ULB UNIVERSITÉ LIBRE DE BRUXELLES

SeqBio 2018
Rouen, 19-20 novembre

Using chromosome conformation capture to assemble genomes to perfection

Nadège Guiglielmoni, Antoine Limasset, Romain Koszul, Jean-François Flot

SeqBio, 19-20 Novembre 2018

Our goal

Using chromosome conformation capture

= information on DNA 3D structure

To assemble genomes

= obtain less fragmented genomes

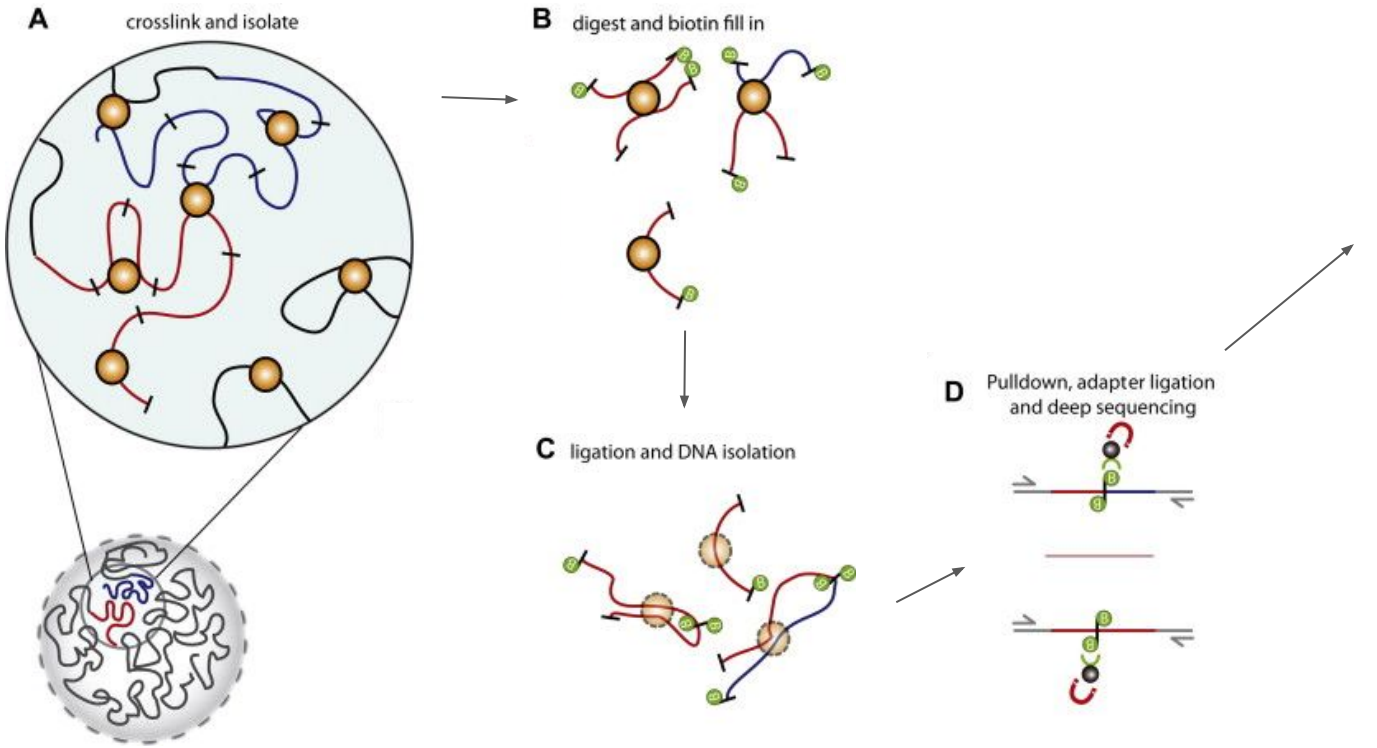
= phase haplotypes

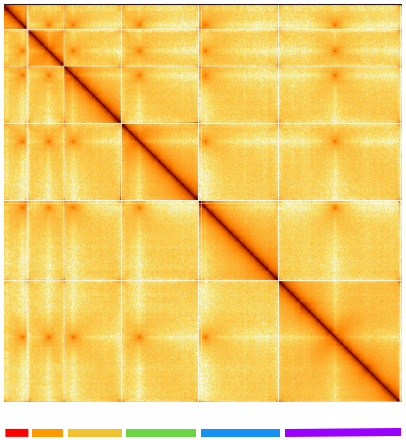
What is chromosome conformation capture ?

Chromosome Conformation Capture (3C)
Technique to study DNA 3D structure
Published in 2002 by Job Dekker
⟶ Hi-C : whole genome contacts (2009)

Capturing chromosome conformation. Dekker et al., 2002
Comprehensive mapping of long-range interactions reveals folding principles of the human genome.Lieberman-Aiden et al., 2009
Exploring the three-dimensional organization of genomes: Interpreting chromatin interaction data. Dekker et al., 2013

Chromosome conformation capture technique : Hi-C





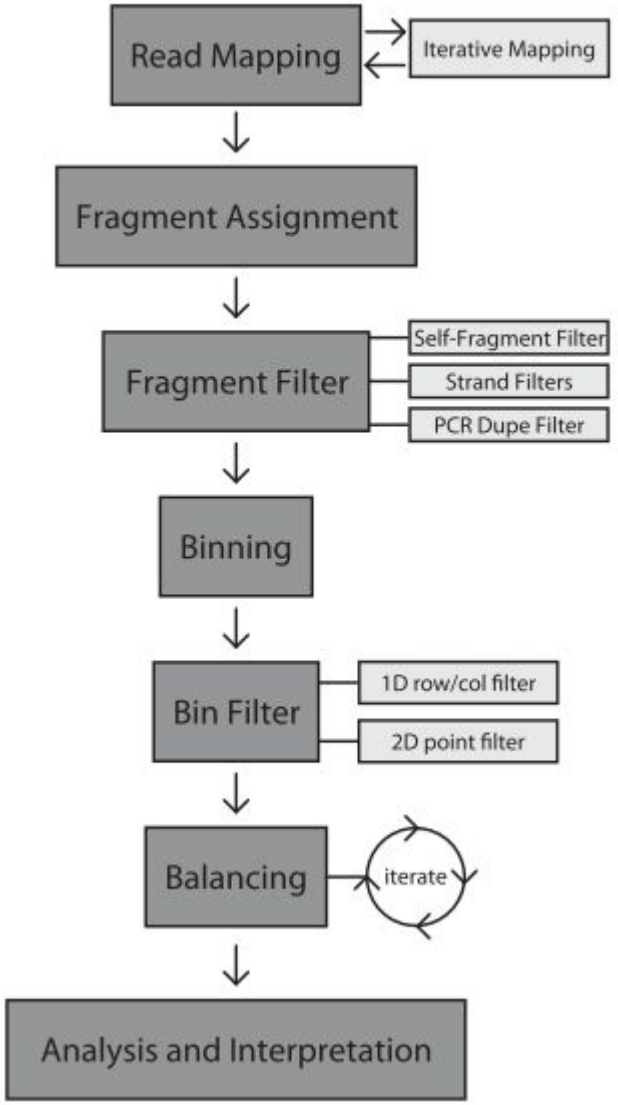
Partial map of
Lachancea cidri

Hi–C: A comprehensive technique to capture the conformation of genomes. Belton et al., 2012

Data processing

Steps :

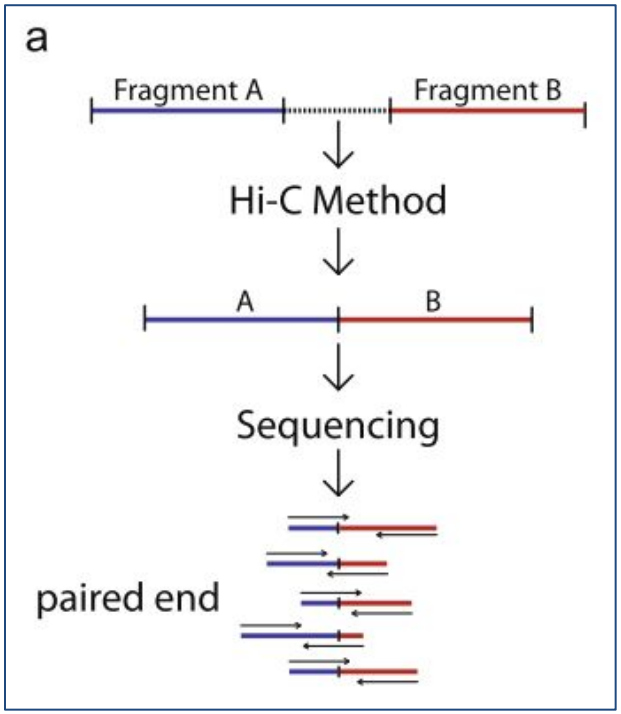
- ➔ Map reads on the genome
- ➔ Attribute fragments
- ➔ Eliminate non-chimeric reads
- ➔ Bin reads and build a map
- ➔ Normalize maps



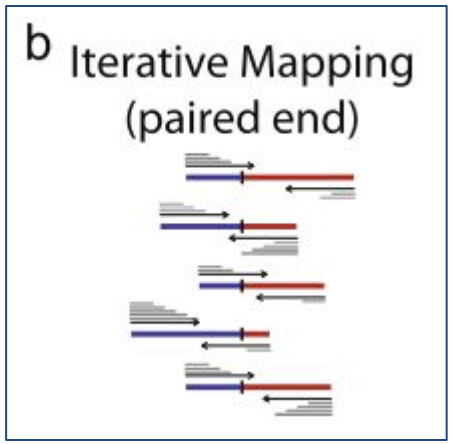
The Hitchhiker's guide to Hi-C analysis: Practical guidelines. Lajoie et al., 2015

Data processing

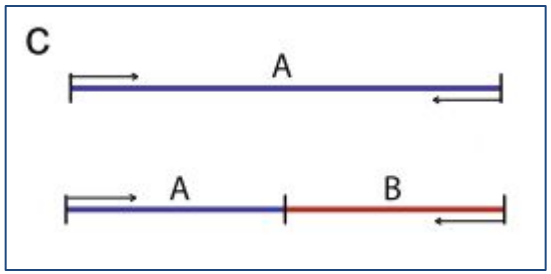
Raw data



Mapping

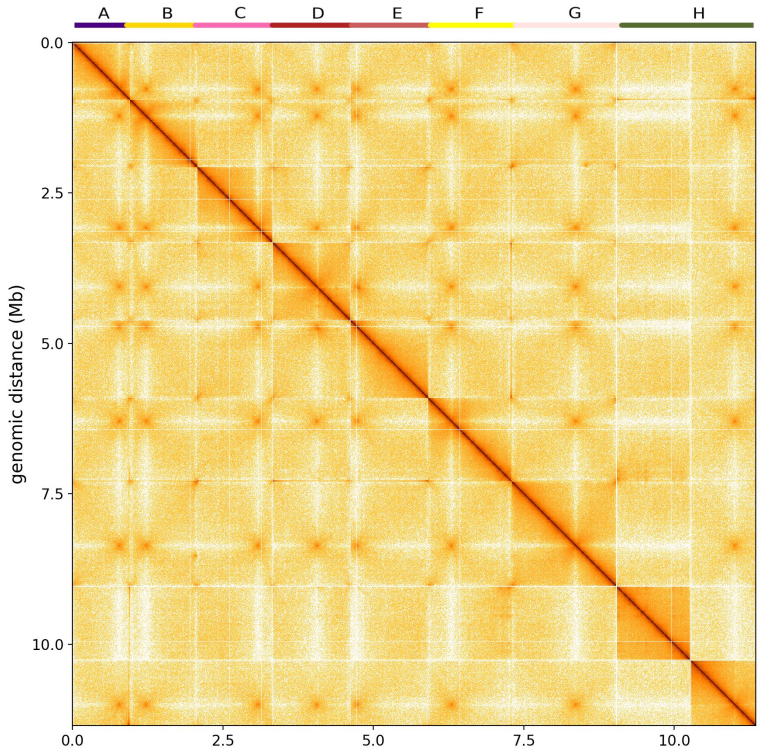


Filtering



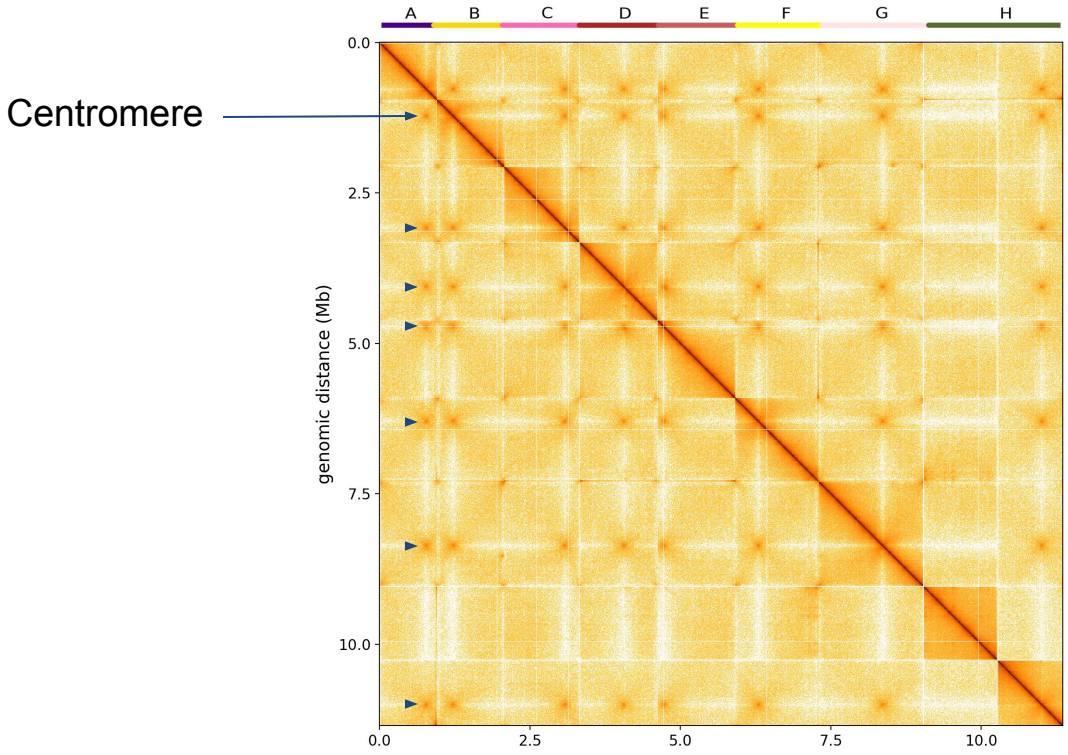
The Hitchhiker's guide to Hi-C analysis: Practical guidelines. Lajoie et al., 2015

Data processing : contact map



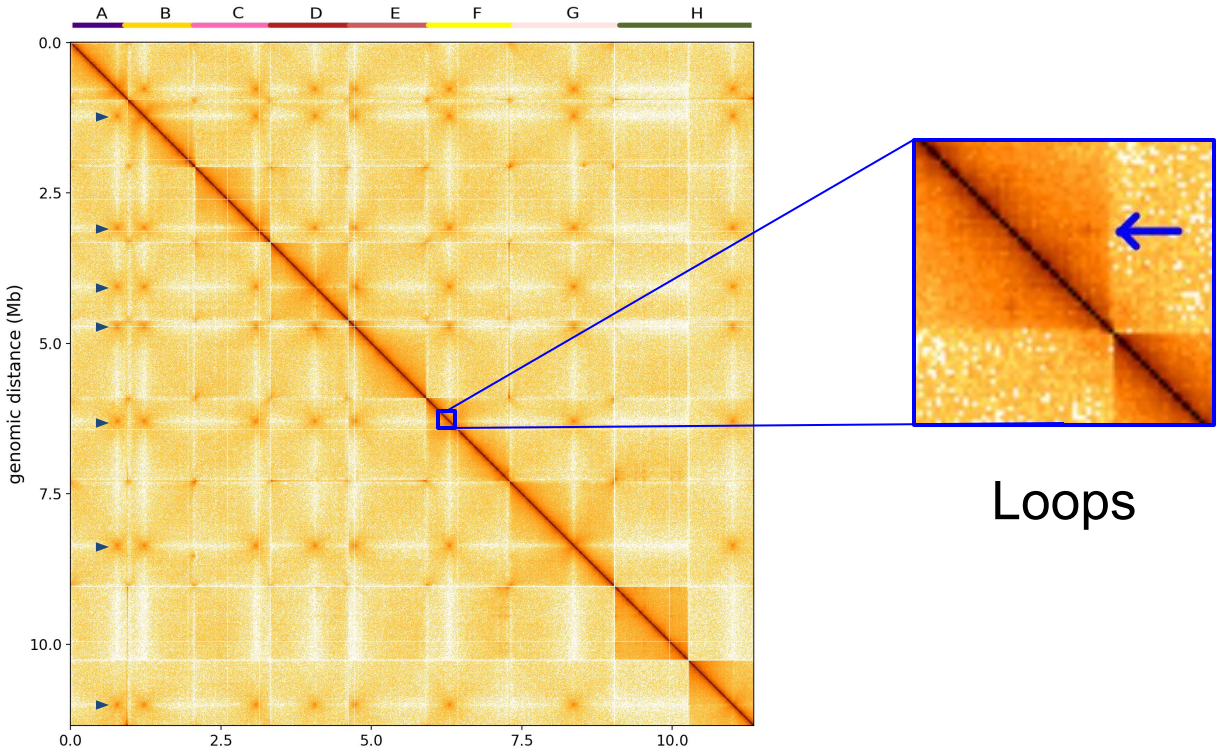
Yeast : *Lachancea kluyveri*

Data processing : contact map



Yeast : *Lachancea kluyveri*

Data processing : contact map



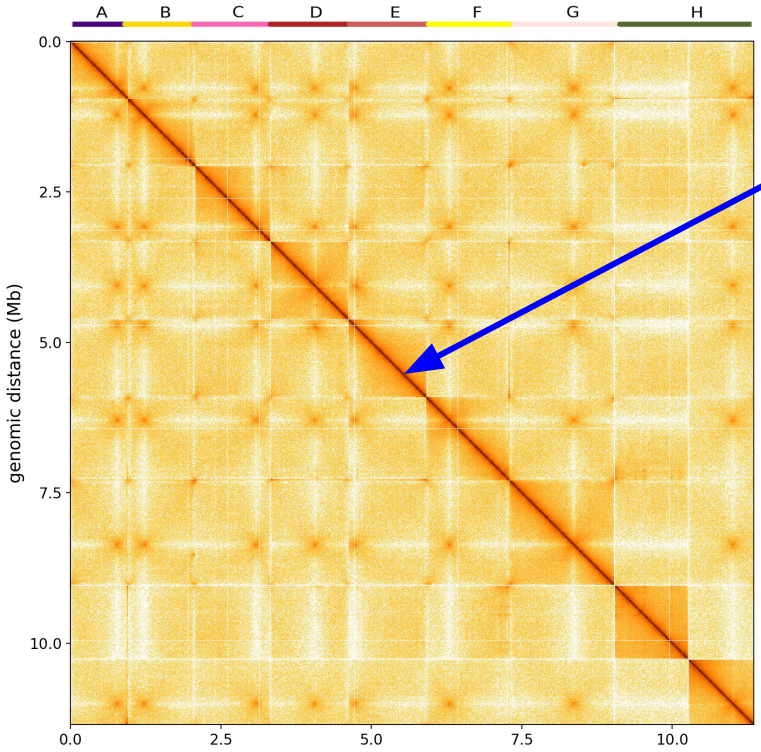
Loops

Yeast : Lachancea kluyveri

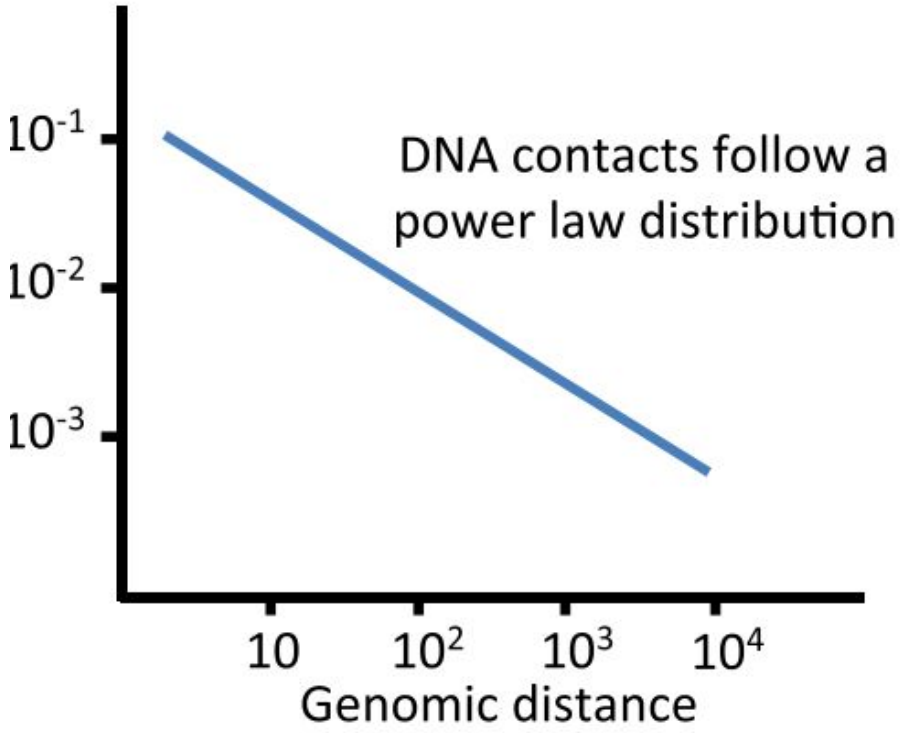
How can chromosome conformation capture data assist in genome assembly ?

➔ DNA 3D structure is linked to 1D structure

Genomic distance law



contact frequency = f(genomic distance)



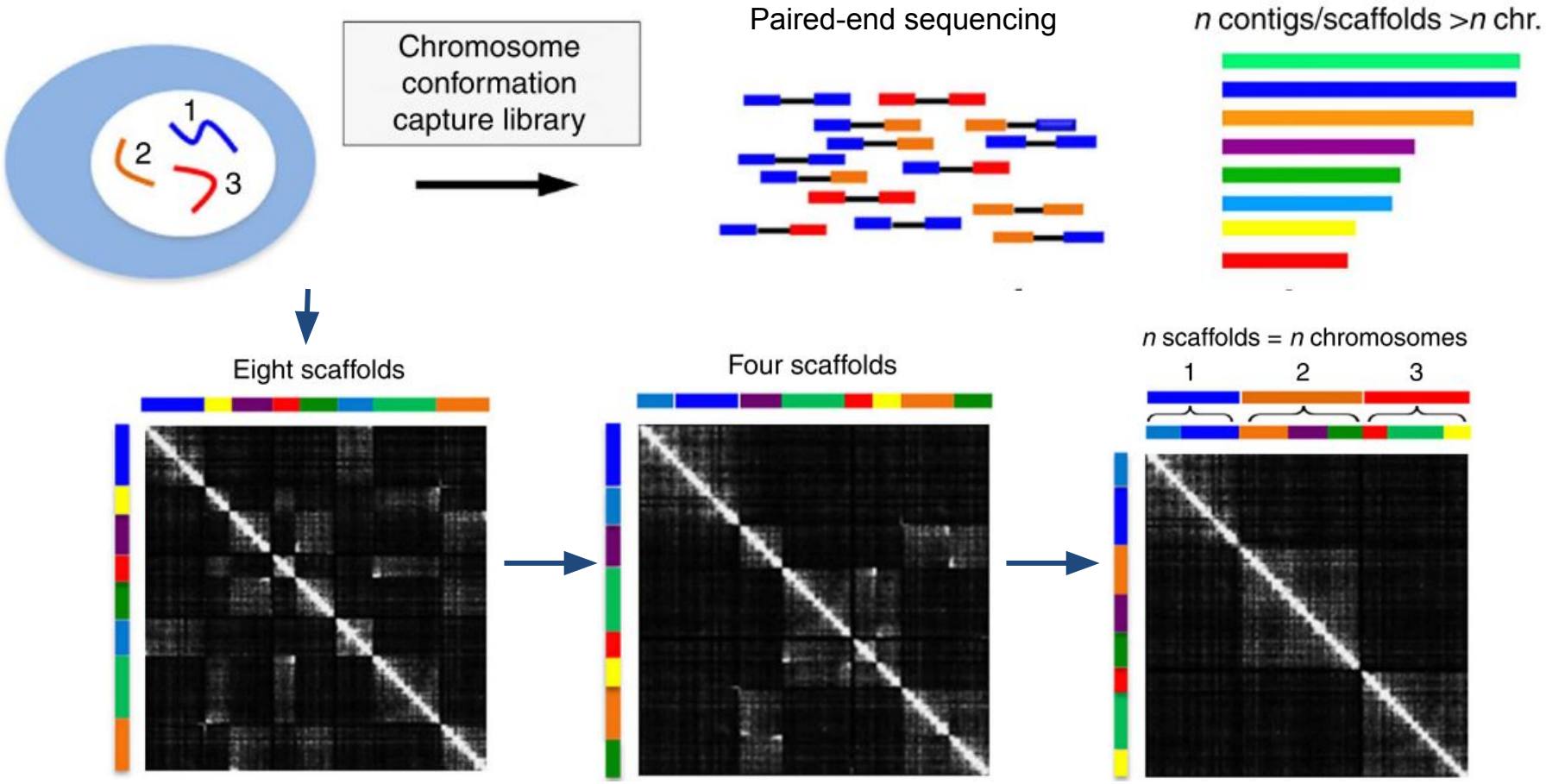
Contact genomics: scaffolding and phasing (meta)genomes using chromosome 3D physical signatures. Flot et al., 2015

Chromosome conformation capture in assembly

Tools to scaffold assemblies or reassemble with Hi-C data :

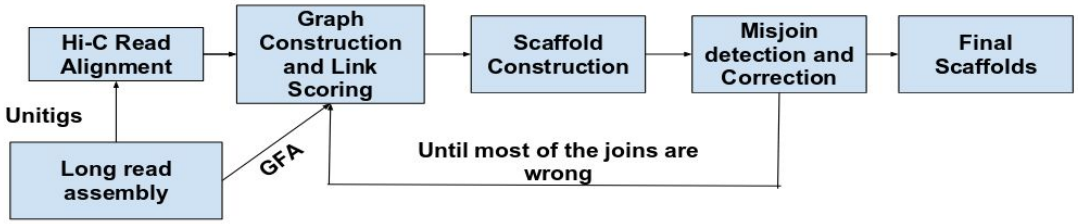
- GRAAL - instaGRAAL : reassemble to obtain a more likely genome according to contact data
- SALSA2 : scaffold by introducing Hi-C links in assembly graphs
- 3D-DNA, Lachesis...

GRAAL principle



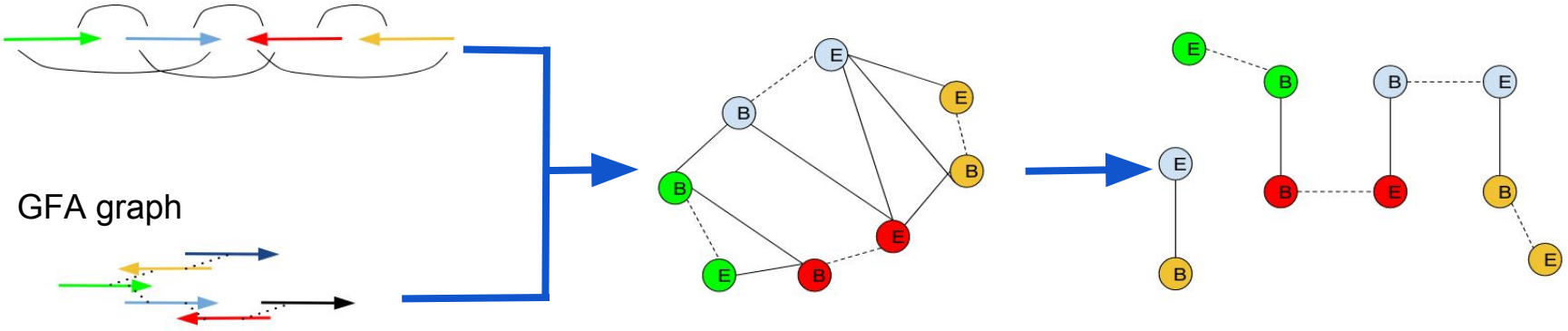
High-quality genome (re)assembly using chromosomal contact data. Marie-Nelly et al., 2014

SALSA2 principle



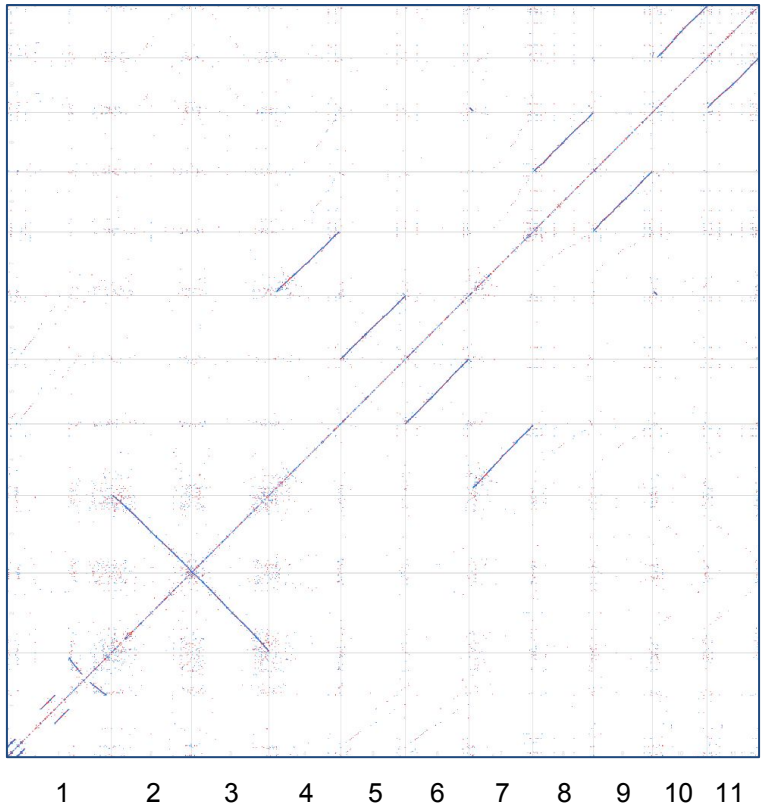
Hi-C data

GFA graph

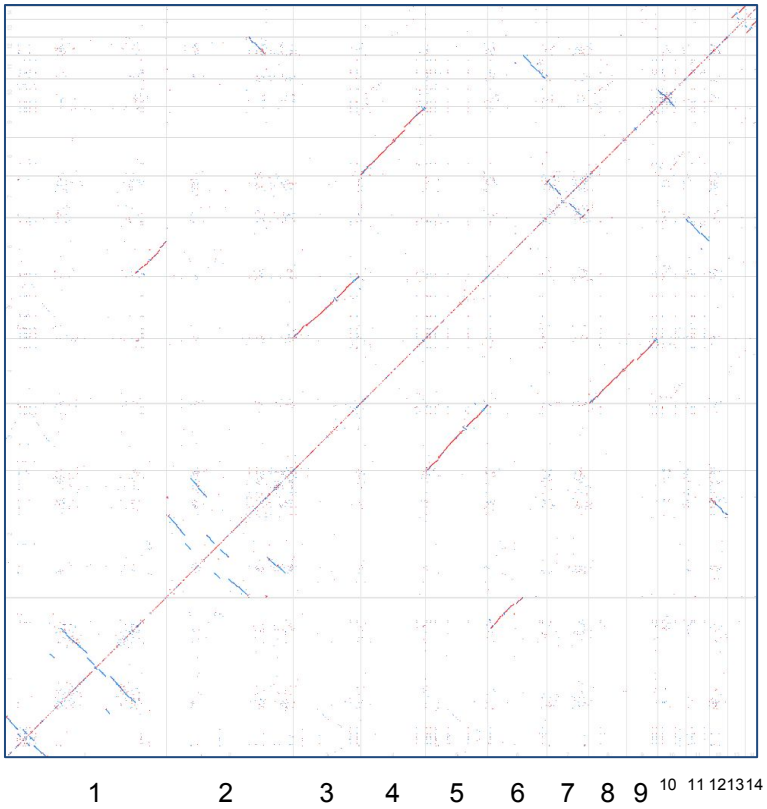


Integrating Hi-C links with assembly graphs for chromosome-scale assembly. Ghurye et al., 2018

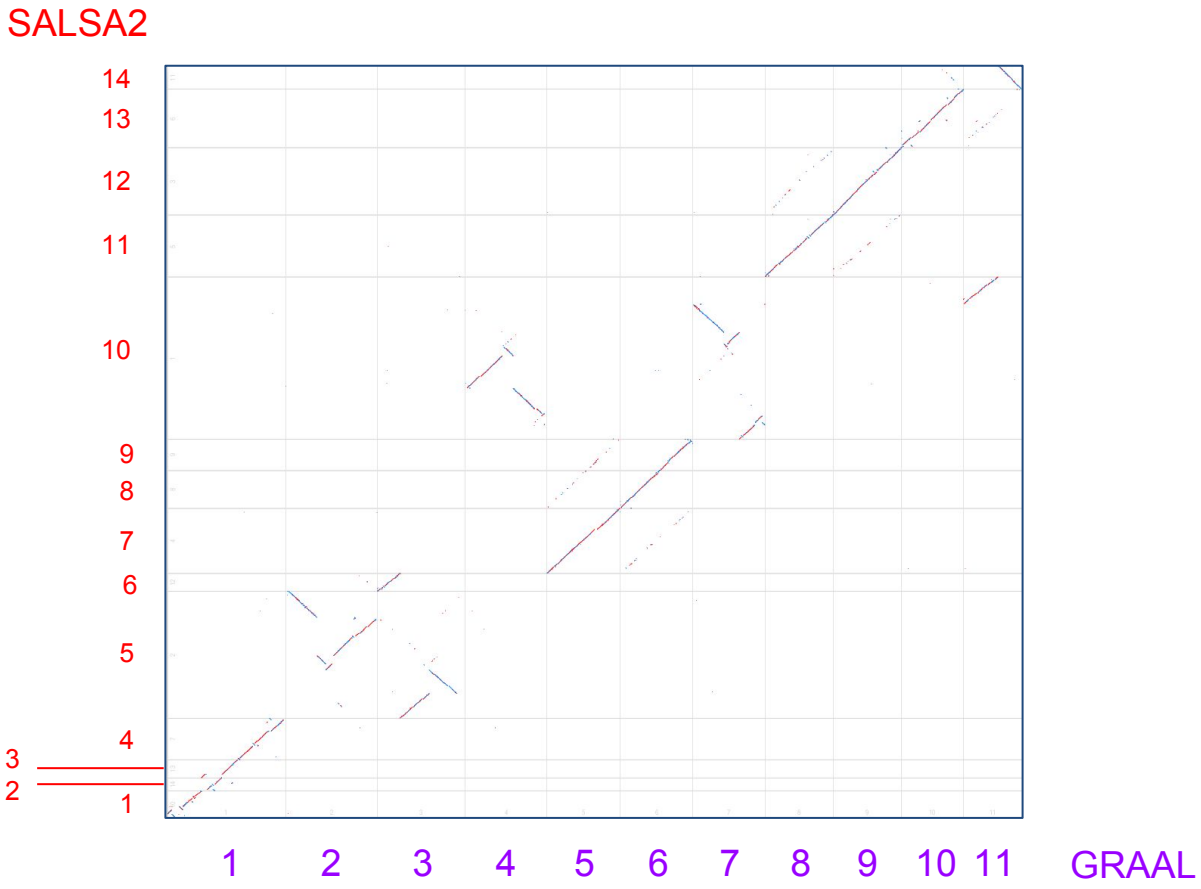
GRAAL reassembly of *Adineta vaga*



SALSA2 scaffolding of *Adineta vaga*



GRAAL vs SALSA2



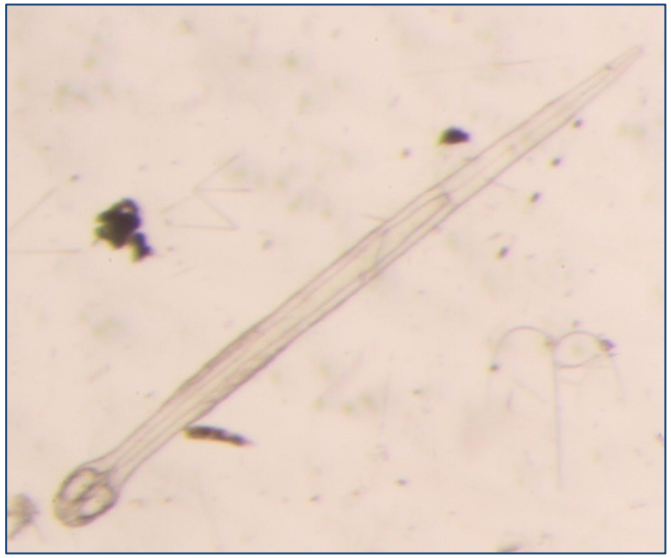
Our project

- Make a tool for genome scaffolding using 3C/Hi-C data

- Test our strategy by generating Hi-C data and assembling the genome of an organism

→ Chaetognaths

Focus on an organism : Chaetognaths



Also known as “arrow worms”

Currently classified as protostomians

Phylum : Chaetognatha
- Class : Sagittida
  - Orders : Aphragmophora
    Phragmophora
- Class: Spadella

IGNITE Innovative Training Network

Goal : study invertebrate non-model organisms

Motivation : invertebrate organisms represent the majority of animal diversity but remain understudied

Gathers biologists, bioinformaticians, computer scientists…

http://www.itn-ignite.eu

@ignite_eu

Innovative Training Network
ignite
Comparative genomics of non-model invertebrates
MARIE CURIE ACTIONS
Paschalis Nitsidis (UC London)
Ferenc Kagan (Sars Centre)
Charles Hillier (EMBL)
Mattia Giacomelli (University Bristol)
Benjamin Bettisworth (HITS)
Kenneth Sandoval (University Galway)
Rati Kar (INRA Rennes)
Juan Antonio Ruiz Santiesteban (University Zagreb)
Yihe Zhao (CIIMAR Porto)
Ramon Rivera
Zeyuan Chen (ZSM Munich)
Maria Dimitrova (Pensoft Sofia)
Killian Smith (Era7 Amilla)
Nadège Guiglielmoni (Universite Libre de Brussels)
Ksenia Juravel (LMU Munich)
QUEENSLAND MUSEUM
Institut Pasteur
ALPHA BIOTOXINE
Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften
INRIA
http://itn-ignite.eu
European Commission
Horizon 2020 European Union funding for Research & Innovation

Acknowledgements

EBE, Université libre de Bruxelles
Jean-François Flot

CRIStAL, Université de Lille
Antoine Limasset

Equipe RSG, Institut Pasteur
Romain Koszul
Axel Cournac
Lyam Baudry
Martial Marbouty
Rémi Montagne
Agnès Thierry

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