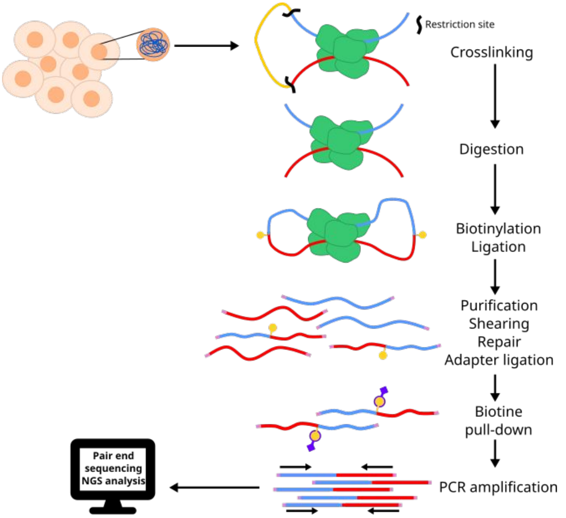
Hi-C BERG : Statistical inference of repeated sequence contacts in Hi-C maps

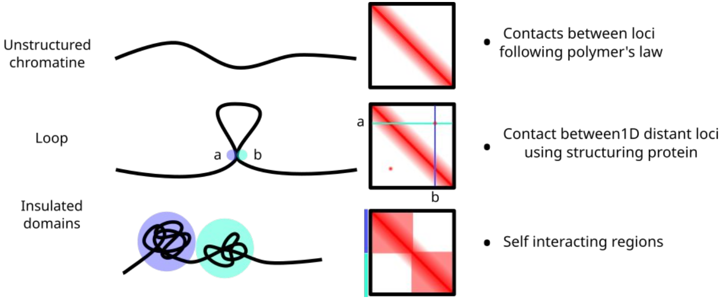
ISMB / ECCB 2023 - Lyon

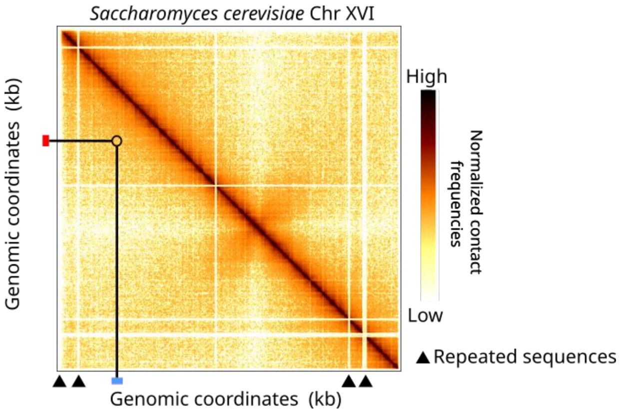
Sébastien Gradit
Spatial Regulation of Genomes - UMR3525

How to investigate genomes conformations ? Chromosome conformation Capture

Indrect observation of chromatine architecture through chimeric DNA construction







How to investigate genomes conformations and track chromatine architecture dynamics ?

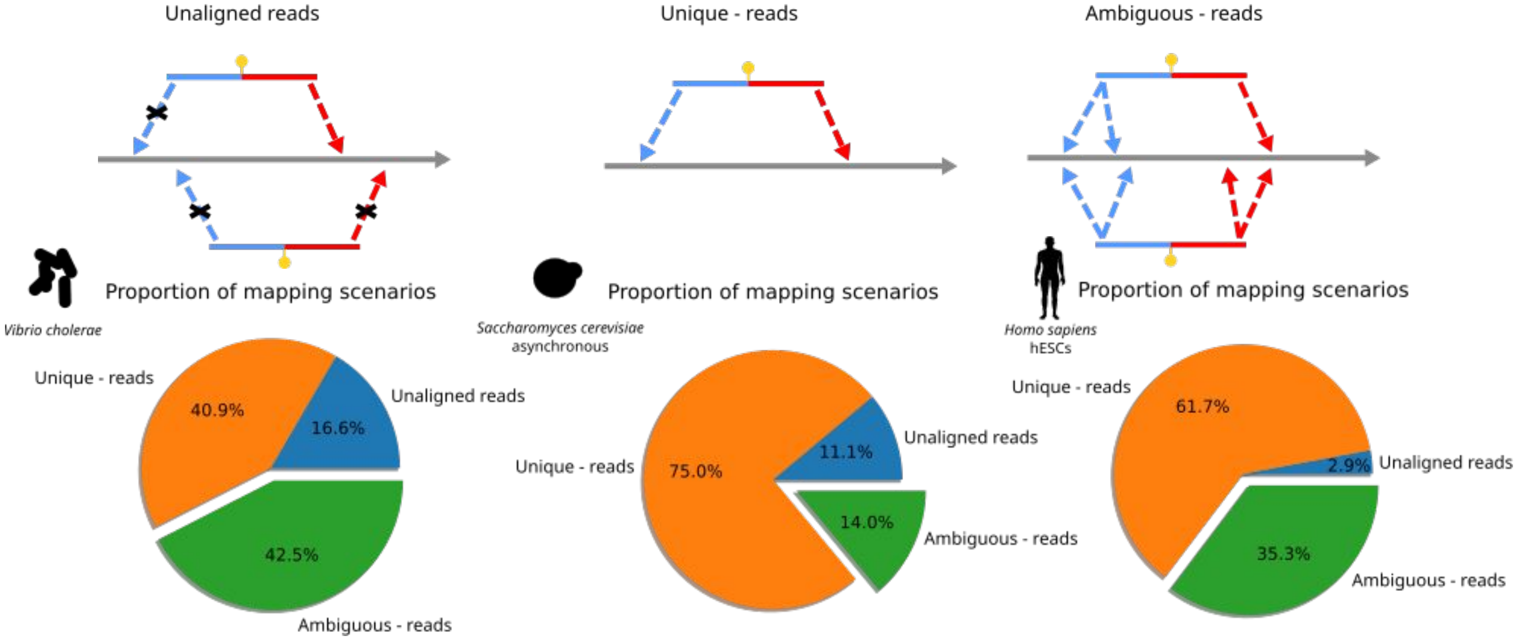
1. Lieberman-Aiden, E. et al. Comprehensive mapping of long range interactions reveals folding principles of the human genome. Science 326, 289–293 (2009).
2. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. Science 295, 1306–1311 (2002).

Repeated elements on Hi-C maps - significant loss of contacts information

Repeated sequences over genomes appear as **empty stripes** on Hi-C maps

Several possible **couples** for a chimere : ***how to choose the right one ?***

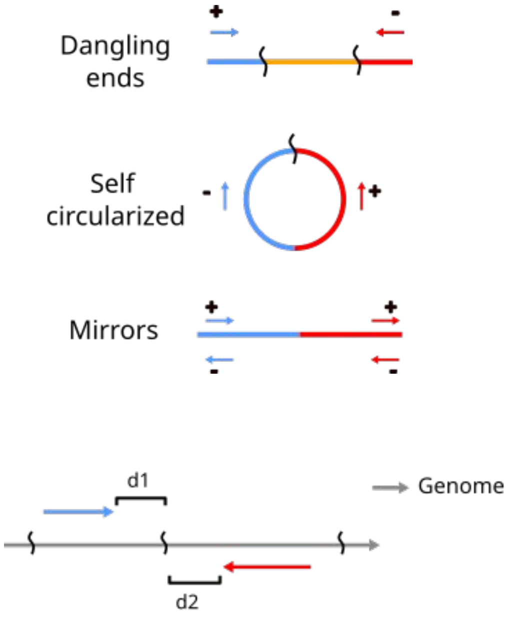
Ambiguous reads are usually **filtered out**

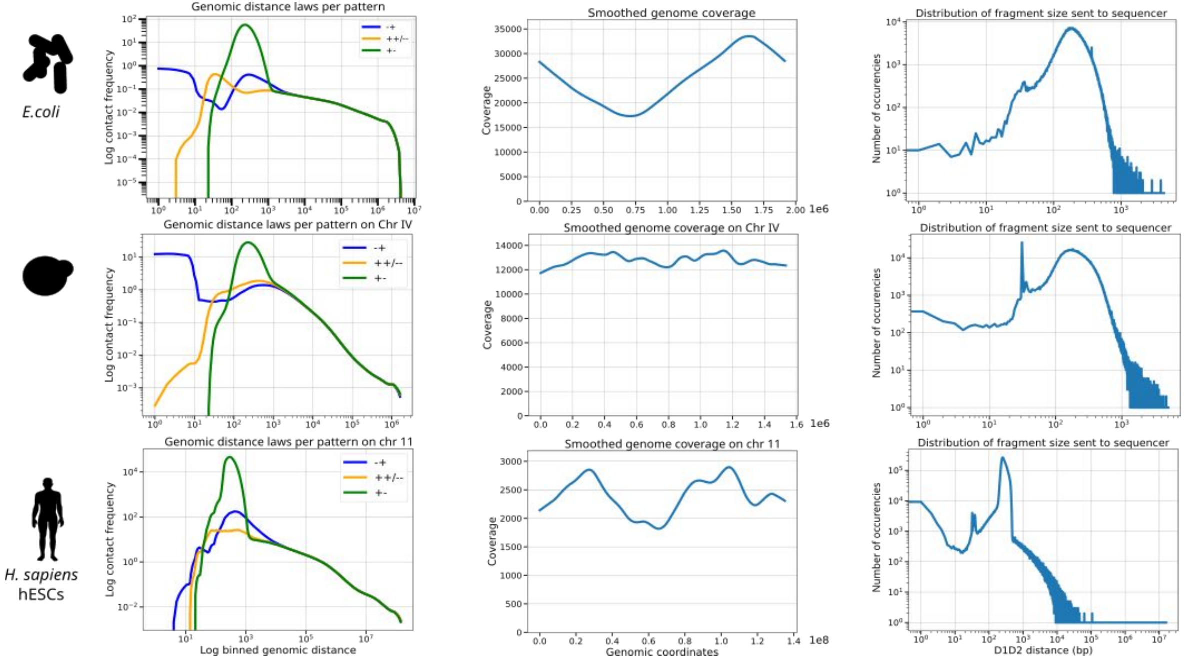


How to investigate genomes structures linked to repeated sequences and their impact ?

Leveraging Hi-C pair-end reads patterns to reduce ambiguity

How to unveil position of such ambiguous pairs corresponding to repeated elements ?

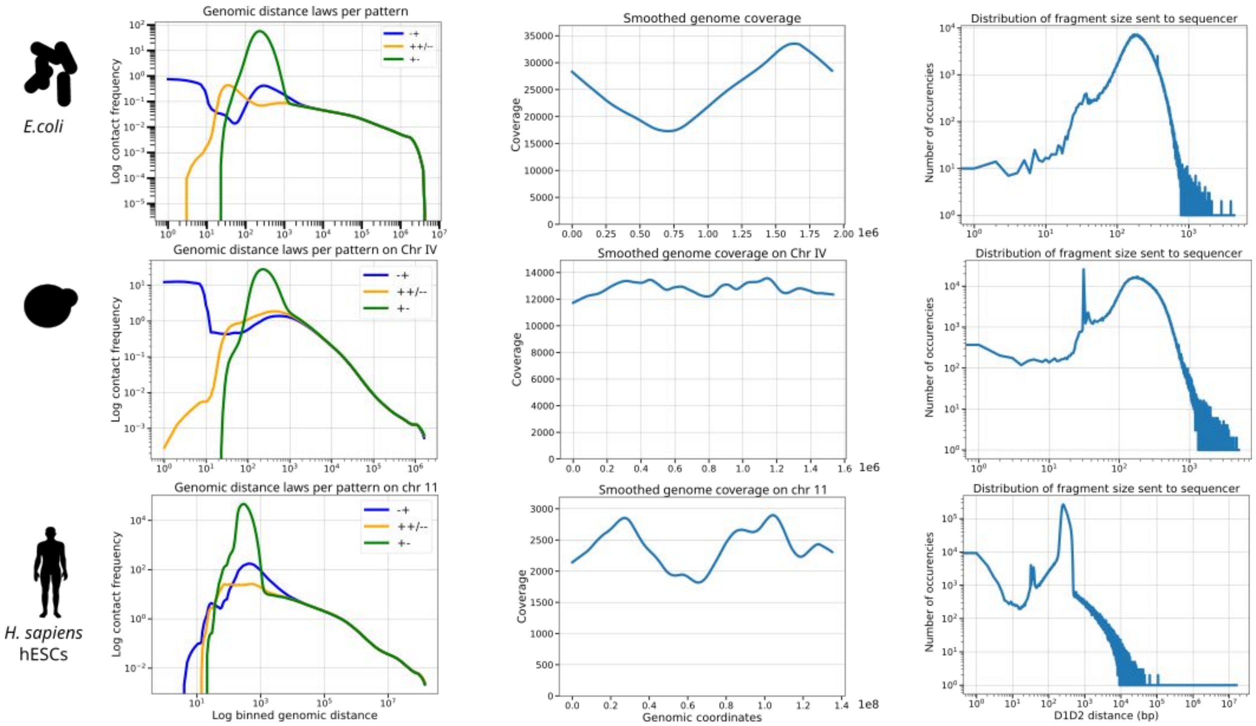




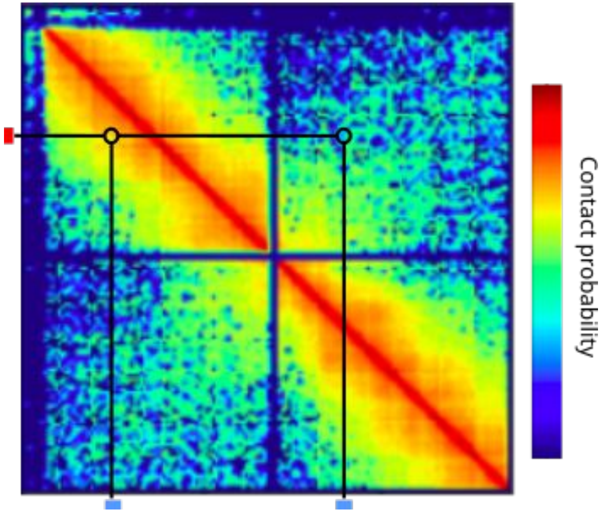
<u>Main idea</u> : extracting **statistical trends** from **unambiguous** chimeres to **infer couple selecting** among ambiguous read-pairs

Leveraging Hi-C pair-end reads patterns to reduce ambiguity

How to unveil position of such ambiguous pairs corresponding to repeated elements ?

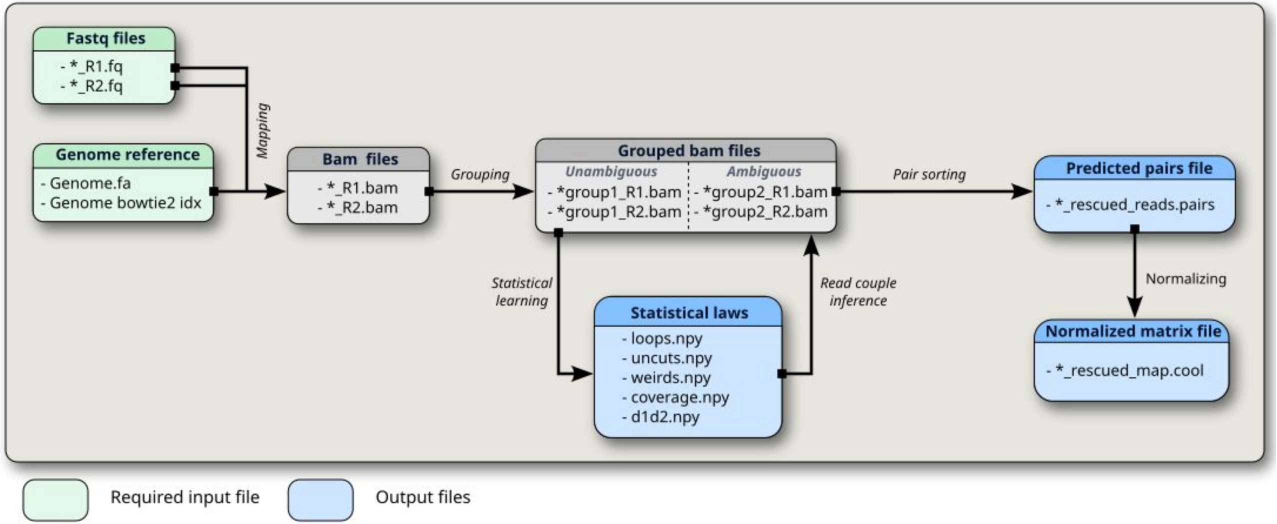


Contact probability map



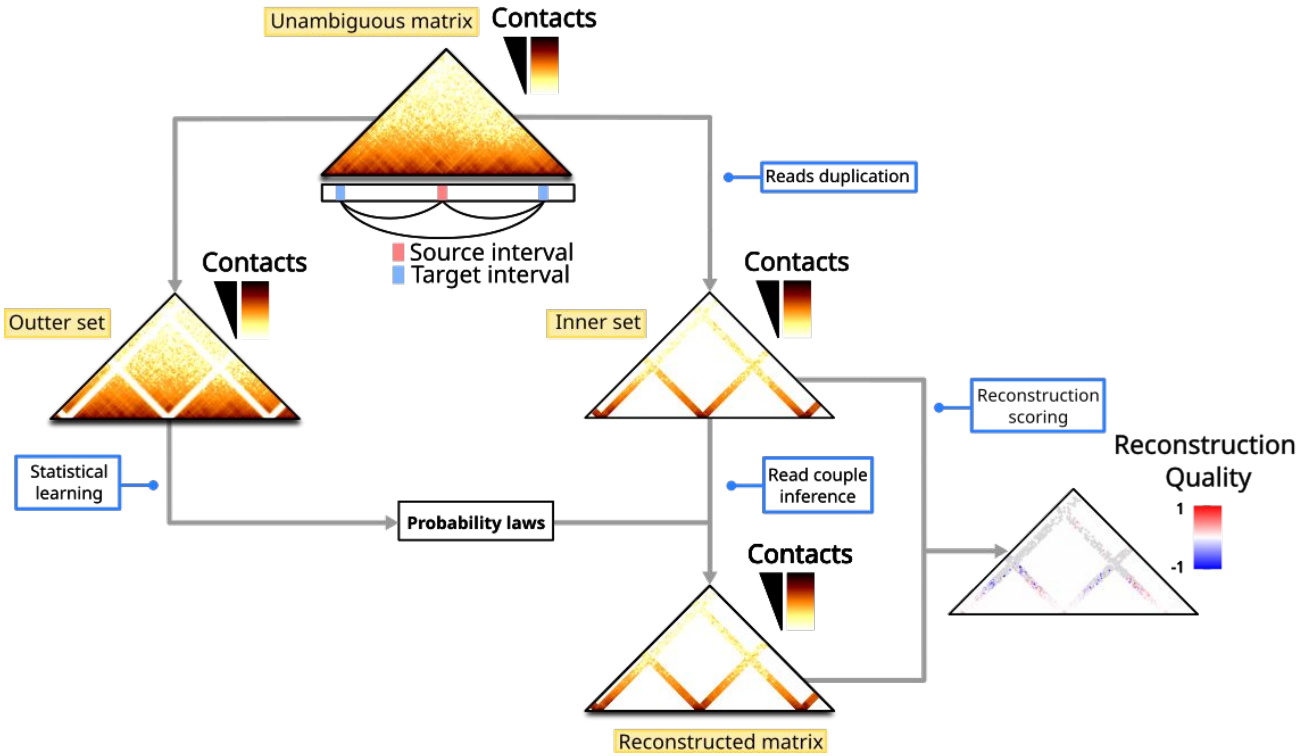
<u>Main idea</u> : extracting **statistical trends** from **unambiguous** chimeres to **infer couple selecting** among ambiguous chimeres

HiC-BERG - combine information to leverage multi-mapping reads



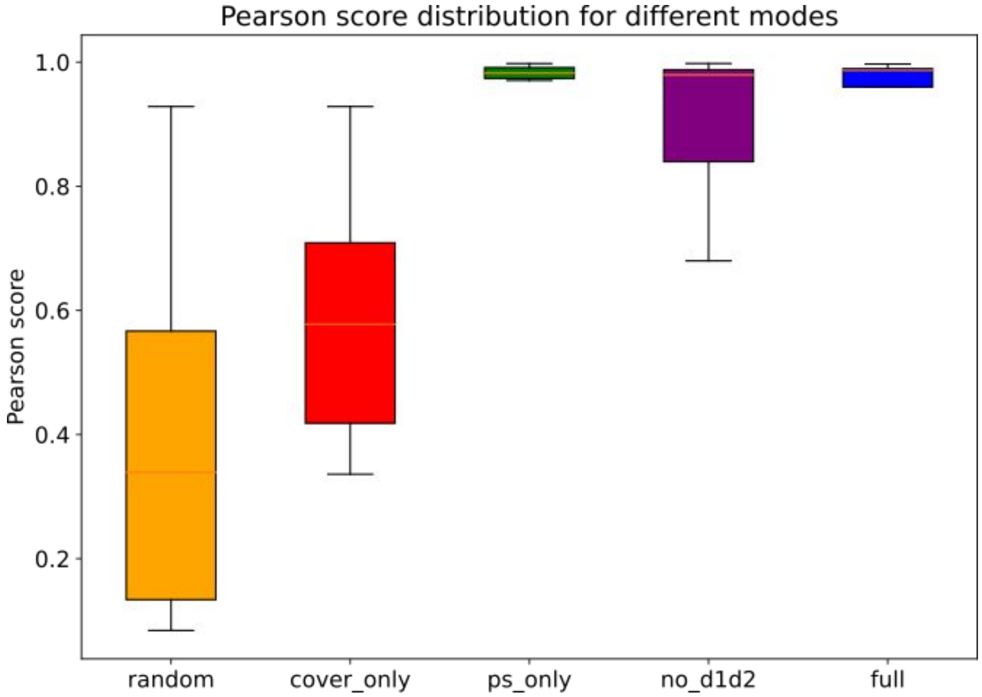
- HiC-BERG : **Python 3.11** package
- **Standard file formats** formats of HiC assays **injectable** at **any step**
- **Computation time** are **genome length** and **richness of repeated sequence** dependant : **40min** for *E.coli* and **2h** for *S.cerevisiae* on respectively 26 millions and 47 millions reads depth banks

HiC-BERG - Global validation strategy



Validation of HIC-BERG through **ambiguity injection** in non ambiguous part of contacts
→ **Controlled strategy** to evaluate Hi-C reconstruction

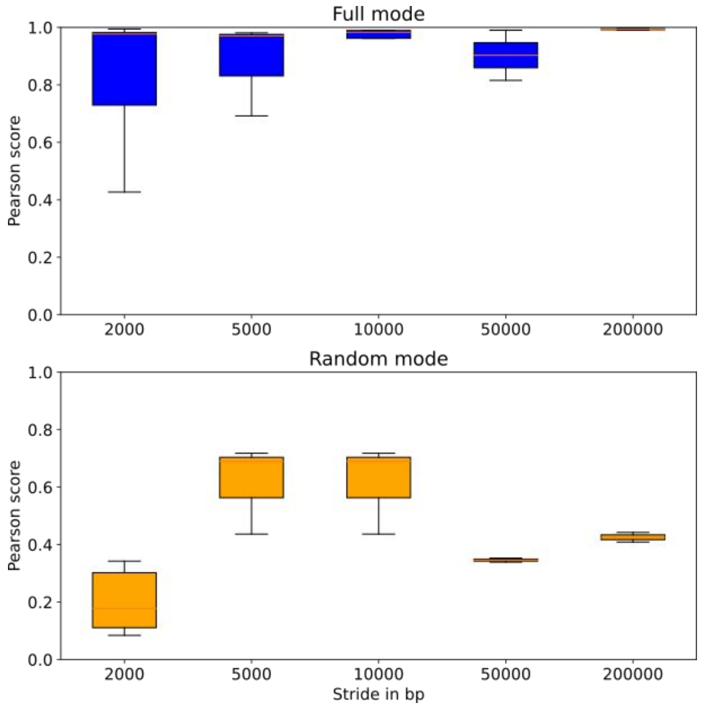
HiC-BERG - Global performances on *S. cerevisiae*

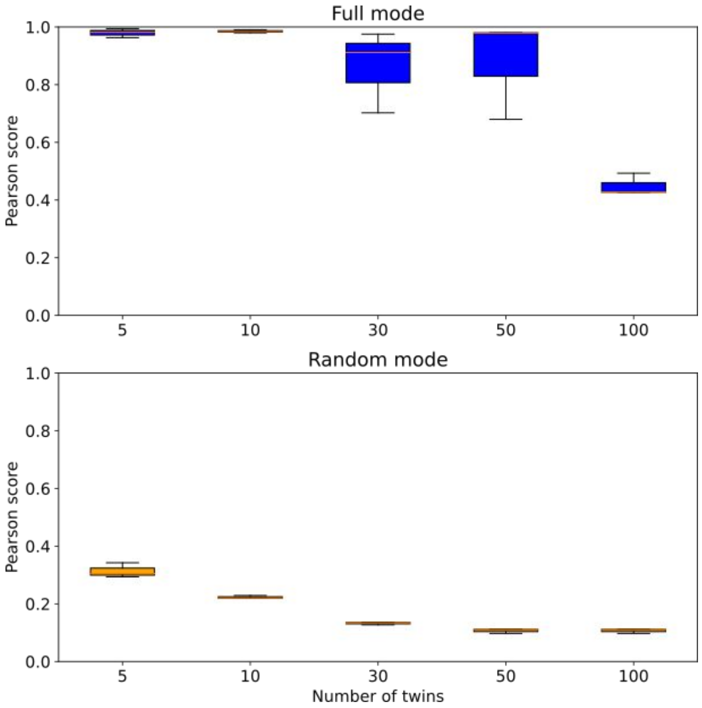


- **Random** model show a mean Pearson correlation ~ 0.4
- **P(s) carry** the **globality** of the information
- Adding **restriction sites information** (d1d2) **reduce dispersion** over reconstruction experiences

HiC-BERG – Global validation

Assessing reconstruction quality through **increasing genomic distance** between fakes (3 repeats on cis and trans chromosomal cases)

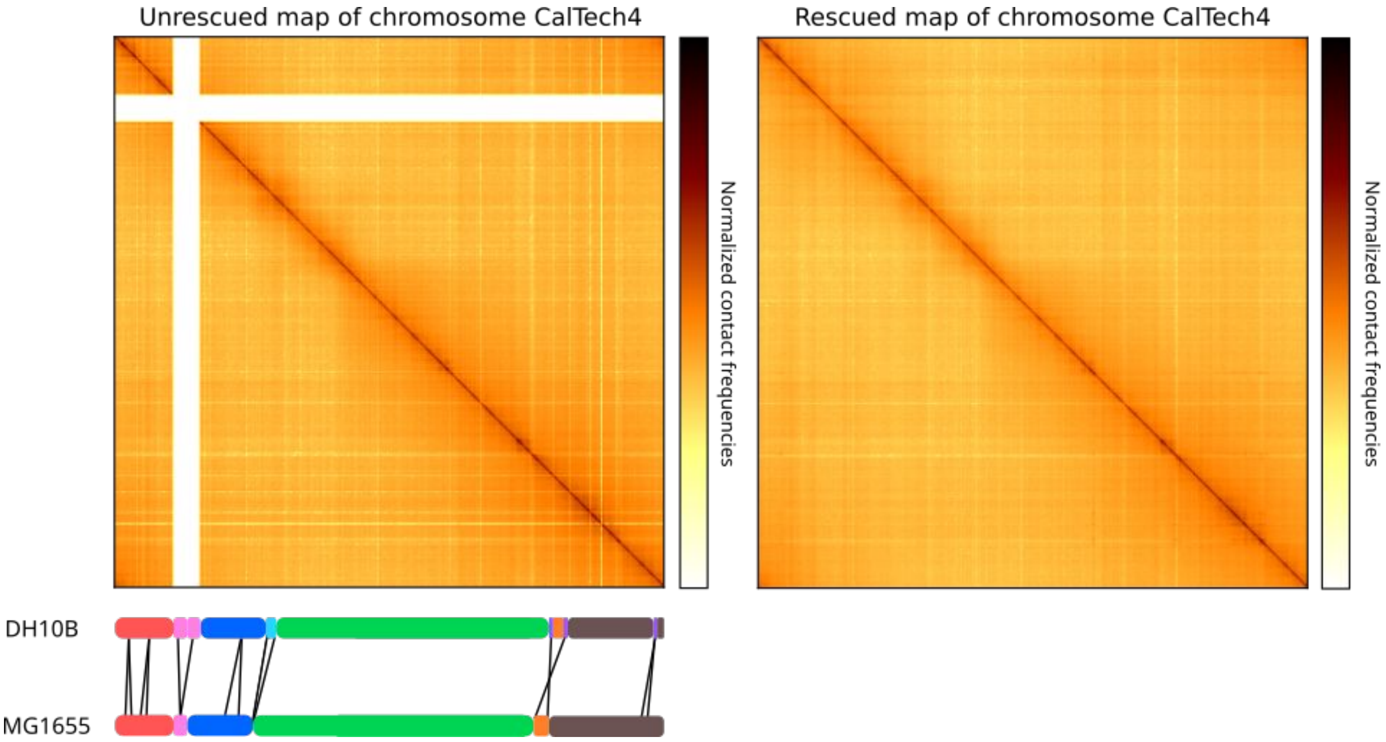




Genomic distance (bp)

HiC-BERG reconstructions appears to be **robust** over **distance between repeated** sequences and **twin repeated sequences number**

Application (I) *E. Coli* 100kb tandem duplication

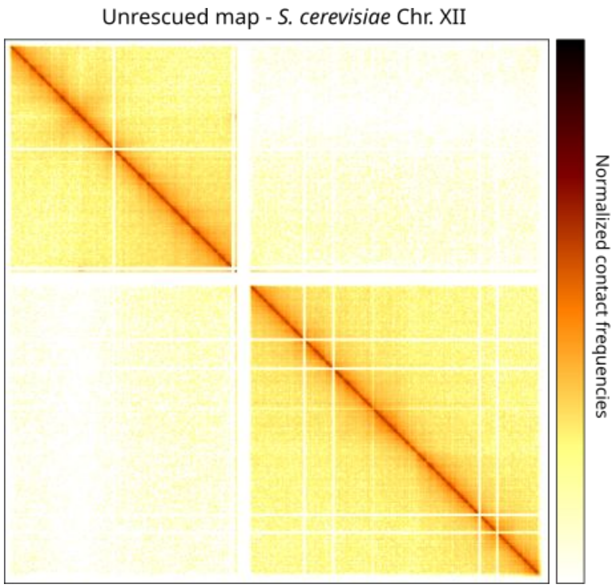


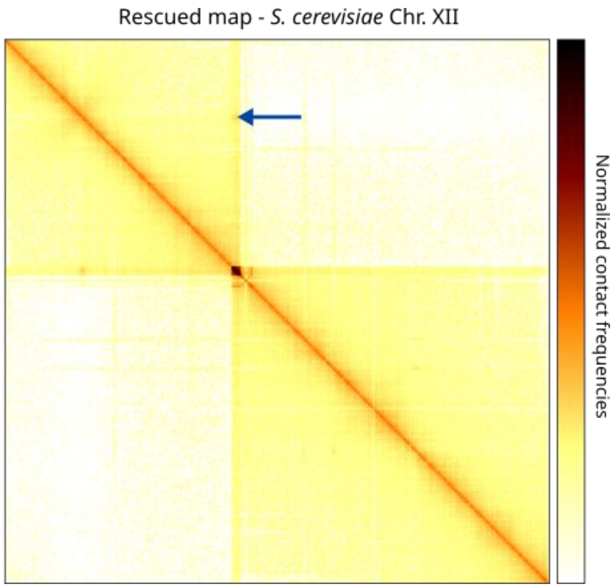
E. coli* DHB10** *strain is known to* ***carry a 100 kb tandem duplicated sequences

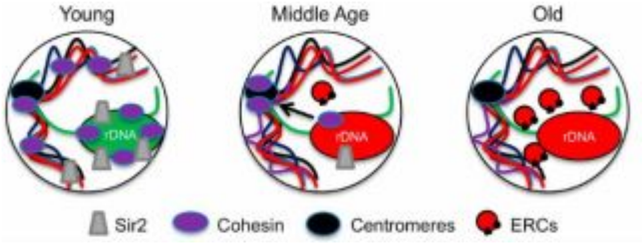
HiC-BERG inference retrieved expected **DNA polymer behavior**

Durfee T, & al. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. J Bacteriol. 2008

Application (II) *S. cerevisiae* rDNA CEN XII


Unrescued map - *S. cerevisiae* Chr. XII

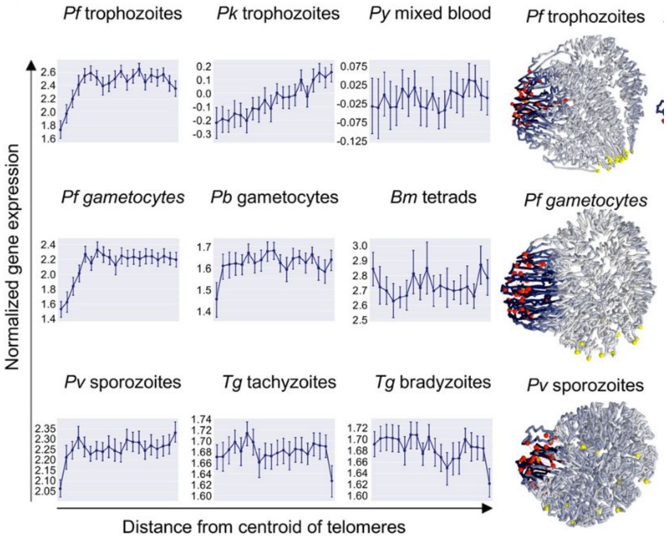

Rescued map - *S. cerevisiae* Chr. XII

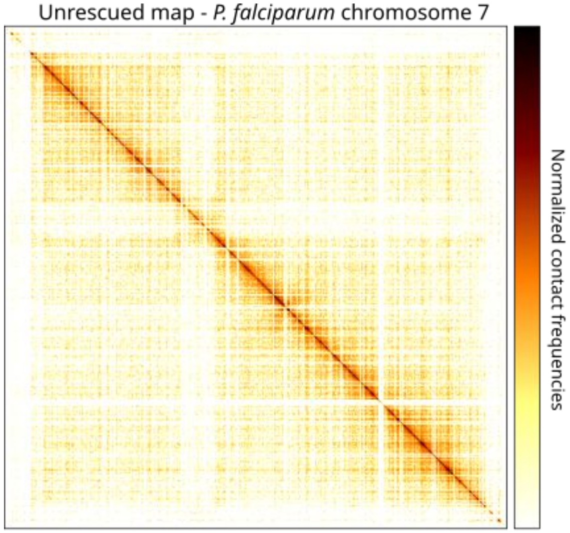


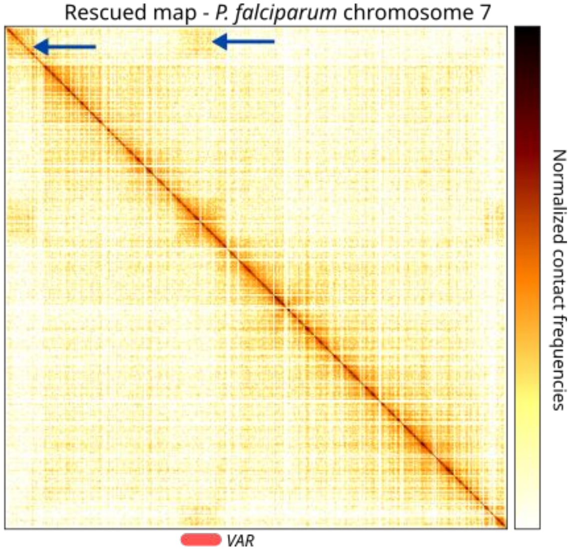
Retrieval of **rDNA complex (200+ copies) interaction** with **centromere** of chromosome XII

Fine, R. D., Maqani, N., Li, M., Franck, E. & Smith, J. S. Depletion of Limiting rDNA Structural Complexes Triggers Chromosomal Instability and Replicative Aging of Saccharomyces cerevisiae. Genetics 212, 75 (2019).

Application (III) *P. falciparum* VAR region clustering







P .falciparum observations showed **VAR gene** and **telomere clusterization** to **repress virulence** expression during parasitic cycle for **host immune system escaping**

Unveil of (sub) **- telomere** / **virulence genes** interactions **previously characterized**

Bunnik, E. M. et al. Comparative 3D genome organization in apicomplexan parasites. Proceedings of the National Academy of Sciences 116, 3183–3192 (2019).