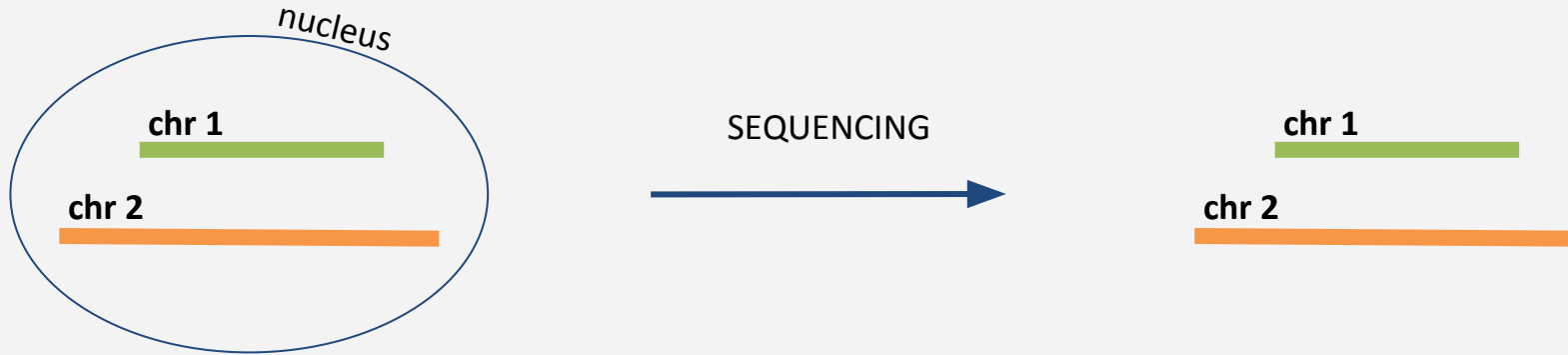


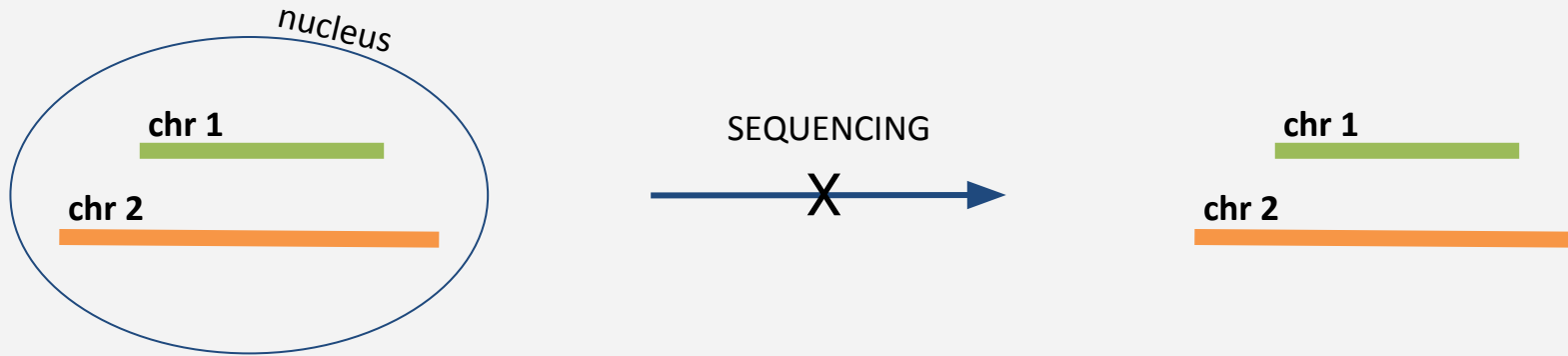
hic2gfa: phasing assembly graphs with chromosome conformation capture

Nadège Guiguelmoni

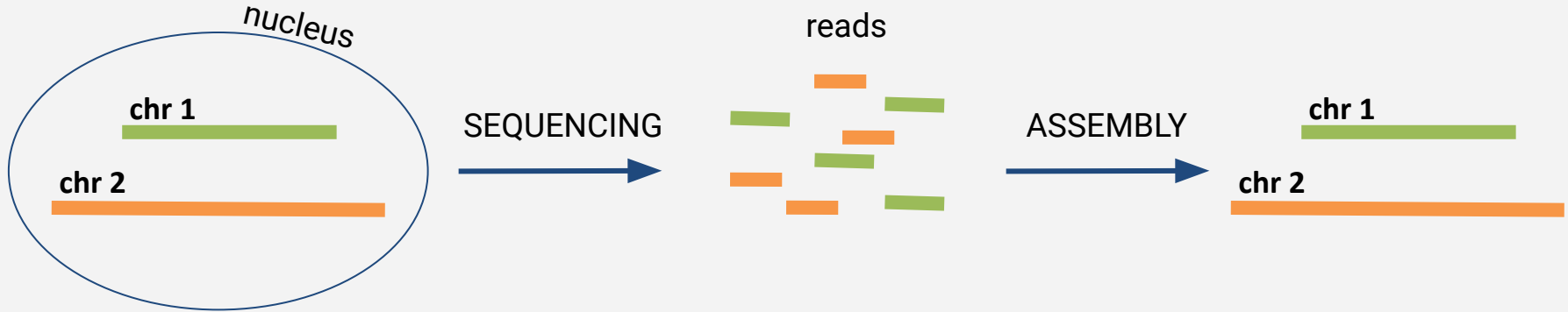
Genome assembly



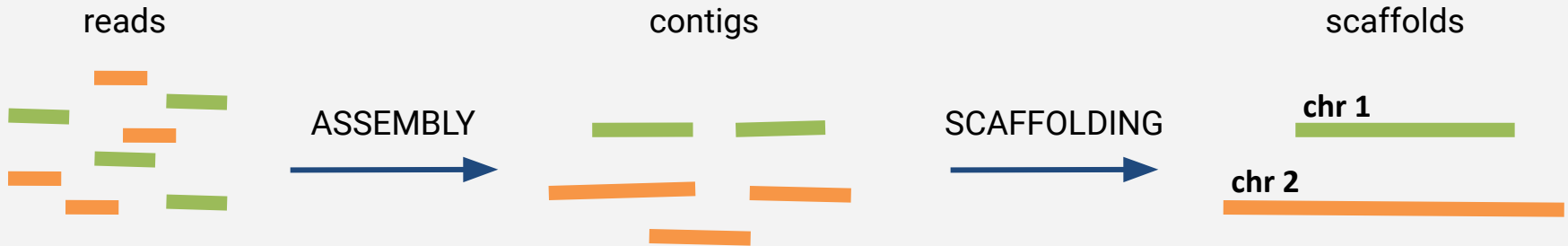
Genome assembly



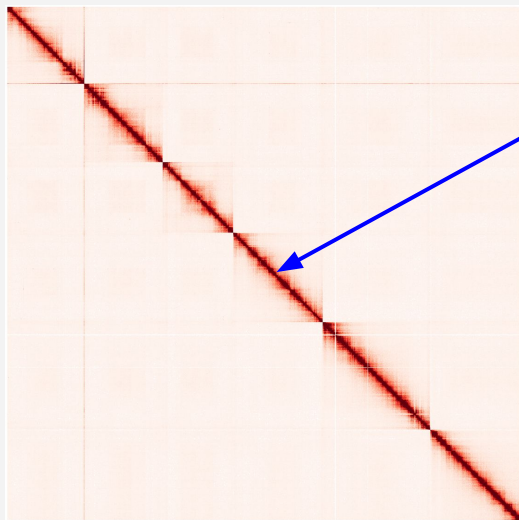
Genome assembly



Genome assembly

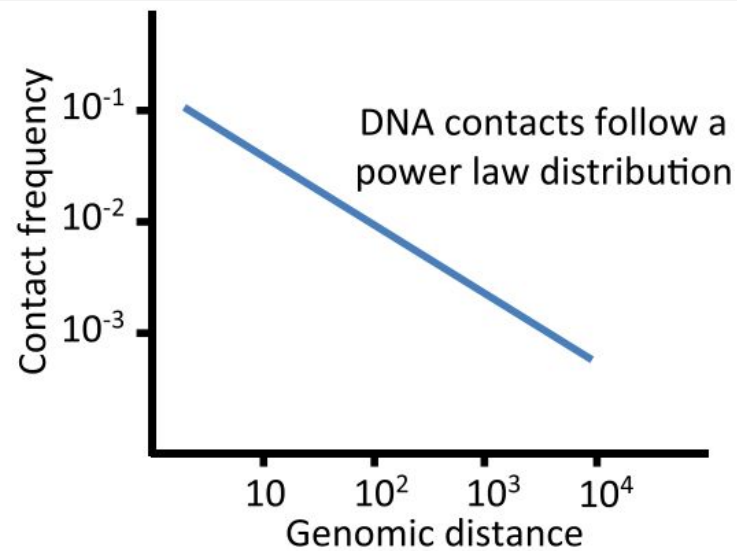


Hi-C scaffolding



Contact map of
Caenorhabditis elegans

contact frequency = $f(\text{genomic distance})$



Hi-C scaffolding

High-throughput genome scaffolding from *in vivo* DNA interaction frequency

Noam Kaplan ✉ & Job Dekker ✉

dnaTri

Lachesis

Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions

Joshua N Burton ✉, Andrew Adey, Rupali P Patwardhan, Ruolan Qiu, Jacob O Kitzman & Jay Shendure

High-quality genome (re)assembly using chromosomal contact data

Hervé Marie-Nelly ✉, Martial Marbouty, Axel Cournac, Jean-François Flot, Gianni Liti, Dante Poggi Parodi, Sylvie Syan, Nancy Guillén, Antoine Margeot, Christophe Zimmer ✉ & Romain Koszul ✉

GRAAL

Hi-C scaffolding

De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds

Olga Dudchenko^{1,2,3,4}, Sanjit S. Batra^{1,2,3,*}, Arina D. Omer^{1,2,3,*}, Sarah K. Nyquist^{1,3}, Marie Hoeger^{1,3}, Neva C. Durand^{1,...}

3D-DNA

SALSA2

Integrating Hi-C links with assembly graphs for chromosome-scale assembly

Jay Ghurye, Arang Rhie, Brian P. Walenz, Anthony Schmitt, Siddarth Selvaraj, Mihai Pop, Adam M. Phillippy, Sergey Koren

instaGRAAL: chromosome-level quality scaffolding of genomes using a proximity ligation-based scaffolder

[Lyam Baudry](#), [Nadège Guiguelmoni](#), [Hervé Marie-Nelly](#), [Alexandre Cormier](#), [Martial Marbouty](#), [Komlan Avia](#), [Yann Loe Mie](#), [Olivier Godfroy](#), [Lieven Sterck](#), [J. Mark Cock](#), [Christophe Zimmer](#), [Susana M. Coelho](#) & [Romain Koszul](#)

instaGRAAL

Limits to current assemblies

→ Scaffolds usually have gaps

Limits to current assemblies

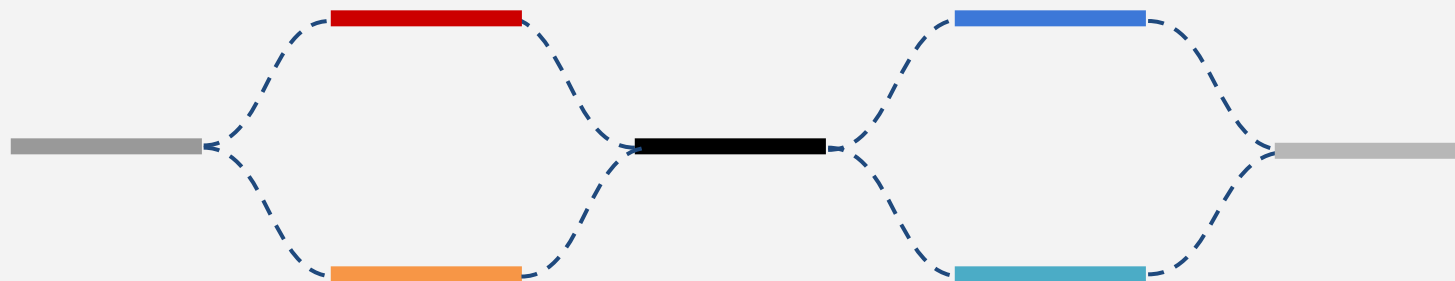
- Scaffolds usually have gaps
- Haploid representations of diploid/polyploid genomes
= partial representation
→ uncollapsed phased assemblies

Limits to current assemblies

- Scaffolds usually have gaps
- Haploid representations of diploid/polyploid genomes
= partial representation
→ uncollapsed phased assemblies
- Phased assembly = correctly associating alleles

Unzipping assembly graphs with Hi-C

Example of assembly graph for a diploid organism



— contig

- - - link

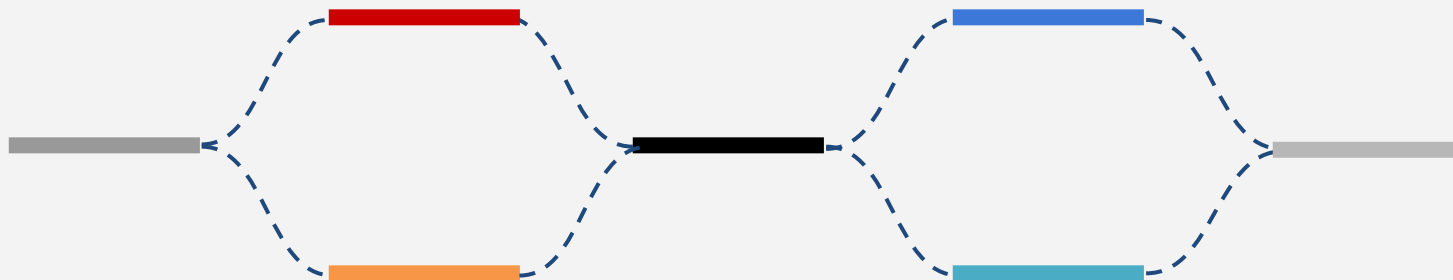
Unzipping assembly graphs with Hi-C

hic2gfa

- ★ input: assembly graph + Hi-C data
- ★ output: phased supercontigs

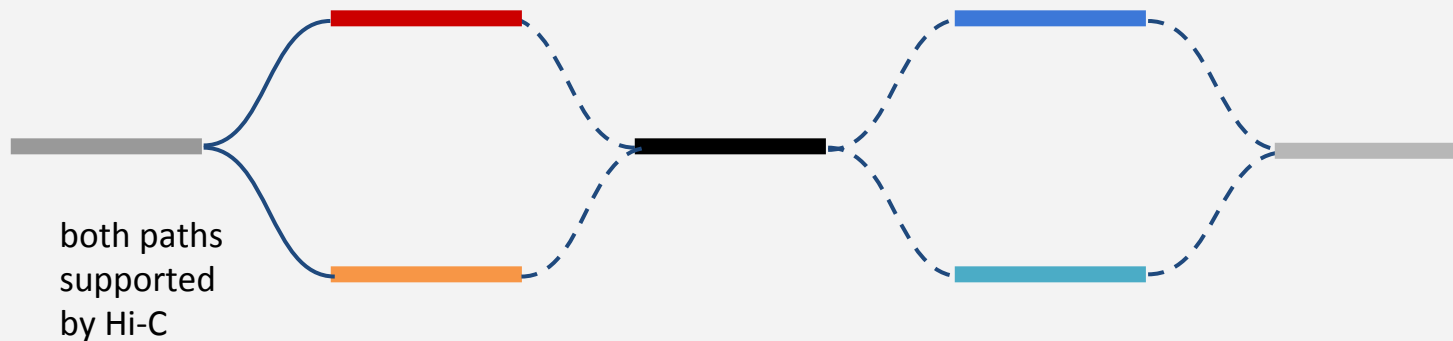
Unzipping assembly graphs with Hi-C

hic2gfa



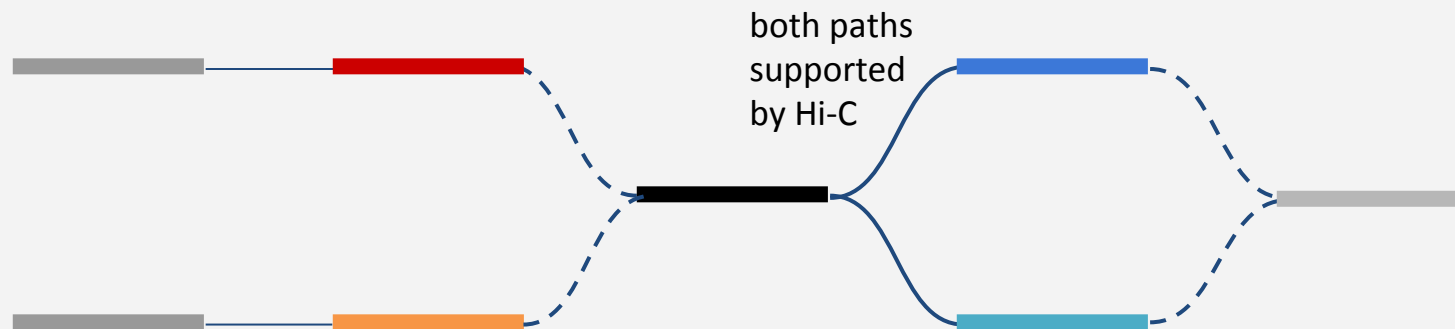
Unzipping assembly graphs with Hi-C

hic2gfa



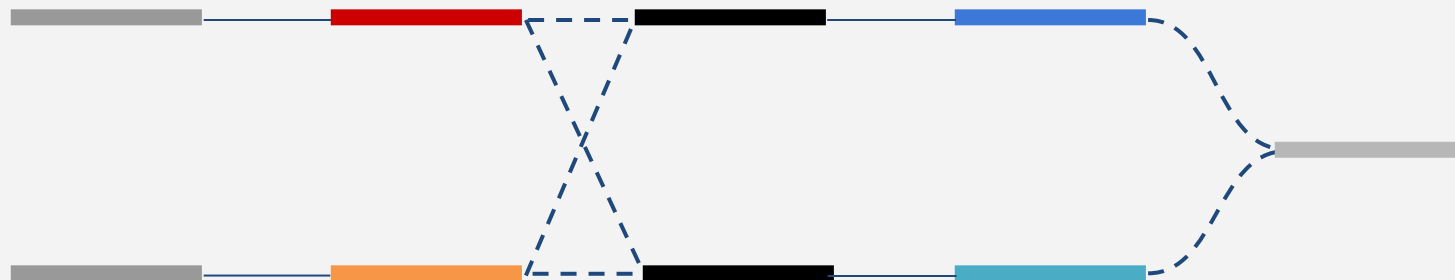
Unzipping assembly graphs with Hi-C

hic2gfa



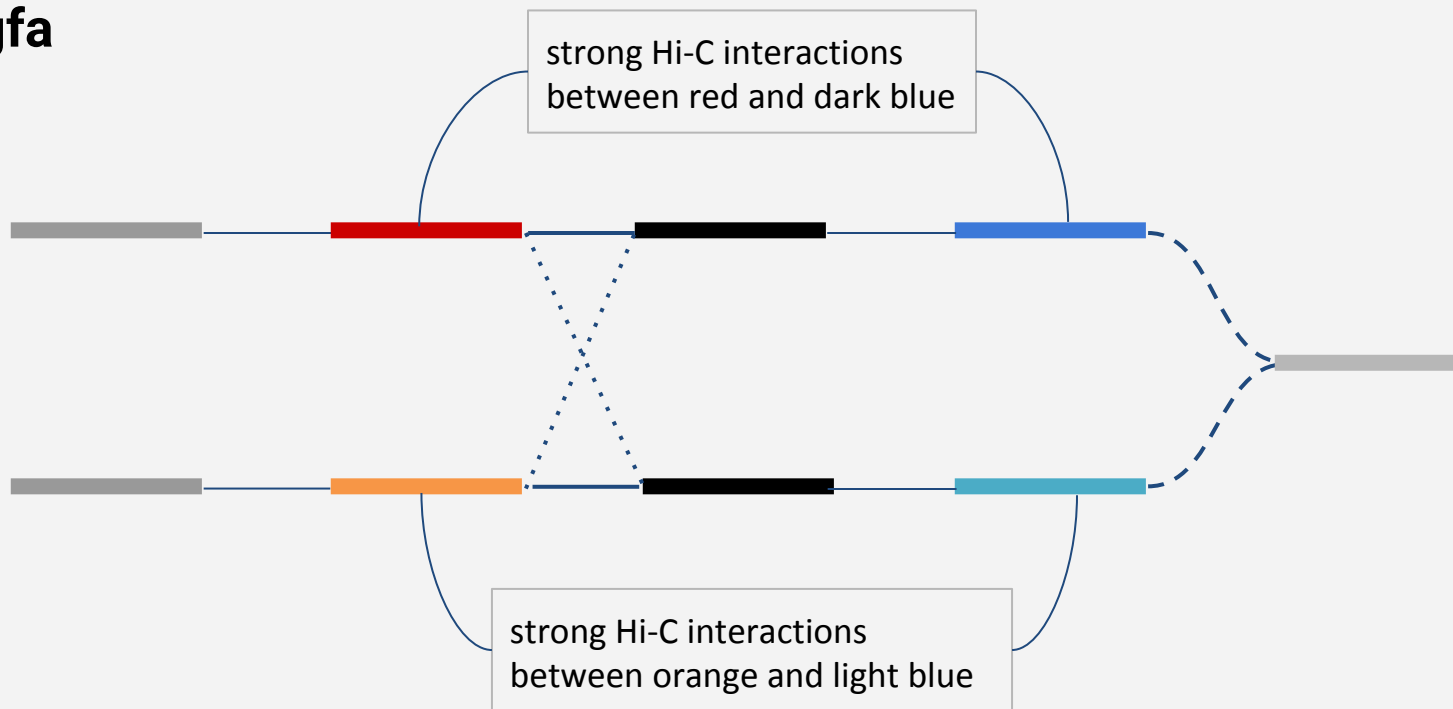
Unzipping assembly graphs with Hi-C

hic2gfa



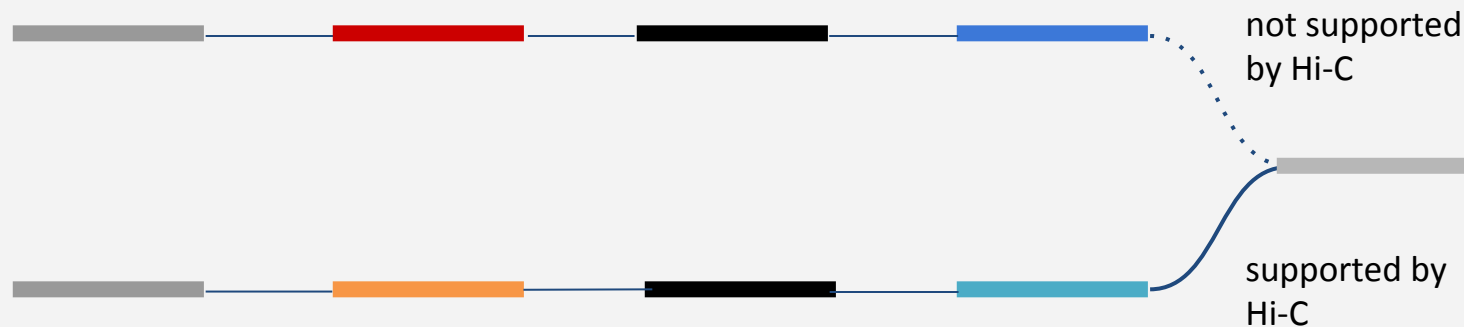
Unzipping assembly graphs with Hi-C

hic2gfa



Unzipping assembly graphs with Hi-C

hic2gfa



Unzipping assembly graphs with Hi-C

hic2gfa

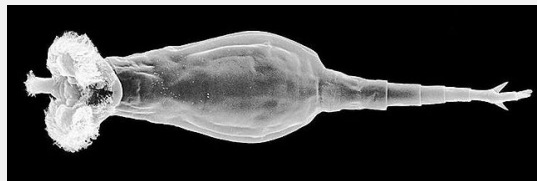


phased supercontigs



Unzipping assembly graphs with Hi-C

Adineta vaga

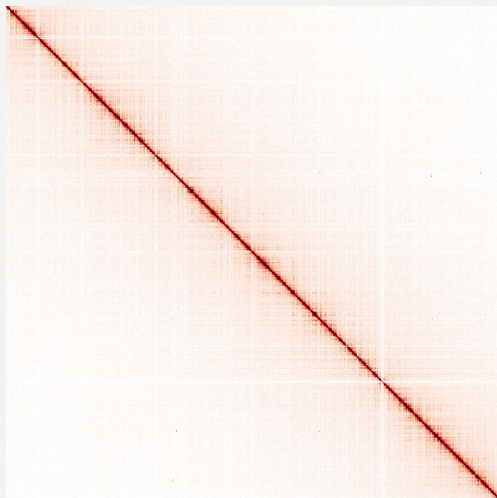


Reads	Contigs total length	Contig N50	Supercontigs total length	Supercontigs N50
Corrected PacBio	182 Mb	269 kb	205 Mb	4.2 Mb
Corrected Nanopore	199 Mb	4.0 Mb	209 Mb	6.3 Mb
HiFi	190 Mb	4.8 Mb	207 Mb	13.5 Mb

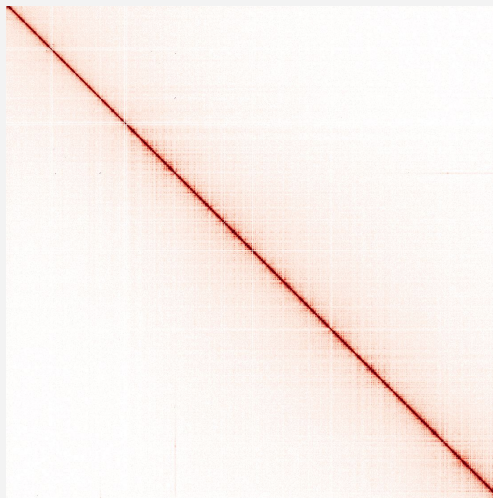
Unzipping assembly graphs with Hi-C

Corrected PacBio contigs + hic2gfa

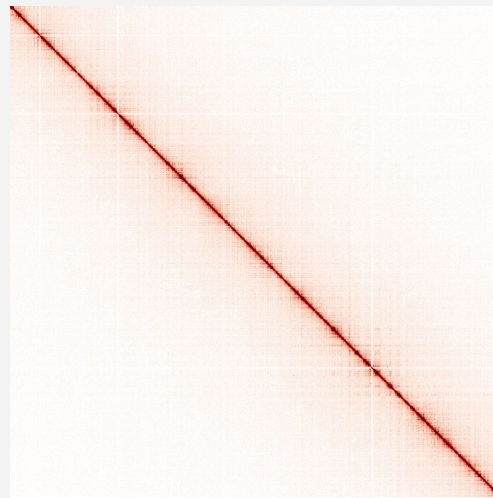
13 contigs



98 contigs



124 contigs



Unzipping assembly graphs with Hi-C

Further works

★ hic2gfa + long reads = GraphUnzip

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

★ Testing, testing, testing

Acknowledgements

EBE, Université libre de Bruxelles

Jean-François Flot

Roland Faure

Université de Namur

Karine van Doninck

Alessandro Derzelle

Antoine Houtain

Paul Simion

Looking for a post-doc from August 1st, 2021



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?