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

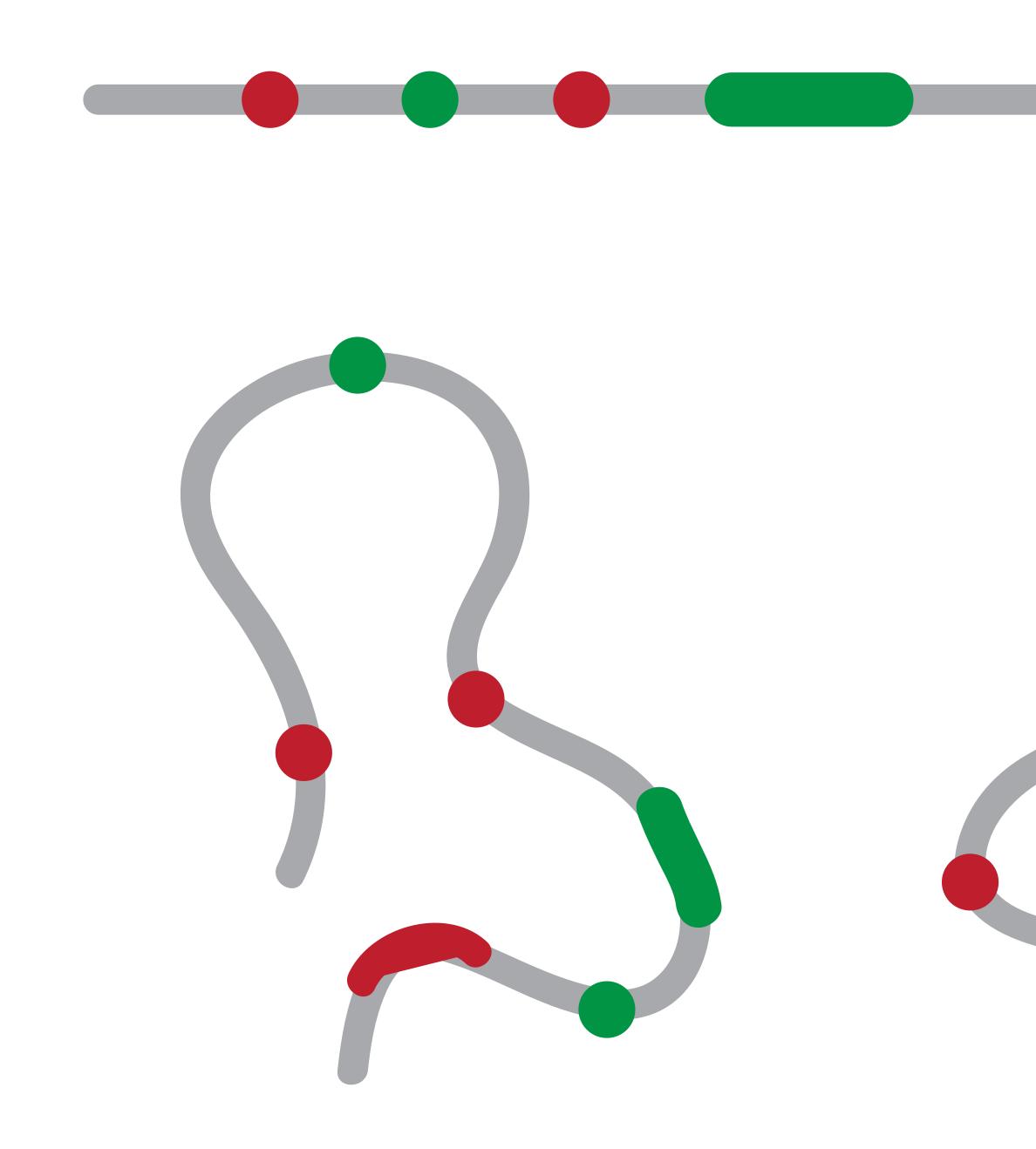
- Model assessment
- Mycoplasma 3D models

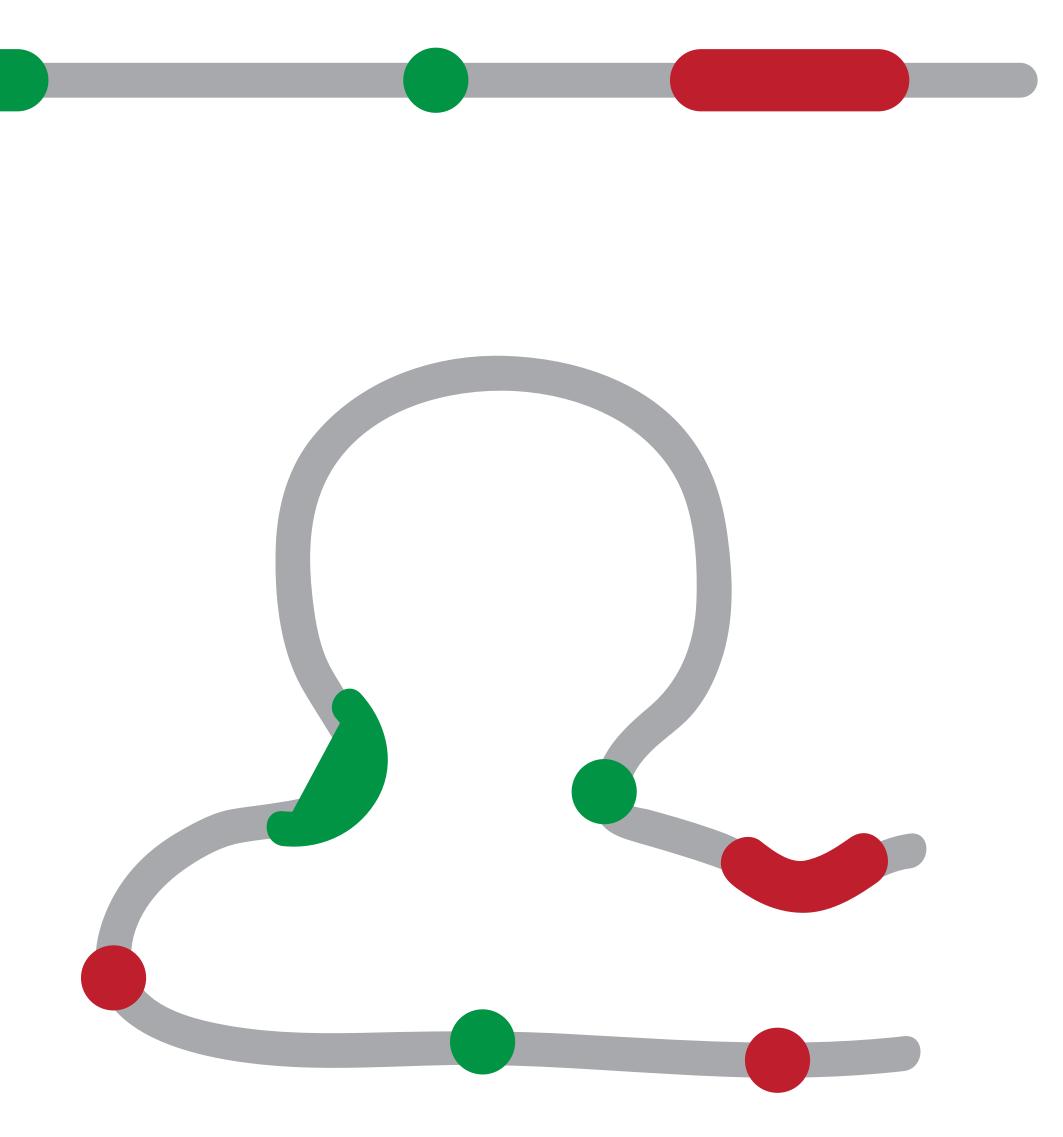
Marc A. Marti-Renom Structural Genomics Group (ICREA, CNAG-CRG)

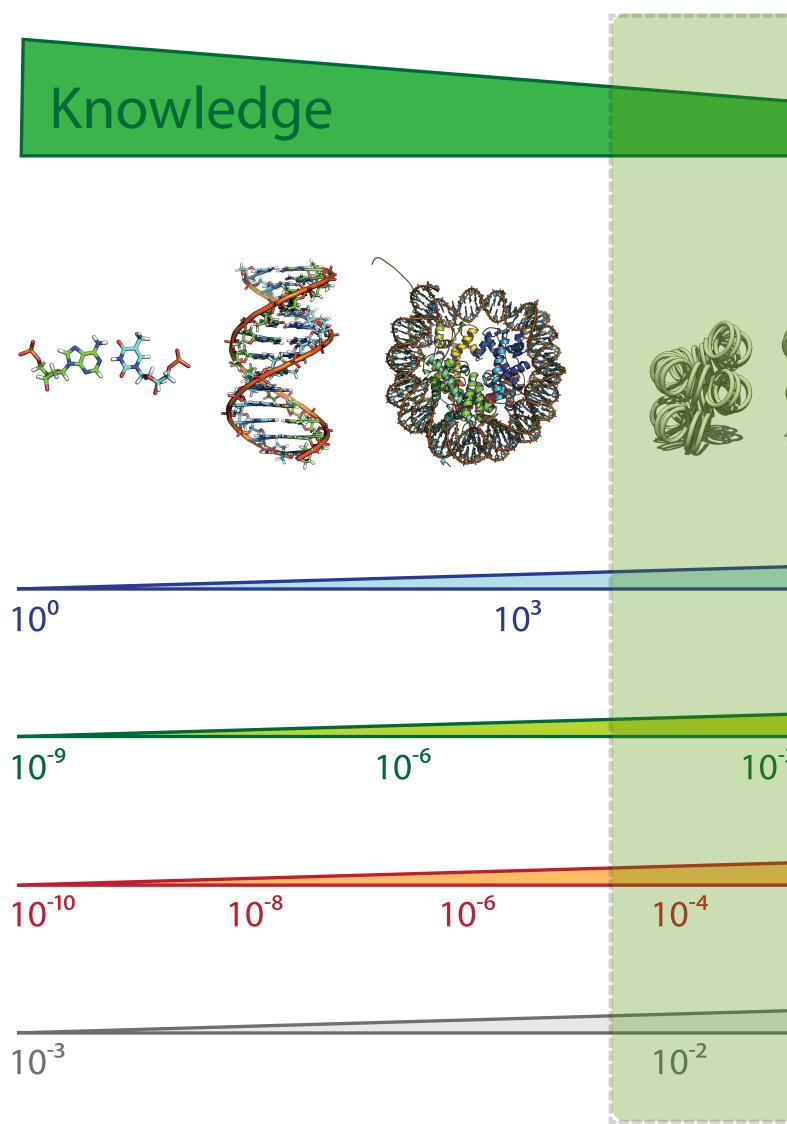
> Serrano and Marti-Renom Labs @CRG Now postdoc @The Walter and Elisa Institute. Melbourne, Australia.

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

Marie Trussart (PhD)





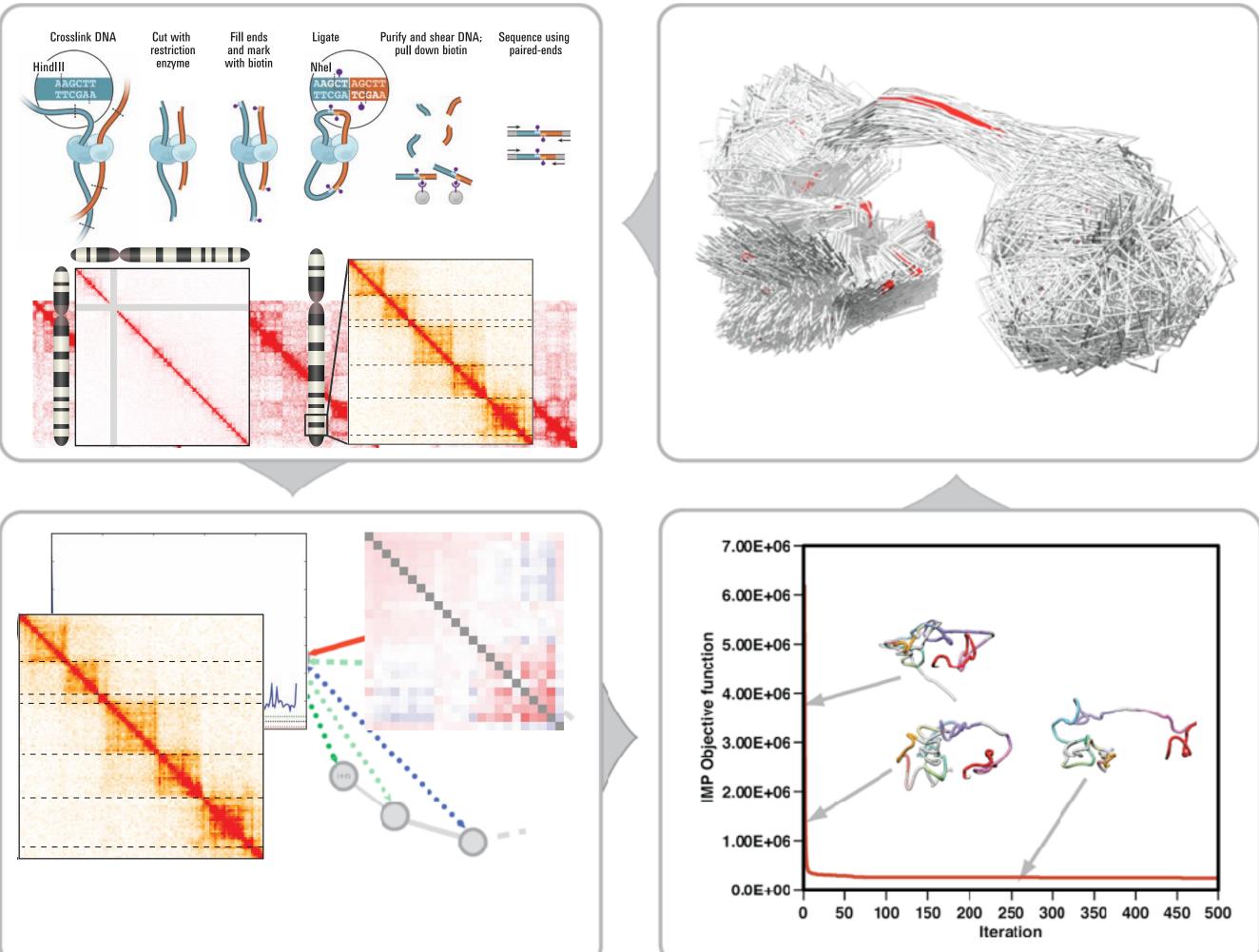


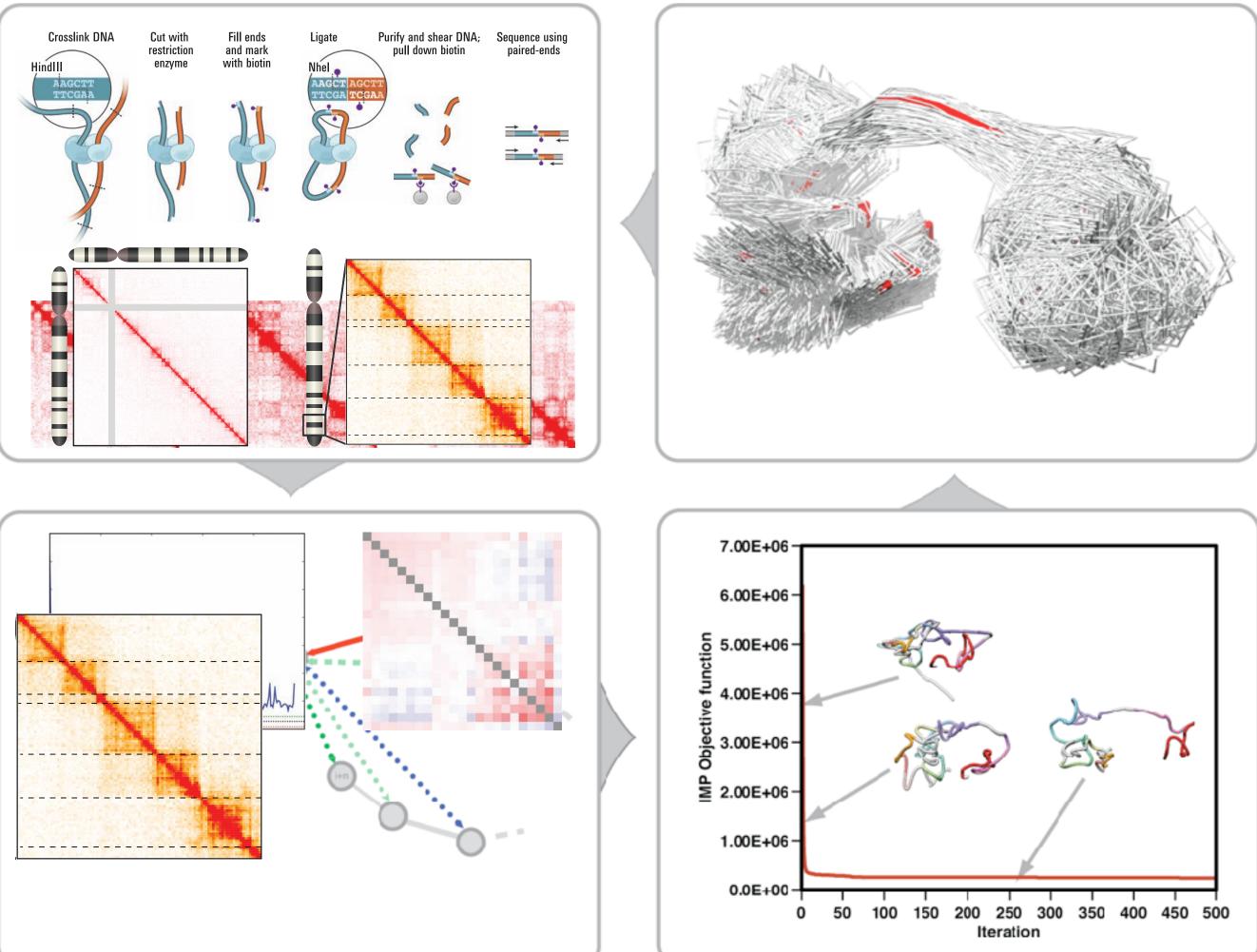
Resolution Gap Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)

	IDM			$\begin{array}{c} & & & & \\ & & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ &$	7
				DNA length	
	10 ⁶			10 ⁹	nt
				Volume]
10 ⁻³		10 ⁰		10 ³	μm³
				Time	
10 ⁻²		10 ⁰	10 ²	10 ³	S
				Resolution]
			10 ⁻¹		μ

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

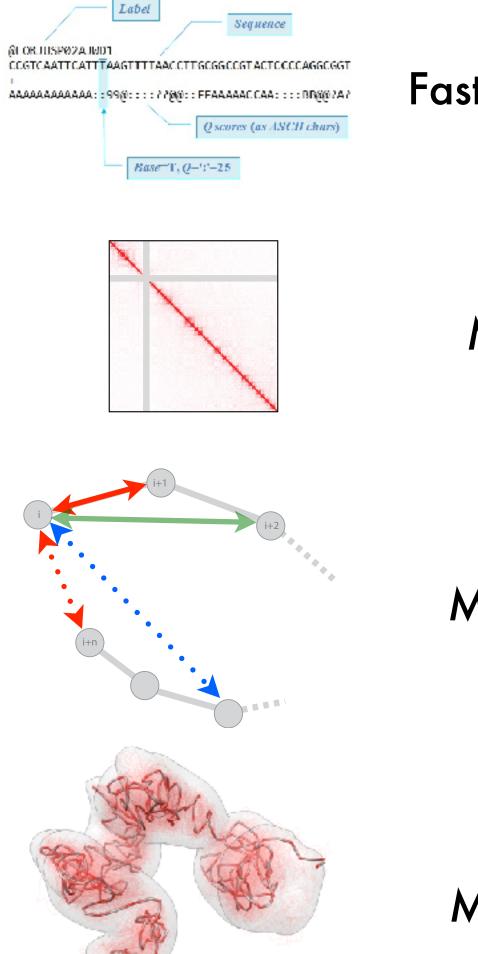
Experiments





Computation



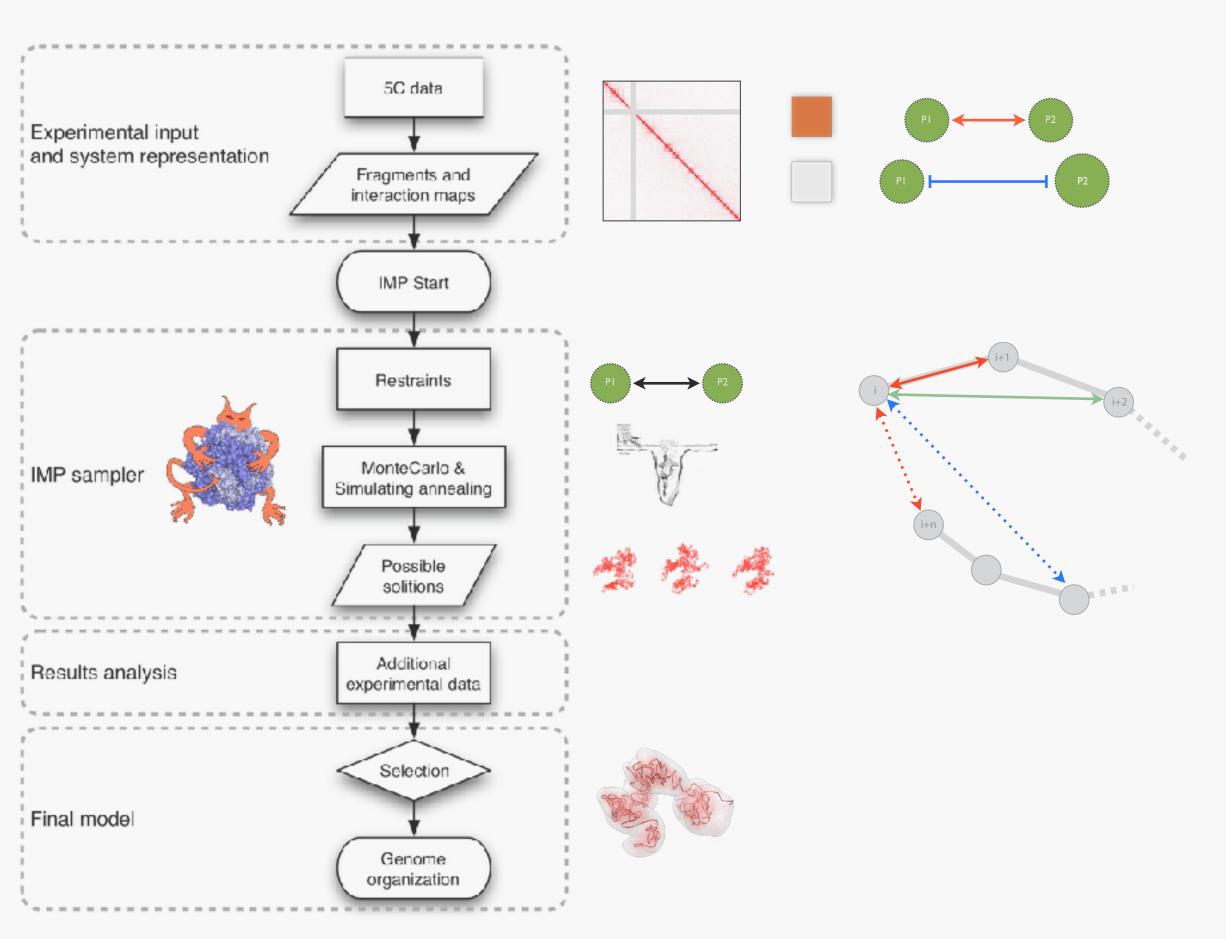


FastQ files to Maps

Map analysis

Model building

Model analysis



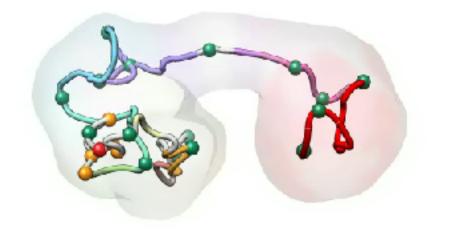


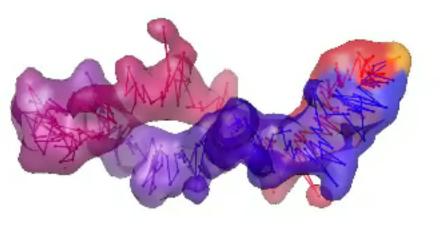


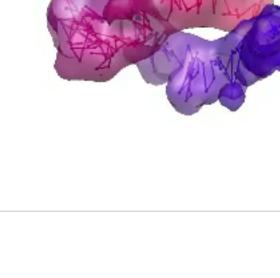
The Three-Dimensional Architecture of a Bacterial Genome and Its Alteration by Genetic Perturbation x A. Umbarger,^{1,8,*} Esteban Toro,^{2,8} Matthew A. Wright,¹ Gregory J. Porreca,¹ Davide Baù,⁴ Sun-Hae Hong,^{2,1} aael J. Fero,² Lihua J. Zhu,⁶ Marc A. Marti-Renom,^{4,*} Harley H. McAdams,² Lucy Shapiro,² Job Dekker,^{6,6,7,*} Converse M. Courses I.

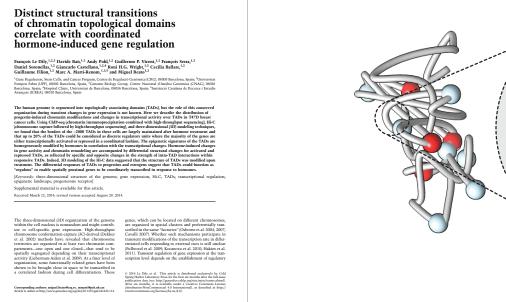
GENES & DEVELOPMENT 28:2151-2162 Published by Cold Spring Harbor Laboratory Pross, ISSN 0890-9369/14/, www.genesdev.org 2151

Molecular Cell Article





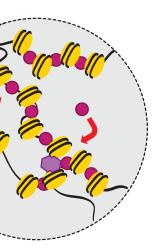




TADbit previous applications...

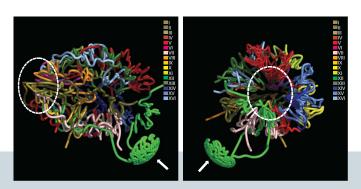
Baù, D. et al. Nat Struct Mol Biol (2011)

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

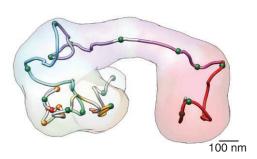


Le Dily, F. et al. Genes & Dev (2014)

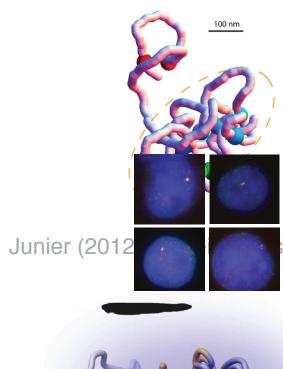
Are the models correct? Trussart et al. NAR (2015)

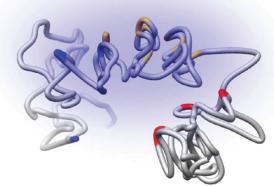




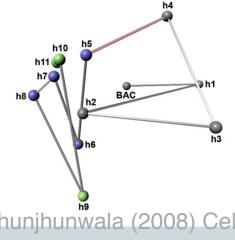


Baù (2011) Nature Structural & Molecular Biology

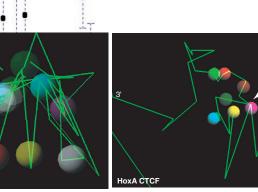




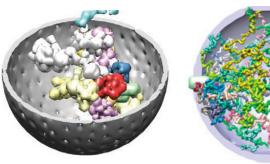
Acemel (2016) Nature Genetics



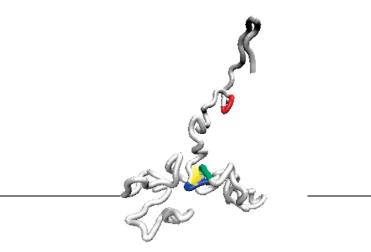
Jhunjhunwala (2008) Cell



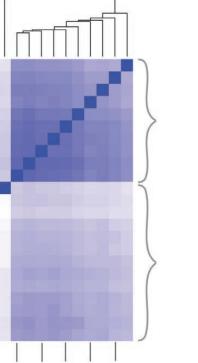
Fraser (2009) Genome Biology Ferraiuolo (2010) Nucleic Acids Research



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



Giorgetti, (2014) Cell







Research

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,*}

¹EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Barcelona, Spain, ²Universitat Pompeu Fabra (UPF), Barcelona, Spain, ³Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Barcelona, Spain, ⁴Genome Biology Group, Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Spain and ⁵Institució 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 meanfield 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 epresentation 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).

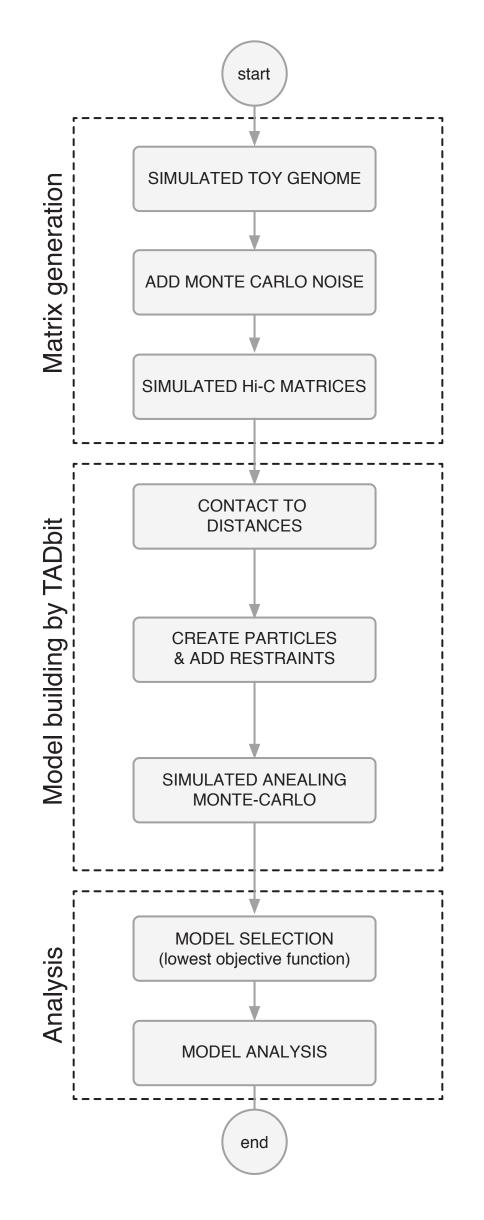
We recently developed a mean-field method for modeling 3D structures of genomes and genomic domains based on 3C interaction data (9). Our approach, called TADbit, was developed around the Integrative Modeling Platform (IMP, http://integrativemodeing.org), a general framework for restraint-based modeling 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, he final models were partially validated by assessing their

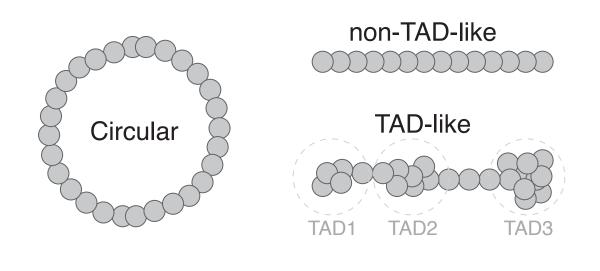
*To whom correspondence should be addressed. Tel: +34 934 020 542; Fax: +34 934 037 279; Email: mmarti@pcb.ub.cat

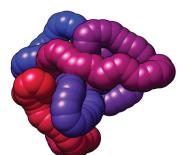
© The Author(s) 2015. Published by Oxford University Press on behalf of Nucleic Acids Research.

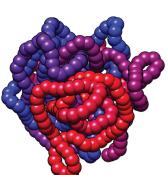
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creative.com/doc/access/article/ac ons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

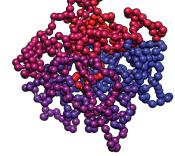
Toy models (168)



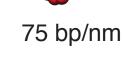


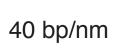


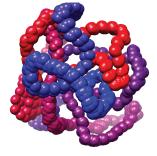




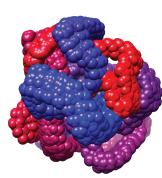
150 bp/nm



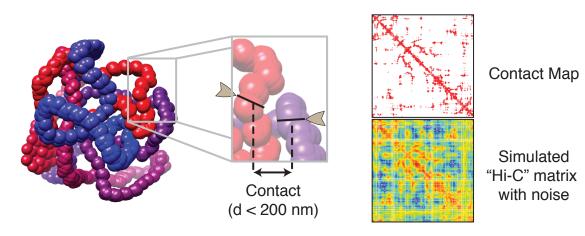




set 0 (Δ ts = 10⁰)

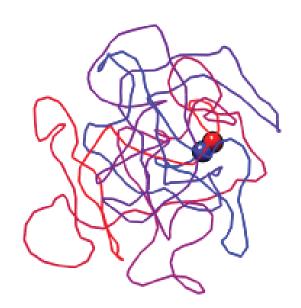


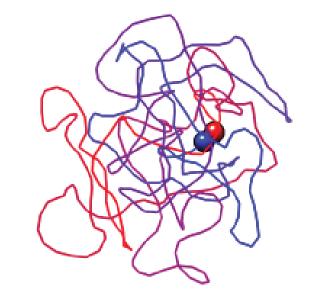
- set 1 ($\Delta ts = 10^{1}$)
- set 2 (Δ ts = 10²)

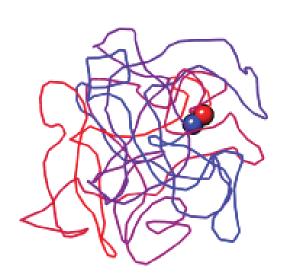


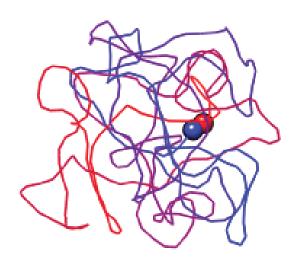


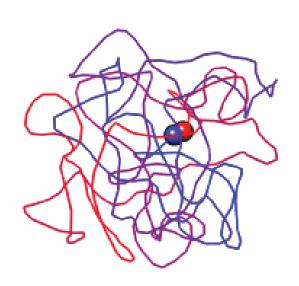
Toy interaction matrices

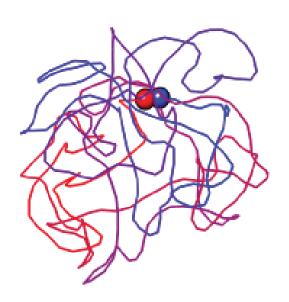


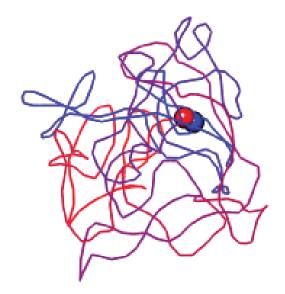




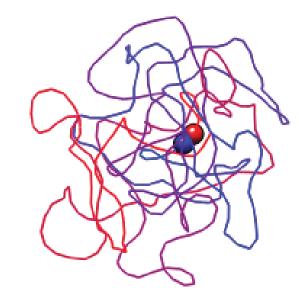


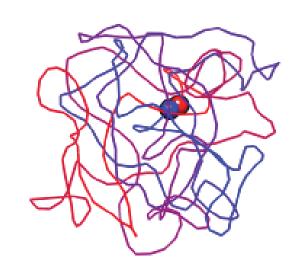


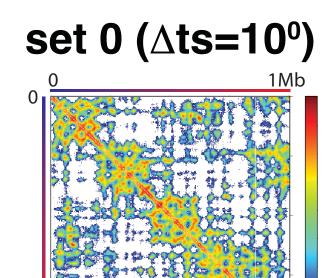






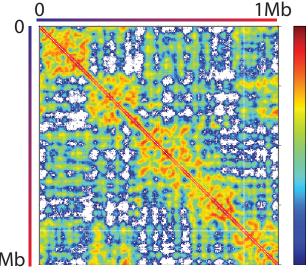


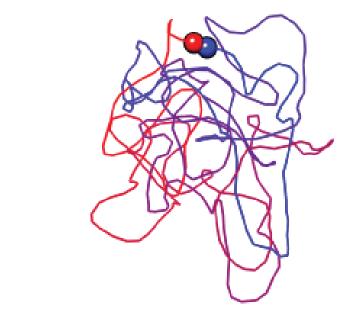


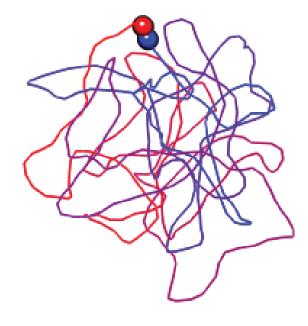


1Mb

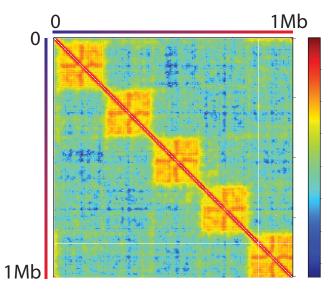
set 4 (∆ts=10⁴)



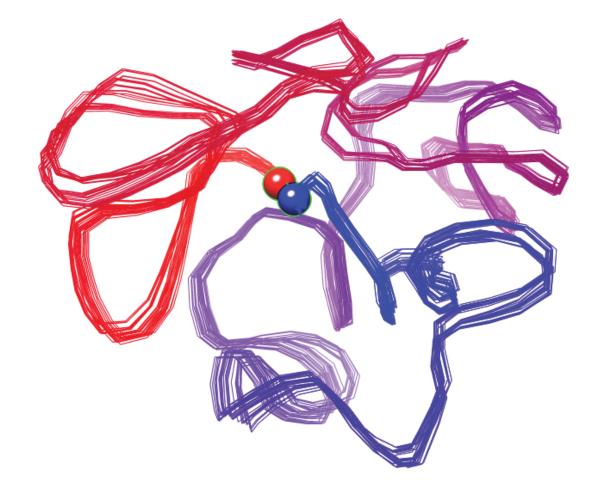


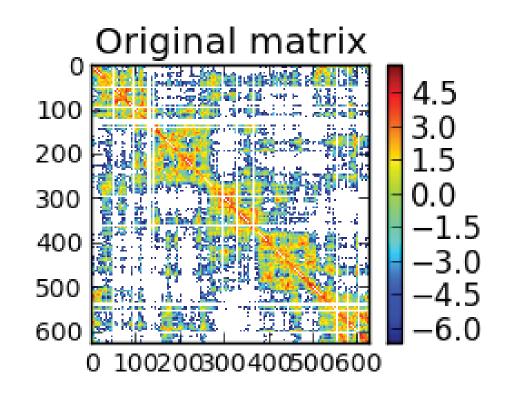


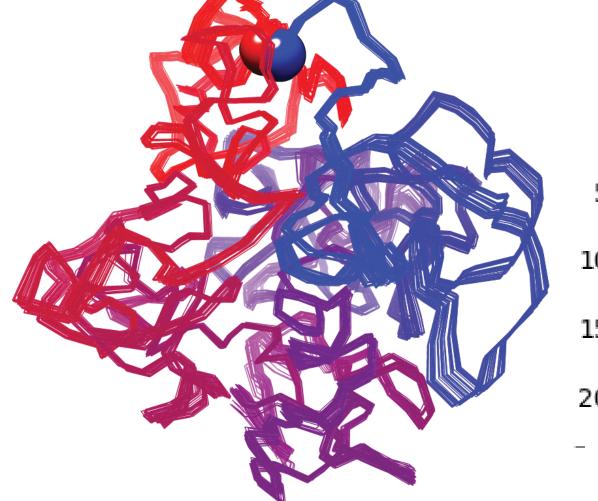


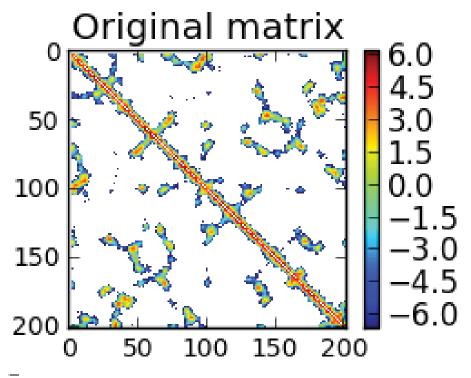


Reconstructing toy models









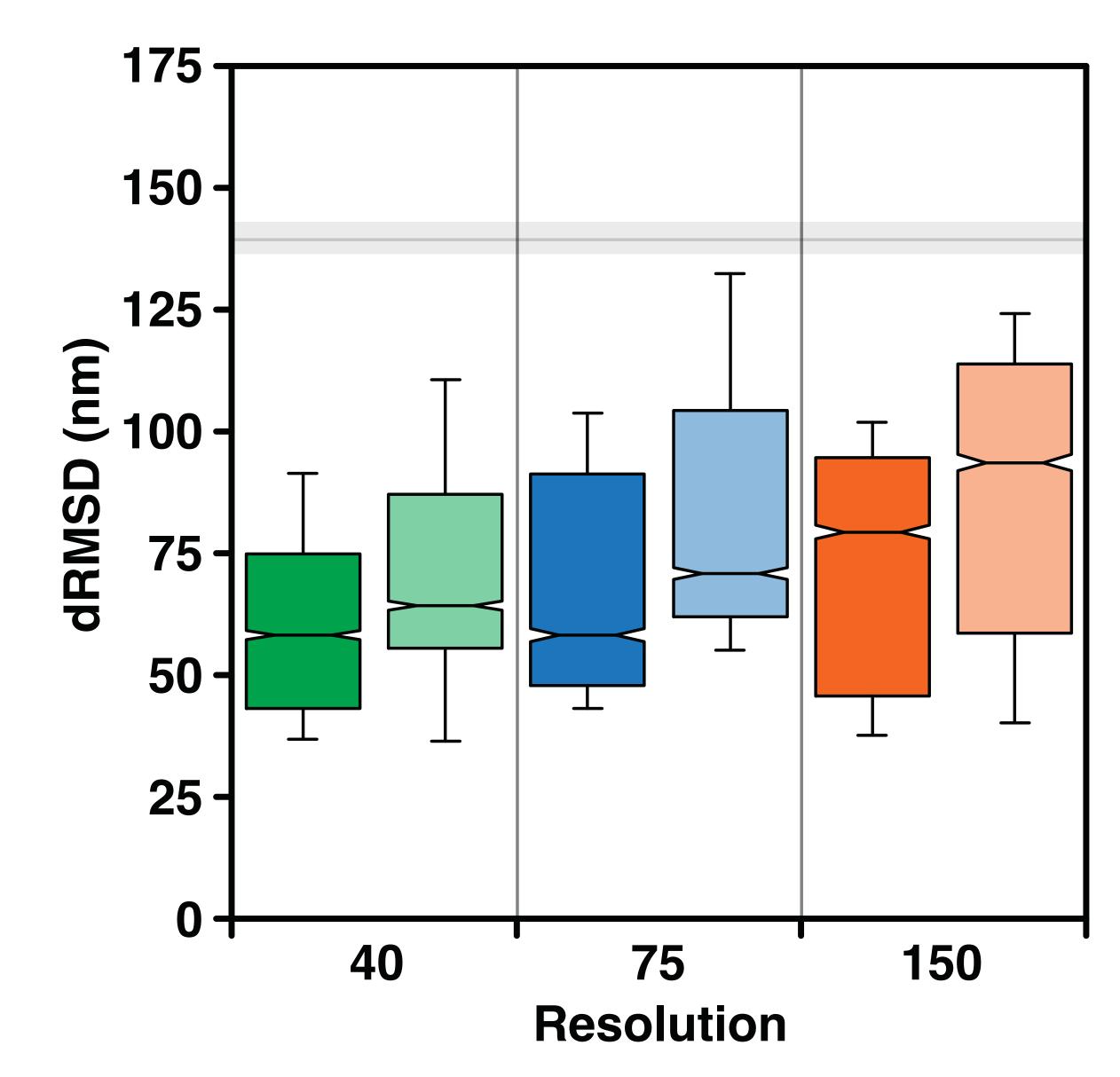


chr40_TAD **α=100 ∆ts=10** TADbit-SCC: 0.91 <dRMSD>: 32.7 nm <dSCC>: 0.94

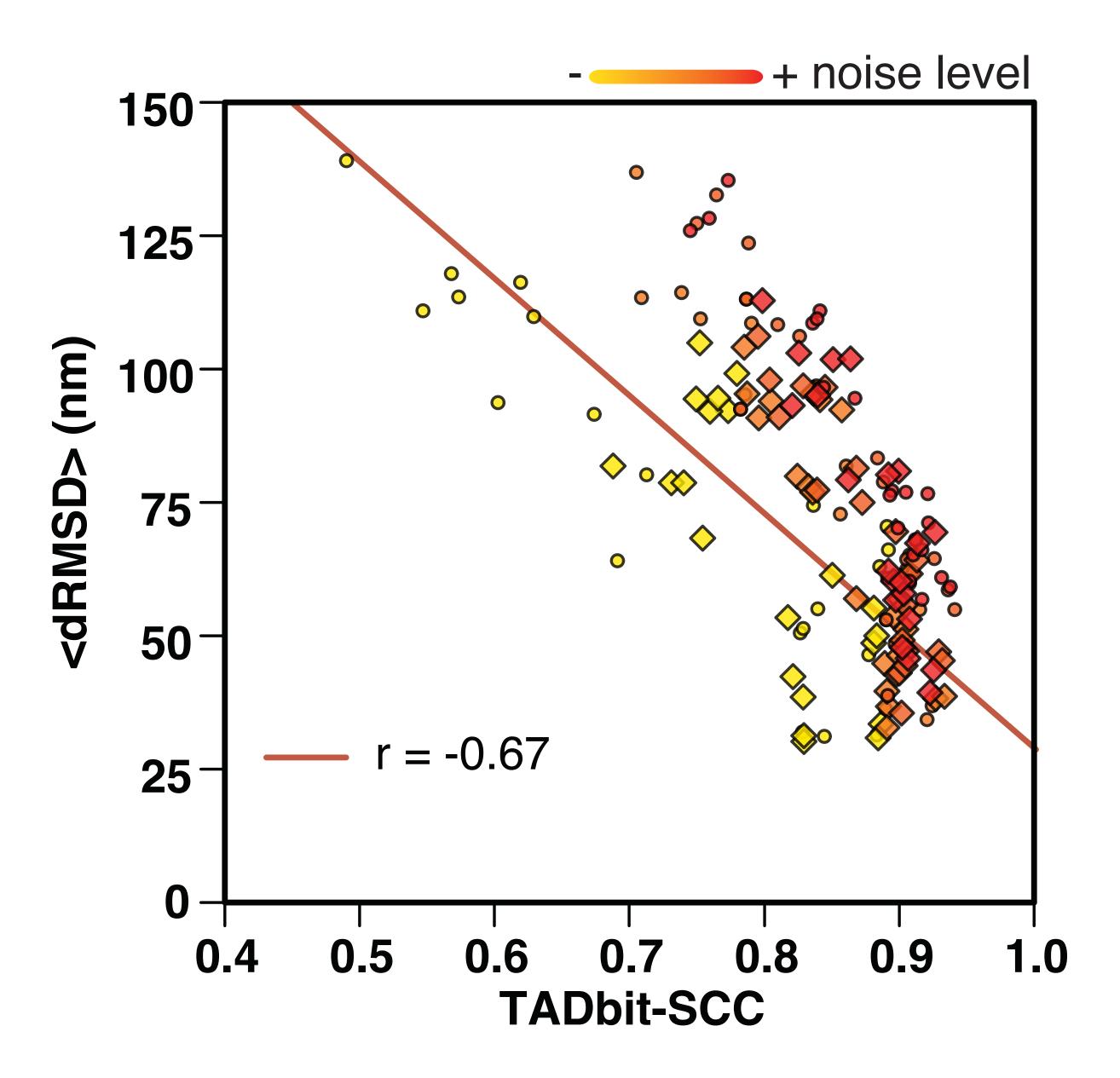
chr150_TAD **α=50** ∆ts=1 TADbit-SCC: 0.82 <dRMSD>: 45.4 nm <dSCC>: 0.86



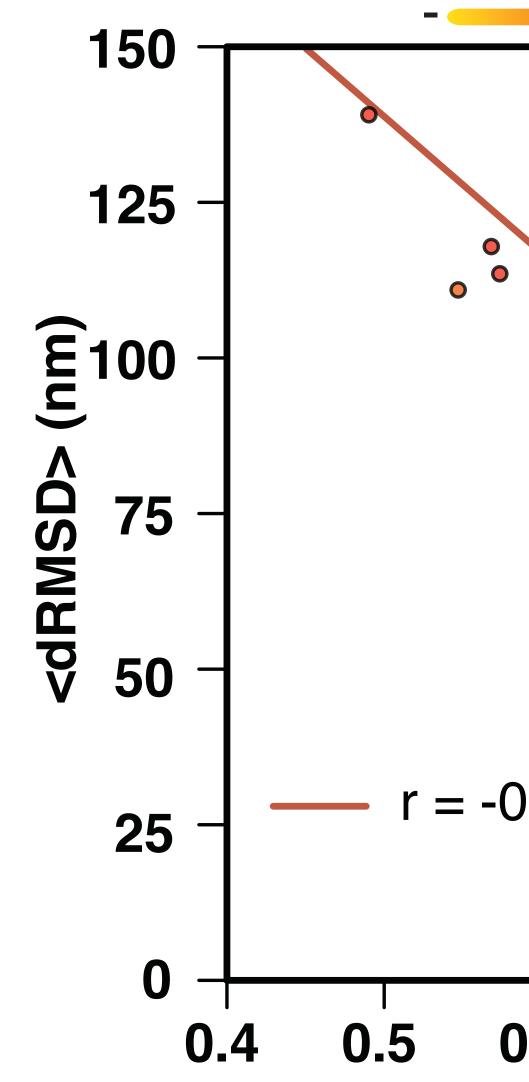
TADs & higher-res are "good"



Noise is "OK"



Structural variability is "NOT OK" + structural variability 150 125 -0 0 0 0 0 ♦ • \diamond 0 r = -0.6725 0 _ 0.6 0.7 0.8 0.9 1.0 0.4 0.5 **TADbit-SCC**







homogenize your cell population!

Higher-res is "good"

put your \$\$ in sequencing

Noise is "OK"

no need to worry much

Structural variability is "NOT OK"

Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae

Trussart et al. Nature Communications (2017) 8 14665



ARTICLE

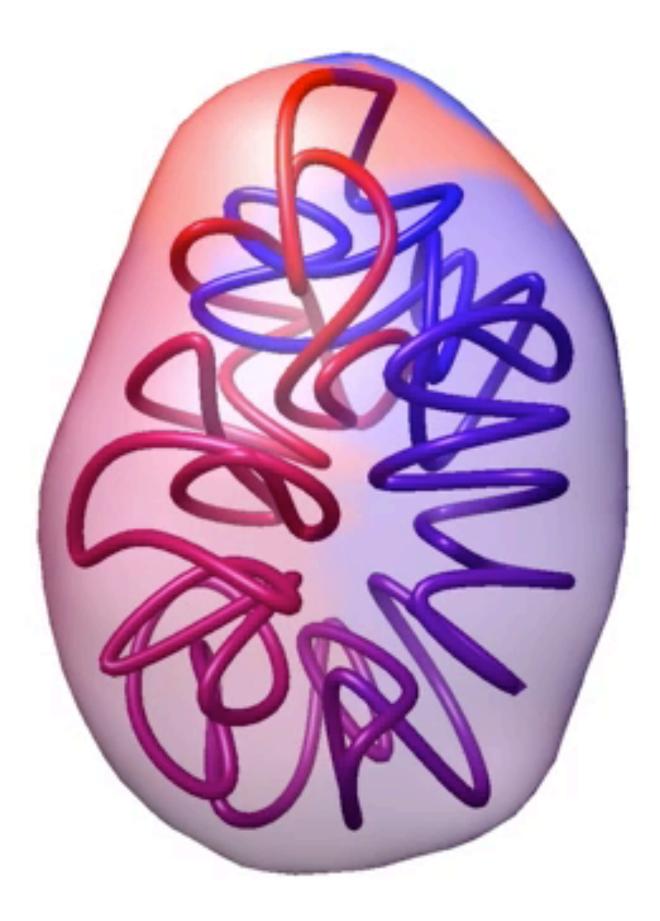
Received 4 Nov 2016 | Accepted 20 Jan 2017 | Published 8 Mar 2017 Dol: 10.1038/ncomms14665 OPEN Defined chromosome structure in the genomereduced bacterium Mycoplasma pneumoniae

Marie Trussart^{1,2}, Eva Yus^{1,2}, Sira Martinez¹, Davide Baù^{3,4}, Yuhei O. Tahara^{5,6}, Thomas Pengo^{1,7}, Michael Widjaja⁸, Simon Kretschmer⁹, Jim Swoger^{1,2}, Steven Djordjevic⁸, Lynne Turnbull⁸, Cynthia Whitchurch⁸, Makoto Miyata^{5,6}, Marc A. Marti-Renom^{2,3,4,10}, Maria Lluch-Senar^{1,2} & Luís 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. This study extends the current understanding of bacterial genome organization and demonstrates that a defined chromosomal structure is a universal feature of living systems.

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

NATURE COMMUNICATIONS | 8:14665 | DOI: 10.1038/ncomms14665 | www.nature.com/naturecommunication



Mycoplasma is a small genome with few structural factors

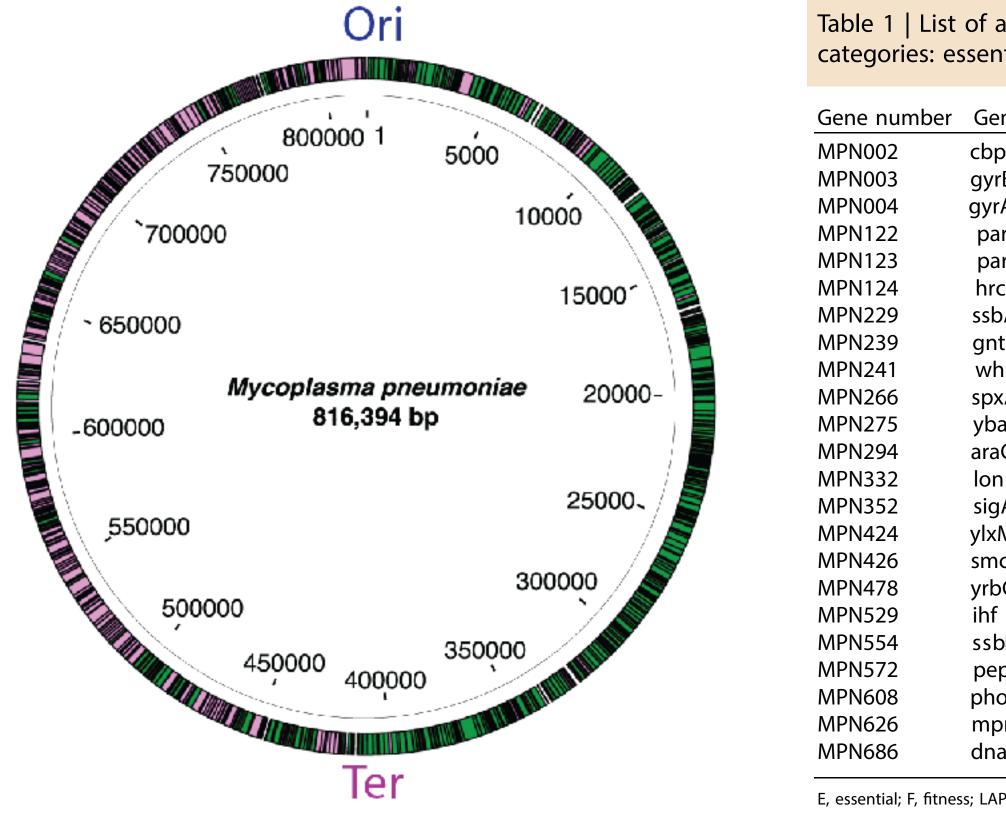
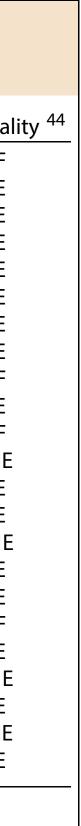
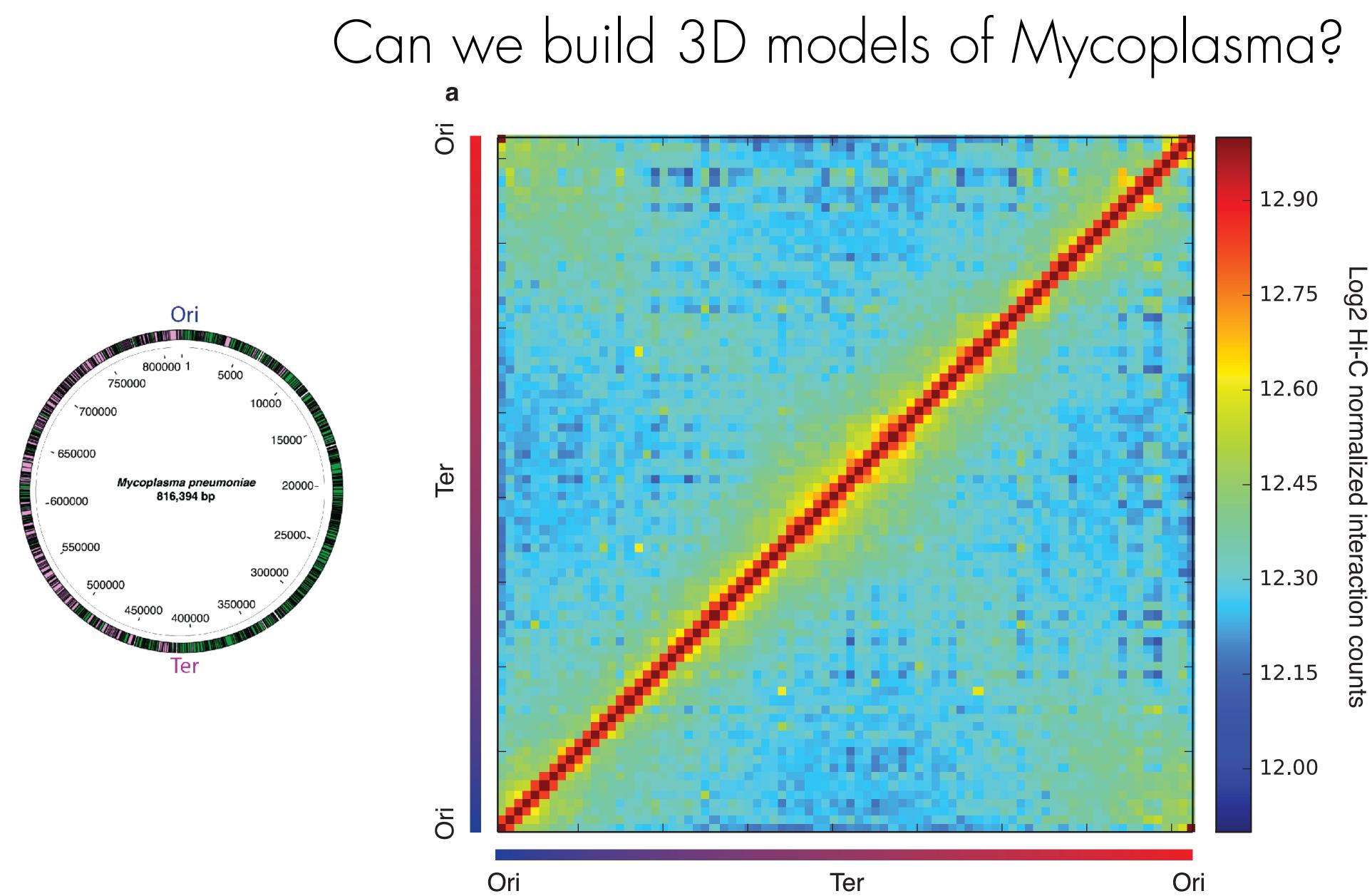


Table 1 | List of assigned transcription factors, sigma factors and structural proteins and essentiality with three distinct categories: essential (E), non -essential (NE) and fitness (F).

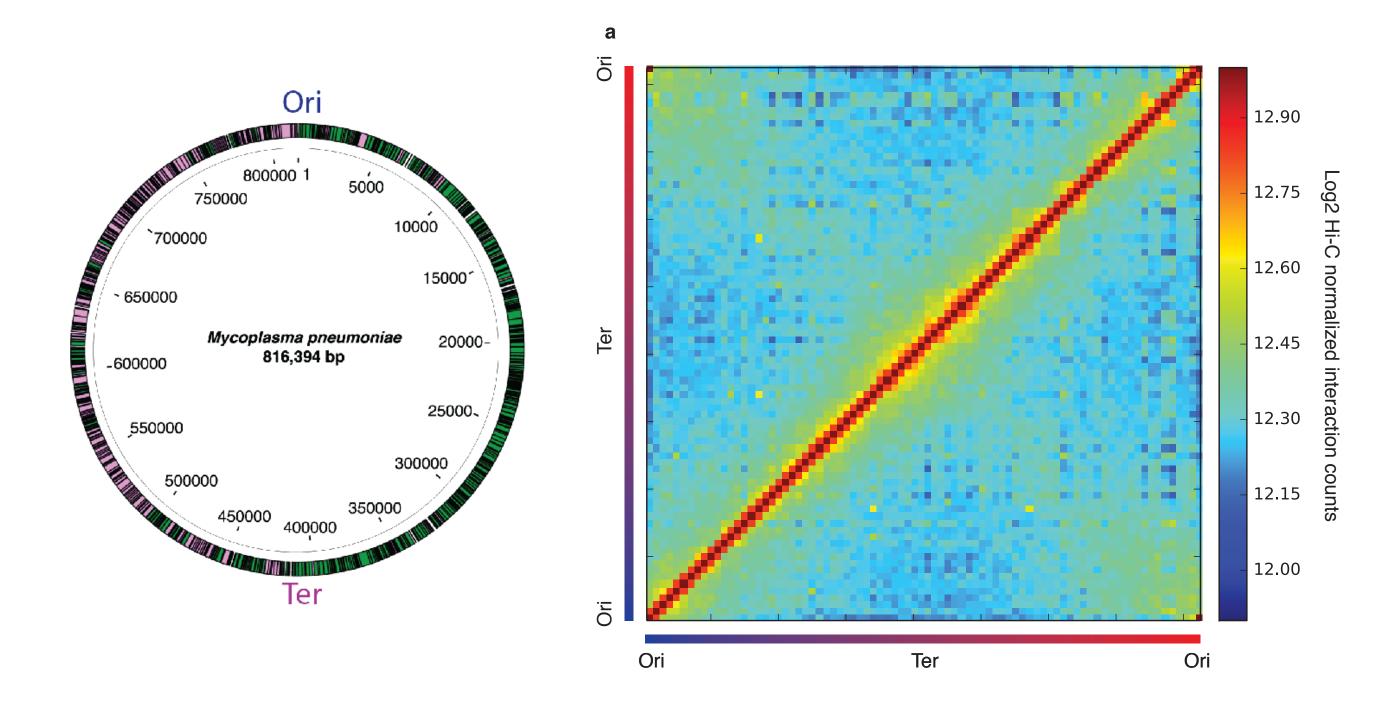
ene name	Protein name	Essentiali
рА	Curved DNA-binding protein CbpA	F
/rB	DNA gyrase subunit B	Е
٧rA	DNA gyrase subunit A	Е
arB	DNA topoisomerase 4 subunit B	Е
arC	DNA topoisomerase 4 subunit A	Е
rcA	Heat-inducible transcription repressor hrcA	Е
bA	SSB-binding ssDNA	Е
ntR	Probable HTH-type transcriptional regulator gntR	Е
/hiA	Transcription factor with WhiA C-terminal domain	F
bхA	Transcriptional regulator Spx	Е
заВ	DNA-binding protein, YbaB/EbfC family	F
aC	AraC-like transcriptional regulator	NE
n	ATP-dependent protease La (EC 3.4.21.53)	Е
gA	RNA polymerase sigma factor rpoD (Sigma-A) (EC 2.7.7.6)	Е
хM	Putative helix-turn-helix protein, YlxM/p13-like protein	NE
nc	SMC family, chromosome/DNA binding/protecting functions	Е
bC	YebC family protein (transcription factor of the tetR family)	Е
f	Histone-like bacterial DNA-binding protein	F
ъbВ	Putative single-stranded DNA-binding protein	Е
epA	Probable cytosol aminopeptidase (EC 3.4.11.1) (leucine aminopeptidase) (LAP) (leucyl aminopeptidase)	Е
noU	Transcriptional regulator involved in phosphate transport system	Е
pn626	Alternative sigma factor	NE
naA	Chromosomal replication initiator protein dnaA	E

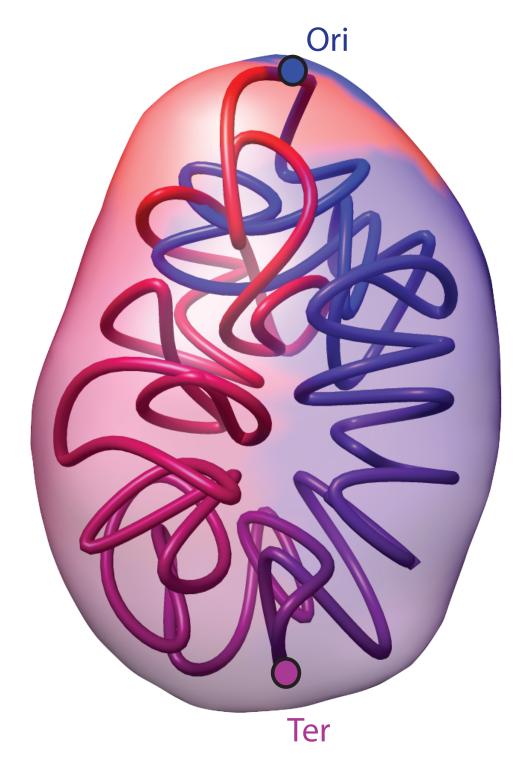
E, essential; F, fitness; LAP, leucine aminopeptidase; NE, non-essential; ssDNA, single-stranded DNA⁴⁴.



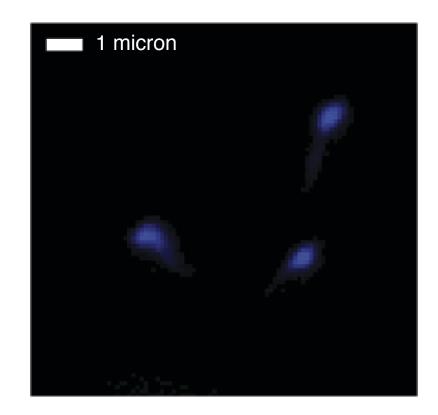


Can we build 3D models of Mycoplasma?



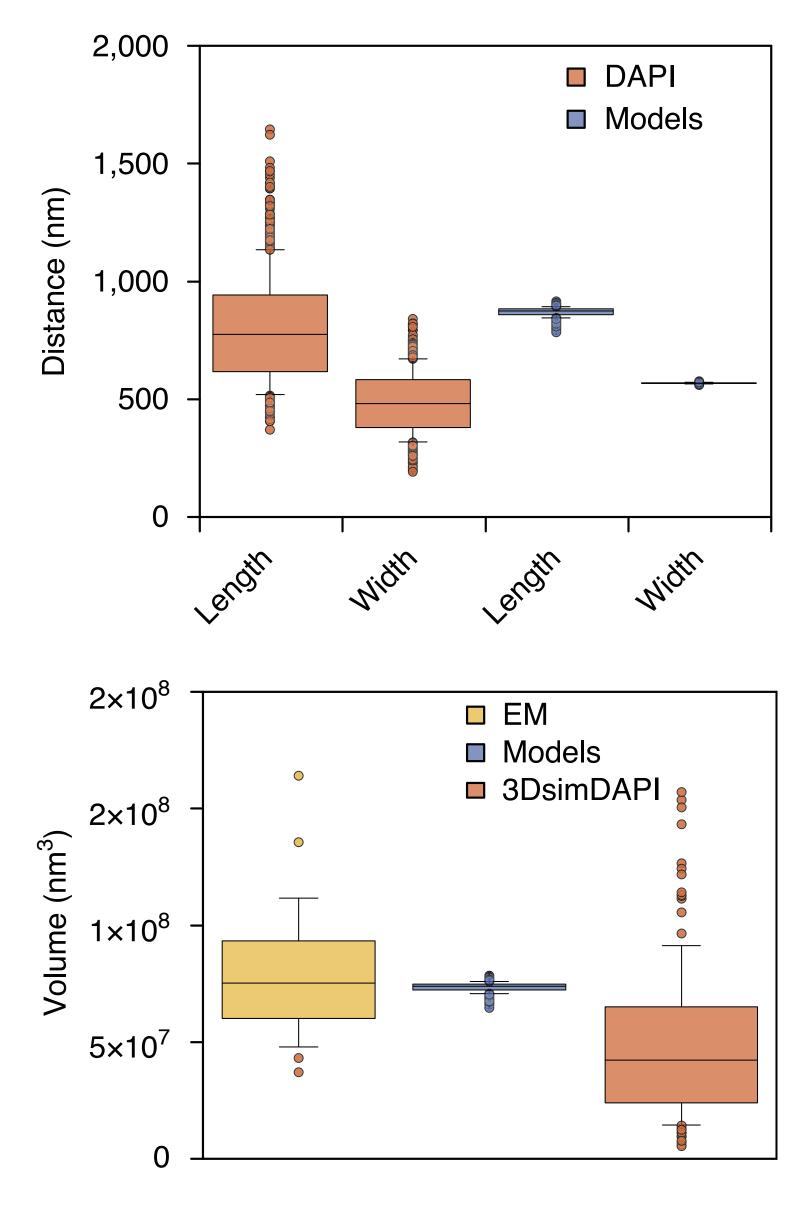


Is the overall 3D model accurate?

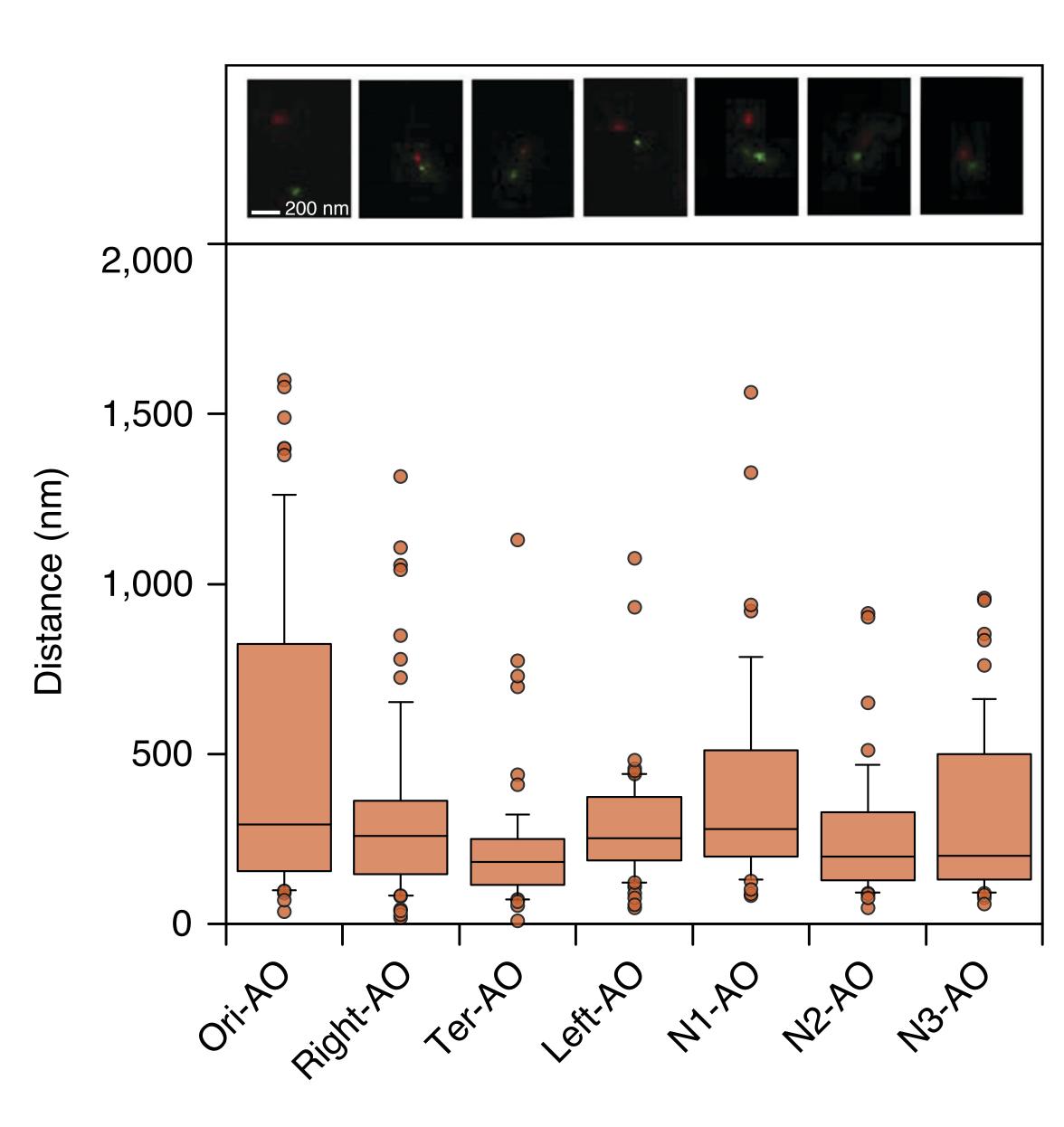


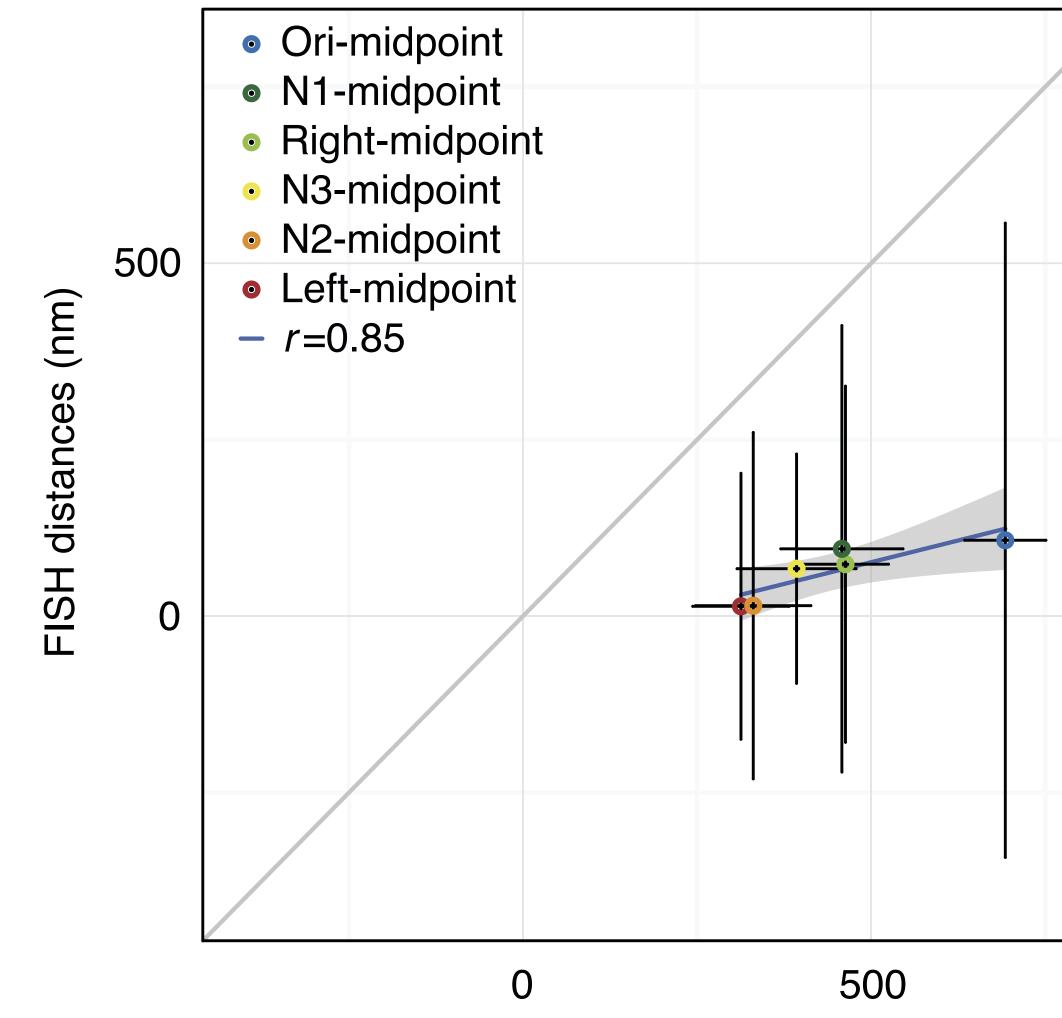


Makoto Miyata Group Osaka City University



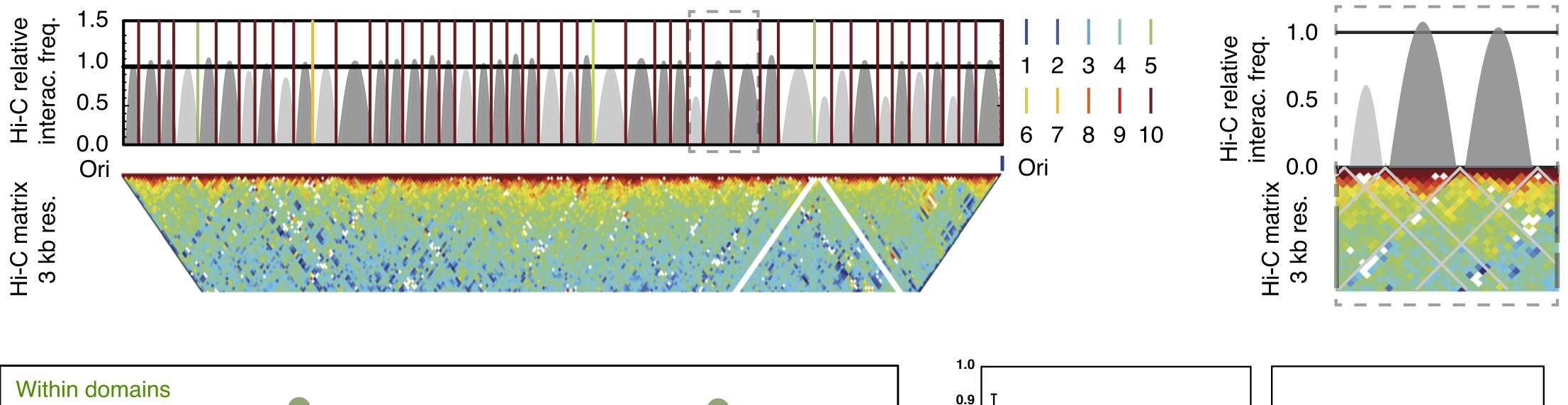
Are the details of the 3D model accurate?

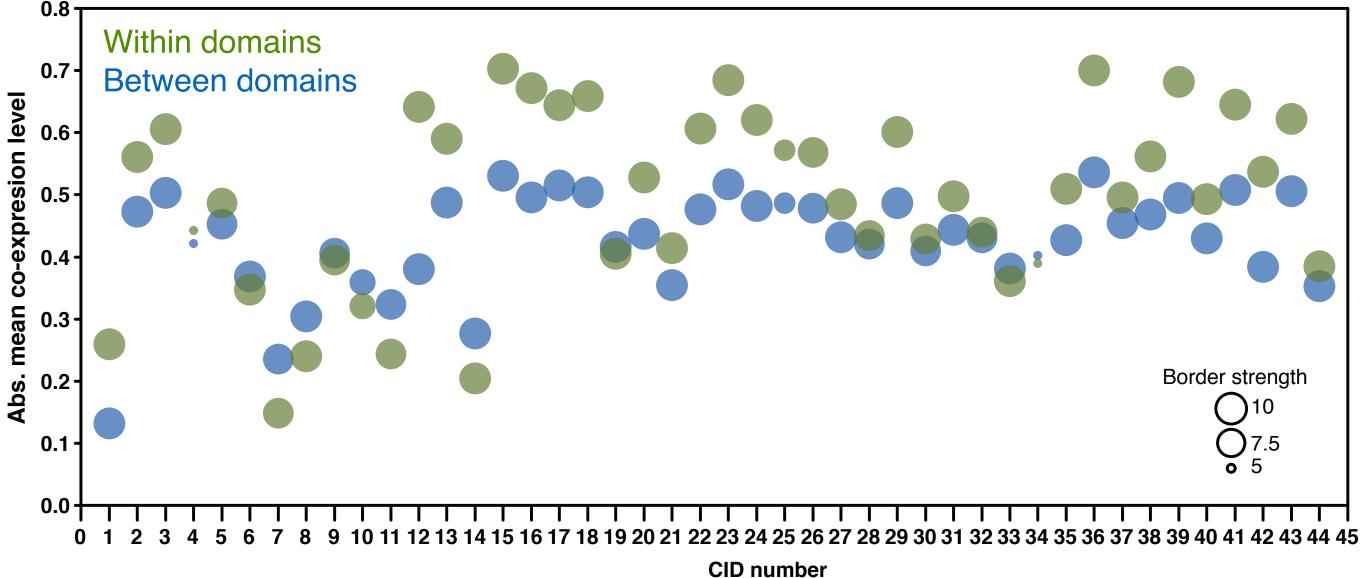




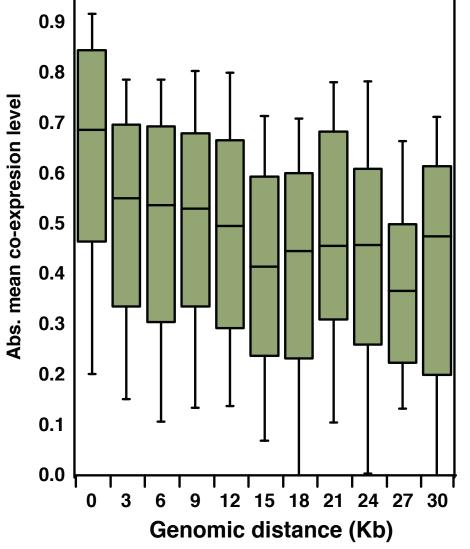
Model distances (nm)

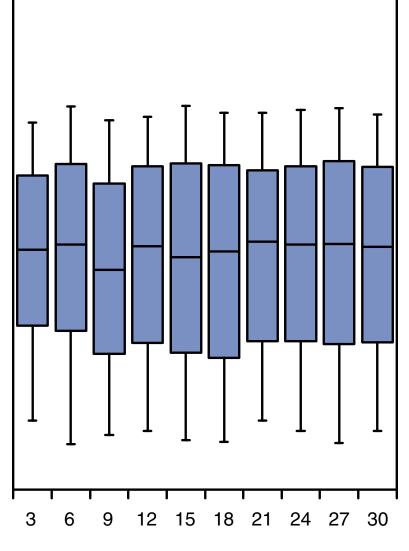




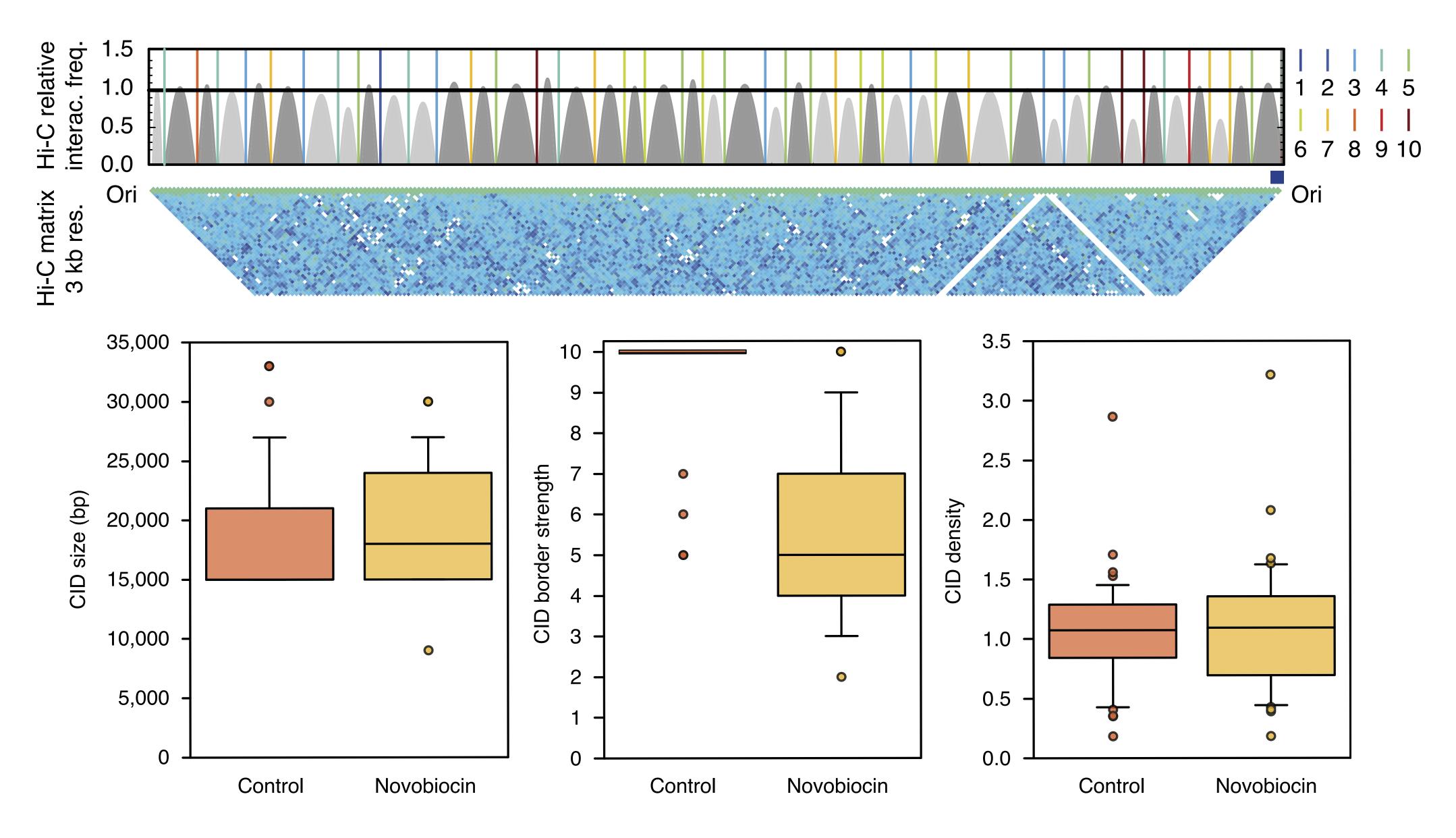


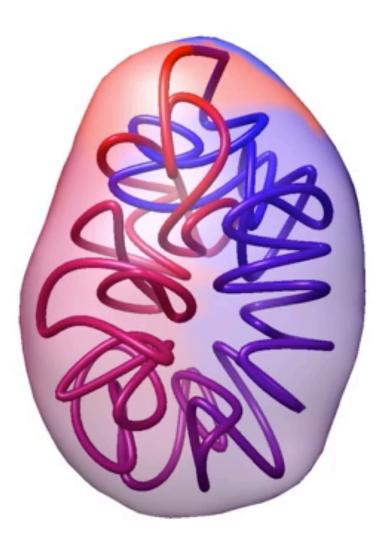
Mycoplasma genome is partitioned into co-regulated CIDs





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

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.

Mycoplasma has CIDs (TADs)



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

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



Marie Trussart Davide Baù Gireesh K. Bogu David Castillo Yasmina Cuartero Irene Farabella Silvia Galan Mike Goodstadt Julen Mendieta Juan Rodríguez-Pérez François Serra Paula Soler Yannick Spill Marco di Stefano

