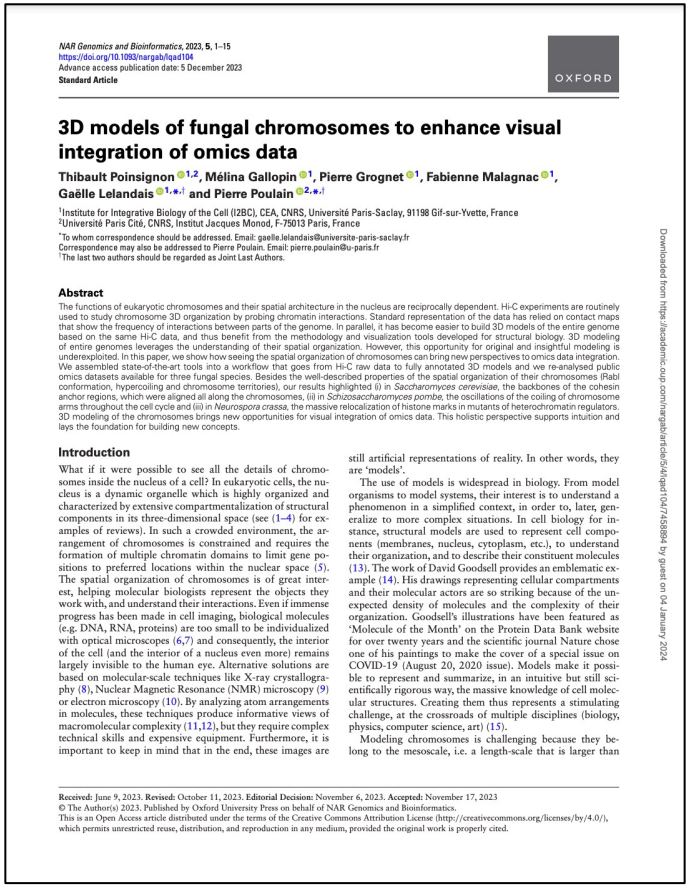
3D models of fungal chromosomes to enhance visual integration of omics data

Thibault Poinsignon, Mélina Gallopin, Pierre Grognet, Fabienne Malagnac, **Gaëlle Lelandais*** and Pierre Poulain*

New article published in NAR Genomics and Bioinformatics

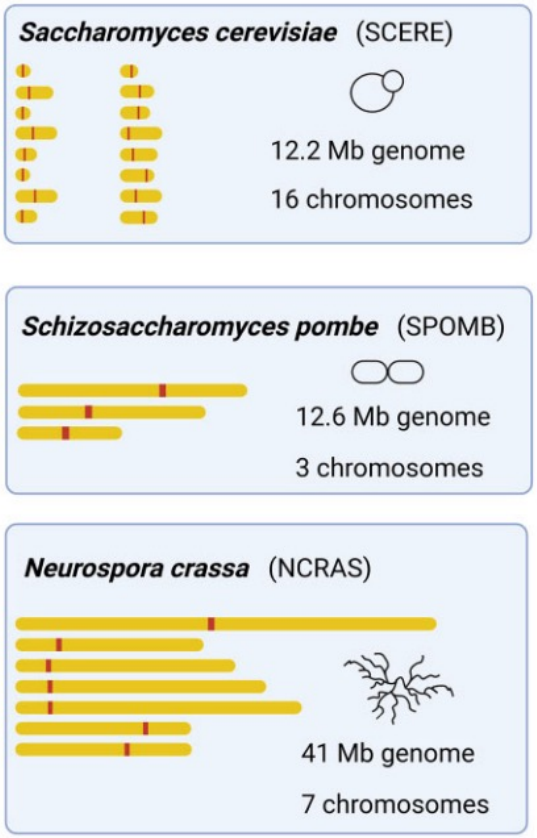


https://academic.oup.com/nargab/article/5/4/lqad104/7458894

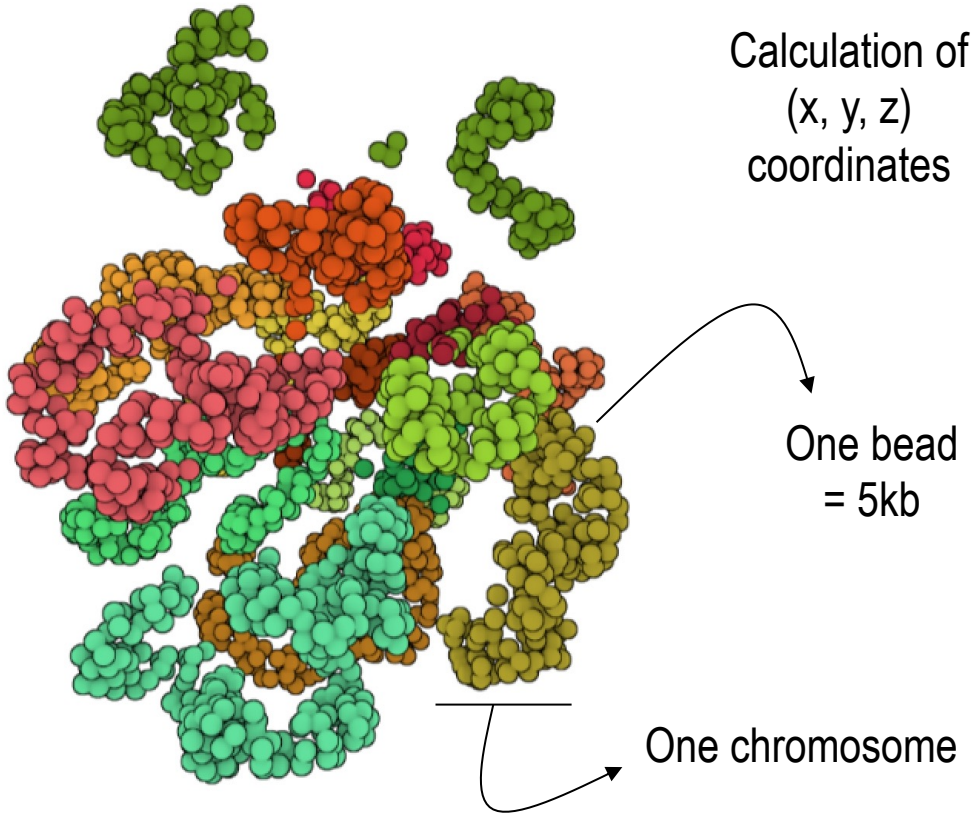
https://doi.org/10.1093/nargab/lqad104

3D fungal chromosome models from public HiC data

- Three different species:



- Example of SCERE genome:
  (HiC data from public database)



Poinsignon et al. (2023)

Methods to reconstruct 3D models of genome organization

- In the last decade, several strategies have been presented in the literature (reviewed in Oluwadare et al., 2019).
- In this work, we chose to use the Pastis-NB method, which is a probability-based method, recently published (Varoquaux et al., 2023).

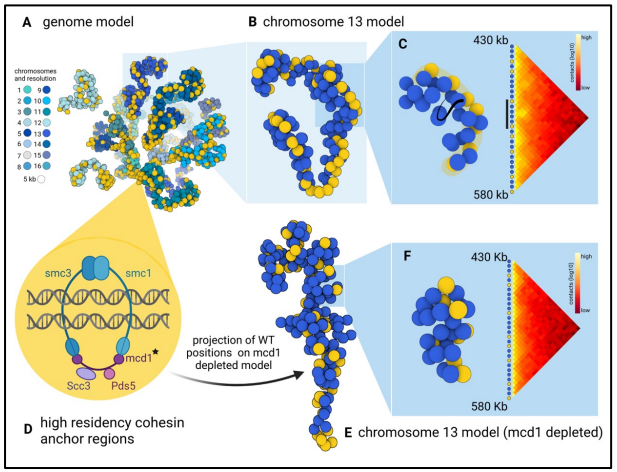
- We explored the potential of using 3D models as complementary information to Hi-C contact maps, to better understand the spatial organization of fungal chromosomes and to improve the visual integration of omics data.

Three examples are presented in the article

Detailed view of chromosome organization → General view of chromosome organization

Poinsignon et al. (2023)

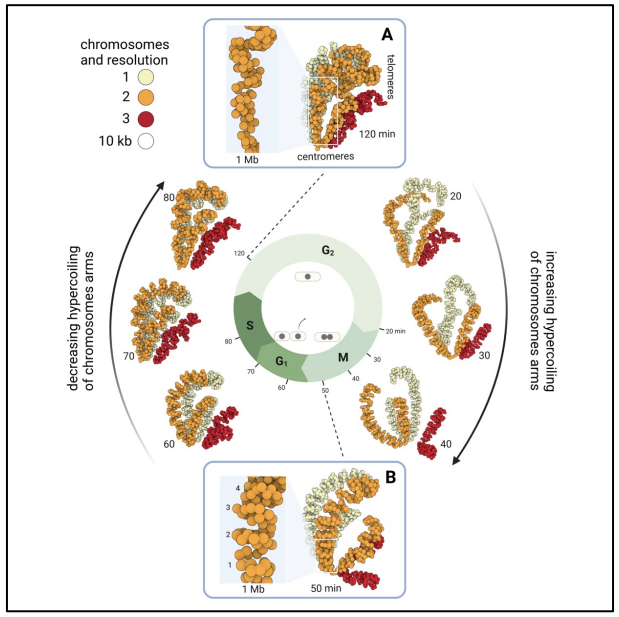
➢ SCERE



Example #1: Spatial alignment of cohesin binding sites along chromosomes in *S. cerevisiae.*

HiC (WT and mutant) + ChIPseq

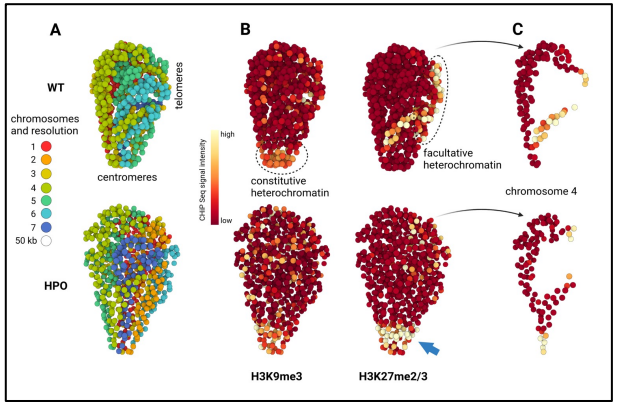
➢ SPOMBE



Example #2: Oscillations in coiling of chromosome arms during the cell cycle in *S. pombe.*

HiC (WT only, at several time points)

➢ NCRASSA



Example #3: Massive relocalisation of histone marks in *N. crassa* heterochromatin regulator mutants.

HiC (WT and mutant) + ChIPseq

Re-analysis of a large omics dataset

- More than 40 FASTQ files were collected:

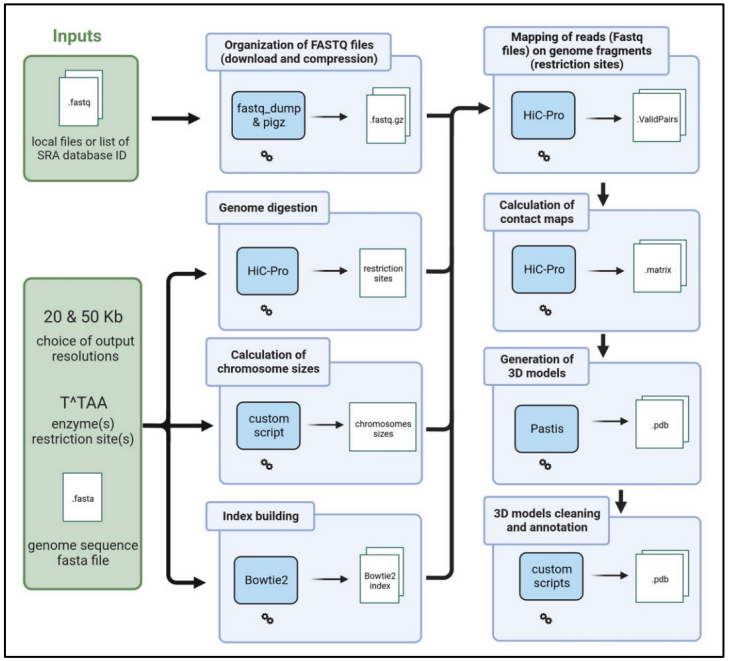
Table 1. Main characteristics of raw data used in this study to create 3D models of fungal chromosomes

Model	Fungal species (genome version)	Strain	Hi-C data sources	SRA IDs	Final number of valid read pairs	Number of beads in the final model	Figure (model resolution in bp)
1	*N. crassa (nc14)*	WT	Rodriguez *et al.* 2022 (31)	SRR14362684 SRR14362685 SRR16761088 SRR16761089 SRR16761090 SRR16761091 SRR16761092	107 636 304	3976	2 (10 000)
2		WT	Galazka *et al.* 2016 (30)	SRR2105869 SRR2105870 SRR2105871 SRR2105872	5 030 121	800	5 (50 000)
3		hpo		SRR2105876 SRR2105877 SRR2105878	2 922 526	808	
4	*S. pombe* (ASM294v2.19)	WT	Tanizawa *et al.* 2017 (27)	SRR5149251 SRR5149252	15 511 107	1223	4 (10 000)
5				SRR5149253 SRR5149254	14 377 600	1211	
6				SRR5149255 SRR5149256	14 394 280	1204	
7				SRR5149257 SRR5149258	15 497 423	1200	
8				SRR5149259 SRR5149260 SRR5942526	19 035 344	1213	
9				SRR5149261 SRR5149263	13 841 928	1224	
10				SRR5149264 SRR5149265 SRR5942527	21 936 238	1201	
11				SRR5149266 SRR5149267 SRR5942528	21 727 635	1189	
12	*S. cerevisiae (R64, S288C)*	WT	Costantino *et al.* 2020 (64)	SRR11893084 SRR11893085	16 847 912	2332	2 and 3 (5000)
13		Mcd1 depleted		SRR11893086 SRR11893087	14 180 445	2323	

The sources of Hi-C data (original articles and SRA identifiers) are given. All datasets were obtained on wild type strains, with the exception of models #3 and #13 which correspond to the *N. crassa* hpo mutant (used in Figure 5) and *S. cerevisiae* mcd1 depleted (used in Figure 3). The number of valid pairs of reads gives an estimation of the overall quality of the data and indicates the density of the contact frequencies measured in the experiment. The number of beads in the 3D model is determined by the chosen resolution during Hi-C data analysis and the length of the reference genome.

Poinsignon et al. (2023)

- Standardized bioinformatics protocols for the creation of 3D models:



Workflow 3D genome builder (3DGB)

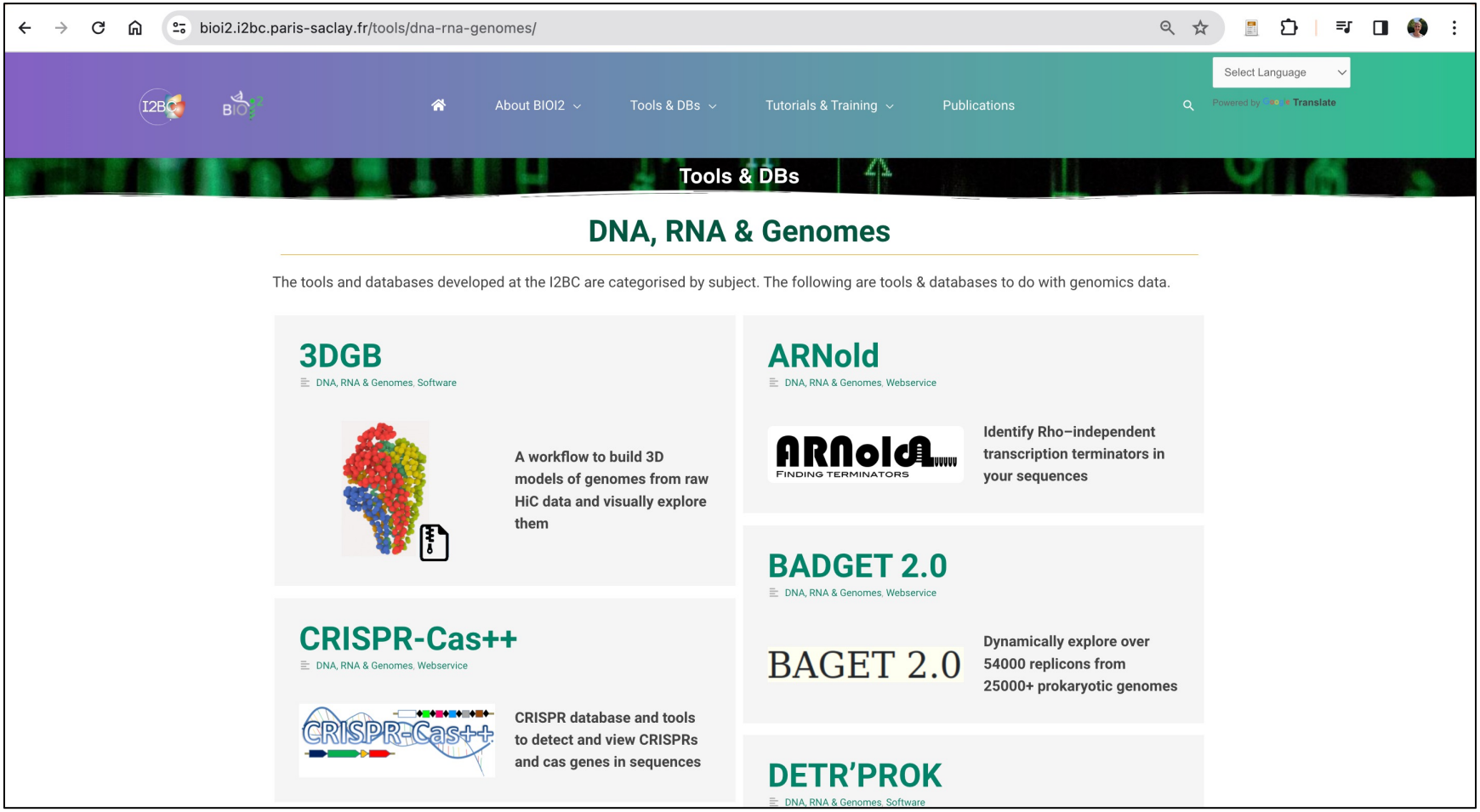
The workflow 3DGB

- It automatically performs the critical bioinformatics steps required to:
  - (i) compute Hi-C contact frequencies,
  - (ii) infer associated 3D models of the chromatin organization,
  - (iii) annotate and control the quality of the 3D models.

- It also adds further processing of the 3D model output as PDB files, suitable for advanced visualization with molecular viewer software (Mol*, PyMol, etc.).

- It is available to other scientists who whish to add 3D models to their analyses of Hi-C data.
  - It requires only four basic inputs to be specified (FASTQ files, a FASTA file, restriction sites, targeted resolutions).

A link is available from the Integrative Bioinformatics platform of the I2BC



https://bioi2.i2bc.paris-saclay.fr/tools/dna-rna-genomes/

Contact : gaelle.lelandais@universite-paris-saclay.fr

THANK YOU FOR YOUR ATTENTION