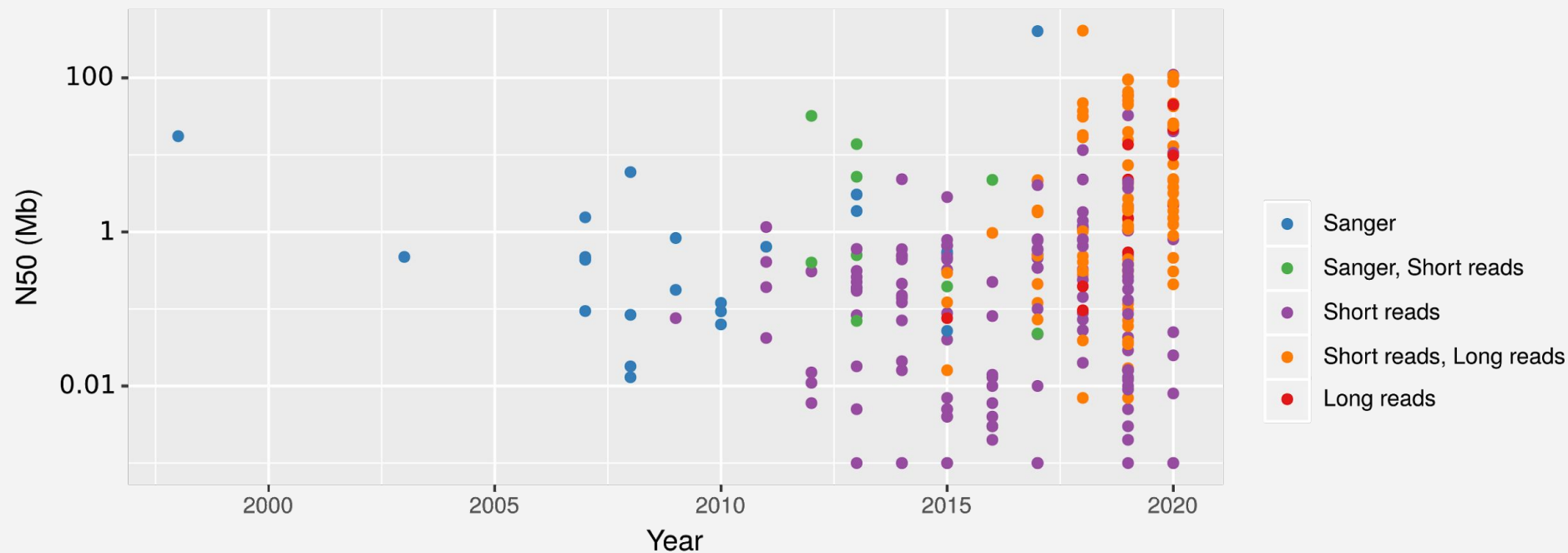


How to crack the genomes of non-model invertebrates: lessons from coral and rotifer genome projects

Nadège Guiglielmoni

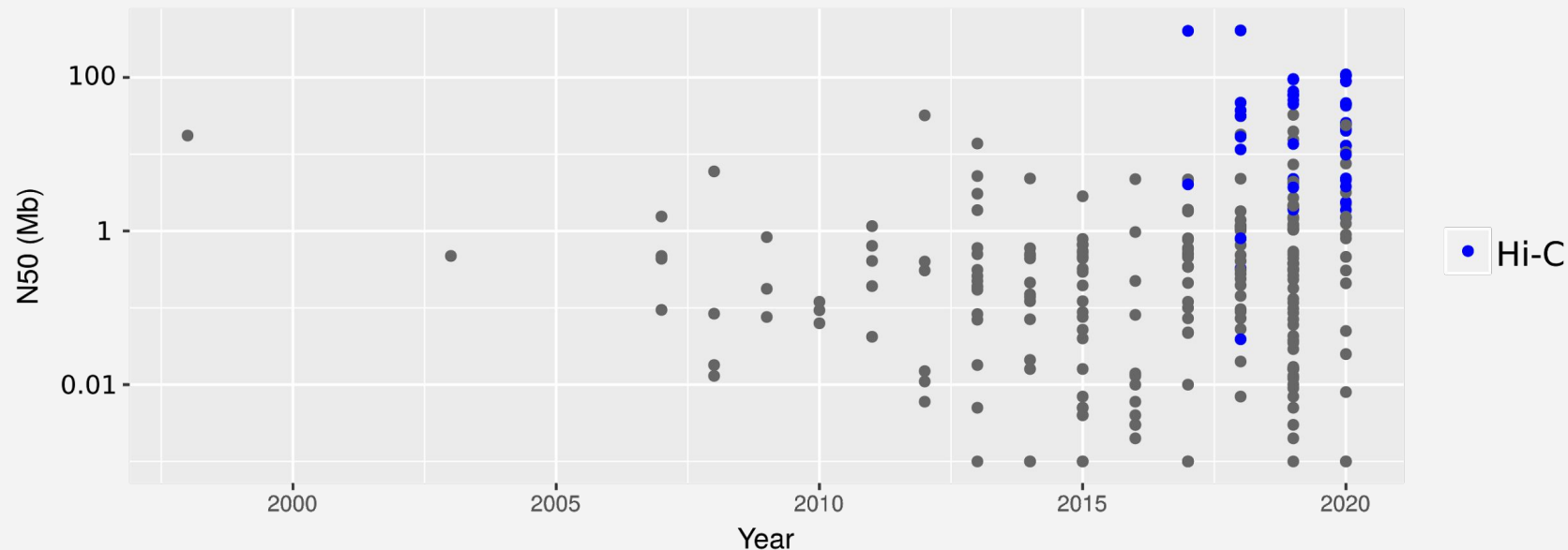
Assemblies of non-vertebrate genomes

Non-vertebrate animals assemblies



Assemblies of non-vertebrate genomes

Non-vertebrate animals assemblies



Two challenging genome assemblies

Two challenging genomes:

- **the coral *Astrangia poculata***: an undersized assembly
initial assembly: 50% of the expected size
- **the rotifer *Adineta vaga***: an oversized assembly
initial assembly: up to 160% of the expected size

Genome assembly of *Astrangia poculata*

Haploid genome size estimation: 453 Mb

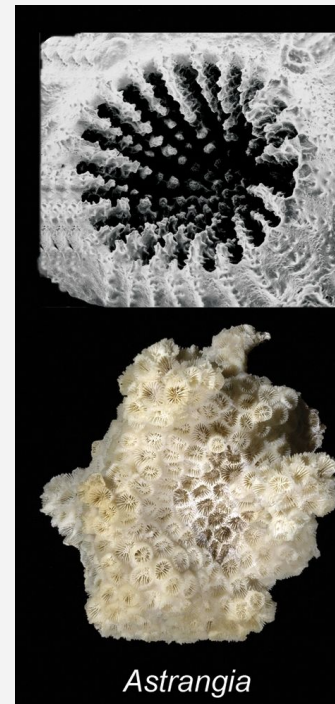
Assembly of short reads + CHICAGO + Hi-C: 252 Mb



Iliana Baums



Kathryn Stankiewicz

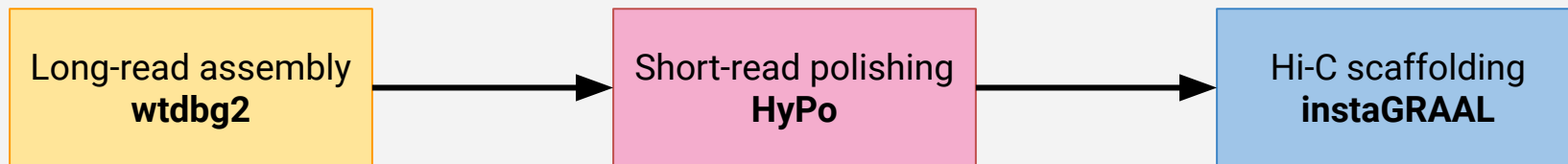


Picture: Cairns S, Kitahara M (2012) An illustrated key to the genera and subgenera of the Recent azooxanthellate Scleractinia (Cnidaria, Anthozoa)

Genome assembly of *Astrangia poculata*

Datasets :

- **Illumina** reads → 430X
- **Nanopore** reads → 15X
- **Hi-C** → 721 million pairs

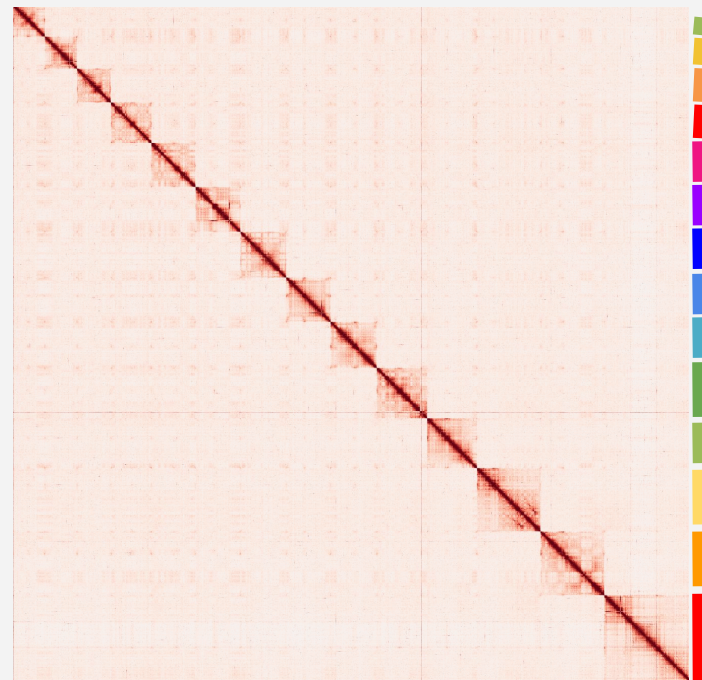


Genome assembly of *Astrangia poculata*

14 scaffolds

455 Mb

BUSCO completeness: 90.4%



Hi-C contact map of *Astrangia poculata*

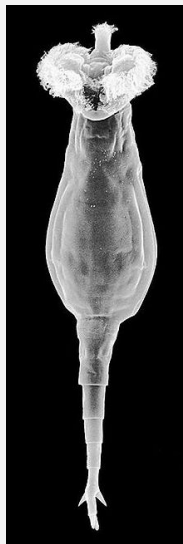
Genome assembly of *Astrangia poculata*

	Old assembly	New assembly
Assembly size (Mb)	252	455
# scaffolds	7848	14
BUSCO completeness (%)	60.2	90.4

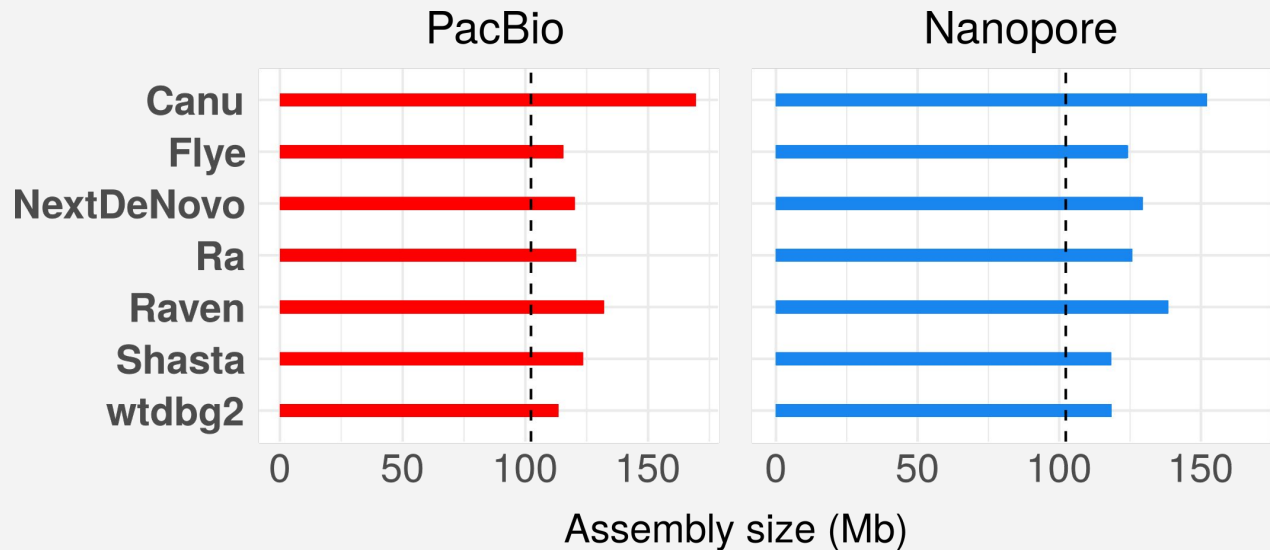
→ 15X of Nanopore reads resolved regions that were not solved by 430X Illumina

Genome assembly of *Adineta vaga*

Expected haploid size 102 Mb



Who Needs Sex (or Males) Anyway?
Liza Gross, PLoS Biology, 2007



Genome assembly of *Adineta vaga*

Haplotype 1 ATTACCAGTCTCAAT**GGATGGCTACTCTTTGACGATAGCT**

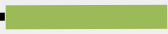
Haplotype 2 ATTACCAGTCTCAA**AGGCTGCTAGTGTTTGACGATAGCT**

Assembly process




Assembly output

Good haploid assemblies

contig 1 ———  ——— ✓

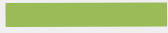
OR


contig 1 ———  ——— ✓

Problematic assembly

contig 1 ———

contig 2 ———

contig 3  X

contig 4 

Genome assembly of *Adineta vaga*

→ **Strategy 1:** choose a better assembler

Assemblers: Canu, Flye, NextDeNovo, Ra, Raven, Shasta, wtdbg2

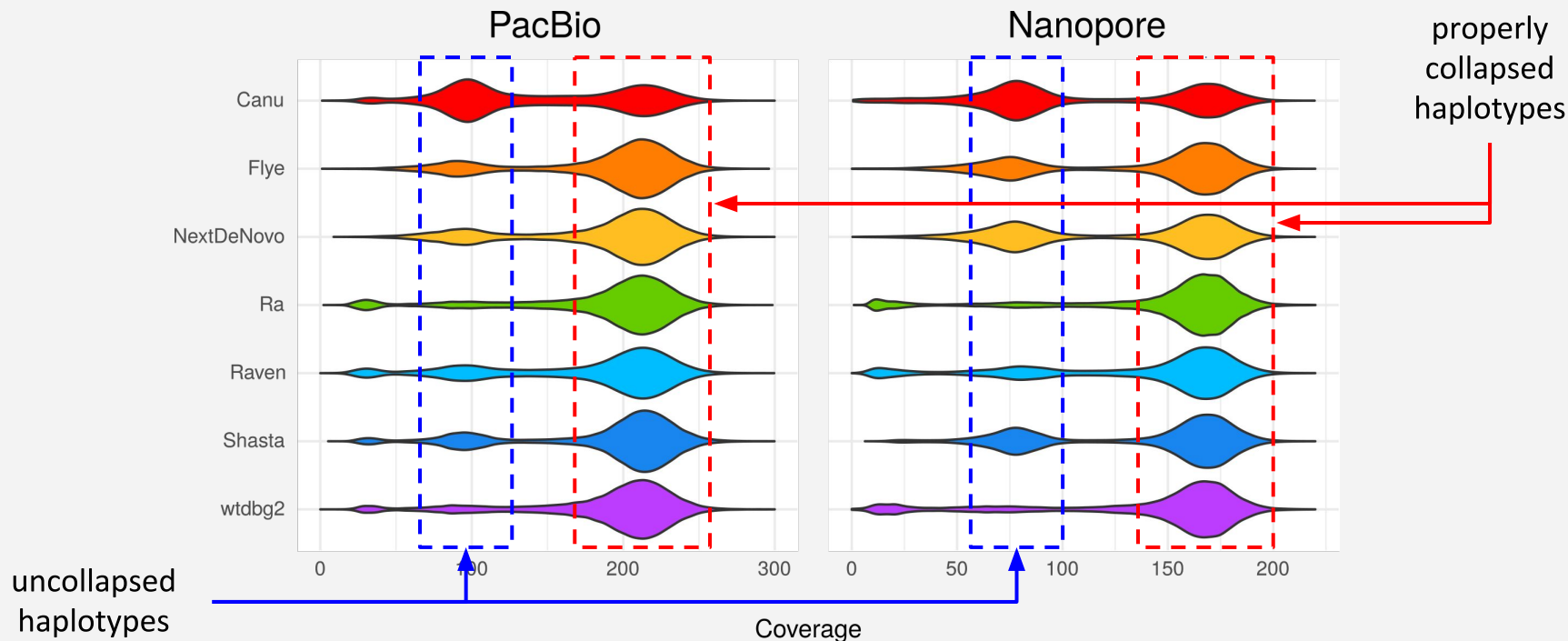
→ **Strategy 2:** select longest reads for assembly

→ **Strategy 3:** removing uncollapsed haplotypes

Tools: HaploMerger2, purge_dups, purge_haplotigs

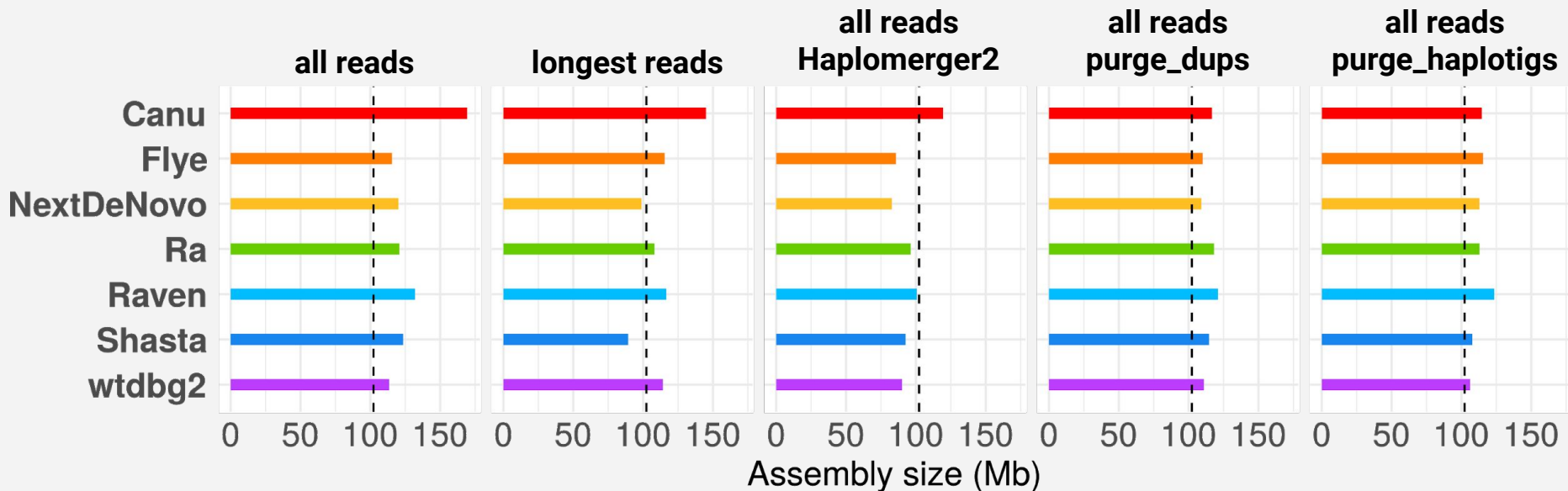
2 long-read datasets for <i>Adineta vaga</i> :	PacBio	23.5 Gb, 230X
	Nanopore	17.5 Gb, 171X

Genome assembly of *Adineta vaga*



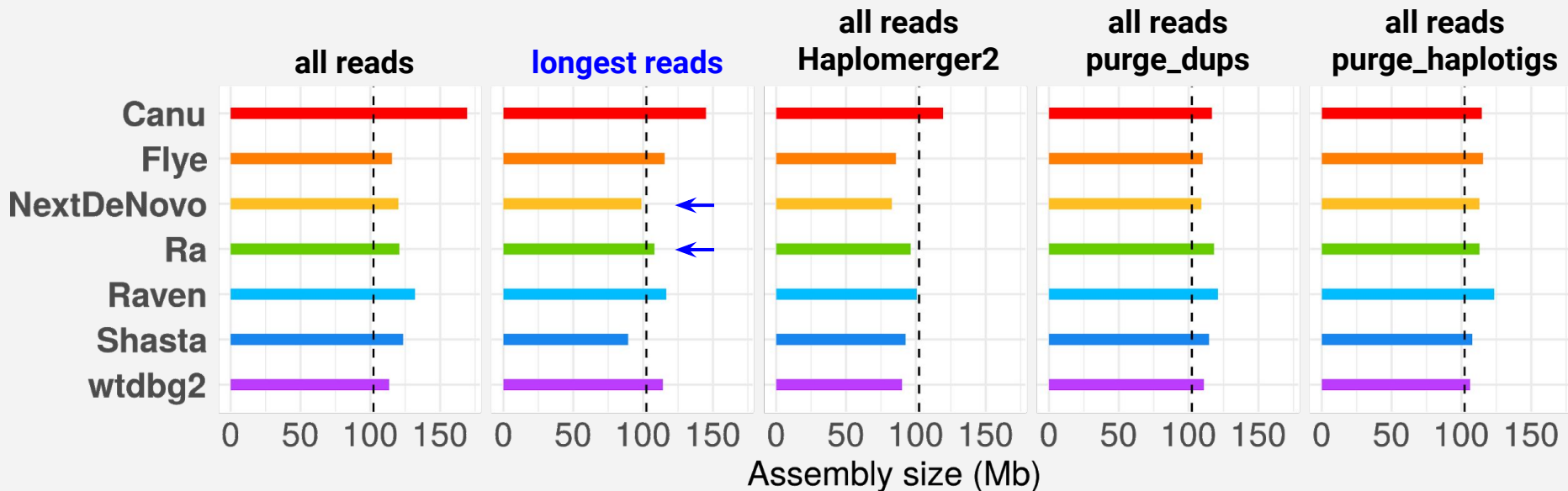
Genome assembly of *Adineta vaga*: PacBio assemblies

Assembly size: compared to the estimated size of 102 Mb



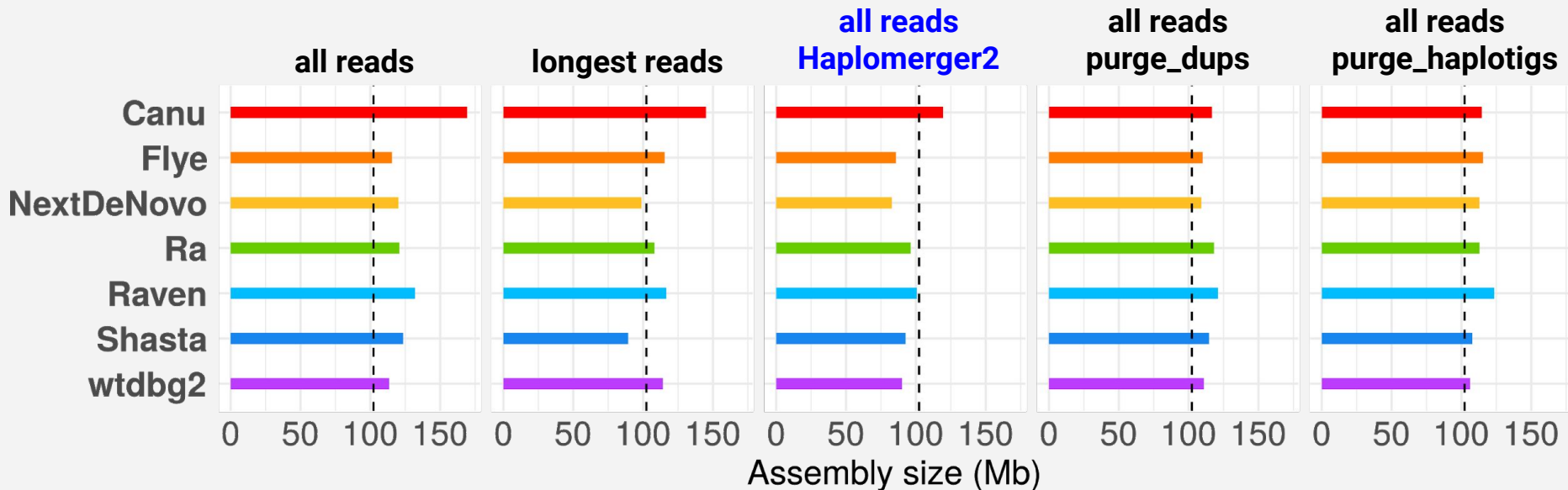
Genome assembly of *Adineta vaga*: PacBio assemblies

Assembly size: compared to the estimated size of 102 Mb



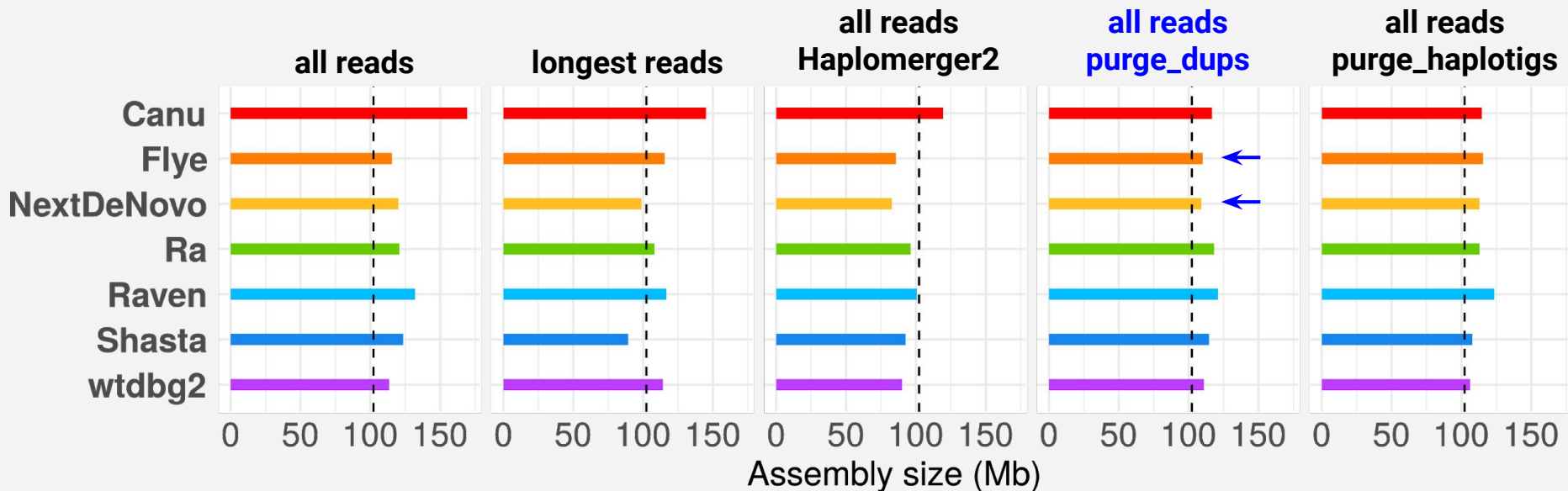
Genome assembly of *Adineta vaga*: PacBio assemblies

Assembly size: compared to the estimated size of 102 Mb



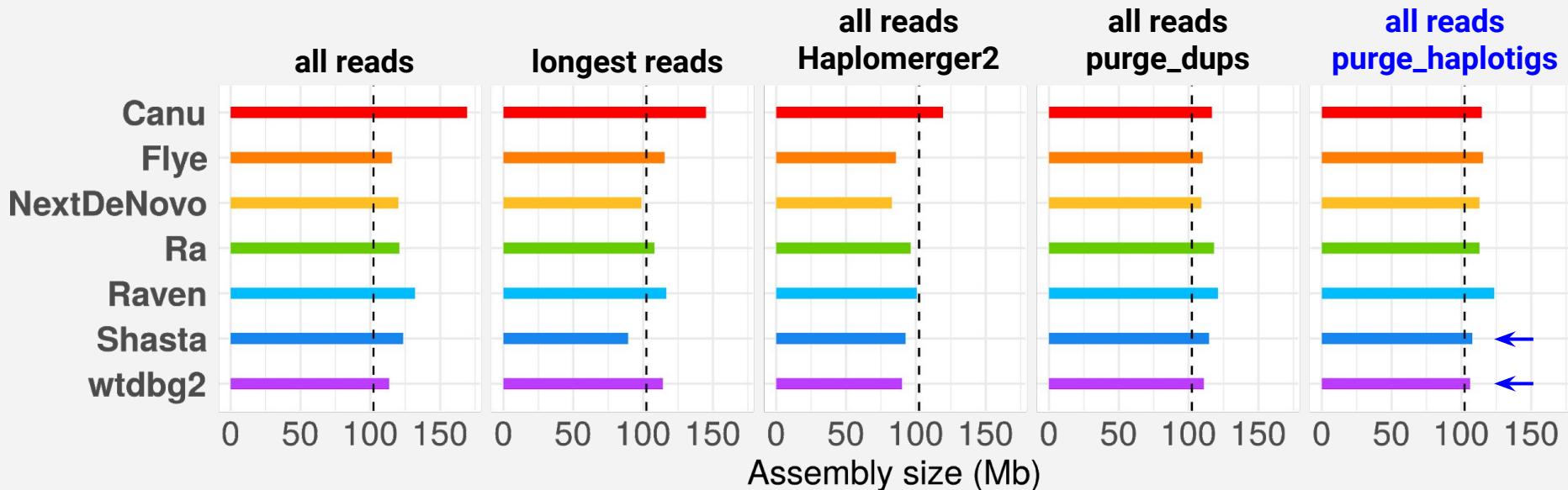
Genome assembly of *Adineta vaga*: PacBio assemblies

Assembly size: compared to the estimated size of 102 Mb



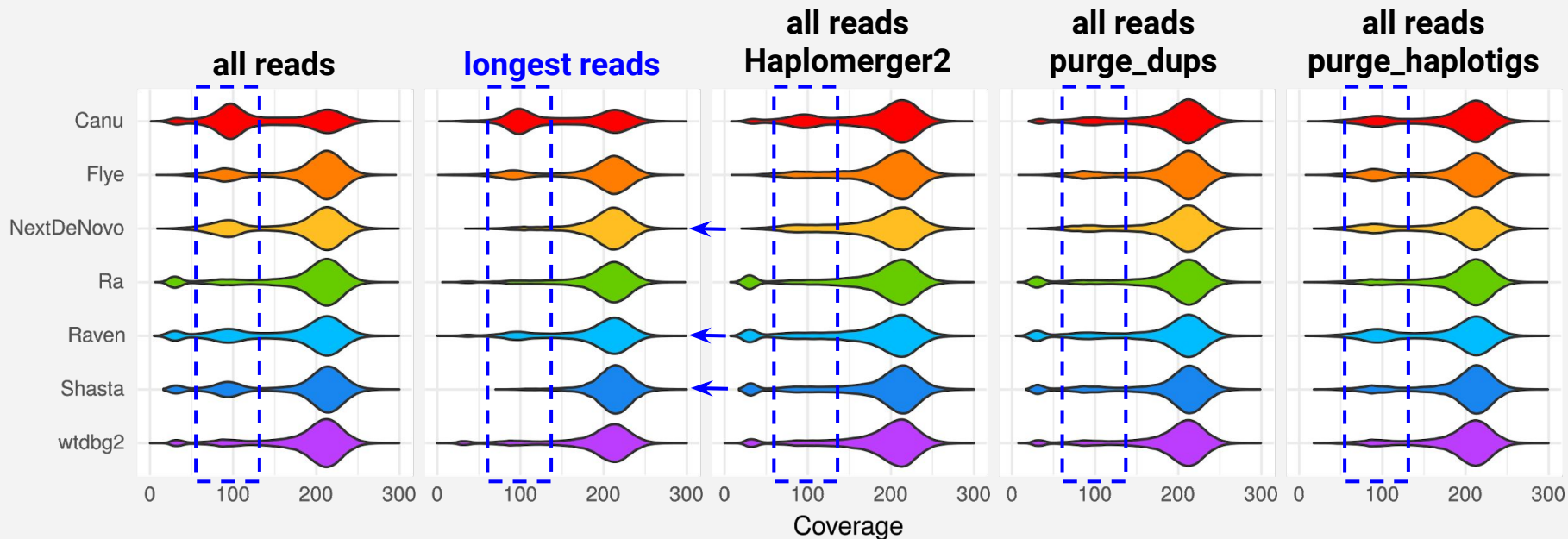
Genome assembly of *Adineta vaga*: PacBio assemblies

Assembly size: compared to the estimated size of 102 Mb



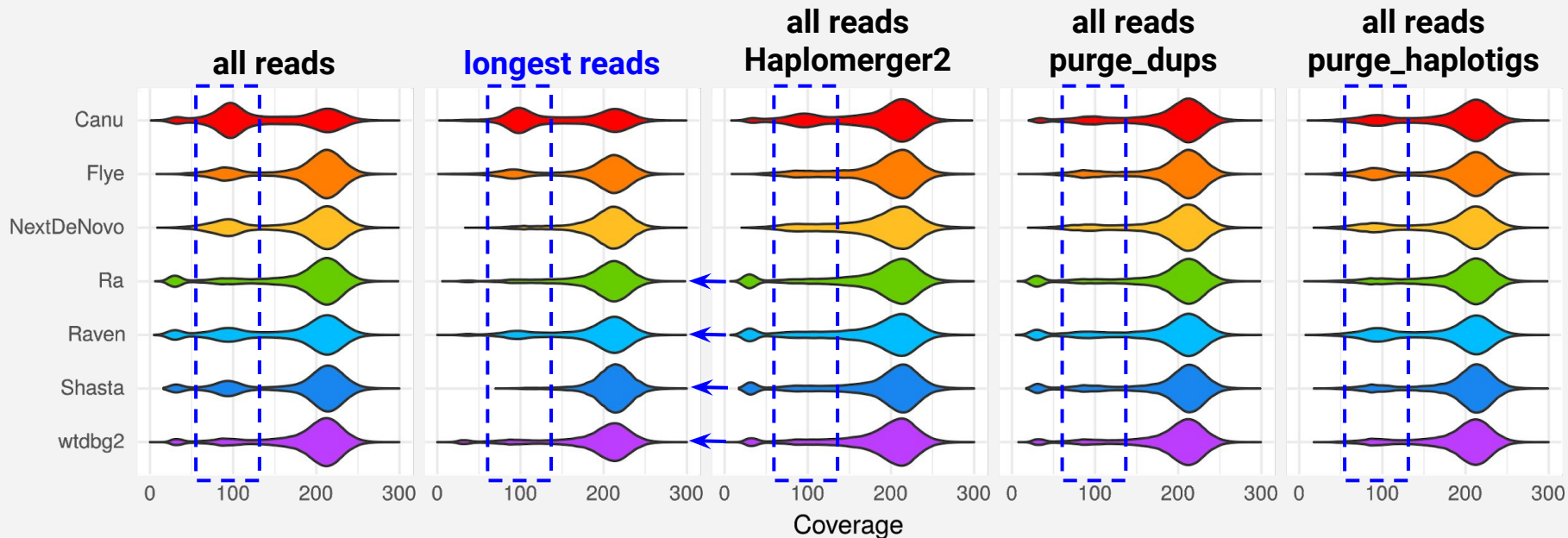
Genome assembly of *Adineta vaga*: PacBio assemblies

Coverage distribution: should only have one high-coverage peak



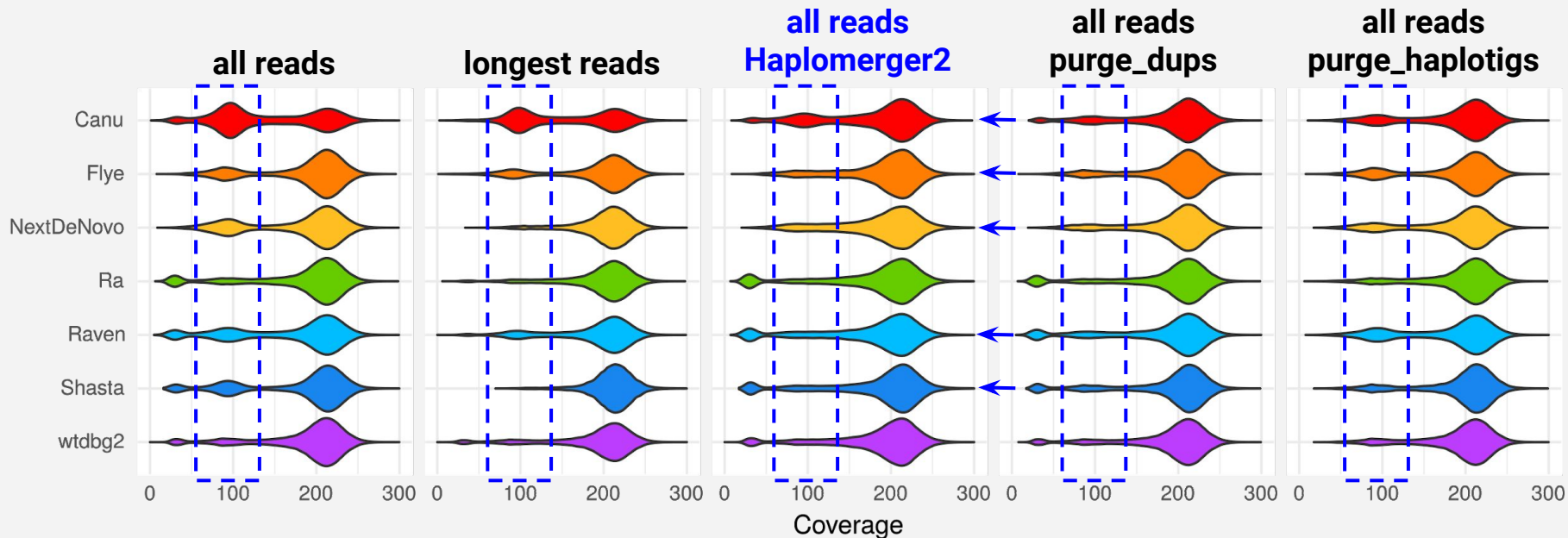
Genome assembly of *Adineta vaga*: PacBio assemblies

Coverage distribution: should only have one high-coverage peak



Genome assembly of *Adineta vaga*: PacBio assemblies

Coverage distribution: should only have one high-coverage peak



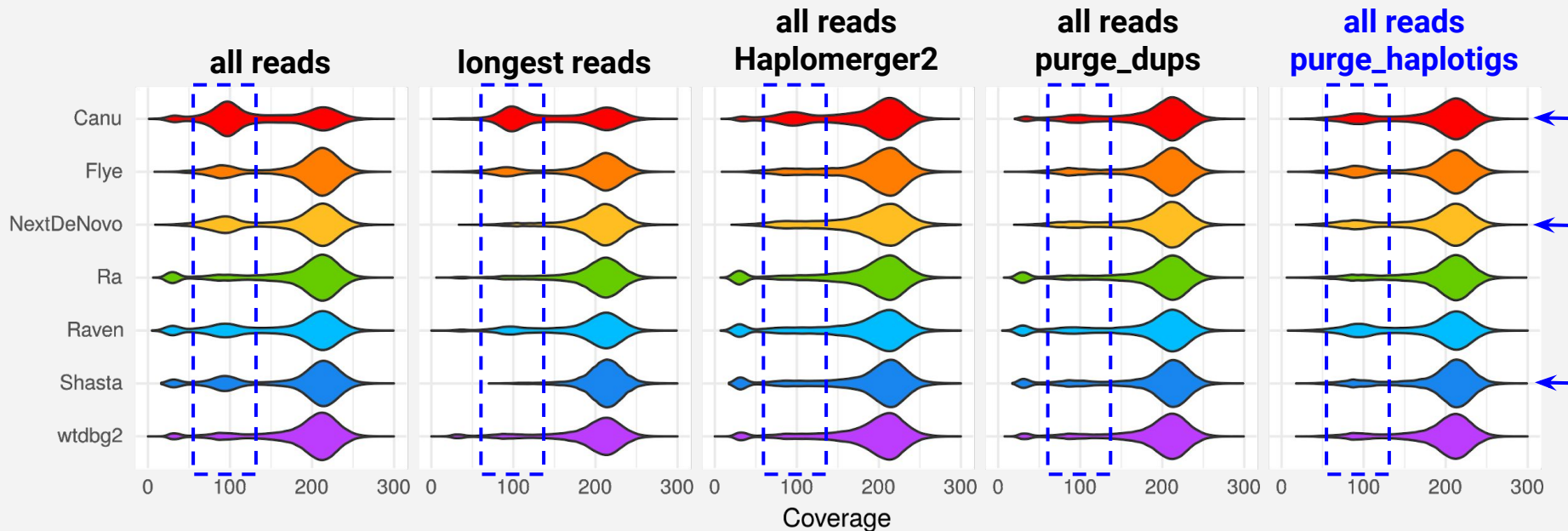
Genome assembly of *Adineta vaga*: PacBio assemblies

Coverage distribution: should only have one high-coverage peak



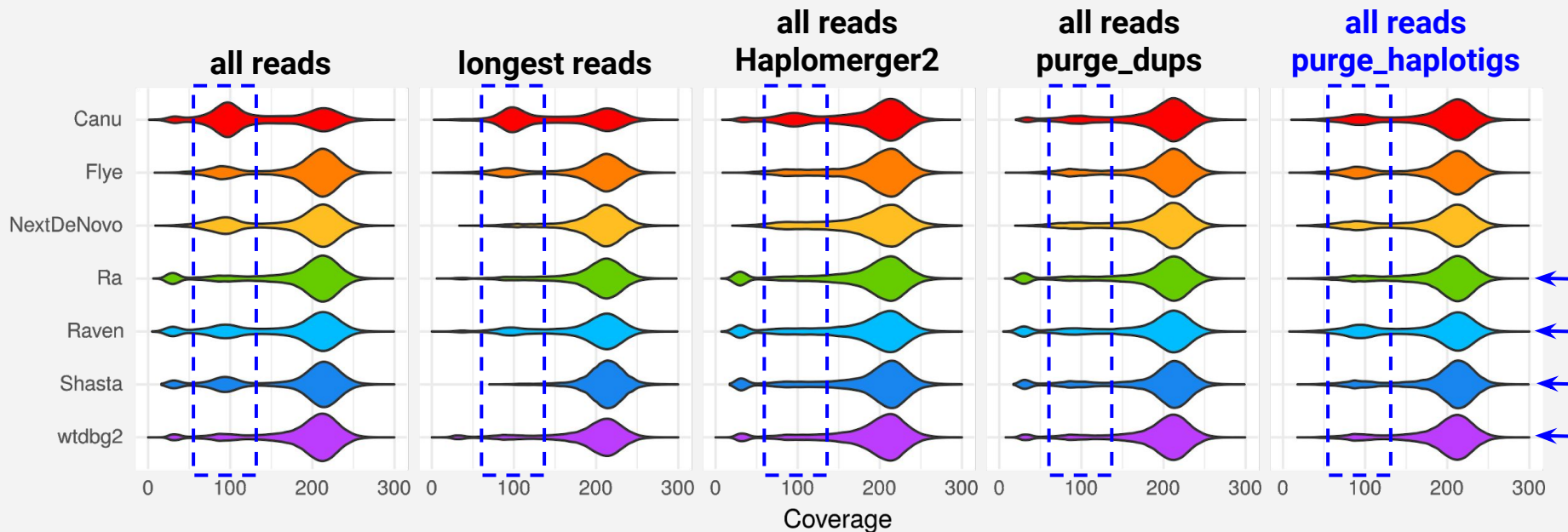
Genome assembly of *Adineta vaga*: PacBio assemblies

Coverage distribution: should only have one high-coverage peak



Genome assembly of *Adineta vaga*: PacBio assemblies

Coverage distribution: should only have one high-coverage peak



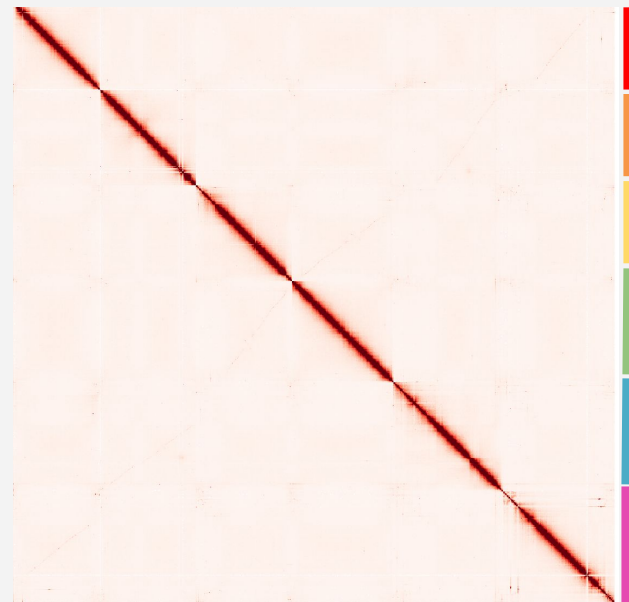
Genome assembly of *Adineta vaga*

- **Strategy 1:** some assemblers are better at collapsing haplotypes (Ra, wtdbg2)
- **Strategy 2:** read filtering can improve structure
- **Strategy 3:** haplotigs-purging tools work better combined with specific assemblers
ex: Flye + purge_dups, wtdbg2 + purge_haplotigs
- Haplotigs-purging tools can be combined for better results

Genome assembly of *Adineta vaga*

Example of Hi-C scaffolding
with **instaGRAAL**

6 chromosome-level scaffolds



Longest reads + Ra

Perspectives

→ **Long reads** can efficiently build structurally **correct haploid assemblies**

Perspectives

- **Long reads** can efficiently build structurally **correct haploid assemblies**
- **Hi-C scaffolding** is a promising technique to **obtain chromosome-level scaffolds for non-model invertebrates**

Perspectives

- **Long reads** can efficiently build structurally **correct haploid assemblies**
- **Hi-C scaffolding** is a promising technique to **obtain chromosome-level scaffolds for non-model invertebrates**
- Strategies applied to other genomes projects in IGNITE:
 - the mollusk ***Arion vulgaris***, project of Zeyuan Chen
 - 26 chromosome-level scaffolds**
 - the chaetognath ***Flaccisagitta enflata***
 - 9 chromosome-level scaffolds**

Acknowledgements

Université libre de Bruxelles

Jean-François Flot

Karine van Doninck

Alessandro Derzelle

Antoine Houtain

Paul Simion

PennState University

Iliana Baums

Kathryn Stankiewicz

Ludwig Maximilian Universität

Ramón E. Rivera-Vicéns



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764840

Thank you for your attention!

Questions?

<https://github.com/nadegeguiglielmoni/presentations>