Structure determination of genomes and genomic domains by satisfaction of spatial restraints

- Model assessment
- Mycoplasma 3D models

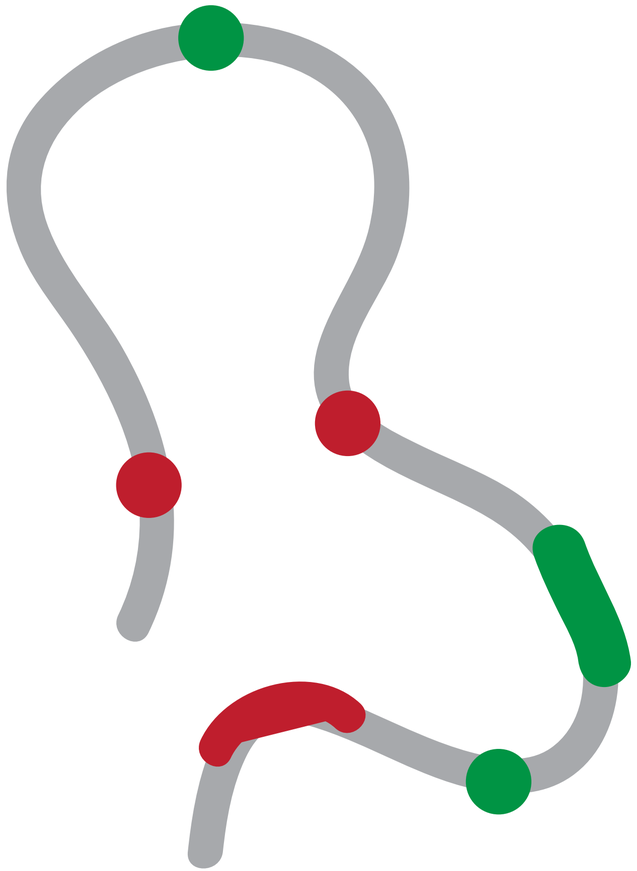
Marc A. Marti-Renom
Structural Genomics Group (ICREA, CNAG-CRG)

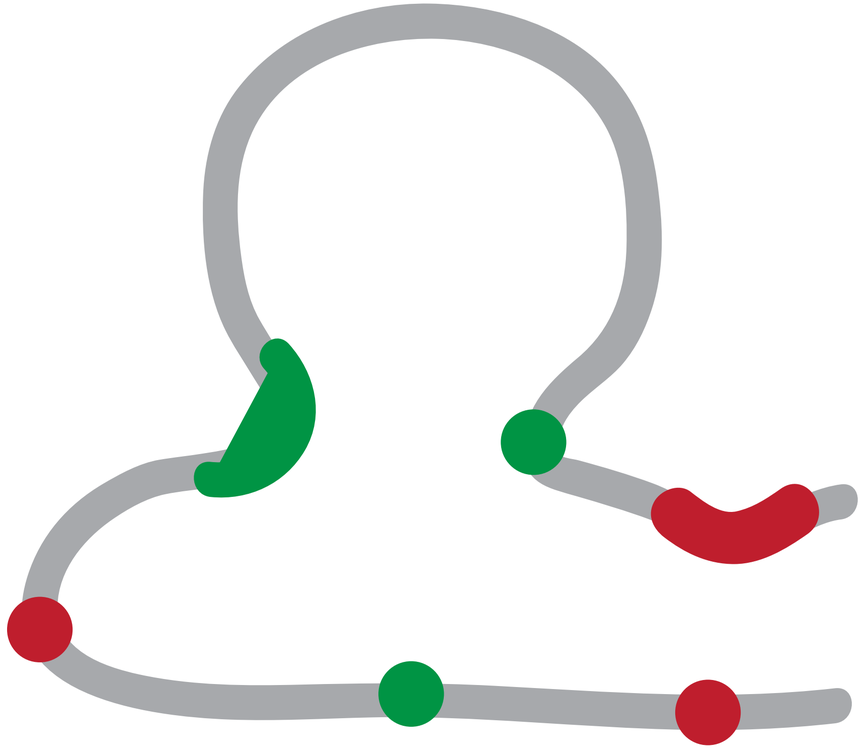
Marie Trussart (PhD)

Serrano and Marti-Renom Labs @CRG
Now postdoc @The Walter and Elisa Institute. Melbourne, Australia.

http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu

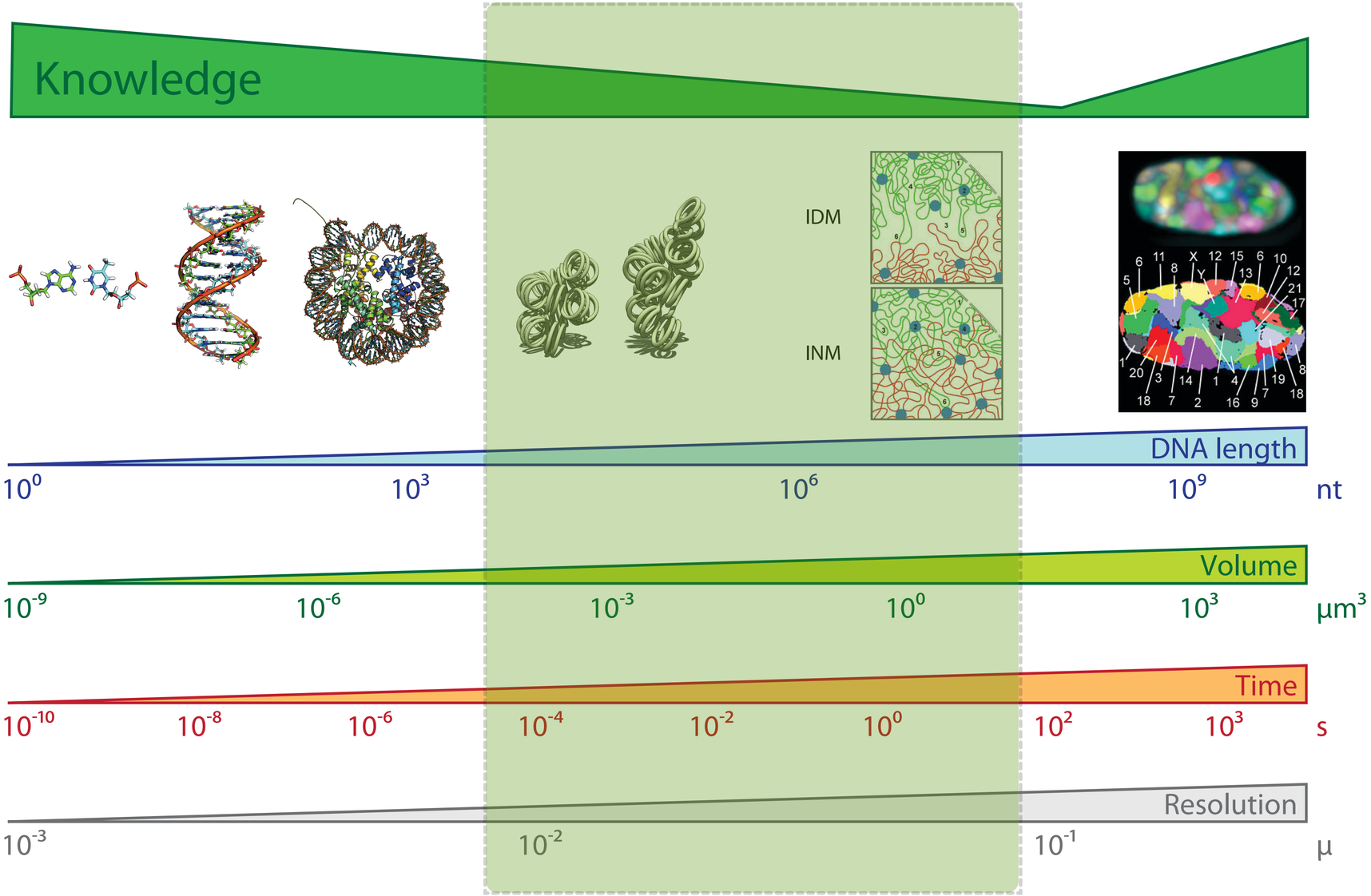
cnag CRG ICREA





Resolution Gap

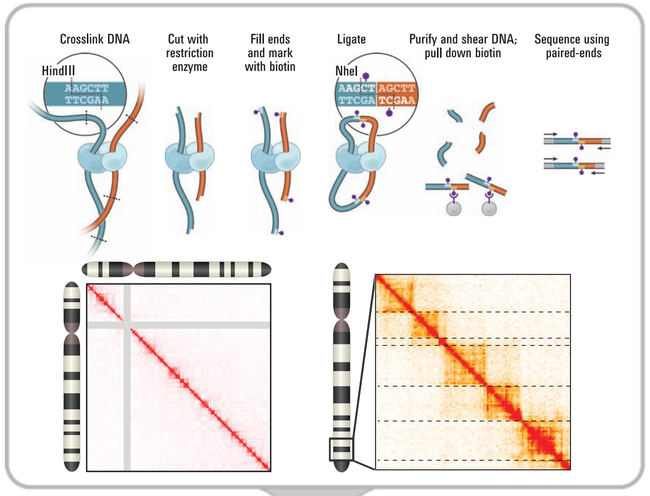
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)

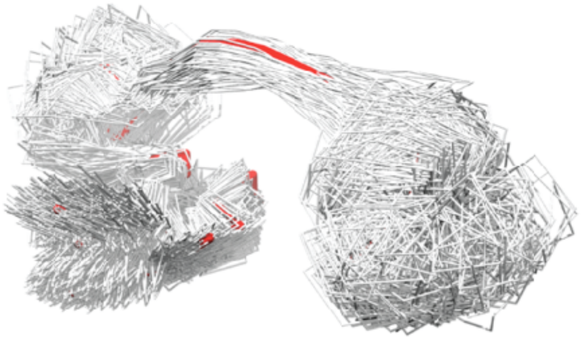


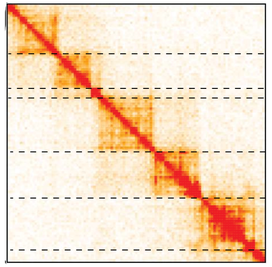
Hybrid Method

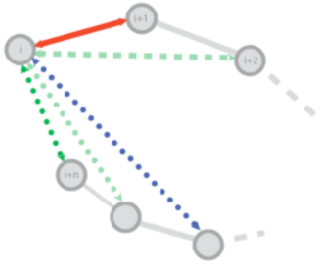
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

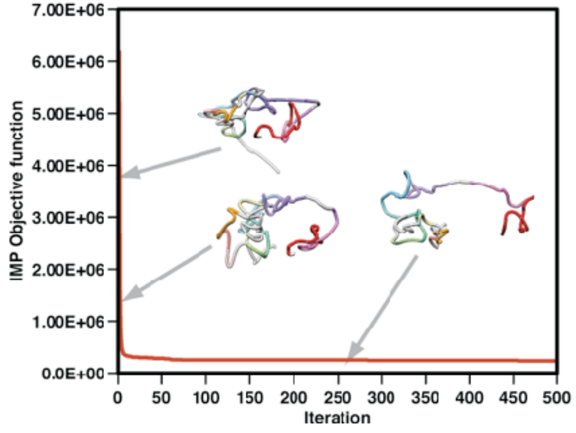
Experiments







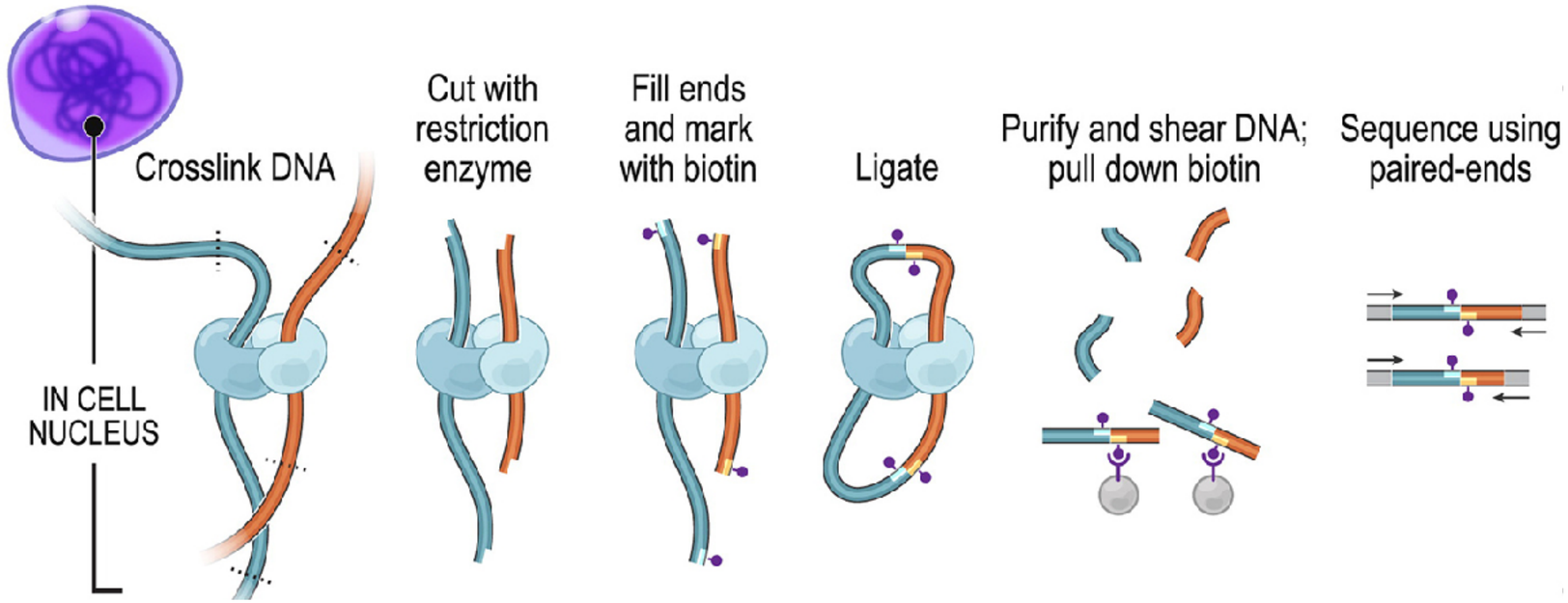


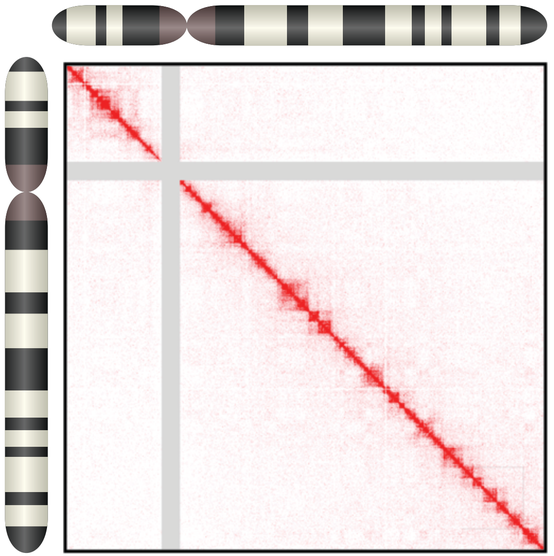


Computation

Chromosome Conformation Capture

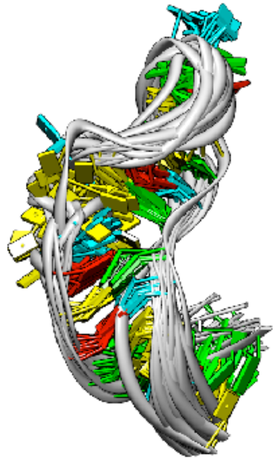
Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Science, 295(5558), 1306–1311.
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.

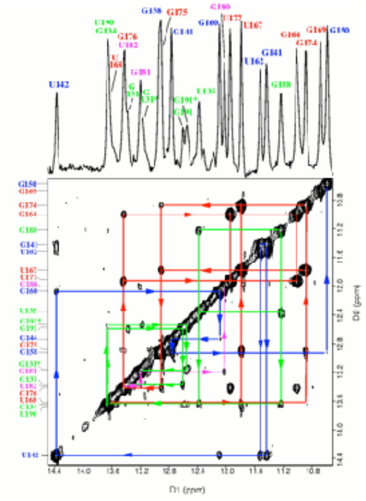




Restraint-based Modeling

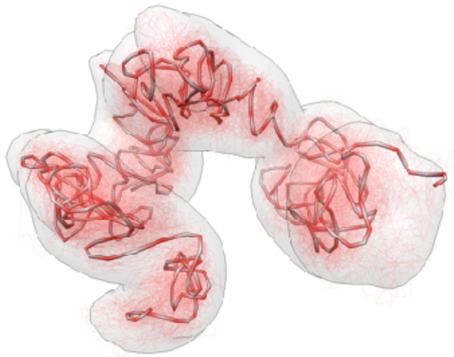
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

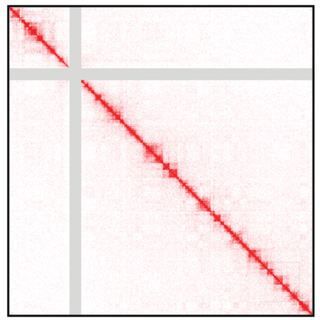




Biomolecular structure determination

2D-NOESY data





Chromosome structure determination

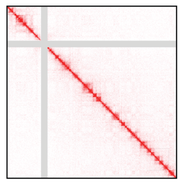
3C-based data

TADbit

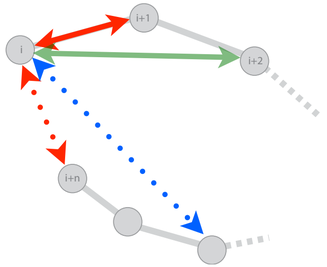
http://3DGenomes.org

Serra et al. BioRxiv (2016)

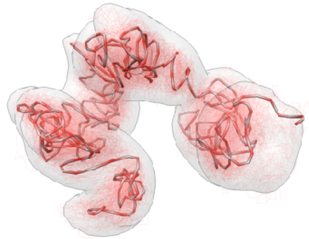
FastQ files to Maps



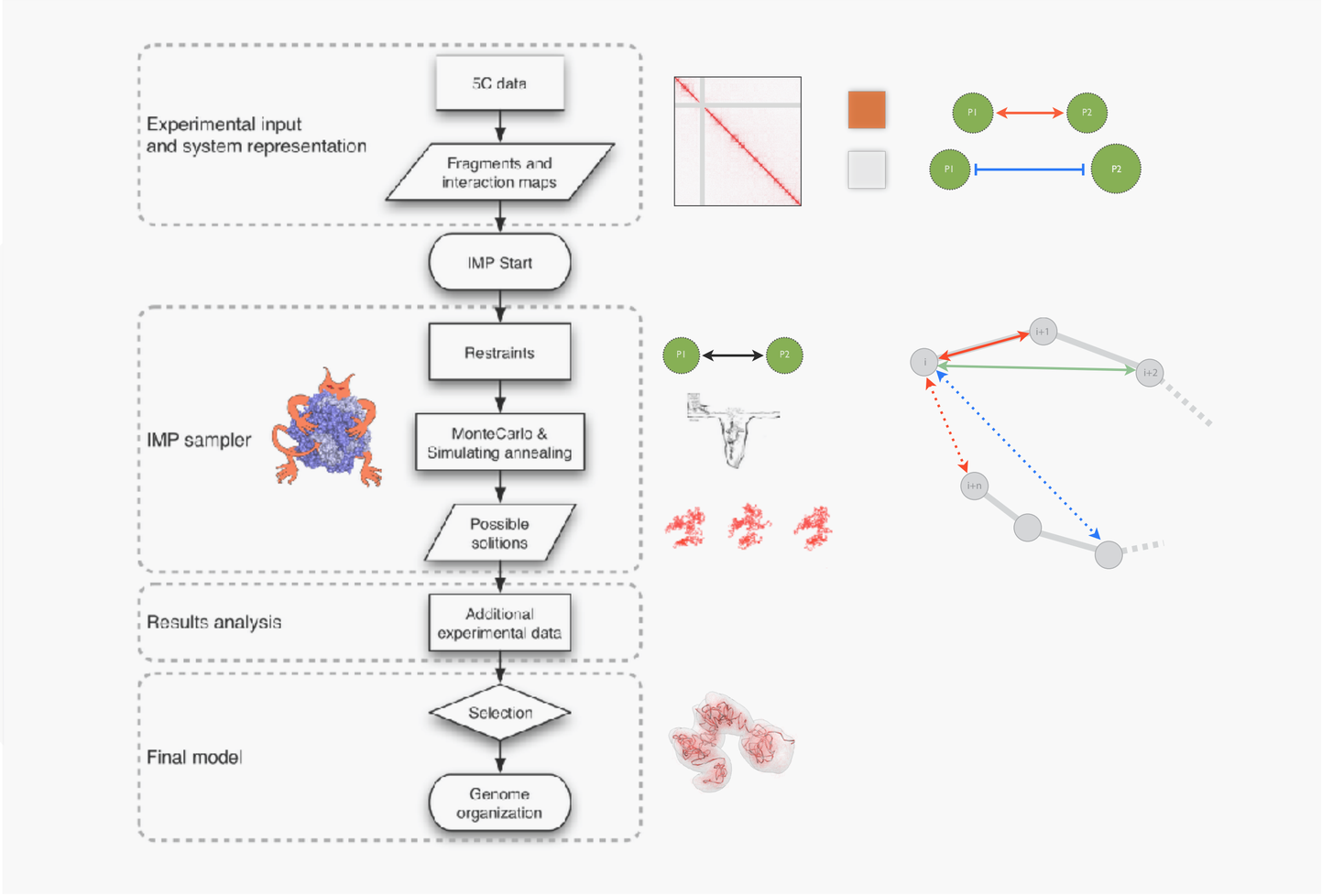
Map analysis



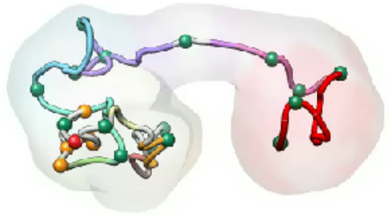
Model building



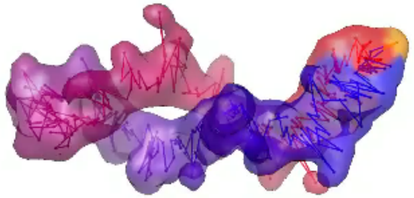
Model analysis



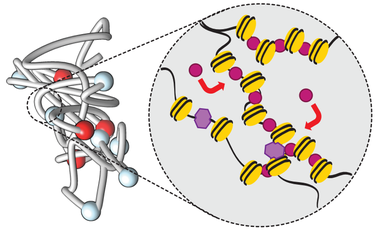
TADbit previous applications…



Baù, D. et al. Nat Struct Mol Biol (2011)



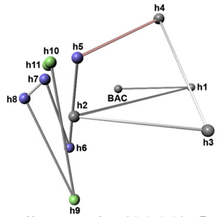
Umbarger, M. A. et al. Mol Cell (2011)



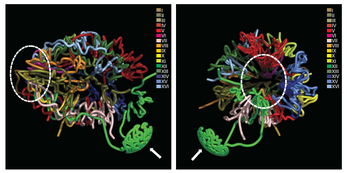
Le Dily, F. et al. Genes & Dev (2014)

Are the models correct?

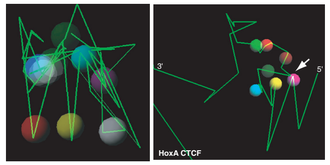
Trussart et al. NAR (2015)



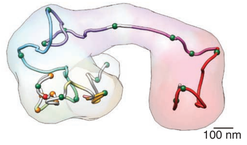
Jhunjhunwala (2008) Cell



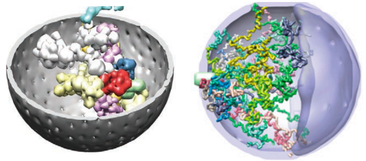
Duan (2010) Nature



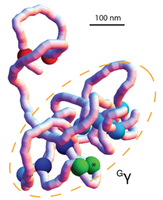
Fraser (2009) Genome Biology
Ferraiuolo (2010) Nucleic Acids Research



Baù (2011) Nature Structural & Molecular Biology

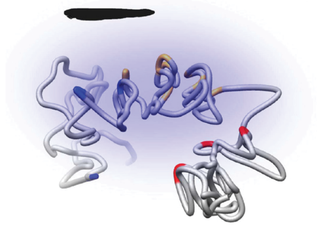


Kalhor (2011) Nature Biotechnology
Tjong (2012) Genome Research



Junier (2012) Nucleic Acids Research

Giorgetti, (2014) Cell



Acemel (2016) Nature Genetics

Nucleic Acids Research Advance Access published March 23, 2015

Nucleic Acids Research, 2015 **1**
doi: 10.1093/nar/gkv221

Assessing the limits of restraint-based 3D modeling of genomes and genomic domains

Marie Trussart[1,2], François Serra[3,4], Davide Baù[3,4], Ivan Junier[2,3], Luís Serrano[1,2,5] and Marc A. Marti-Renom[3,4,5,*]

[1]EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Barcelona, Spain, [2]Universitat Pompeu Fabra (UPF), Barcelona, Spain, [3]Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Barcelona, Spain, [4]Genome Biology Group, Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Spain and [5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Received January 16, 2015; Revised February 16, 2015; Accepted February 22, 2015

ABSTRACT

Restraint-based modeling of genomes has been recently explored with the advent of Chromosome Conformation Capture (3C-based) experiments. We previously developed a reconstruction method to resolve the 3D architecture of both prokaryotic and eukaryotic genomes using 3C-based data. These models were congruent with fluorescent imaging validation. However, the limits of such methods have not systematically been assessed. Here we propose the first evaluation of a mean-field restraint-based reconstruction of genomes by considering diverse chromosome architectures and different levels of data noise and structural variability. The results show that: first, current scoring functions for 3D reconstruction correlate with the accuracy of the models; second, reconstructed models are robust to noise but sensitive to structural variability; third, the local structure organization of genomes, such as Topologically Associating Domains, results in more accurate models; fourth, to a certain extent, the models capture the intrinsic structural variability in the input matrices and fifth, the accuracy of the models can be *a priori* predicted by analyzing the properties of the interaction matrices. In summary, our work provides a systematic analysis of the limitations of a mean-field restrain-based method, which could be taken into consideration in further development of methods as well as their applications.

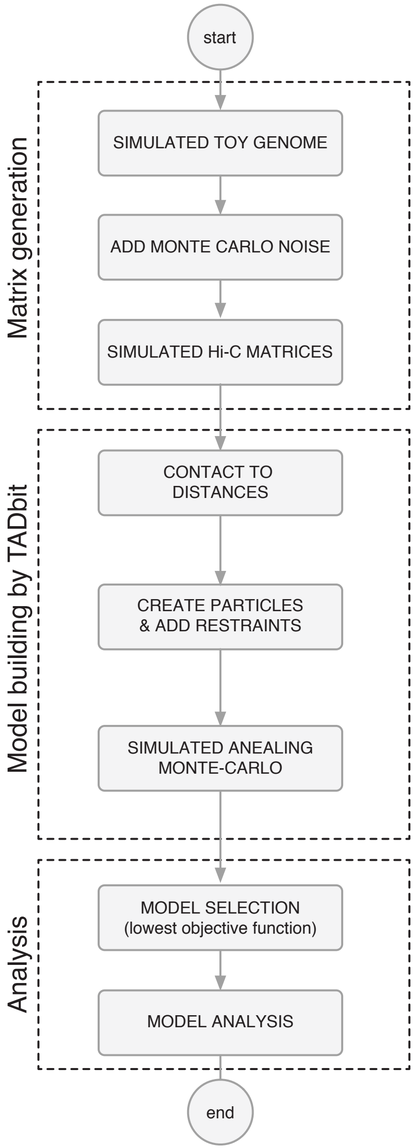
INTRODUCTION

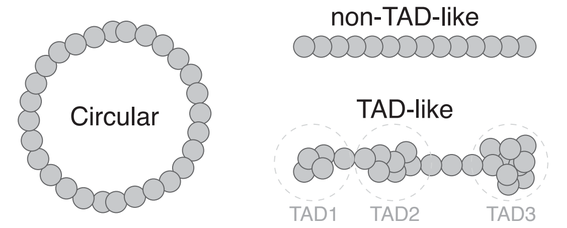
Recent studies of the three-dimensional (3D) conformation of genomes are revealing insights into the organization and the regulation of biological processes, such as gene expression regulation and replication (1–6). The advent of the so-called Chromosome Conformation Capture (3C) assays (7), which allowed identifying chromatin-looping interactions between pairs of loci, helped deciphering some of the key elements organizing the genomes. High-throughput derivations of genome-wide 3C-based assays were established with Hi-C technologies (8) for an unbiased identification of chromatin interactions. The resulting genome interaction matrices from Hi-C experiments have been extensively used for computationally analyzing the organization of genomes and genomic domains (5). In particular, a significant number of new approaches for modeling the 3D organization of genomes have recently flourished (9–14). The main goal of such approaches is to provide an accurate 3D representation of the bi-dimensional interaction matrices, which can then be more easily explored to extract biological insights. One type of methods for building 3D models from interaction matrices relies on the existence of a limited number of conformational states in the cell. Such methods are regarded as mean-field approaches and are able to capture, to a certain degree, the structural variability around these mean structures (15).

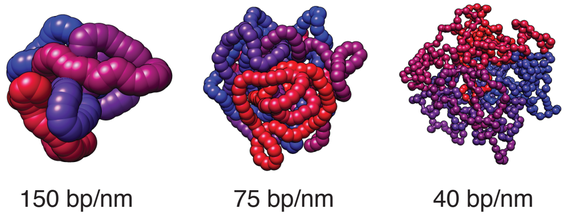
We recently developed a mean-field method for modeling 3D structures of genomes and genomic domains based on 3C interaction data (9). Our approach, called TADbit, was developed around the Integrative Modeling Platform (IMP, http://integrativemodeing.org), a general framework for restraint-based modeling of 3D bio-molecular structures (16). Briefly, our method uses chromatin interaction frequencies derived from experiments as a proxy of spatial proximity between the ligation products of the 3C libraries. Two fragments of DNA that interact with high frequency are dynamically placed close in space in our models while two fragments that do not interact as often will be kept apart. Our method has been successfully applied to model the structures of genomes and genomic domains in eukaryote and prokaryote organisms (17–19). In all of our studies, the final models were partially validated by assessing their

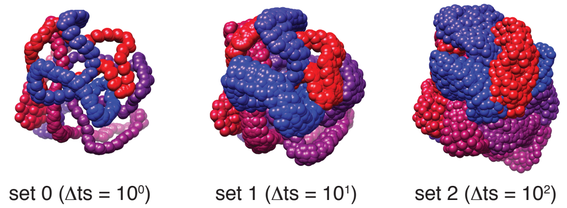
*To whom correspondence should be addressed. Tel: +34 934 020 542; Fax: +34 934 037 279; Email: mmarti@pcb.ub.cat

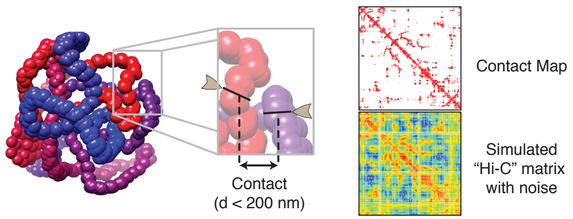
Toy models (168)





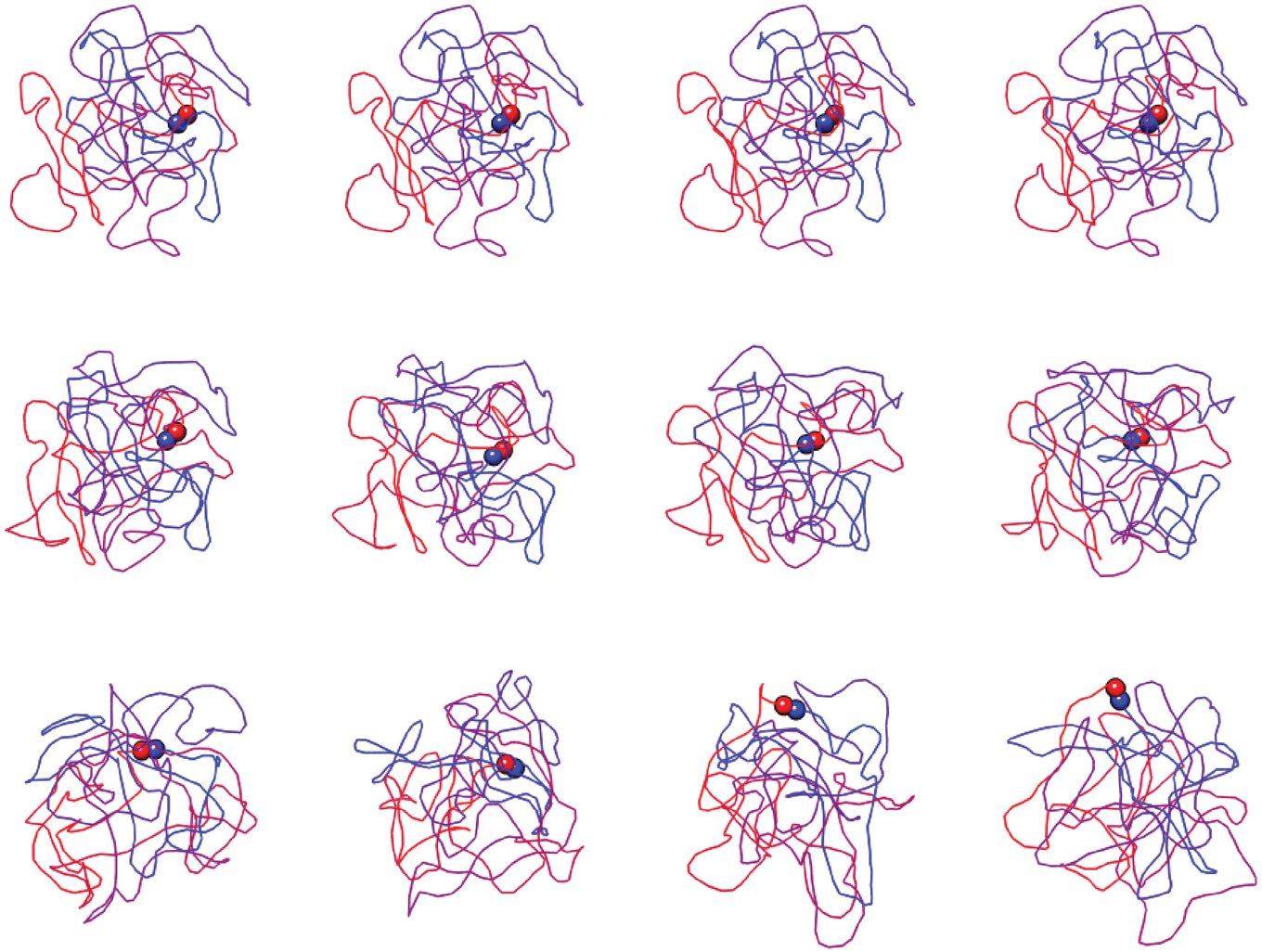




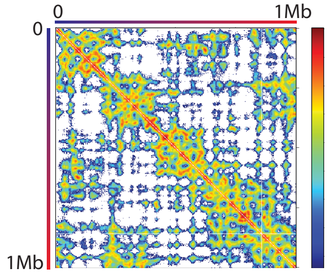


by Ivan Junier

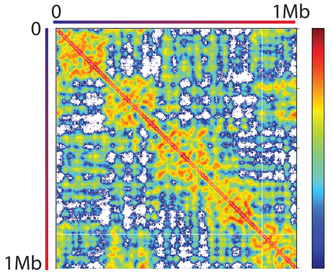
Toy interaction matrices



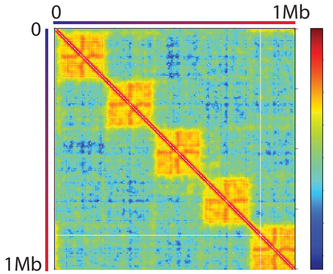
set 0 ($\Delta ts=10^0$)



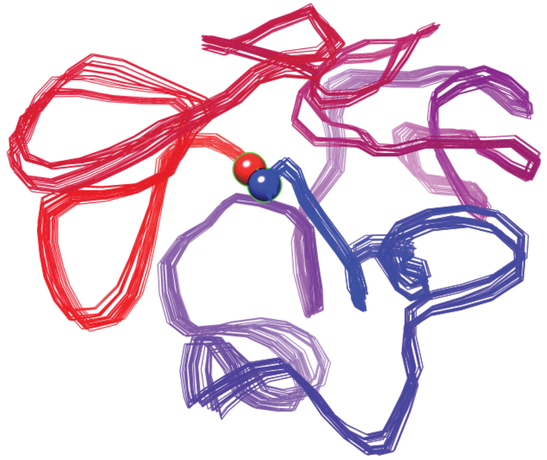
set 4 ($\Delta ts=10^4$)

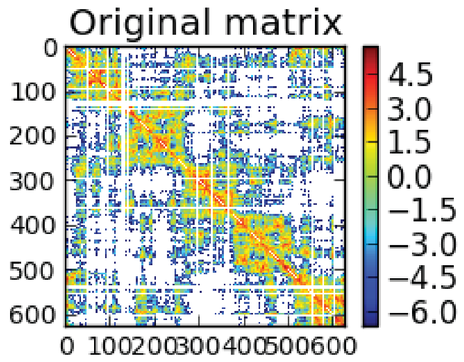


set 6 ($\Delta ts=10^6$)

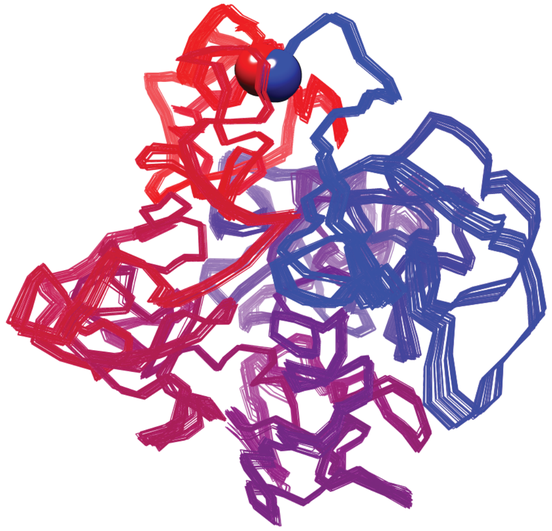


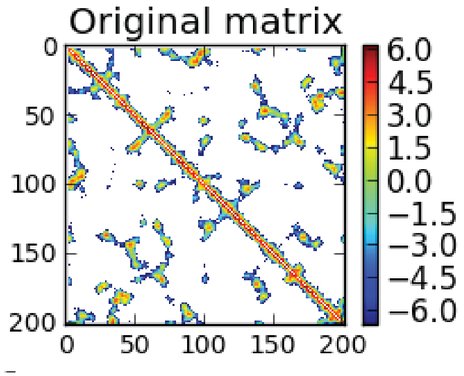
Reconstructing toy models

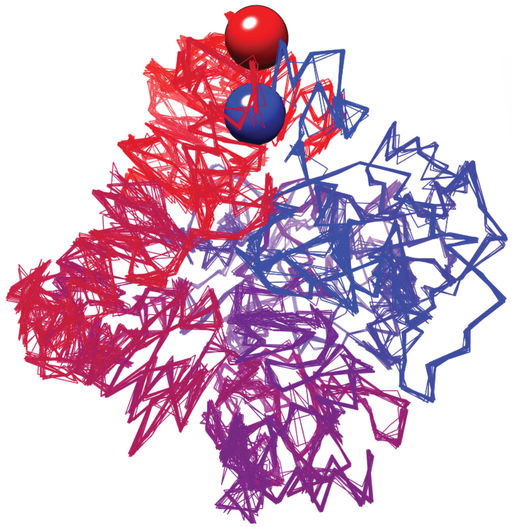




chr40_TAD
α=100
Δts=10
TADbit-SCC: 0.91
<dRMSD>: 32.7 nm
<dSCC>: 0.94

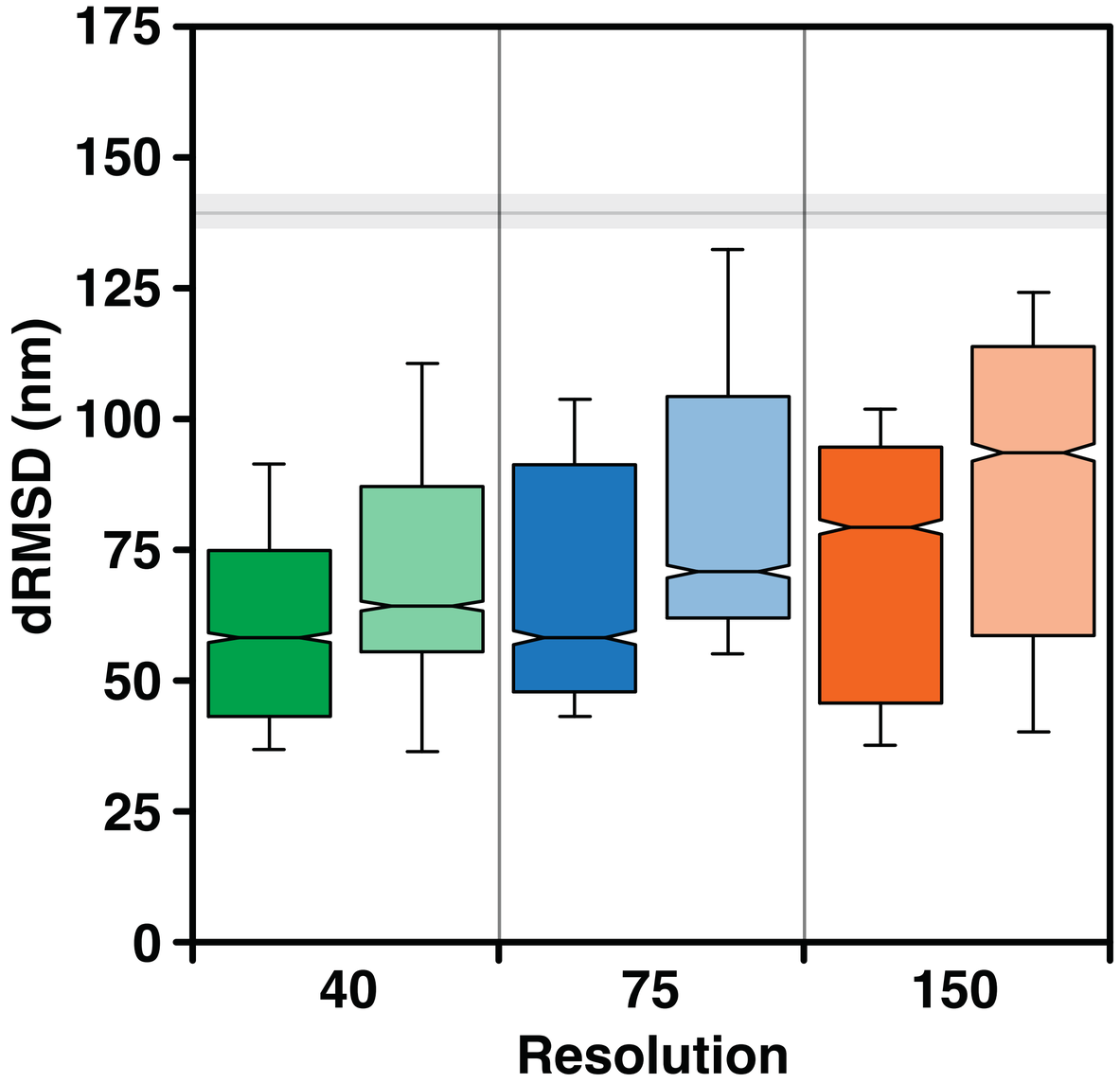




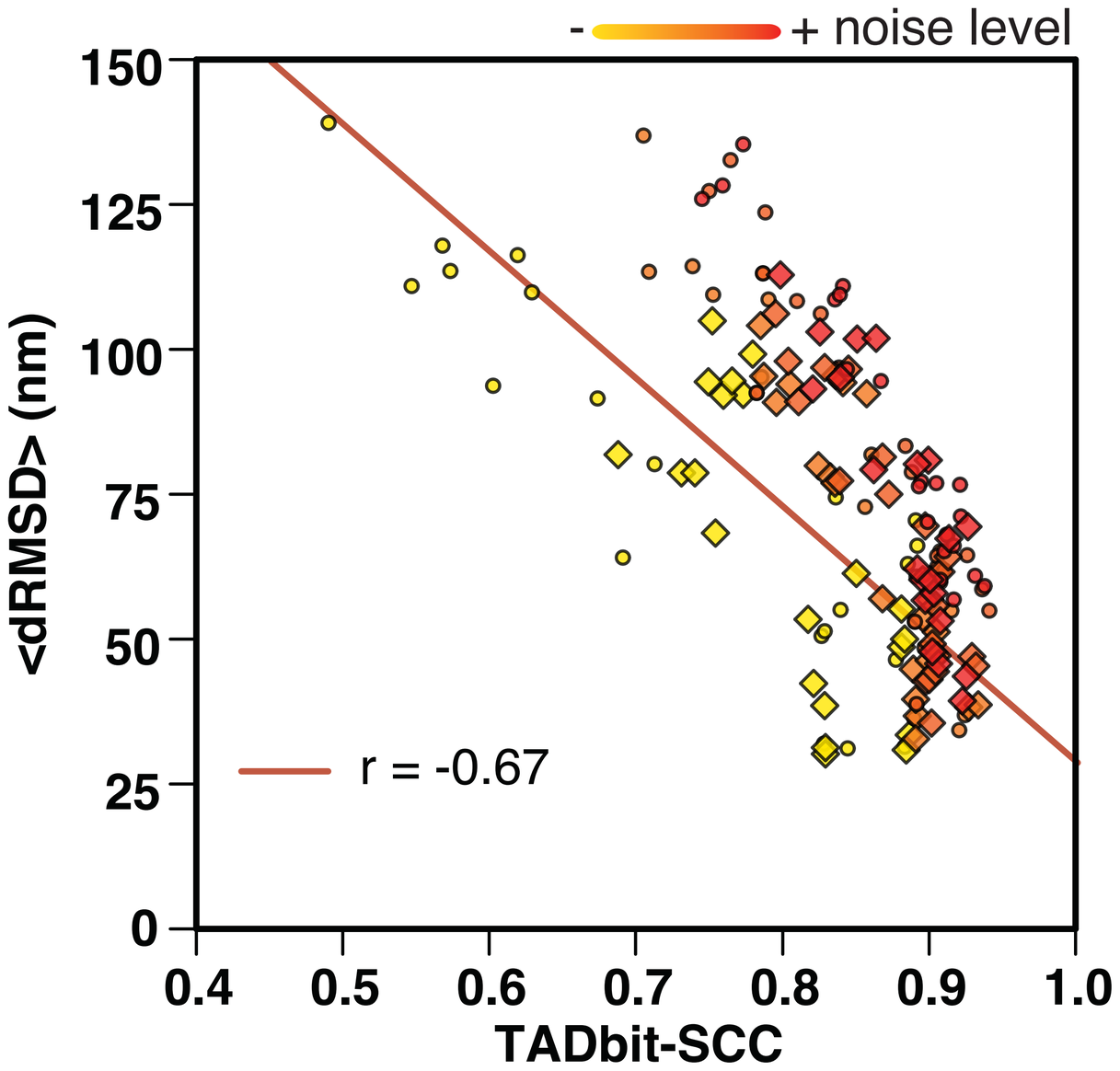


chr150_TAD
α=50
Δts=1
TADbit-SCC: 0.82
<dRMSD>: 45.4 nm
<dSCC>: 0.86

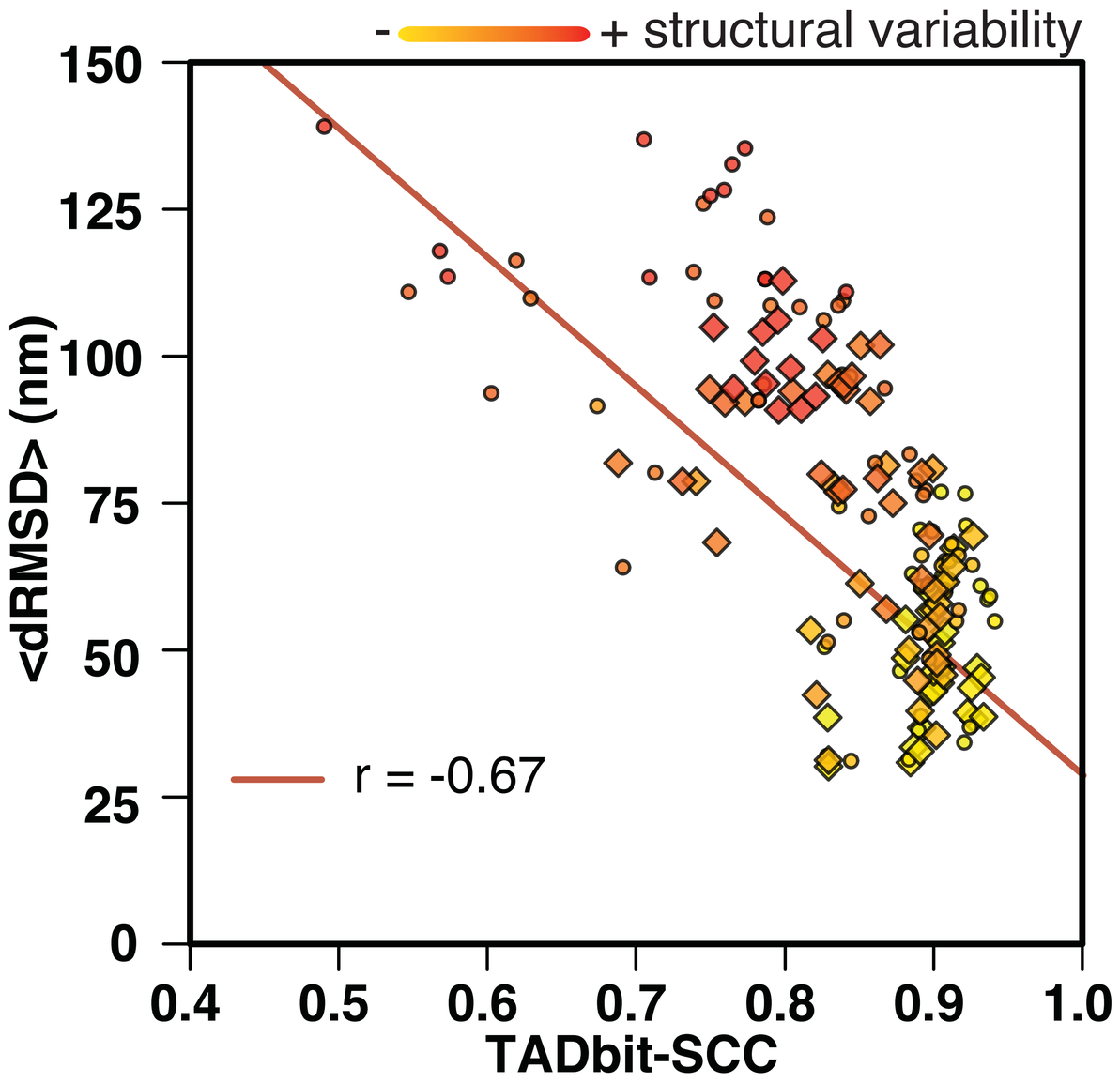
TADs & higher-res are "good"



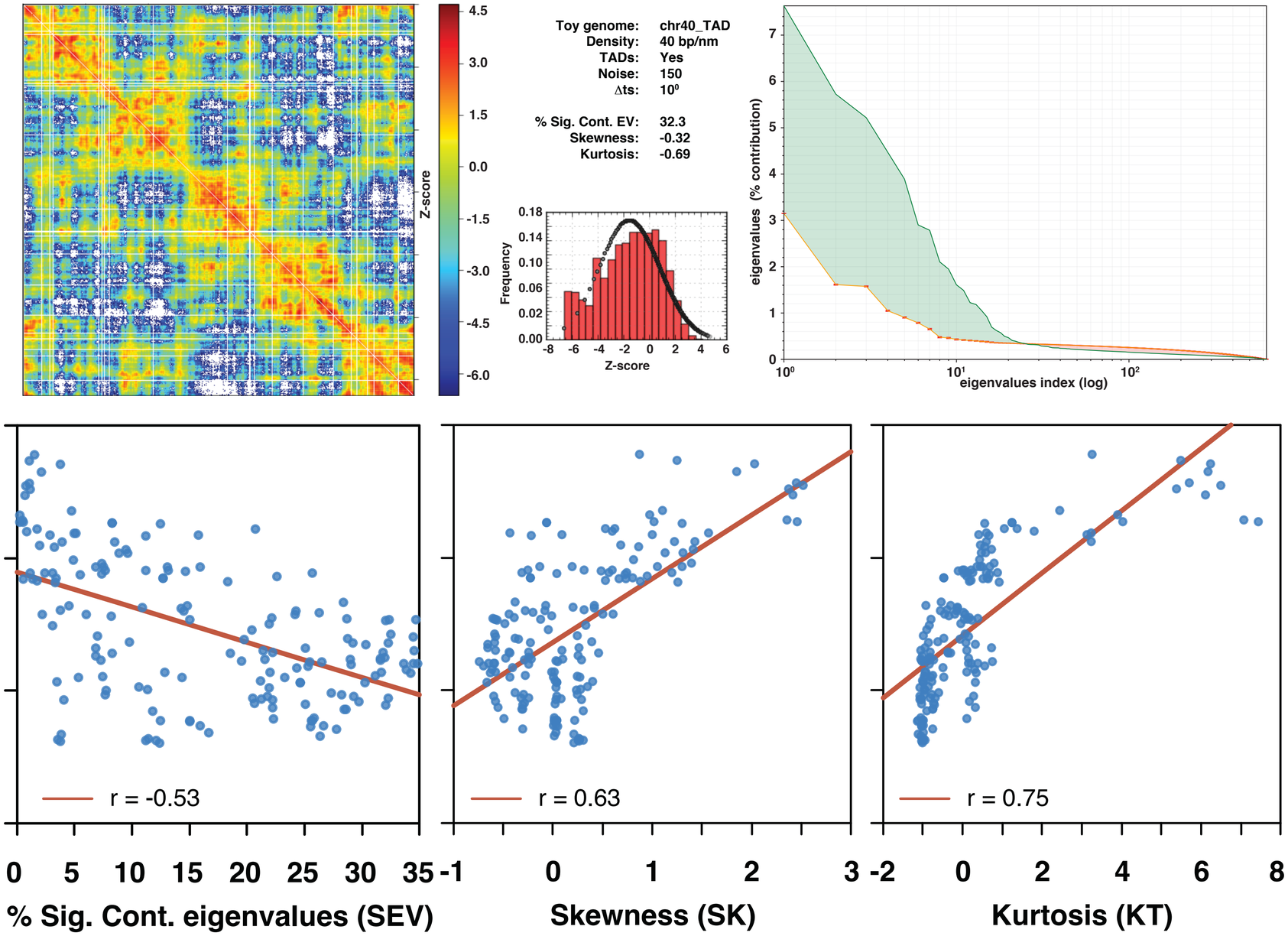
Noise is "OK"


- + noise level
150
125
100
75
50
25
0
<dRMSD> (nm)
r = -0.67
0.4
0.5
0.6
0.7
0.8
0.9
1.0
TADbit-SCC

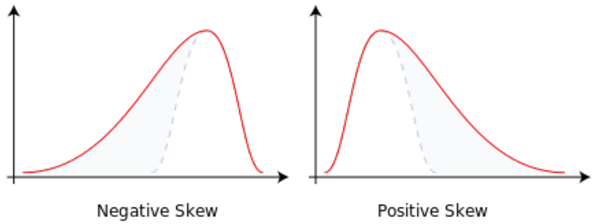
Structural variability is "NOT OK"

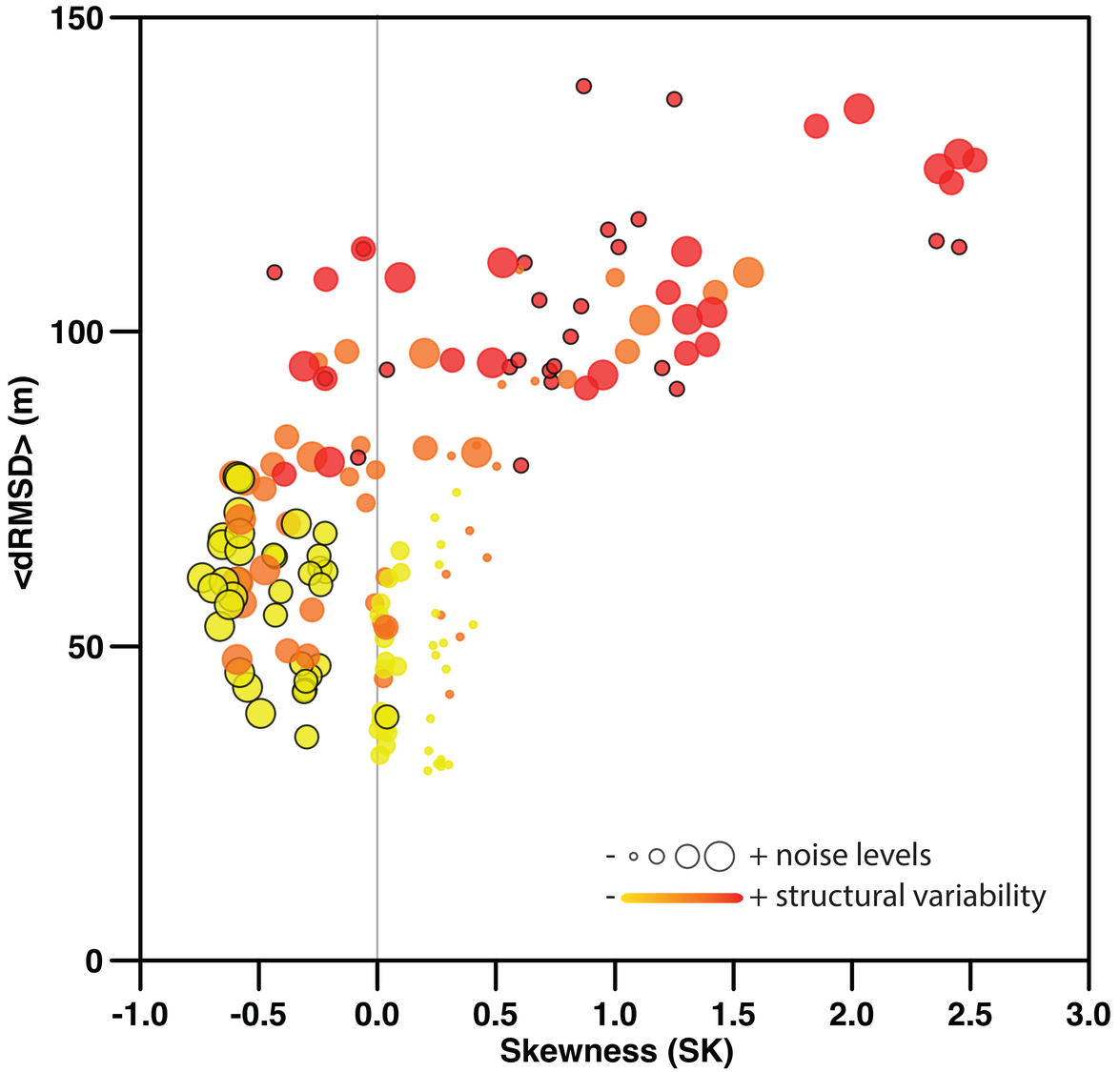


Can we predict the accuracy of the models?



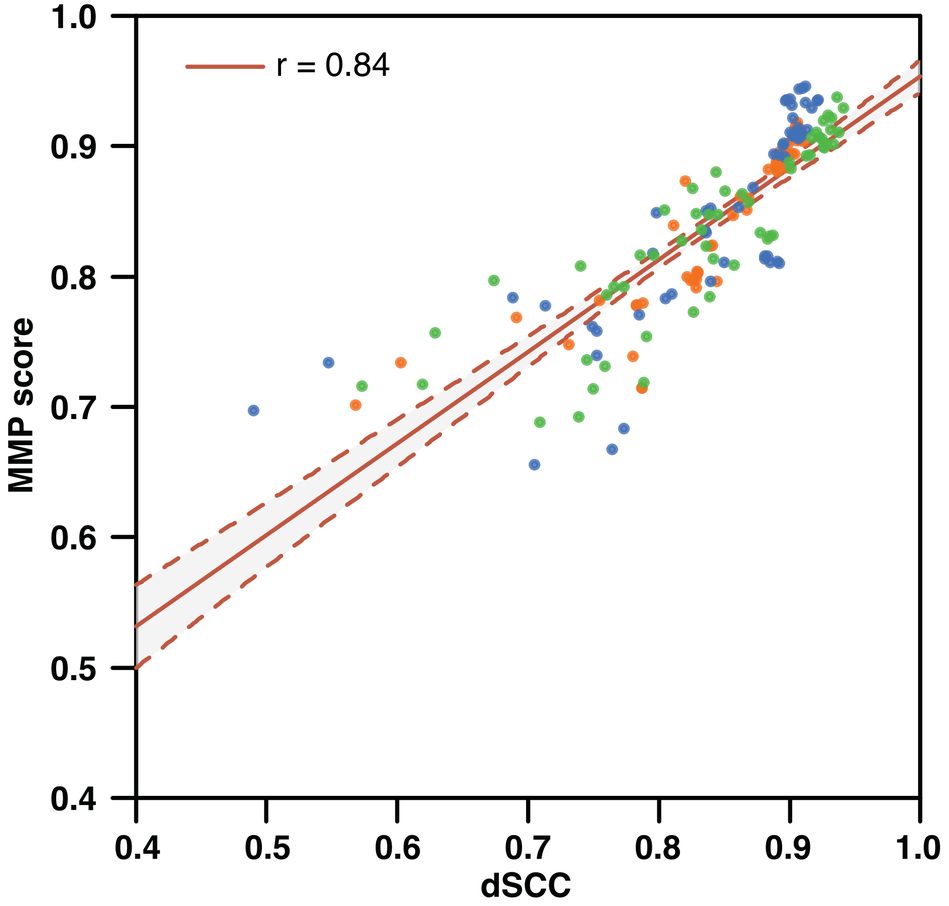
Skewness "side effect"





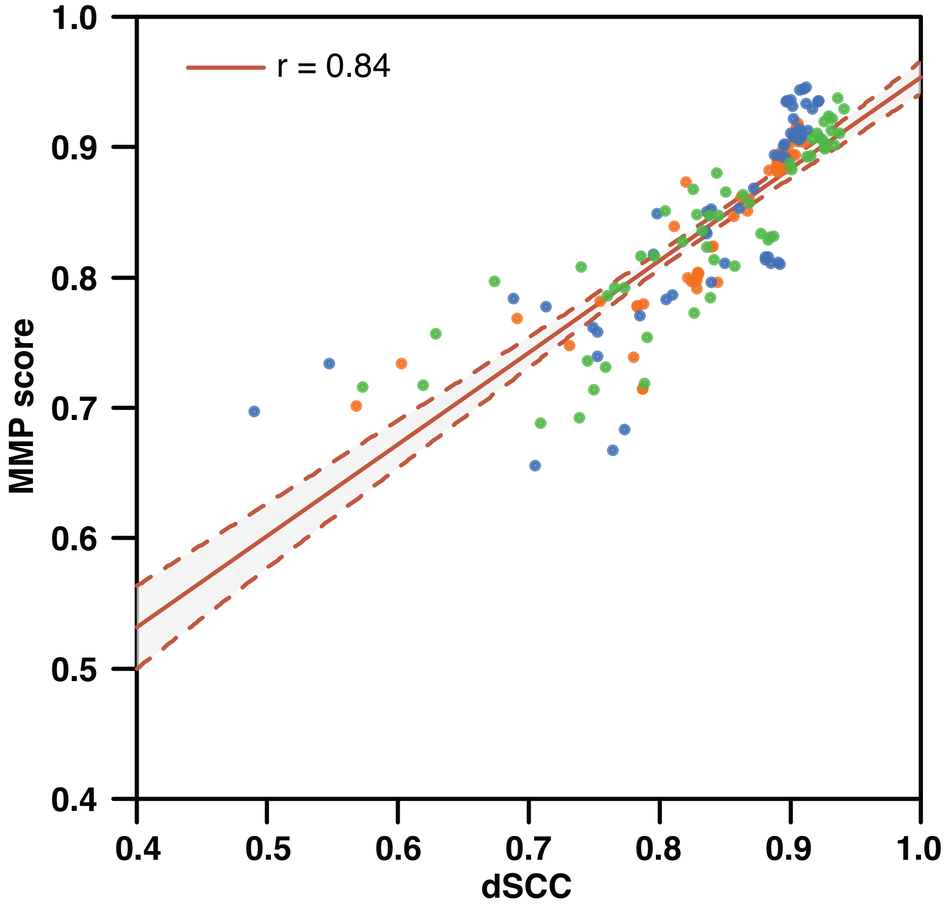
Can we predict the accuracy of the models?

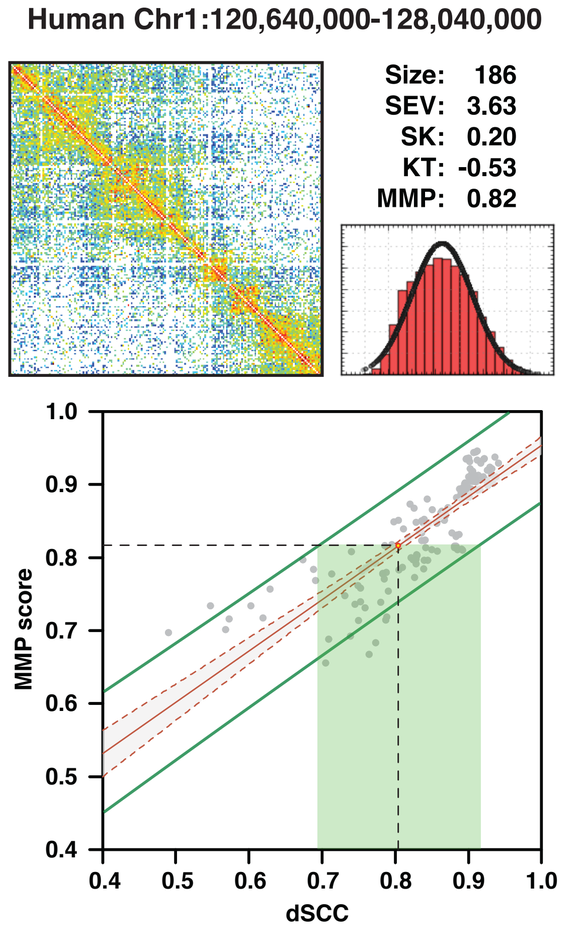
$$\mathrm{MMP} = -0.0002 * \mathrm{Size} + 0.0335 * \mathrm{SK} - 0.0229 * \mathrm{KU} + 0.0069 * \mathrm{SEV} + 0.8126$$



Can we predict the accuracy of the models?

$$\mathrm{MMP} = -0.0002 * \mathrm{Size} + 0.0335 * \mathrm{SK} - 0.0229 * \mathrm{KU} + 0.0069 * \mathrm{SEV} + 0.8126$$





Higher-res is "good"

put your $$ in sequencing

Noise is "OK"

no need to worry much

Structural variability is "NOT OK"

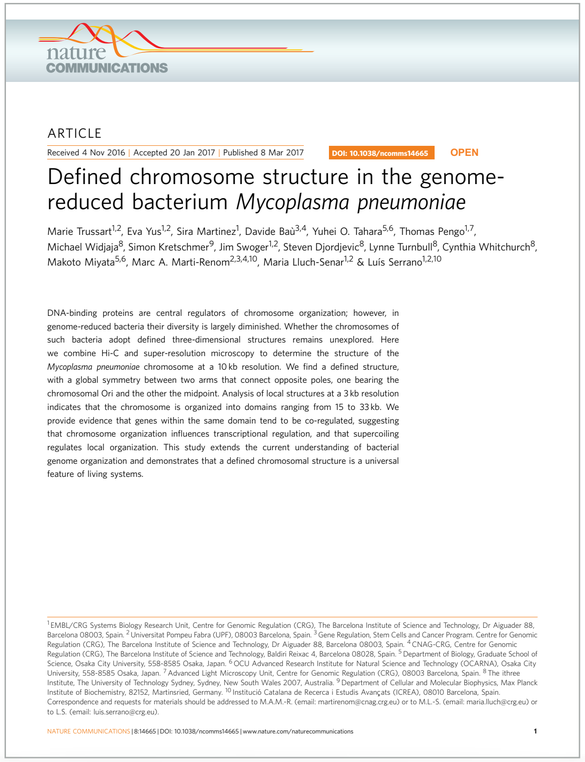
homogenize your cell population!

…but we can differentiate between noise and structural variability

and we can a priori predict the accuracy of the models

Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae

Trussart et al. Nature Communications (2017) 8 14665



nature COMMUNICATIONS

ARTICLE

Received 4 Nov 2016 | Accepted 20 Jan 2017 | Published 8 Mar 2017

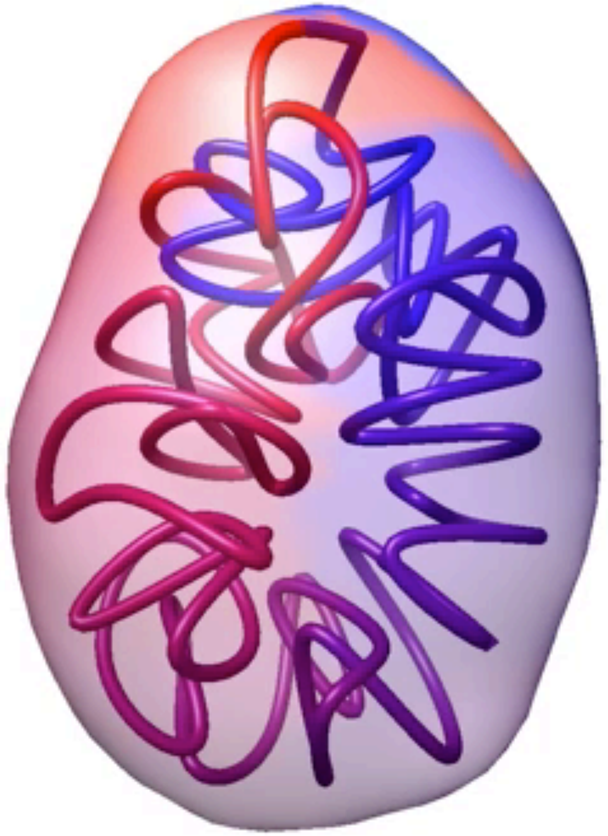
DOI: 10.1038/ncomms14665 OPEN

Defined chromosome structure in the genome-reduced bacterium *Mycoplasma pneumoniae*

Marie Trussart[1,2], Eva Yus[1,2], Sira Martinez[1], Davide Baù[3,4], Yuhei O. Tahara[5,6], Thomas Pengo[1,7], Michael Widjaja[8], Simon Kretschmer[9], Jim Swoger[1,2], Steven Djordjevic[8], Lynne Turnbull[8], Cynthia Whitchurch[8], Makoto Miyata[5,6], Marc A. Marti-Renom[2,3,4,10], Maria Lluch-Senar[1,2] & Luis Serrano[1,2,10]

DNA-binding proteins are central regulators of chromosome organization; however, in genome-reduced bacteria their diversity is largely diminished. Whether the chromosomes of such bacteria adopt defined three-dimensional structures remains unexplored. Here we combine Hi-C and super-resolution microscopy to determine the structure of the *Mycoplasma pneumoniae* chromosome at a 10 kb resolution. We find a defined structure, with a global symmetry between two arms that connect opposite poles, one bearing the chromosomal Ori and the other the midpoint. Analysis of local structures at a 3 kb resolution indicates that the chromosome is organized into domains ranging from 15 to 33 kb. We provide evidence that genes within the same domain tend to be co-regulated, suggesting that chromosome organization influences transcriptional regulation, and that supercoiling regulates local organization. This study extends the current understanding of bacterial genome organization and demonstrates that a defined chromosomal structure is a universal feature of living systems.

[1] EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr Aiguader 88, Barcelona 08003, Spain. [2] Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. [3] Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr Aiguader 88, Barcelona 08003, Spain. [4] CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Baldiri Reixac 4, Barcelona 08028, Spain. [5] Department of Biology, Graduate School of Science, Osaka City University, 558-8585 Osaka, Japan. [6] OCU Advanced Research Institute for Natural Science and Technology (OCARNA), Osaka City University, 558-8585 Osaka, Japan. [7] Advanced Light Microscopy Unit, Centre for Genomic Regulation (CRG), 08003 Barcelona, Spain. [8] The three Institute, The University of Technology Sydney, Sydney, New South Wales 2007, Australia. [9] Department of Cellular and Molecular Biophysics, Max Planck Institute of Biochemistry, 82152, Martinsried, Germany. [10] Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain. Correspondence and requests for materials should be addressed to M.A.M.-R. (email: martinenom@cnag.crg.eu) or to M.L.-S. (email: maria.lluch@crg.eu) or to L.S. (email: luis.serrano@crg.eu).



Mycoplasma is a small genome with few structural factors

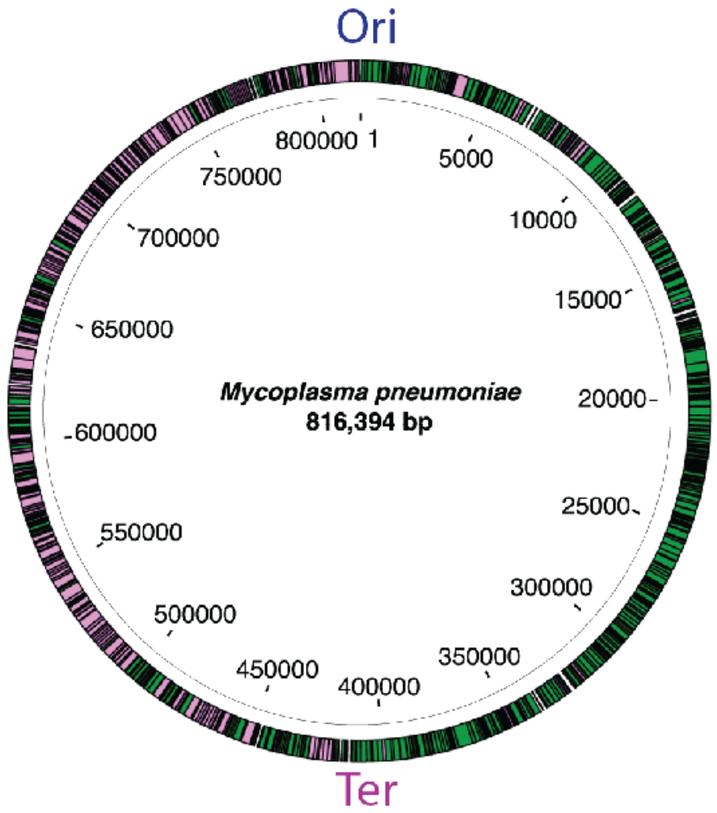
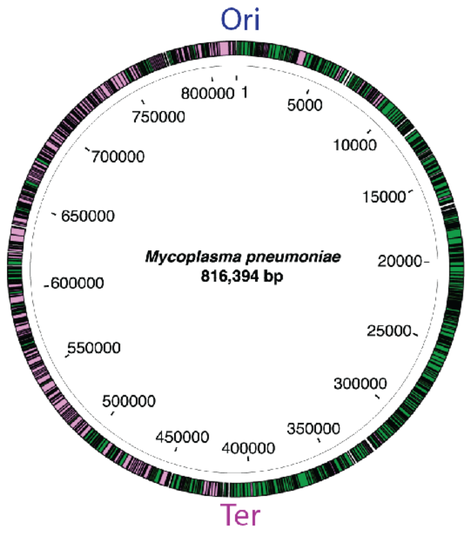


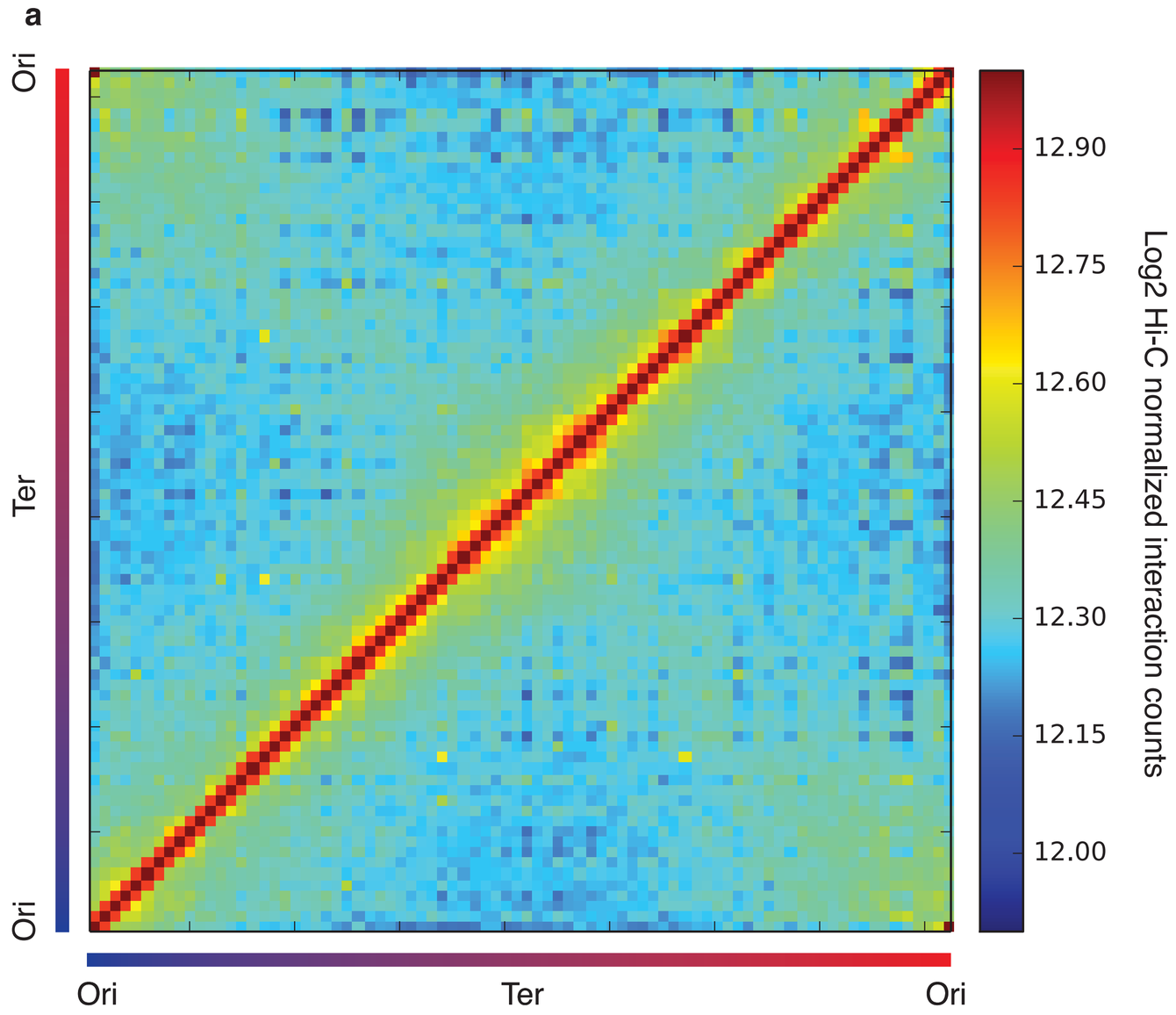
Table 1 | List of assigned transcription factors, sigma factors and structural proteins and essentiality with three distinct categories: essential (E), non -essential (NE) and fitness (F).

Gene number	Gene name	Protein name	Essentiality [44]
MPN002	cbpA	Curved DNA-binding protein CbpA	F
MPN003	gyrB	DNA gyrase subunit B	E
MPN004	gyrA	DNA gyrase subunit A	E
MPN122	parB	DNA topoisomerase 4 subunit B	E
MPN123	parC	DNA topoisomerase 4 subunit A	E
MPN124	hrcA	Heat-inducible transcription repressor hrcA	E
MPN229	ssbA	SSB-binding ssDNA	E
MPN239	gntR	Probable HTH-type transcriptional regulator gntR	E
MPN241	whiA	Transcription factor with WhiA C-terminal domain	F
MPN266	spxA	Transcriptional regulator Spx	E
MPN275	ybaB	DNA-binding protein, YbaB/EbfC family	F
MPN294	araC	AraC-like transcriptional regulator	NE
MPN332	lon	ATP-dependent protease La (EC 3.4.21.53)	E
MPN352	sigA	RNA polymerase sigma factor rpoD (Sigma-A) (EC 2.7.7.6)	E
MPN424	ylxM	Putative helix-turn-helix protein, YlxM/p13-like protein	NE
MPN426	smc	SMC family, chromosome/DNA binding/protecting functions	E
MPN478	yrbC	YebC family protein (transcription factor of the tetR family)	E
MPN529	ihf	Histone-like bacterial DNA-binding protein	F
MPN554	ssbB	Putative single-stranded DNA-binding protein	E
MPN572	pepA	Probable cytosol aminopeptidase (EC 3.4.11.1) (leucine aminopeptidase) (LAP) (leucyl aminopeptidase)	E
MPN608	phoU	Transcriptional regulator involved in phosphate transport system	E
MPN626	mpn626	Alternative sigma factor	NE
MPN686	dnaA	Chromosomal replication initiator protein dnaA	E

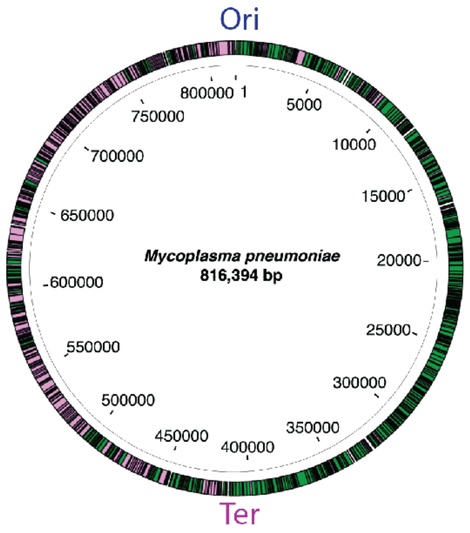
E, essential; F, fitness; LAP, leucine aminopeptidase; NE, non-essential; ssDNA, single-stranded DNA[44].

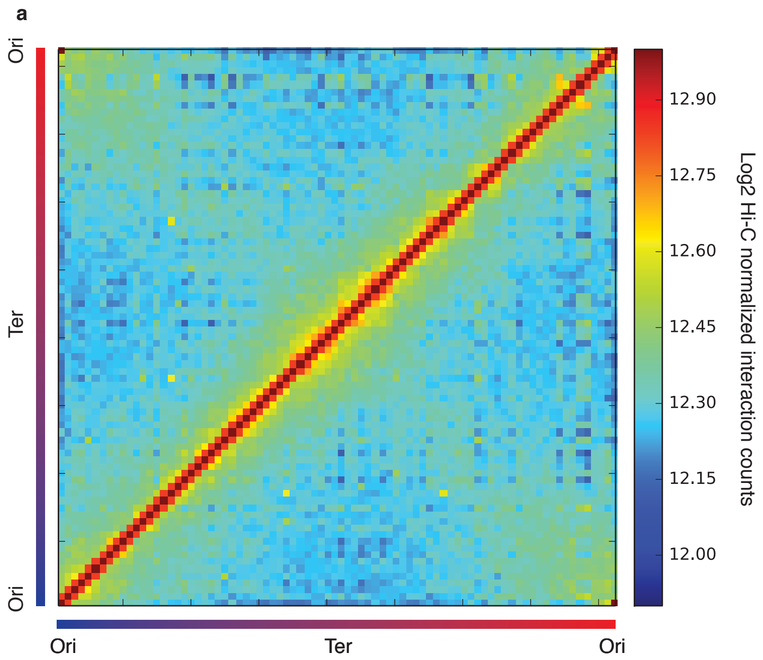
Can we build 3D models of Mycoplasma?

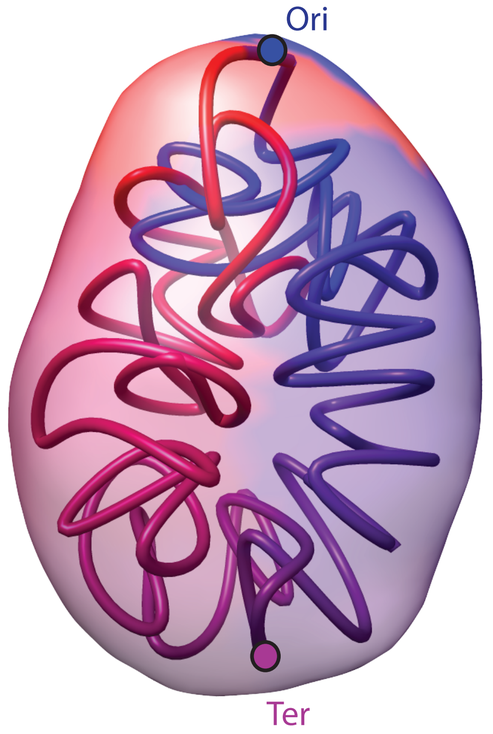


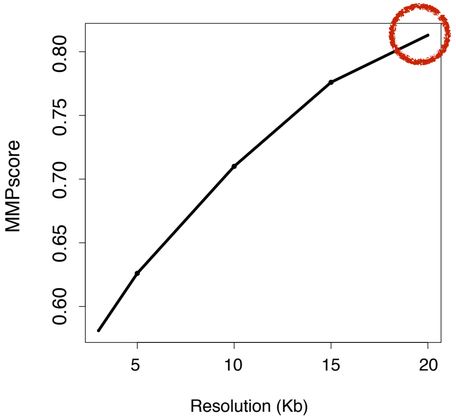


Can we build 3D models of Mycoplasma?

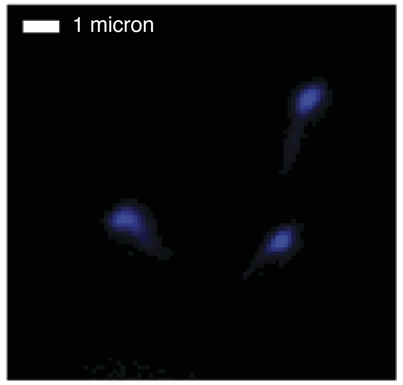


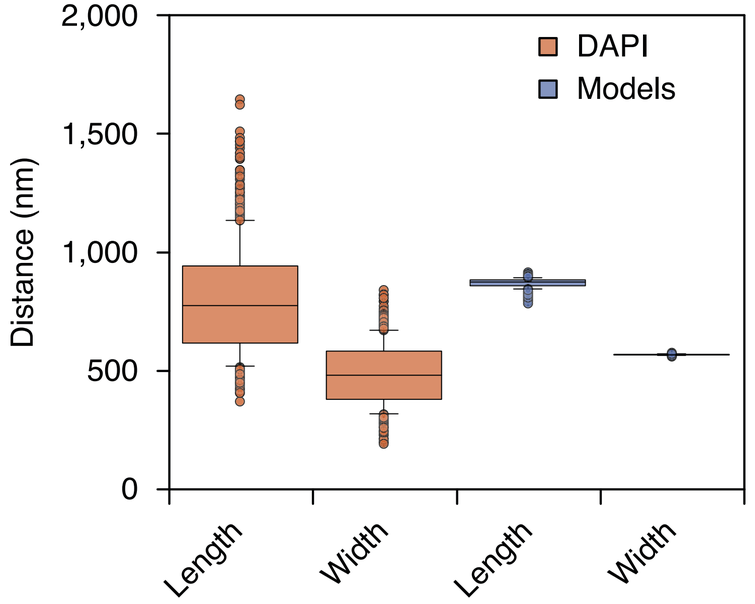


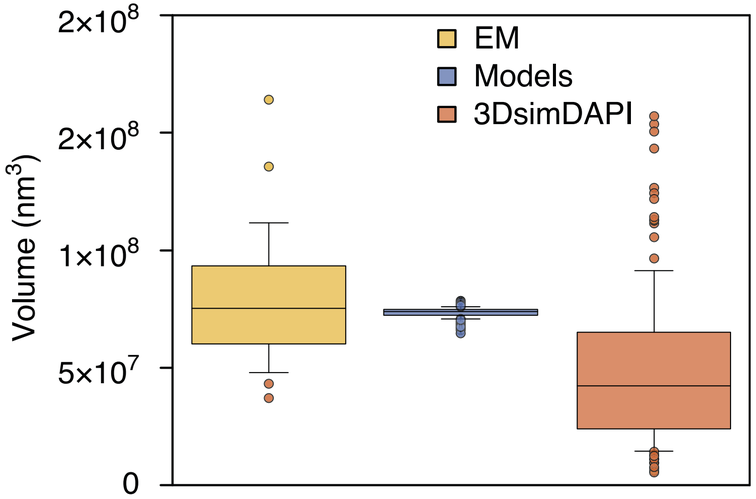




Is the overall 3D model accurate?

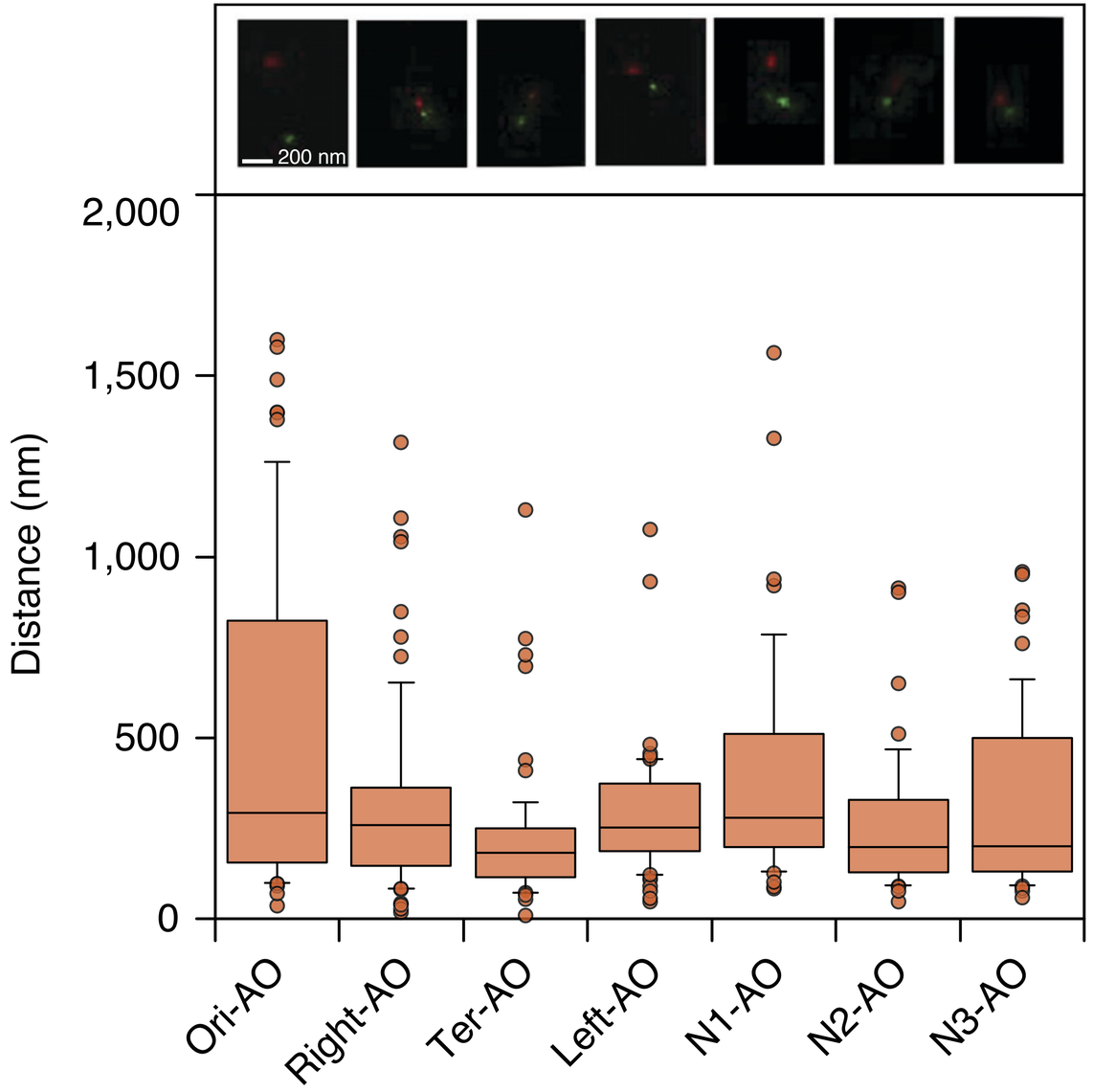


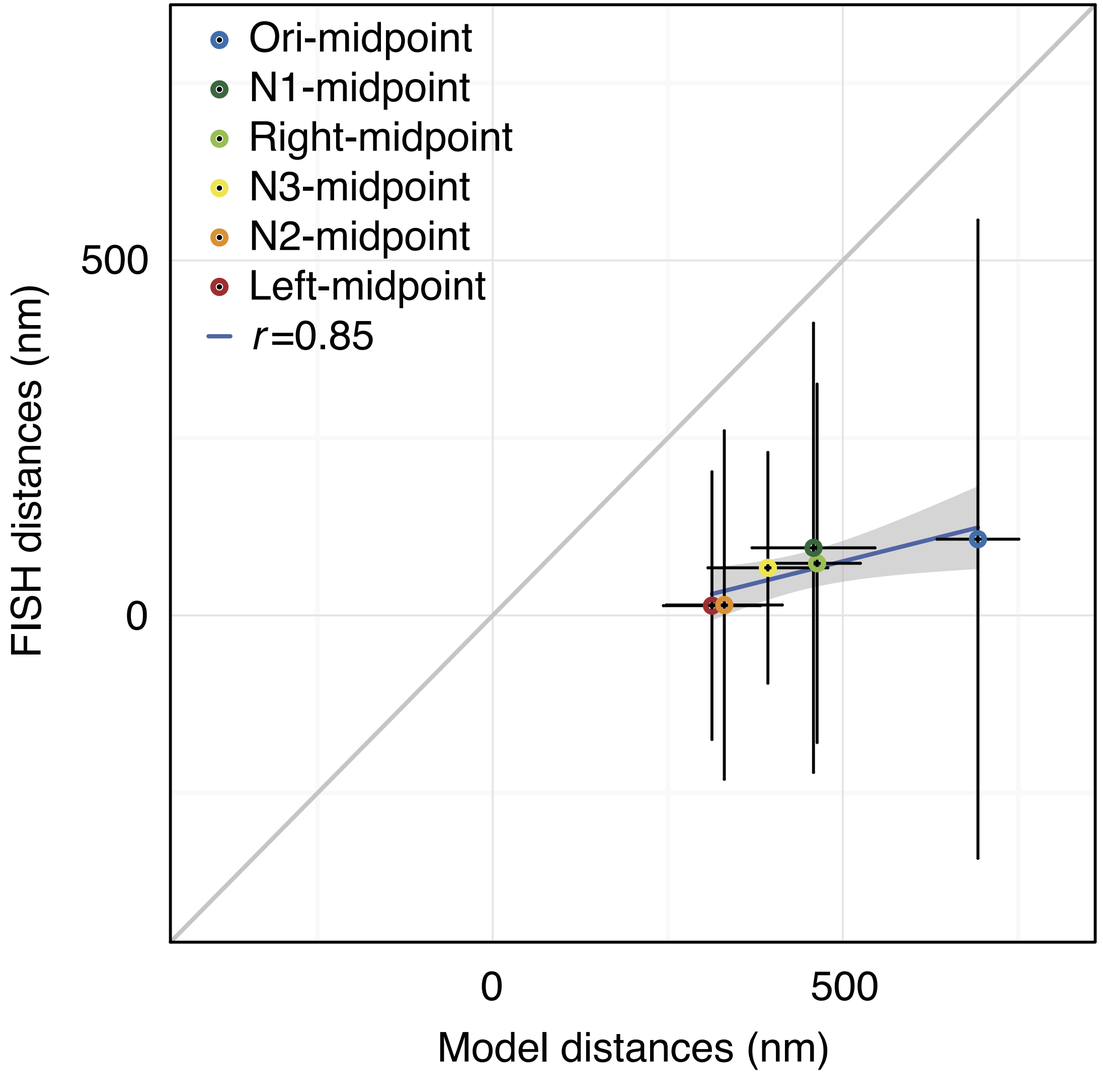




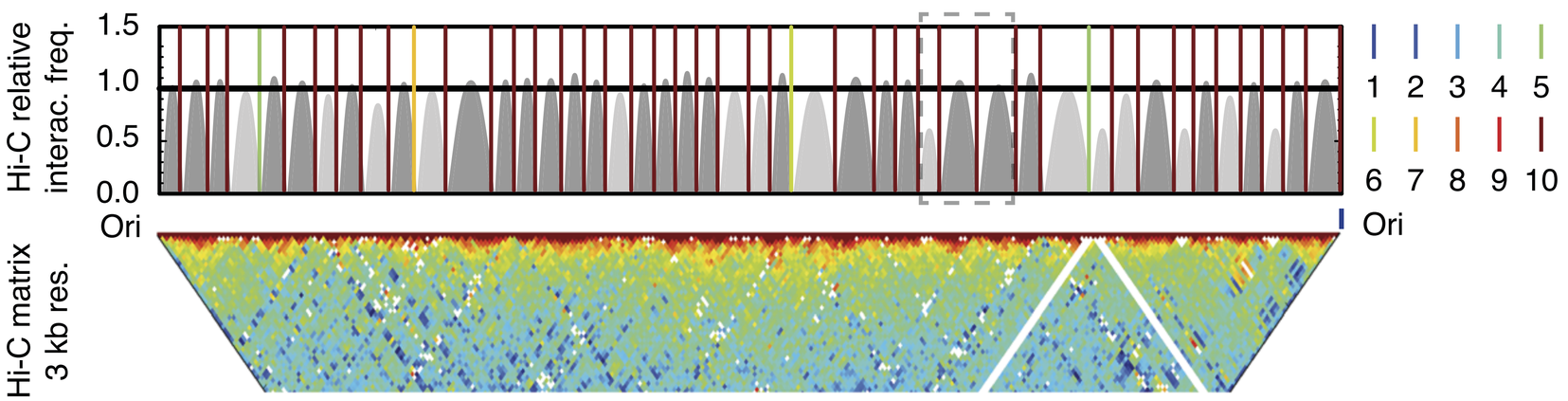
Makoto Miyata Group
Osaka City University

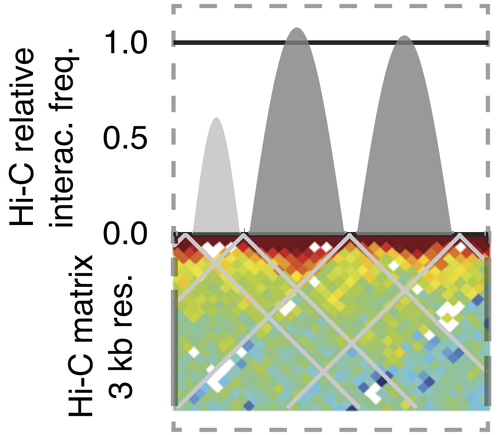
Are the details of the 3D model accurate?

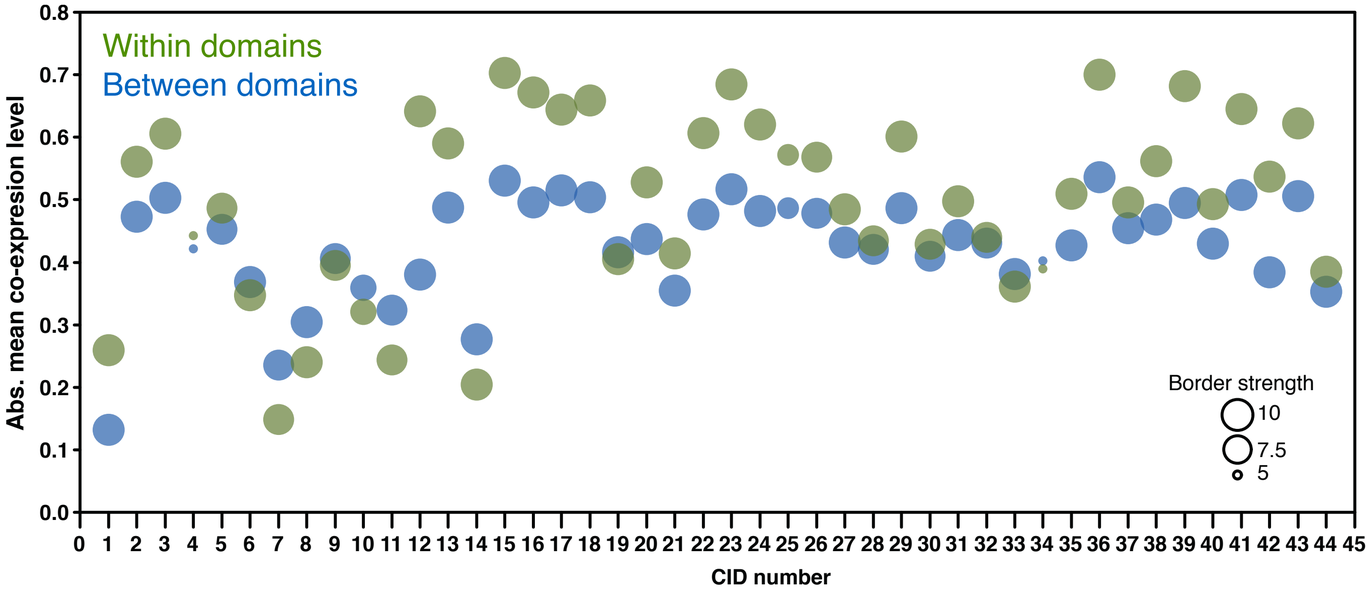


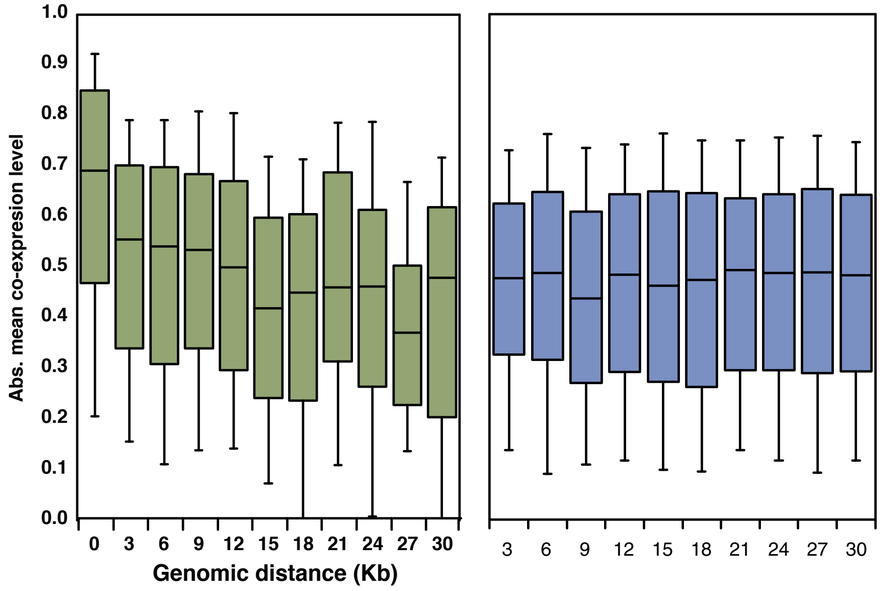


Mycoplasma genome is partitioned into co-regulated CIDs

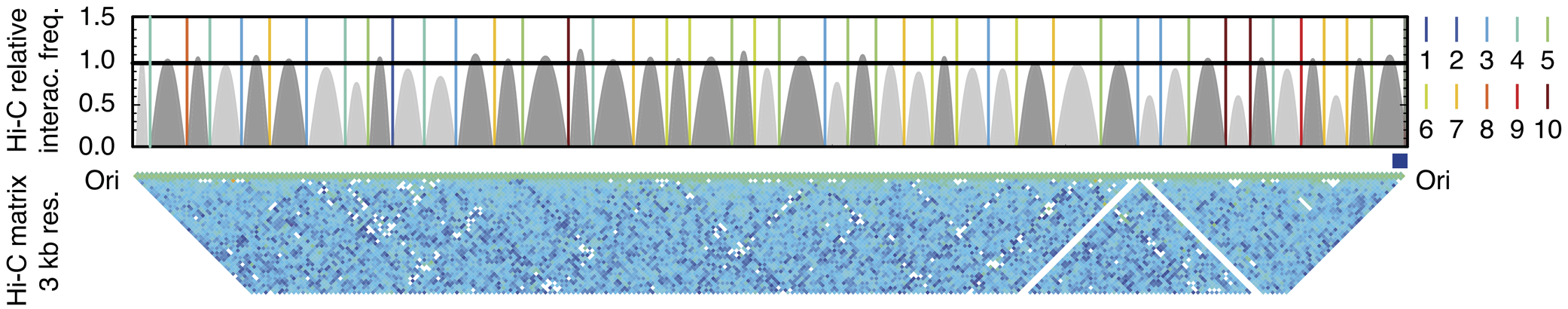


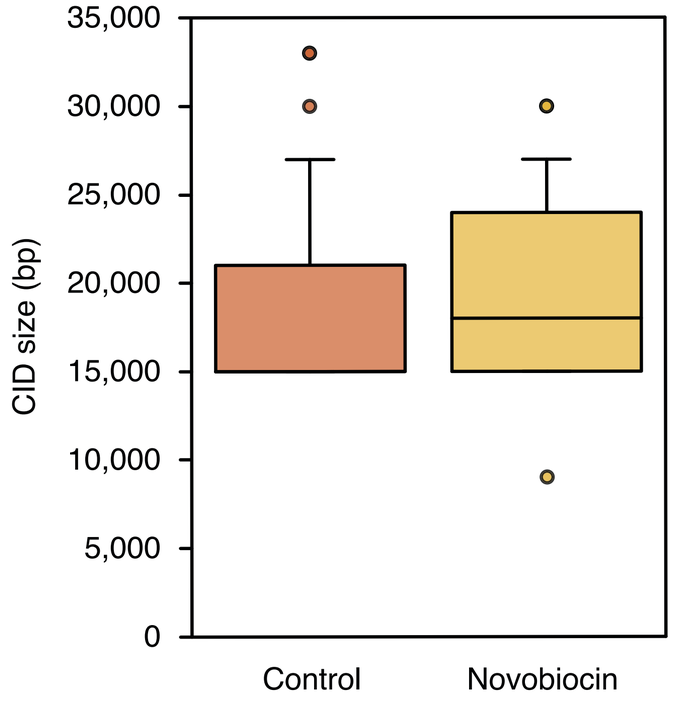


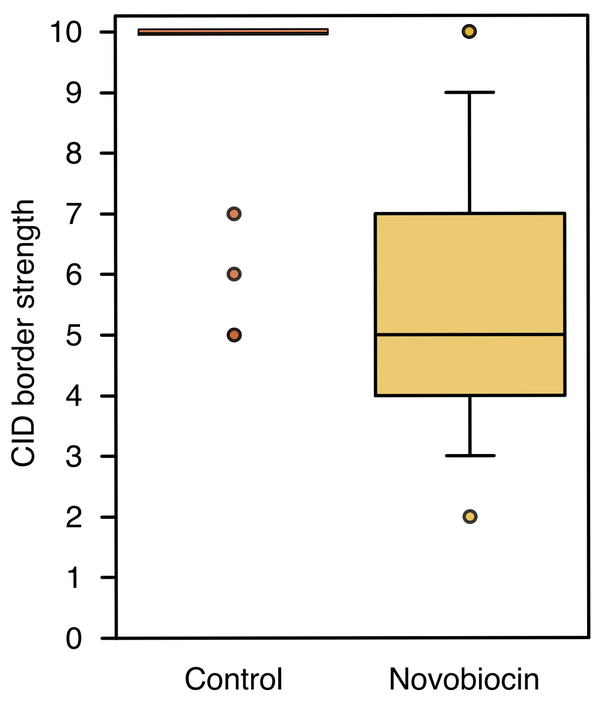


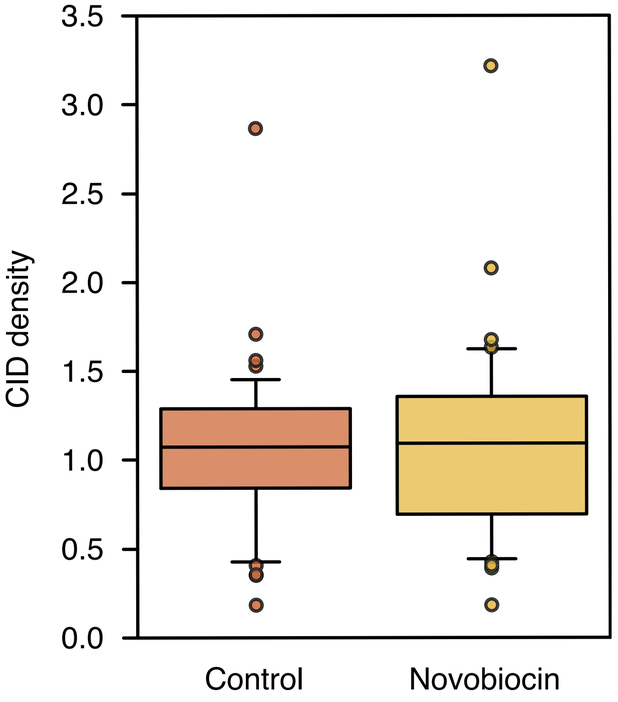


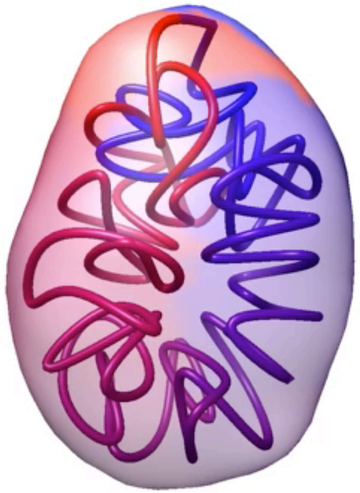
Inhibiting supercoiling decreases the sharpness of domain borders











Mycoplasma reduced-genome has a “3D structure”

Similar to Caulobacter, Mycoplasma has a double diagonal intersecting near the centre of the genome

Mycoplasma has CIDs (TADs)

CIDs contain co-regulated genes.

Inhibition of supercoiling by novobiocin significantly reduced the sharpness of CID borders.

Very few factors may be necessary to define a 3D structure

Other elements like supercoiling could regulate these domain boundaries.

Marie Trussart
Davide Baù

Gireesh K. Bogu
David Castillo
Yasmina Cuartero
Irene Farabella
Silvia Galan
Mike Goodstadt
Julen Mendieta
François Serra
Paula Soler
Yannick Spill
Marco di Stefano

in collaboration with Ivan Junier (Université Joseph Fourier) & Luís Serrano (CRG)

http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu