COMPLEX
TRAIT
COMMUNITY

Multi-platform genome assembly of an SHR/OlaIpcv X BN-Lx/Cub F1 rat “trio”

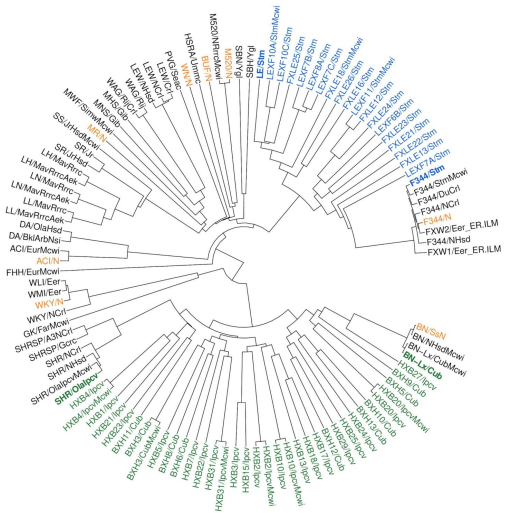
Andrea Guarracino
Postdoctoral Scholar

@AndresGuarahino

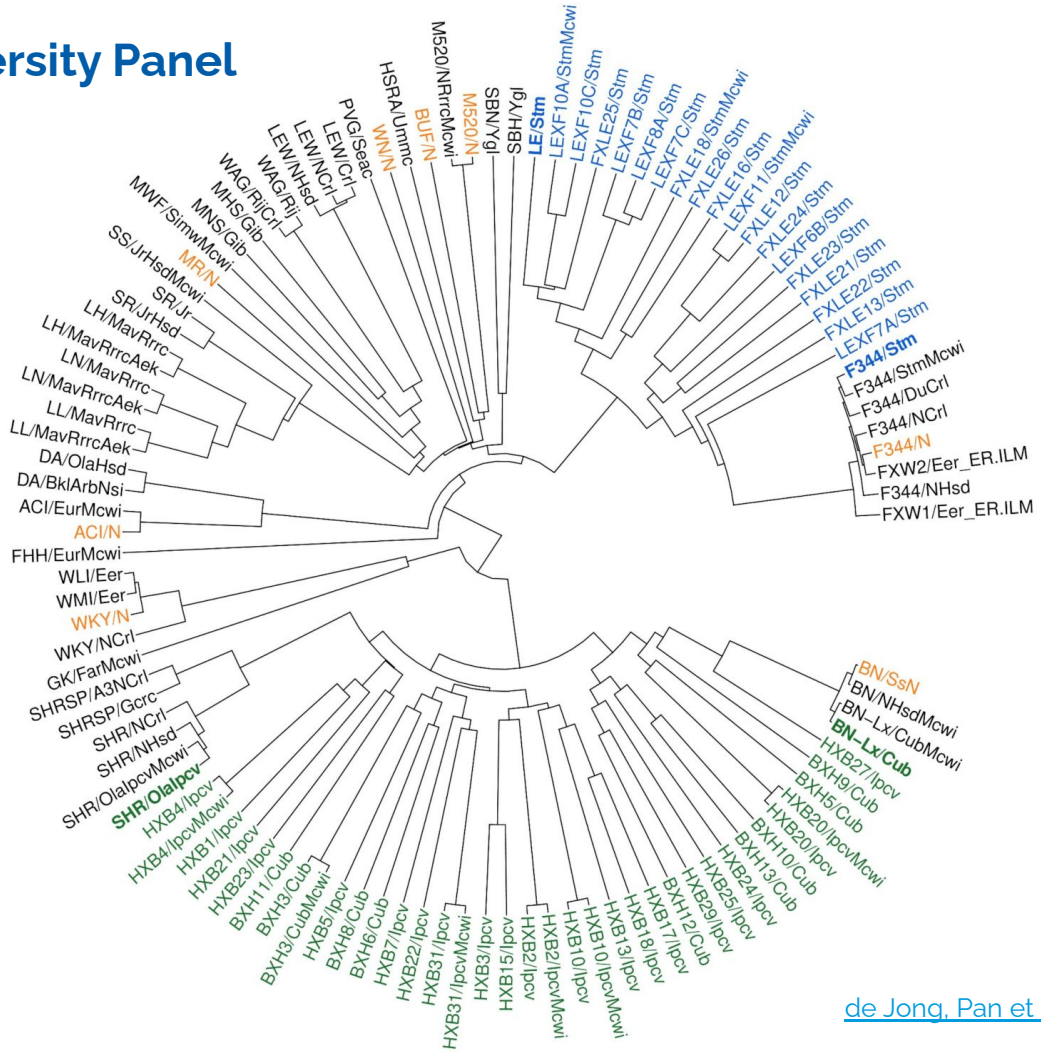
CTC-RG2024
Medical College of Wisconsin, Milwaukee, USA
2024/10/04

Outline

- Introduction
- Assembly
- Quality control
- Adaptive sequencing
- ONT reads correction



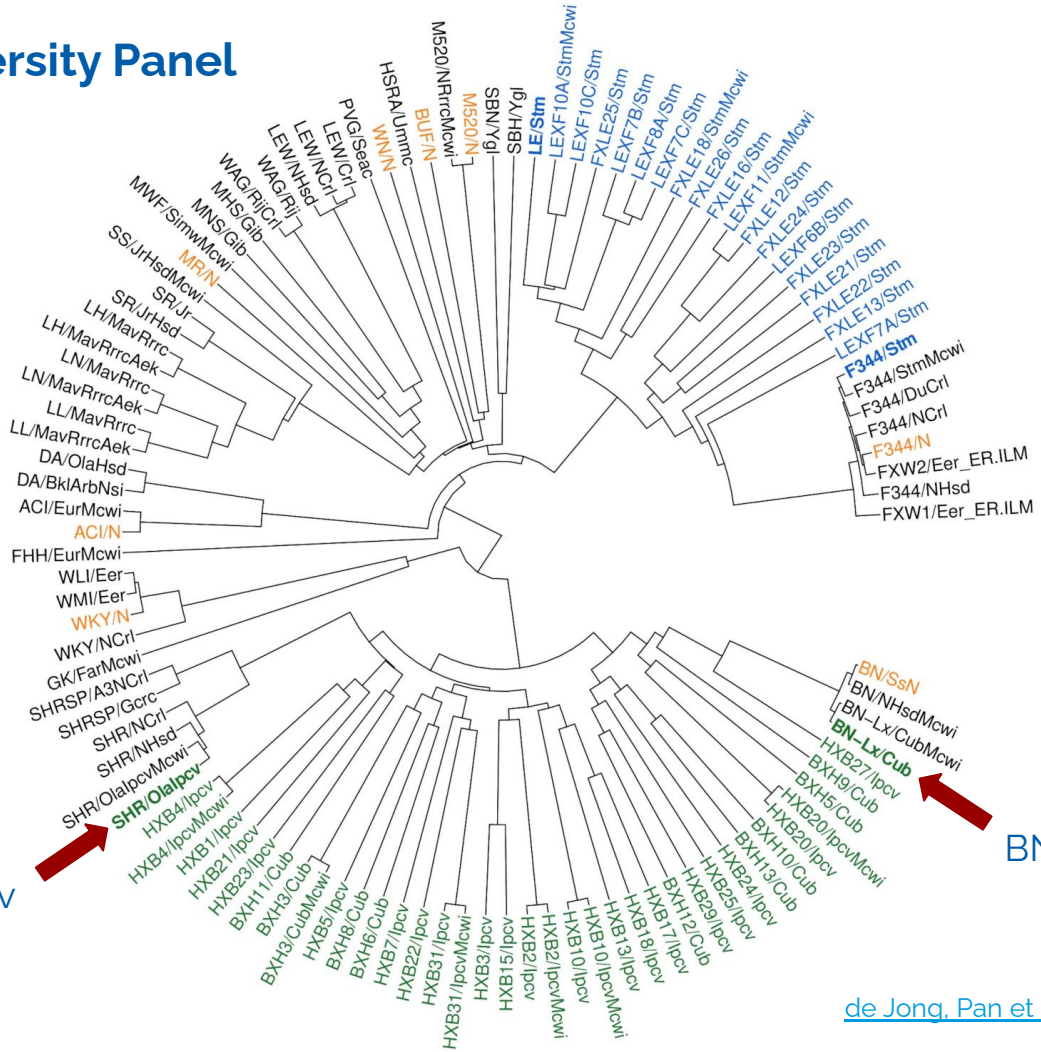
Hybrid Rat Diversity Panel



THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER.

de Jong, Pan et al. 2024, Cell Genomics

Hybrid Rat Diversity Panel



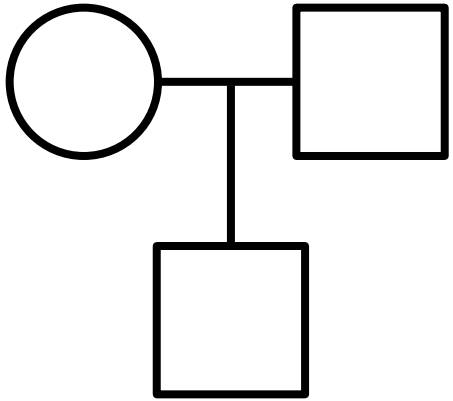
SHR/OlaIpcv

BN-Lx/Cub

de Jong, Pan et al. 2024, Cell Genomics

Trio

SHR/OlaIpcv

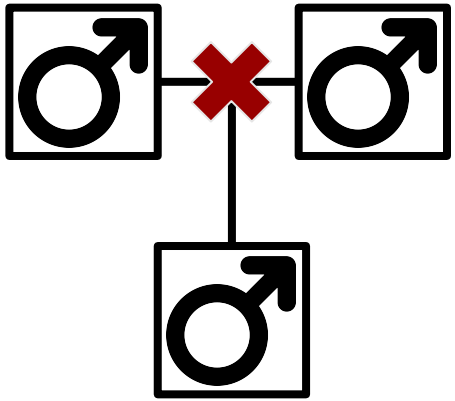


BN-Lx/Cub

SHR/OlaIpcv X BN-Lx/Cub F1

“Trio”

SHR/OlaIpcv



BN-Lx/Cub

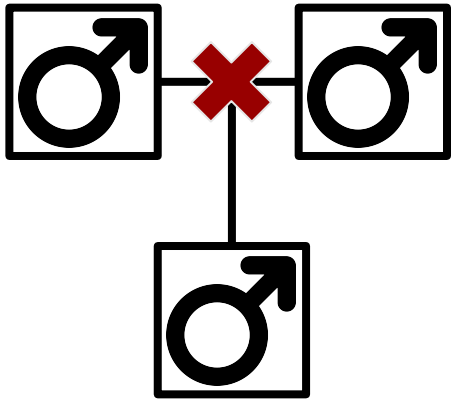
SHR/OlaIpcv X BN-Lx/Cub F1

“Trio”

SHR/OlaIpcv

PacBio HiFi reads: 44X coverage

PacBio



BN-Lx/Cub

PacBio HiFi reads: 41X coverage

PacBio

SHR/OlaIpcv X BN-Lx/Cub F1

ONT reads: 81X coverage (of which 36X Ultra-Long)

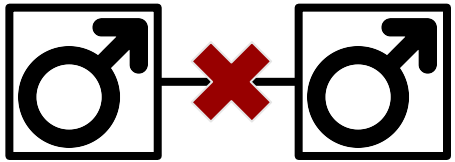
Oxford NANOPORE Technologies

Outline

- Introduction
- Assembly
- Quality control
- Adaptive sequencing
- ONT reads correction

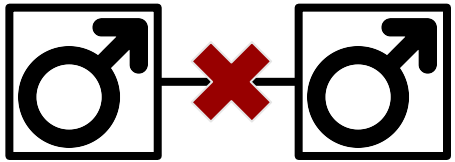
Assembly

SHR/Olalpcv



BN-Lx/Cub

Assembly



SHR/OlaIpcv

PacBio

hifiasm

Haploid assembly

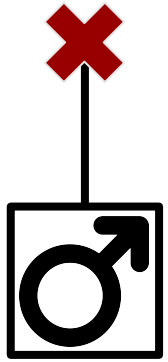
BN-Lx/Cub

PacBio

hifiasm

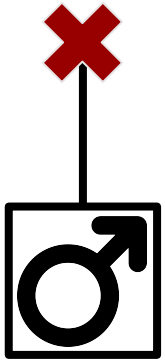
Haploid assembly

Assembly



SHR/OlaIpcv X BN-Lx/Cub F1

Assembly



SHR/OlaIpcv X BN-Lx/Cub F1

SHR/OlaIpcv PacBio

BN-Lx/Cub PacBio

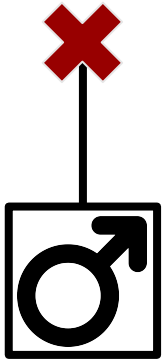
50kbps Oxford NANOPORE Technologies

hifiasm

Diploid assembly

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Assembly



SHR/OlaIpcv X BN-Lx/Cub F1

SHR/OlaIpcv PacBio
BN-Lx/Cub PacBio
50kbps Oxford NANOPORE Technologies

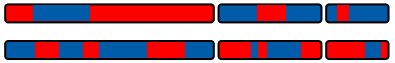
HERRO-corrected Oxford NANOPORE Technologies
50kbps Oxford NANOPORE Technologies

hifiasm

Diploid assembly

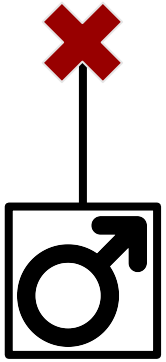
hifiasm

Diploid assembly



THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Assembly



SHR/OlaIpcv X BN-Lx/Cub F1

SHR/OlaIpcv PacBio
BN-Lx/Cub PacBio
50kbps Oxford NANOPORE Technologies

HERRO-corrected Oxford NANOPORE Technologies
50kbps Oxford NANOPORE Technologies

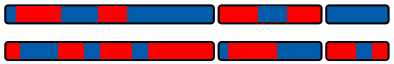
HERRO-corrected Oxford NANOPORE Technologies
50kbps Oxford NANOPORE Technologies
Trio information PacBio

hifiasm

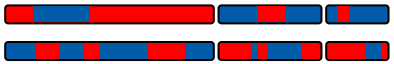
hifiasm

verkko

Diploid assembly



Diploid assembly

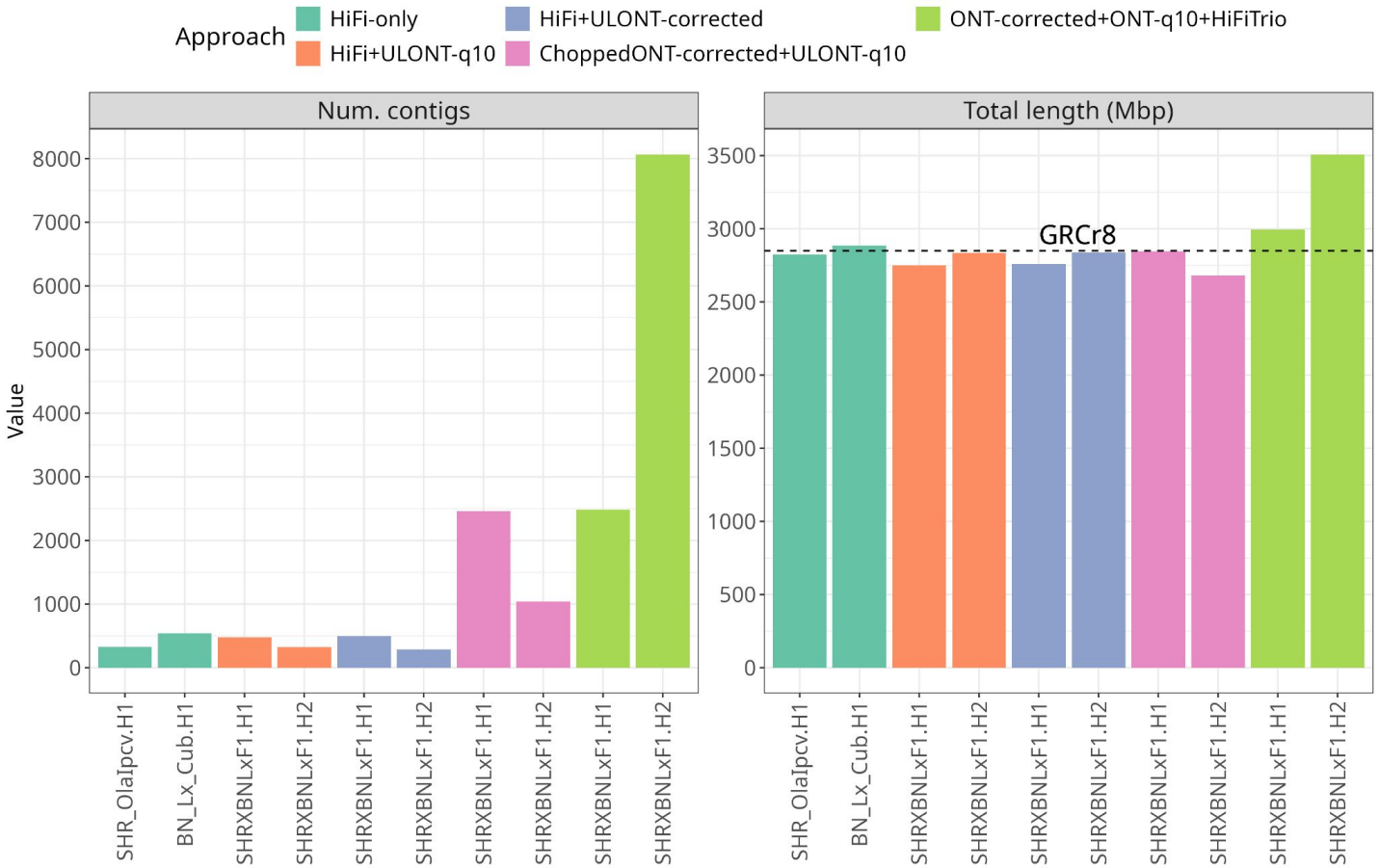


Phased diploid assembly

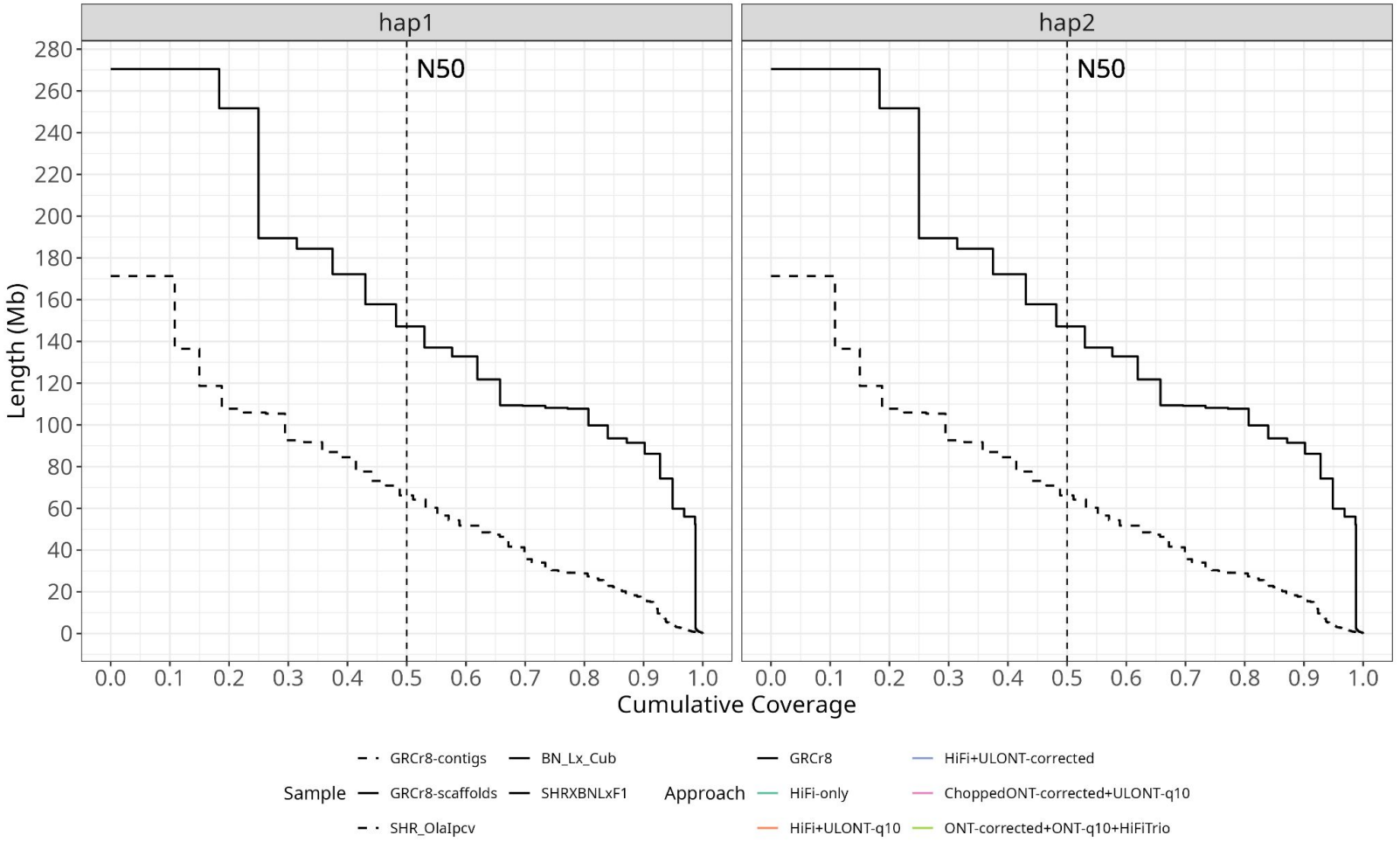
Outline

- Introduction
- Assembly
- Quality control
- Adaptive sequencing
- ONT reads correction

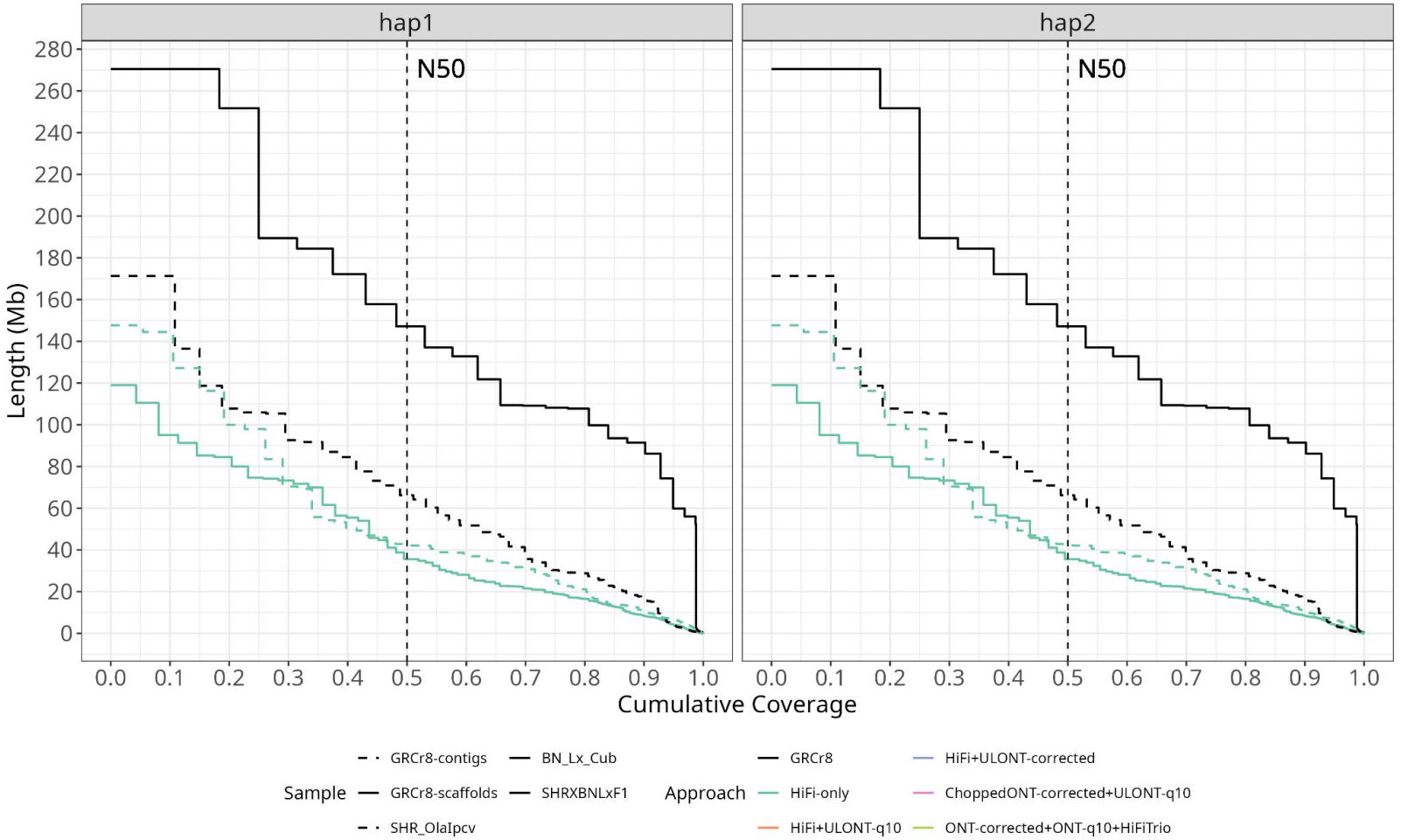
Assembly contiguity



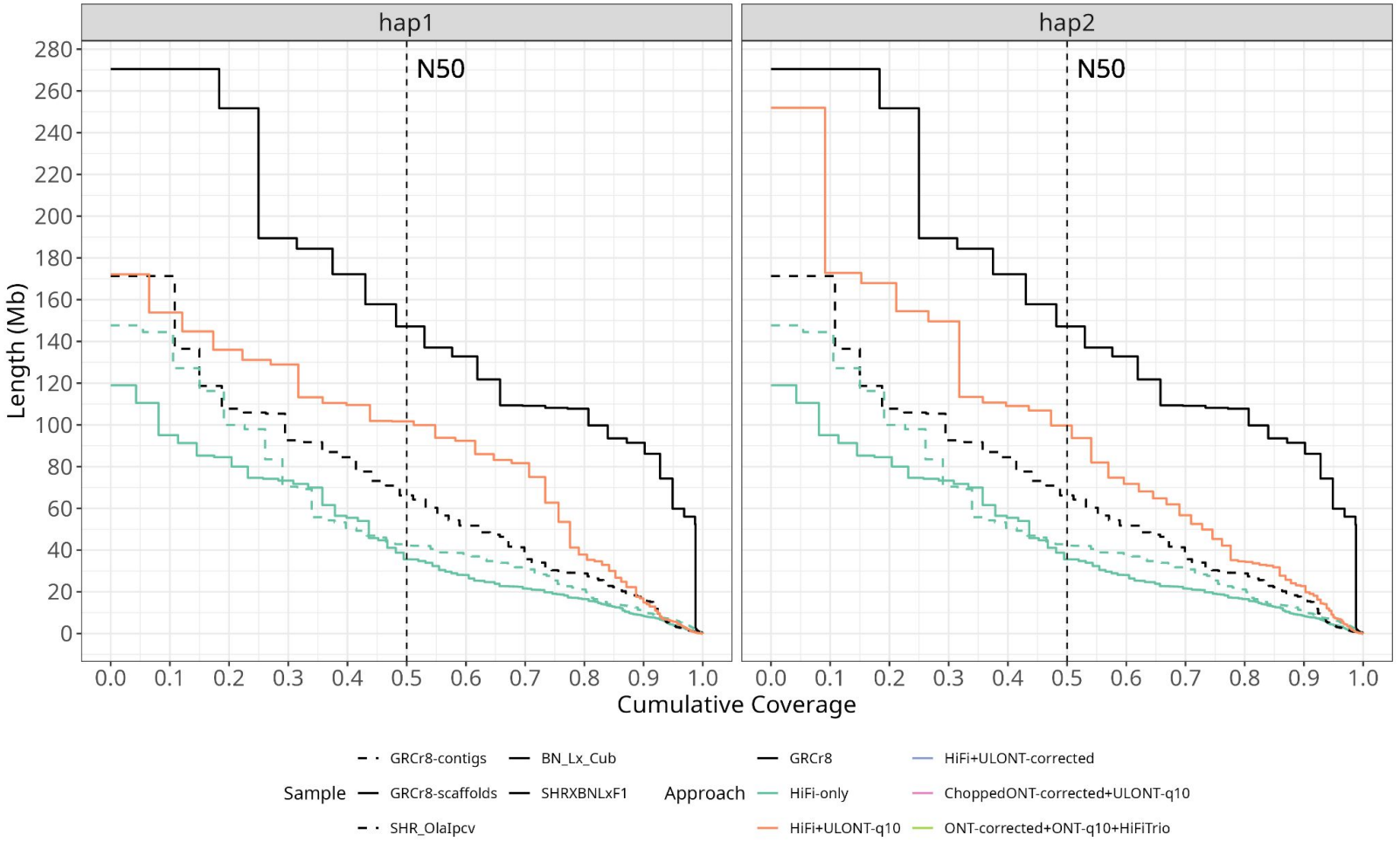
Assembly contiguity



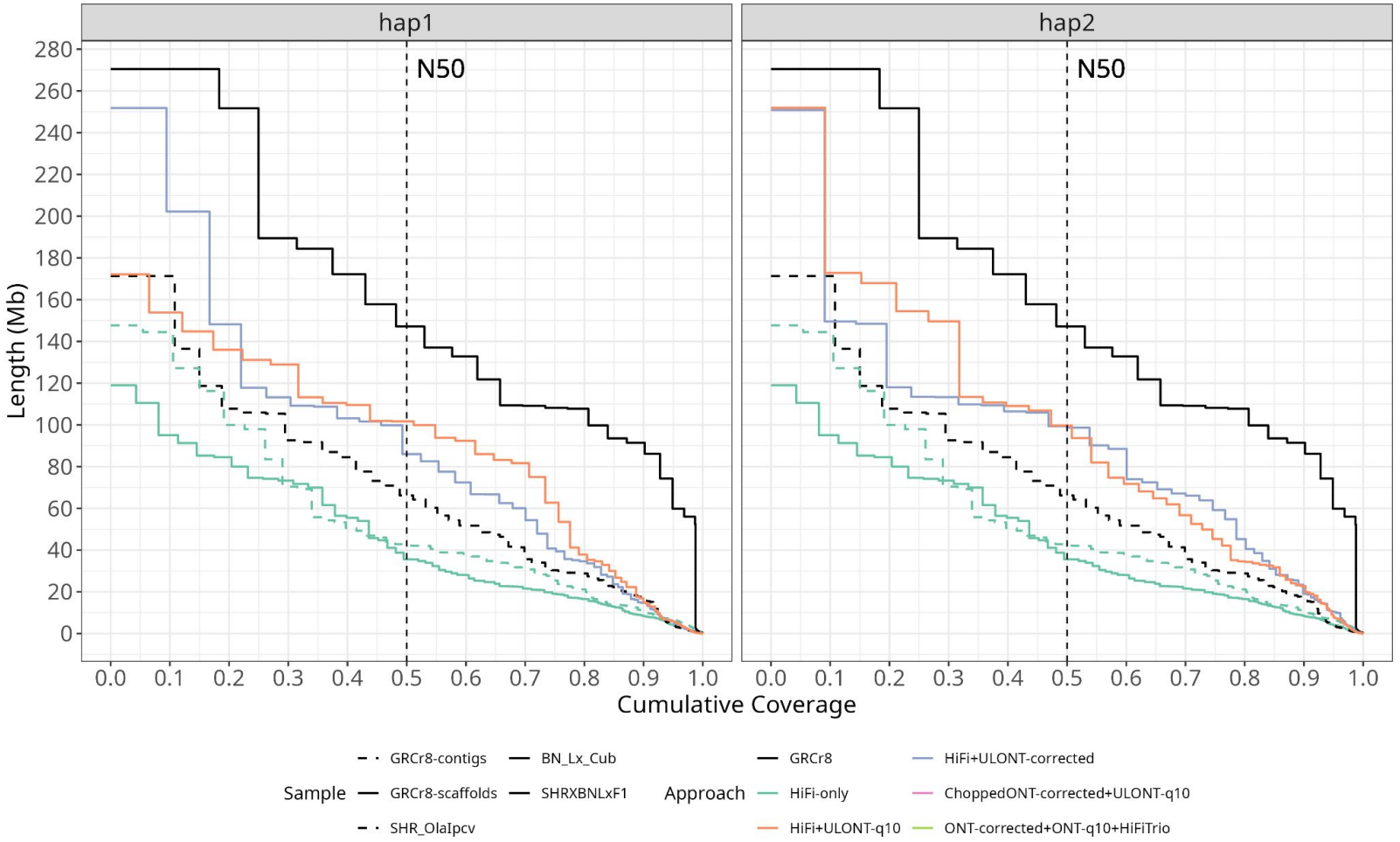
Assembly contiguity



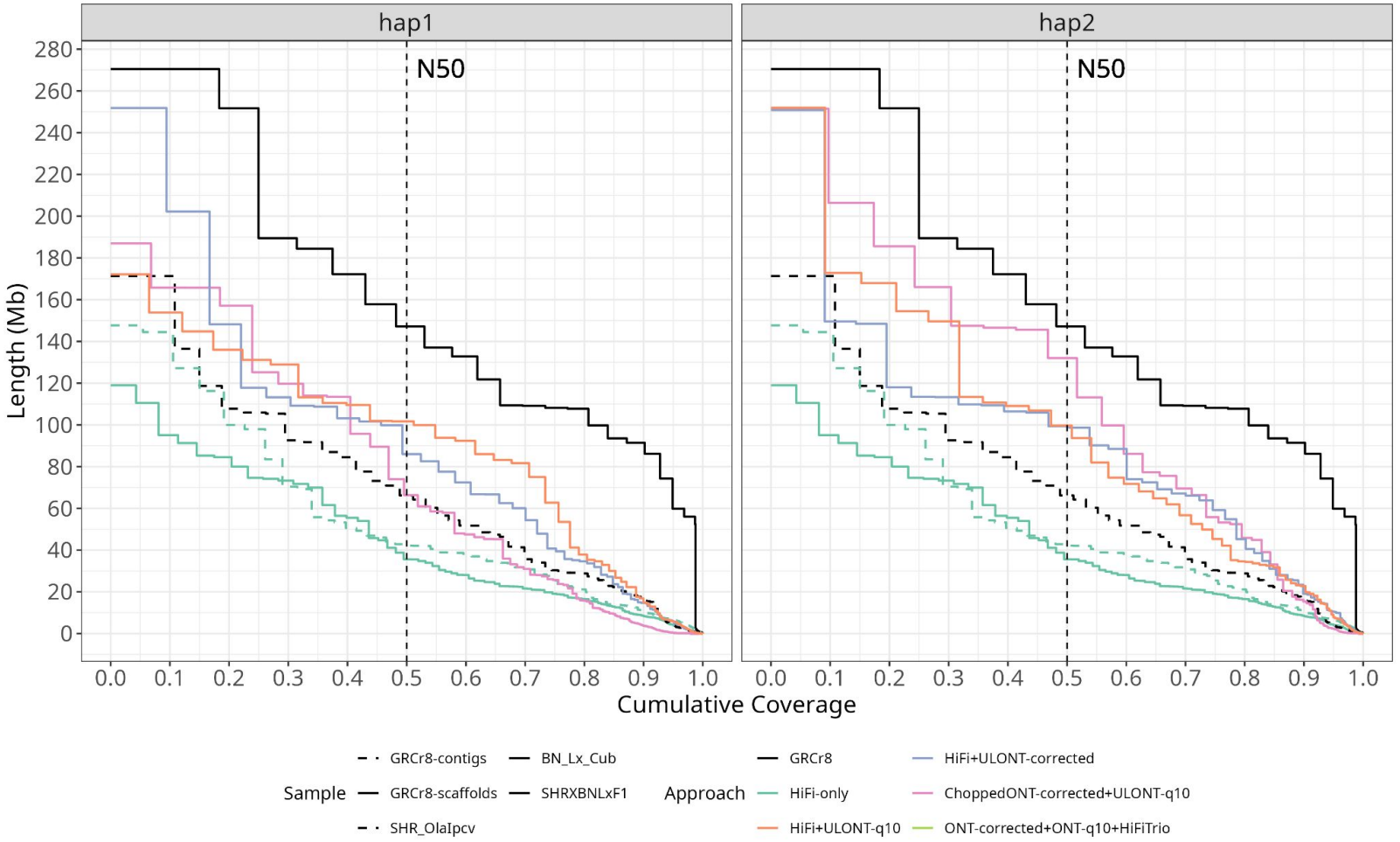
Assembly contiguity



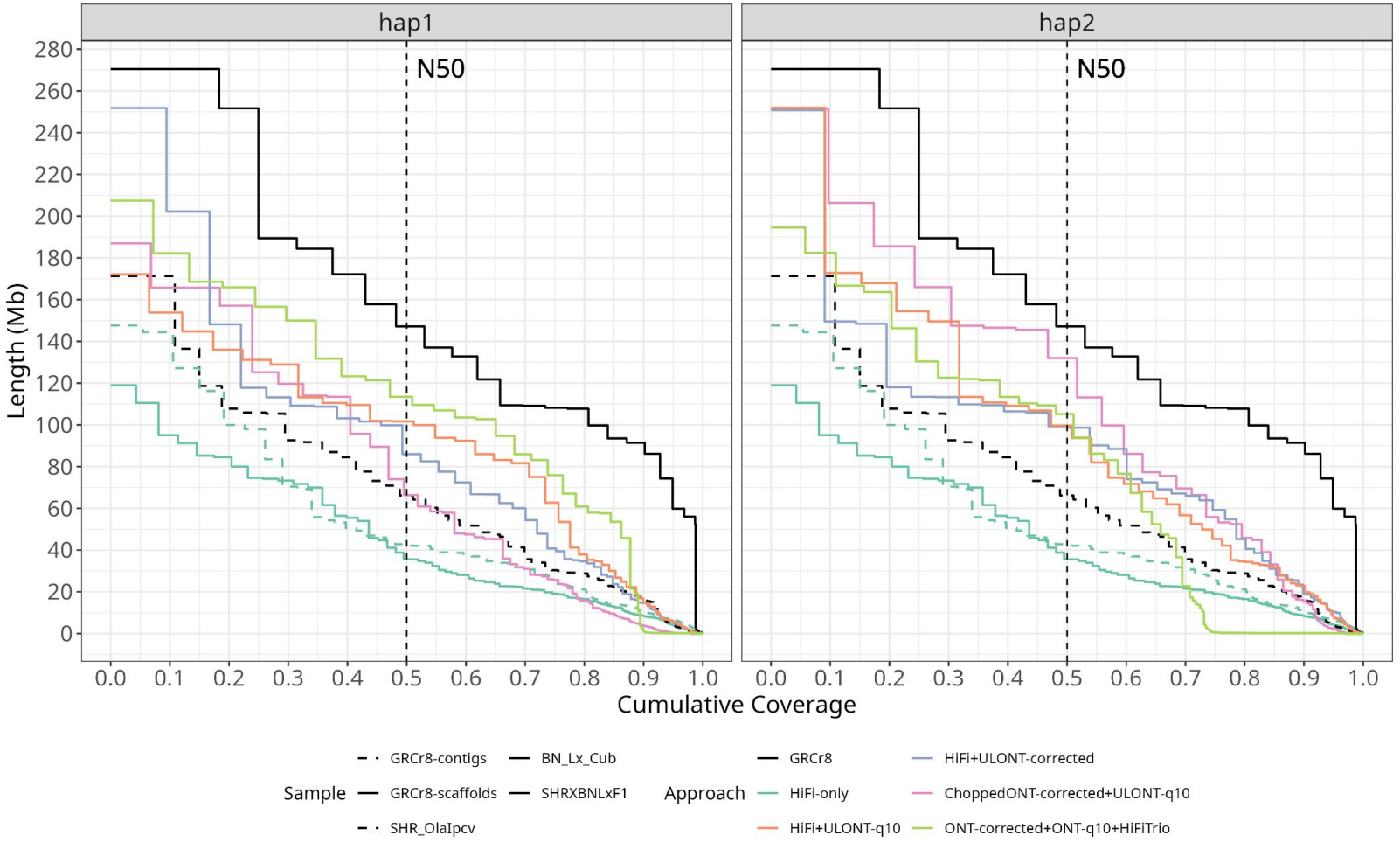
Assembly contiguity



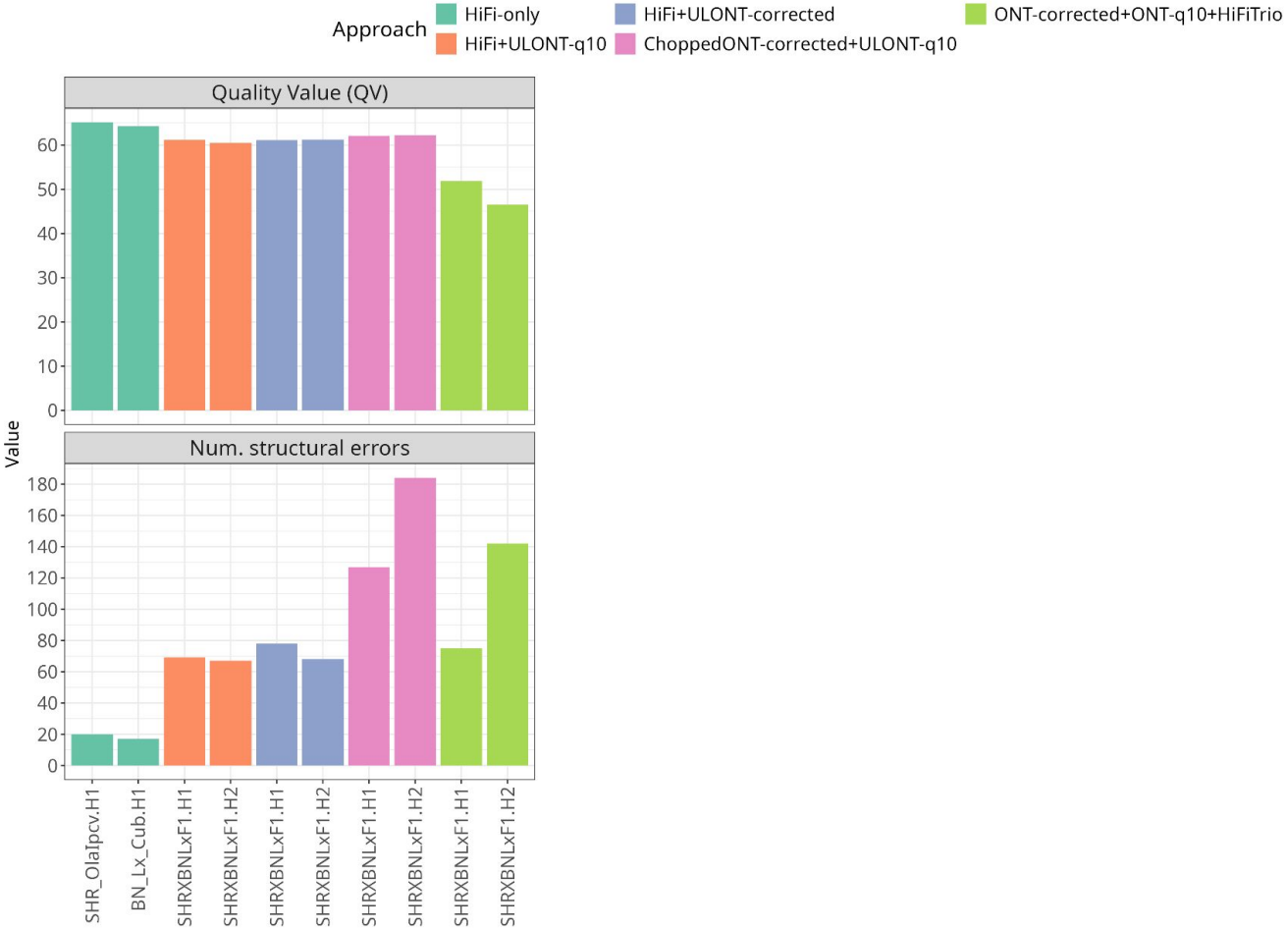
Assembly contiguity



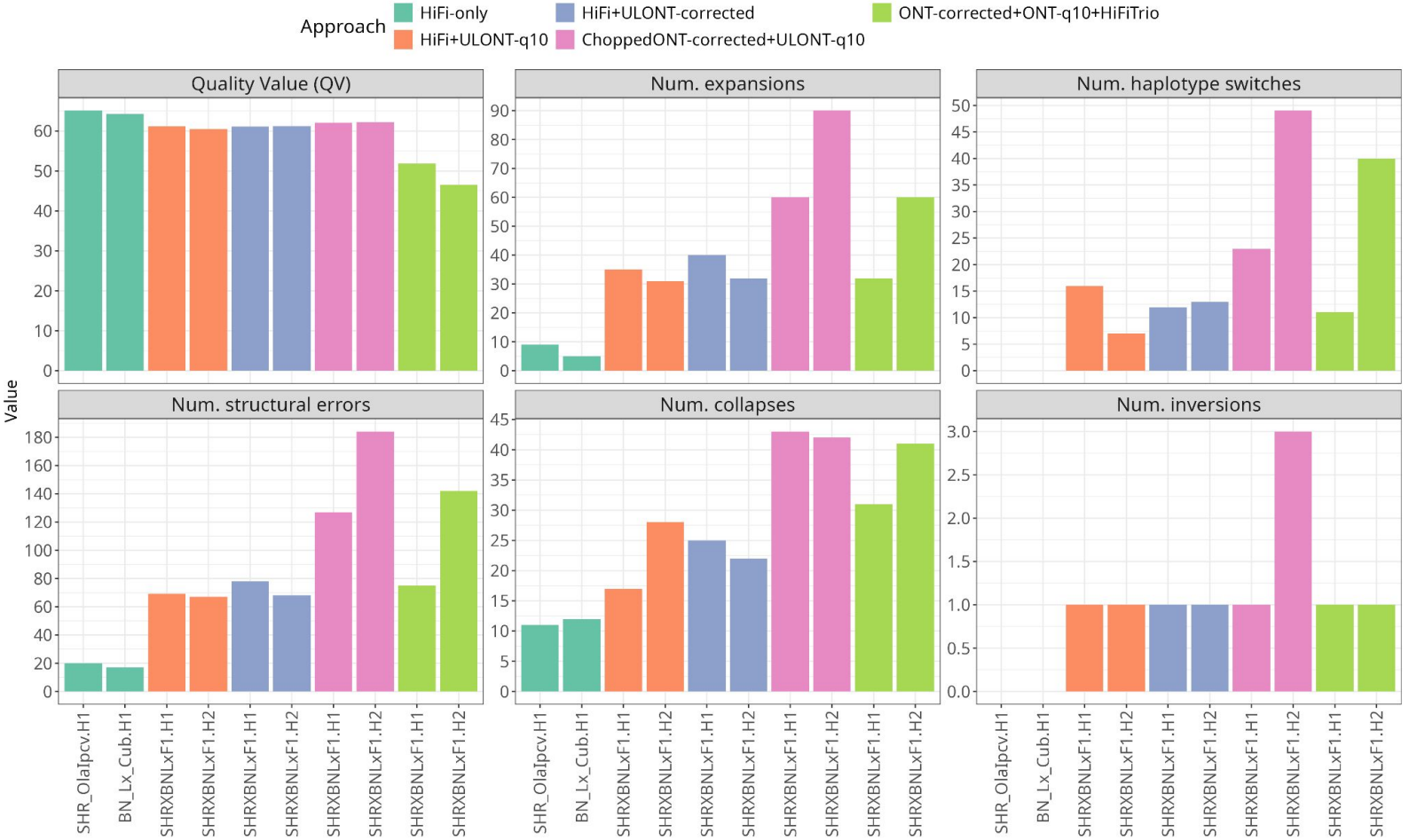
Assembly contiguity



Assembly correctness



Assembly correctness



Outline

- Introduction
- Assembly
- Quality control
- Adaptive sequencing
- ONT reads correction

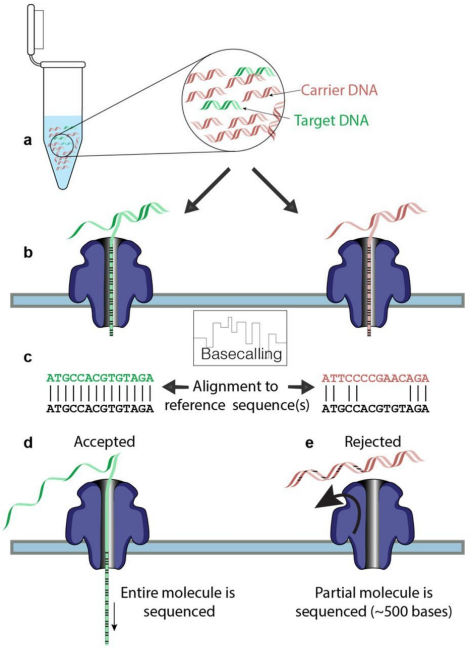
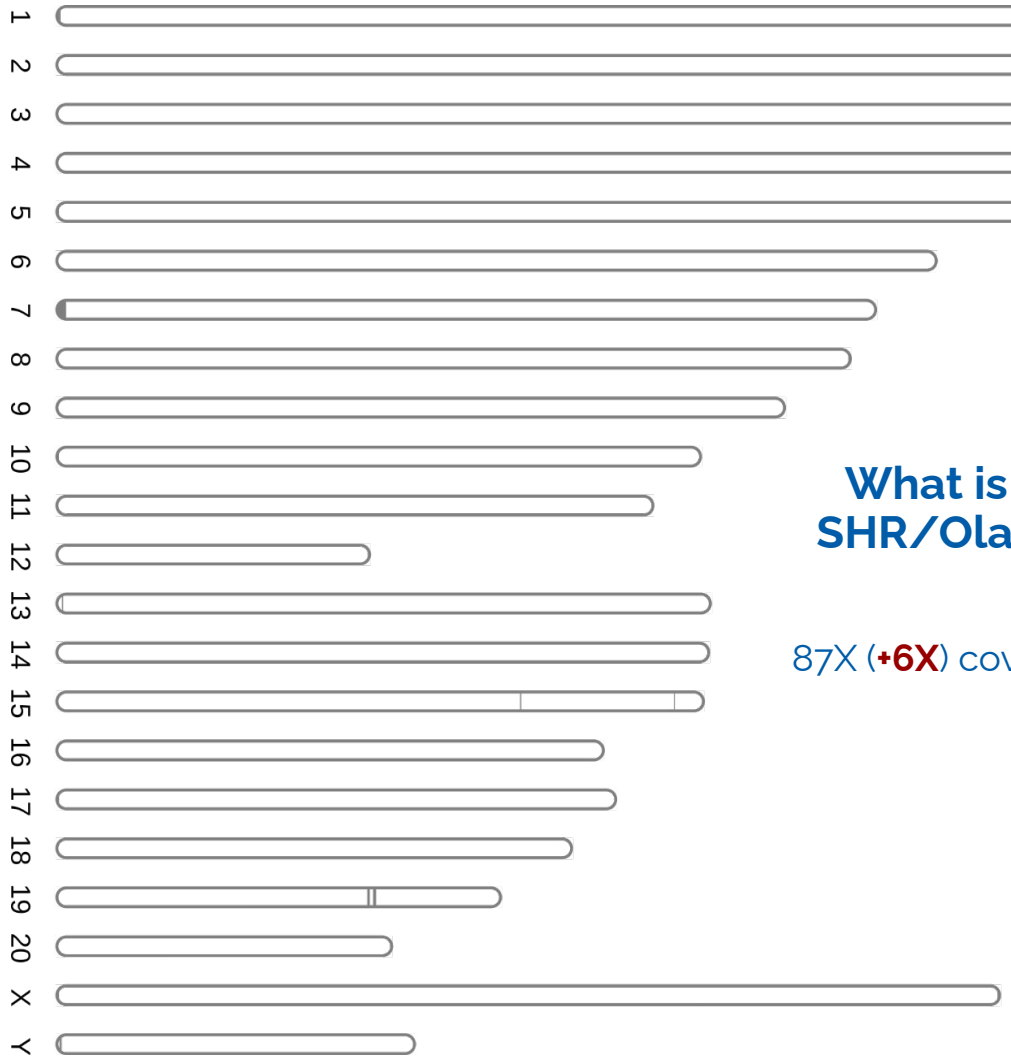


Image from Miani et al., 2023

What is missing in the Nanopore-only SHR/OlaIpcv X BN-Lx/Cub F1 assembly

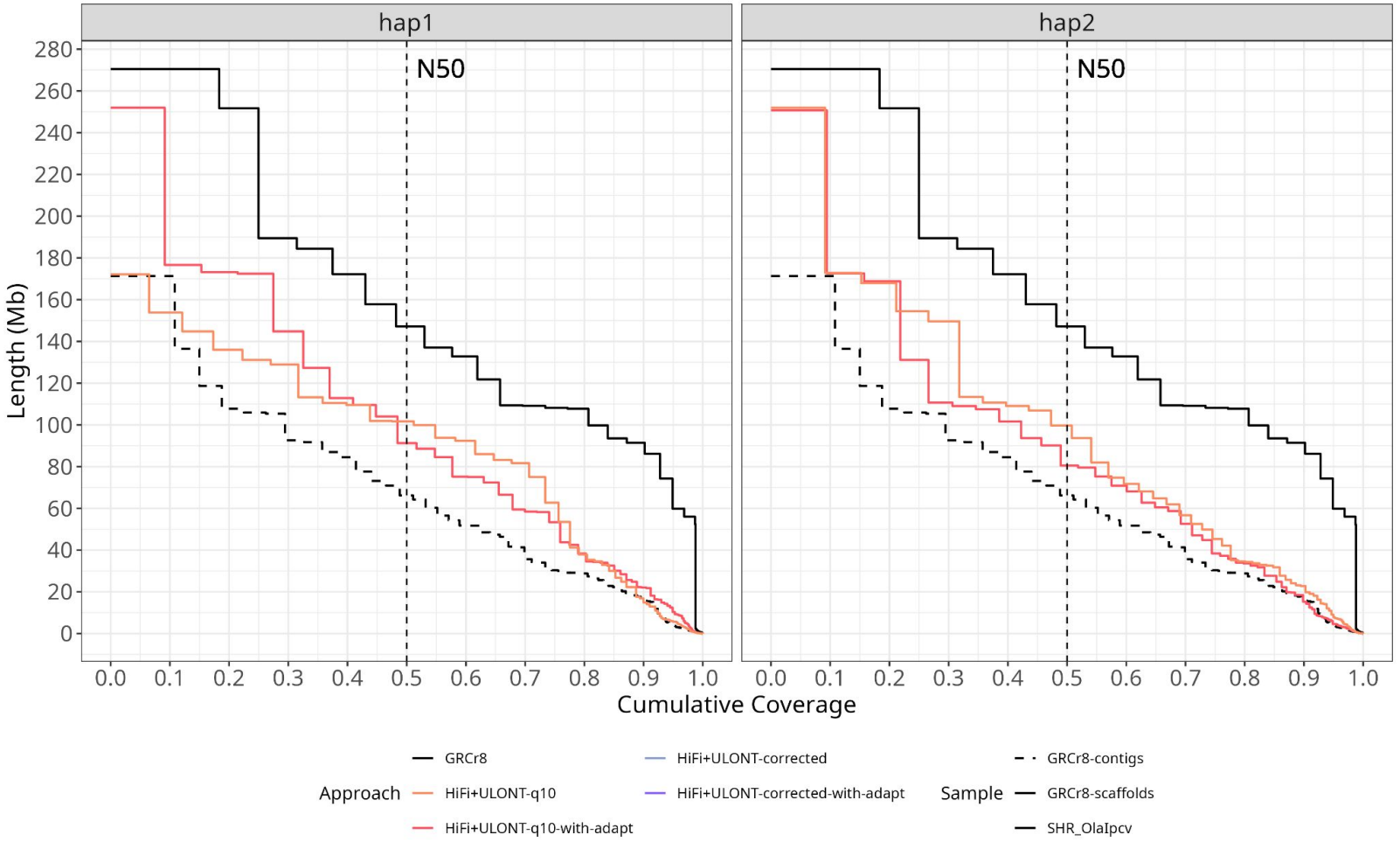


What is missing in the Nanopore-only SHR/OlaIpcv X BN-Lx/Cub F1 assembly

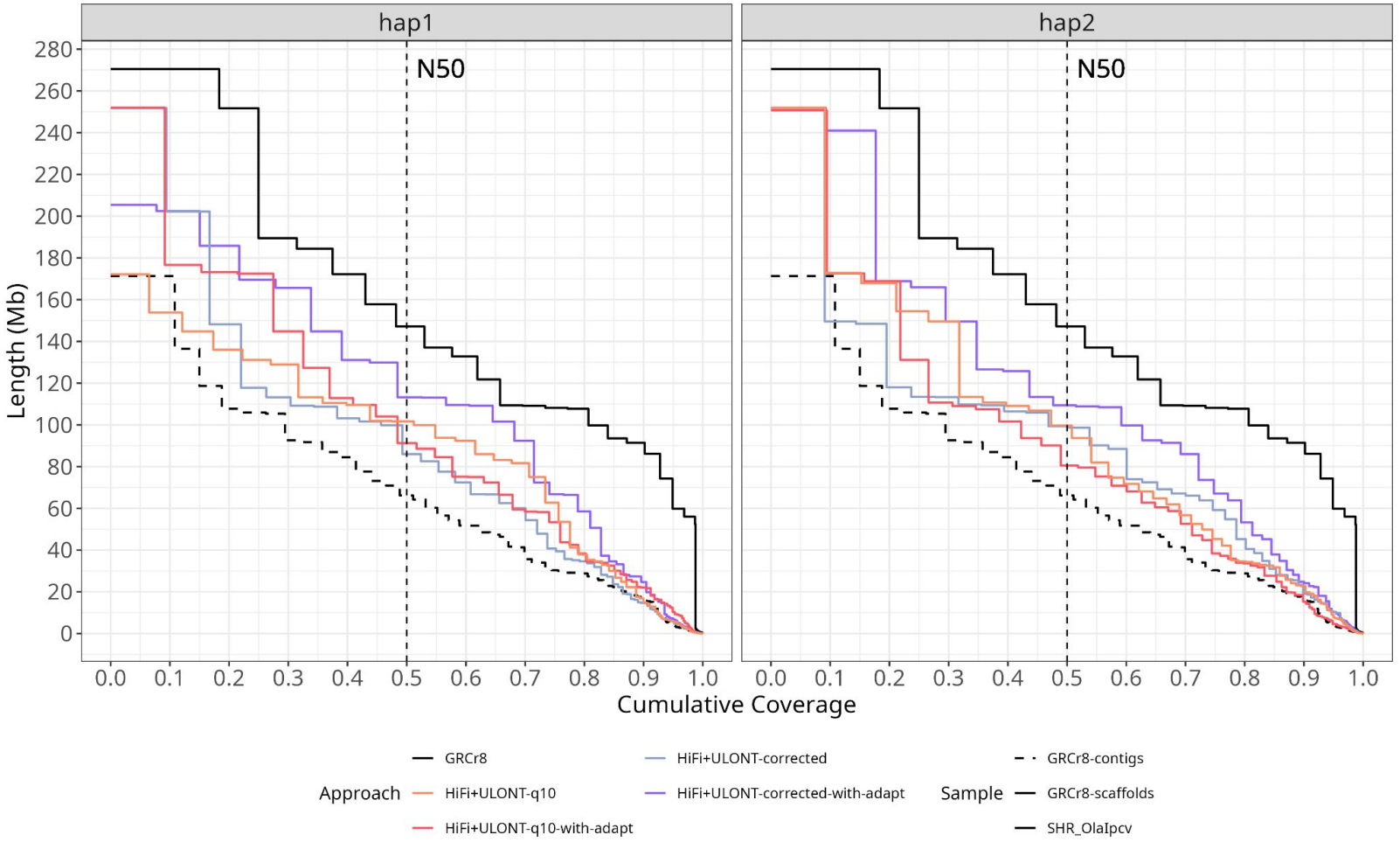
87X (**+6X**) coverage (of which 37.8X (**+1.8X**) Ultra-Long)

Oxford NANOPORE Technologies

Assembly contiguity including adaptive sequencing

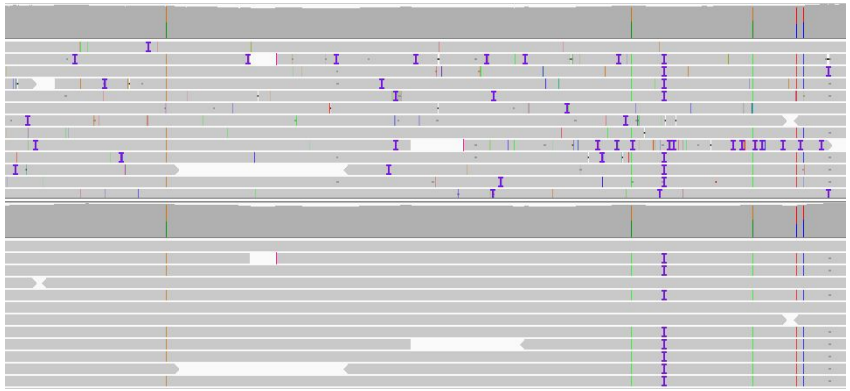


Assembly contiguity including adaptive sequencing

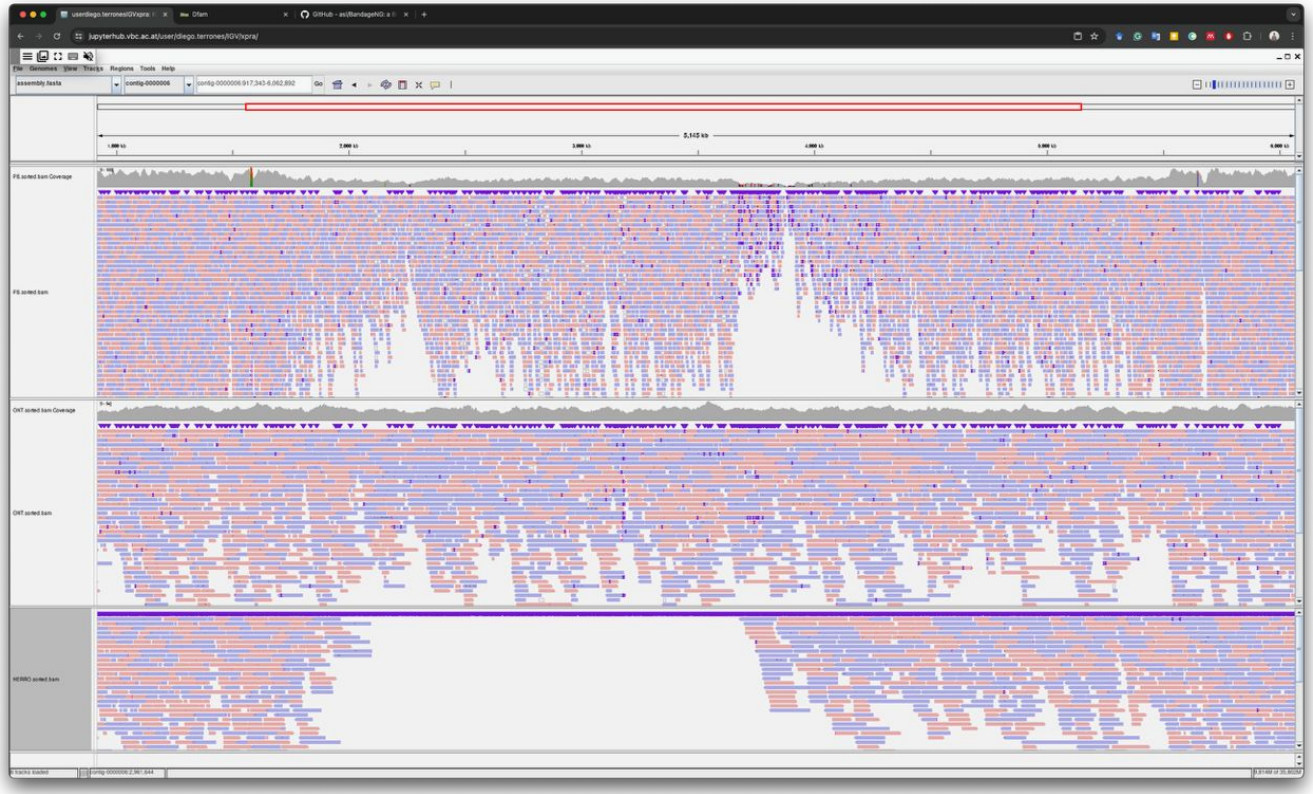


Outline

- Introduction
- Assembly
- Quality control
- Adaptive sequencing
- ONT reads correction

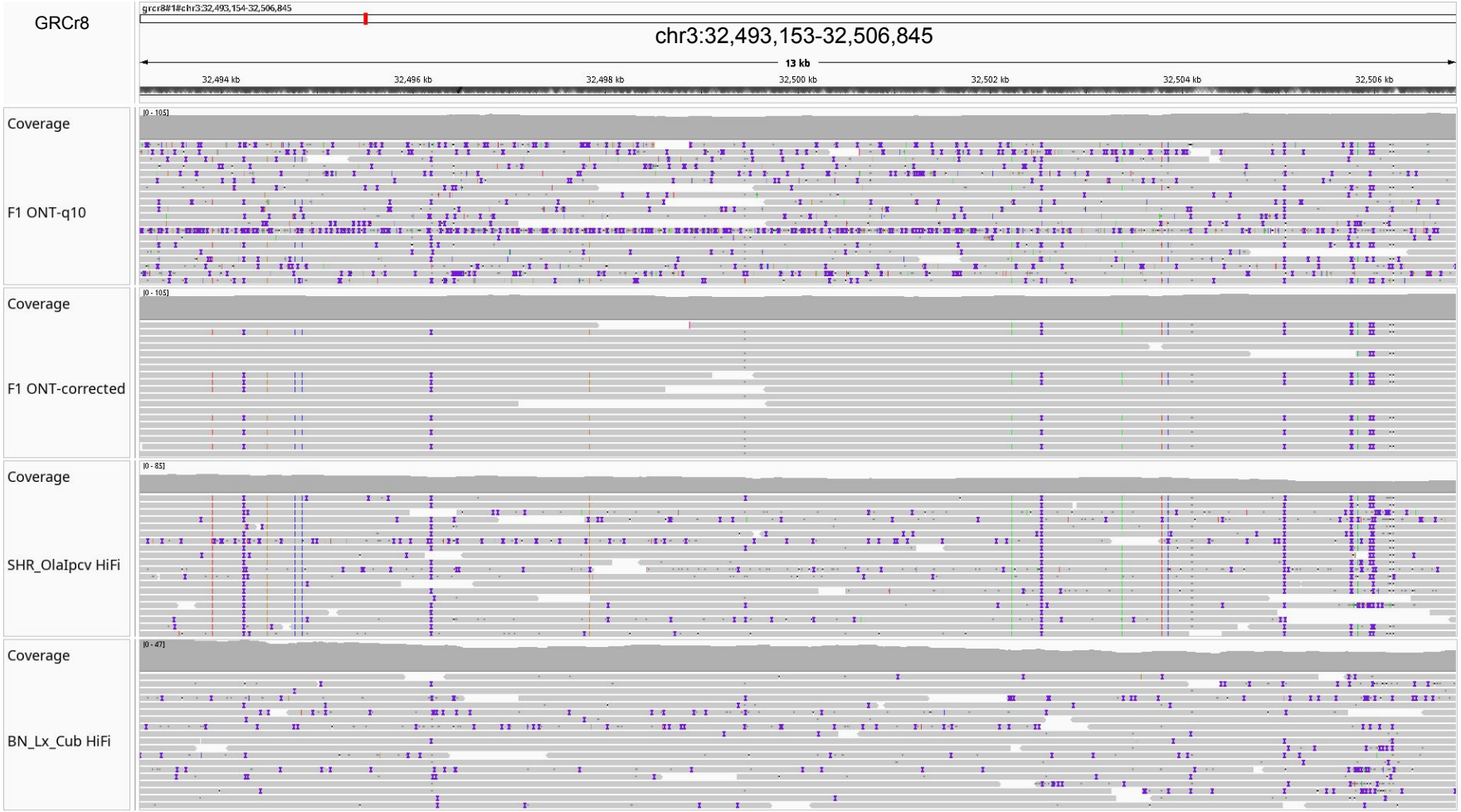


Is HERRO-correction working?

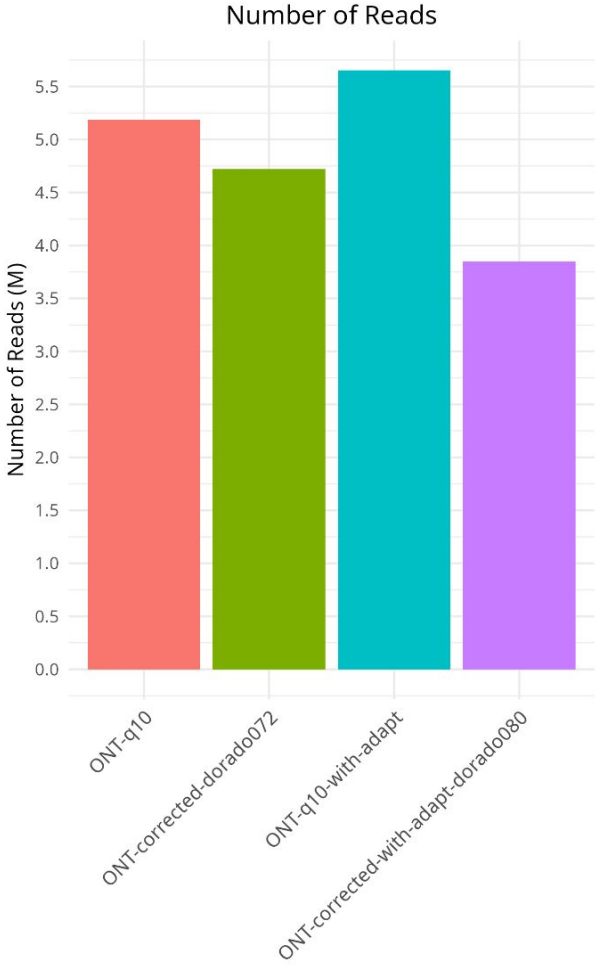


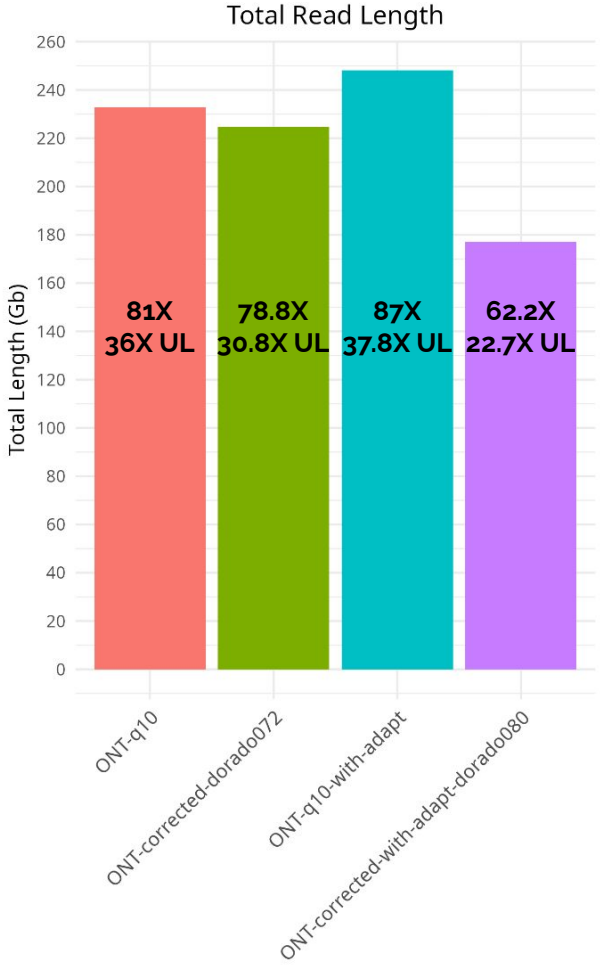
https://github.com/nanoporetech/dorado/issues/851

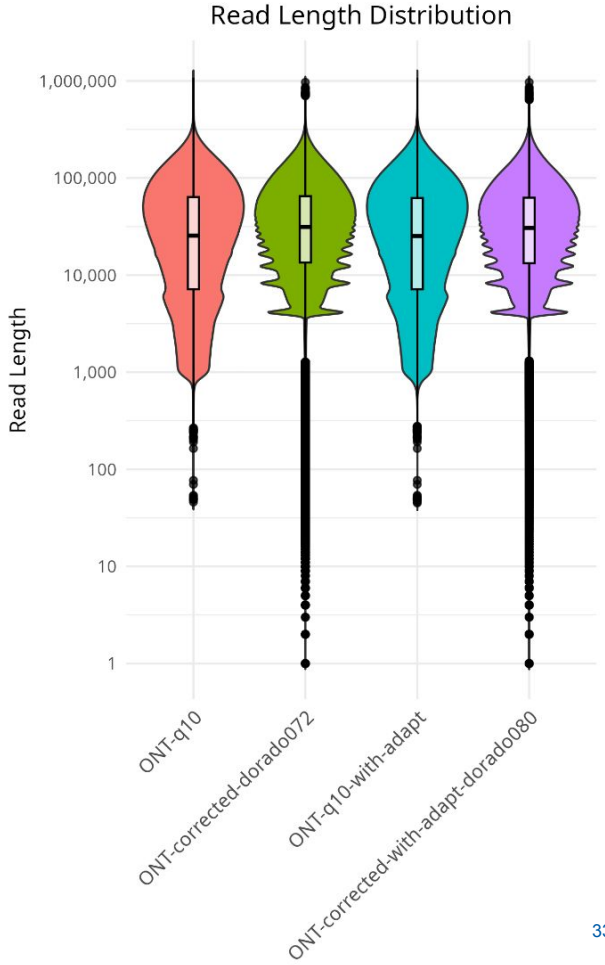
HERRO-correction is working



HERRO-correction statistics

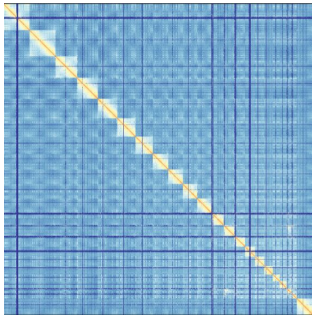




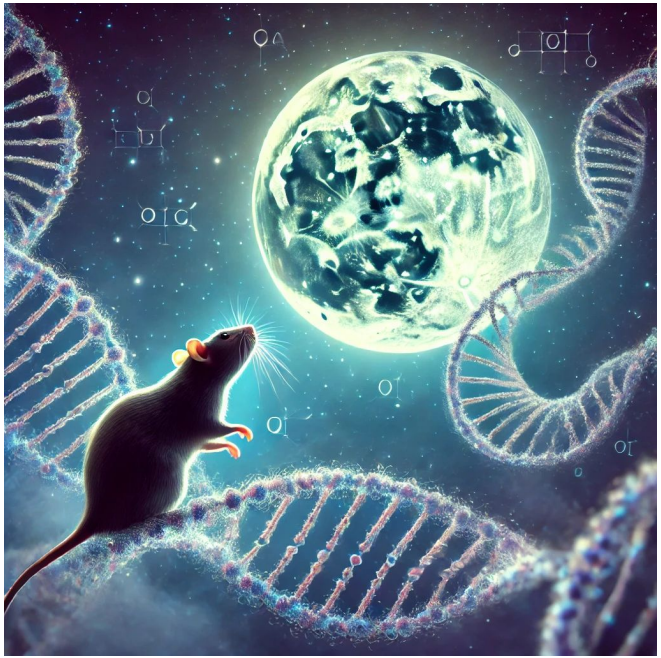


Conclusions

- ONT reads improve assembly contiguity while paying for correctness
- ONT-only assemblies have more contigs and more errors
- HERRO-correction works, but it is still not enough (and is very slow)



Hi-C data is coming!



Thanks!

Rachel Ward
- Flavia Villani

Burt Sharp
- Farnaz Salehi

Denghui Chen
- Abraham A. Palmer

Robert W. Williams
- Vincenza Colonna

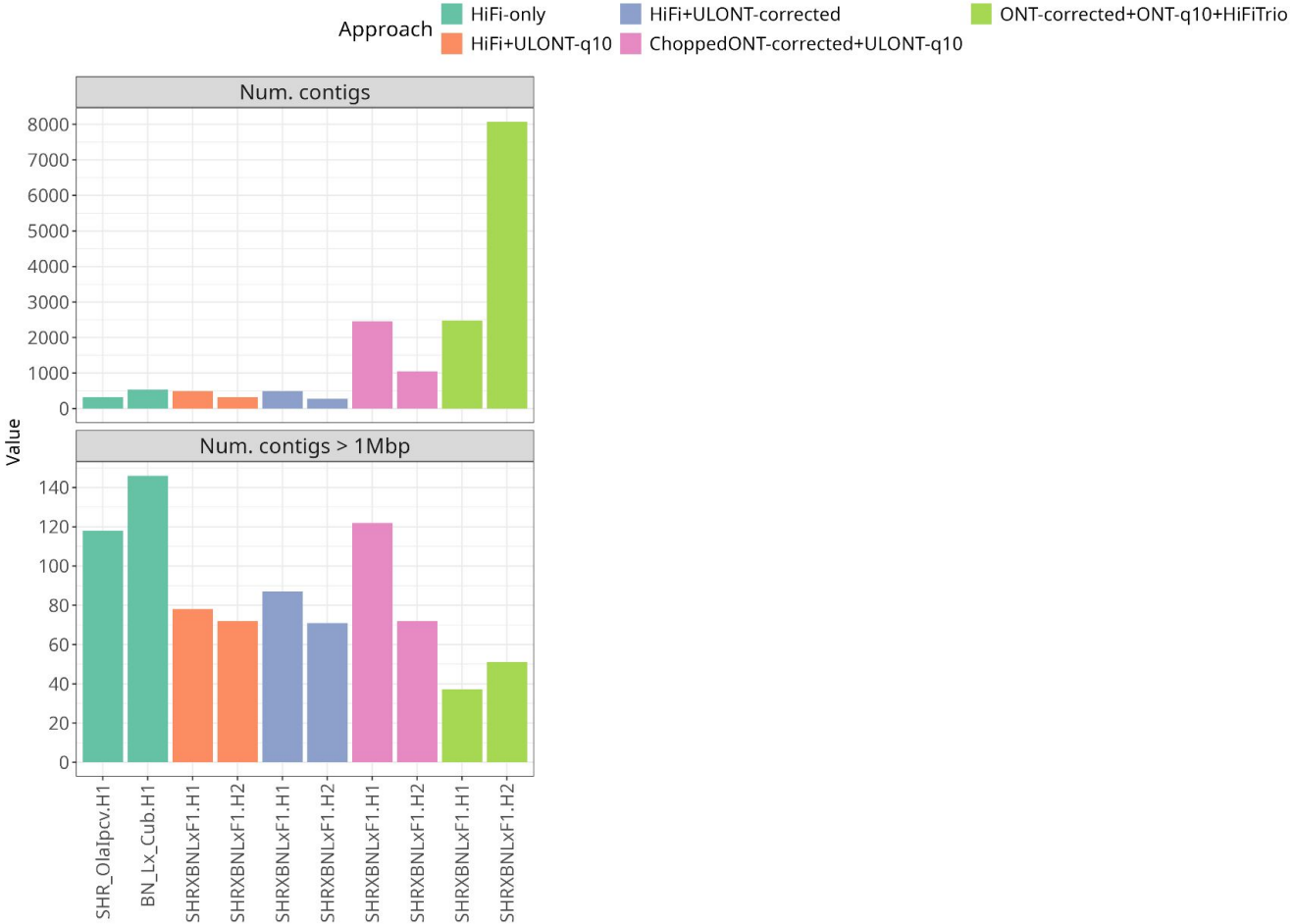
David G. Ashbrook
- Hao Chen

Pjotr Prins
- Erik Garrison

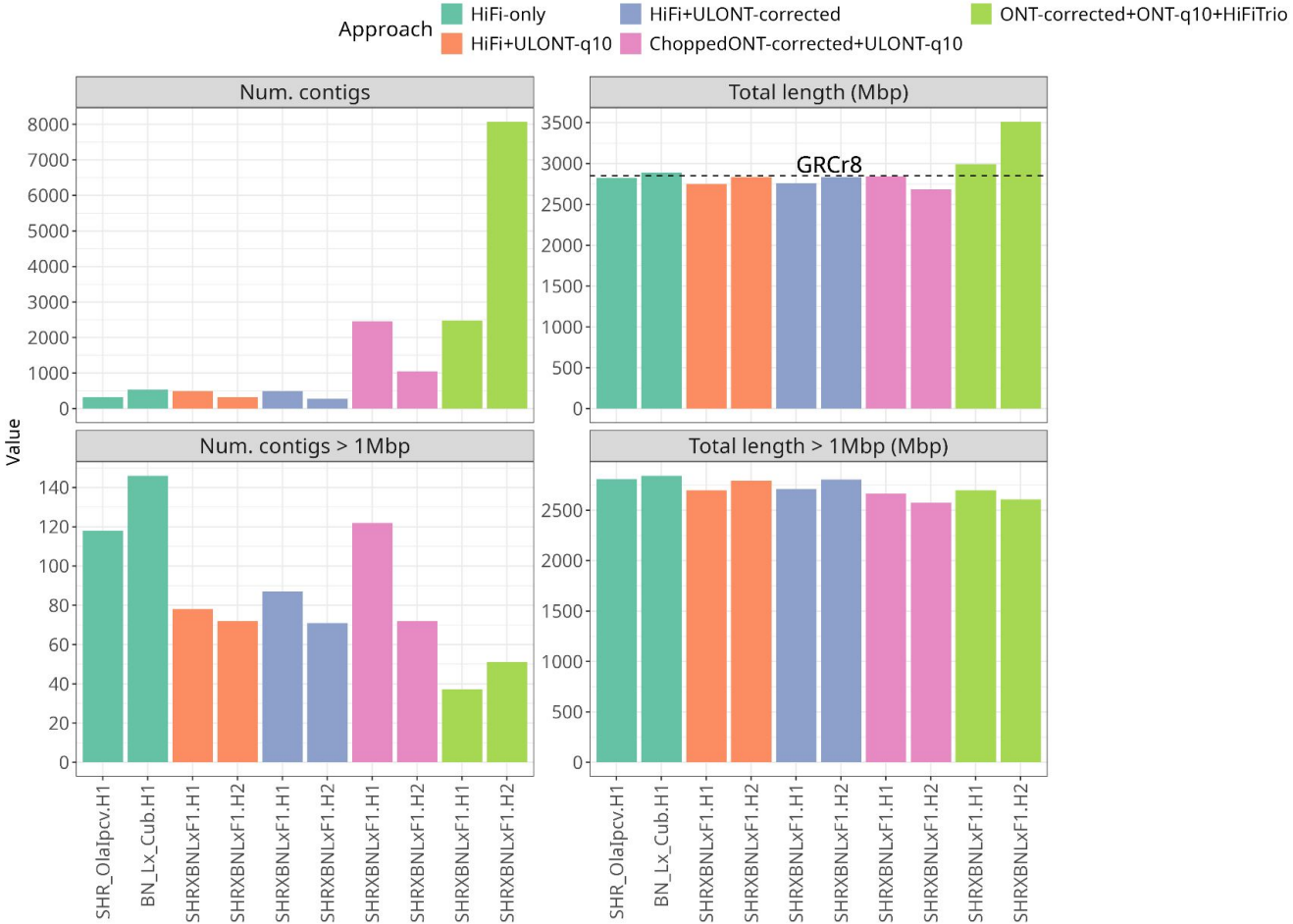
HiFi data generated in collaborations with
Abraham Palmer (NIDA P50)
Robert W. Williams (NIDA P30)

Pangenomics of nicotine abuse in the hybrid rat diversity panel
U01DA057530-02 (MPIs Burt Sharp, Hao Chen, Rob Williams)

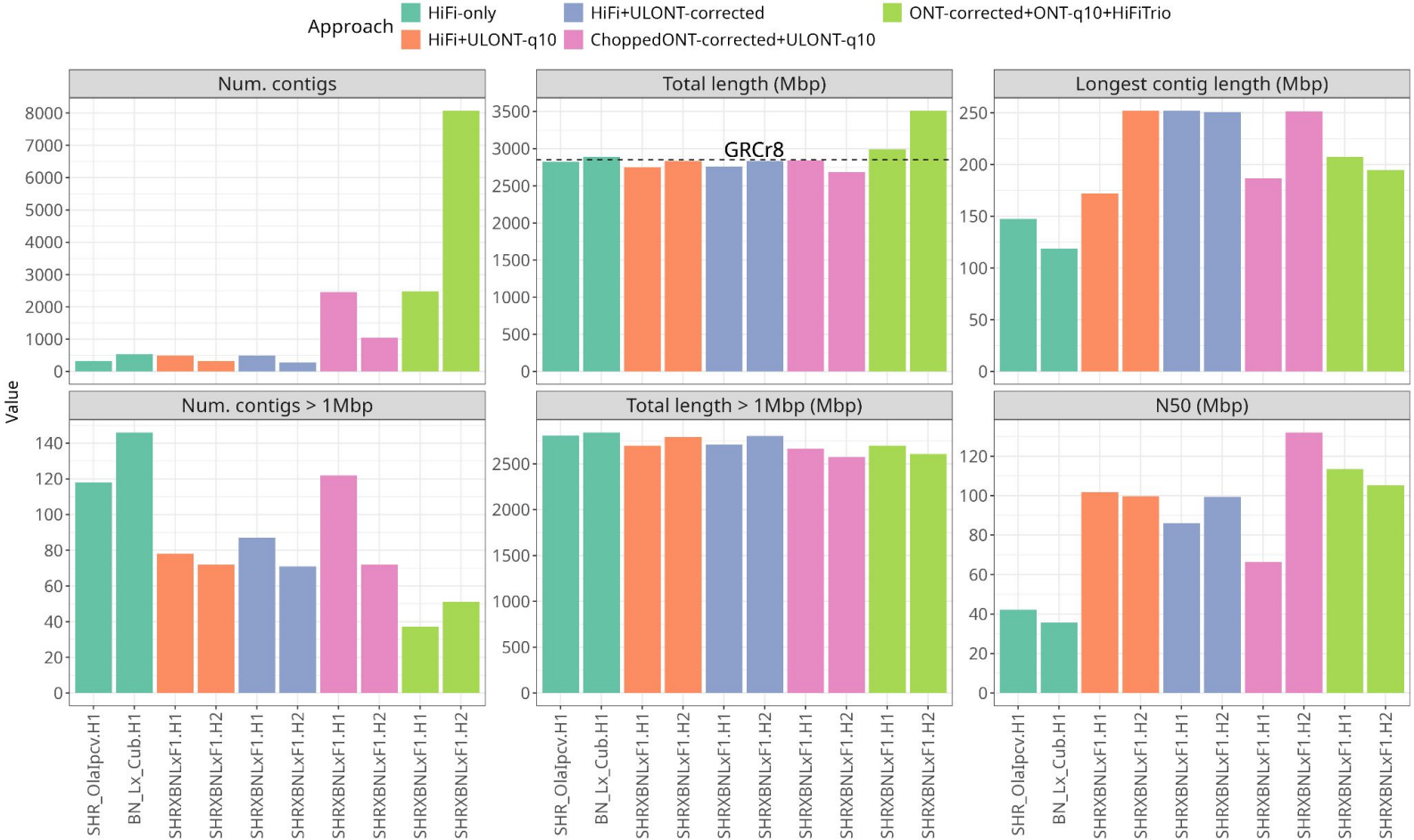
Assembly contiguity



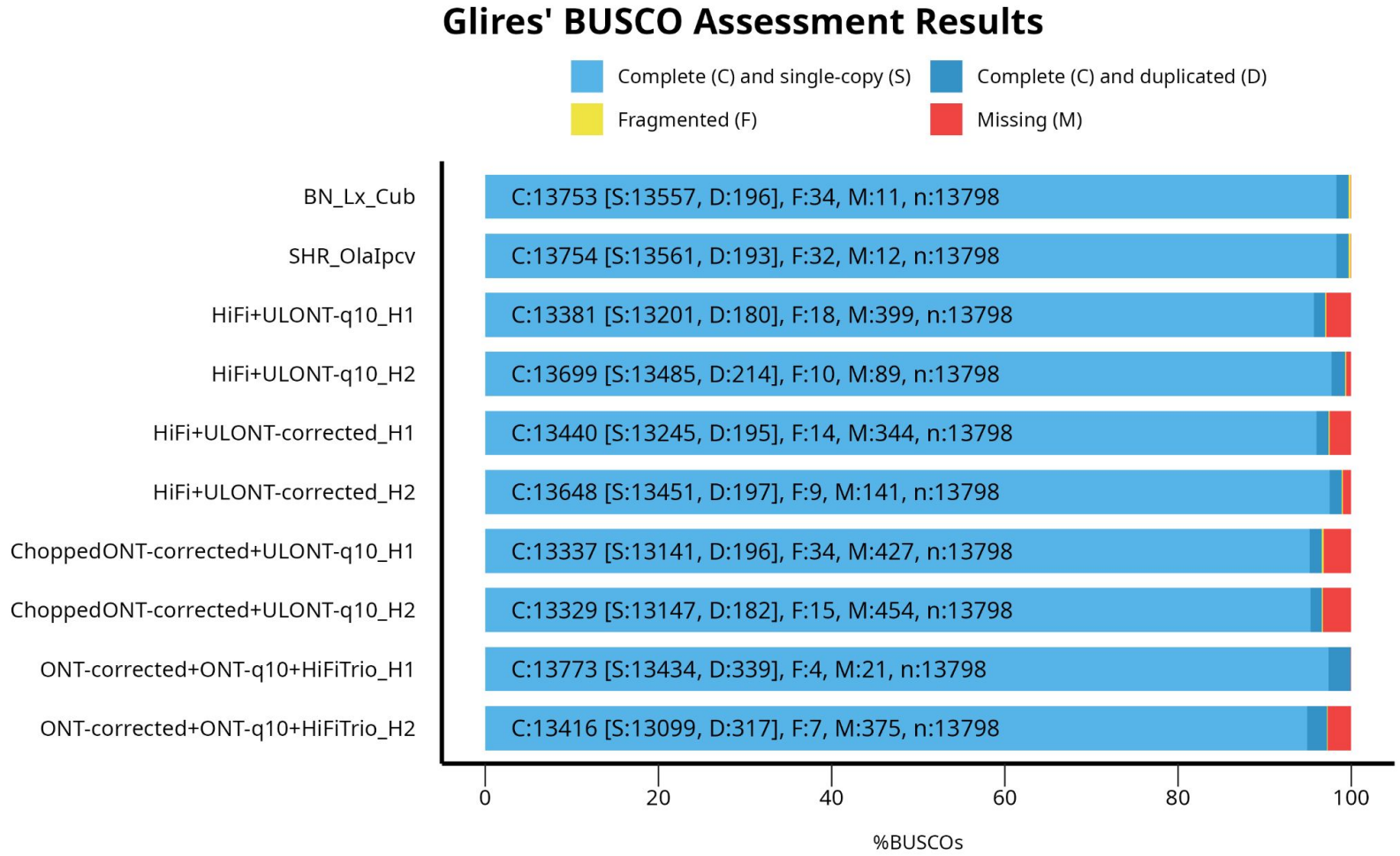
Assembly contiguity



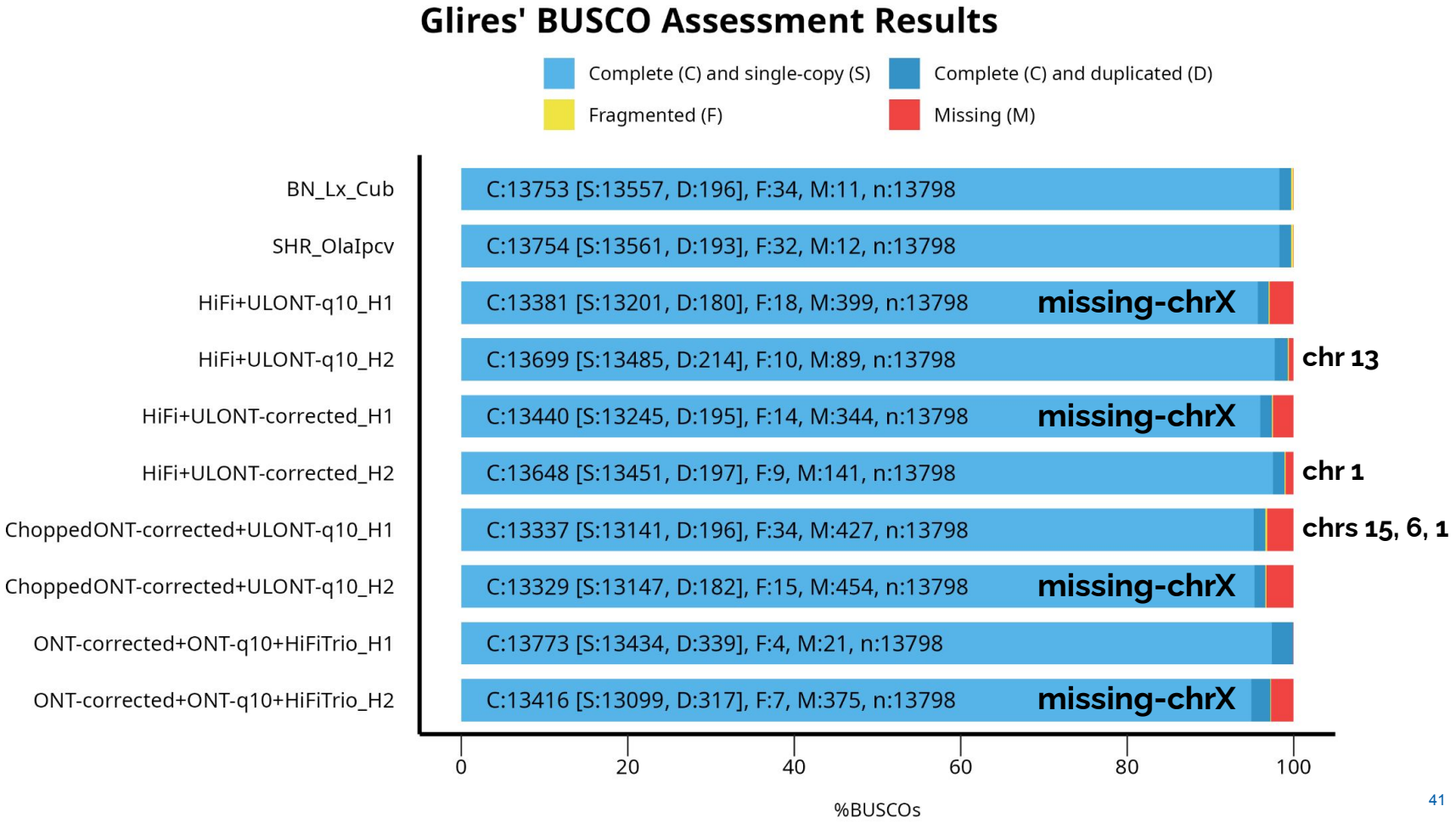
Assembly contiguity



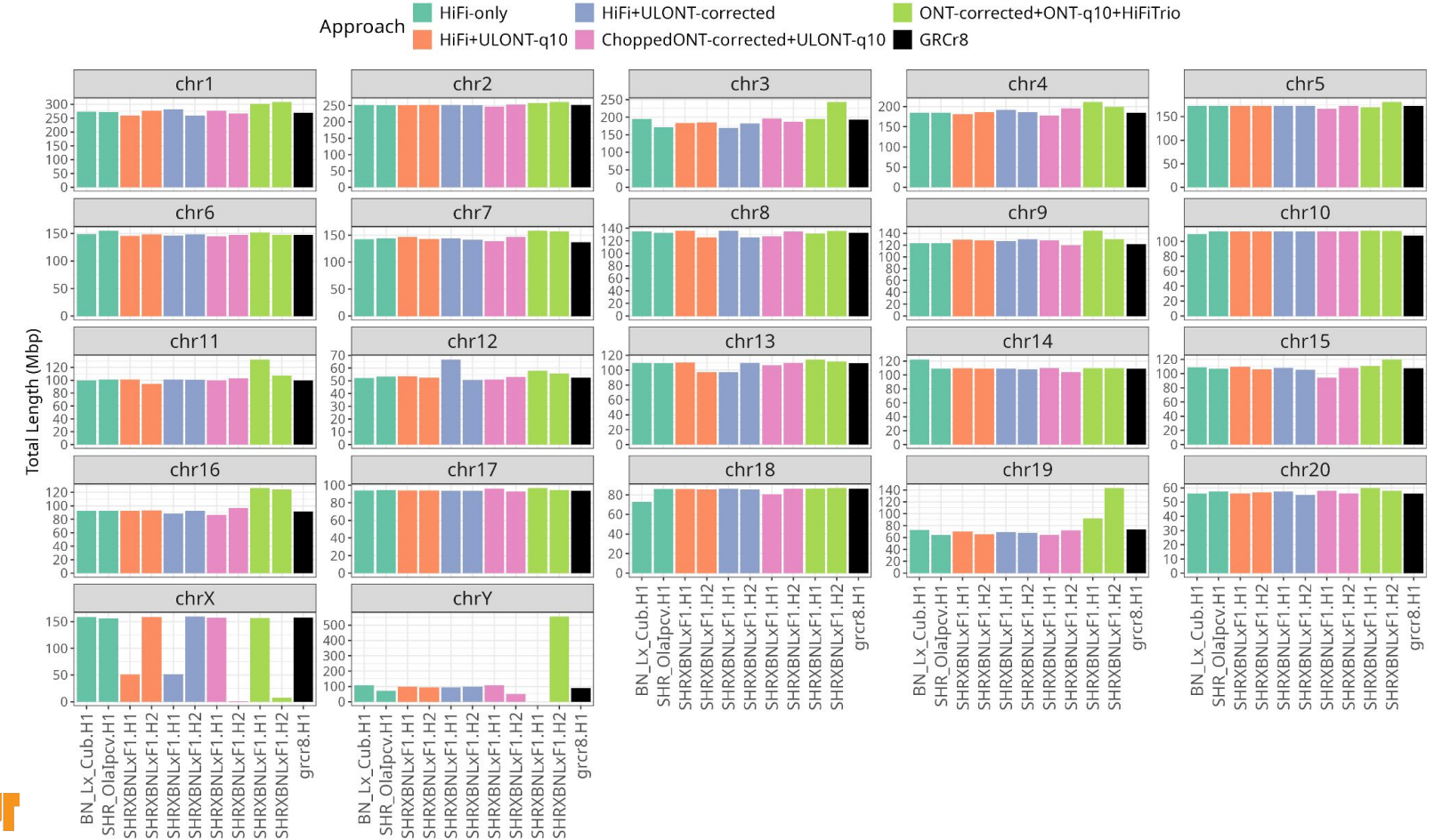
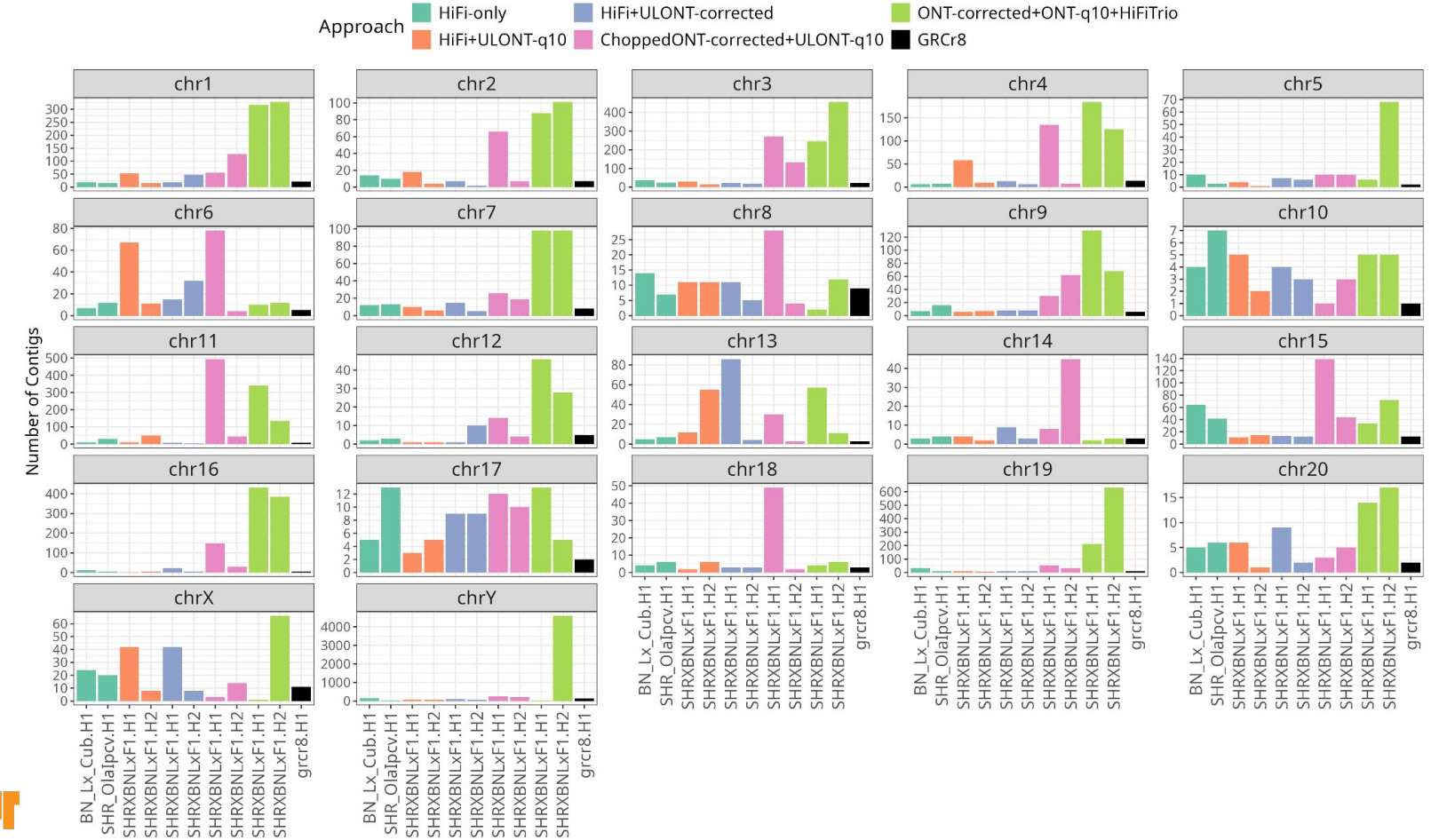
BUSCO


Glires' BUSCO Assessment Results
Complete (C) and single-copy (S)
Complete (C) and duplicated (D)
Fragmented (F)
Missing (M)
BN_Lx_Cub
C:13753 [S:13557, D:196], F:34, M:11, n:13798
SHR_OlaIpcv
C:13754 [S:13561, D:193], F:32, M:12, n:13798
HiFi+ULONT-q10_H1
C:13381 [S:13201, D:180], F:18, M:399, n:13798
HiFi+ULONT-q10_H2
C:13699 [S:13485, D:214], F:10, M:89, n:13798
HiFi+ULONT-corrected_H1
C:13440 [S:13245, D:195], F:14, M:344, n:13798
HiFi+ULONT-corrected_H2
C:13648 [S:13451, D:197], F:9, M:141, n:13798
ChoppedONT-corrected+ULONT-q10_H1
C:13337 [S:13141, D:196], F:34, M:427, n:13798
ChoppedONT-corrected+ULONT-q10_H2
C:13329 [S:13147, D:182], F:15, M:454, n:13798
ONT-corrected+ONT-q10+HiFiTrio_H1
C:13773 [S:13434, D:339], F:4, M:21, n:13798
ONT-corrected+ONT-q10+HiFiTrio_H2
C:13416 [S:13099, D:317], F:7, M:375, n:13798
0
20
40
60
80
100
%BUSCOs

BUSCO


Glires' BUSCO Assessment Results
Complete (C) and single-copy (S)
Complete (C) and duplicated (D)
Fragmented (F)
Missing (M)
BN_Lx_Cub
C:13753 [S:13557, D:196], F:34, M:11, n:13798
SHR_OlaIpcv
C:13754 [S:13561, D:193], F:32, M:12, n:13798
HiFi+ULONT-q10_H1
C:13381 [S:13201, D:180], F:18, M:399, n:13798
missing-chrX
HiFi+ULONT-q10_H2
C:13699 [S:13485, D:214], F:10, M:89, n:13798
chr 13
HiFi+ULONT-corrected_H1
C:13440 [S:13245, D:195], F:14, M:344, n:13798
missing-chrX
HiFi+ULONT-corrected_H2
C:13648 [S:13451, D:197], F:9, M:141, n:13798
chr 1
ChoppedONT-corrected+ULONT-q10_H1
C:13337 [S:13141, D:196], F:34, M:427, n:13798
chrs 15, 6, 1
ChoppedONT-corrected+ULONT-q10_H2
C:13329 [S:13147, D:182], F:15, M:454, n:13798
missing-chrX
ONT-corrected+ONT-q10+HiFiTrio_H1
C:13773 [S:13434, D:339], F:4, M:21, n:13798
ONT-corrected+ONT-q10+HiFiTrio_H2
C:13416 [S:13099, D:317], F:7, M:375, n:13798
missing-chrX
0
20
40
60
80
100
%BUSCOs

Partitioning by chromosome - Chromosome lengths



Partitioning by chromosome - Contig counts