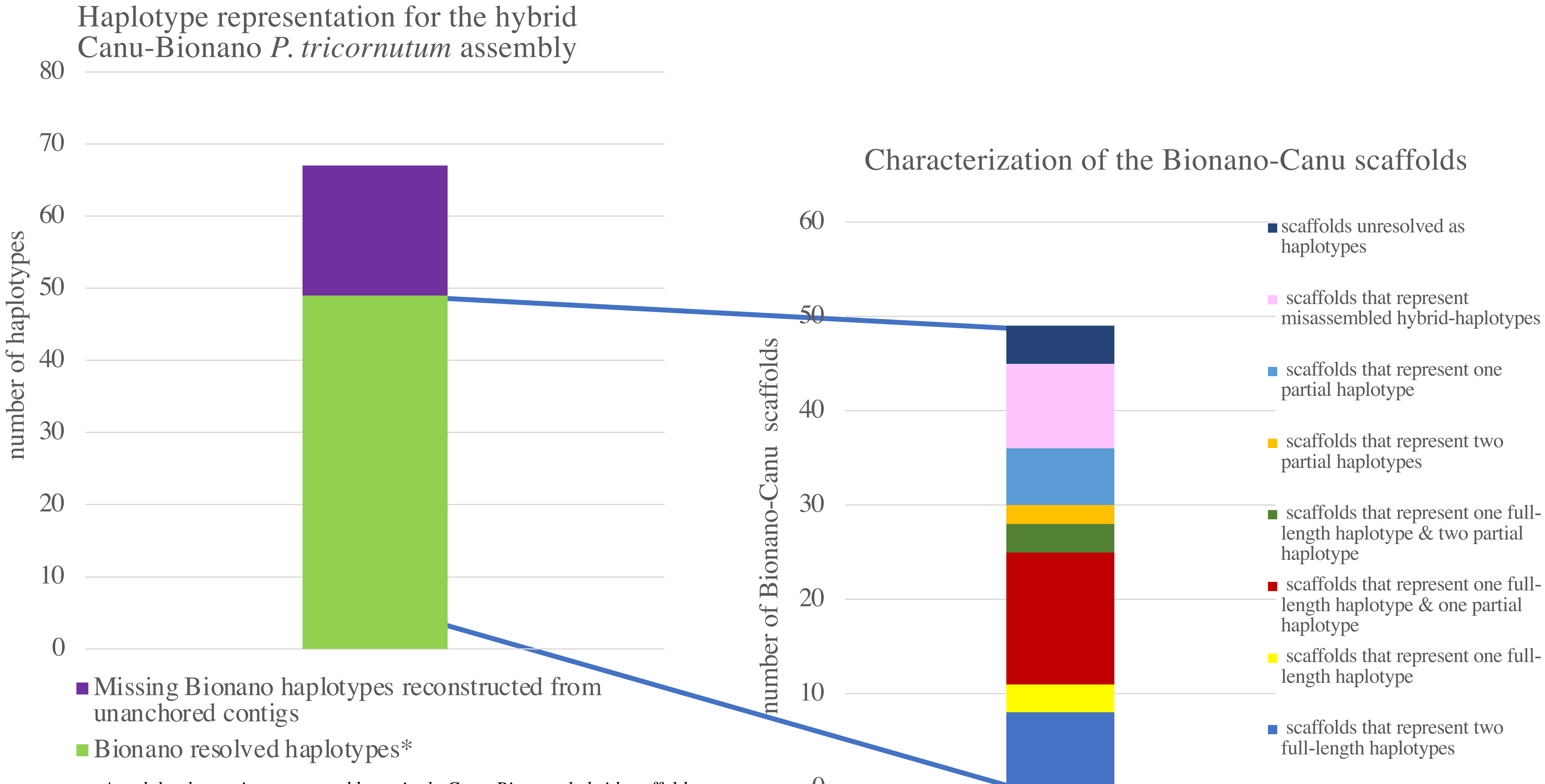
Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5



**each haplotype is represented by a single Canu-Bionano hybrid scaffold*

Supplementary Figure 6

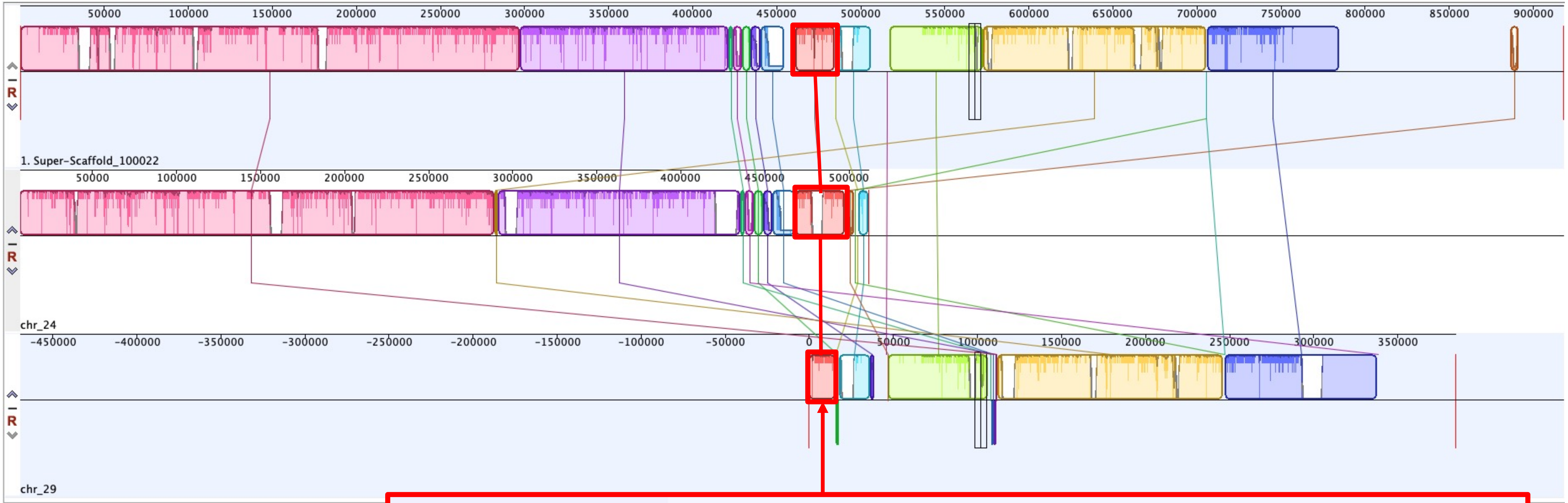
Super-Scaffold_100002

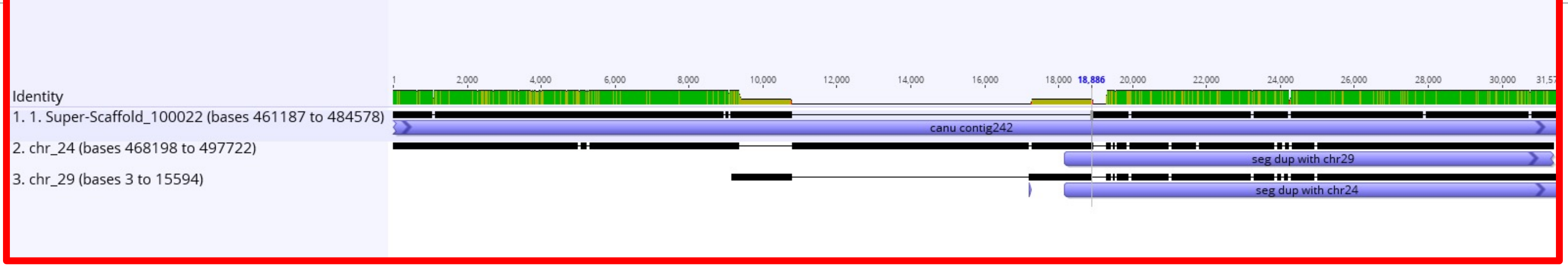
chr28-chr7-chr13 hybrid
contigs are erroneously linked
owing to segmental duplications
between those chromosome pairs

Supplementary Figure 7A

Super-Scaffold_100022

chr24-chr29 hybrid
contigs were erroneously linked owing to segmental duplications
between those chromosome pairs



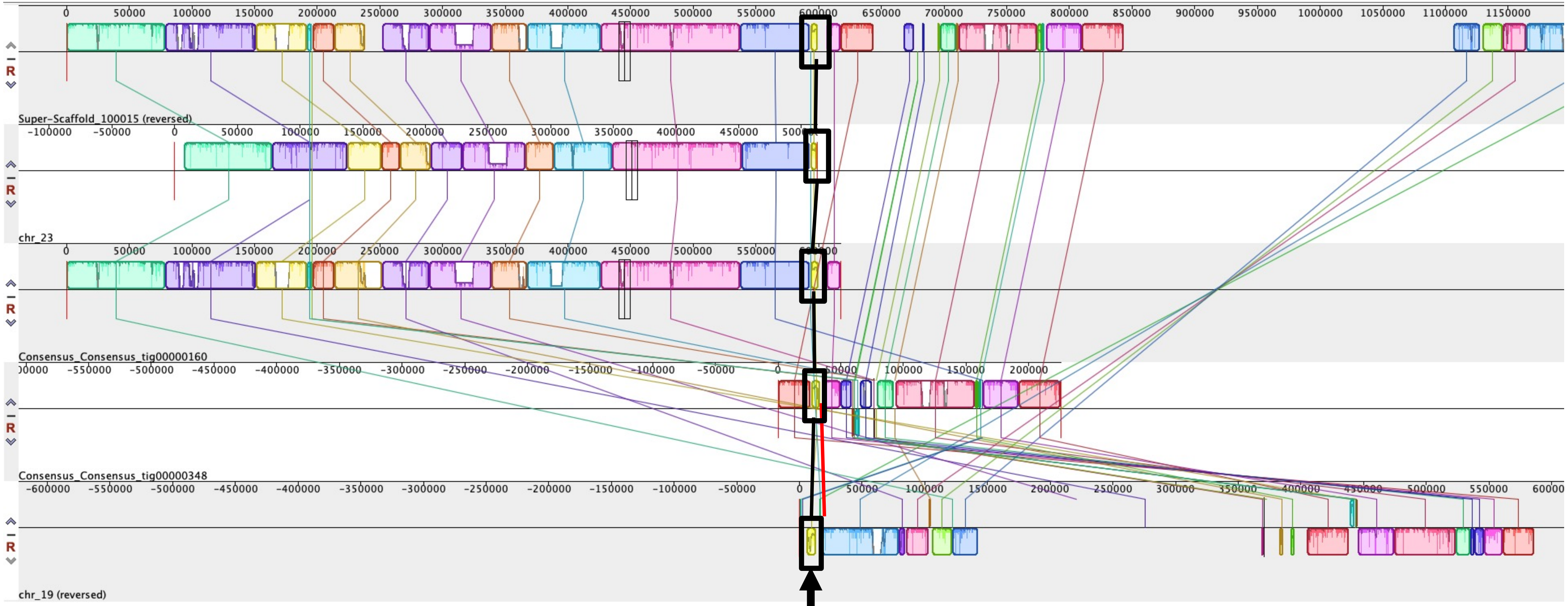


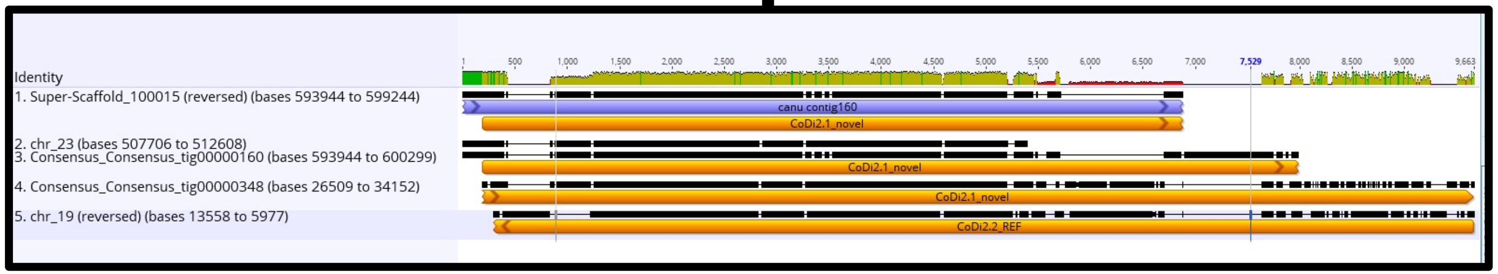
1. 1. Super-Scaffold_100022 (bases 461187 to 484578)
2. chr_24 (bases 468198 to 497722)
3. chr_29 (bases 3 to 15594)

Supplementary Figure 7B

Super-Scaffold_100015

chr23-chr19 hybrid
contigs were erroneously linked owing to CoDi2.1 repetitive elements





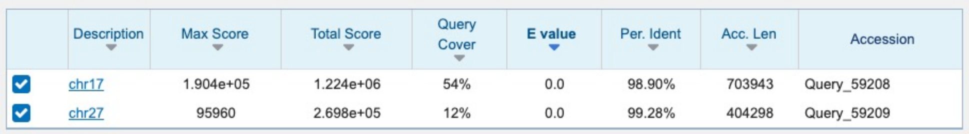
1. Super-Scaffold_100015 (reversed) (bases 593944 to 599244)
2. chr_23 (bases 507706 to 512608)
3. Consensus_Consensus_tig00000160 (bases 593944 to 600299)
4. Consensus_Consensus_tig00000348 (bases 26509 to 34152)
5. chr_19 (reversed) (bases 13558 to 5977)

Supplementary Figure 7C

Super-Scaffold_100016

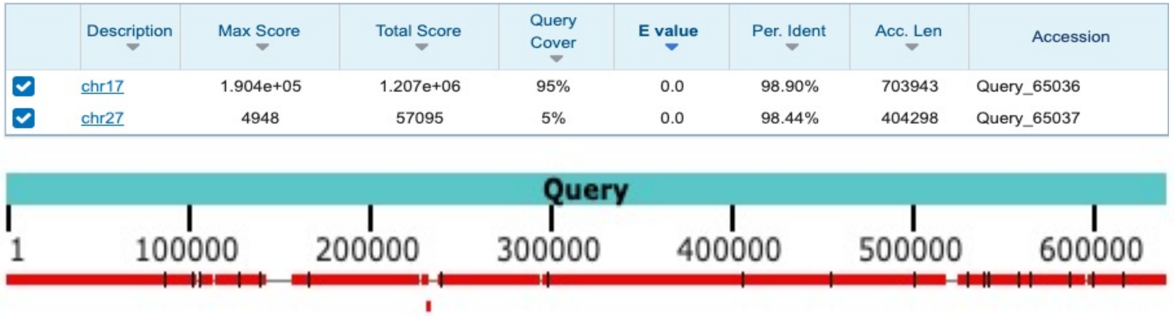
chr17—chr27 hybrid

Reference chr17 and chr27 blasted against Super-Scaffold_100016



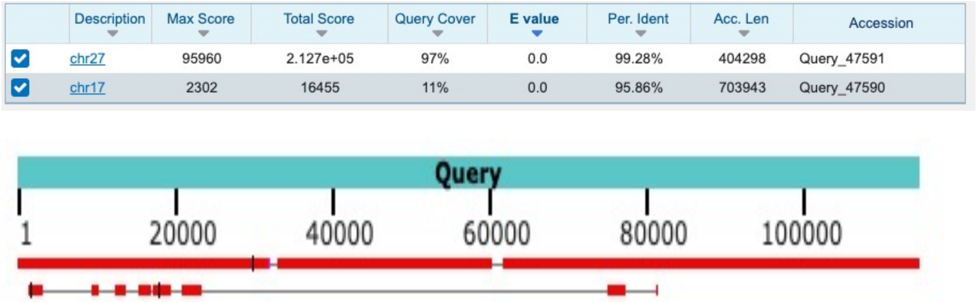
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	chr17	1.904e+05	1.224e+06	54%	0.0	98.90%	703943	Query_59208
☑	chr27	95960	2.698e+05	12%	0.0	99.28%	404298	Query_59209

Portions of Canu contig2816 blasts with high query coverage to chr17. The small sequence that aligns to chr27 is part of the CoDi3.2 locus.



	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	chr17	1.904e+05	1.207e+06	95%	0.0	98.90%	703943	Query_65036
☑	chr27	4948	57095	5%	0.0	98.44%	404298	Query_65037

Almost the entirety of Canu contig2803 blasts strongly to chr27



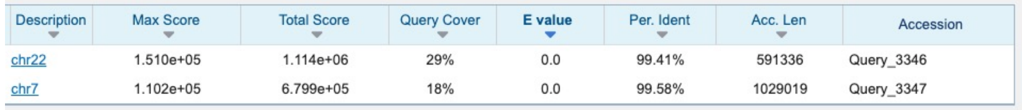
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	chr27	95960	2.127e+05	97%	0.0	99.28%	404298	Query_47591
☑	chr17	2302	16455	11%	0.0	95.86%	703943	Query_47590

Supplementary Figure 8A

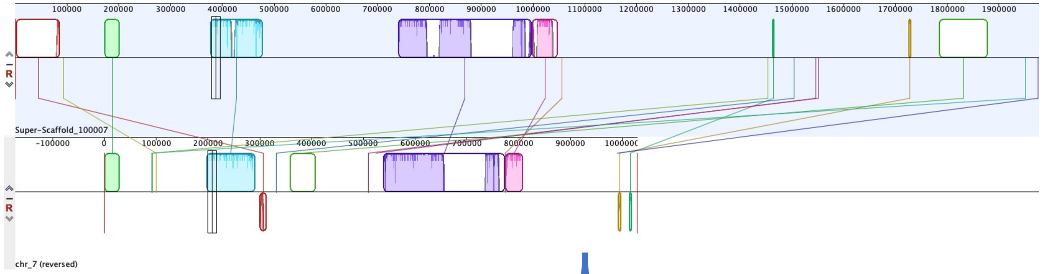
Super-Scaffold_100007

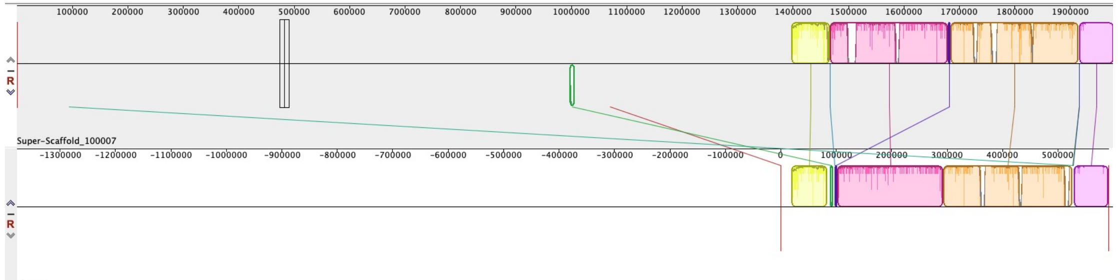
chr7—chr22 hybrid

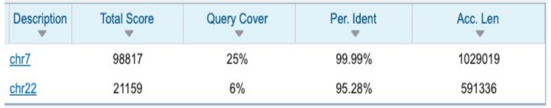
Reference chr22 and chr7 blasted against Super-Scaffold_100007



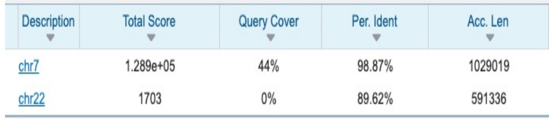
Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
chr22	1.510e+05	1.114e+06	29%	0.0	99.41%	591336	Query_3346
chr7	1.102e+05	6.799e+05	18%	0.0	99.58%	1029019	Query_3347



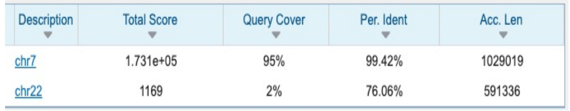




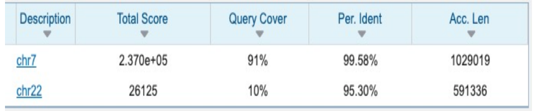
Description	Total Score	Query Cover	Per. Ident	Acc. Len
chr7	98817	25%	99.99%	1029019
chr22	21159	6%	95.28%	591336



Description	Total Score	Query Cover	Per. Ident	Acc. Len
chr7	1.289e+05	44%	98.87%	1029019
chr22	1703	0%	89.62%	591336

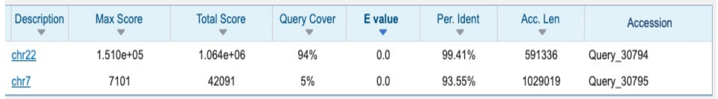


Description	Total Score	Query Cover	Per. Ident	Acc. Len
chr7	1.731e+05	95%	99.42%	1029019
chr22	1169	2%	76.06%	591336

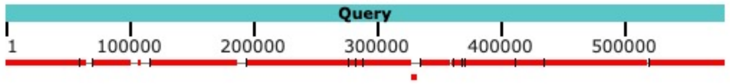


Description	Total Score	Query Cover	Per. Ident	Acc. Len
chr7	2.370e+05	91%	99.58%	1029019
chr22	26125	10%	95.30%	591336

Almost the entirety of Canu contig336 is homologous to chr22.



Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
chr22	1.510e+05	1.064e+06	94%	0.0	99.41%	591336	Query_30794
chr7	7101	42091	5%	0.0	93.55%	1029019	Query_30795



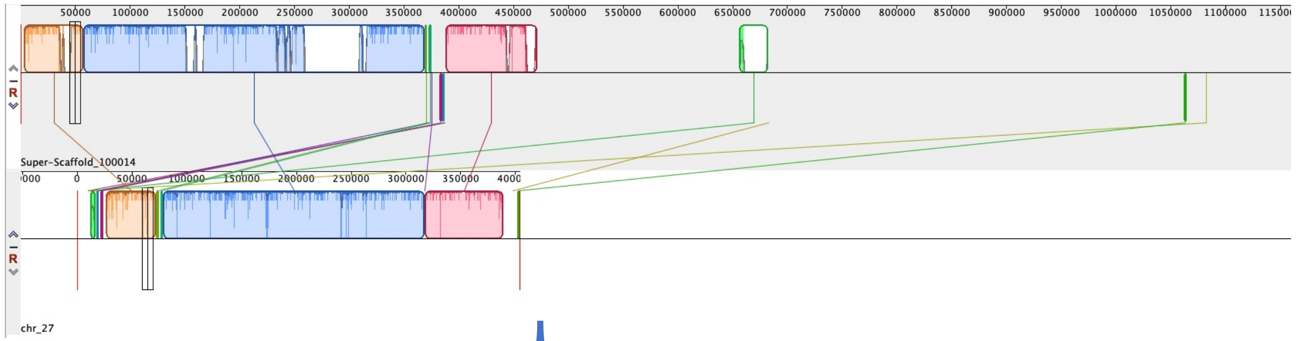
Supplementary Figure 8B

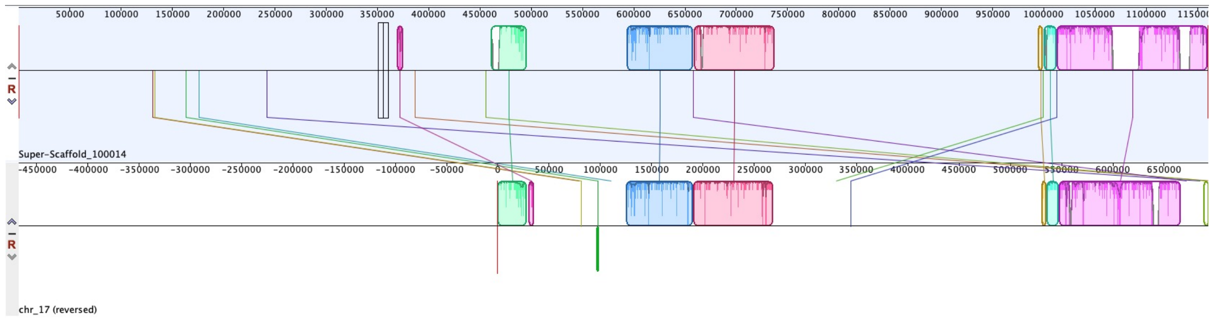
Super-Scaffold_100014

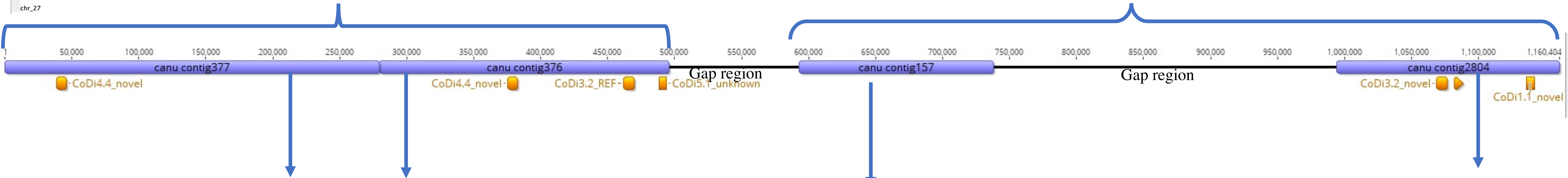
chr27—chr17 hybrid

Reference chr27 and chr17 blasted against Super-Scaffold_100014

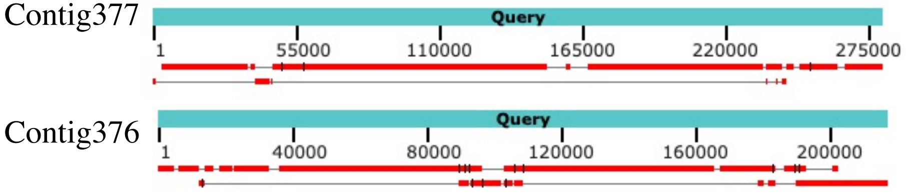
	Description	Total Score	Query Cover	E value	Per. Ident	Acc. Len
☑	chr27	7.802e+05	37%	0.0	99.03%	404298
☑	chr17	6.373e+05	28%	0.0	99.42%	703943



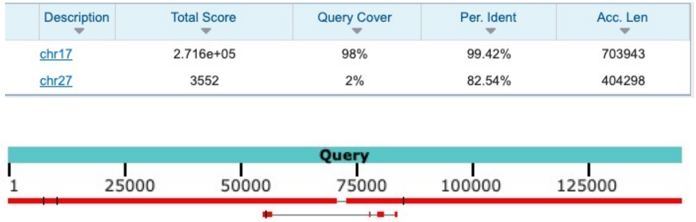




Canu contig377 and contig 376 have 99% identity to chr27 with a qcov against chr2 of 88% and 80%, respectively. Contig377 has a qcov to chr17 of 3% while the qcov to chr17 for contig 376 is 23%.

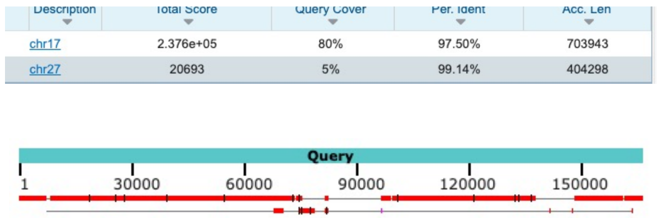


Canu contig157 blasts to chr17 along its entire length.



Description	Total Score	Query Cover	Per. Ident	Acc. Len
chr17	2.716e+05	98%	99.42%	703943
chr27	3552	2%	82.54%	404298

Canu contig2804 blasts to chr17 along the majority of its length.



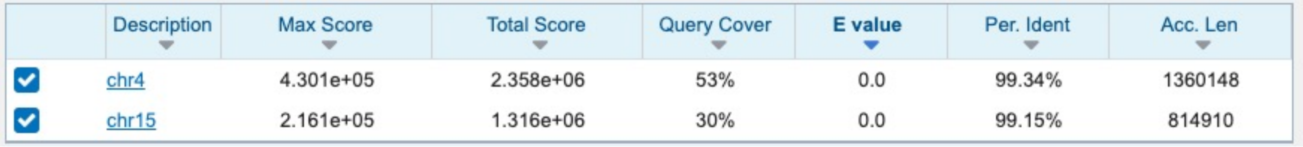
Description	Total Score	Query Cover	Per. Ident	Acc. Len
chr17	2.376e+05	80%	97.50%	703943
chr27	20693	5%	99.14%	404298

Supplementary Figure 8C

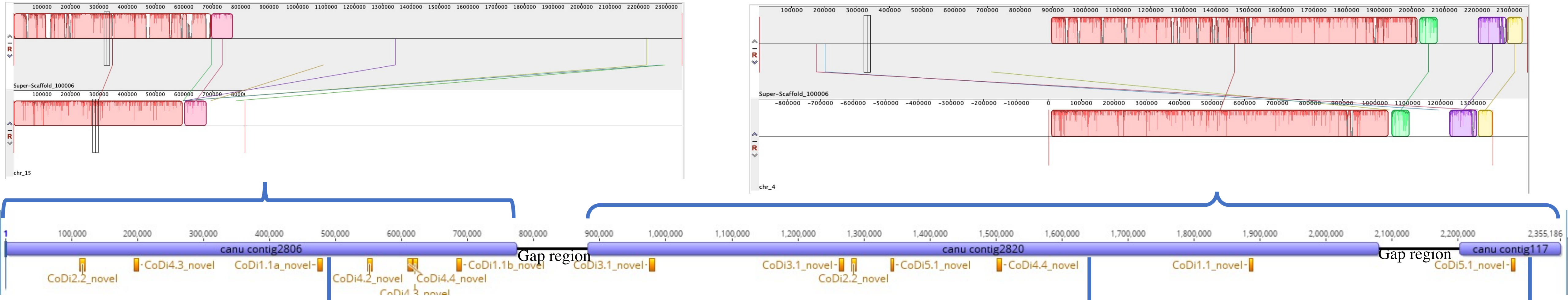
Super-Scaffold_100006

chr15—chr4 hybrid

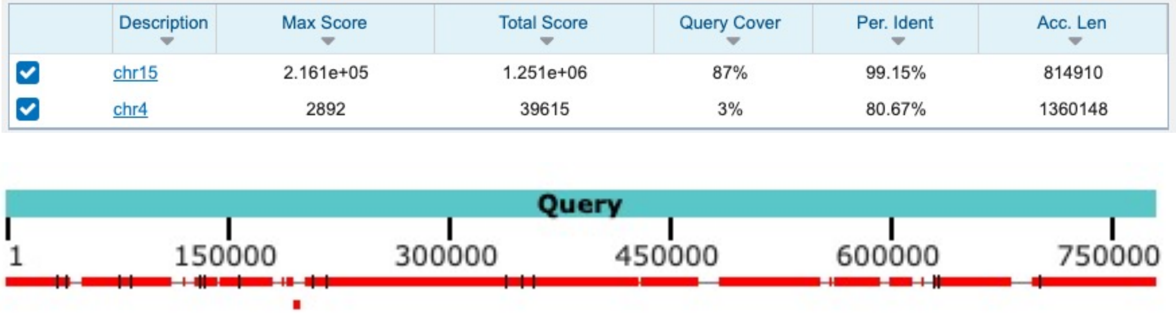
Reference chr4 and chr15 blasted against Super-Scaffold_100006



☐	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len
☑	chr4	4.301e+05	2.358e+06	53%	0.0	99.34%	1360148
☑	chr15	2.161e+05	1.316e+06	30%	0.0	99.15%	814910

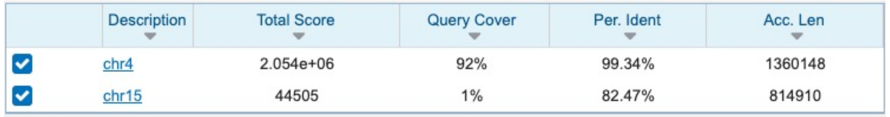


Almost the entirety of Canu contig2806 blasts to chr15 with low query coverage to chr4.



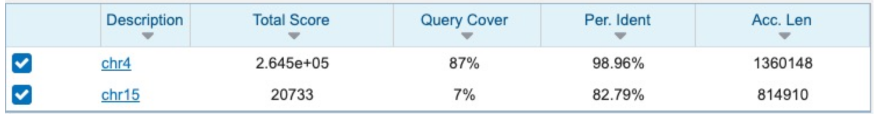
☐	Description	Max Score	Total Score	Query Cover	Per. Ident	Acc. Len
☑	chr15	2.161e+05	1.251e+06	87%	99.15%	814910
☑	chr4	2892	39615	3%	80.67%	1360148

Almost the entirety of Canu contig2820 blasts to chr4 with low query coverage to chr15.



☐	Description	Total Score	Query Cover	Per. Ident	Acc. Len
☑	chr4	2.054e+06	92%	99.34%	1360148
☑	chr15	44505	1%	82.47%	814910

Almost the entirety of Canu contig117 blasts to chr4 with low query coverage to chr15.



☐	Description	Total Score	Query Cover	Per. Ident	Acc. Len
☑	chr4	2.645e+05	87%	98.96%	1360148
☑	chr15	20733	7%	82.79%	814910

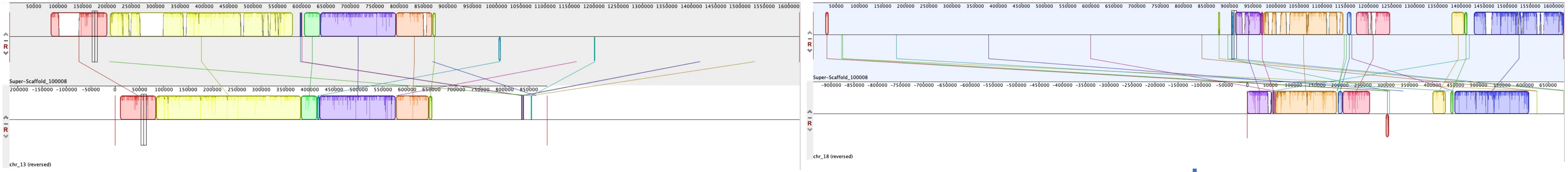
Supplementary Figure 8D

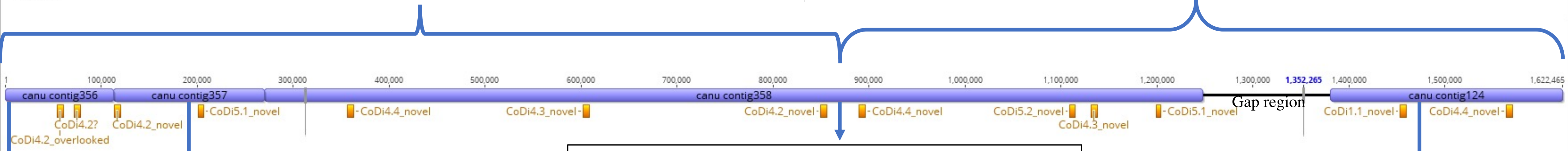
Super-Scaffold_100008

chr13—chr18 hybrid

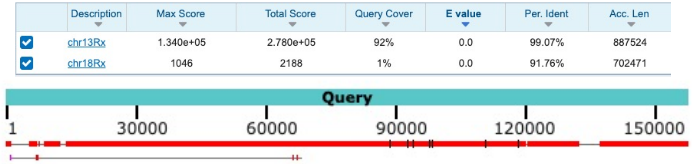
Reference chr13 and chr18 blasted against Super-Scaffold_100008

☐	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	chr13		2.100e+05	1.545e+06	50%	0.0	99.33%	887524	Query_10986
☑	chr18		69653	1.037e+06	33%	0.0	98.98%	702471	Query_10987



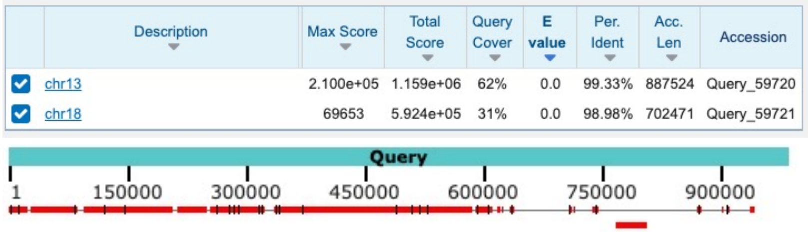


Canu contig357 blasts to chr13 along almost its entire length



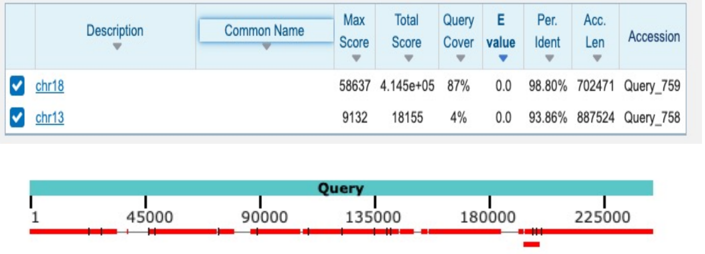
☐	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len
☑	chr13Rx	1.340e+05	2.780e+05	92%	0.0	99.07%	887524
☑	chr18Rx	1046	2188	1%	0.0	91.76%	702471

Portions of Canu contig358 blast to chr13 and chr18. Raw read mapping does not support contig358 as a misassembled contig.



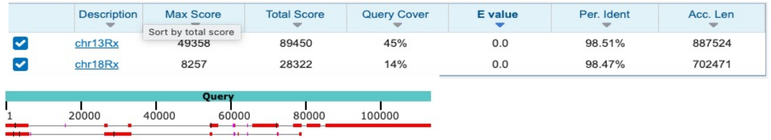
☐	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	chr13	2.100e+05	1.159e+06	62%	0.0	99.33%	887524	Query_59720
☑	chr18	69653	5.924e+05	31%	0.0	98.98%	702471	Query_59721

The majority of Canu contig124 blasts strongly to chr18 and aligns to only to the CoDi4.4 locus of chr13



☐	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	chr18		58637	4.145e+05	87%	0.0	98.80%	702471	Query_759
☑	chr13		9132	18155	4%	0.0	93.86%	887524	Query_758

Canu contig356 strongly identifies a portion of the contig to chr13 with the rest of the contig not well resolved to another reference chromosome



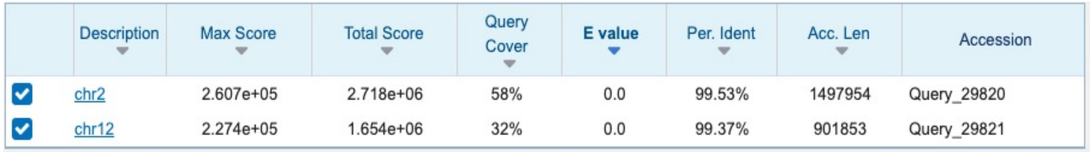
☐	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len
☑	chr13Rx	49358	89450	45%	0.0	98.51%	887524
☑	chr18Rx	8257	28322	14%	0.0	98.47%	702471

Supplementary Figure 8E

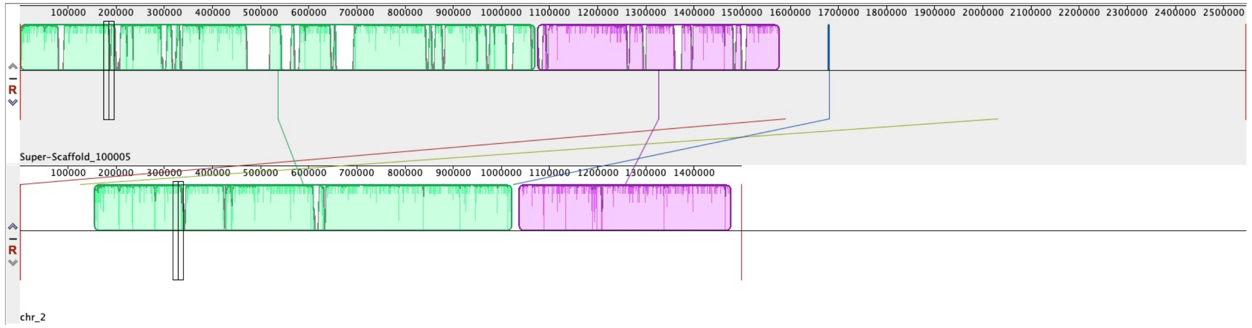
Super-Scaffold_100005

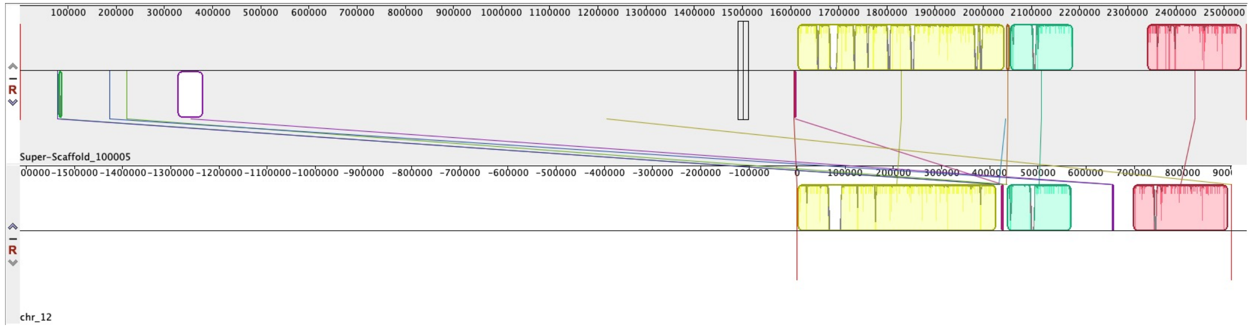
chr2—chr12 hybrid

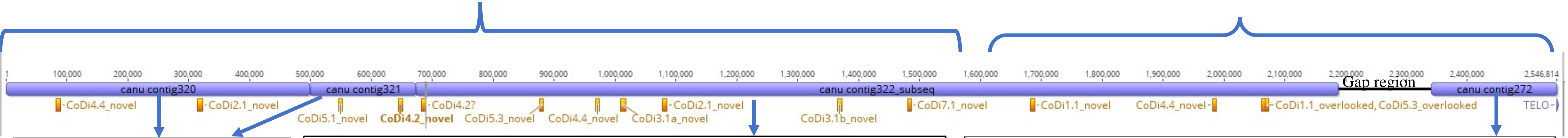
Reference chr2 and chr12 blasted against Super-Scaffold_100005



	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	chr2	2.607e+05	2.718e+06	58%	0.0	99.53%	1497954	Query_29820
☑	chr12	2.274e+05	1.654e+06	32%	0.0	99.37%	901853	Query_29821

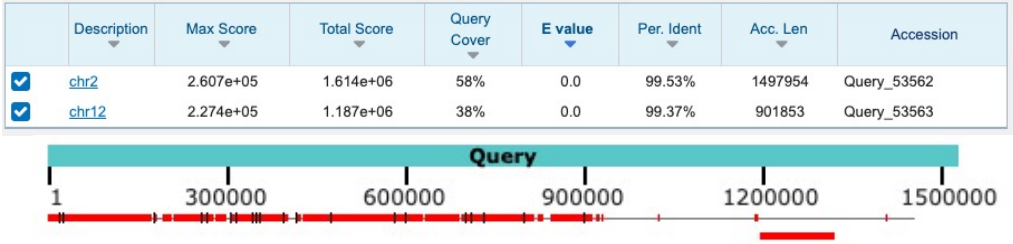






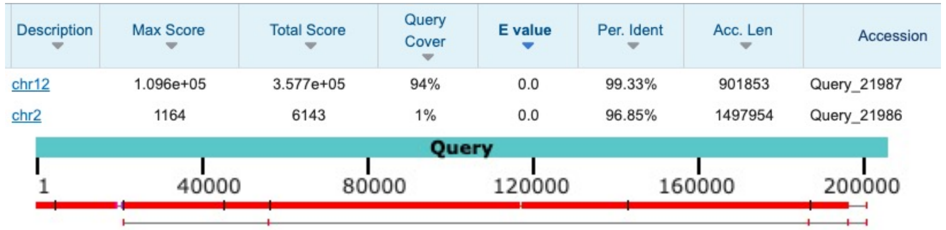
Canu contig320 and contig321 have 99% identity to chr2 with a qcov against chr2 of 91% and 84%, respectively. Both contigs have <10% qcov to chr12

Portions of Canu contig322 blast to chr2 and chr12 but raw read mapping does not indicate contig322 as a misassembled contig and does not support splitting this contig.



	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	chr2	2.607e+05	1.614e+06	58%	0.0	99.53%	1497954	Query_53562
☑	chr12	2.274e+05	1.187e+06	38%	0.0	99.37%	901853	Query_53563

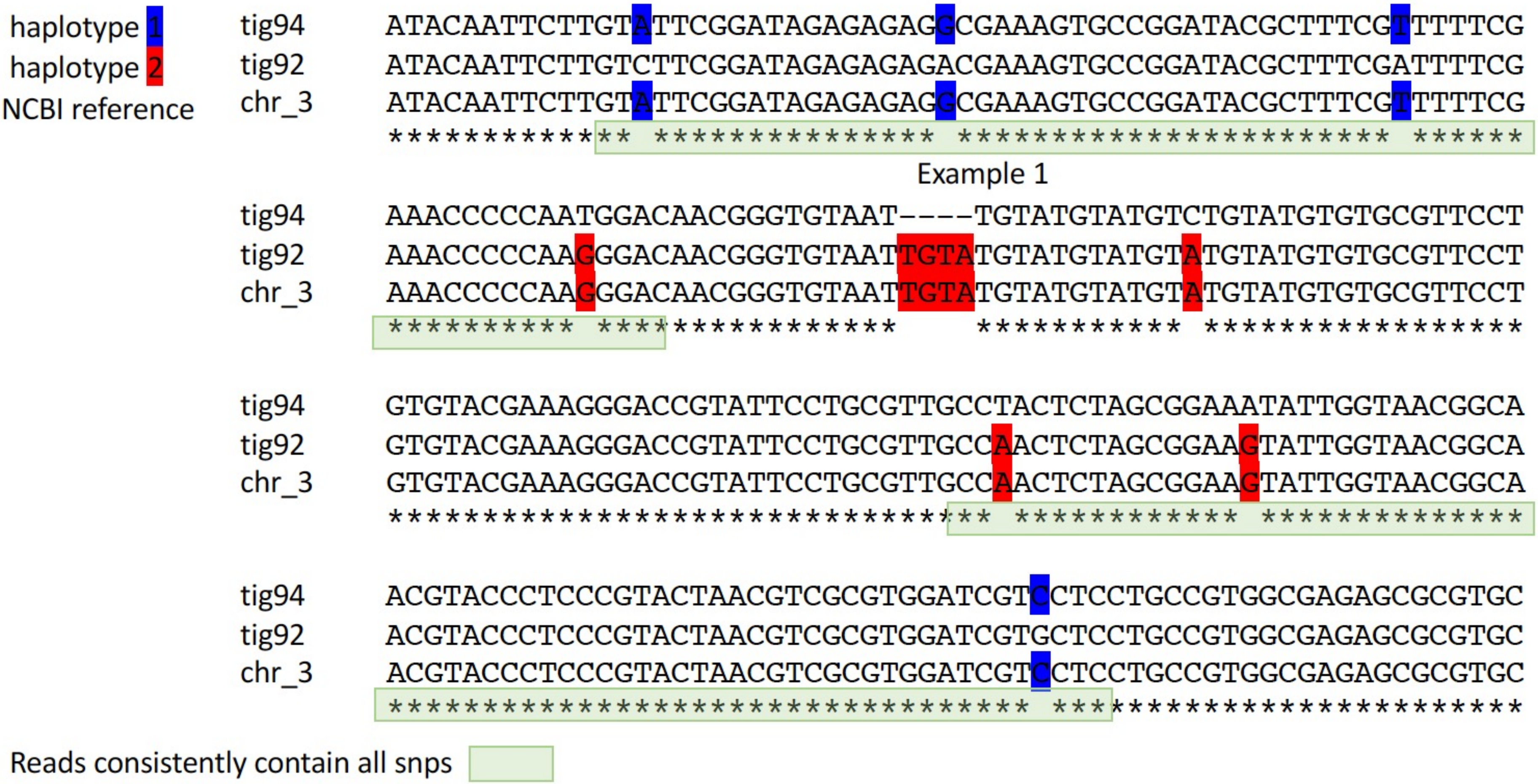
Canu contig272 does not align to chr2 but does show strong identity along its entire length to chr12.



Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
chr12	1.096e+05	3.577e+05	94%	0.0	99.33%	901853	Query_21987
chr2	1164	6143	1%	0.0	96.85%	1497954	Query_21986

Supplementary Figure 8F

Reference consensus is a mixture of haplotypes



haplotype 1
haplotype 2
NCBI reference

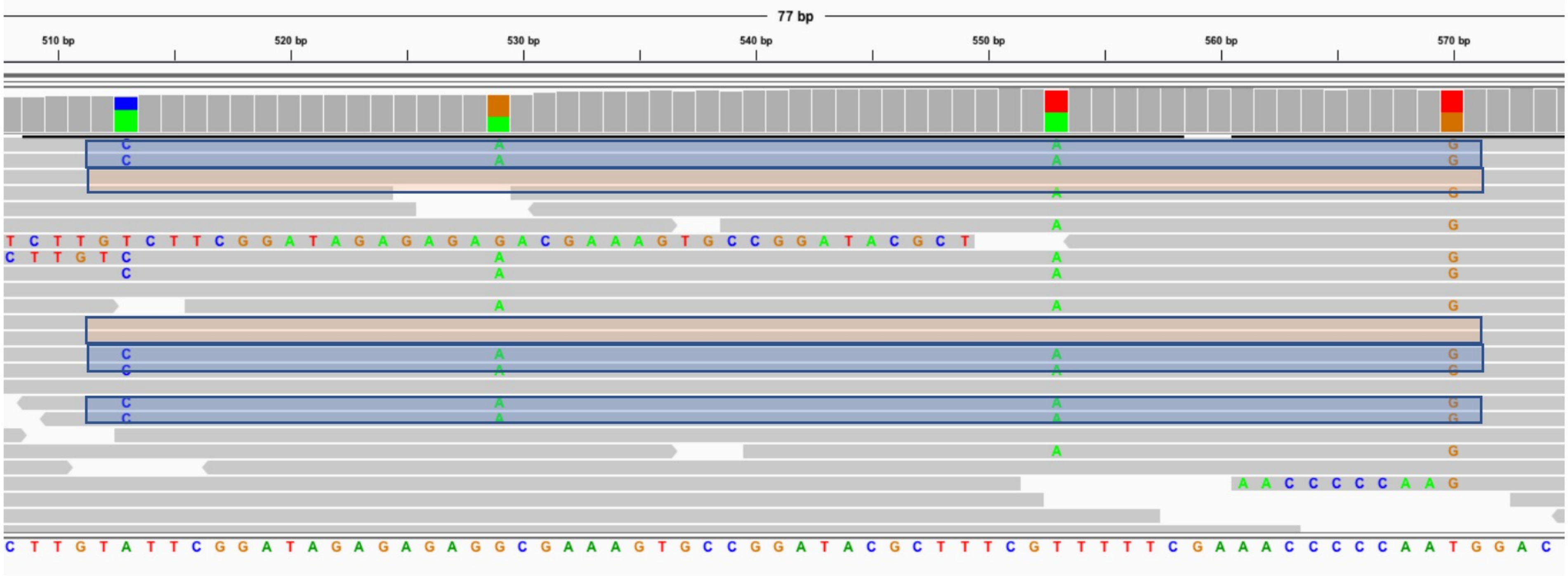
```
tig94    ATACAATTCTTGTATTCGGATAGAGAGAGGCGAAAGTGCCGGATACGCTTTCGTTTTTCG
tig92    ATACAATTCTTGTCTTCGGATAGAGAGAGACGAAAGTGCCGGATACGCTTTCGATTTTCG
chr_3    ATACAATTCTTGTATTCGGATAGAGAGAGGCGAAAGTGCCGGATACGCTTTCGTTTTTCG
         ************* *************** *********************** ******
                                     Example 1

tig94    AAACCCCCAATGGACAACGGGTGTAAT----TGTATGTATGTCTGTATGTGTGCGTTCCT
tig92    AAACCCCCAAGGGACAACGGGTGTAATTGTATGTATGTATGTATGTATGTGTGCGTTCCT
chr_3    AAACCCCCAAGGGACAACGGGTGTAATTGTATGTATGTATGTATGTATGTGTGCGTTCCT
         ********** ****************    *********** *****************

tig94    GTGTACGAAAGGGACCGTATTCCTGCGTTGCCTACTCTAGCGGAAATATTGGTAACGGCA
tig92    GTGTACGAAAGGGACCGTATTCCTGCGTTGCCAACTCTAGCGGAAGTATTGGTAACGGCA
chr_3    GTGTACGAAAGGGACCGTATTCCTGCGTTGCCAACTCTAGCGGAAGTATTGGTAACGGCA
         ******************************** ************ **************

tig94    ACGTACCCTCCCGTACTAACGTCGCGTGGATCGTCCTCCTGCCGTGGCGAGAGCGCGTGC
tig92    ACGTACCCTCCCGTACTAACGTCGCGTGGATCGTGCTCCTGCCGTGGCGAGAGCGCGTGC
chr_3    ACGTACCCTCCCGTACTAACGTCGCGTGGATCGTCCTCCTGCCGTGGCGAGAGCGCGTGC
         ********************************** *************************
```

Reads consistently contain all snps 🟩

Supplementary Figure 9

Reference consensus is a mixture of haplotypes

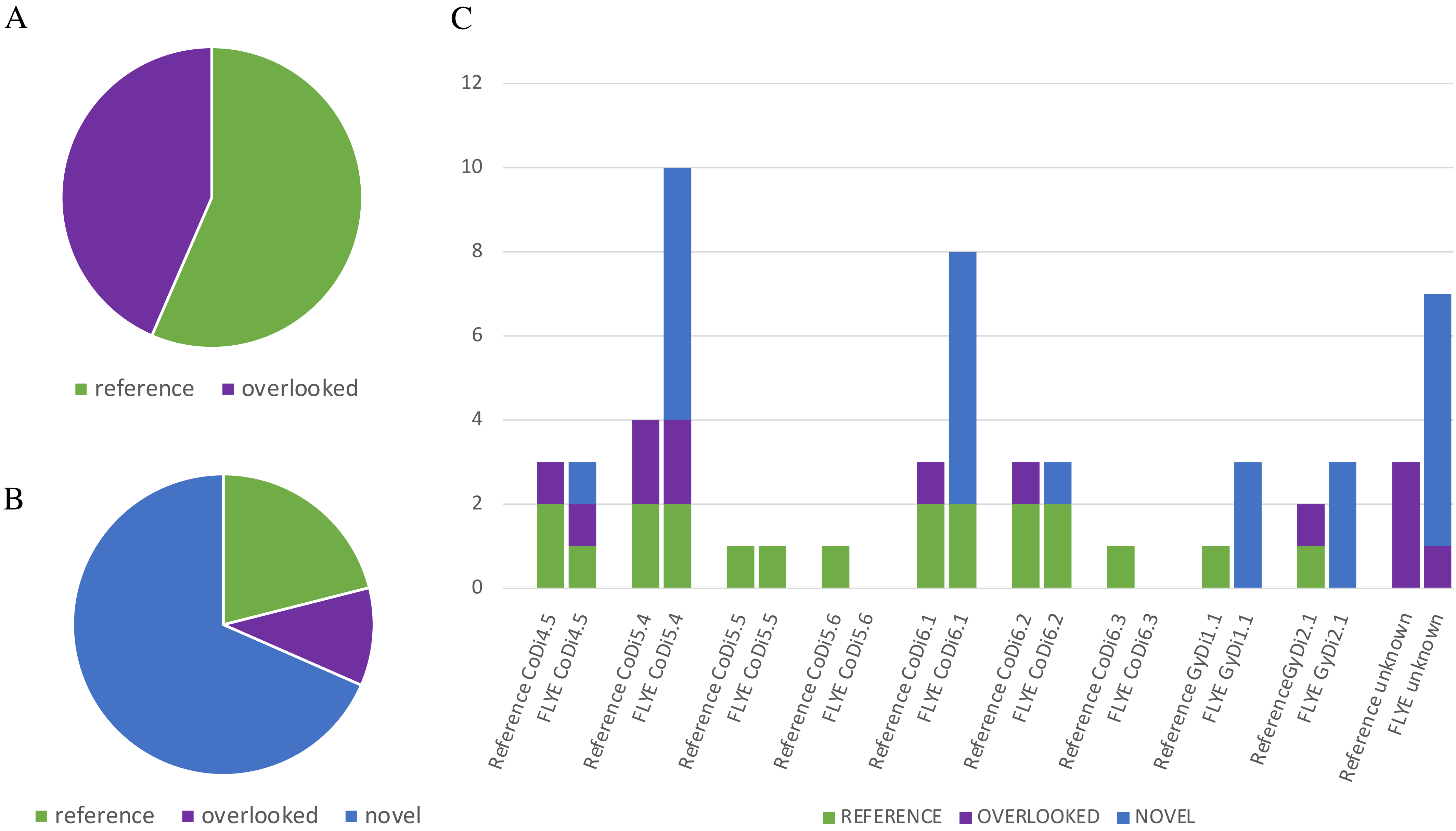
IGV visual multiple alignment of Illumina reads –example 1



32 reads with haplotype 1

30 reads with haplotype 2

Supplementary Figure 10



Supplementary Figure 11