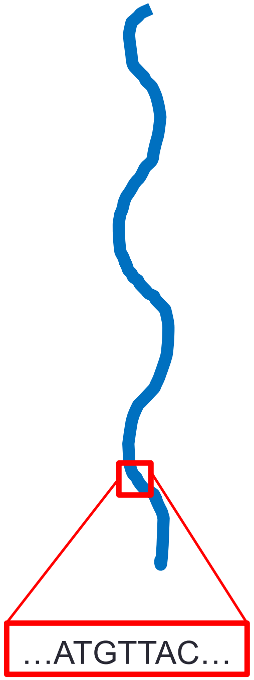
STUDYING 3D GENOME EVOLUTION USING GENOMIC SEQUENCE

Raphaël MOURAD, Assistant Professor,
Center of Integrative Biology
University Paul Sabatier, Toulouse III

INTRODUCTION

3D structure of chromosome

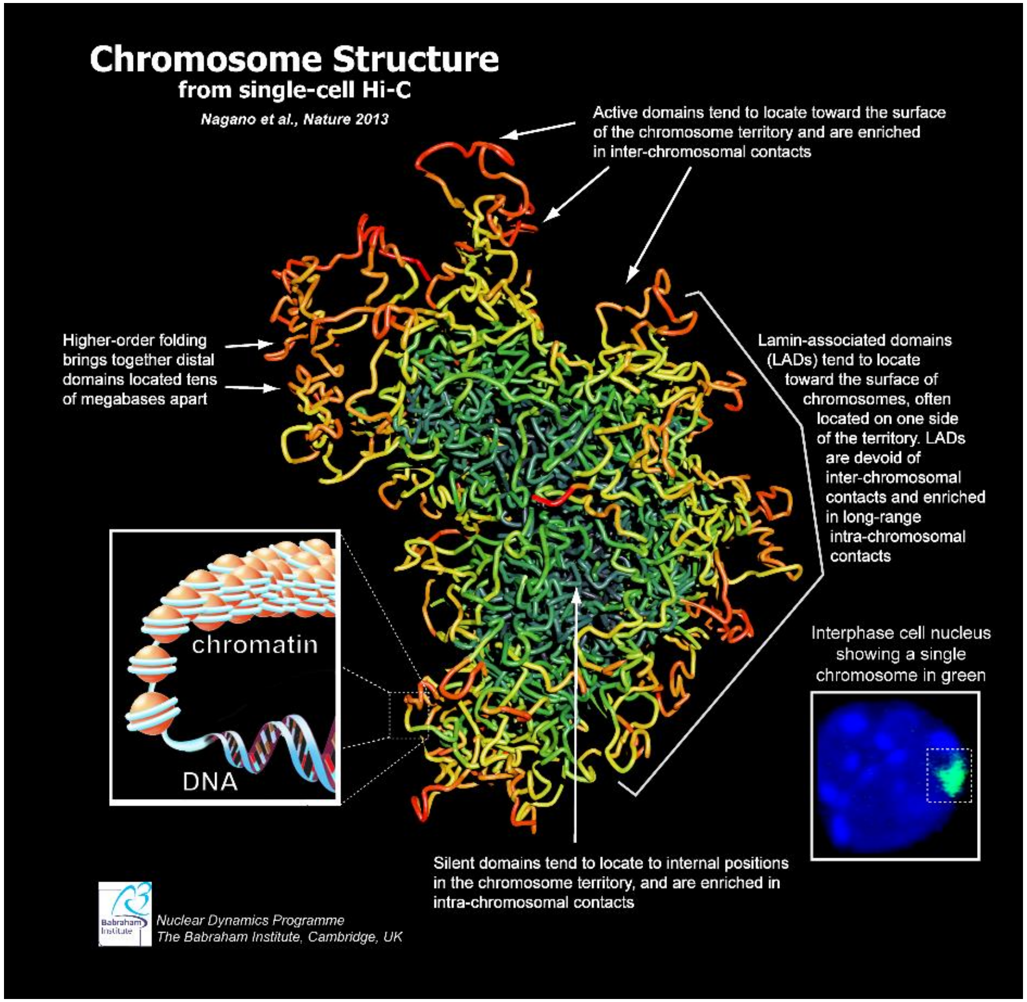
Chromosome
~ 10 cm long



…ATGTTAC…

DNA

Compaction into the cell nucleus (5 µm)



Chromosome Structure
from single-cell Hi-C

Nagano et al., Nature 2013

Active domains tend to locate toward the surface of the chromosome territory and are enriched in inter-chromosomal contacts

Higher-order folding brings together distal domains located tens of megabases apart

Lamin-associated domains (LADs) tend to locate toward the surface of chromosomes, often located on one side of the territory. LADs are devoid of inter-chromosomal contacts and enriched in long-range intra-chromosomal contacts

Interphase cell nucleus showing a single chromosome in green

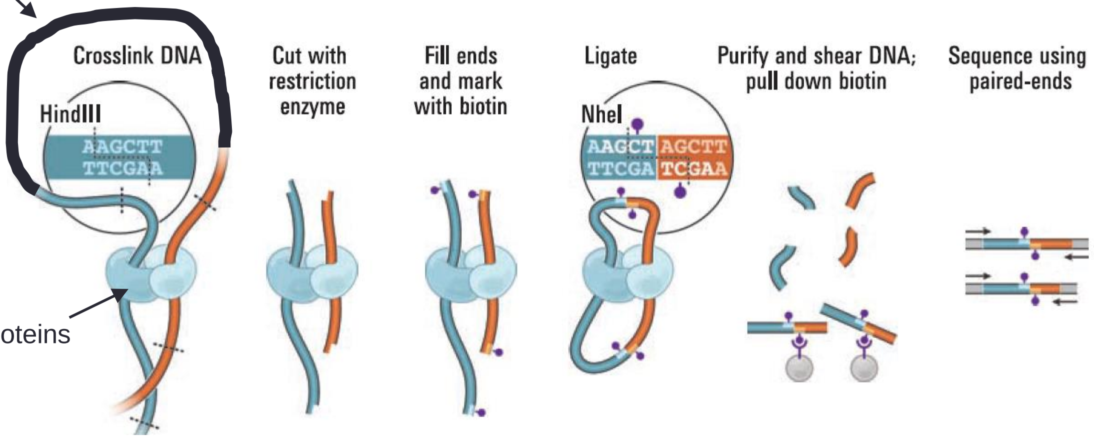
Silent domains tend to locate to internal positions in the chromosome territory, and are enriched in intra-chromosomal contacts

Nuclear Dynamics Programme
The Babraham Institute, Cambridge, UK

Standard analysis of 3D genome using Hi-C

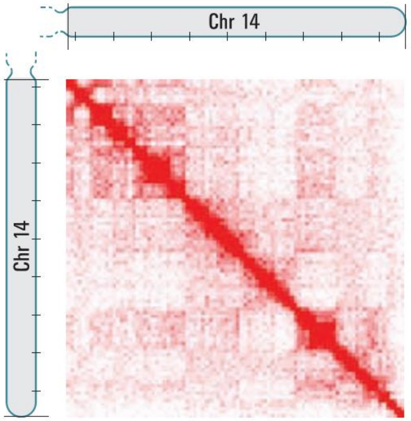
High-throughput chromosome conformation capture (Hi-C)
(Lieberman-Aiden et al, Science, 2009).

Next-Gen Sequencing Experiment

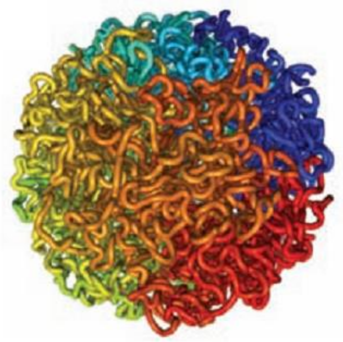


Results

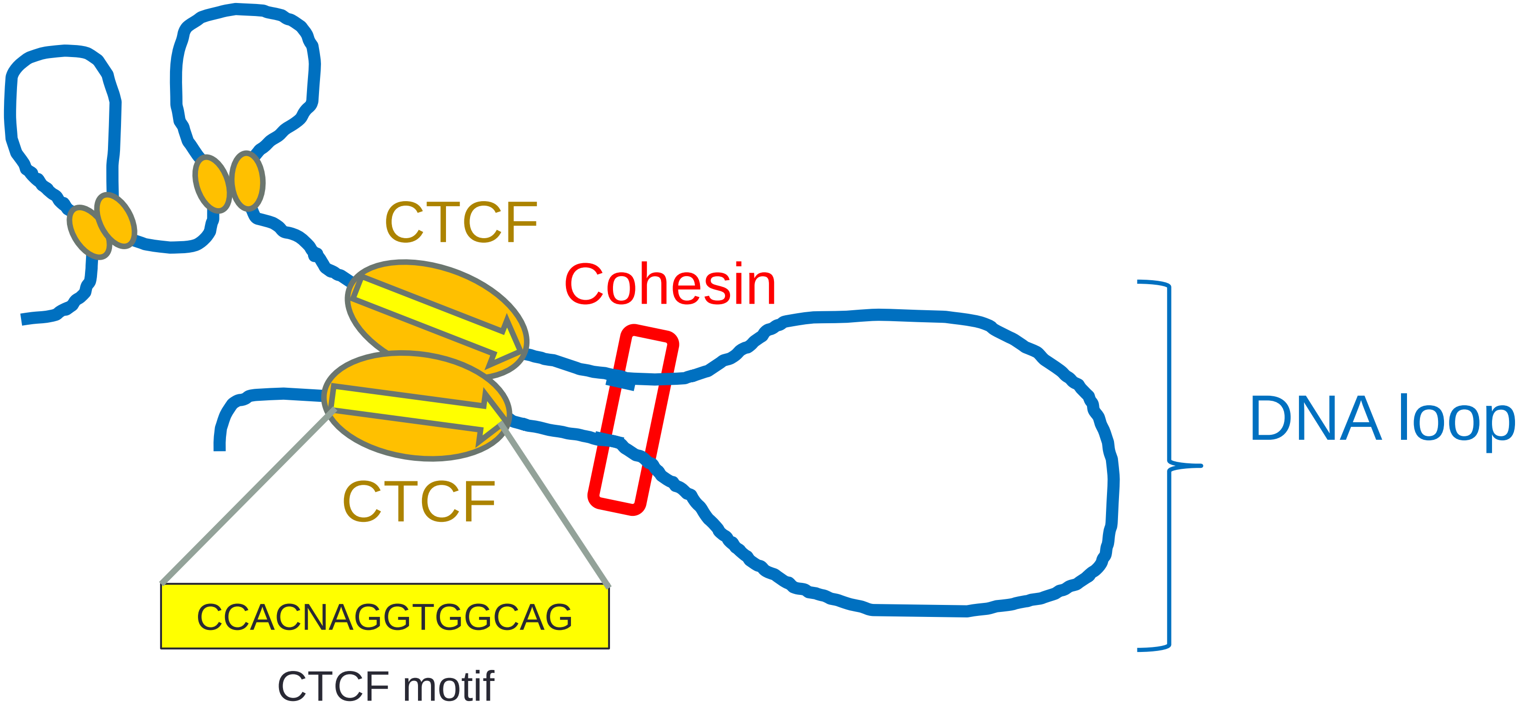
Interaction heatmap



3D structure of chromosome



Chromosomes are organized in loops mediated by CTCF and cohesin in vertebrates

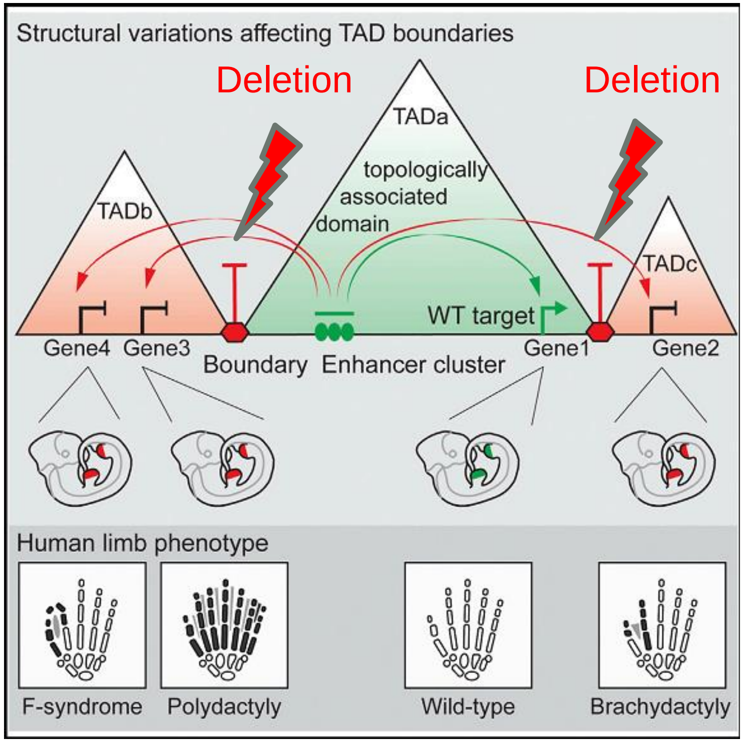


Rao et al., Cell 2015; Sanborn et al., PNAS 2015.

The biological role of 3D genome

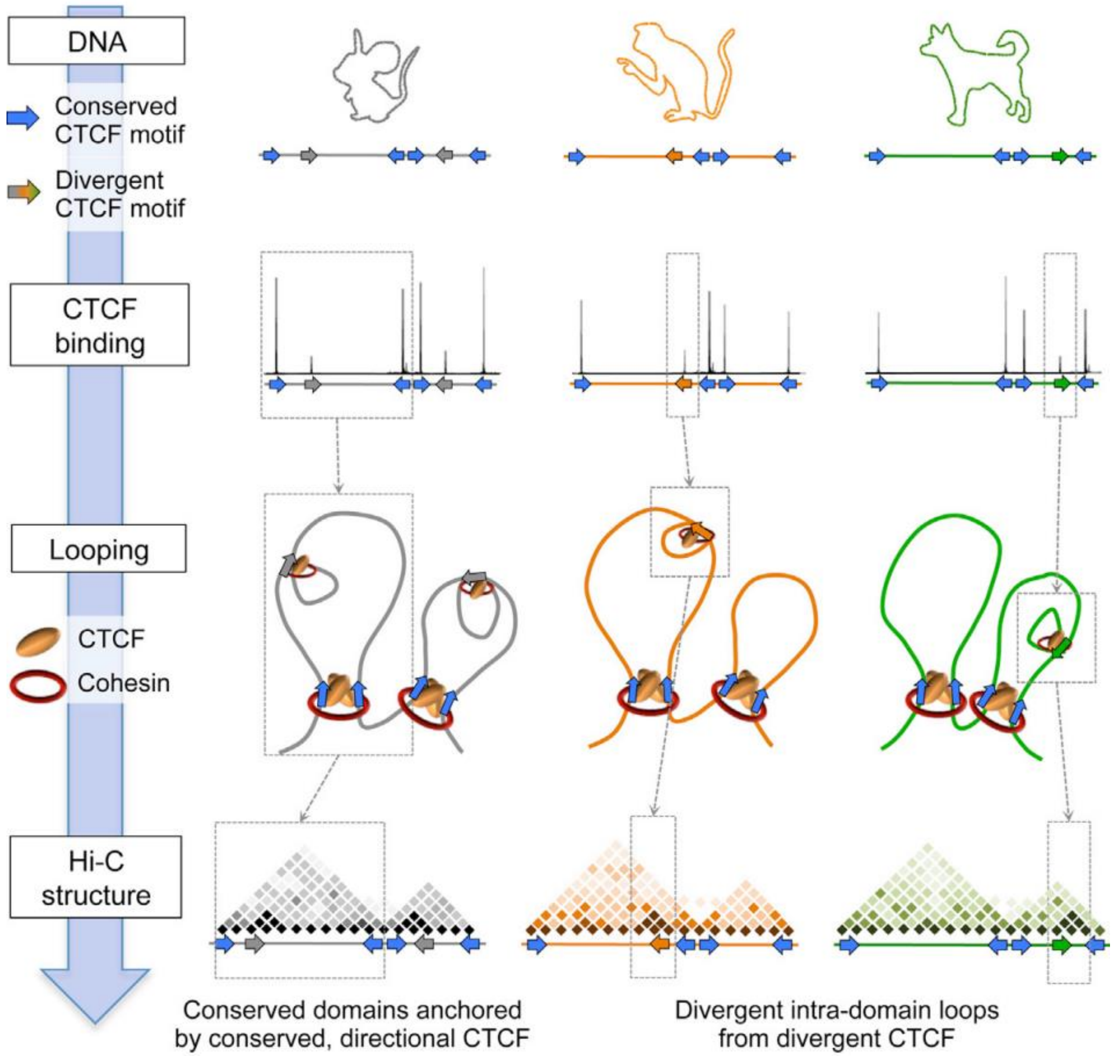
- 3D genome play an important role in:
  - gene expression regulation,
  - DNA replication
  - …

- Deletions of 3D domain boundaries can cause de novo enhancer-promoter interactions and misexpression, and can lead to **genetic diseases and cancer**.



Lupianez et al., Cell, 2015.

3D genome evolves



- Changes in CTCF motif position and orientation leads to 3D genome evolution.

Rudan et al.,
Cell Reports 2015.

HOW TO STUDY 3D GENOME EVOLUTION?

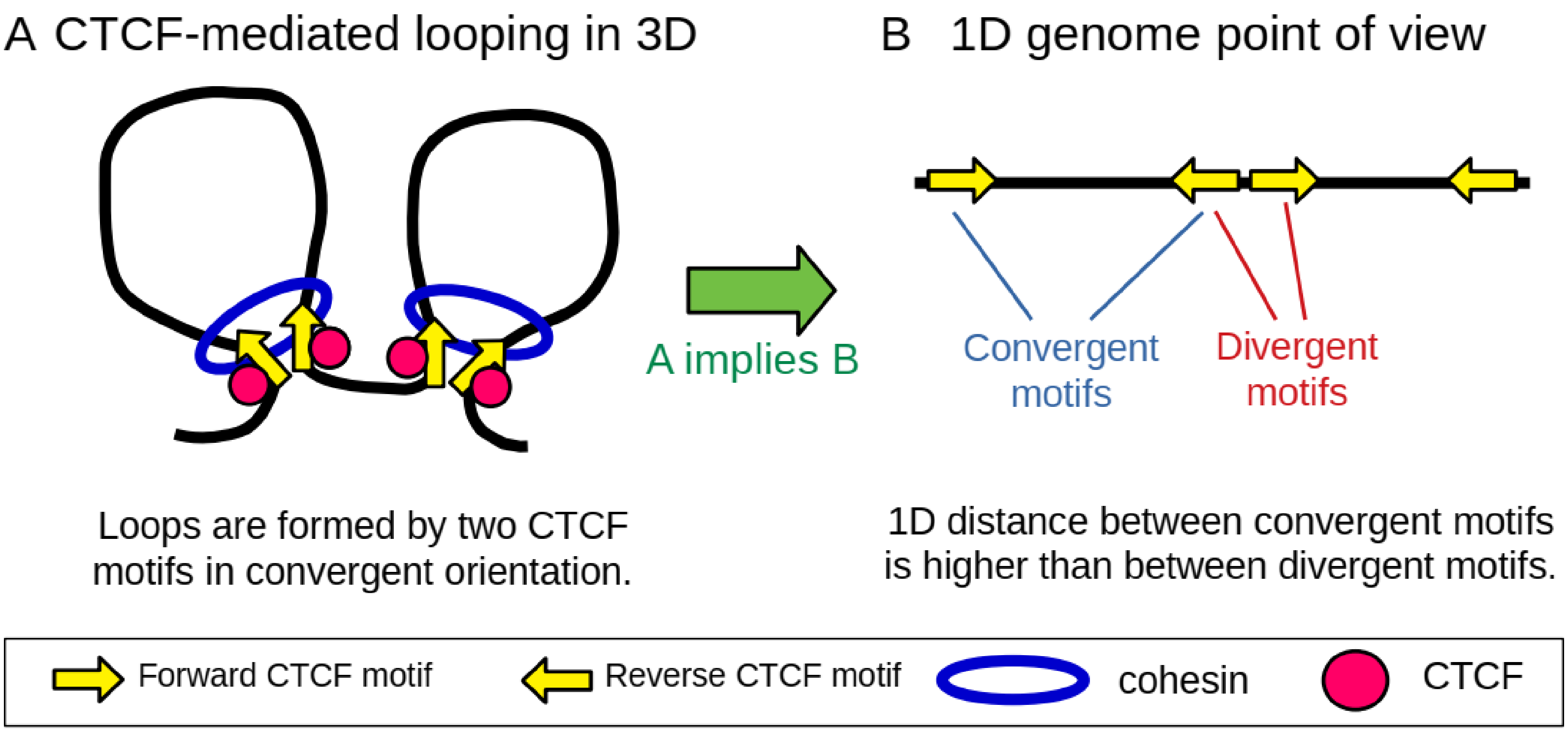
Not enough proper Hi-C data to study the evolution of 3D genome

- There are only 10-20 species whose Hi-C data are available to study the evolution of CTCF-mediated looping in vertebrates (for which CTCF is conserved).

- Available data between species are often from different tissues, which do not allow proper comparisons.

- Most Hi-C data do not have enough resolution to study CTCF-mediated looping.

1D genome point of view of CTCF-mediated looping in 3D



A CTCF-mediated looping in 3D

B 1D genome point of view

A implies B

Convergent motifs

Divergent motifs

Loops are formed by two CTCF motifs in convergent orientation.

1D distance between convergent motifs is higher than between divergent motifs.

⇨ Forward CTCF motif

⇦ Reverse CTCF motif

⬭ cohesin

● CTCF

- The distance between convergent CTCF motifs is expected to be higher than the distance between divergent CTCF motifs.

3D ratio (3DR) to assess CTCF looping

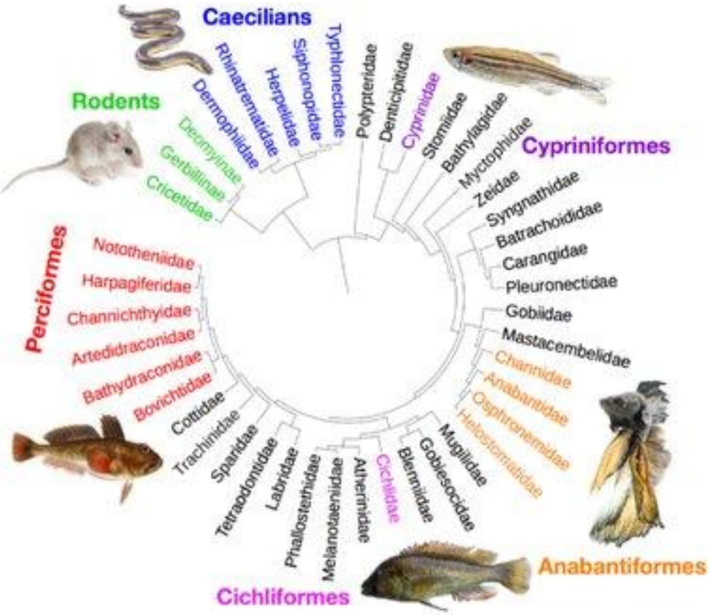
$$3\mathrm{DR} = median(d_{\rightarrow\leftarrow})/median(d_{\leftarrow\rightarrow}),$$

which is the ratio of 2 medians: the median of the distances between two contiguous motifs in convergent orientation (noted “→←"), and the median of the distances between two contiguous motifs in divergent orientation (noted "←→").

3DR is expected to be > 1, if the 3D genome comprises CTCF-mediated CTCF looping!

Thousands of genome sequences can be studied using 3DR!

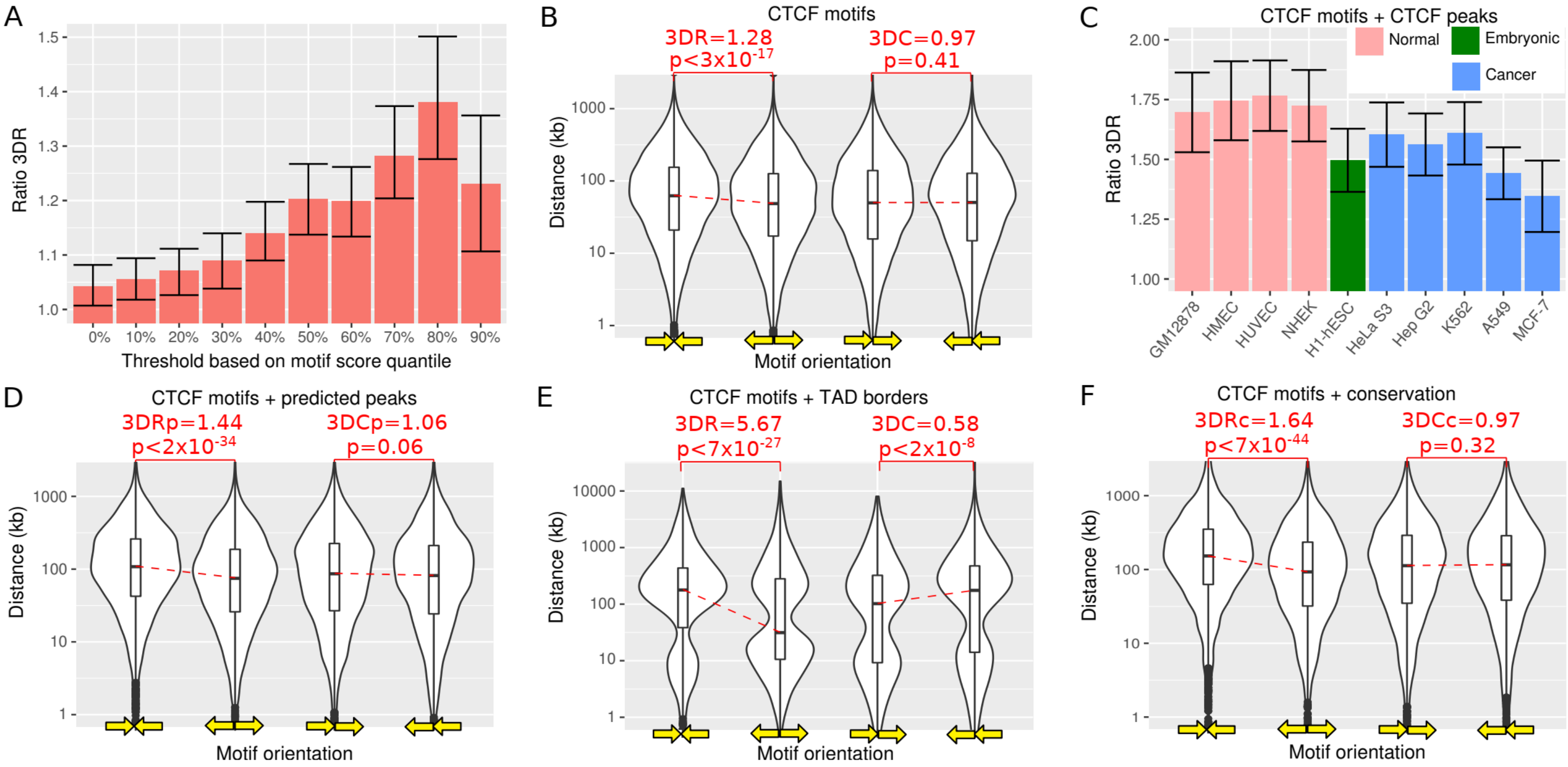
VERTEBRATE GENOMES PROJECT



- CTCF motif is conserved among vertebrates!
- Vertebrate Genomes Project plans to sequence all vertebrates on Earth (66K species)!

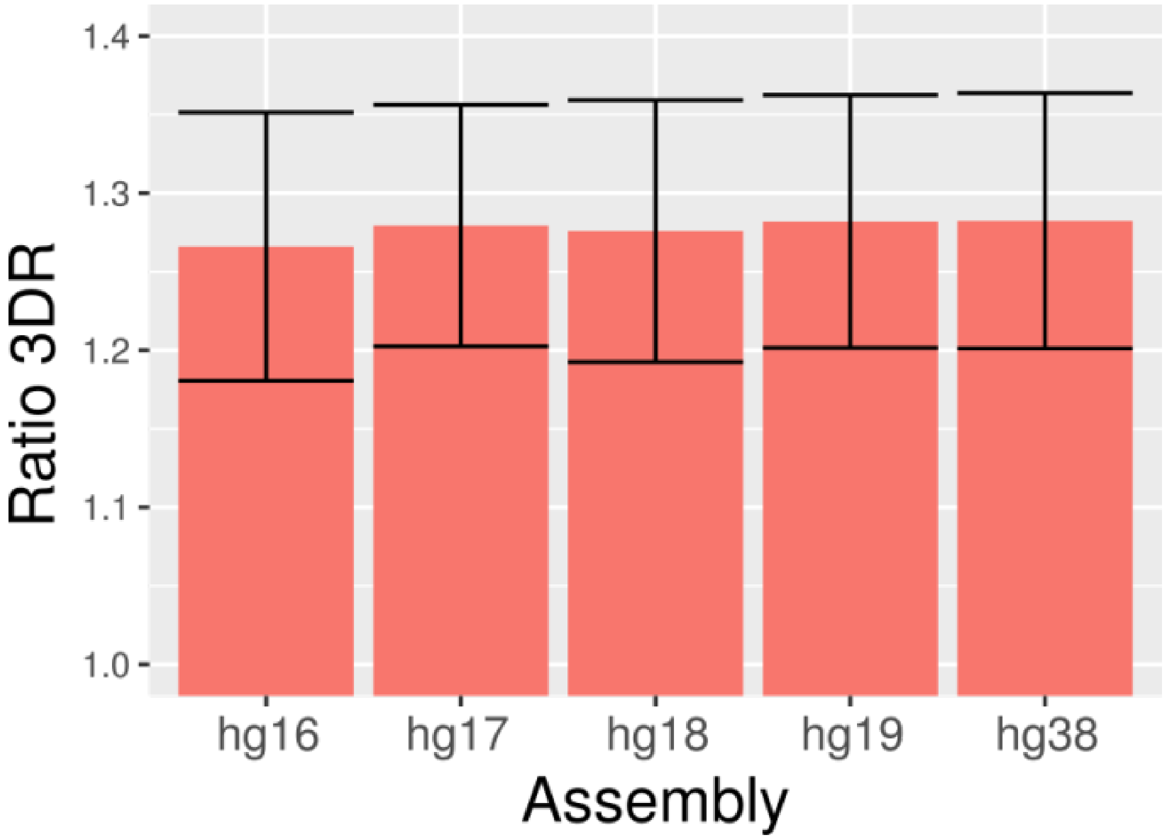
RESULTS

3DR detects CTCF-mediated looping in human

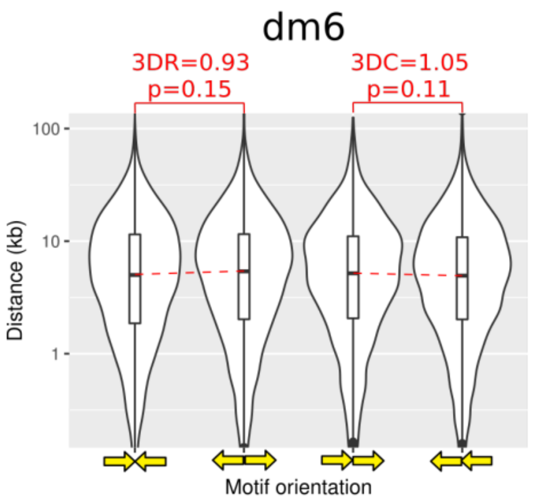


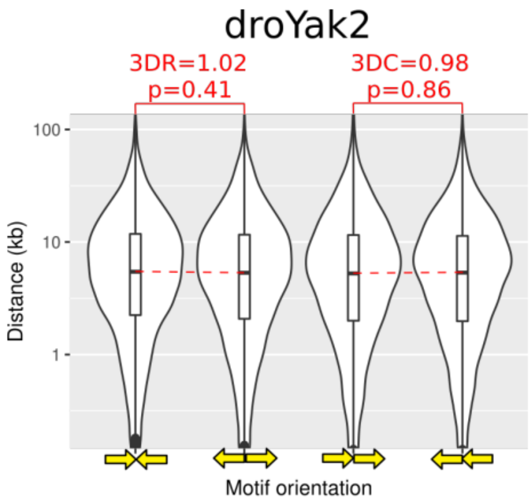
- 3DR can detect CTCF-mediated looping in the human genome.
- 3DR is improved by using ChIP-seq, predicted ChIP-seq or conservation data.

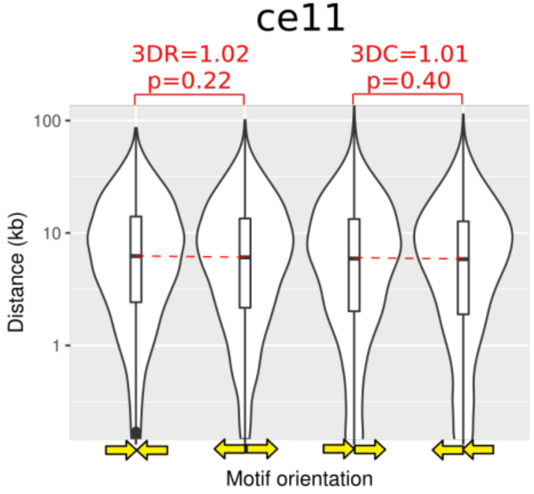
3DR is stable between recent human genome assemblies



3DR = 1 for genomes without CTCF looping

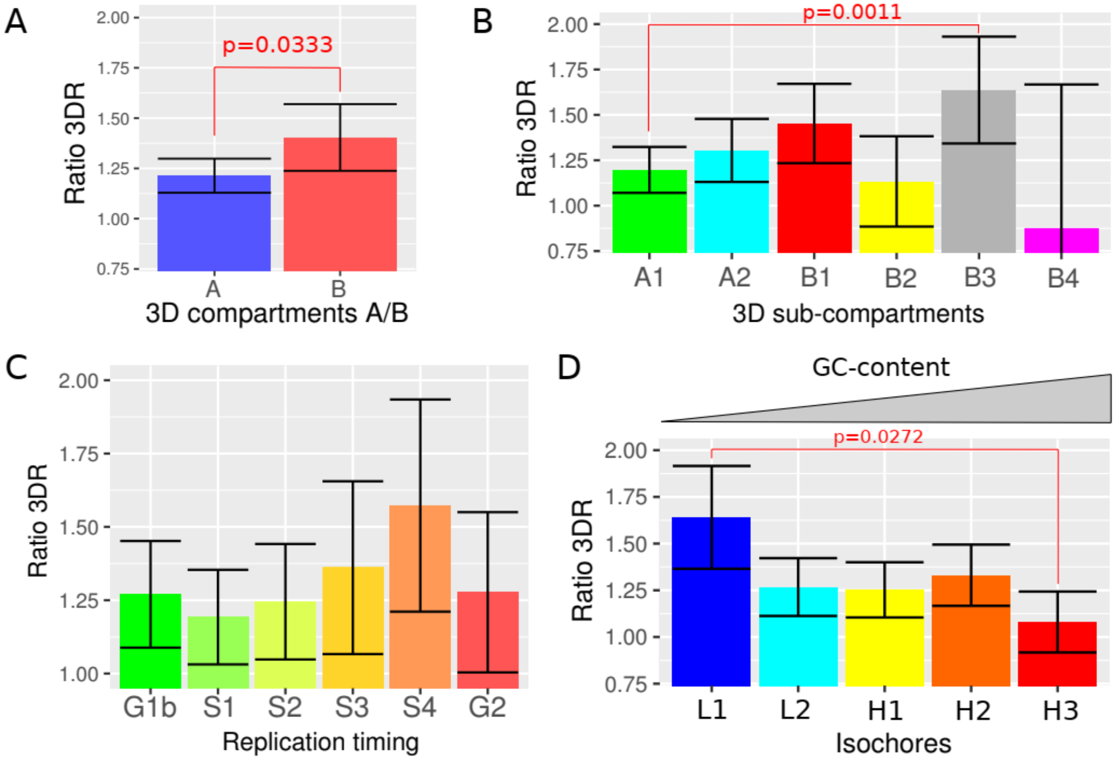






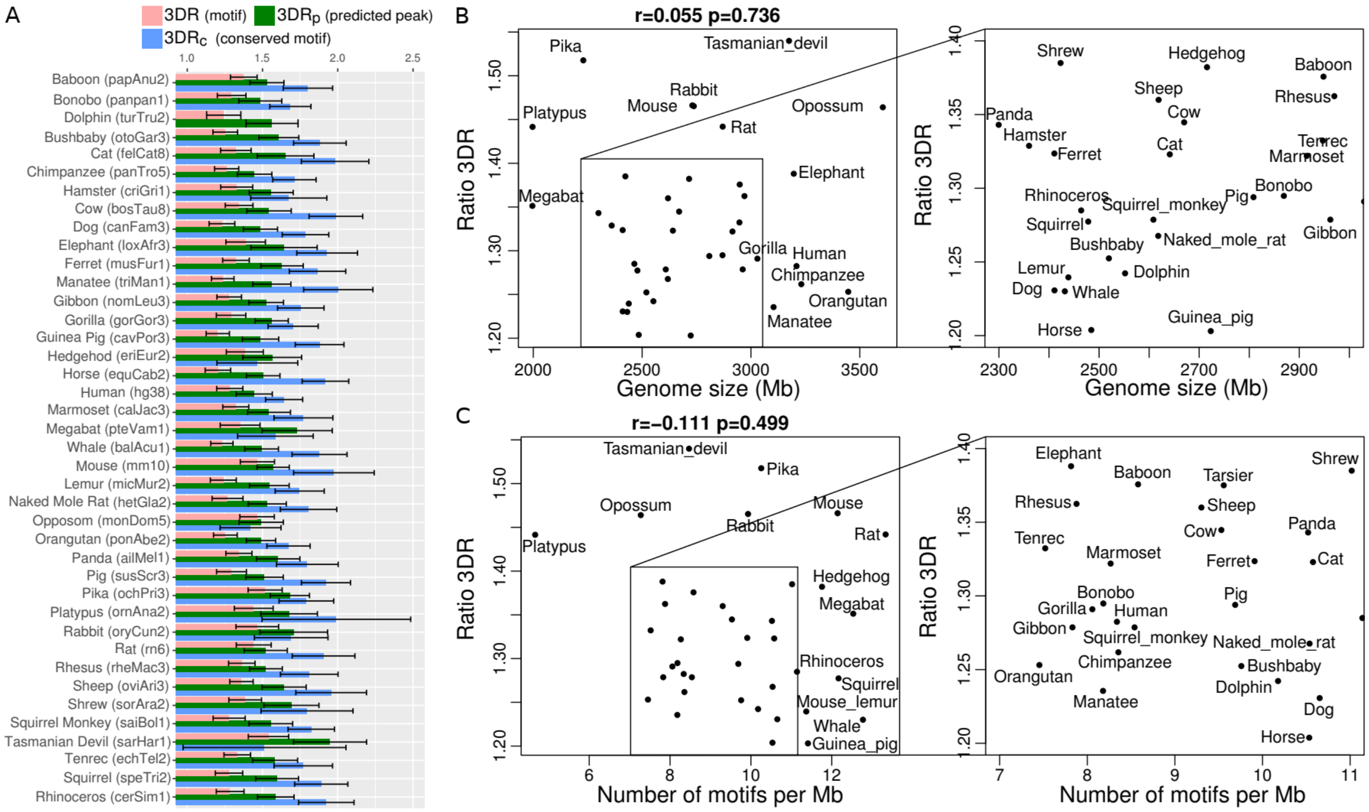
- Control experiment.
- In drosophila (dm6 and droYak2) and C. elegans (ce11) genomes, CTCF looping is absent.
- In agreement, 3DR = 1.

3DR varies with 3D compartments and isochores



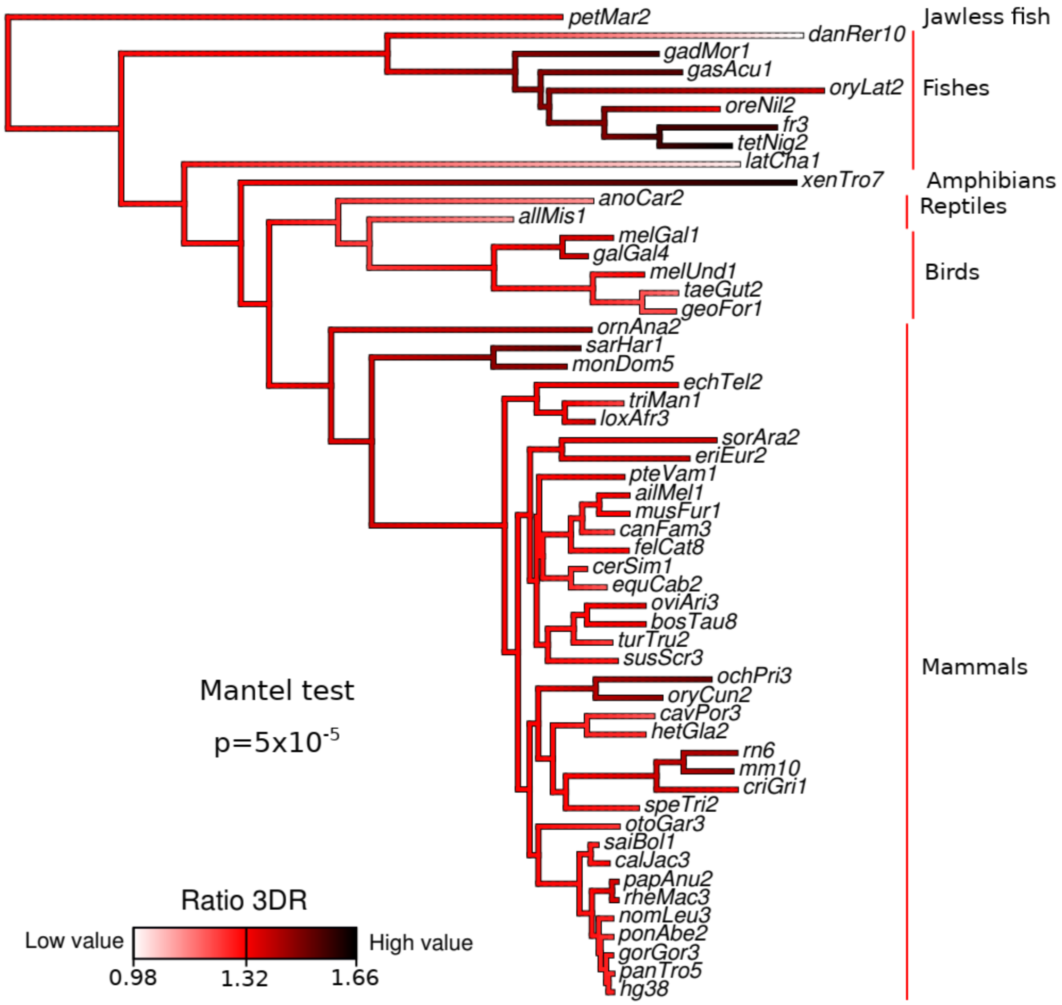
- 3DR is higher in B (sub)compartment, late replication timing and low GC isochores.

3DR detects CTCF-mediated looping in all mammals



- 3DR is independent of genome size and CTCF motif density.

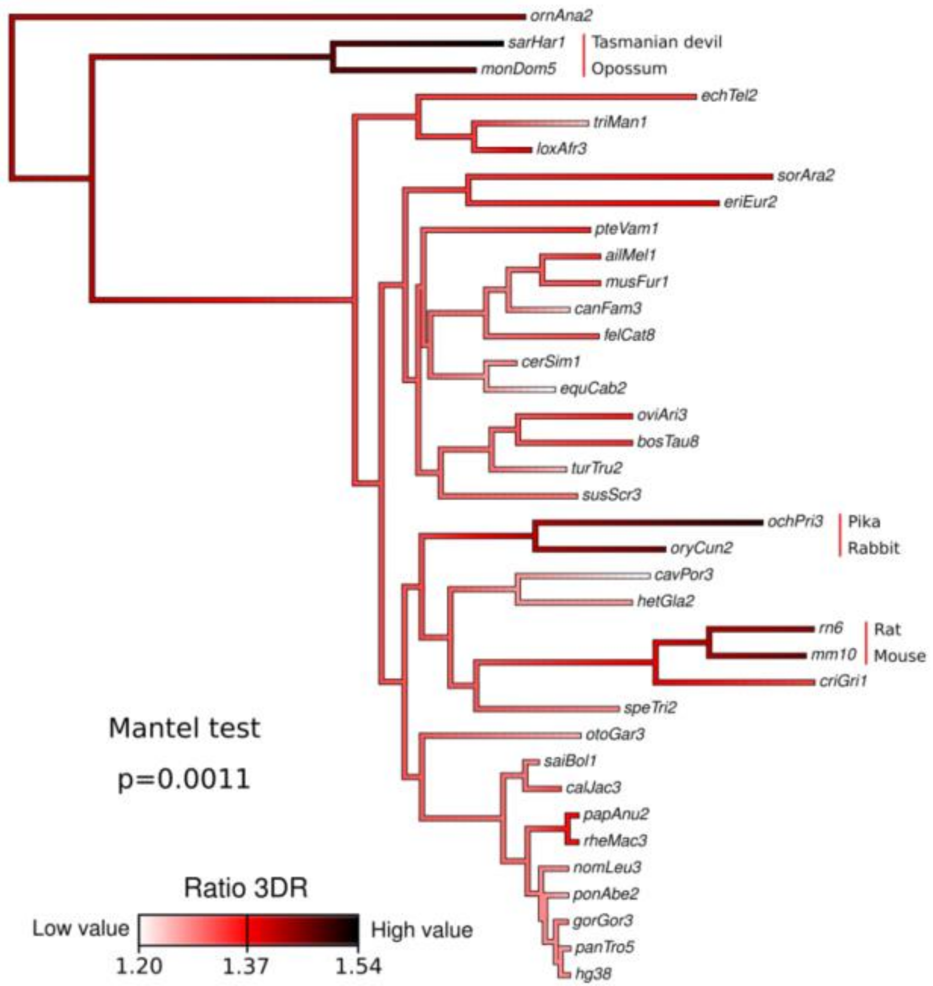
Phylogenetic analysis of 3DR



- Species that are phylogenetically close tend to have a ratio that is closer than species that are phylogenetically far (Mantel test $p=5\times10^{-5}$).

- 3DR is thus conserved!

Phylogenetic analysis of 3DR in mammals



- Among mammals, rat and mouse, pika and rabbit, and Tasmanian devil and opossum show high 3DR value.

For more details, read the article!

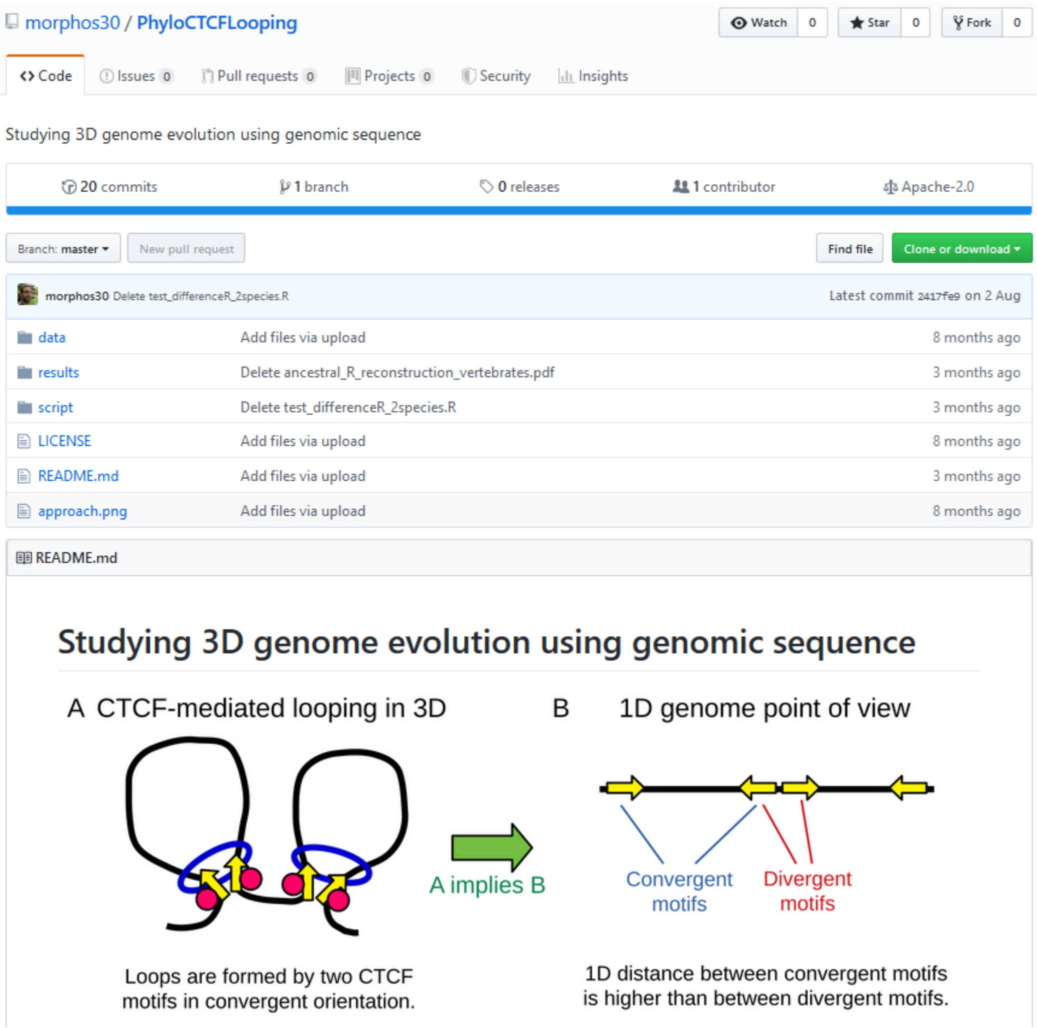
Studying 3D genome evolution using genomic sequence

Raphaël Mourad ✉

Bioinformatics, btz775, https://doi.org/10.1093/bioinformatics/btz775

Published: 12 October 2019 **Article history** ▾

GitHub R code



Studying 3D genome evolution using genomic sequence

morphos30 Delete test_differenceR_2species.R — Latest commit 2417fe9 on 2 Aug

data	Add files via upload	8 months ago
results	Delete ancestral_R_reconstruction_vertebrates.pdf	3 months ago
script	Delete test_differenceR_2species.R	3 months ago
LICENSE	Add files via upload	8 months ago
README.md	Add files via upload	3 months ago
approach.png	Add files via upload	8 months ago

Studying 3D genome evolution using genomic sequence

Loops are formed by two CTCF motifs in convergent orientation.

1D distance between convergent motifs is higher than between divergent motifs.

END

Bibliography

- Erez Lieberman-Aiden, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 326(5950):289-293, October 2009.
- Tom Sexton, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell, 148(3):458-472, February 2012.
- Chunhui Hou, et al. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Molecular Cell, 48:471-484, November 2012.
- Jesse R. Dixon, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature, 485(7398):376-380, May 2012.
- Jennifer E. Phillips-Cremins and Victor G. Corces. Chromatin insulators: Linking genome organization to cellular function. Molecular Cell, 50(4):461-474, May 2013.
- Jun Liang, et al. Chromatin immunoprecipitation indirect peaks highlight functional long-range interactions among insulator proteins and RNAII pausing. Molecular Cell, 53(4):672-681, February 2014.
- Kevin Van Bortle, et al. Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biology, 15(5):R82+, June 2014.
- Li Li, et al. Widespread rearrangement of 3D chromatin organization underlies Polycomb-mediated stress-induced silencing. Molecular Cell, (15):S1097-2765, March 2015.
- Suhas S. P. Rao, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell, 159(7):1665-1680, February 2015.
- D. G. Lupianez et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell, 161(5):1012-1025, May 2015.