

Improving *de novo* assemblies with Hi-C

Nadège Guiglielmoni

About me

PhD student at ULB and in the ITN IGNITE www.itn-ignite.eu

euraxess.ec.europa.eu/jobs/search

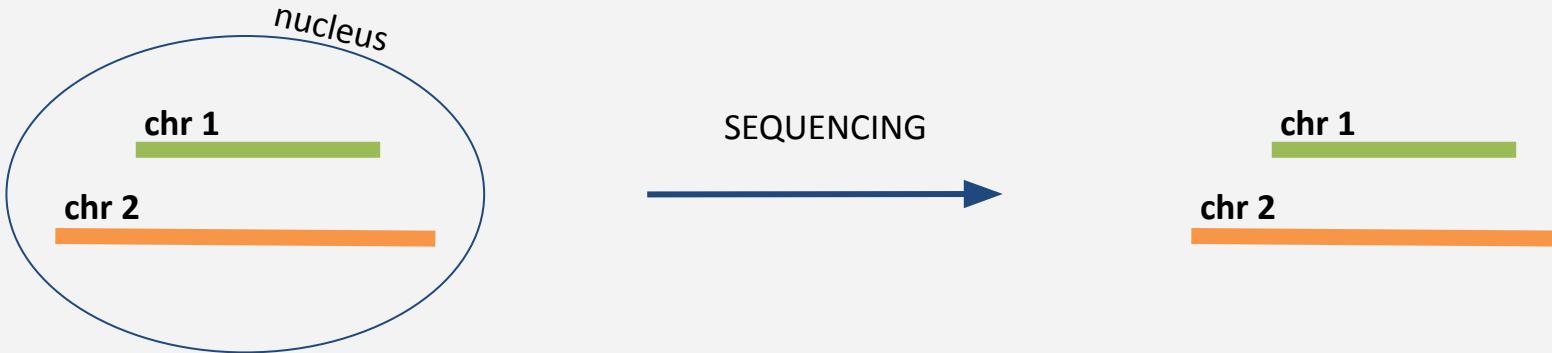
- Long-read assembly
- Hi-C scaffolding
- Genomics of non-vertebrates

nadege.guiglielmoni@ulb.be

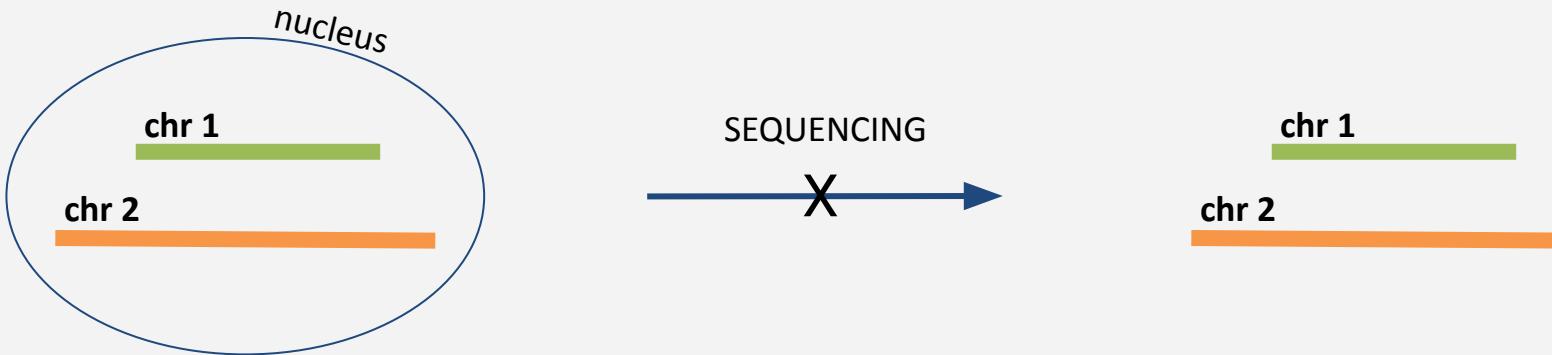
@NGuiglielmoni



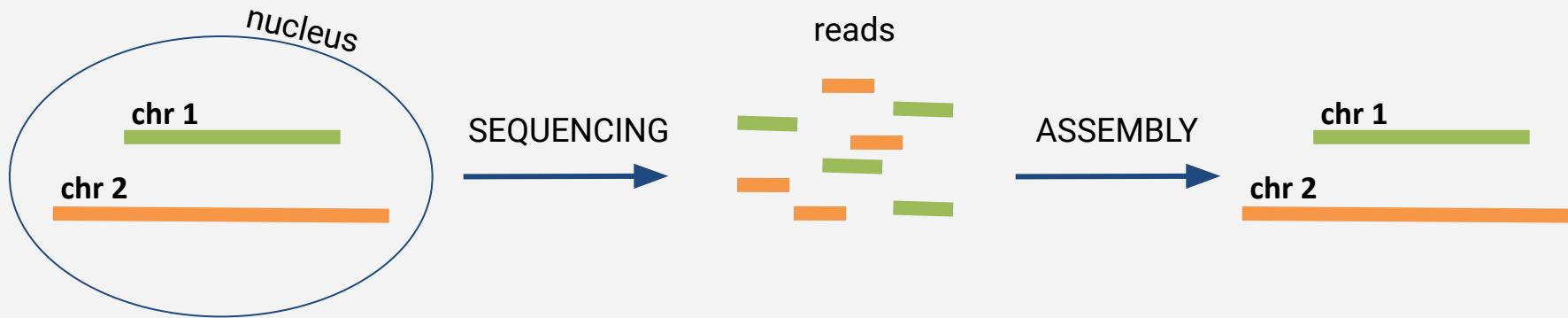
Introduction



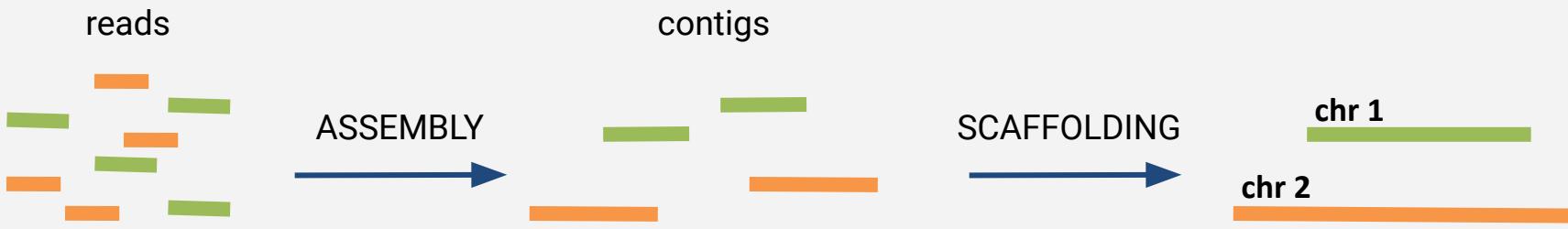
Introduction



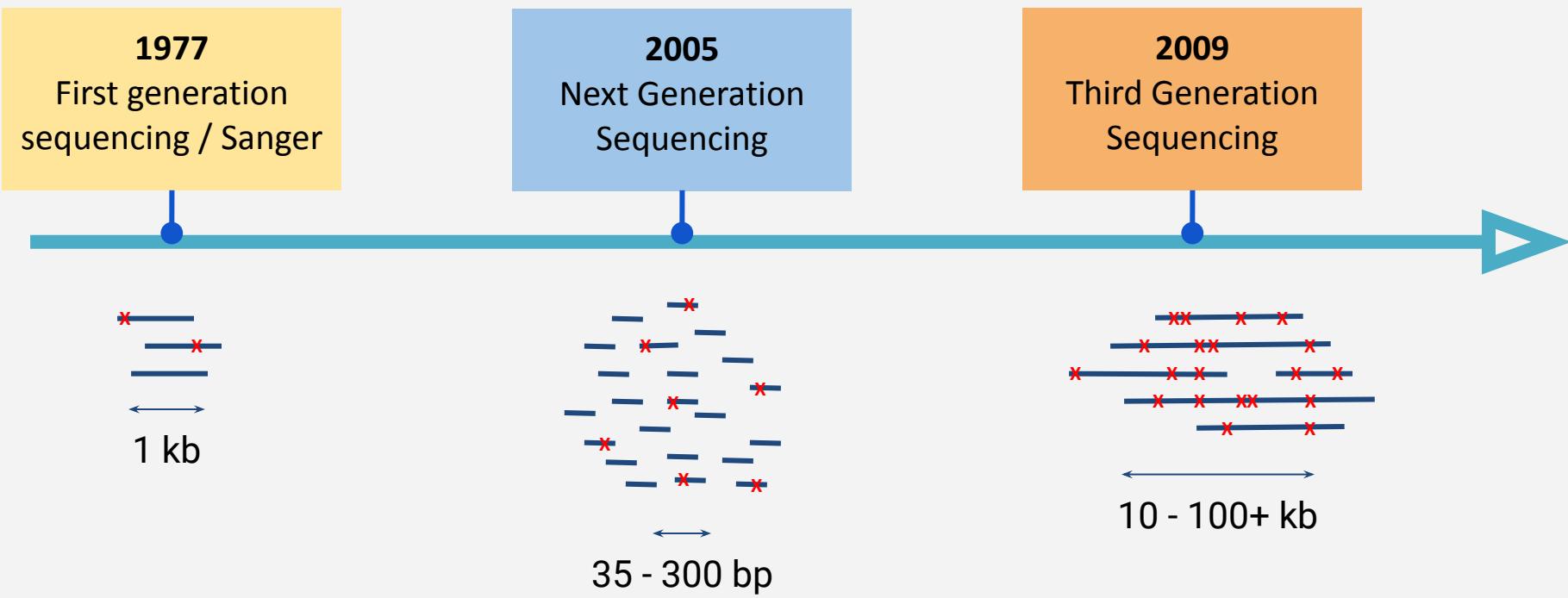
Introduction



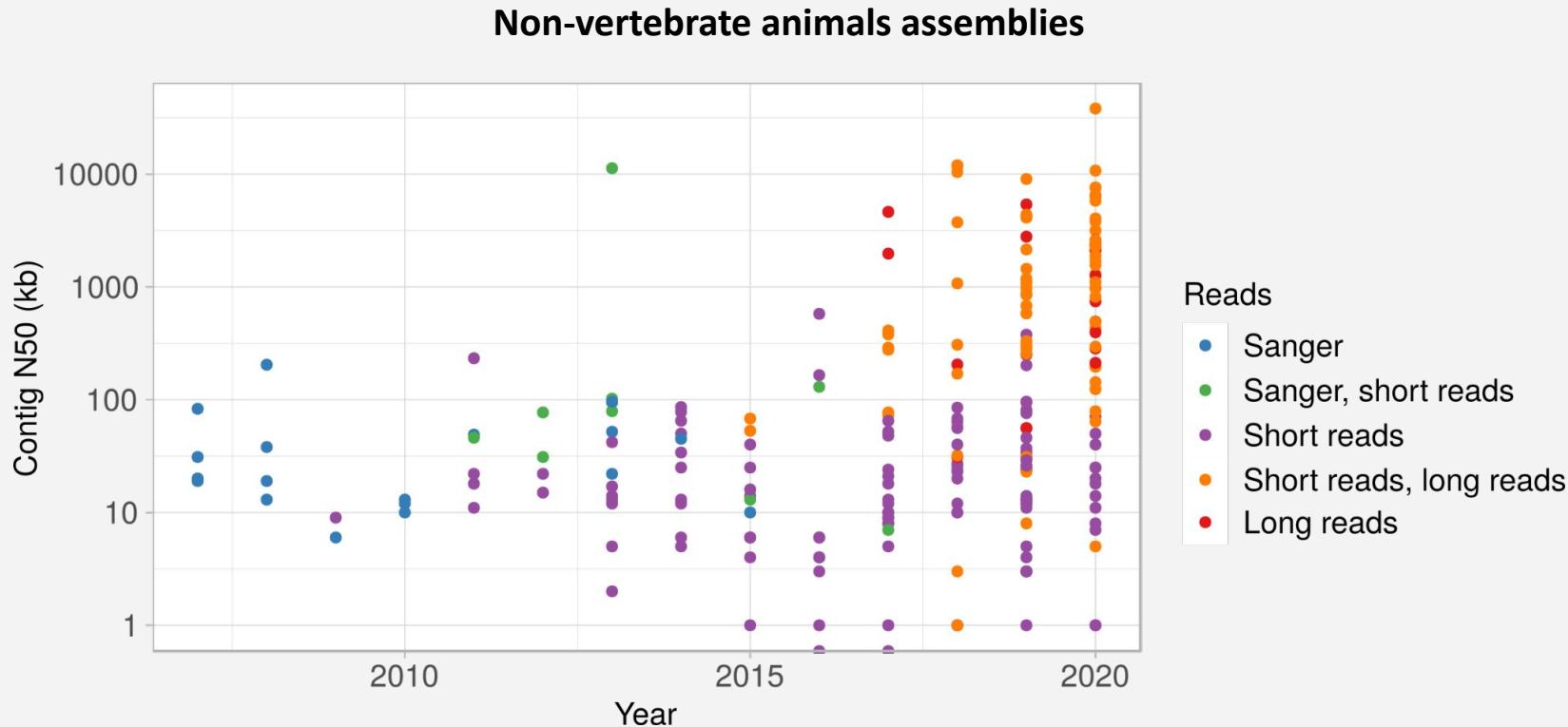
Introduction



Introduction



Introduction

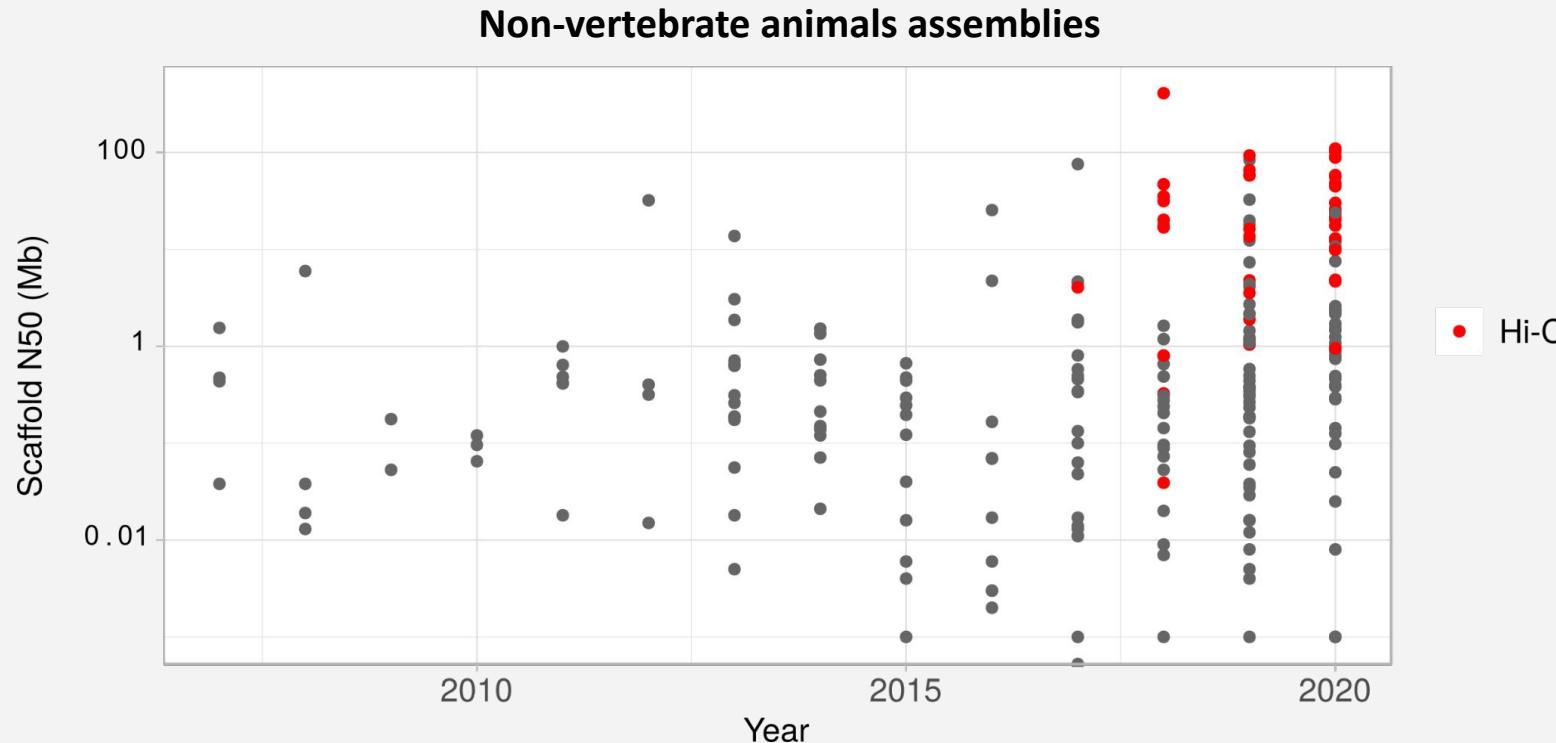


Introduction

Scaffolding: grouping and orienting contigs to build chromosome-level scaffolds

- **Mate-pairs:** short reads with a long insert
- **Long reads**
- **Genetic maps:** ordered markers
- **Optical maps:** ordered markers
- **Linked reads:** barcoded short reads
- **Hi-C/3C/Proximity ligation**

Introduction

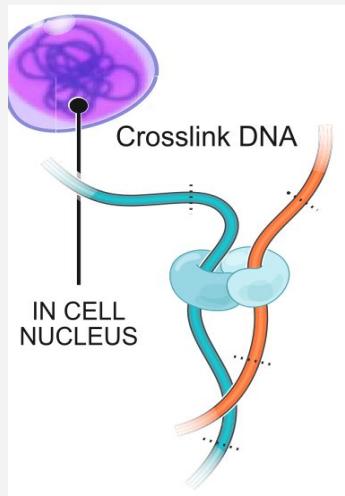


Introduction

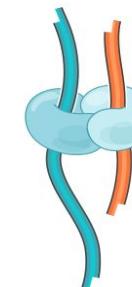
- Hi-C data
- Hi-C scaffolding
- Metagenomes
- Phasing assemblies

Hi-C data

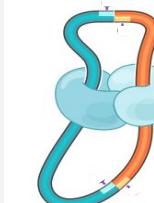
3C



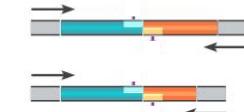
Cut with
restriction
enzyme



Ligate

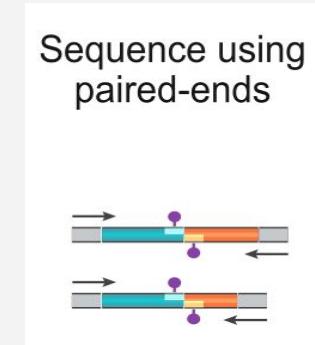
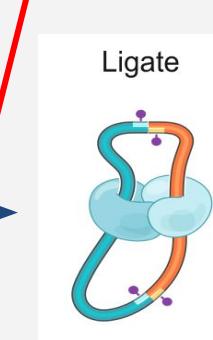
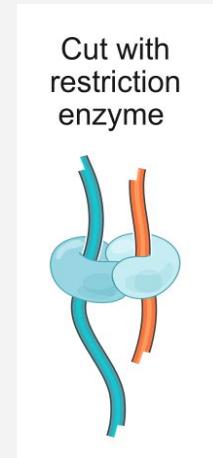
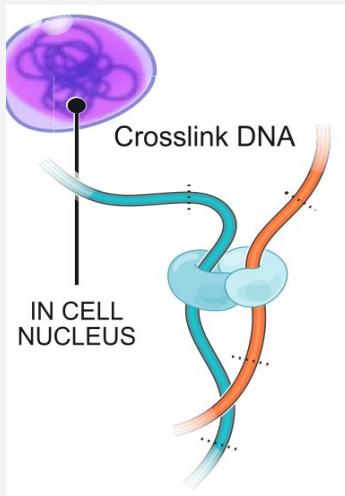


Sequence using
paired-ends



Hi-C data

Hi-C



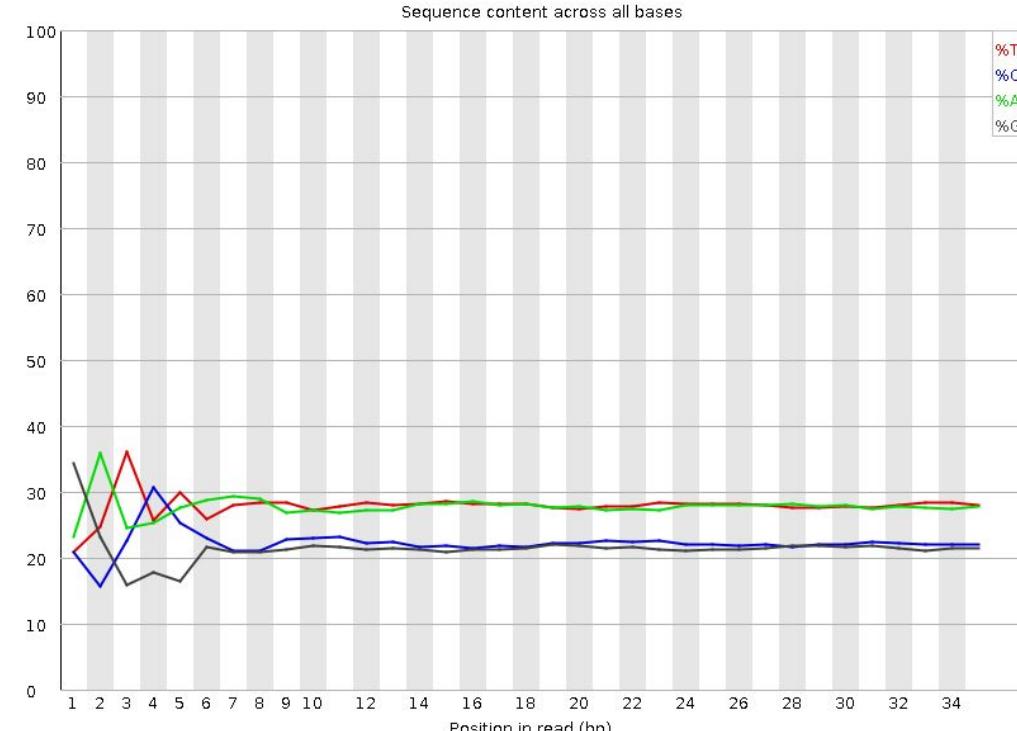
Hi-C data

Summary

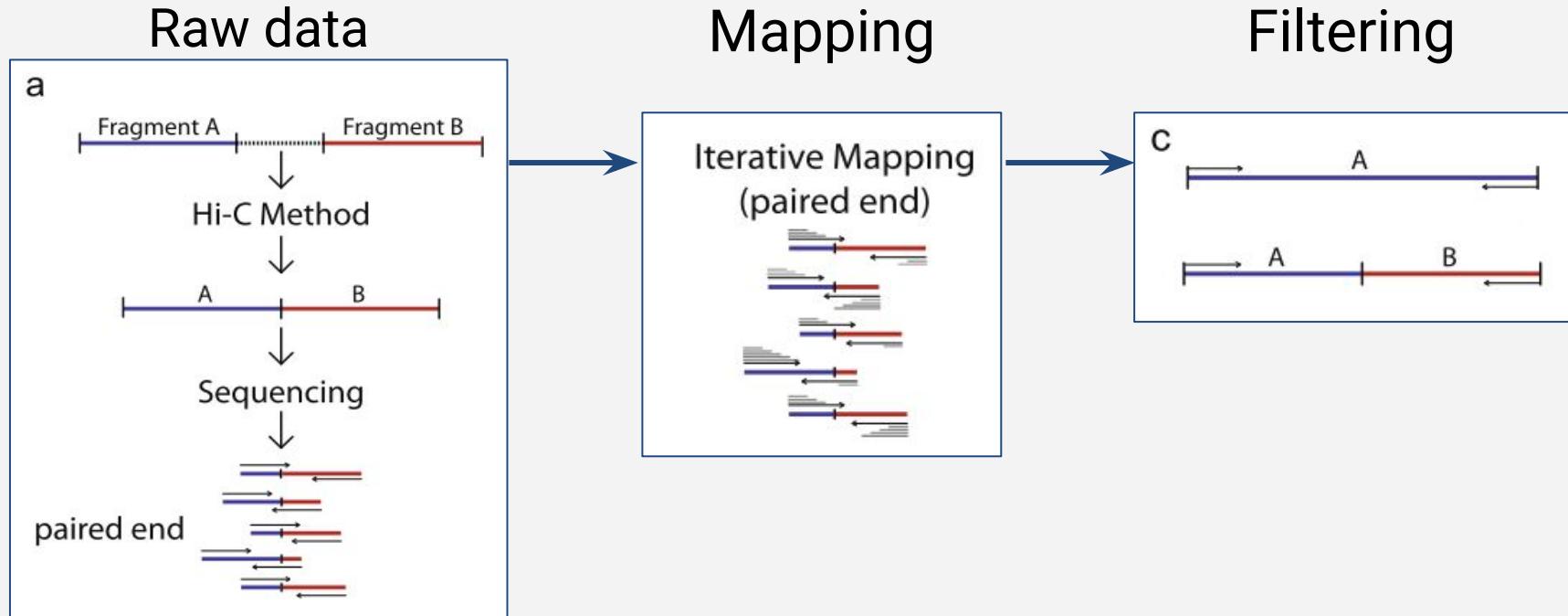
- ✓ [Basic Statistics](#)
- ✓ [Per base sequence quality](#)
- ✓ [Per tile sequence quality](#)
- ✓ [Per sequence quality scores](#)
- ! [Per base sequence content](#)
- ✓ [Per sequence GC content](#)
- ✓ [Per base N content](#)
- ✓ [Sequence Length Distribution](#)
- ! [Sequence Duplication Levels](#)
- ✓ [Overrepresented sequences](#)
- ✓ [Adapter Content](#)



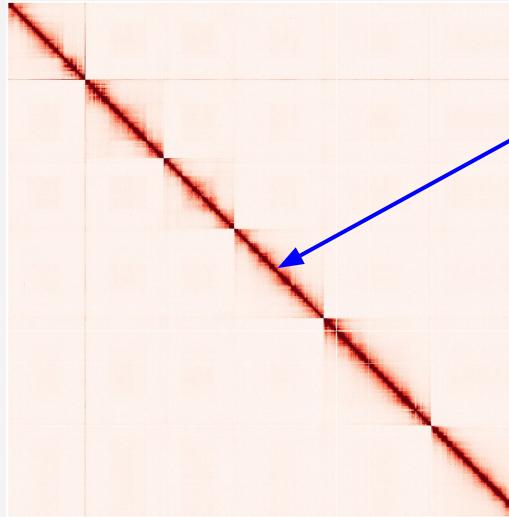
Per base sequence content



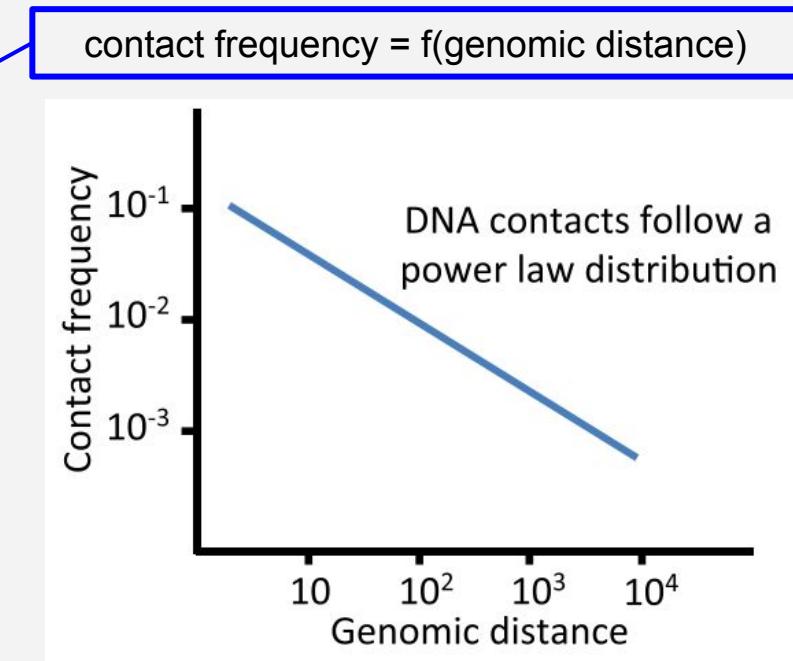
Hi-C data



Hi-C scaffolding



Contact map of
Caenorhabditis elegans



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

3D-DNA

✉ 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,...}

SALSA2

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

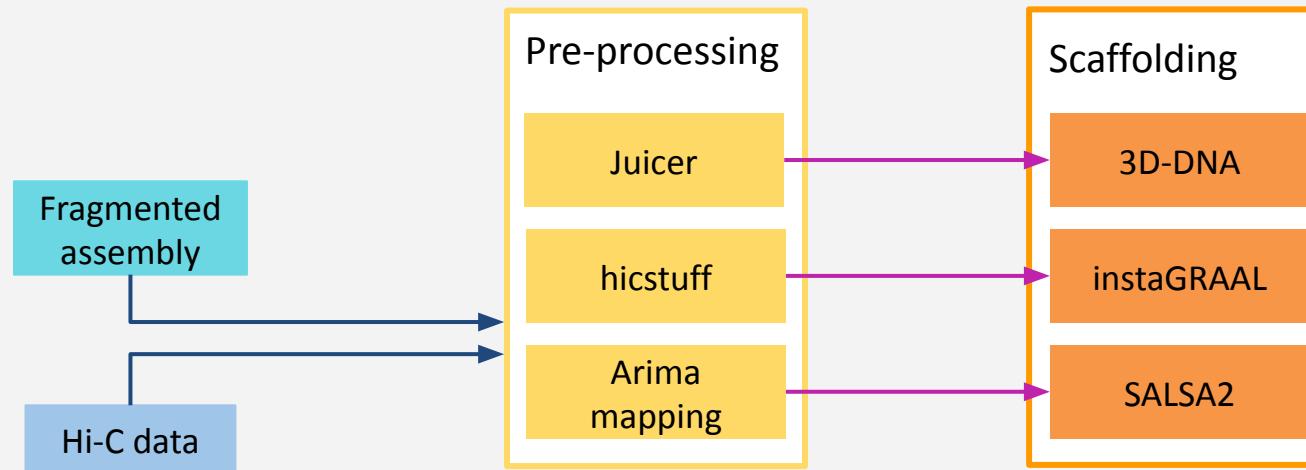
Jay Ghurye, Arang Rhie, Brian P. Walenz, Anthony Schmitt, Siddarth Selvaraj, Mihai Pop, Adam M. Phillippy ✉,
Sergey Koren ✉

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

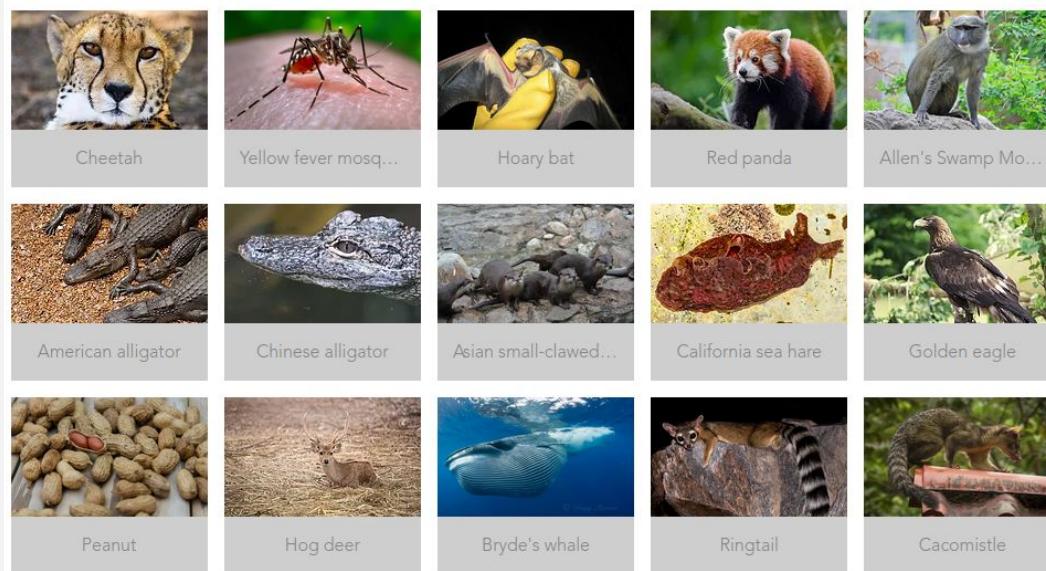
Lyam Baudry, Nadège Guiglielmoni, Hervé Marie-Nelly, Alexandre Cormier, Martial Marbouty, Komlan
Avia, Yann Loe Mie, Olivier Godfroy, Lieven Sterck, J. Mark Cock, Christophe Zimmer, Susana M. Coelho
✉ & Romain Koszul ✉

instaGRAAL

Hi-C scaffolding



Hi-C scaffolding



www.dnazoo.org

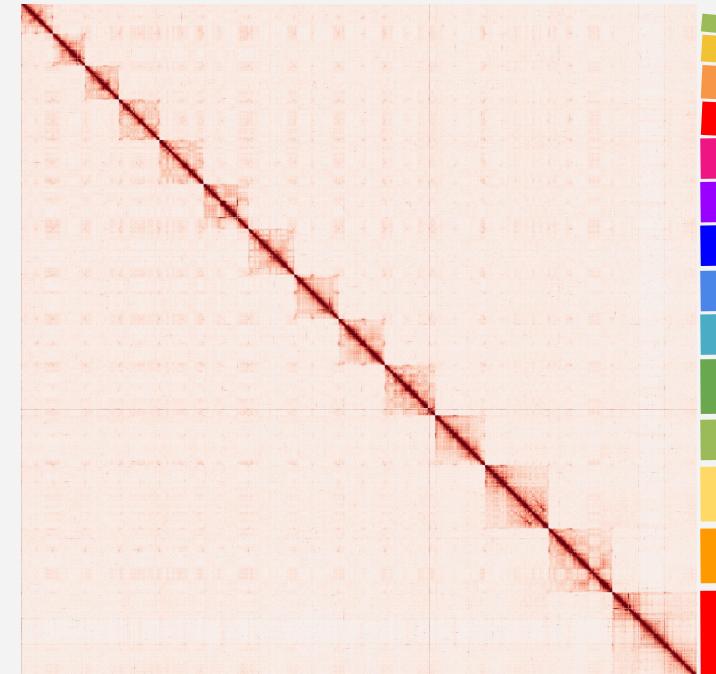
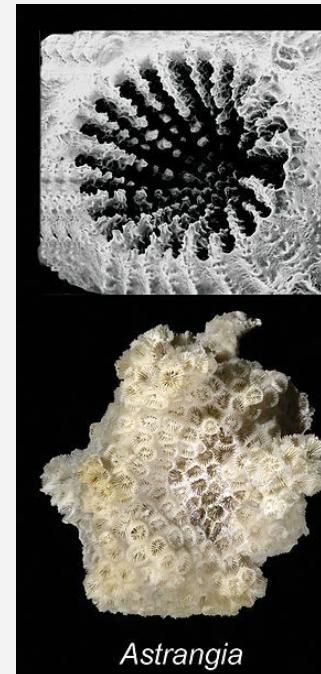


Hi-C scaffolding

Astrangia poculata
(coral)

14 scaffolds

455 Mb

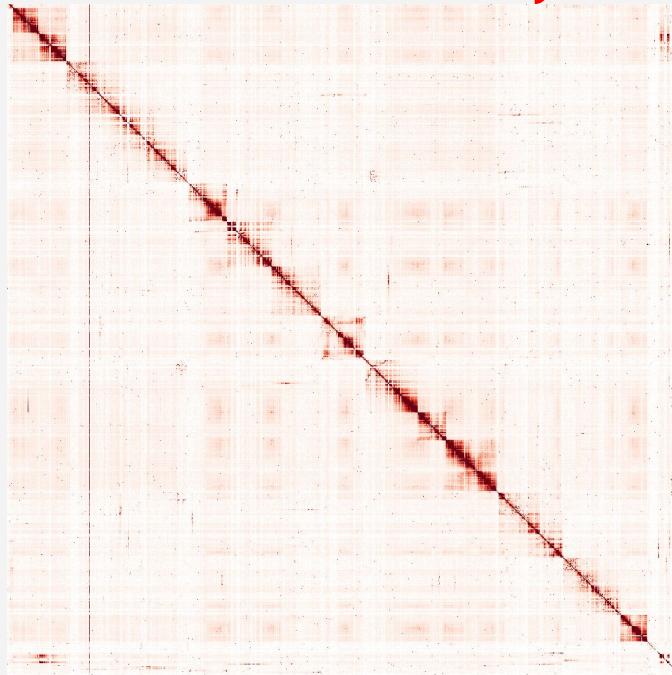


Hi-C scaffolding

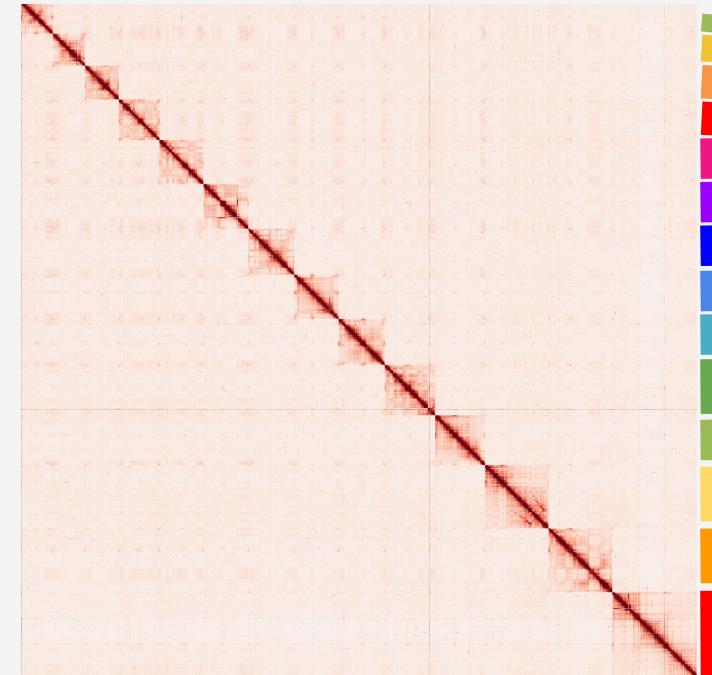
Lineage dynamics of the endosymbiotic cell type in the soft coral *Xenia*

Minjie Hu, Xiaobin Zheng, Chen-Ming Fan & Yixian Zheng

Nature 582, 534–538(2020) | Cite this article



Hi-C contact map of *Xenia sp.*

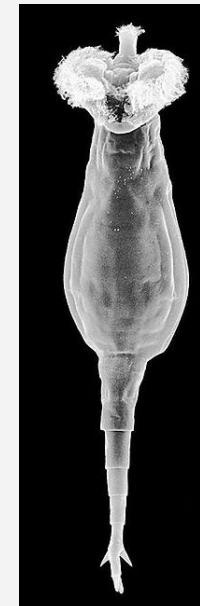


Hi-C contact map of *Astrangia poculata*

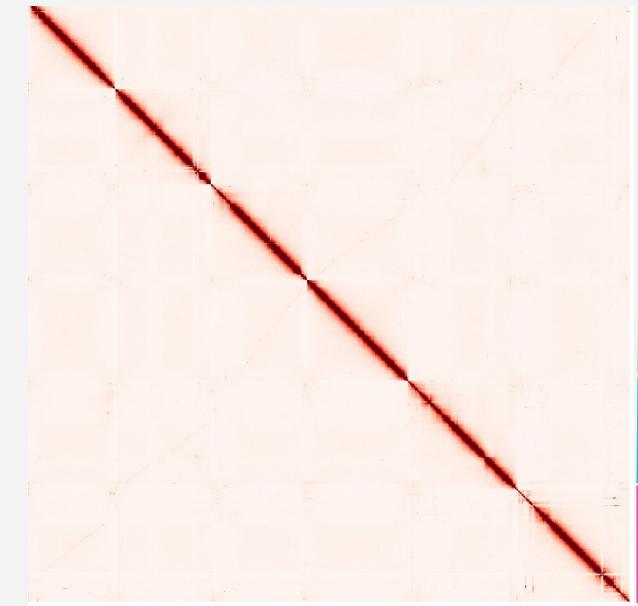
Hi-C scaffolding

Adineta vaga (rotifer)

6 scaffolds



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



Hi-C contact map of *Adineta vaga*

Hi-C scaffolding

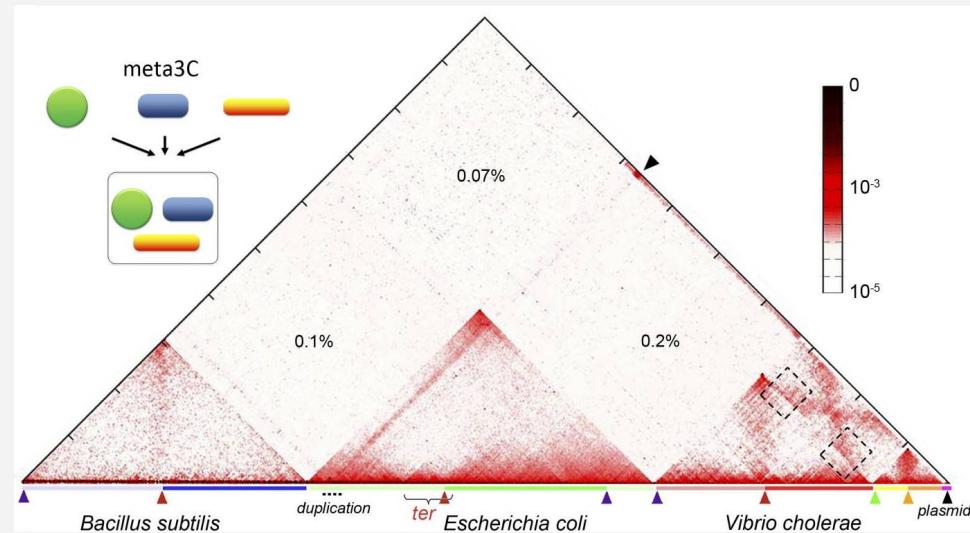
“What coverage should I get?”

→ Arima recommends 200 millions pairs per Gb

Species	Size	# fragments	# Hi-C pairs	Hi-C mapping
<i>Adineta vaga</i>	101 Mb	30	55 millions	83%
<i>Astrangia poculata</i>	455 Mb	2995	723 millions	67%
<i>Flaccisagitta enflata</i>	929 Mb	6612	489 millions	37%
<i>Mercenaria mercenaria</i>	1.86 Gb	5118	455 millions	55%

Metagenomes

Mixture of organisms, but different nuclei = no 3D contacts



Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms, Marbouty et al, 2014

Metagenomes

Metator

INPUT

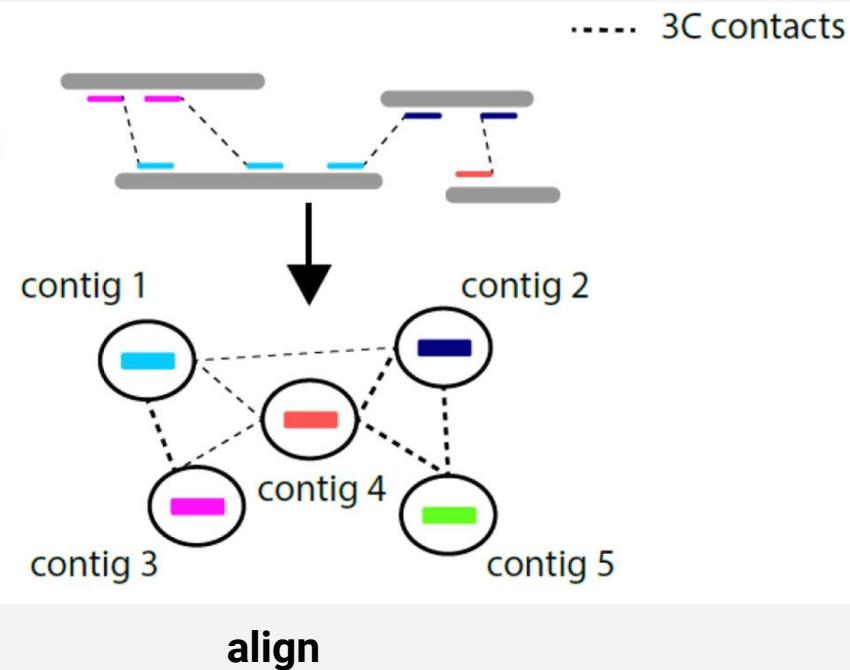
meta3C library
PE reads



metagenome assembly

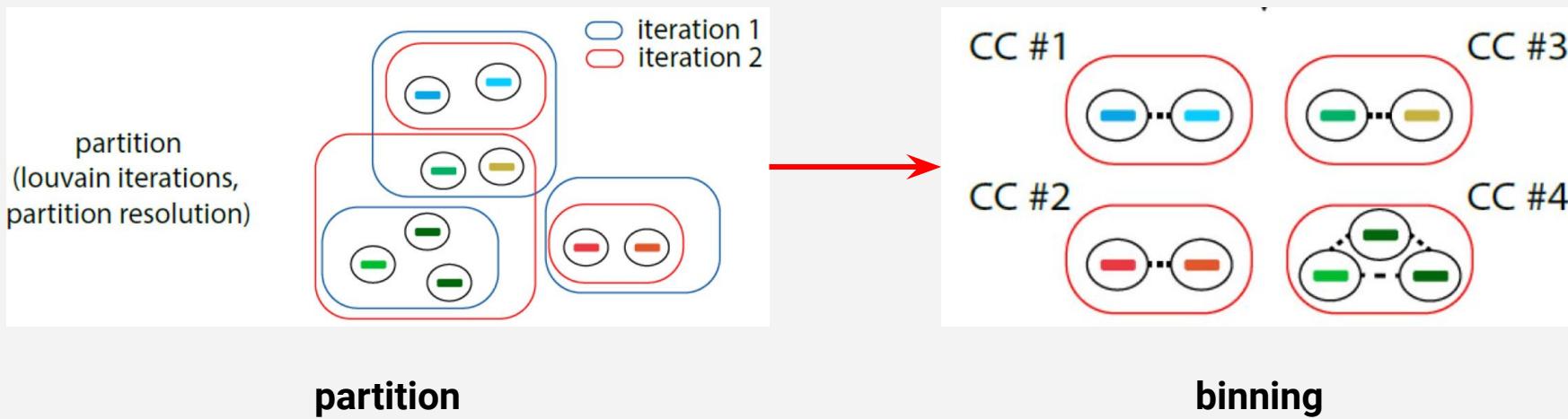
alignment
(mapping, filtering,
sorting merging)

network
(counting,
normalization)



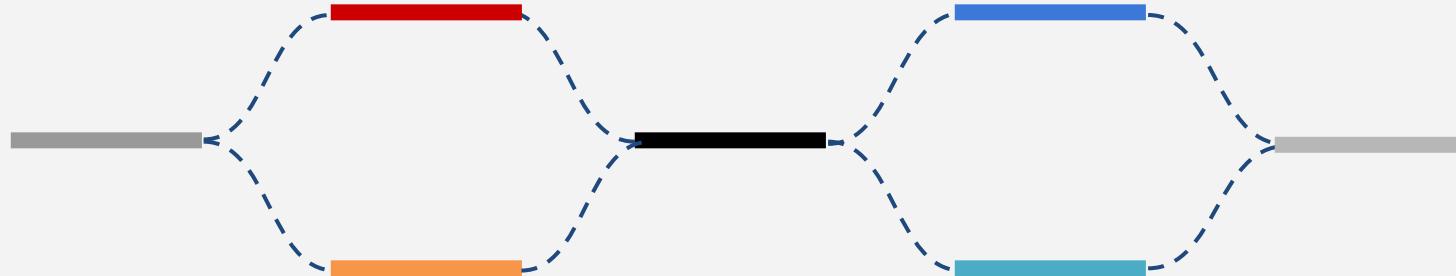
Metagenomes

Metator



Phasing assemblies

Diploid genome
= 2 haplotypes



Phasing assemblies

GraphUnzip

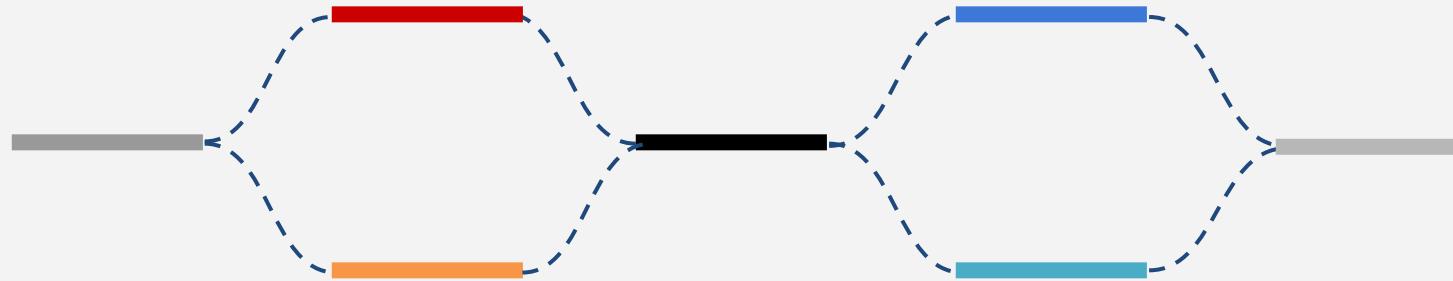
- ★ input: assembly graph + Hi-C data
- ★ output: phased supercontigs

GraphUnzip: unzipping assembly graphs with long reads and Hi-C

Roland Faure,  Nadège Guiglielmoni,  Jean-François Flot
doi: <https://doi.org/10.1101/2021.01.29.428779>

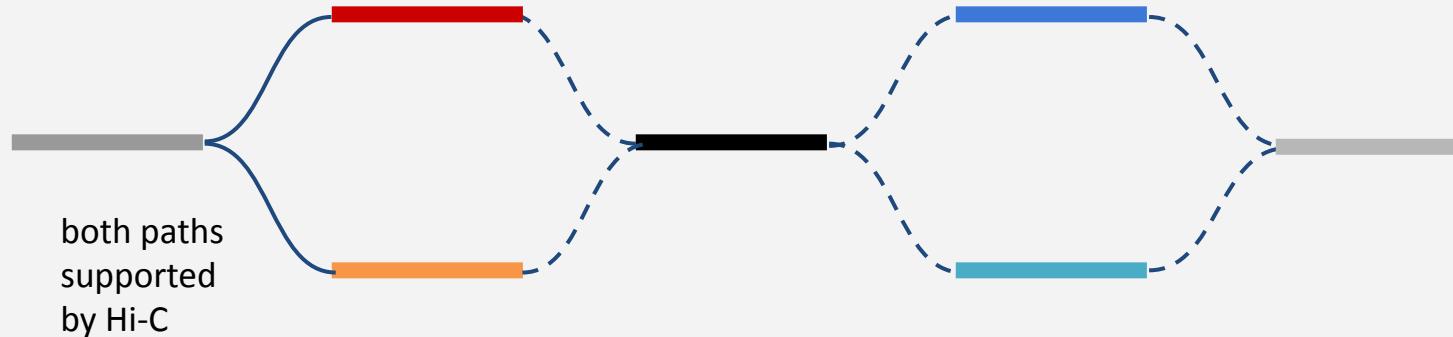
Phasing assemblies

GraphUnzip



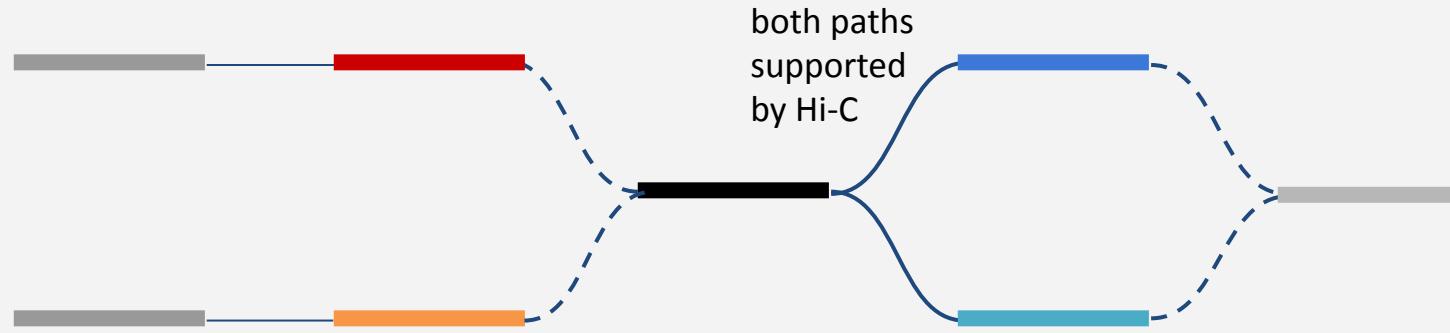
Phasing assemblies

GraphUnzip



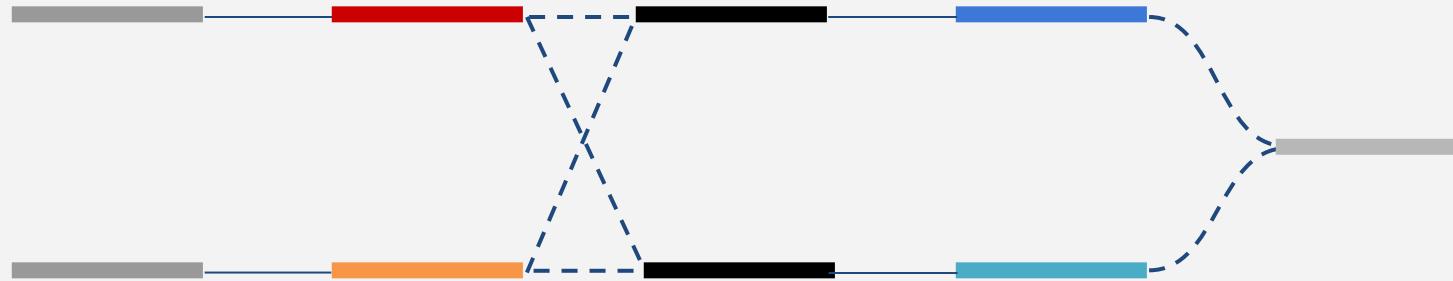
Phasing assemblies

GraphUnzip



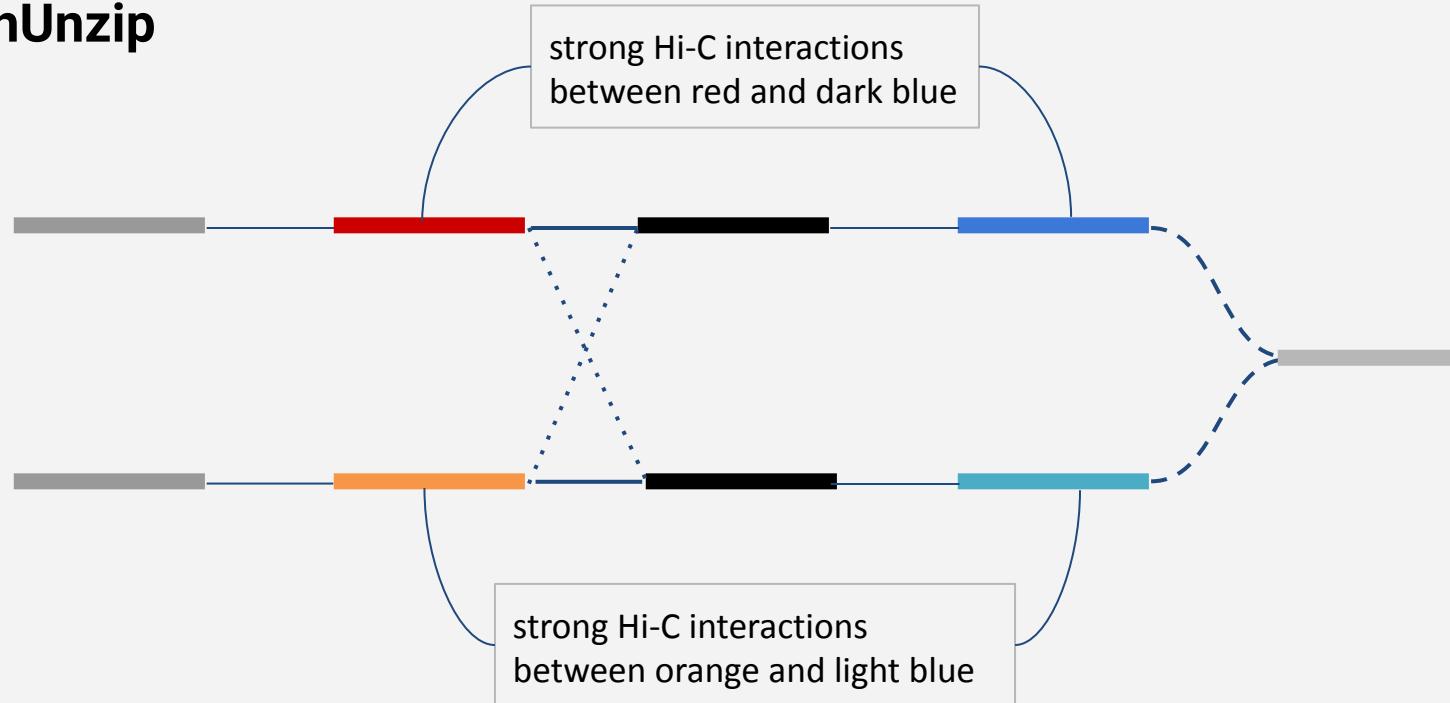
Phasing assemblies

GraphUnzip



Phasing assemblies

GraphUnzip



Phasing assemblies

GraphUnzip



Phasing assemblies

GraphUnzip

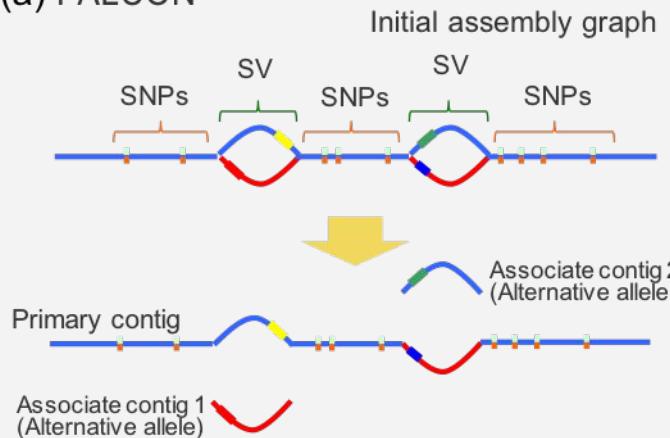


phased supercontigs

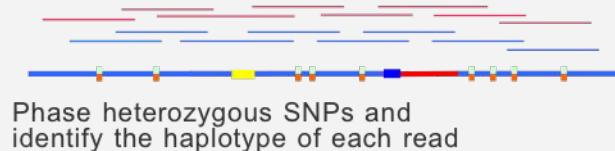


Phasing assemblies

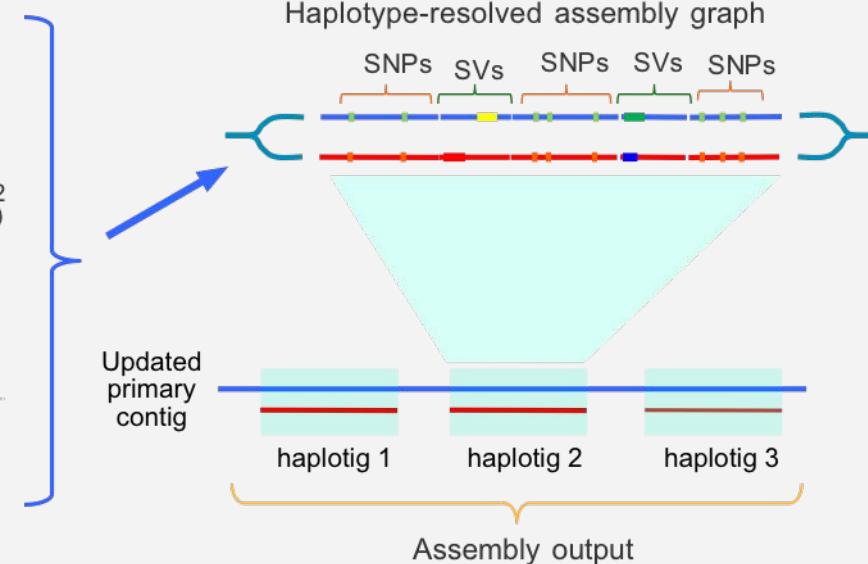
(a) FALCON



(b)



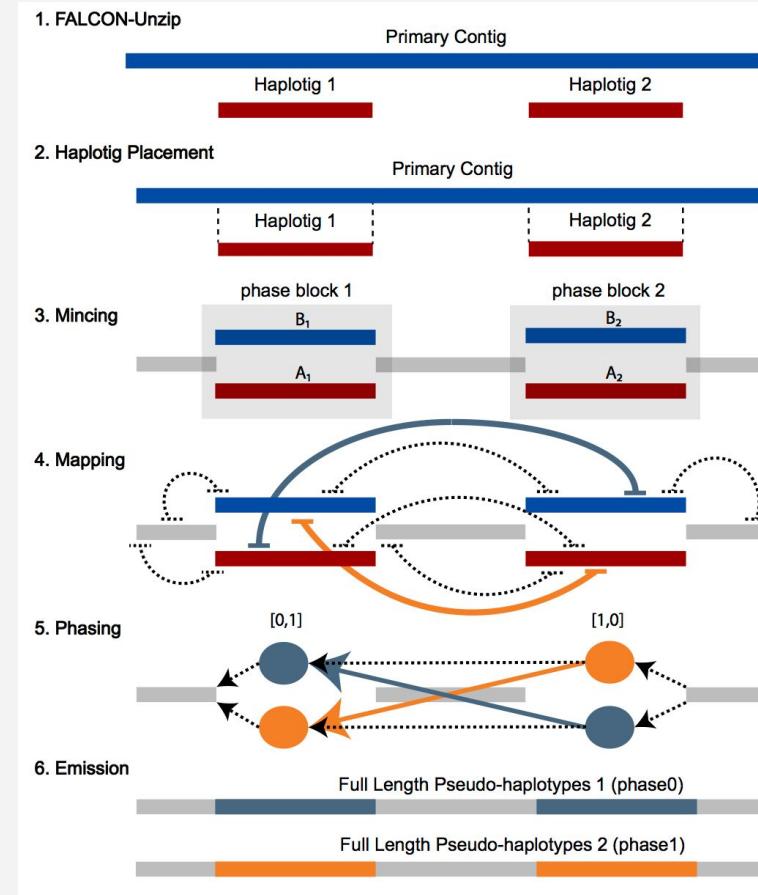
(c) FALCON-Unzip



Phasing assemblies

FALCON-Phase

Extended haplotype phasing of *de novo* genome assemblies with FALCON-Phase,
Kronenberg et al., 2019



Thank you for your attention!
Questions?

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