



COMPLEX
TRAIT
COMMUNITY

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

Andrea Guaracino
Postdoctoral Scholar

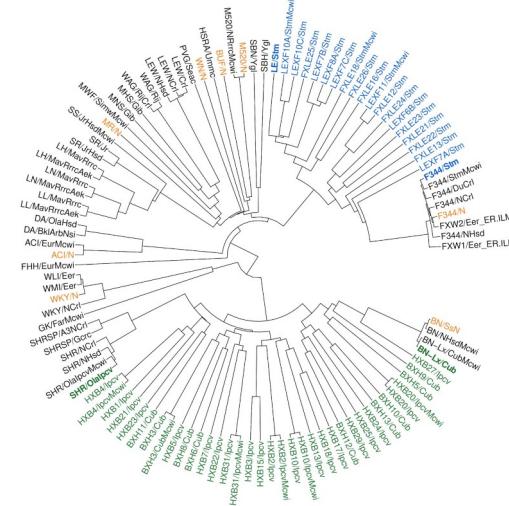
 [@AndresGuarhino](https://twitter.com/AndresGuarhino)

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

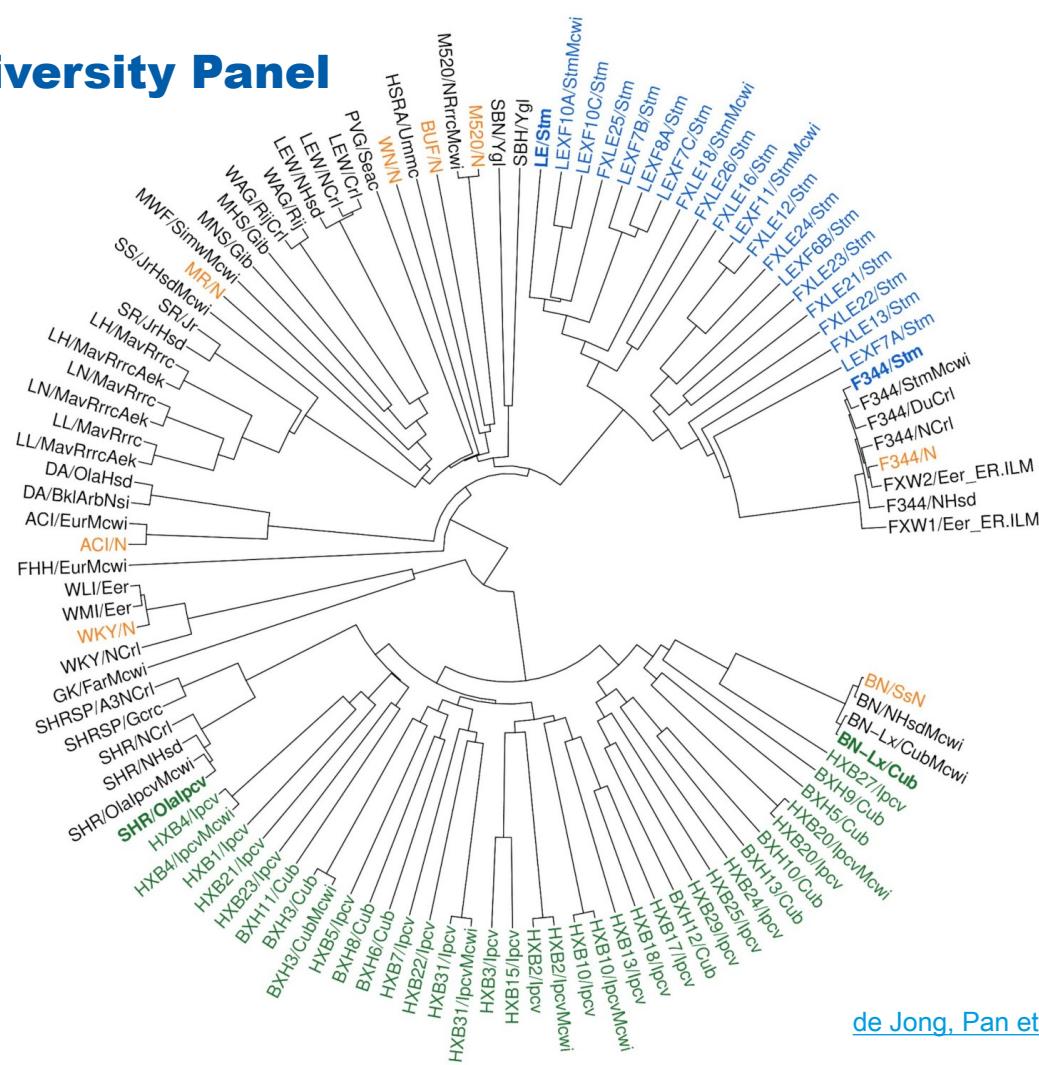


Outline

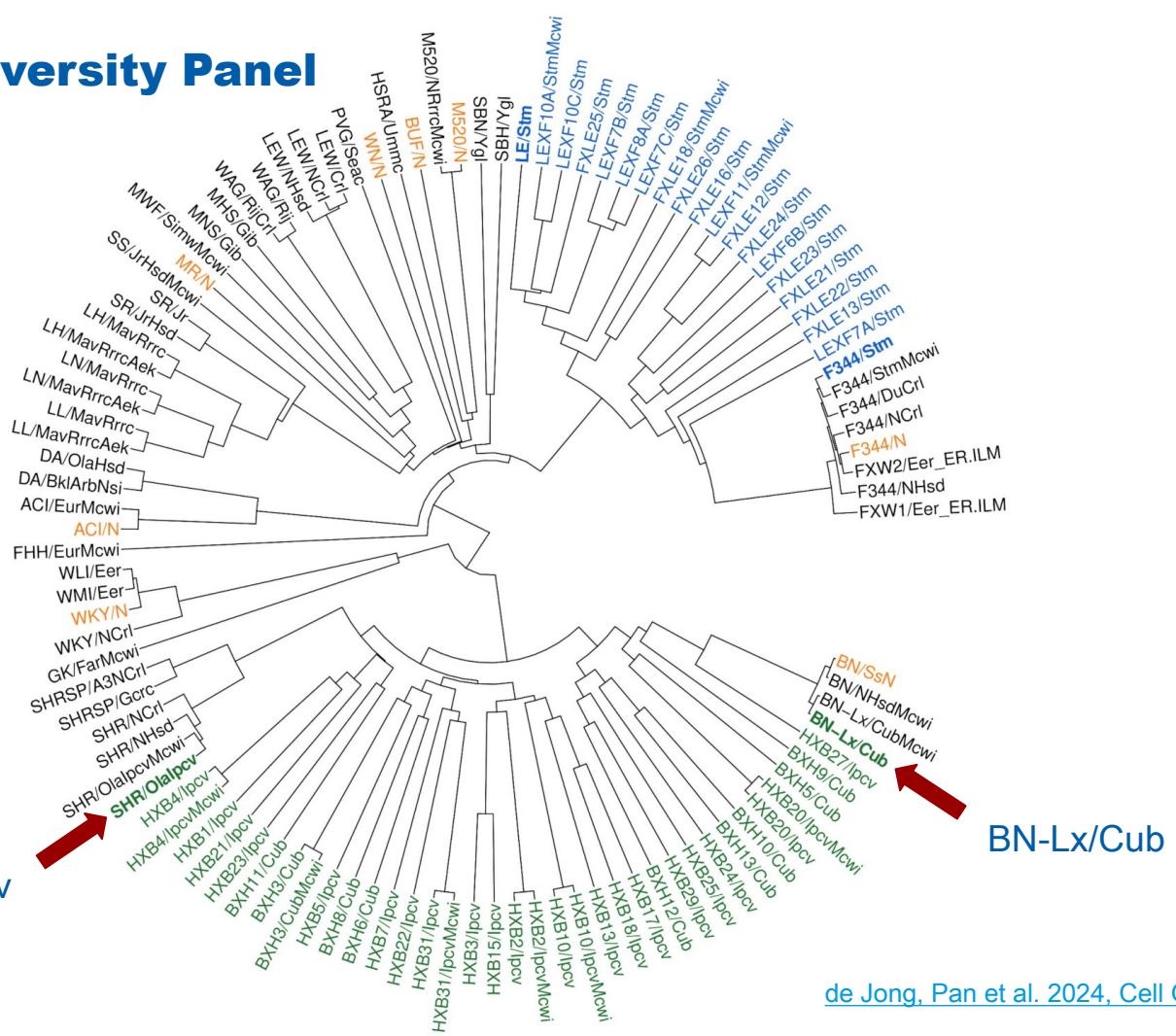
- Introduction
 - Assembly
 - Quality control
 - Adaptive sequencing
 - ONT reads correction



Hybrid Rat Diversity Panel

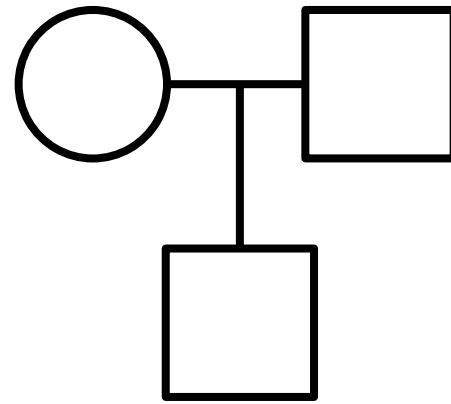


Hybrid Rat Diversity Panel



Trio

SHR/OlaPcv



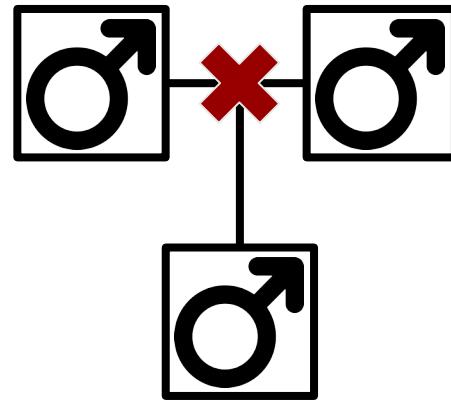
BN-Lx/Cub

SHR/OlaPcv X BN-Lx/Cub F1

“Trio”



SHR/OlaLpcv



BN-Lx/Cub

SHR/OlaLpcv X BN-Lx/Cub F1

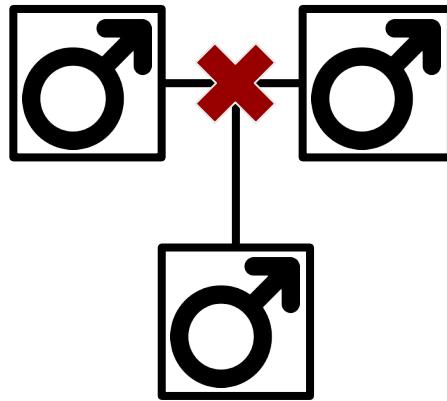
“Trio”



SHR/OlaLpcv

PacBio HiFi reads: 44X coverage

PacBio



BN-Lx/Cub

PacBio HiFi reads: 41X coverage

PacBio

SHR/OlaLpcv X BN-Lx/Cub F1

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



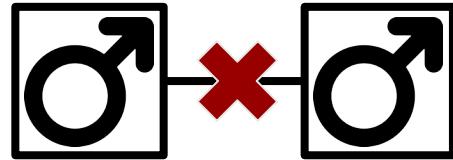
Outline

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



Assembly

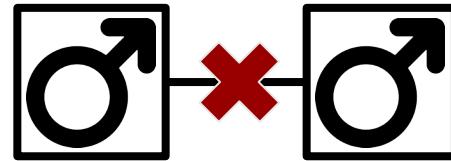
SHR/Olalpcv



BN-Lx/Cub

Assembly

SHR/OlaIpcv



BN-Lx/Cub

PacBio

hifiasm

Haploid assembly



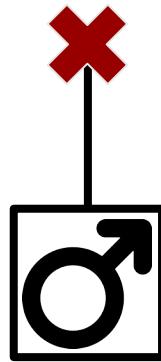
PacBio

hifiasm

Haploid assembly

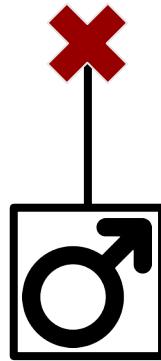


Assembly



SHR/Olaalpcv X BN-Lx/Cub F1

Assembly



SHR/Olalpcv X BN-Lx/Cub F1

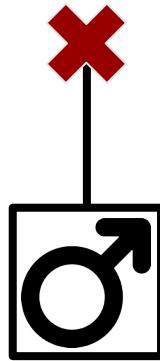
SHR/Olalpcv **PacBio**
BN-Lx/Cub **PacBio**
50kbps  **NANOPORE**
Technologies

hifiasm

Diploid assembly



Assembly



SHR/Olalpcv X BN-Lx/Cub F1

SHR/Olalpcv **PacBio**
BN-Lx/Cub **PacBio**
50kbps  **NANOPORE**
Technologies

HERRO-corrected  **NANOPORE**
Technologies
50kbps  **NANOPORE**
Technologies

hifiasm

Diploid assembly

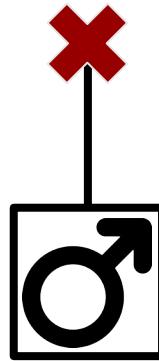


hifiasm

Diploid assembly



Assembly



SHR/Olalpcv X BN-Lx/Cub F1

SHR/Olalpcv **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

Diploid assembly



hifiasm

Diploid assembly



verkko

Phased diploid assembly

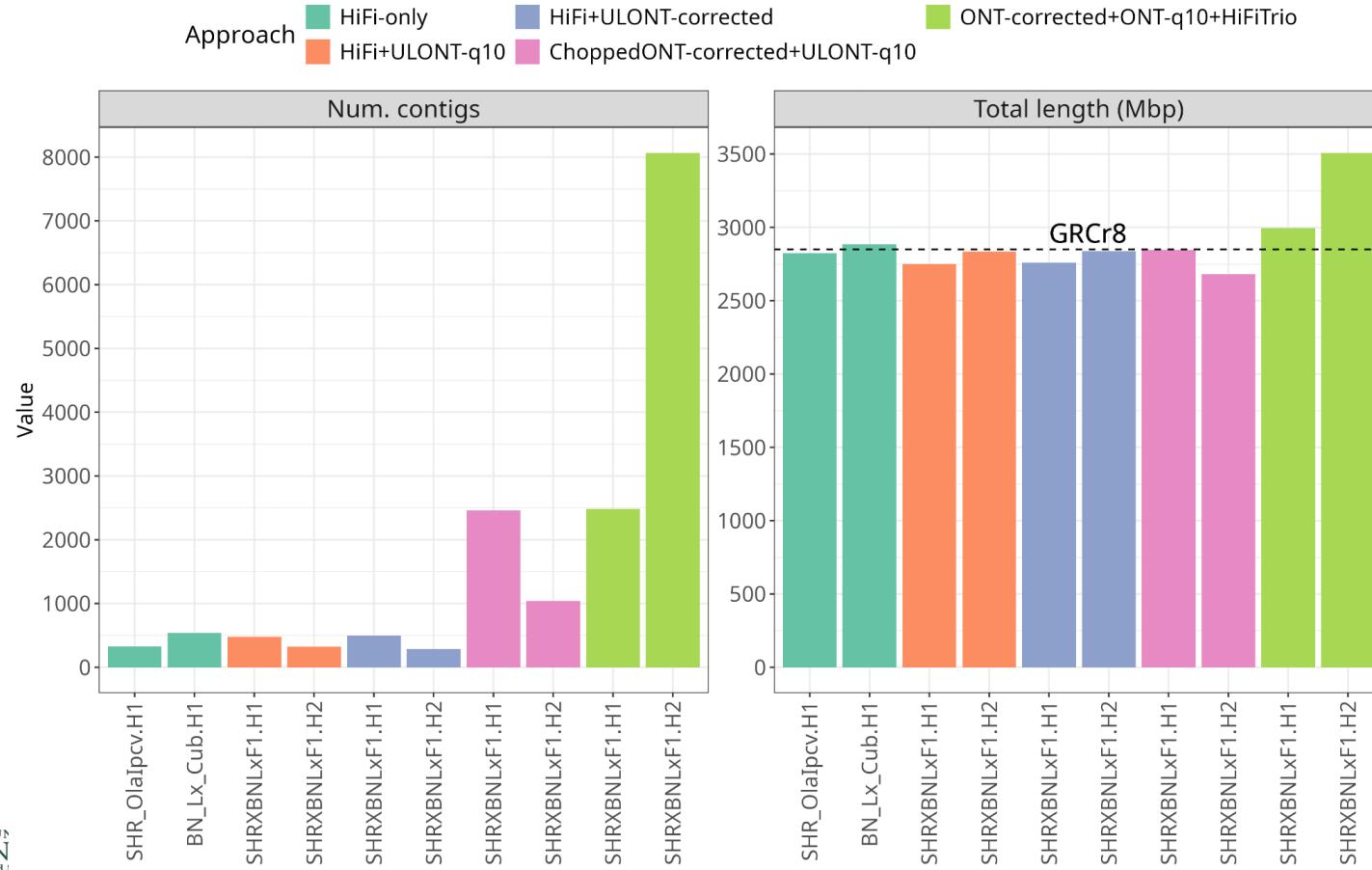


Outline

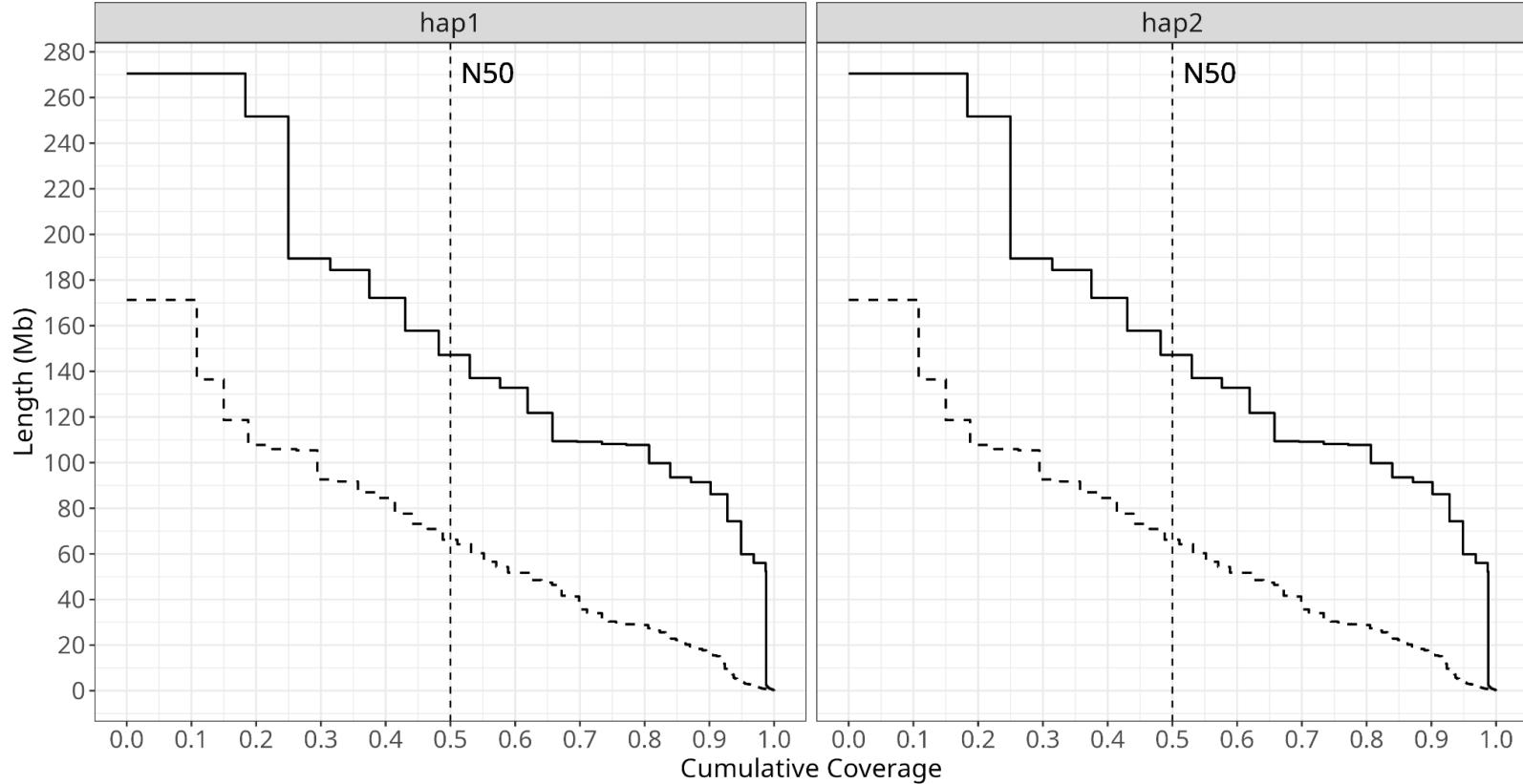
- Introduction
- Assembly
- **Quality control**
- Adaptive sequencing
- ONT reads correction



Assembly contiguity

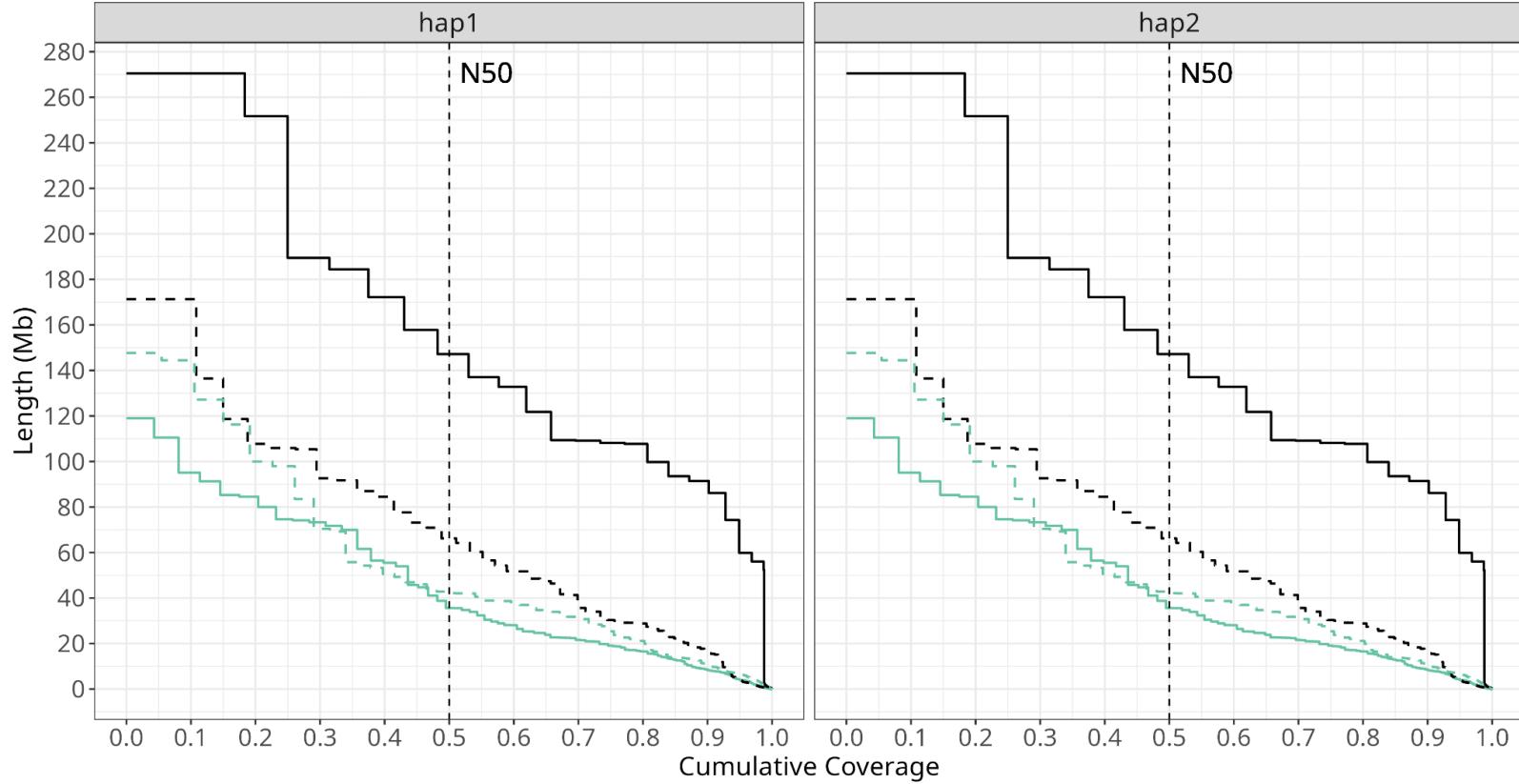


Assembly contiguity



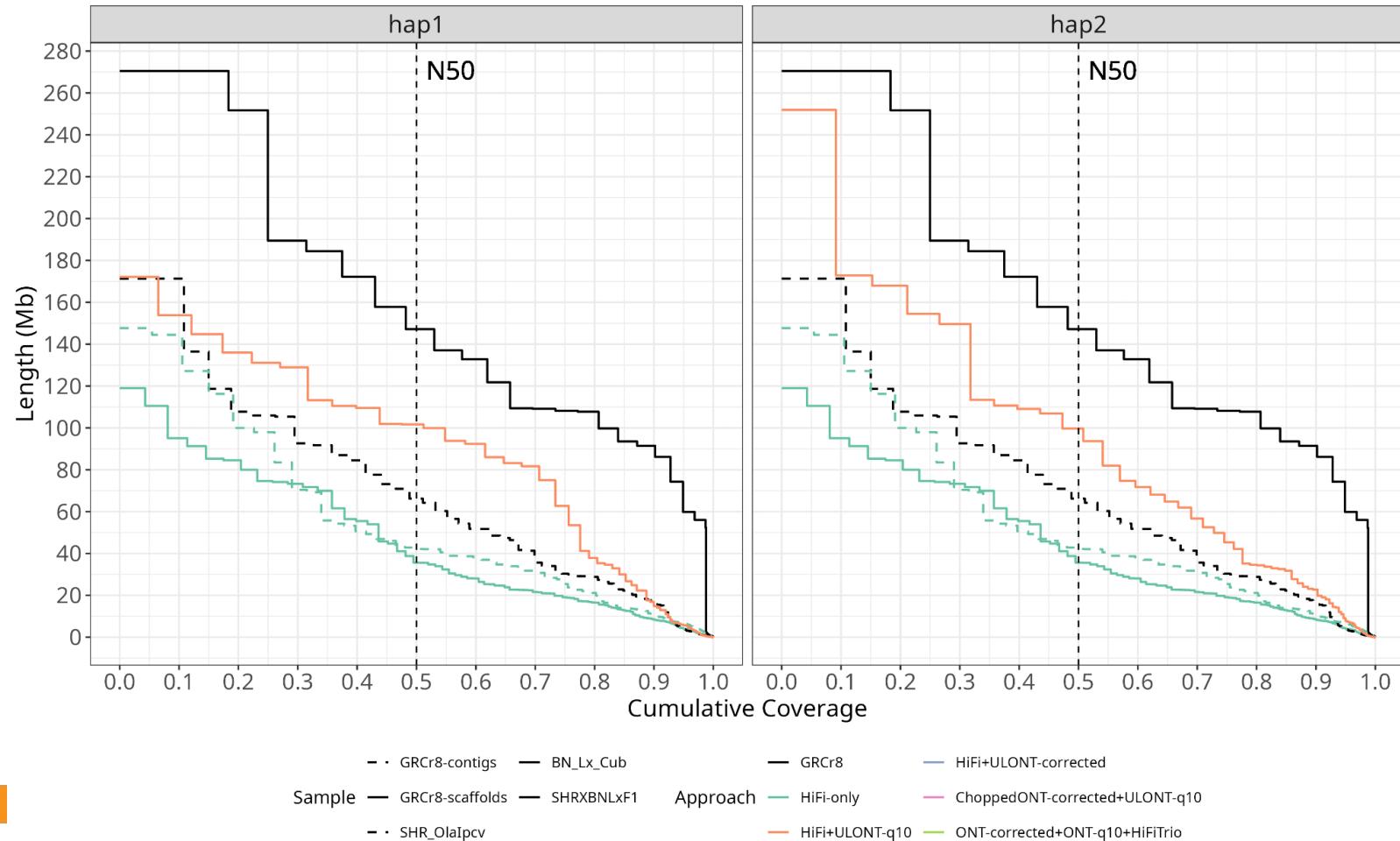
Sample	GRCr8-scaffolds	SHR_XBNLx_F1	Approach	GRCr8	HiFi+ULONT-corrected
- - - SHR_OlaIpcv	-	-	-	-	-
- - - GRCr8-contigs	-	-	-	-	-

Assembly contiguity

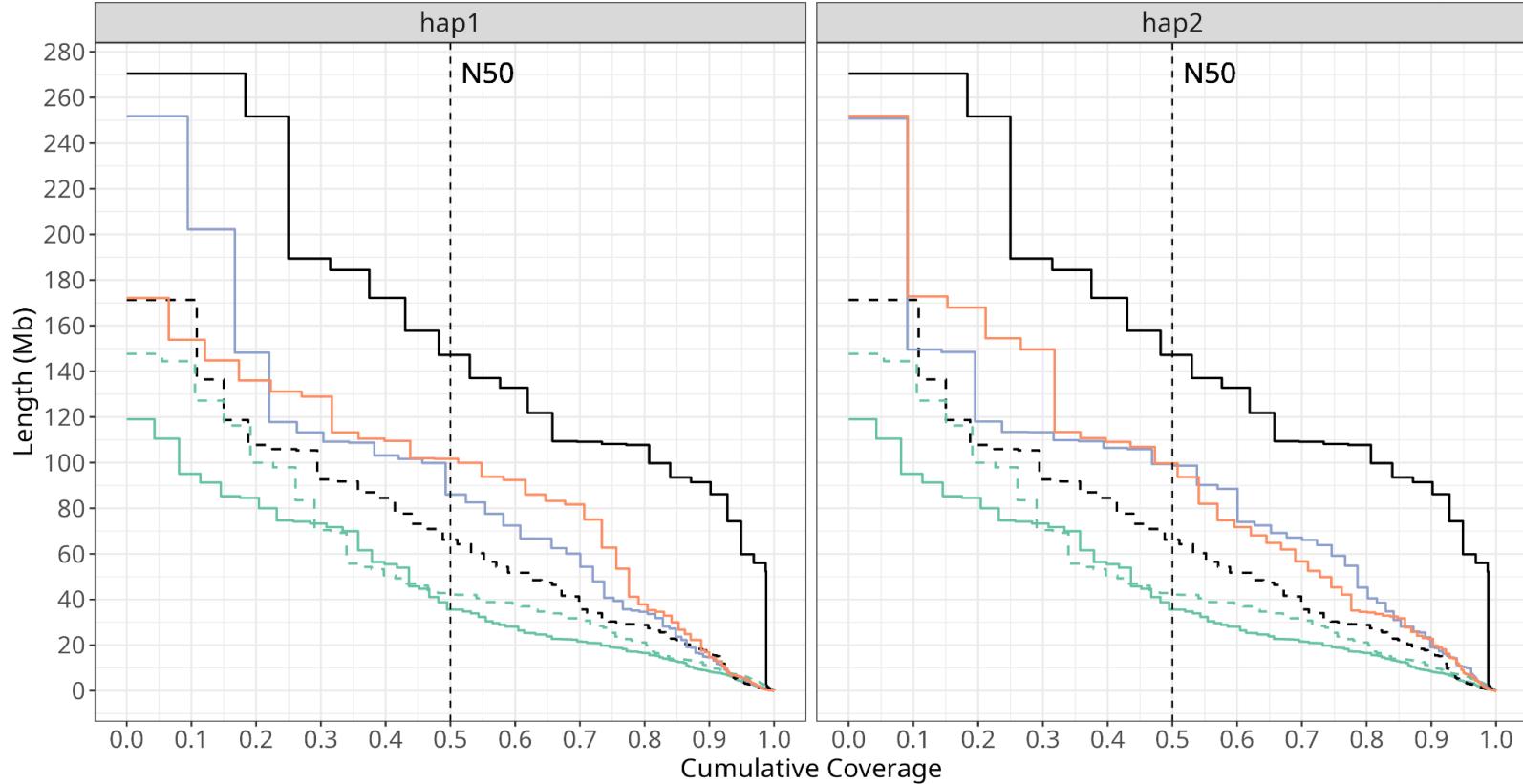


Sample — GRCr8-scaffolds - - - SHR_OlaIpcv	Approach — BN_Lx_Cub — GRCr8 — HiFi+ULONT-corrected — HiFi-only — ChoppedONT-corrected+ULONT-q10 — HiFi+ULONT-q10 — ONT-corrected+ONT-q10+HiFiTrio
--	---

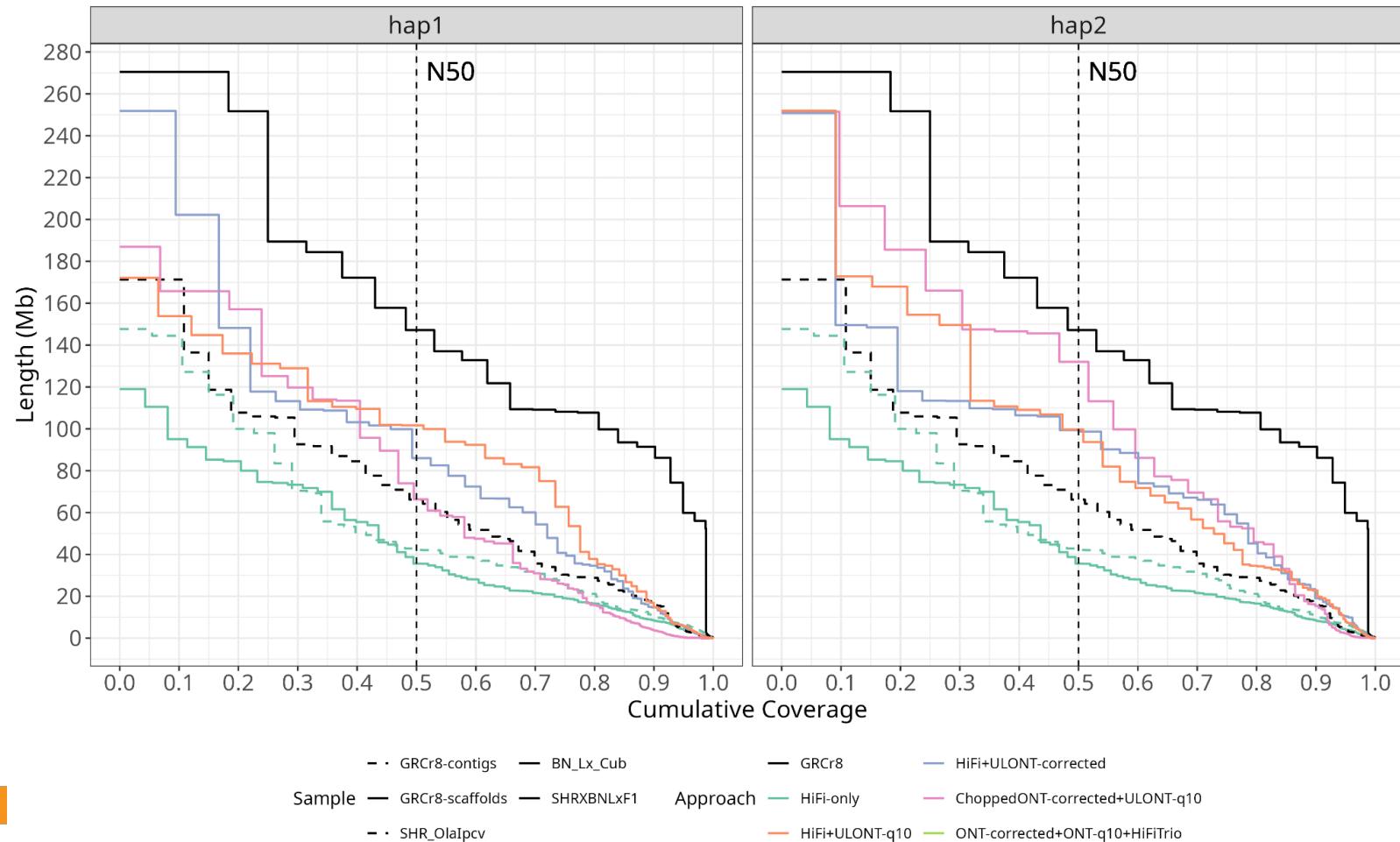
Assembly contiguity



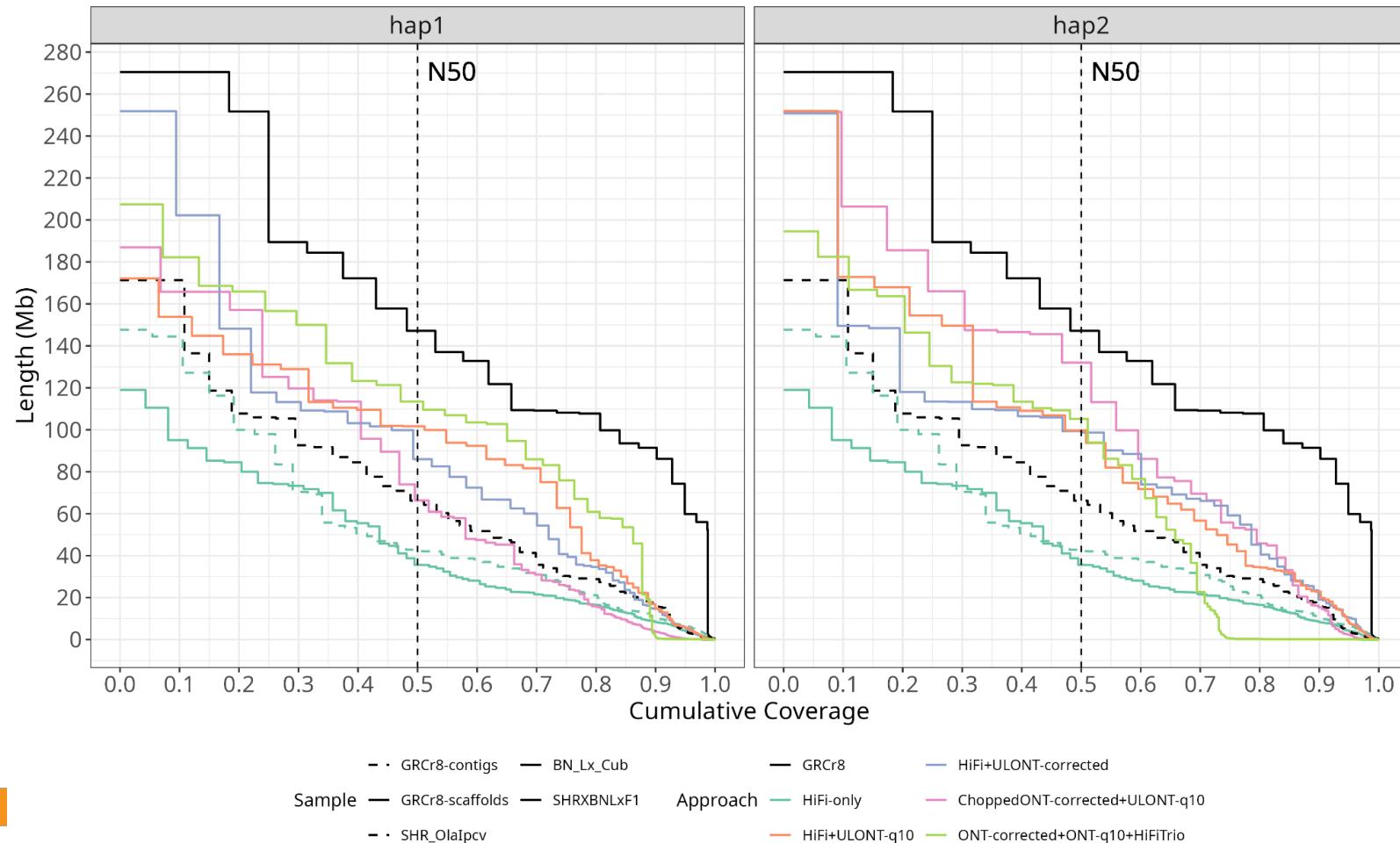
Assembly contiguity



Assembly contiguity



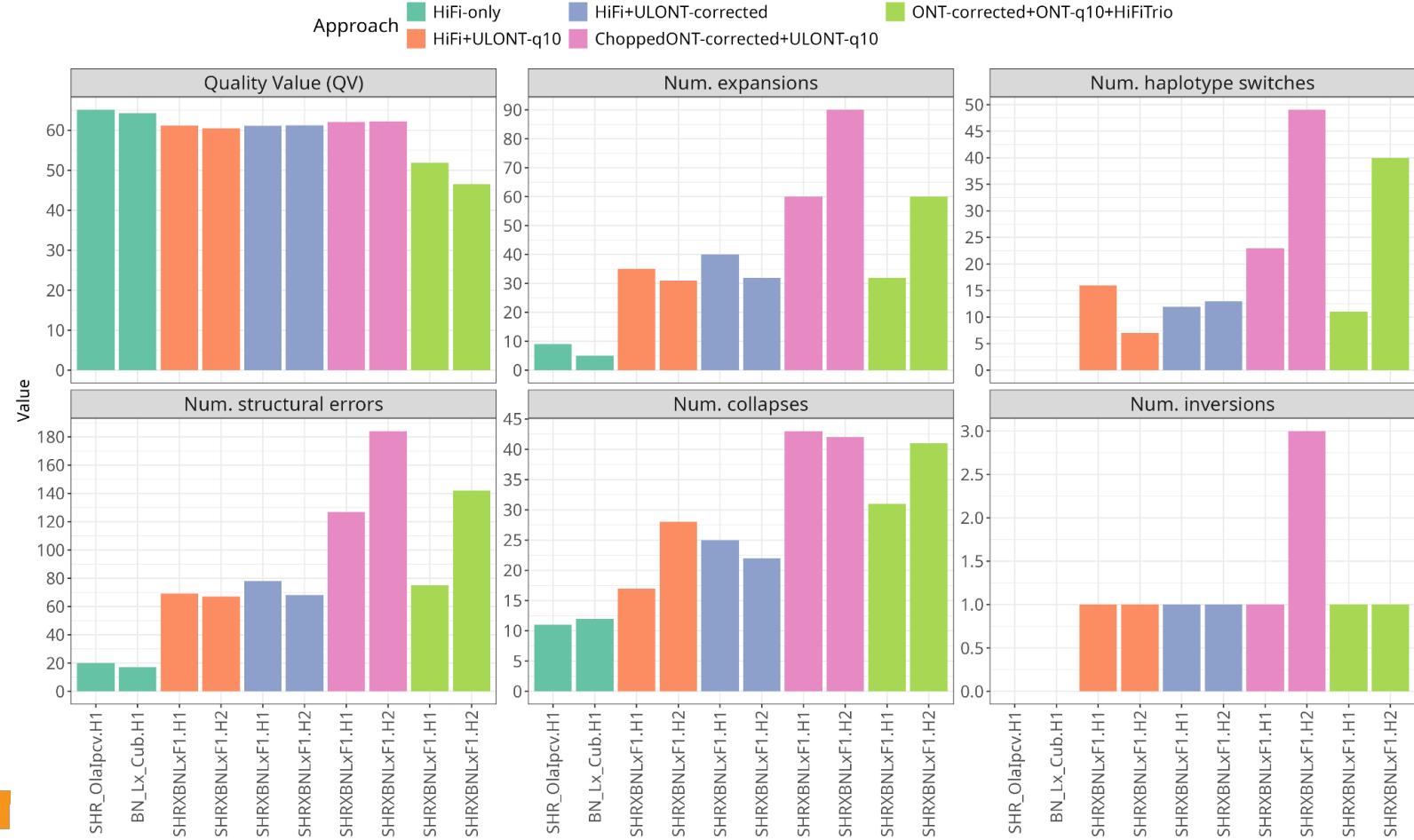
Assembly contiguity



Assembly correctness



Assembly correctness



Outline

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

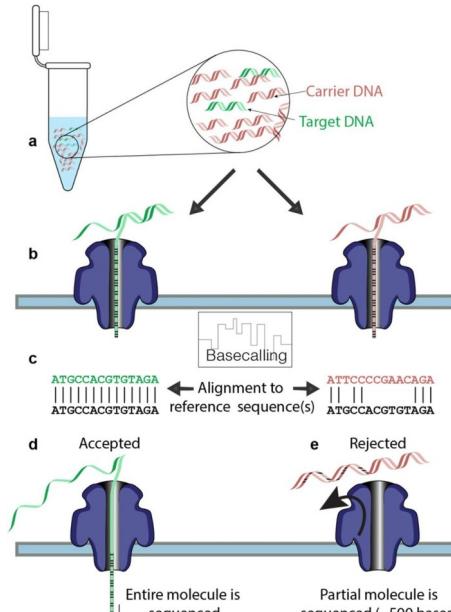
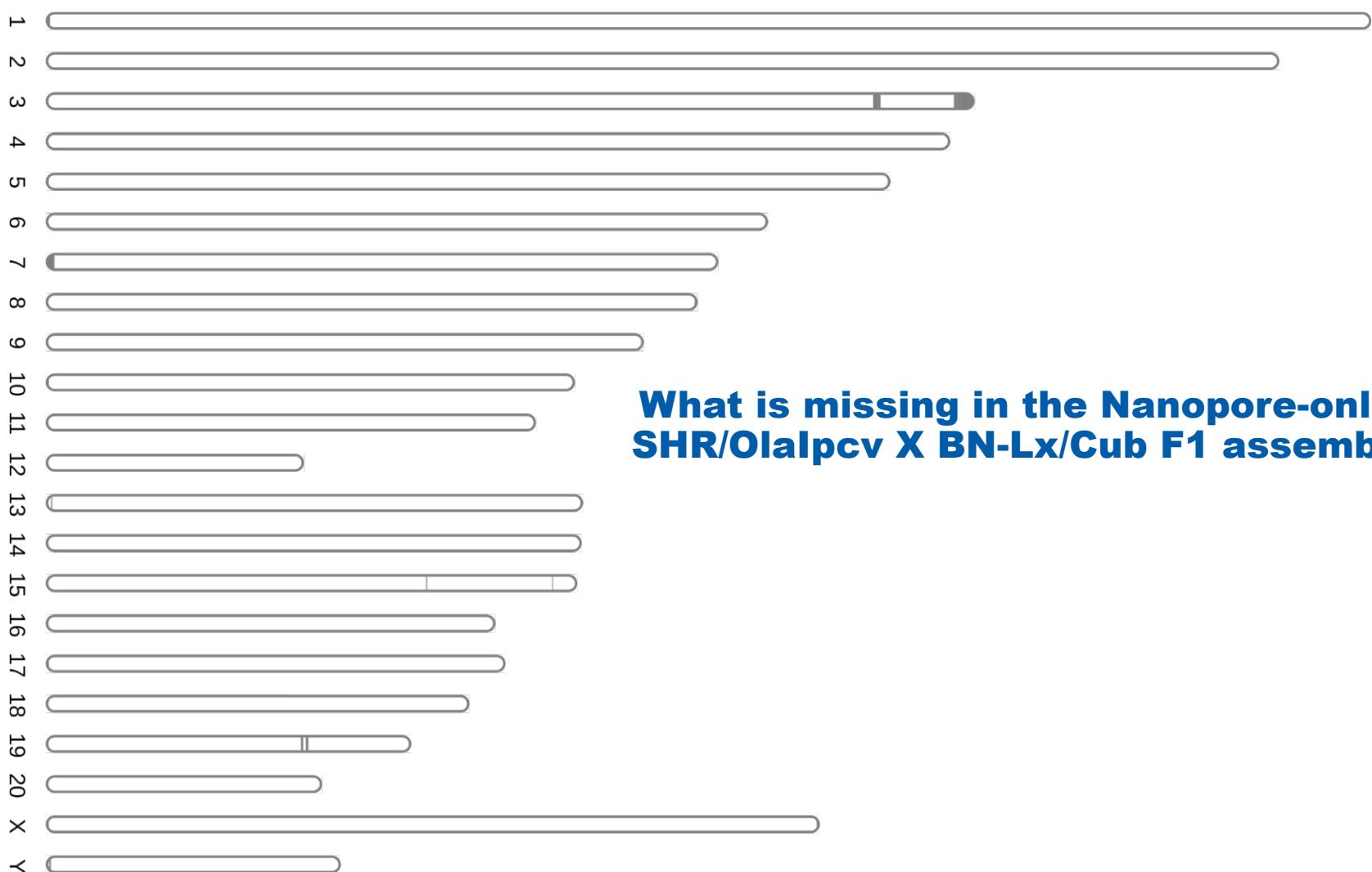
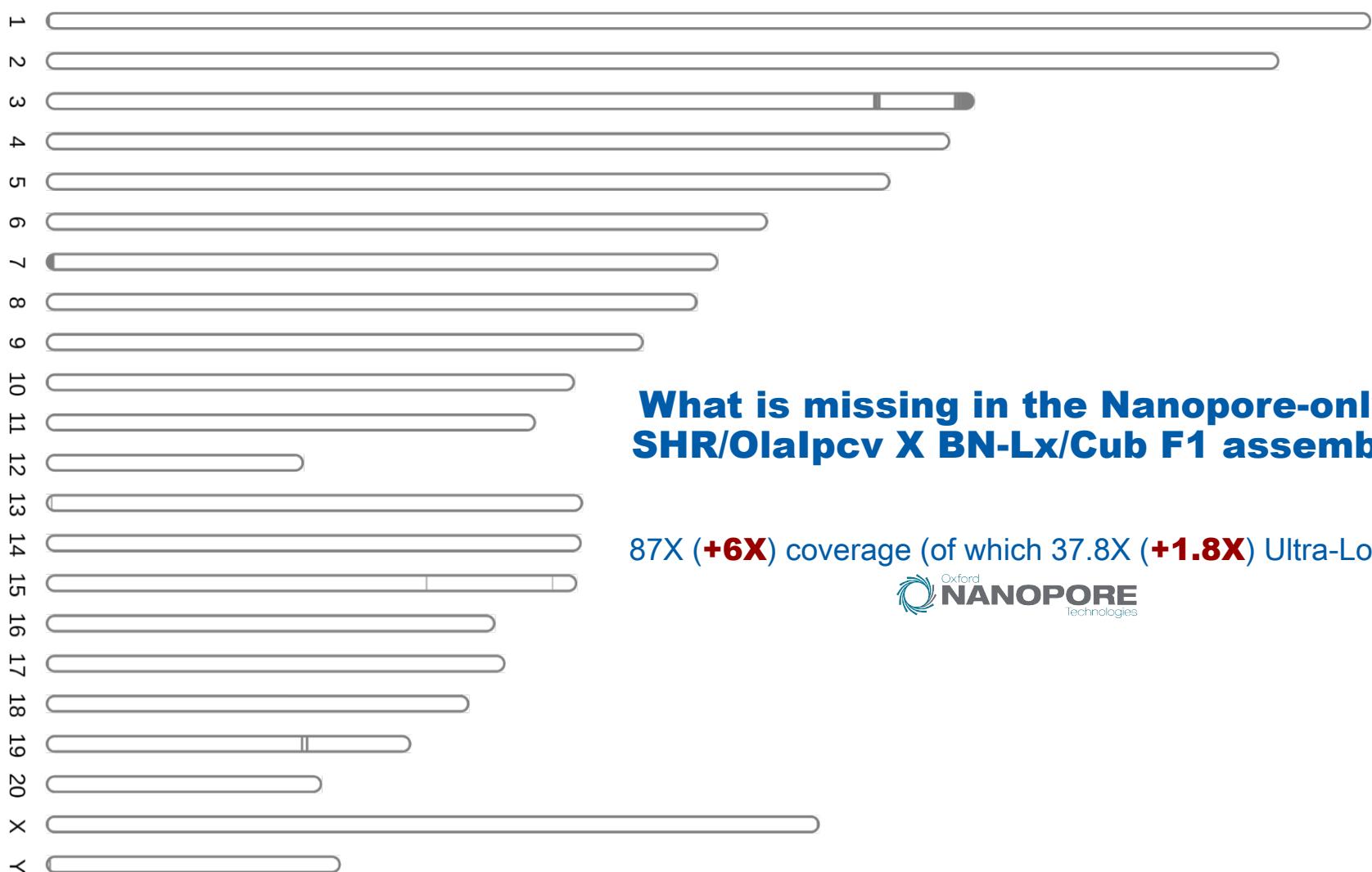


Image from [Miani et al., 2023](#)



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

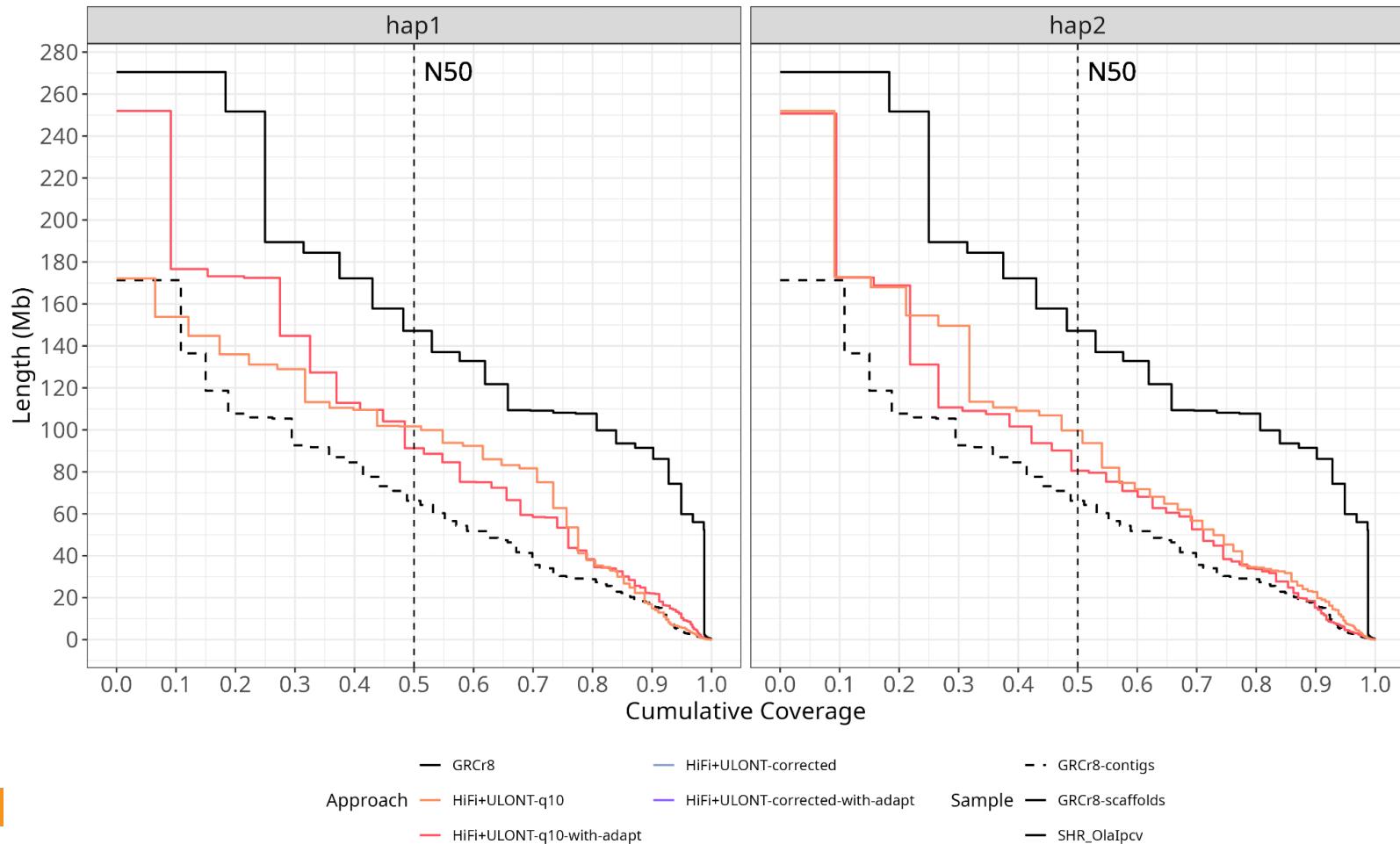


What is missing in the Nanopore-only SHR/Olapcv X BN-Lx/Cub F1 assembly

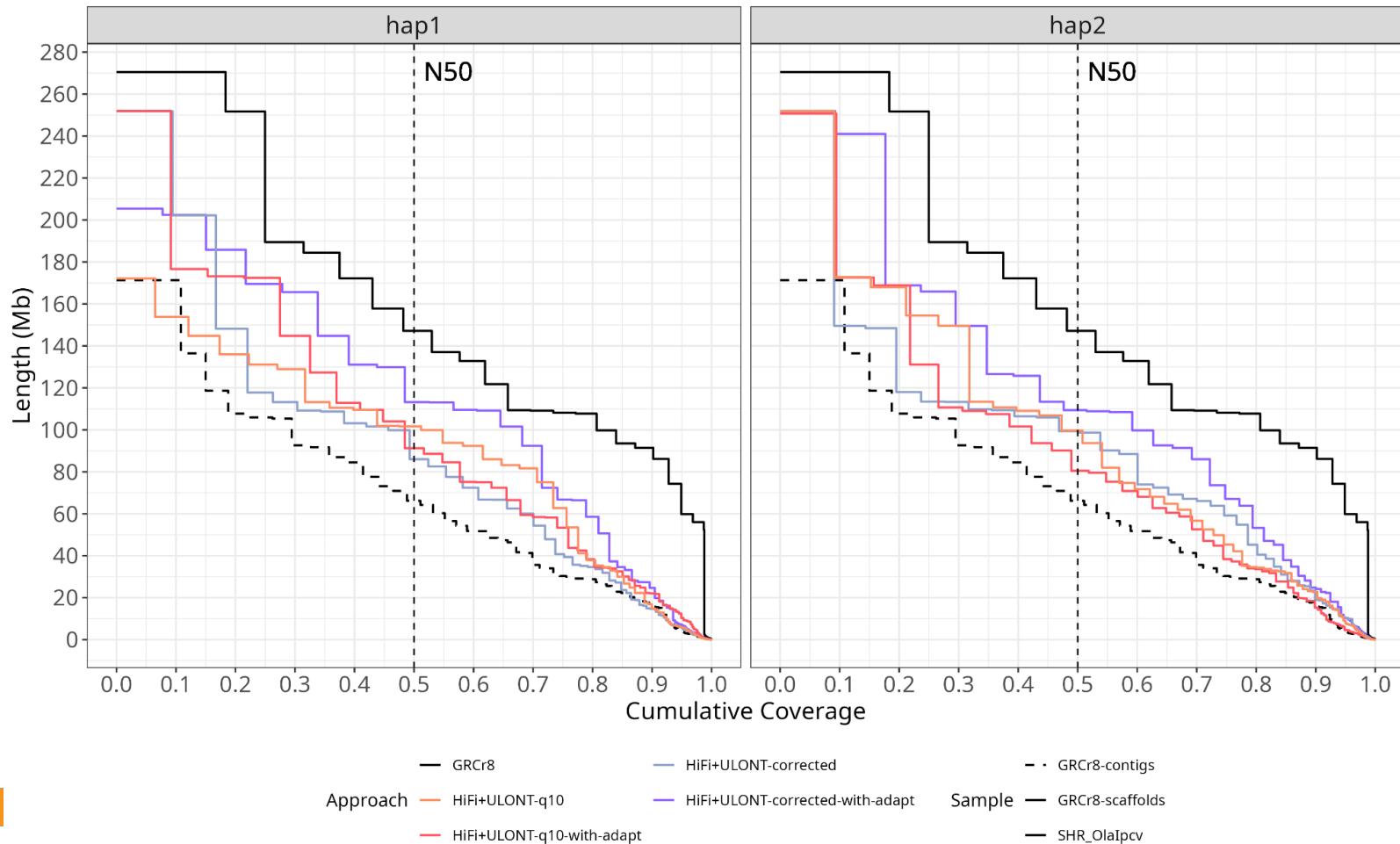
87X (**+6X**) coverage (of which 37.8X (**+1.8X**) Ultra-Long)



Assembly contiguity including adaptive sequencing

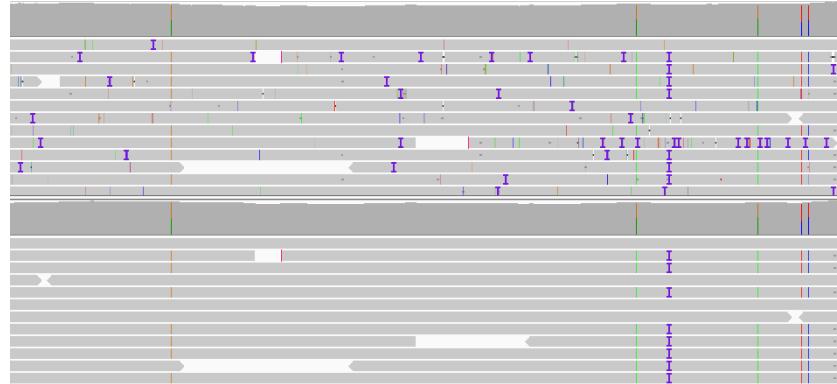


Assembly contiguity including adaptive sequencing

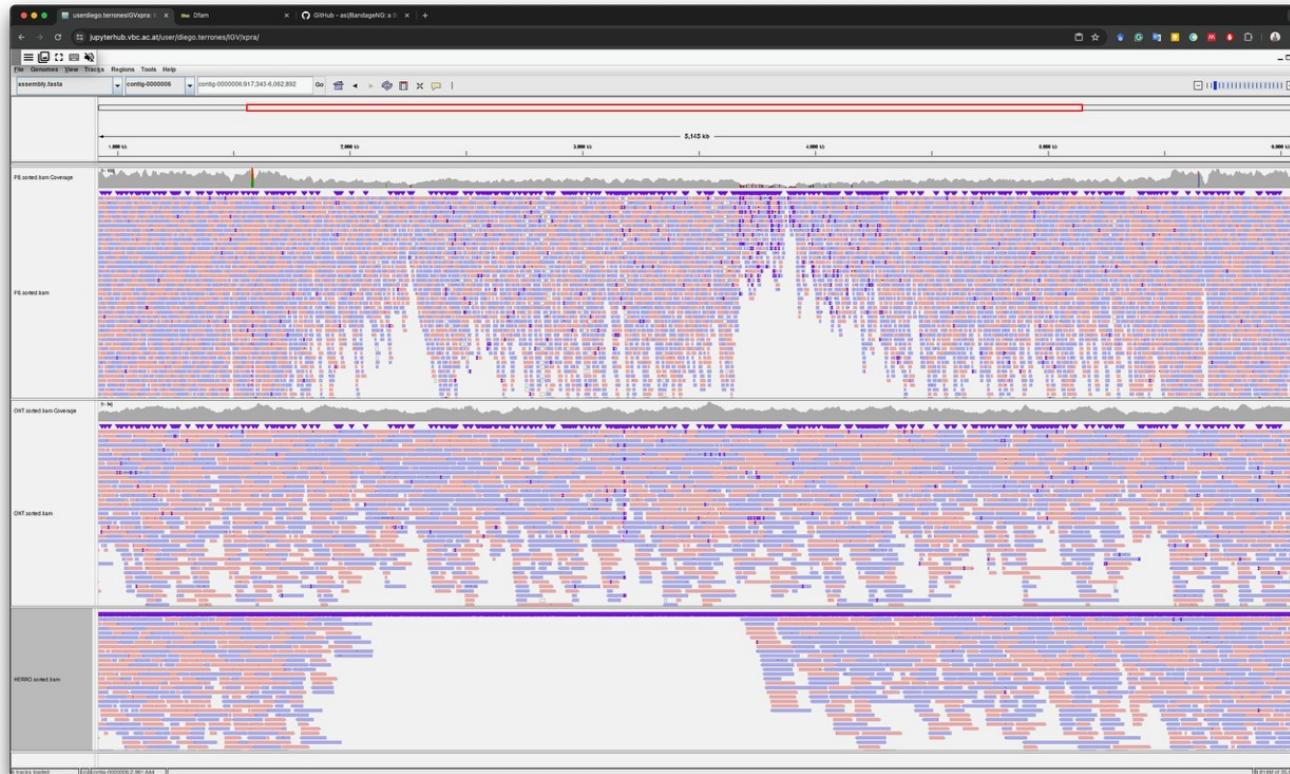


Outline

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



Is HERRO-correction working?

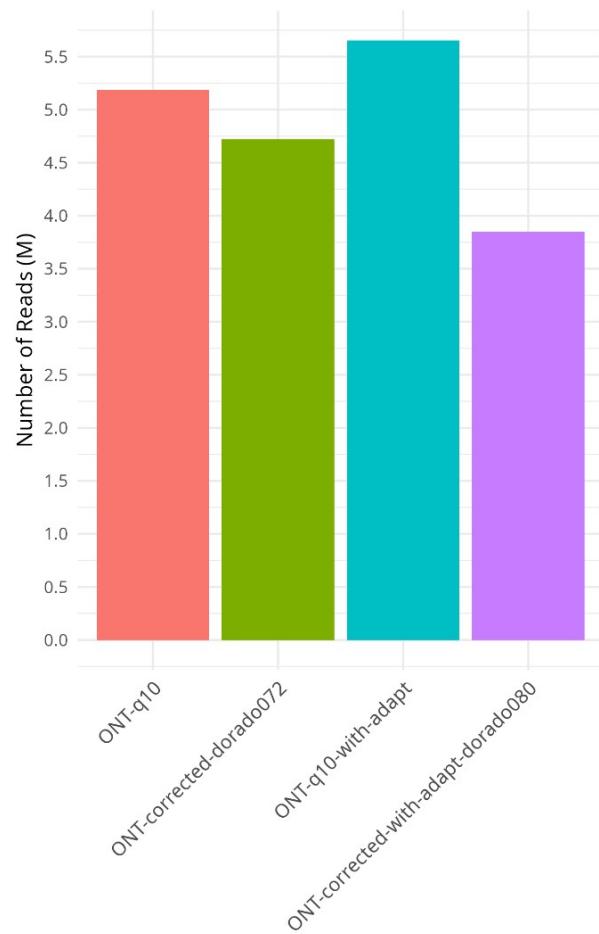


HERRO-correction is working

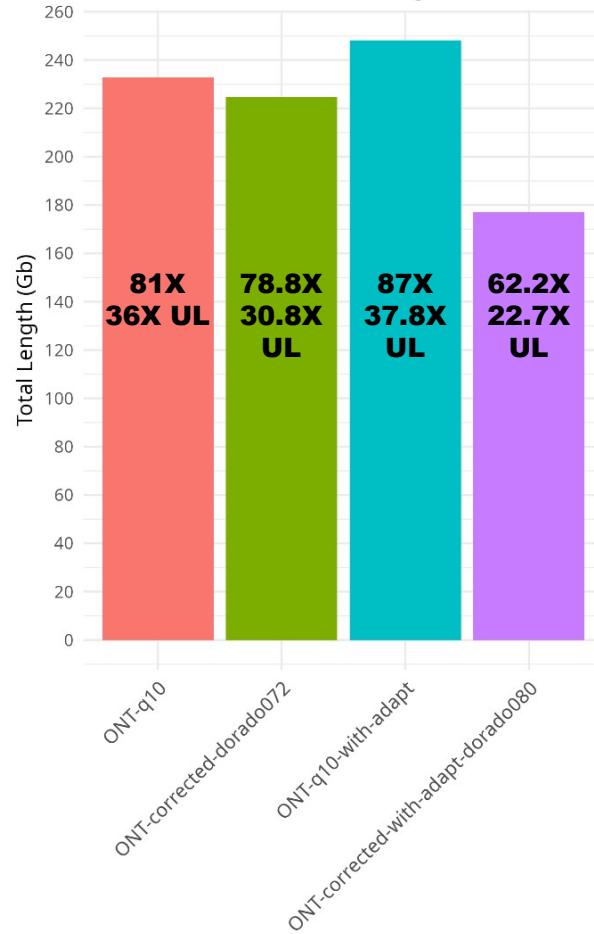


HERRO-correction statistics

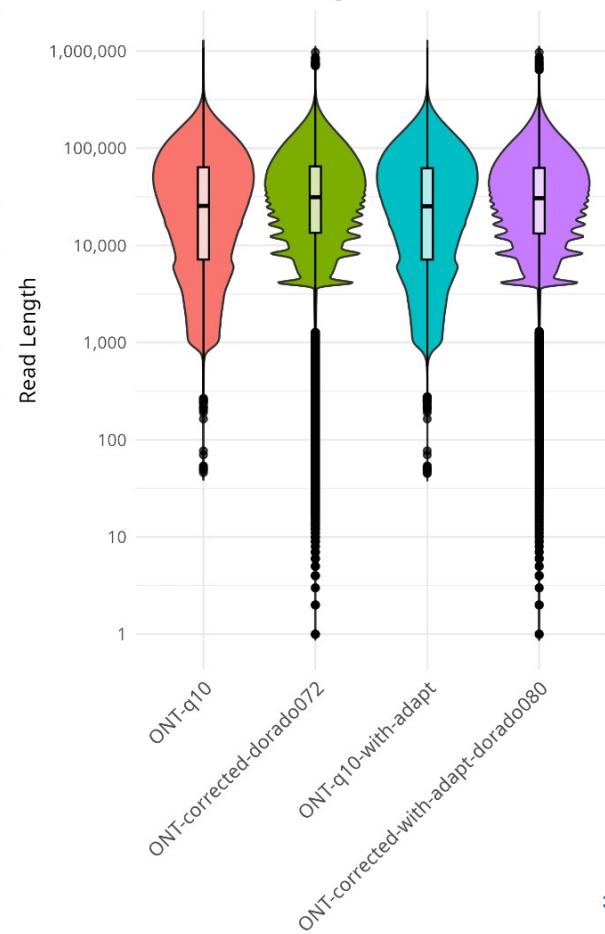
Number of Reads



Total Read Length

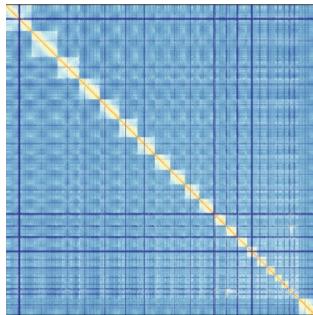


Read Length Distribution



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

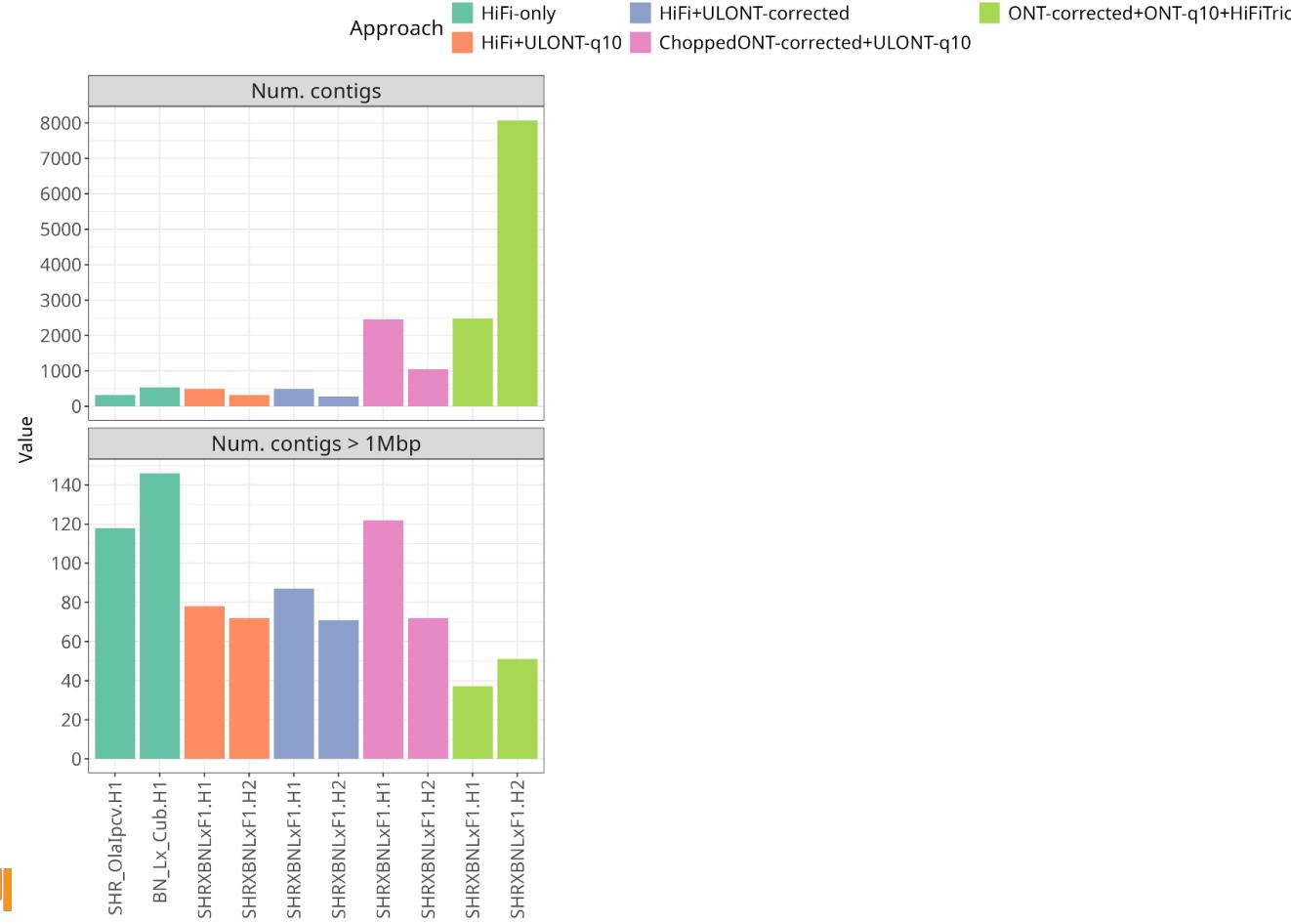
Pjotr Prins

Erik Garrison

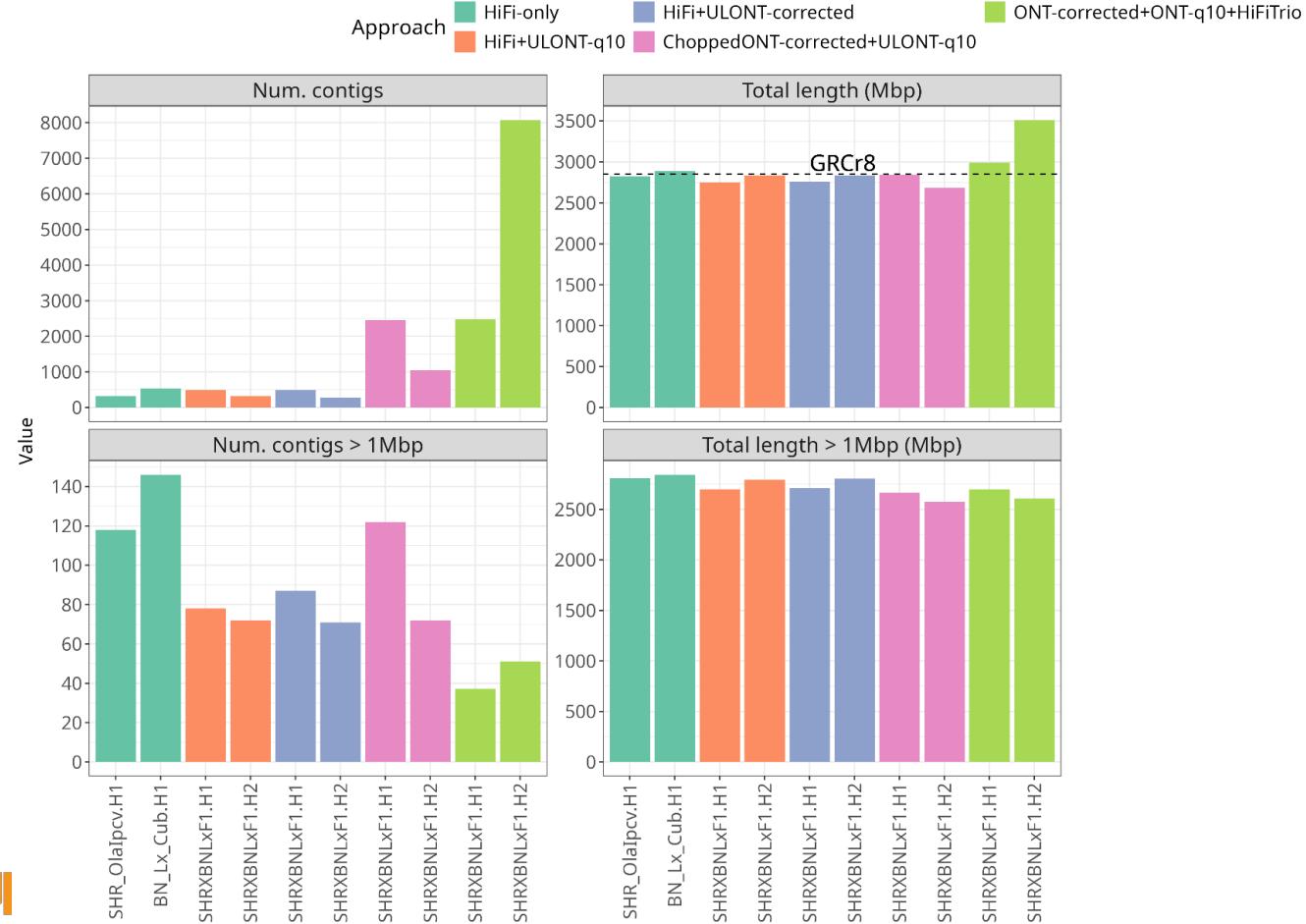
HiFi data generated in collaborations with
Abraham Palmer (NIDA P50)
Robert W. Williams (NIDA P30)

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

Assembly contiguity



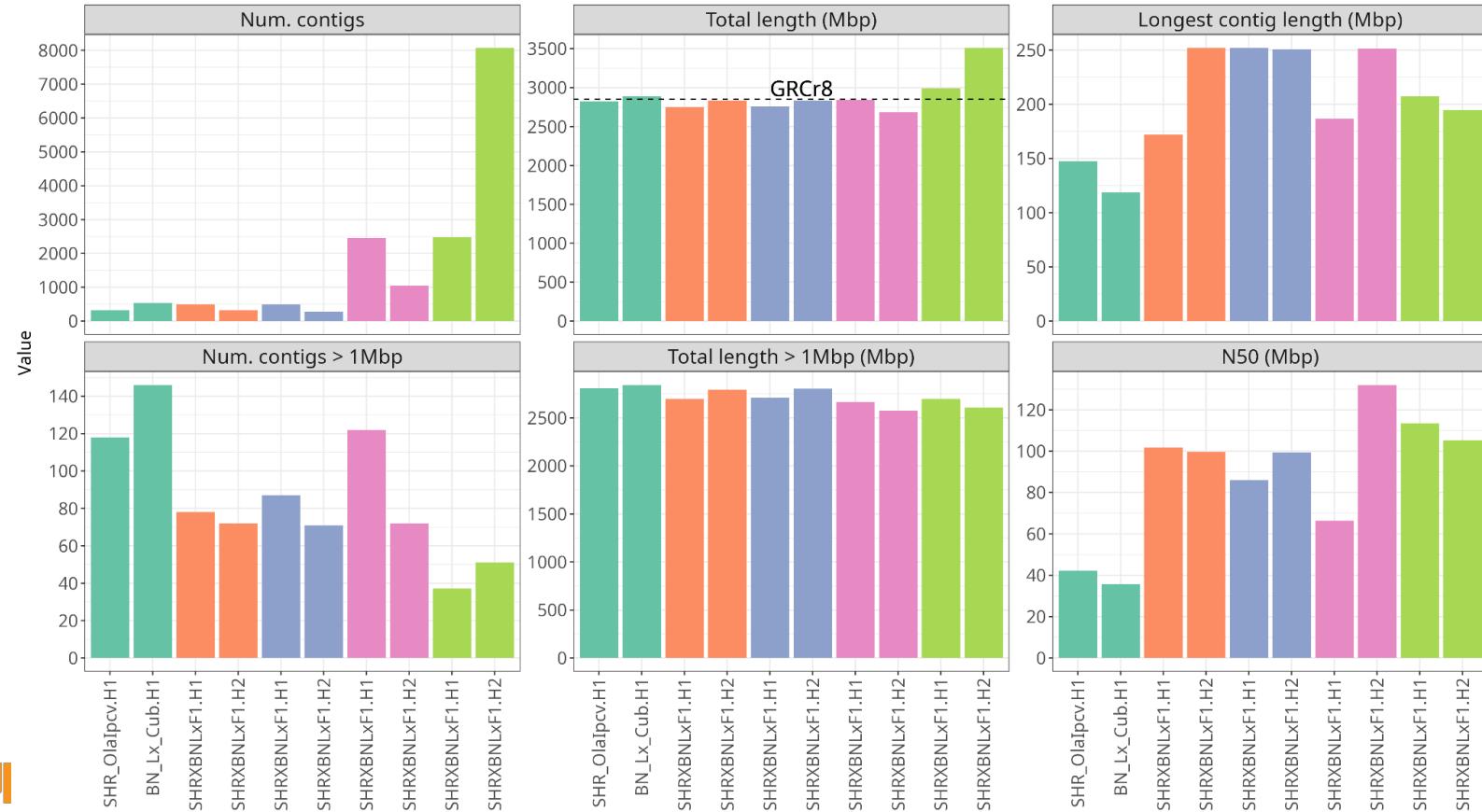
Assembly contiguity



Assembly contiguity

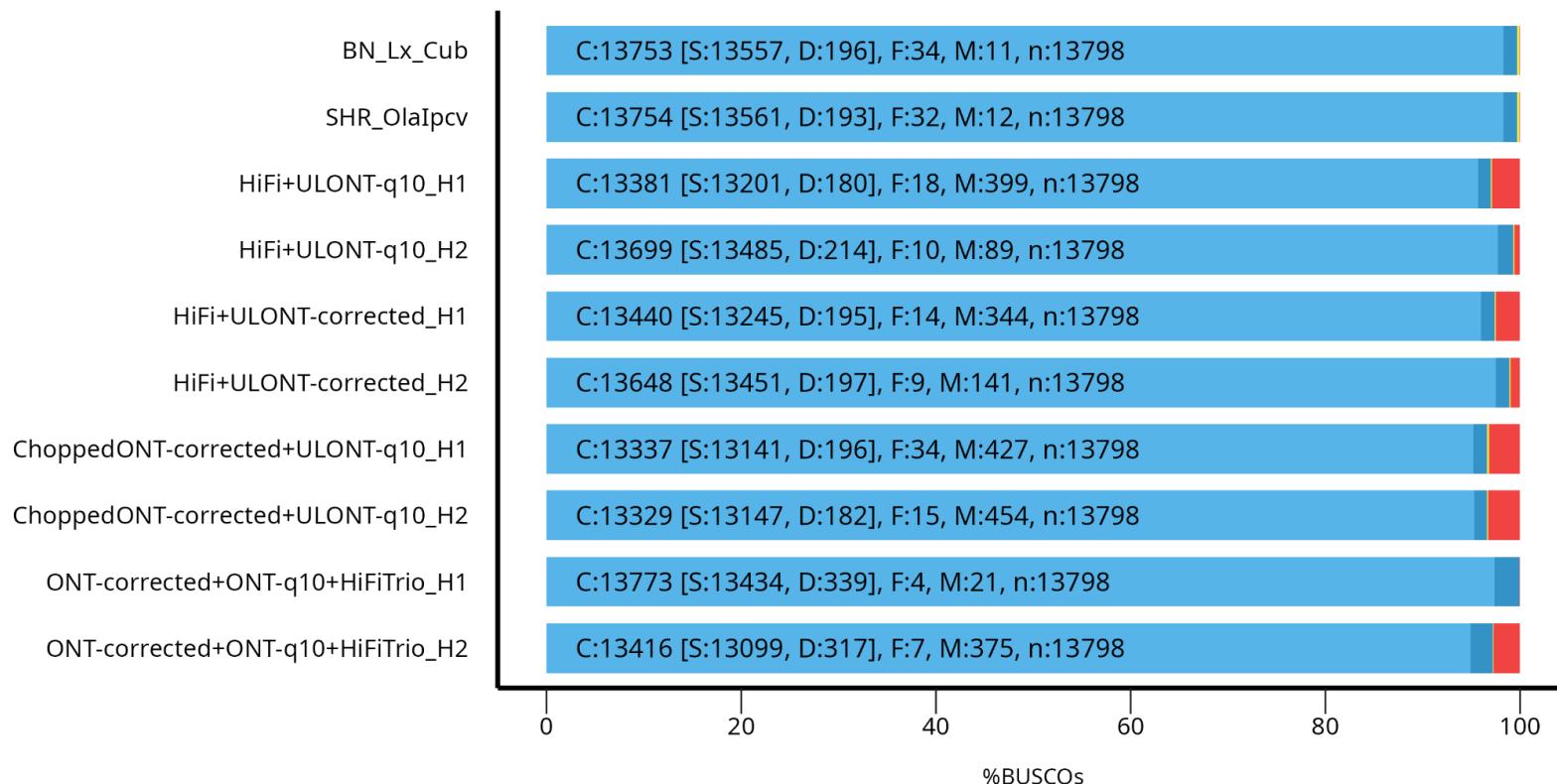
Approach

- HiFi-only
- HiFi+ULONT-corrected
- ONT-corrected+ONT-q10+HiFiTrio
- HiFi+ULONT-q10
- ChoppedONT-corrected+ULONT-q10

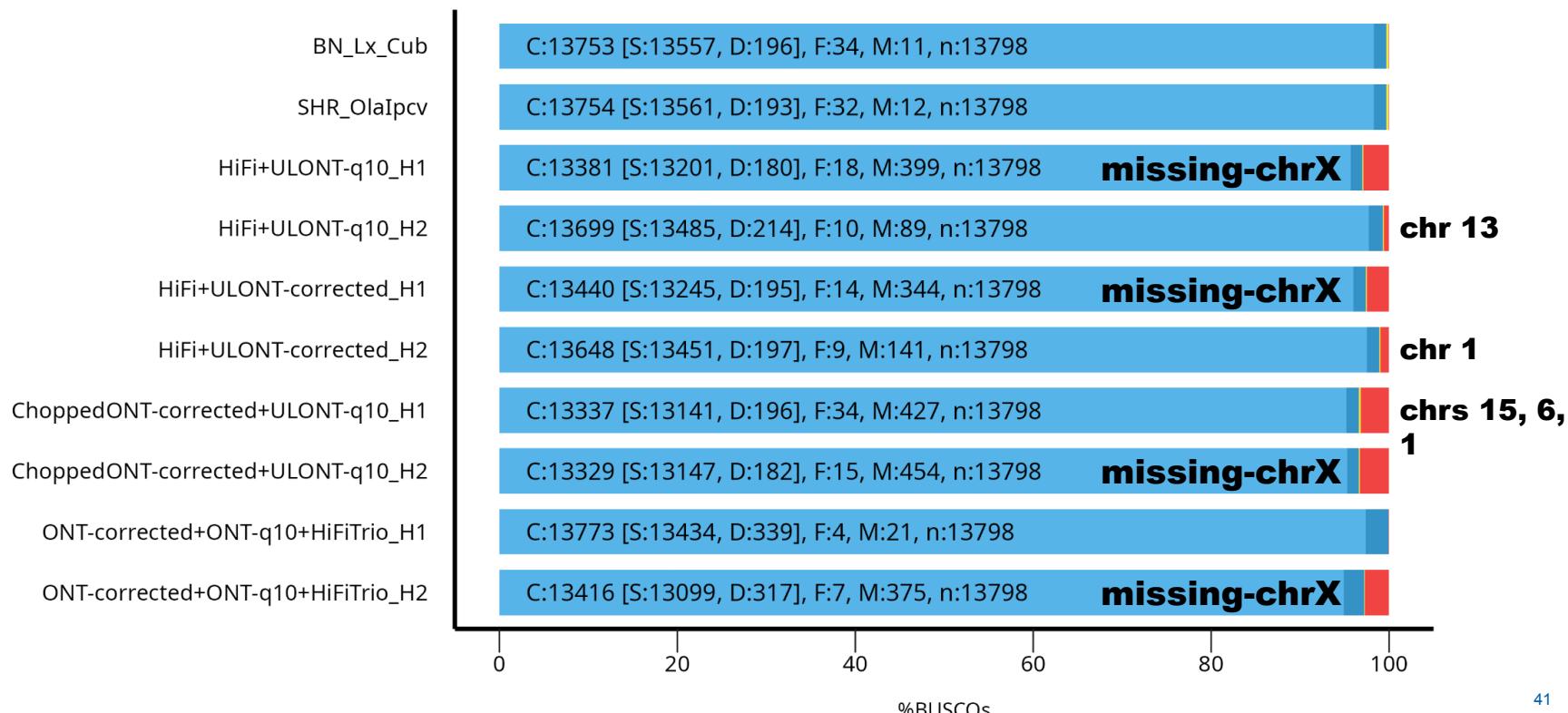


Glires' BUSCO Assessment Results

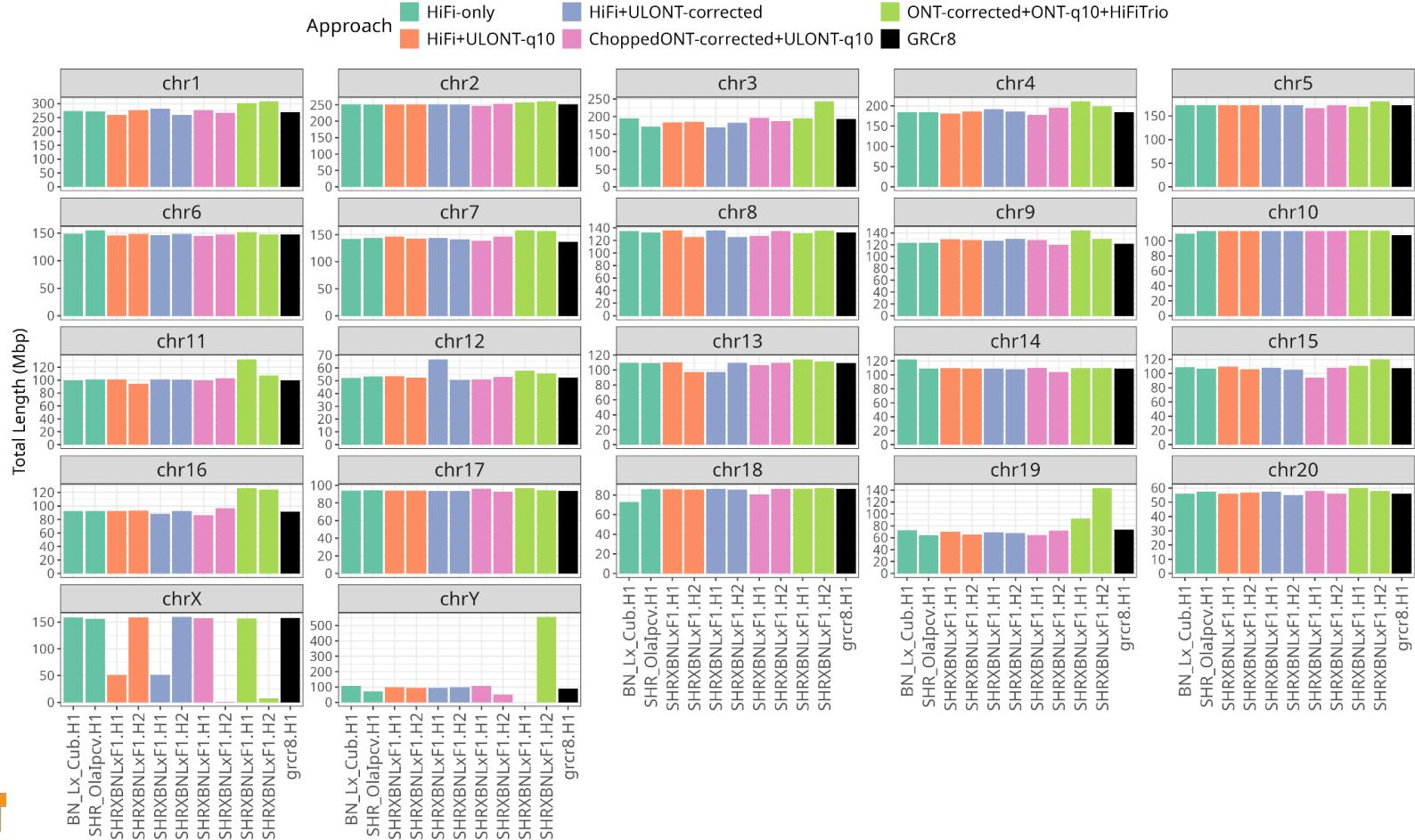
Complete (C) and single-copy (S) Complete (C) and duplicated (D)
Fragmented (F) Missing (M)



Glires' BUSCO Assessment Results



Partitioning by chromosome - Chromosome lengths



Partitioning by chromosome - Contig counts

