Toward complete, T2T, genome inference with nanopore sequencing

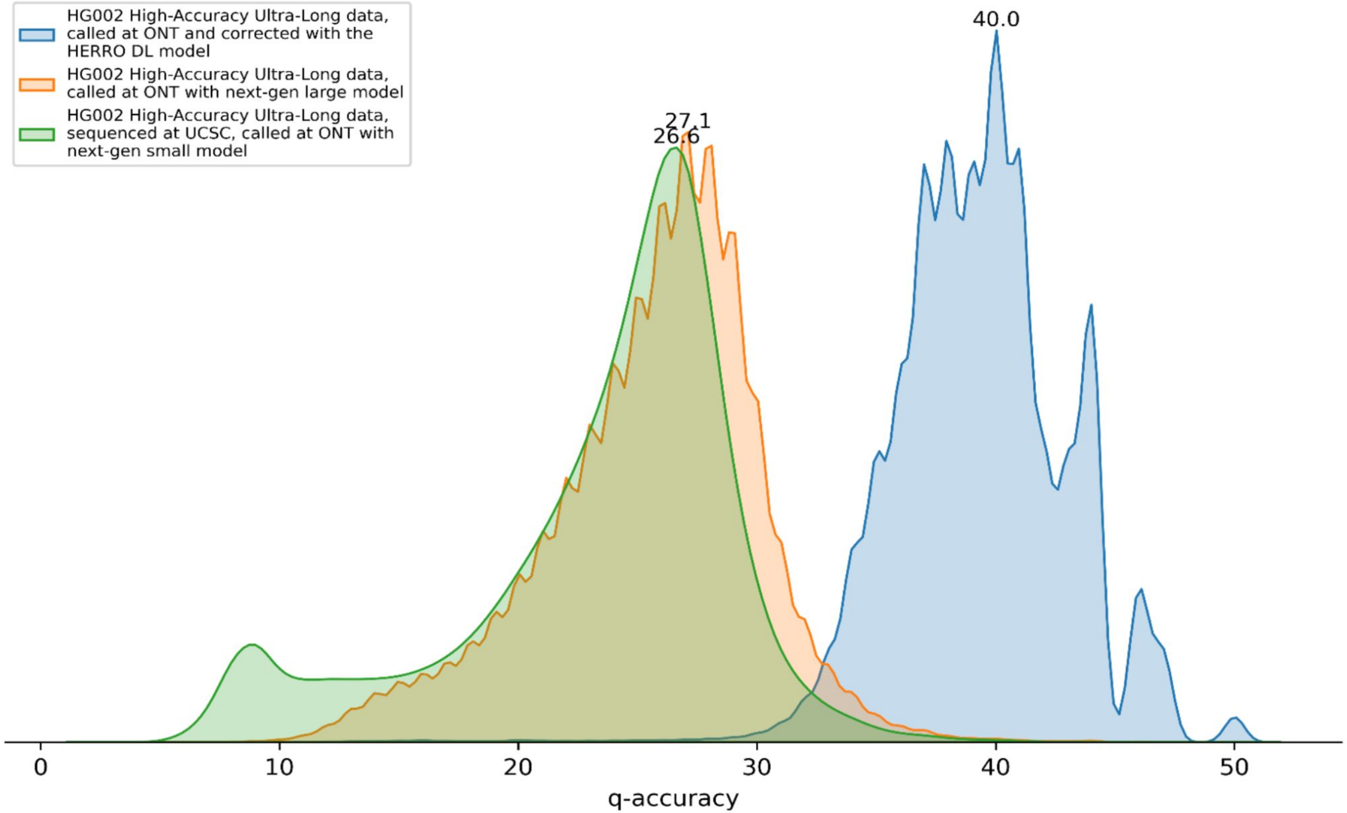
Benedict Paten, Associate Professor, Biomolecular Engineering
Associate Director, UC Santa Cruz Genomics Institute

Santa Cruz Breakwater Lighthouse, photo courtesy Kishwar Shafin

UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics Institute

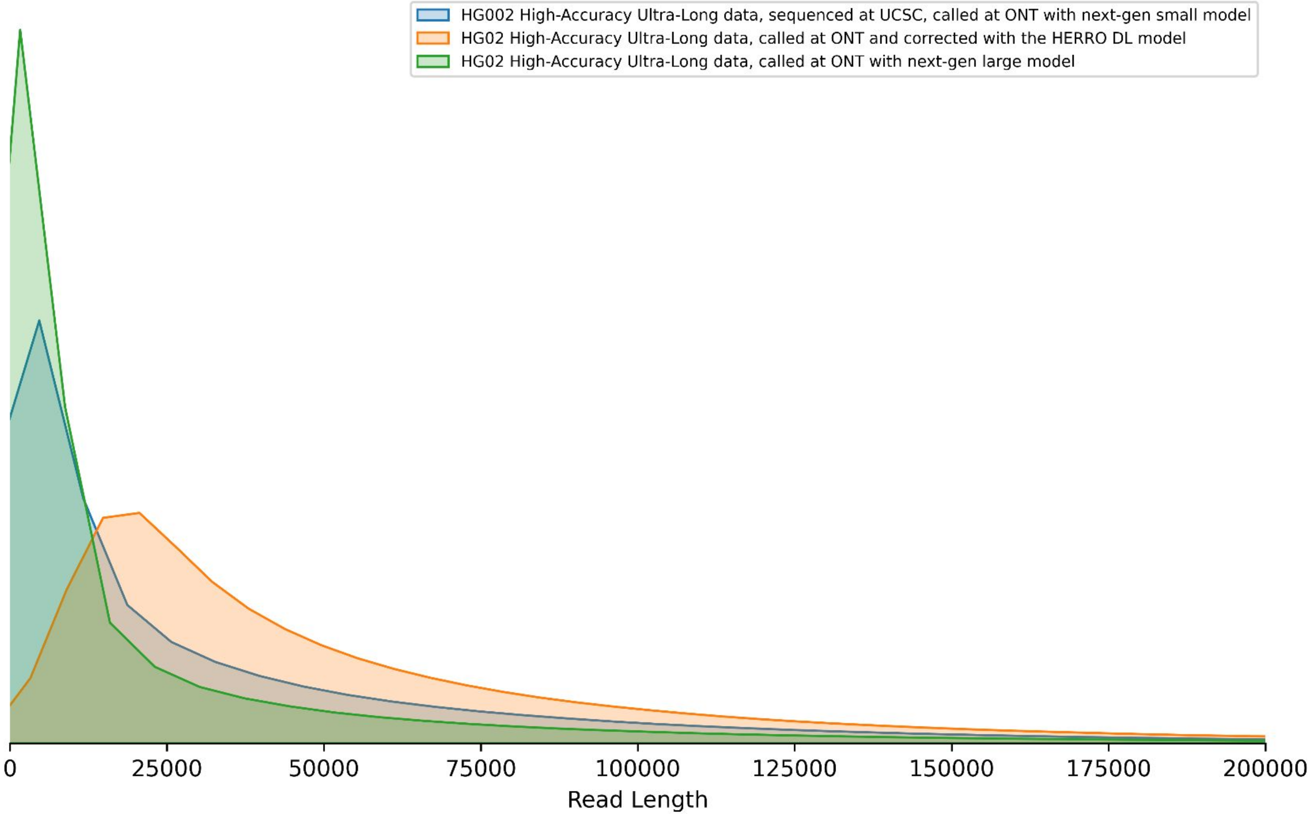
New ONT Q27 Chemistry (pre-release)

- Improved (unreleased) base-caller and updated chemistry for R10



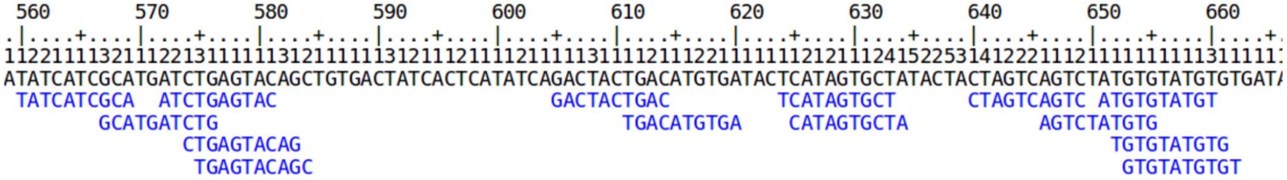
New ONT Q27 Chemistry (pre-release)

- Using ultra-long prep works nicely



Shasta, simplified

(1) Represent reads as “markers”, k=XX



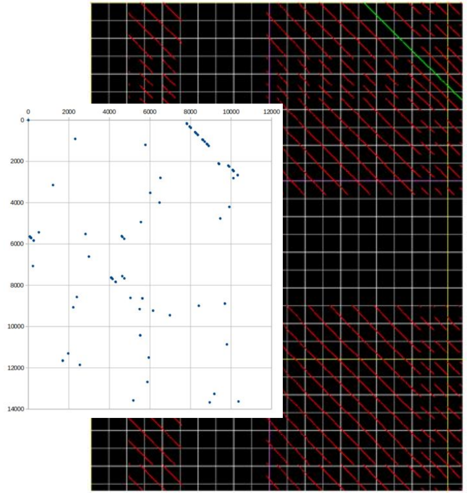
```
 560       570       580       590       600       610       620       630       640       650       660
.|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+.
112211113211122131111113121111113121112111121111131111211122111111121211124152253141222111211111111113111111
ATATCATCGCATGATCTGAGTACAGCTGTGACTATCACTCATATCAGACTACTGACATGTGATACTCATAGTGCTATACTACTAGTCAGTCTATGTGTATGTGTGATA
 TATCATCGCA  ATCTGAGTAC                       GACTACTGAC         TCATAGTGCT      CTAGTCAGTC ATGTGTATGT
        GCATGATCTG                                  TGACATGTGA    CATAGTGCTA           AGTCTATGTG
               CTGAGTACAG                                                                    TGTGTATGTG
                TGAGTACAGC                                                                    GTGTATGTGT
```

Shasta, simplified

(1) Represent reads as “markers”, k=XX

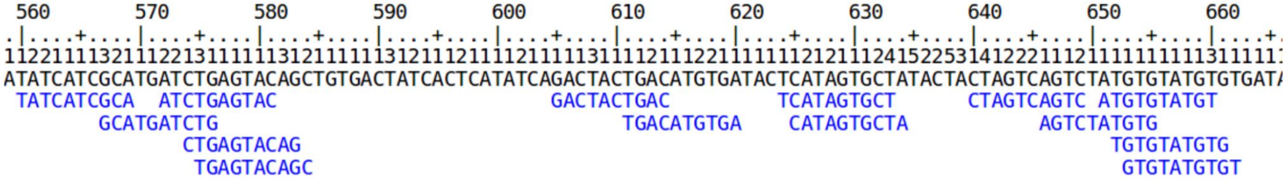
```
 560       570       580       590       600       610       620       630       640       650       660
.|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+.
112211113211122131111111312111111312111211112111113111121112211111112121112415225314122211121111111111311111
ATATCATCGCATGATCTGAGTACAGCTGTGACTATCACTCATATCAGACTACTGACATGTGATACTCATAGTGCTATACTACTAGTCAGTCTATGTGTATGTGTGATA
 TATCATCGCA  ATCTGAGTAC                       GACTACTGAC         TCATAGTGCT      CTAGTCAGTC ATGTGTATGT
        GCATGATCTG                                  TGACATGTGA    CATAGTGCTA           AGTCTATGTG
               CTGAGTACAG                                                                    TGTGTATGTG
                TGAGTACAGC                                                                    GTGTATGTGT
```

(2) MinHash/align reads as marker sequences



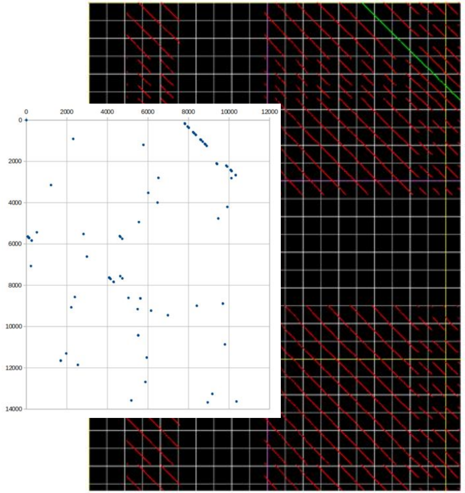
Shasta, simplified

(1) Represent reads as “markers”, k=XX

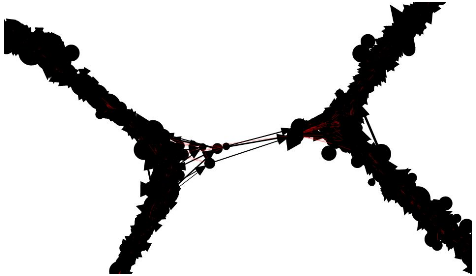


```
 560       570       580       590       600       610       620       630       640       650       660
.|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+.
11221111321112213111111131211111131211121111211111311112111221111111212111241522531412221112111111111311111
ATATCATCGCATGATCTGAGTACAGCTGTGACTATCACTCATATCAGACTACTGACATGTGATACTCATAGTGCTATACTACTAGTCAGTCTATGTGTATGTGTGATA
 TATCATCGCA  ATCTGAGTAC                       GACTACTGAC         TCATAGTGCT      CTAGTCAGTC ATGTGTATGT
        GCATGATCTG                                  TGACATGTGA    CATAGTGCTA           AGTCTATGTG
               CTGAGTACAG                                                                    TGTGTATGTG
                TGAGTACAGC                                                                    GTGTATGTGT
```

(2) MinHash/align reads as marker sequences

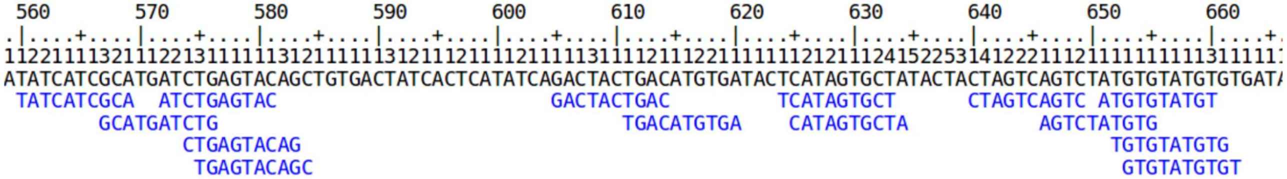


(3) Construct read overlap graph to prune overlaps



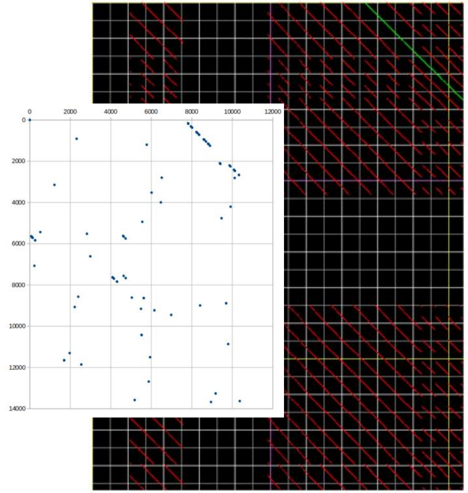
Shasta, simplified

(1) Represent reads as “markers”, k=XX

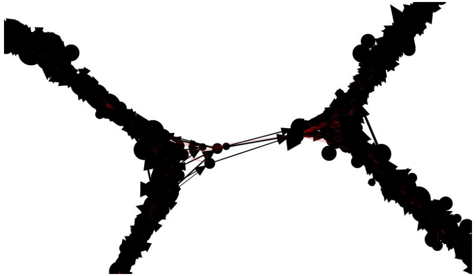


```
 560       570       580       590       600       610       620       630       640       650       660
.|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+.
112211113211122131111113121111113121112111211111311112111221111111212111241522531412221112111111111131111111
ATATCATCGCATGATCTGAGTACAGCTGTGACTATCACTCATATCAGACTACTGACATGTGATACTCATAGTGCTATACTACTAGTCAGTCTATGTGTATGTGTGATA
 TATCATCGCA  ATCTGAGTAC                       GACTACTGAC         TCATAGTGCT      CTAGTCAGTC ATGTGTATGT
        GCATGATCTG                                  TGACATGTGA    CATAGTGCTA           AGTCTATGTG
               CTGAGTACAG                                                                    TGTGTATGTG
                TGAGTACAGC                                                                    GTGTATGTGT
```

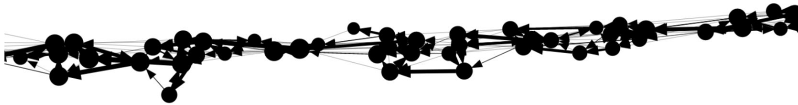
(2) MinHash/align reads as marker sequences



(3) Construct read overlap graph to prune overlaps

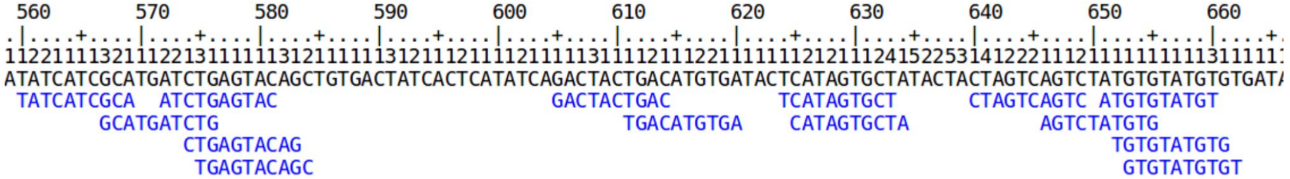


(4) Construct marker graph (MG) representing aligned reads



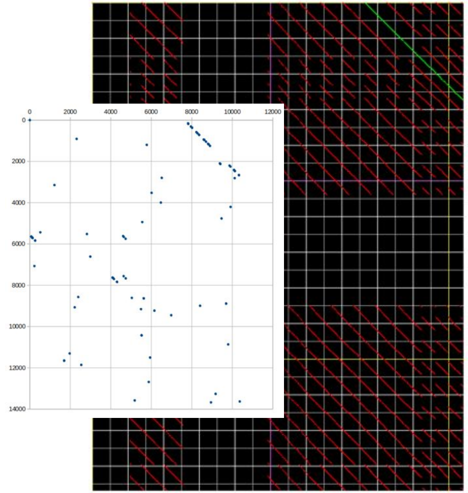
Shasta, simplified

(1) Represent reads as “markers”, k=XX

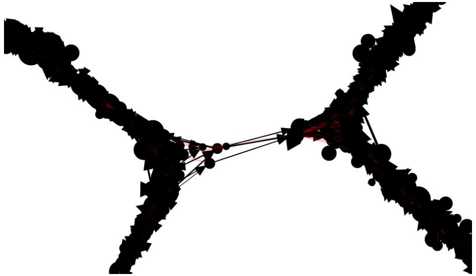


```
 560       570       580       590       600       610       620       630       640       650       660
.|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+....|....+.
112211113211122131111131211111113121112111121111131112111221111111212111241522531412221112111111111131111111
ATATCATCGCATGATCTGAGTACAGCTGTGACTATCACTCATATCAGACTACTGACATGTGATACTCATAGTGCTATACTACTAGTCAGTCTATGTGTATGTGTGATA
 TATCATCGCA  ATCTGAGTAC                       GACTACTGAC         TCATAGTGCT      CTAGTCAGTC ATGTGTATGT
        GCATGATCTG                                  TGACATGTGA    CATAGTGCTA           AGTCTATGTG
               CTGAGTACAG                                                                    TGTGTATGTG
                TGAGTACAGC                                                                    GTGTATGTGT
```

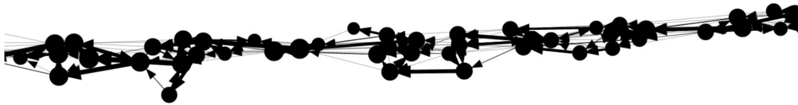
(2) MinHash/align reads as marker sequences



(3) Construct read overlap graph to prune overlaps



(4) Construct marker graph (MG) representing aligned reads



(5) NEW: Trace haplotypes in MG to assemble sequence - aka “Mode 3”

Shasta "Mode 3" assembly

- Released in preliminary form with Shasta 0.12.0.
- Despite known issues (to be improved on in future releases), produces useful phased assemblies using high accuracy nanopore reads from the ONT December 2023 data release (https://labs.epi2me.io/gm24385_ncm23_preview/) (referred to here as *ncm23*)
- Like previous Shasta releases, uses markers, MinHash, read graph, marker graph.
- Final sequence assembly is new.
    - Uses the marker graph to locate features that are unique to a single location+haplotype in the assembly.
    - "Read following" on these unique features.
    - Then uses local assemblies to assemble sequences between unique features.
- Invoked with *--config Nanopore-ncm23-May2024*
- Sequence assembly for a human genome takes 2-5 hours on a machine of appropriate size, depending on coverage.
- Memory requirement is currently 6 bytes per input base.
    - A 1 TB machine can run a human assembly at 50x.

Shasta assemblies

Two assemblies:

- An assembly at 38x using only the reads from the ONT release, with a 10 Kb read length cutoff.
- Total sequence assembled
- An assembly at 58x which also uses, in addition, a dataset sequenced at UCSC.
- “Single haplotype” sequence assembled is estimated based on assembled coverage

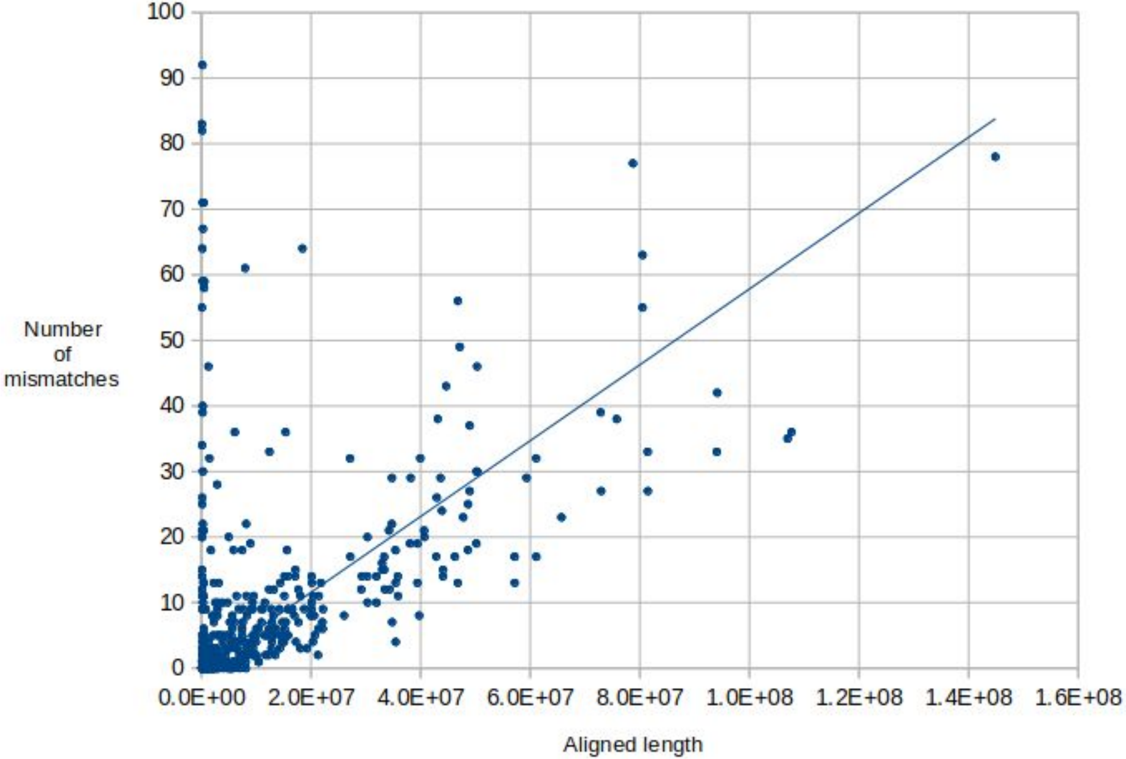
Coverage	Total sequence assembled (Mb)	N50 (Mb)	Total “single haplotype” sequence assembled (Mb)
38x	5885	16.5	5682
58x	5856	35.5	5675

Base level sequence quality

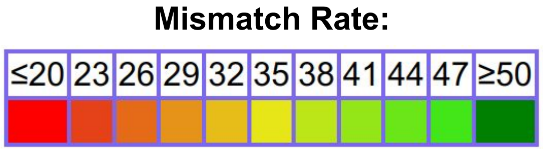
- “Single haplotype” assembled segments are mapped to the hg002v1.0.1 reference haplotypes w/Minimap2 asm10.
- Most segments map in a single mapping.
- Count the number of mismatched, inserted, deleted bases in each alignment.
- Least square fit with constrained origin gives an estimate of mismatch, insert, delete rate.
- Mismatch rate is an overestimate because of mismatches that occur in alignments as part or complex indels.
- Insert/delete rates are dominated by long homopolymer runs.

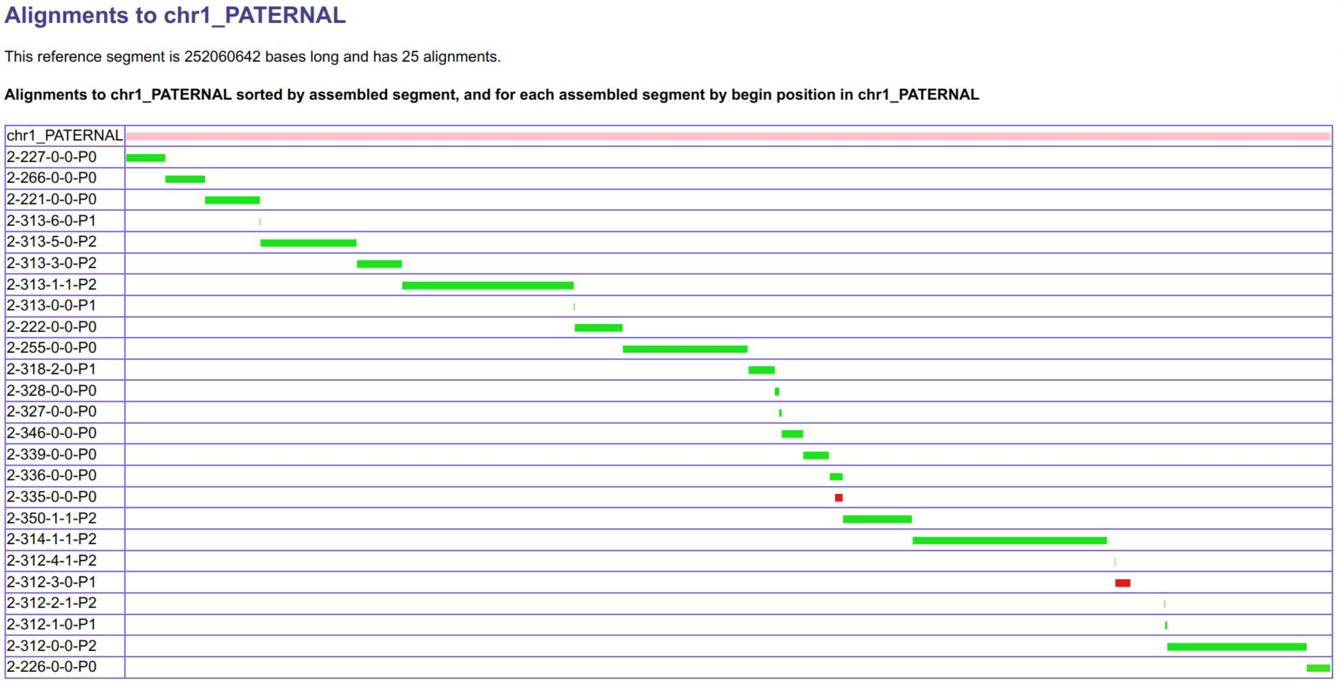
	38x	58x
Mismatch Q	60.0	62.4
Insert Q	44.4	44.9
Delete Q	38.4	36.0

Scatter plot for mismatches (58x assembly)



Assembled contig mappings to the T2T Assembly

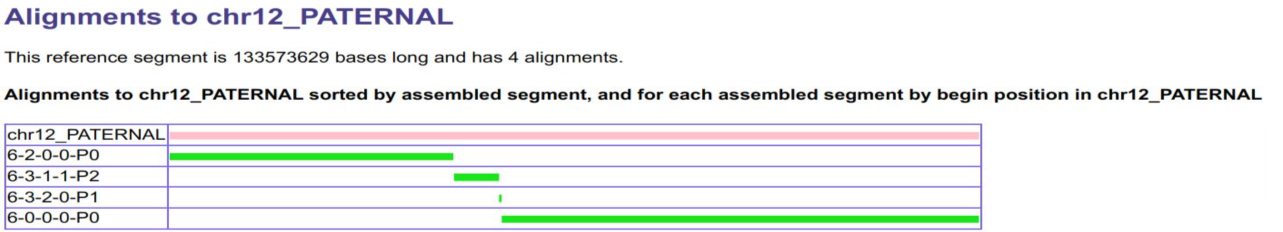




Alignments to chr1_PATERNAL

This reference segment is 252060642 bases long and has 25 alignments.

Alignments to chr1_PATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr1_PATERNAL



Alignments to chr12_PATERNAL

This reference segment is 133573629 bases long and has 4 alignments.

Alignments to chr12_PATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr12_PATERNAL

Comparing to Hifiasm with ONT

We used **38x** and **58x** coverage ONT Ultra-Long datasets.

All Hifiasm assemblies where generated using the latest Hifiasm-0.19.9-r616 release.

1. Hifiasm was first used to generate error corrected reads (using the --write-ec parameter) and coverage estimates.
2. Hifiasm was then invoked with --dbg-ovec to generate all-vs-all read overlaps
3. Then, cis and trans overlaps were merged
4. The *RAFT algorithm fragments the error corrected reads. The RAFT (Repeat Aware Fragmentation Tool) is an algorithm designed to improve assembly quality by rescuing contained reads.
5. The final Hifiasm run generates the assembly of the fragmented error-corrected reads using a single round of error correction (-r1 parameter). The newly announced parameter “--telo-m CCCTAA” is also used to keep telomeres at the ends of contigs/scaffolds.
6. Hi-C data can optionally be integrated during the final assembly step

* Sudhanva Shyam Kamath, Mehak Bindra, Debnath Pal, Chirag Jain, Telomere-to-telomere assembly by preserving contained reads. bioRxiv 2023.11.07.565066; doi:10.1101/2023.11.07.565066

Assembly Stats

		Coverage: 38x		
	HG002 T2T	**SHASTA**	**HIFIASM RAFT HERRO**	**HIFIASM RAFT**
Assembled Length (Mb)	6,000	5,885	6,049	6,063
N50 (Mb)	147	16.4	82.8	64
L50	16	102	27	30
# of sequences	48	21,859*	395	1,613

* Shasta uses a philosophy of outputting everything regardless of length and local complexity of the assembly graph, leaving it to the user to decide what is meaningful.

Assembly Stats

		Coverage: 58x	
	HG002 T2T	**SHASTA**	**HIFIASM RAFT HERRO**
Assembled Length (Mb)	6,000	5,856	6,011
N50 (Mb)	147	35.4	84
L50	16	53	29
# of sequences	48	16,542*	818

* Shasta uses a philosophy of outputting everything regardless of length and local complexity of the assembly graph, leaving it to the user to decide what is meaningful.

Mapping to the T2T Assembly

We mapped the assembled contigs back to the T2T HG002 v1.0.1 reference genome with the latest Minimap2 v2.28 using the “asm10” preset and evaluated the primary alignments

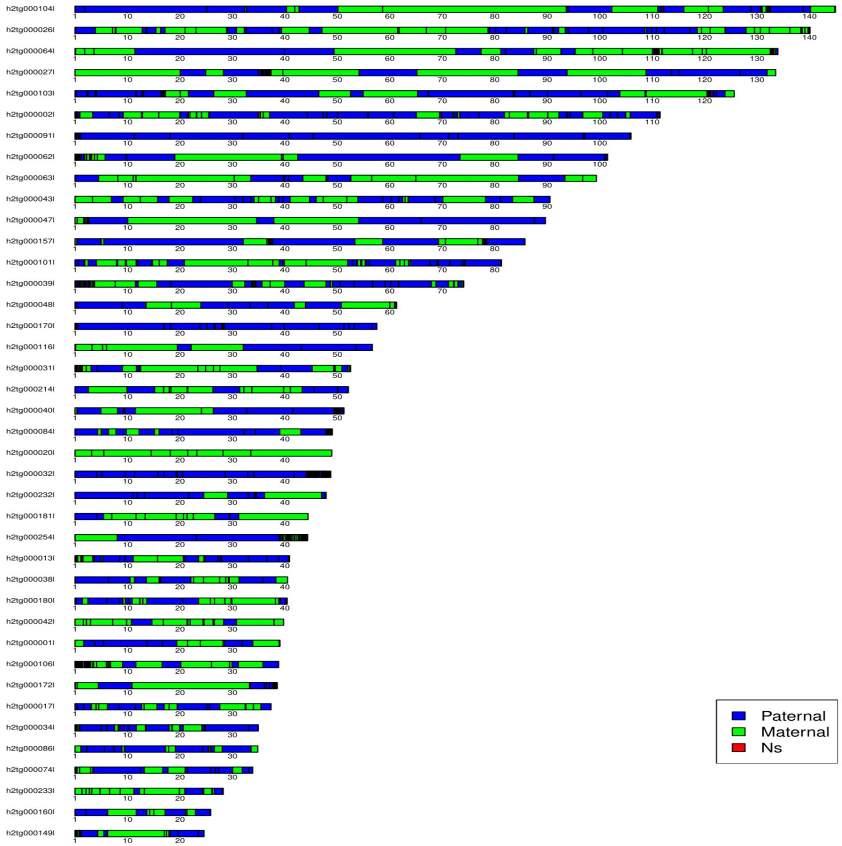
If a contig has a mix of maternal and paternal alleles, it might align to either the maternal or the paternal chromosome

OR

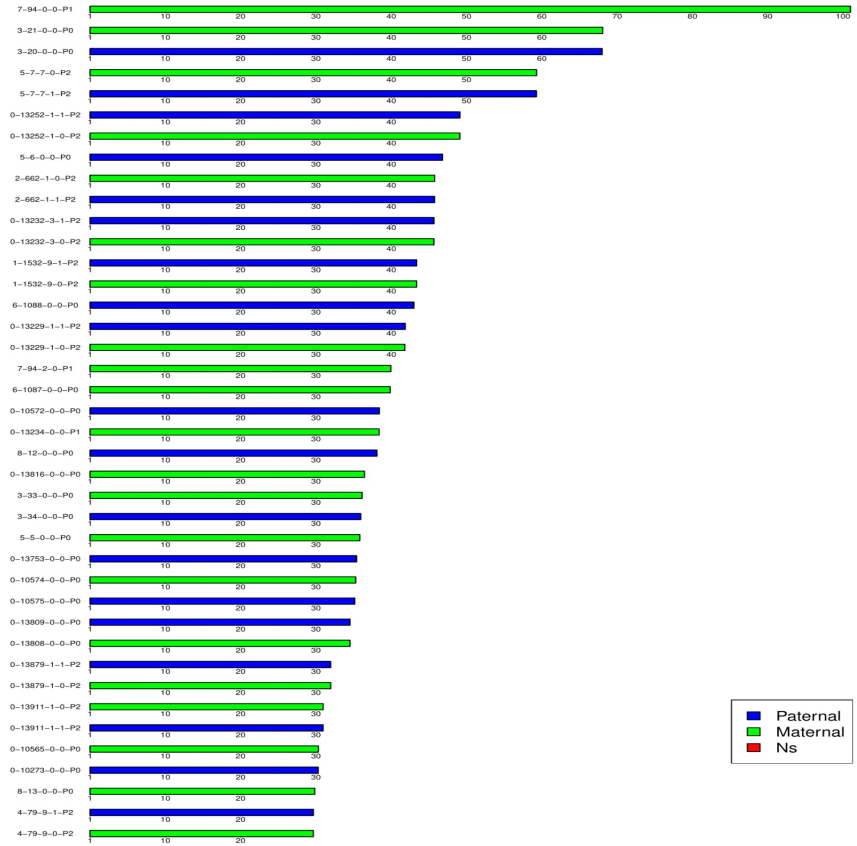
it could split and have parts of it aligned to one haplotype and parts to the other haplotype

Assembled contig mappings to the T2T Assembly

Hifiasm with 38X ONT UL



Shasta with 38X ONT UL



Hifiasm 38X ONT UL contig mappings to the T2T assembly

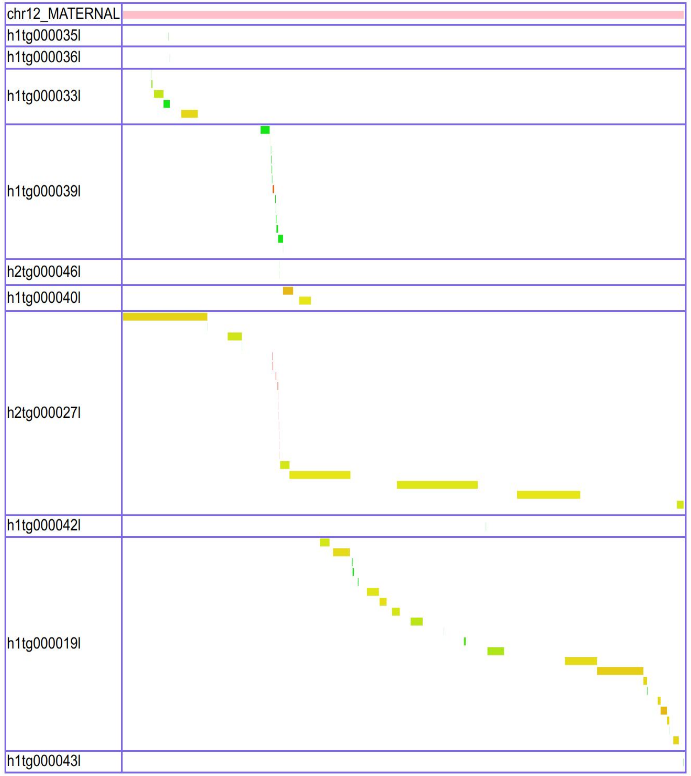
Mismatch Rate:

≤20	23	26	29	32	35	38	41	44	47	≥50
🟥	🟥	🟧	🟧	🟧	🟨	🟨	🟩	🟩	🟩	🟩

Alignments to chr12_MATERNAL

This reference segment is 133580598 bases long and has 67 alignments.

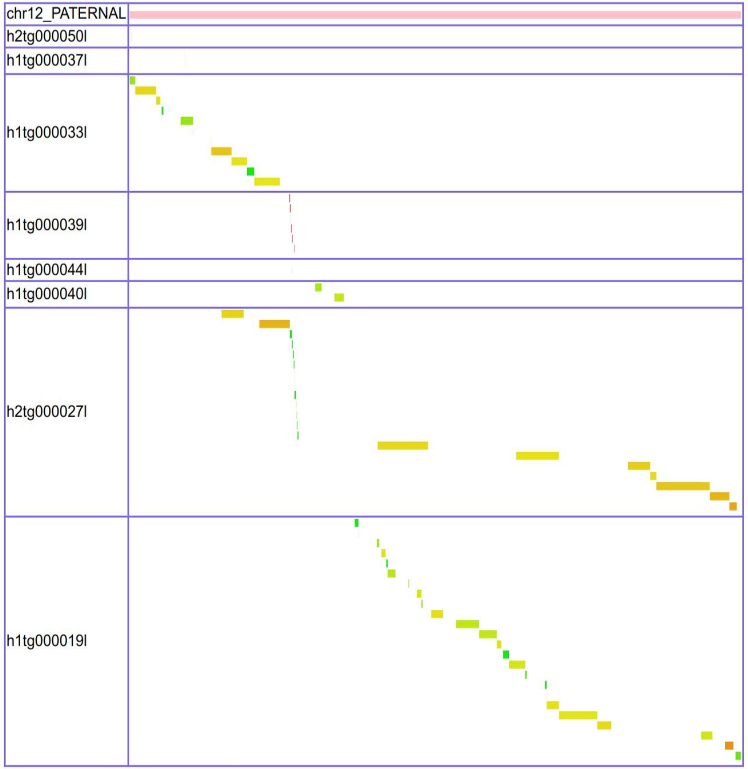
Alignments to chr12_MATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr12_MATERNAL



Alignments to chr12_PATERNAL

This reference segment is 133573629 bases long and has 67 alignments.

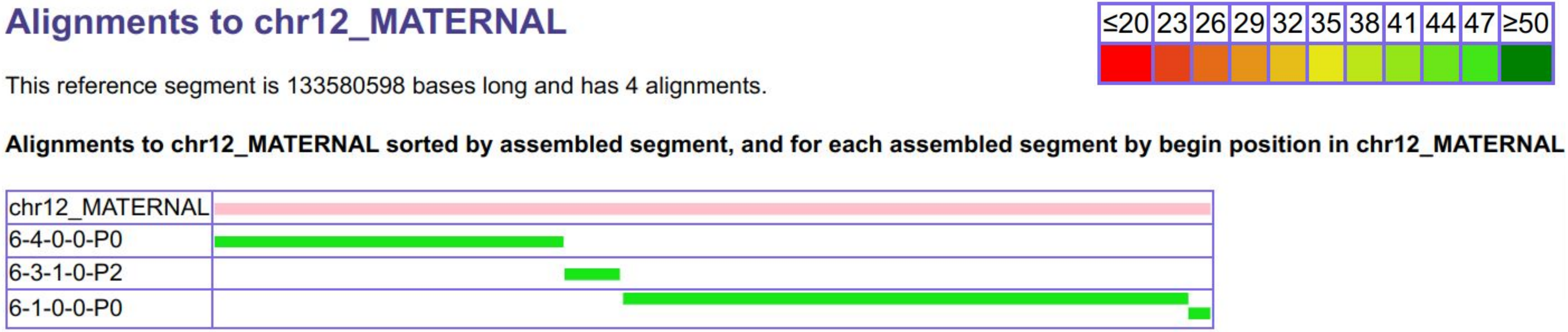
Alignments to chr12_PATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr12_PATERNAL



Shasta 38X ONT UL contig mappings to the T2T assembly

Mismatch Rate:

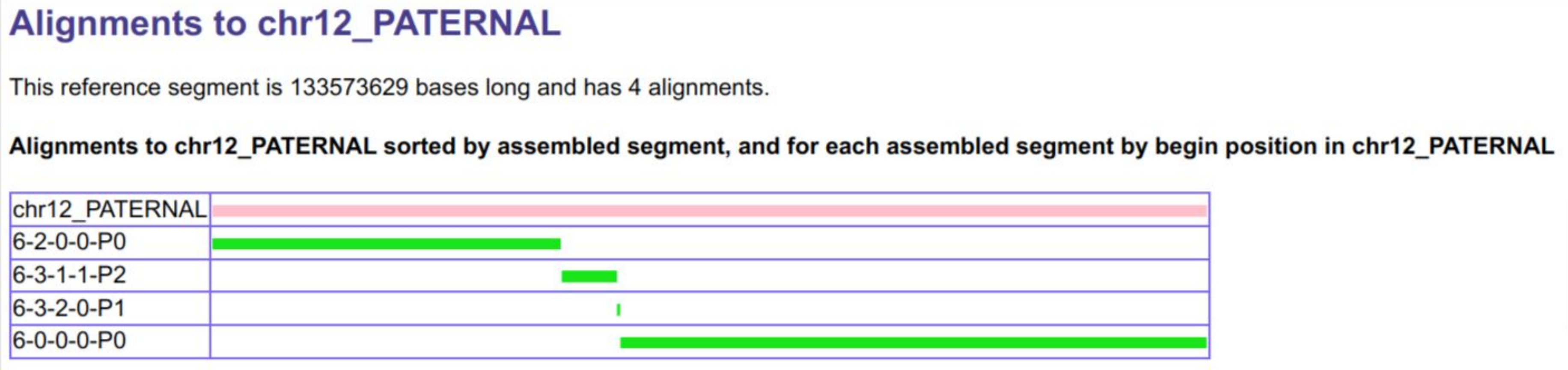
≤20	23	26	29	32	35	38	41	44	47	≥50
🟥	🟥	🟧	🟧	🟨	🟨	🟩	🟩	🟩	🟩	🟩



Alignments to chr12_MATERNAL

This reference segment is 133580598 bases long and has 4 alignments.

Alignments to chr12_MATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr12_MATERNAL



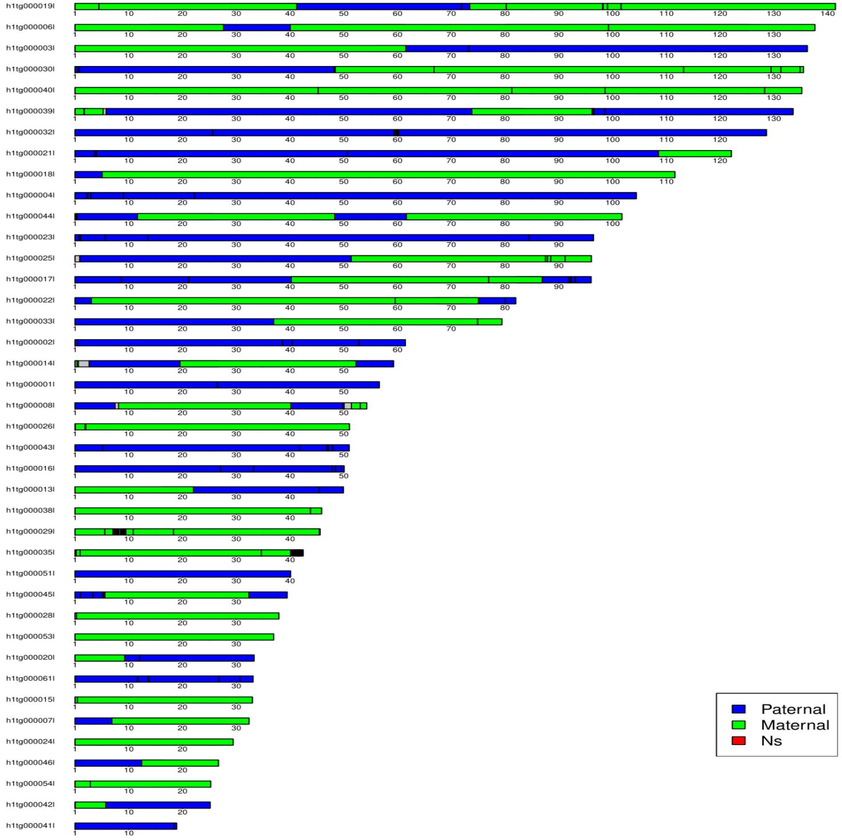
Alignments to chr12_PATERNAL

This reference segment is 133573629 bases long and has 4 alignments.

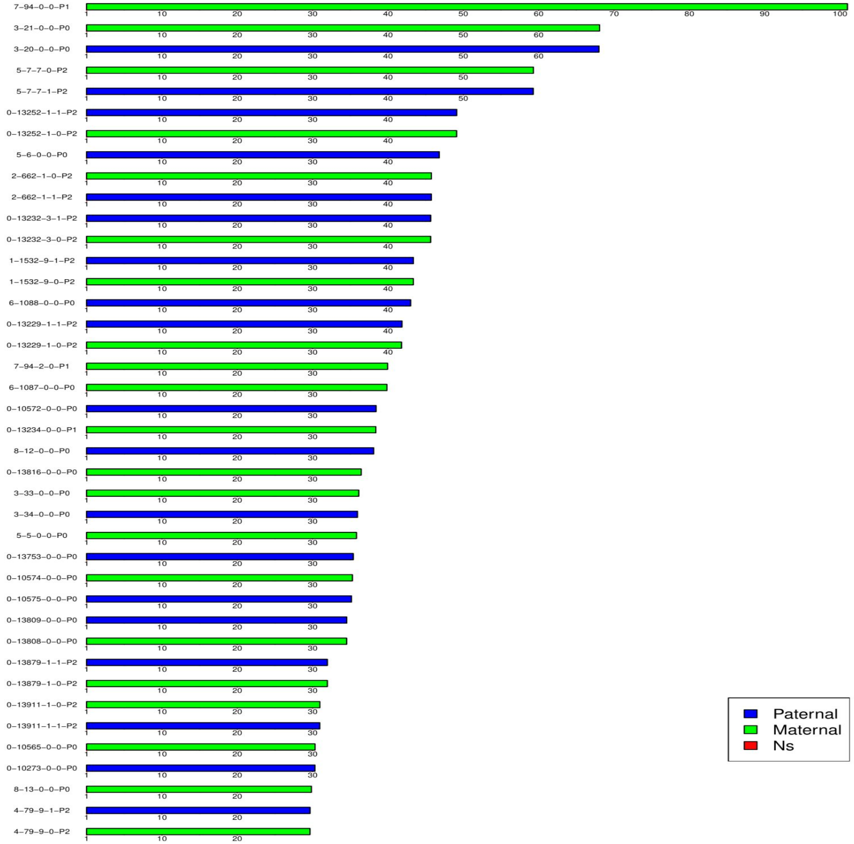
Alignments to chr12_PATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr12_PATERNAL

Assembled contig mappings to the T2T Assembly

Hifiasm with 38X ONT UL Data HERRO Corrected



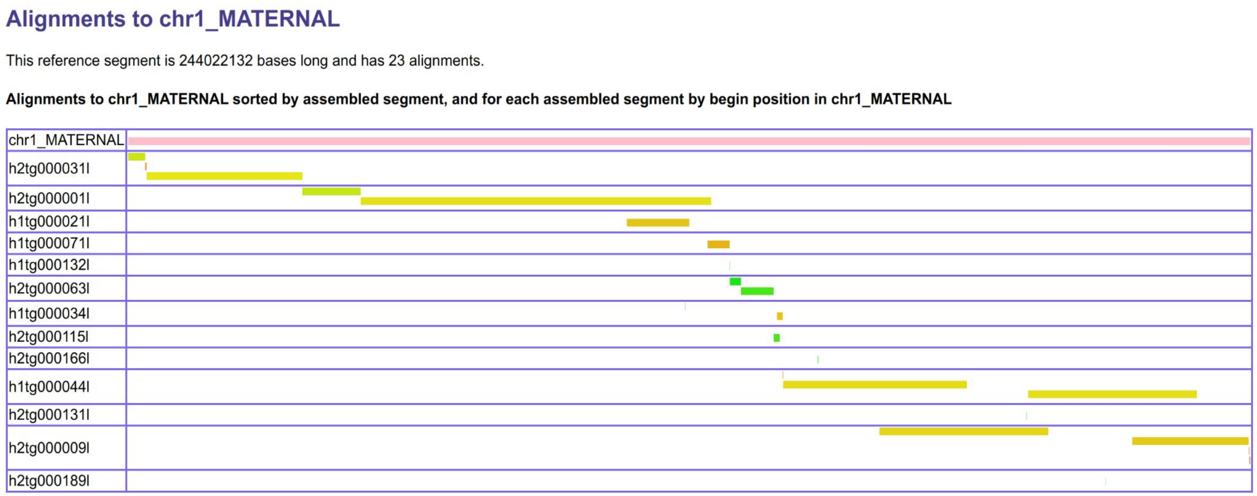
Shasta with 38X ONT UL



Mismatch Rate:

≤20	23	26	29	32	35	38	41	44	47	≥50
🟥	🟥	🟧	🟧	🟨	🟨	🟩	🟩	🟩	🟩	🟩

Hifiasm 38X ONT UL HERRO Corrected



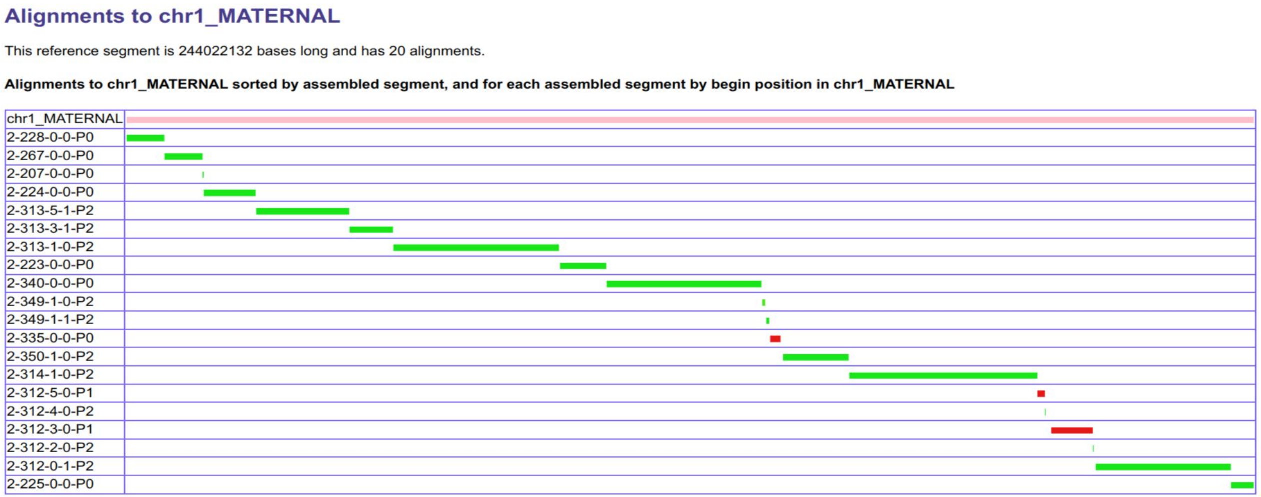
Alignments to chr1_MATERNAL

This reference segment is 244022132 bases long and has 23 alignments.

Alignments to chr1_MATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr1_MATERNAL

chr1_MATERNAL
h2tg000031l
h2tg000001l
h1tg000021l
h1tg000071l
h1tg000132l
h2tg000063l
h1tg000034l
h2tg000115l
h2tg000166l
h1tg000044l
h2tg000131l
h2tg000009l
h2tg000189l

Shasta 38X ONT UL



Alignments to chr1_MATERNAL

This reference segment is 244022132 bases long and has 20 alignments.

Alignments to chr1_MATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr1_MATERNAL

chr1_MATERNAL
2-228-0-0-P0
2-267-0-0-P0
2-207-0-0-P0
2-224-0-0-P0
2-313-5-1-P2
2-313-3-1-P2
2-313-1-0-P2
2-223-0-0-P0
2-340-0-0-P0
2-349-1-0-P2
2-349-1-1-P2
2-335-0-0-P0
2-350-1-0-P2
2-314-1-0-P2
2-312-5-0-P1
2-312-4-0-P2
2-312-3-0-P1
2-312-2-0-P2
2-312-0-1-P2
2-225-0-0-P0

Compleasm*

Model Organism:
H. sapiens
Lineage Gene Set:
primates_odb10

		38x ONT Ultra-Long reads		
N = 13780	**HG002 T2T**	**SHASTA**	**HIFIASM RAFT HERRO**	**HIFIASM RAFT**
Single Copy	470 (3.41%)	951 (6.90%)	340 (2.47%)	452 (3.28%)
Duplicated	13,299 (95.51%)	12,779 (92.74%)	13,428 (97.45%)	13,317 (96.64%)
Fragmented	7 (0.05%)	26 (0.19%)	8 (0.06%)	7 (0.05%)
Missing	4 (0.03%)	24 (0.17%)	4 (0.03%)	4 (0.03%)

* Neng Huang, Heng Li, compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics, 39, btad595, 2023. doi:10.1093/bioinformatics/btad595

Compleasm*

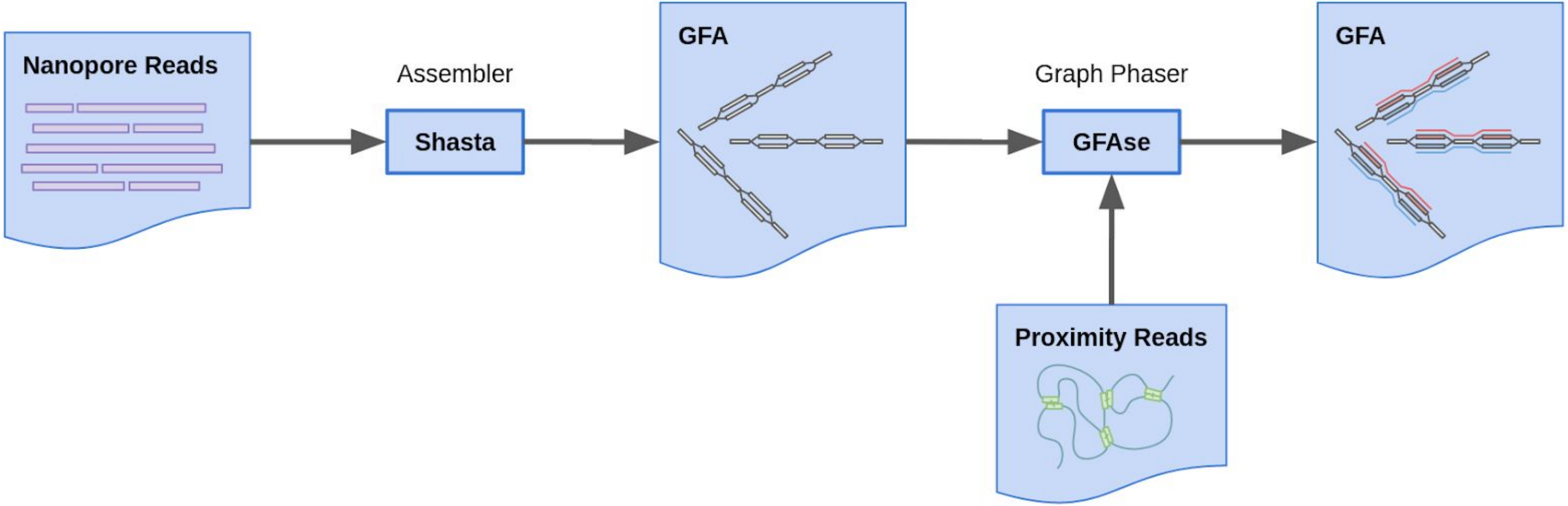
Model Organism:
H. sapiens
Lineage Gene Set:
primates_odb10

		58x ONT Ultra-Long reads	
N = 13780	**HG002 T2T**	**SHASTA**	**HIFIASM RAFT HERRO**
Single Copy	470 (3.41%)	997 (7.24%)	519 (3.77%)
Duplicated	13,299 (95.51%)	12,756 (92.57%)	13,250 (96.15%)
Fragmented	7 (0.05%)	13 (0.10%)	7 (0.05%)
Missing	4 (0.03%)	14 (0.10%)	4 (0.03%)

* Neng Huang, Heng Li, compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics, 39, btad595, 2023. doi:10.1093/bioinformatics/btad595

Shasta + GFAse

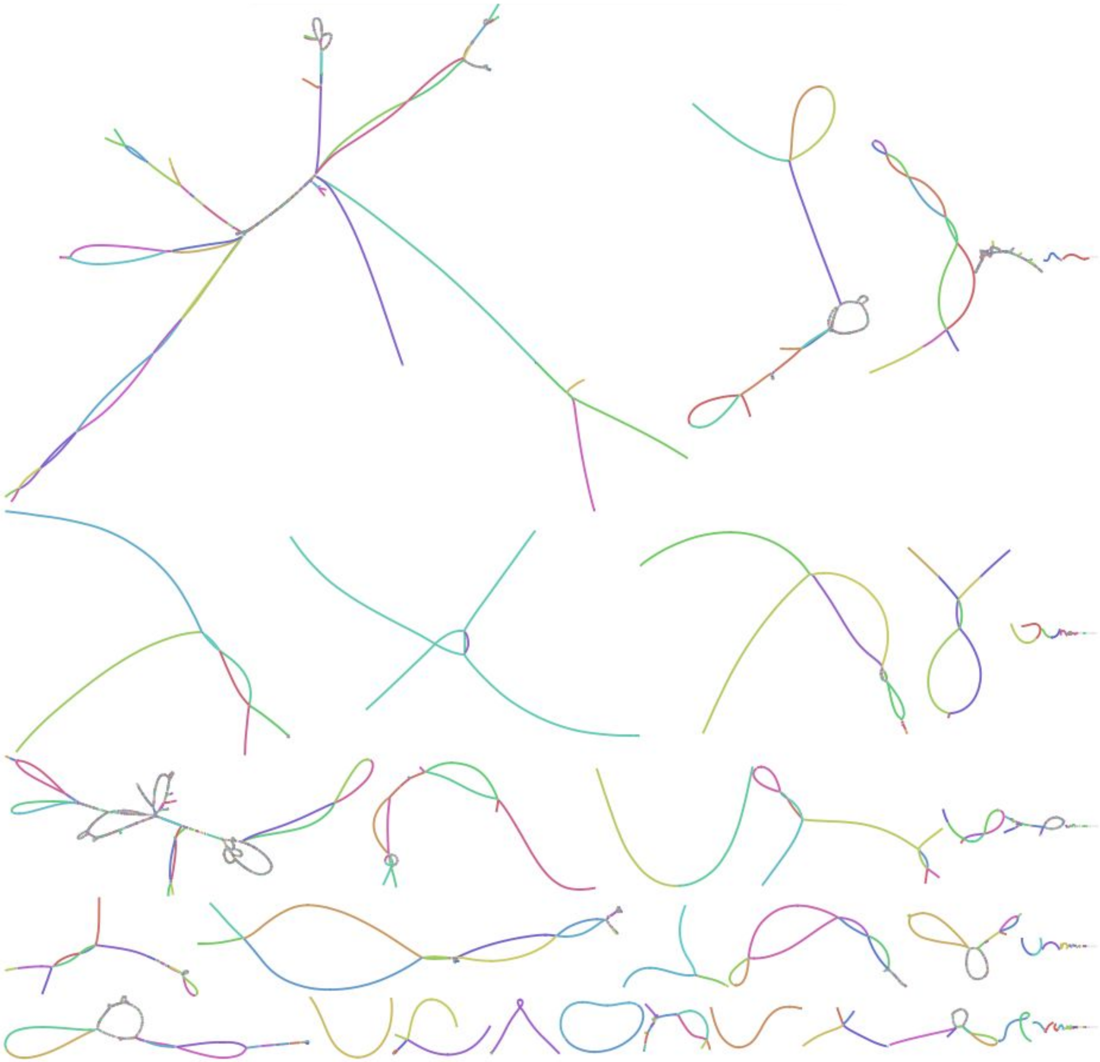
We further phased the assemblies with Hi-C using GFAse



See: Phased nanopore assembly with Shasta and modular graph phasing with GFAse, Lorig-Roach et al. Genome Research, 2024

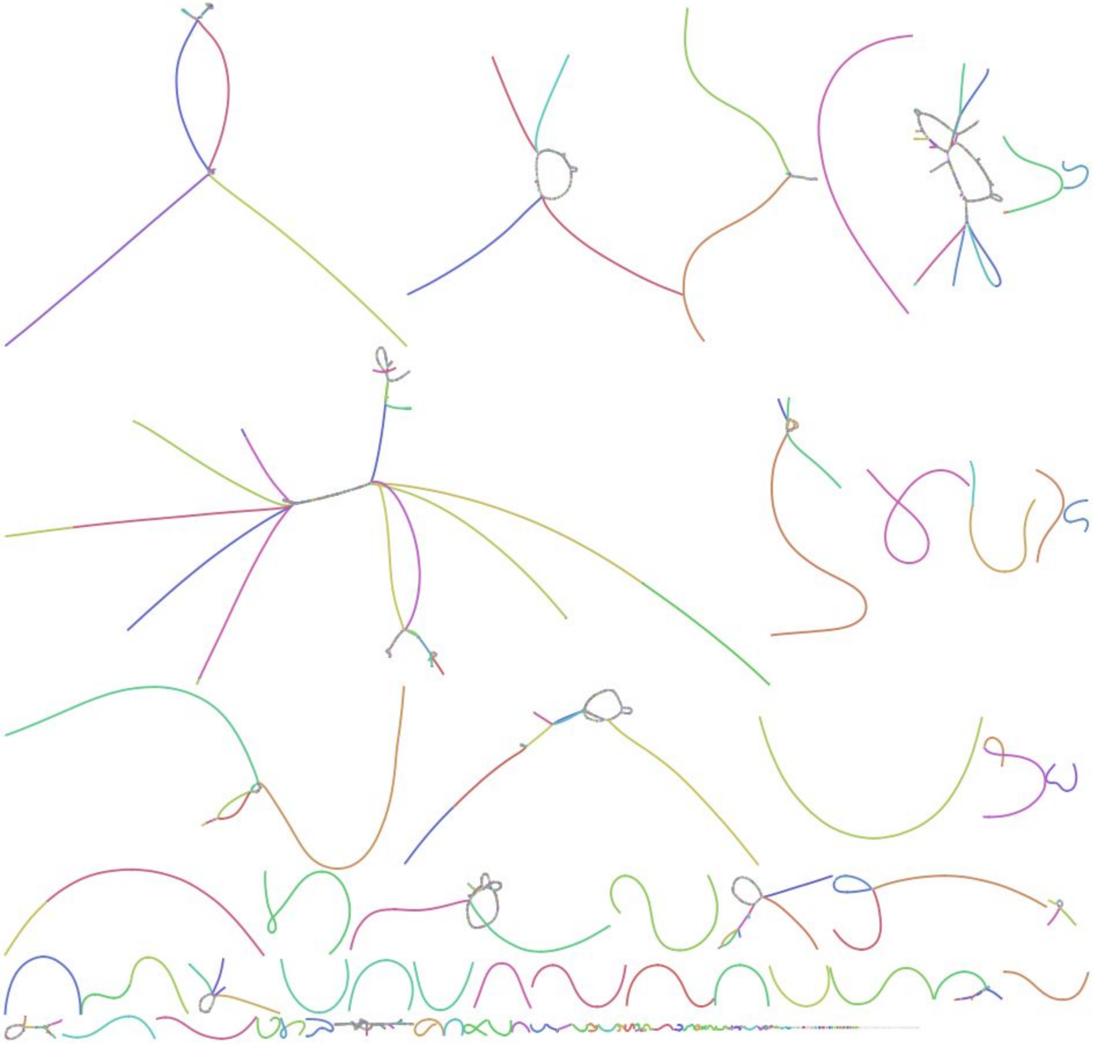
Shasta (58x)

- Bandage plot of assembly graph
- Before GFAse

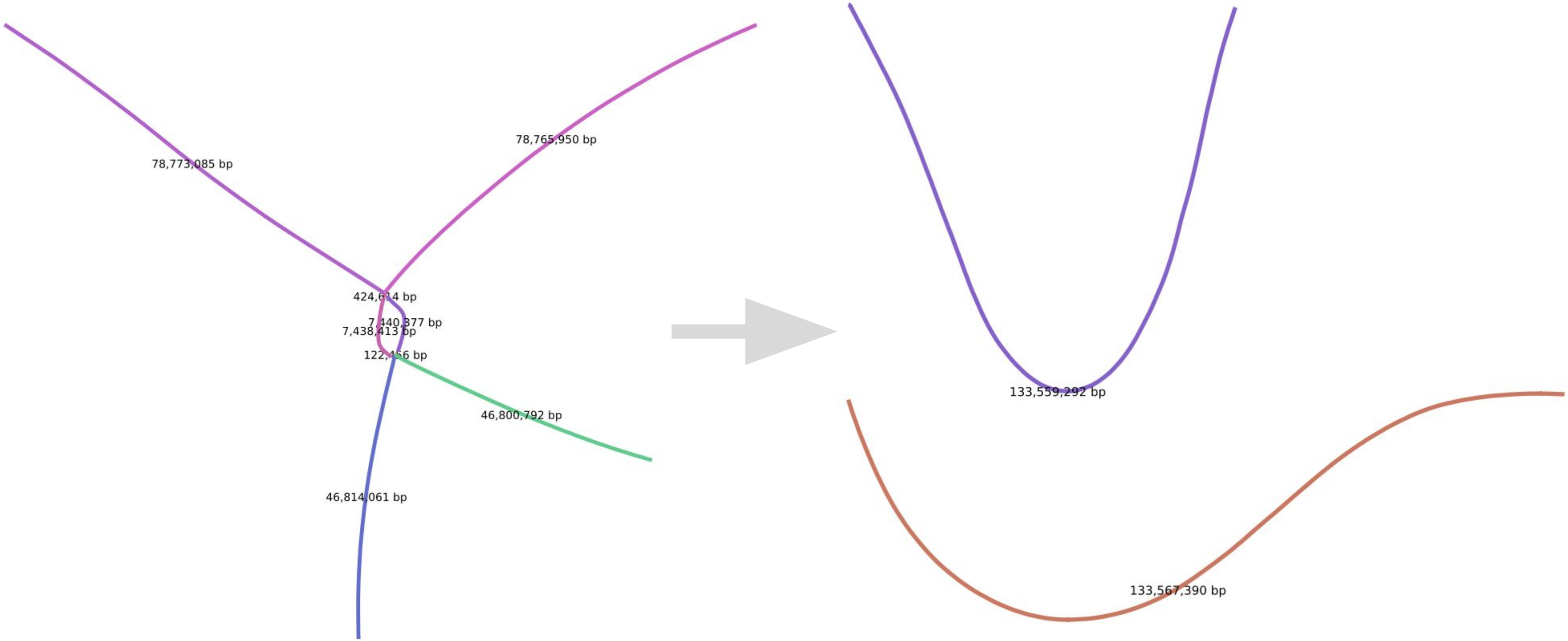


Shasta (58x)

- Bandage plot of assembly graph
- After GFAse



Shasta + GFAse



Assembly Stats with Hi-C

		Coverage: 38x		
	HG002 T2T	**SHASTA + GFAse**	**HIFIASM RAFT HERRO**	**HIFIASM RAFT**
Assembled Length (Mb)	6,000	5,966	5,997	6,044
N50 (Mb)	147	54,9	79,3	61,7
L50	16	33	29	32
# of sequences	48	21,130*	401	1,673

* Shasta uses a philosophy of outputting everything regardless of length and local complexity of the assembly graph, leaving it to the user to decide what is meaningful.

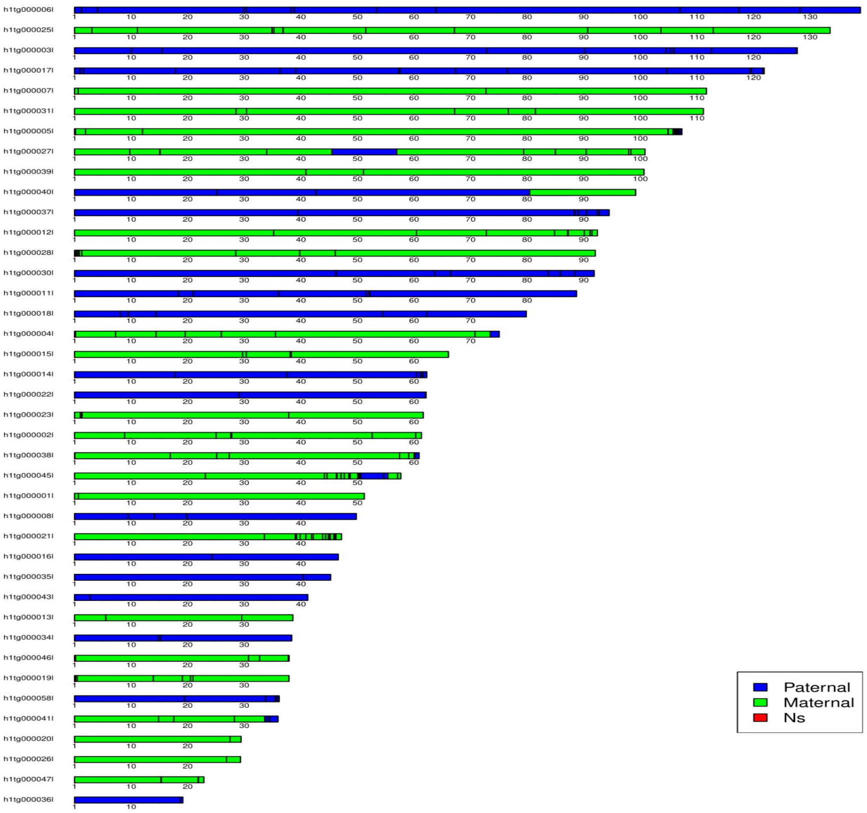
Assembly Stats with Hi-C

		Coverage: 58x	
	HG002 T2T	**SHASTA + GFAse**	**HIFIASM RAFT HERRO**
Assembled Length (Mb)	6,000	5,951	6,022
N50 (Mb)	147	70.5	79.8
L50	16	29	29
# of sequences	48	16,127*	846

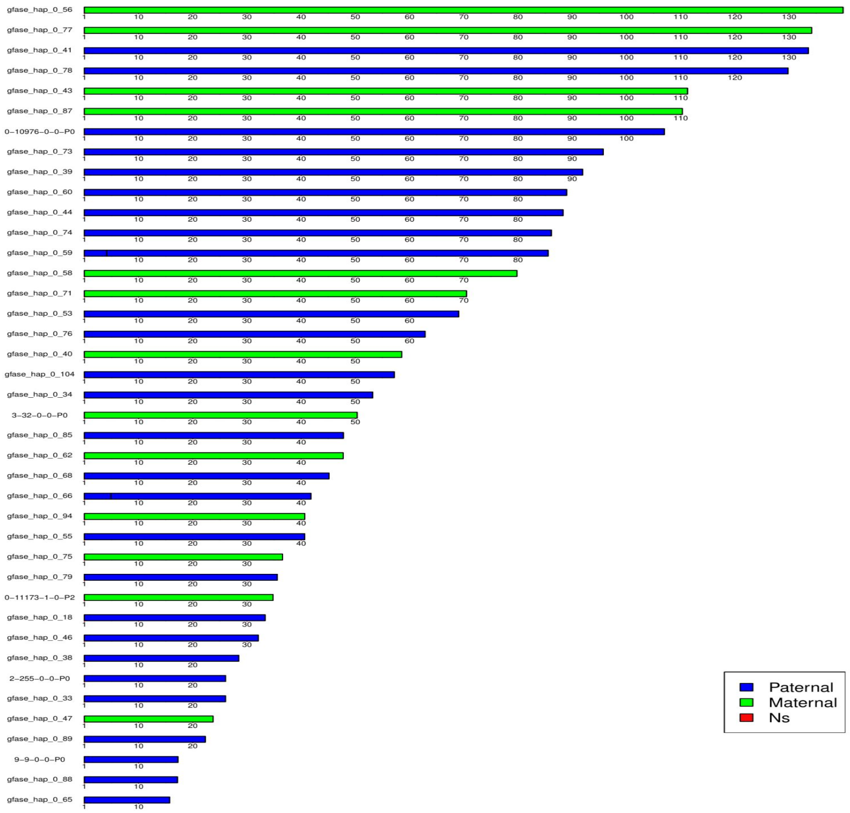
* Shasta uses a philosophy of outputting everything regardless of length and local complexity of the assembly graph, leaving it to the user to decide what is meaningful.

Assembled contig mappings to the T2T Assembly

Hifiasm with 58X ONT UL Data HERRO Corrected + HiC



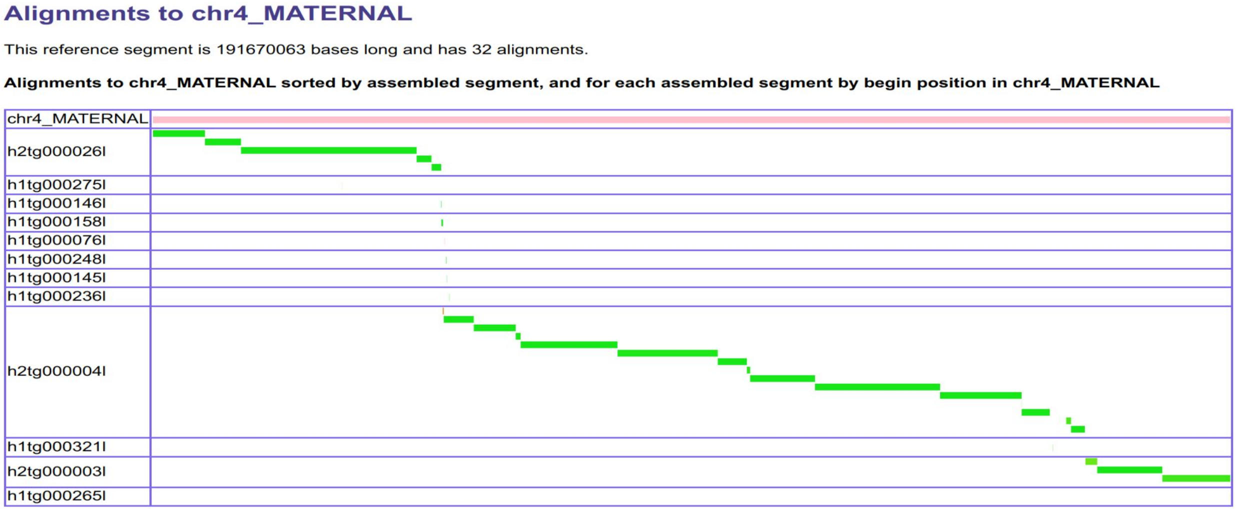
Shasta with 58X ONT UL Data + GFAse with HiC



Mismatch Rate:

≤20	23	26	29	32	35	38	41	44	47	≥50
🟥	🟥	🟧	🟧	🟨	🟨	🟨	🟩	🟩	🟩	🟩

Hifiasm with 58X ONT UL HERRO Corrected + HiC

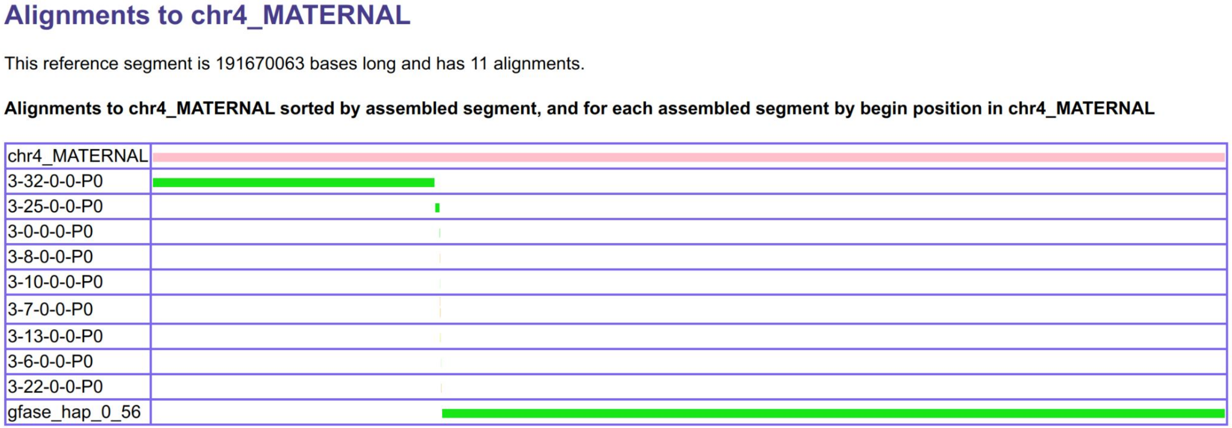


Alignments to chr4_MATERNAL

This reference segment is 191670063 bases long and has 32 alignments.

Alignments to chr4_MATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr4_MATERNAL

Shasta with 58X ONT UL + GFAse with HiC



Alignments to chr4_MATERNAL

This reference segment is 191670063 bases long and has 11 alignments.

Alignments to chr4_MATERNAL sorted by assembled segment, and for each assembled segment by begin position in chr4_MATERNAL

Compleasm*

Model Organism:
H. sapiens
Lineage Gene Set:
primates_odb10

		38x ONT Ultra-Long reads + 2 Hi-C FlowCell libraries		
N = 13780	**HG002 T2T**	**SHASTA**	**HIFIASM RAFT HERRO**	**HIFIASM RAFT**
Single Copy	470 (**3.41%**)	482 (**3.5%**)	480 (**3.48%**)	595 (**4.32%**)
Duplicated	13,299 (**95.51%**)	13,283 (**96.39%**)	13,288 (**96.43%**)	13,174 (**95.60%**)
Fragmented	7 (**0.05%**)	9 (**0.07%**)	8 (**0.06%**)	7 (**0.05%**)
Missing	4 (**0.03%**)	6 (**0.04%**)	4 (**0.03%**)	4 (**0.03%**)

* Neng Huang, Heng Li, compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics, 39, btad595, 2023. doi:10.1093/bioinformatics/btad595

Compleasm*

Model Organism:
H. sapiens
Lineage Gene Set:
primates_odb10

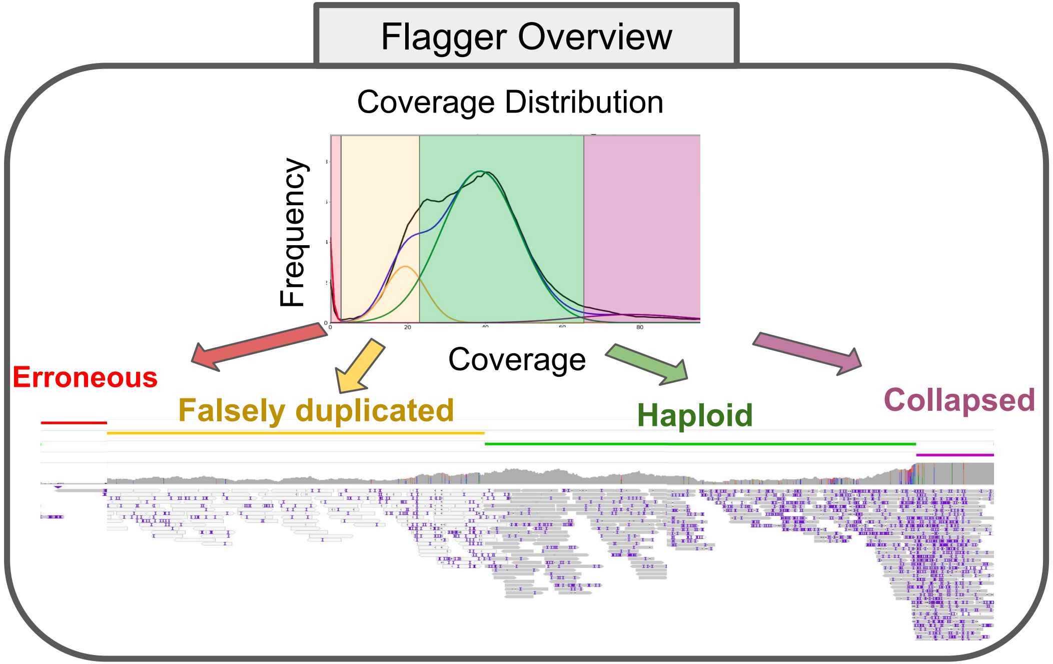
		58x ONT Ultra-Long reads + 2 Hi-C FlowCell libraries	
N = 13780	**HG002 T2T**	**SHASTA**	**HIFIASM RAFT HERRO**
Single Copy	470 (**3.41%**)	471 (**3.42%**)	516 (**3.74%**)
Duplicated	13,299 (**95.51%**)	13,296 (**96.49%**)	13,253 (**96.18%**)
Fragmented	7 (**0.05%**)	8 (**0.06%**)	7 (**0.05%**)
Missing	4 (**0.03%**)	5 (**0.04%**)	4 (**0.03%**)

* Neng Huang, Heng Li, compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics, 39, btad595, 2023. doi:10.1093/bioinformatics/btad595

Assembly QC: Flagger :
A read-mapping-based pipeline for assessing diploid assemblies

- Flagger takes **long reads (ONT or HiFI)** mapped to the diploid assembly in a haplotype-aware manner and finds read depth of coverages along the assembly.

- It then uses a **Gaussian Mixture Model** to infer the coverage boundaries for
  - Well-assembled blocks (**Haploid**)
  - and 3 kinds of unreliable blocks which can be either
    - **Erroneous,**
    - **Falsely duplicated**
    - **Collapsed**



Flagger Overview

Coverage Distribution

Erroneous

Falsely duplicated

Haploid

Collapsed

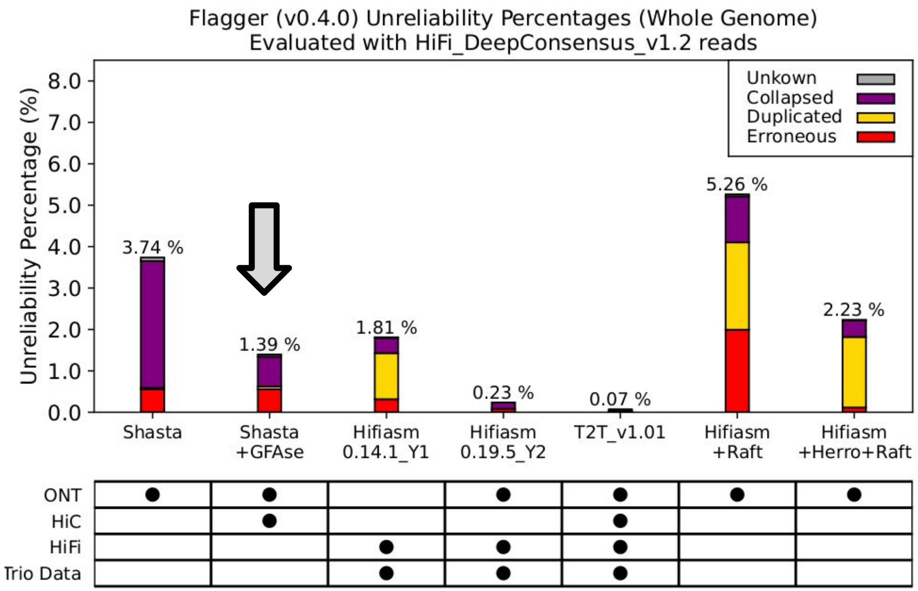
github.com/mobinasri/flagger

Benchmarking Shasta and GFAse assemblies with Flagger

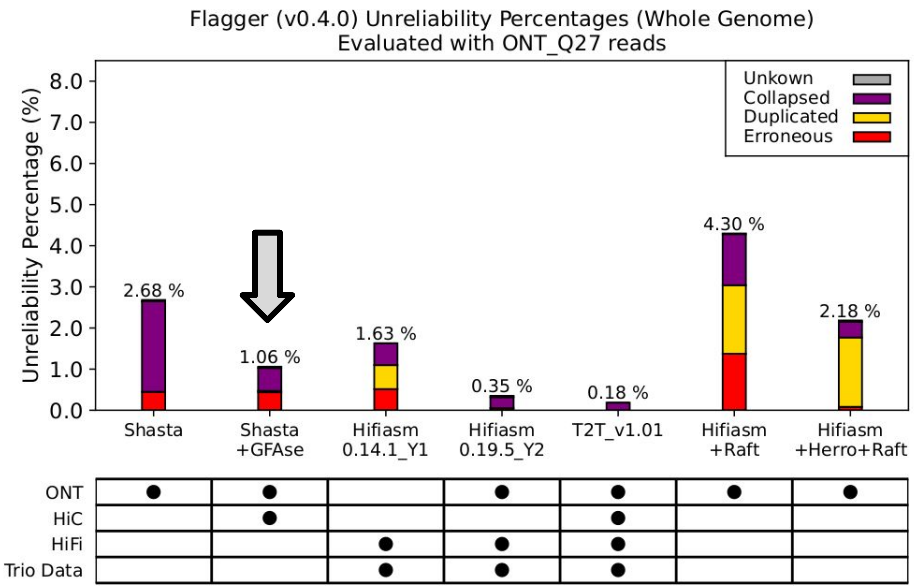
Results For Whole Genome

- Flagger results using both HiFi and ONT reads confirm that Shasta+GFAse assemblies have comparable structural accuracy with HPRC-Year1 assemblies produced with HiFiasm assembler.

- Recent version of Hifiasm assembler outperforms Shasta+GFAse partly due to employing high accuracy HiFi reads and taking phasing information from parental reads, which are not used by Shasta+GFAse.



	Shasta	Shasta +GFAse	Hifiasm 0.14.1_Y1	Hifiasm 0.19.5_Y2	T2T_v1.01	Hifiasm +Raft	Hifiasm +Herro+Raft
ONT	●	●		●	●	●	●
HiC		●			●		
HiFi			●	●	●		
Trio Data			●	●	●		

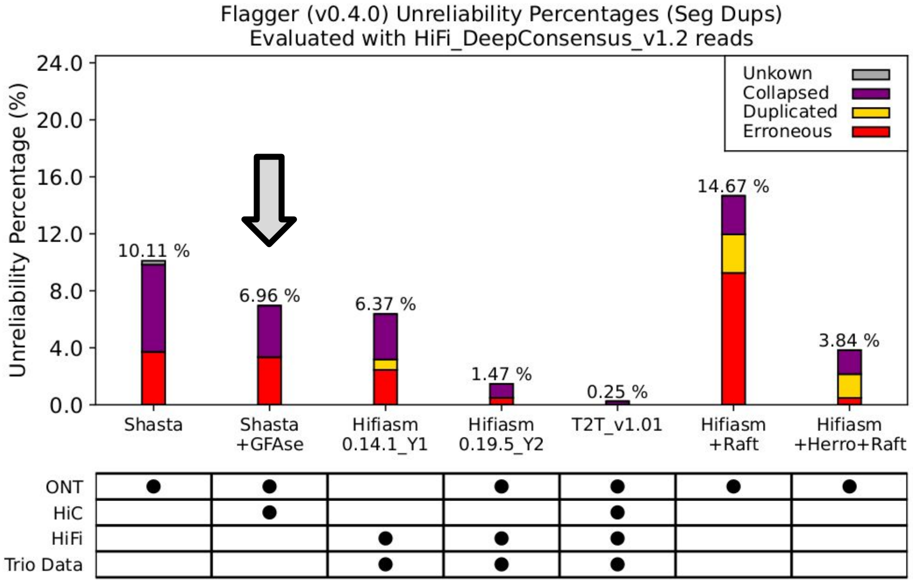


	Shasta	Shasta +GFAse	Hifiasm 0.14.1_Y1	Hifiasm 0.19.5_Y2	T2T_v1.01	Hifiasm +Raft	Hifiasm +Herro+Raft
ONT	●	●		●	●	●	●
HiC		●			●		
HiFi			●	●	●		
Trio Data			●	●	●		

Benchmarking Shasta and GFAse assemblies with Flagger

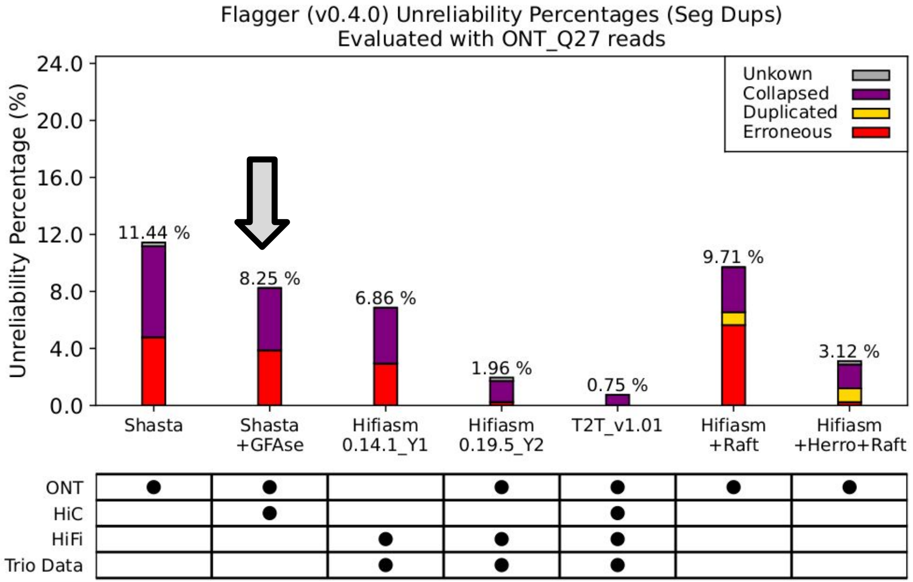
Results For Segmental Duplications

- Similar to whole genome results, in segmental duplications (projected from CHM13-v2.0 annotation) Shasta+GFAse has comparable structural accuracy with HPRC_Y1.



Flagger (v0.4.0) Unreliability Percentages (Seg Dups)
Evaluated with HiFi_DeepConsensus_v1.2 reads

	Shasta	Shasta +GFAse	Hifiasm 0.14.1_Y1	Hifiasm 0.19.5_Y2	T2T_v1.01	Hifiasm +Raft	Hifiasm +Herro+Raft
ONT	●	●		●	●	●	●
HiC		●			●		
HiFi			●	●	●		
Trio Data			●	●	●		



Flagger (v0.4.0) Unreliability Percentages (Seg Dups)
Evaluated with ONT_Q27 reads

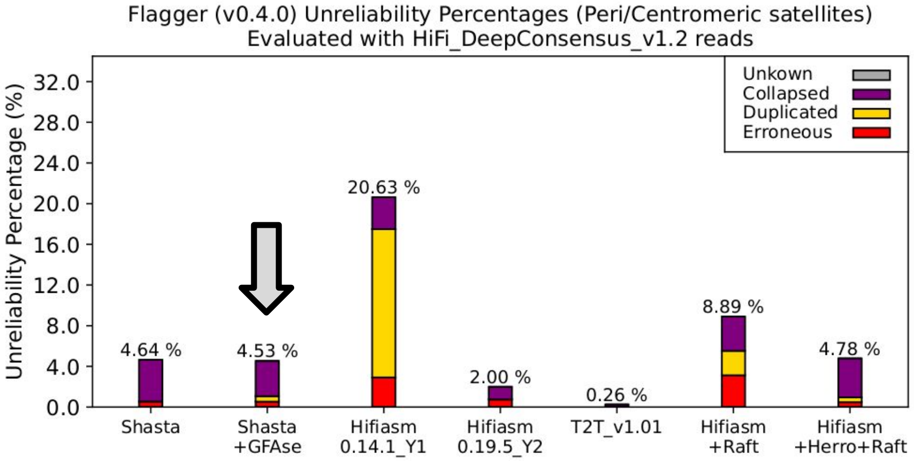
	Shasta	Shasta +GFAse	Hifiasm 0.14.1_Y1	Hifiasm 0.19.5_Y2	T2T_v1.01	Hifiasm +Raft	Hifiasm +Herro+Raft
ONT	●	●		●	●	●	●
HiC		●			●		
HiFi			●	●	●		
Trio Data			●	●	●		

Benchmarking Shasta and GFAse assemblies with Flagger

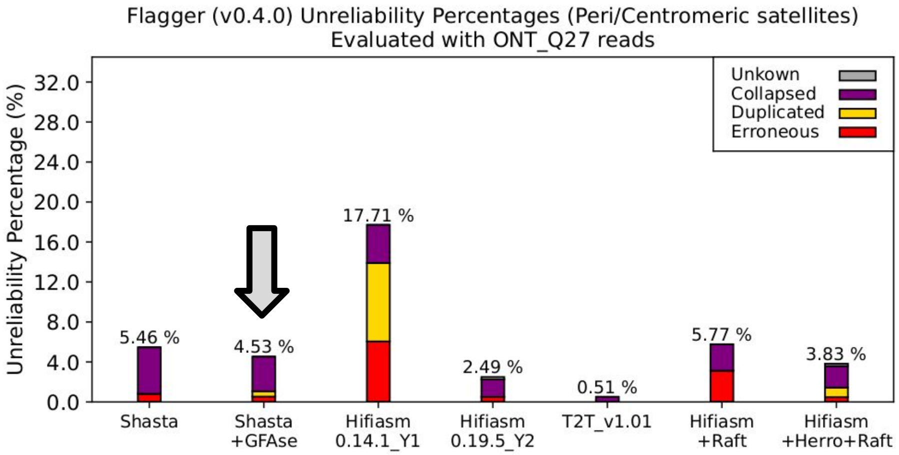
Results For Peri/Centromeric Satellites

- In peri/centromeric satellites (projected from CHM13-v2.0 annotation) Shasta+GFAse is performing better than HPRC_Y1. Long stretches of false duplications were detected in HPRC_Y1.

- This issue in Hifiasm was resolved in later versions of Hifiasm (HPRC_Y2) so that the recent Hifiasm assembly slightly outperforms Shasta+GFAse in satellites.



	Shasta	Shasta +GFAse	Hifiasm 0.14.1_Y1	Hifiasm 0.19.5_Y2	T2T_v1.01	Hifiasm +Raft	Hifiasm +Herro+Raft
ONT	●	●		●	●	●	●
HiC		●			●		
HiFi			●	●	●		
Trio Data			●	●	●		



	Shasta	Shasta +GFAse	Hifiasm 0.14.1_Y1	Hifiasm 0.19.5_Y2	T2T_v1.01	Hifiasm +Raft	Hifiasm +Herro+Raft
ONT	●	●		●	●	●	●
HiC		●			●		
HiFi			●	●	●		
Trio Data			●	●	●		

Future plans

- The initial Shasta release of Mode 3 assembly only includes an assembly configuration for the *ncm23* ONT reads. It may be possible to provide an assembly configuration for ONT R10 reads in a follow up release.
- Fix/improve on current known issues/limitations:
    - Strand separation sometimes leads to haplotype breaks (dangling segments).
    - Inconsistent alignments in satellite-rich regions.
    - Improved detangling could result in increased contiguity.
    - Fix a few gross inefficiencies, which will reduce memory requirements and execution times.

Acknowledgements

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

https://cglgenomics.ucsc.edu/

Computational Genomics Lab,
left-to-right:
Roni Altshuler
Kim Czupil
Uyen Nguyen
Ash O'Farrell
Mira Mastoras
Me
Prajna Hebbar
Nafiseh Jafarzadeh
Nick Keener
Konstantinos Kyriakidis
Jimin Park
Cecilia Cisar
Lon Blauvelt
Mobin Asri
Shloka Negi
Ivo Violich

Adam Novak

Jordan Eizenga

David Haussler

Glenn Hickey

Melissa Meredith

Parsa Eskander

Mark Diekhans

Xian Chang

Jouni Siren

Brandy Baird
Joshua Gardner
Sara O'Rourke

NIH NHGRI

NORTHEASTERN UNIVERSITY

Paolo Carnevali

Miten Jain

BROAD INSTITUTE

Ryan Lorig-Roach

NIH NATIONAL CANCER INSTITUTE

Mikhail Kolmogorov

Inserm
La science pour la santé
From science to health

Jean Monlong

Karen Miga

Julian Lucas