Principles of the 3D genome organization in malaria mosquitoes

Varvara Lukyanchikova[1,2], Miroslav Nuriddinov[2], Nariman Battulin[2], Alexander Tyapkin, Polina Belokopytova[2], Veniamin Fishman[2], Igor V. Sharakhov[1]

[1]Virginia Polytechnic Institute and State University, Department of Entomology, Blacksburg, VA
[2]Institute of Cytology and Genetics SB RAS, Department of Molecular Mechanisms of Development, Novosibirsk, Russia

INTRODUCTION

The spatial organization of the genome plays an important role in cell function. The main principles of the 3D chromosome folding in eukaryotes have been discovered using Hi-C – a groundbreaking technology that exploits *in vivo* chromatin proximity information. This method can also yield dramatically improved genome assemblies. We applied the Hi-C approach to improve the genome assemblies for *Anopheles* species and to understand the principles of spatial genome organization in malaria vectors. We obtained new chromosome-level genome assemblies for *Anopheles albimanus*, *An. atroparvus*, *An. coluzzii*, and *An. stephensi*. Hi-C identified topologically associating domains (TADs), A/B compartments, and high-frequency long-distance contacts in the *Anopheles* genomes. Functional genomics data can be now put in the chromatin organization context. High-frequency long-distance chromatin contacts occur between genes and intergenic (possibly regulatory) regions. Some of these contacts are conserved across anophelines. Heterochromatin lacks typical TADs and has random interactions across the entire region. Hi-C is a robust tool for visualization and discovery of chromosomal inversions. Our results provide new facts for understanding of how architectural genome folding carries into effect within the nuclear space in malaria vectors.

AIMS

The main aim of our research was to investigate the principles of the spatial genome architecture in medically important *Anopheles* species achieved by optimization of genome-wide chromosome conformation capture approach (Hi-C) applied to mosquito embryos. Additionally, we were interested in comparing the 3D genome organization across species of the *Anopheles* genus.

EXPERIMENTAL DESIGN

The species chosen for our project represent broad range of units on evolutionary tree and have diverged through the evolution 0.5-100 mya. In terms of evolution, the most distant species, *An coluzzii* and *An. albimanus*, have been separated from each other on the same evolutionary distance as mouse and human (~100 mya), while the closest pair, *An. coluzzii* and *An. stephensi* are separated by the 30 mya distance (**Fig. 1**). Therefore, we are interested to see if any significant alterations in TADs, A/B compartments or 3D-genome structure in general takes place in the evolution.

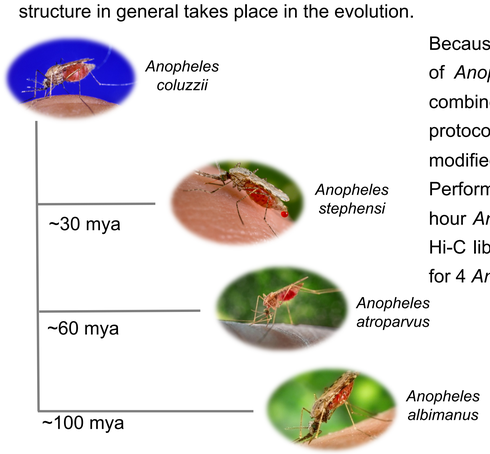


Fig. 1. Phylogeny of species in the genus *Anopheles*.

Because of specific characteristics of *Anopheles* mosquito embryos, we have combined the standard *in situ* Hi-C protocol[1] with the *Drosophila* Hi-C protocol[2], modified, and optimized it for *Anopheles*. Performing the new protocol on 15-18-hour *Anopheles* eggs, we have generated 8 Hi-C libraries, including 2 biological replicas for 4 *Anopheles* species.

The Hi-C libraries were labeled with adaptors, prepared for Illumina sequencing, and sequenced with HiSeq (~60 mln reads per species) using Illumina 150-bp paired-end sequencing and analyzed with Hi-C Pro software[3].

PRINCIPLES OF 3D ORGANIZATION IN MOSQUITOES REVEALED BY Hi-C

Using the Hi-C Pro pipeline we have identified species-specific patterns of short- and long-range chromatin interactions and inter-chromosomal contacts for all studied species. Moreover, we have detected and fixed scaffold orientation errors and scaffold misassemblies within the existing genome references.

In accordance to Rabl-like conformation model we revealed centromere-centromere interactions, inter- and intra-chromosomal telomere-telomere interactions, and stable long-distance interactions between chromosome arms (**Fig. 2**). Centromere-centromere interactions were confirmed by FISH in *An. stephensi* (**Fig. 3**).

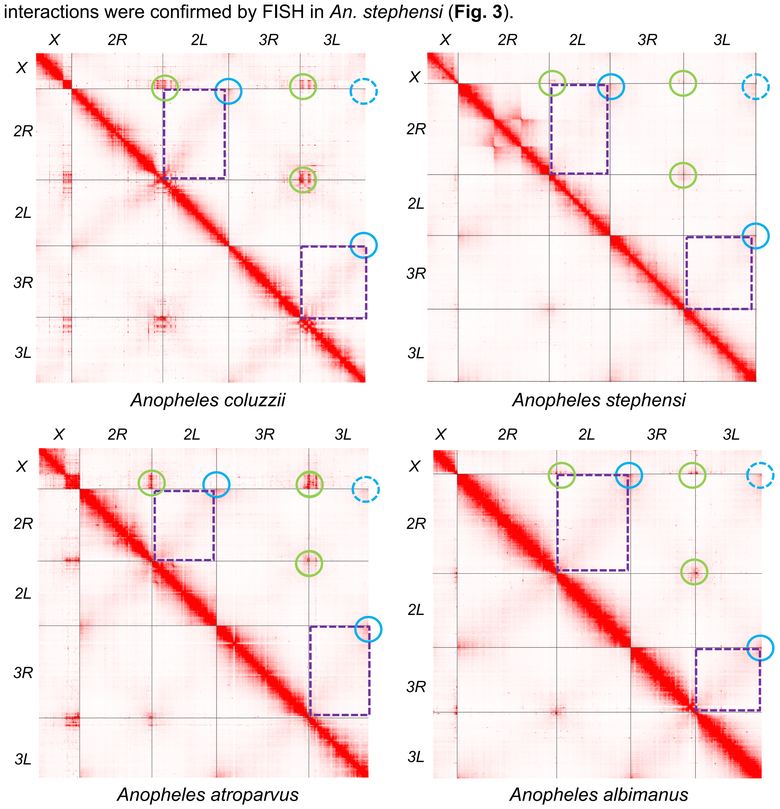
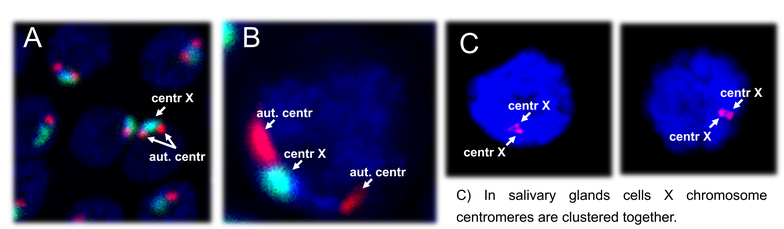


Fig. 2. Visualization of Hi-C heat maps for *Anopheles* species using Juicebox. Centromere-centromere interactions are shown by green circles, inter- and intra-chromosomal telomere-telomere interactions are shown by blue circles. Stable long-distance interactions between chromosome arms are seen as "wings" (purple rectangle).



A) In follicular cells all centromeres cluster together.

B) In nurse cells centromeres are close but not exactly clustered.

C) In salivary glands cells X chromosome centromeres are clustered together.

Fig. 3. Centromere-centromere interactions in *An. stephensi*.

A/B COMPARTMENTS, TADs, LONG-DISTANCE INTERACTIONS

Our data demonstrate that *Anopheline* genomes are fully partitioned into A and B compartments **(Fig. 4)** similar to other insects[4].

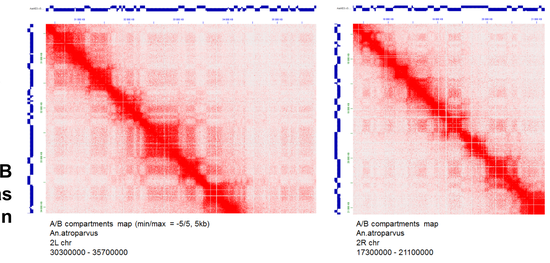


Fig. 4. A and B compartments identified as a "chess board" pattern in *An. atroparvus*.

In all studied *Anopheles* species we observed specific long-distance interactions within the X and 2R chromosomes, which are formed by conserved orthologous genes and intergenic regions (**Fig. 5**). These interactions have been maintained during ~100 million years of *Anopheles* evolution suggesting their importance for gene regulation. While TADs can be identified in euchromatin, heterochromatin regions lack typical TADs (**Fig. 5CD**).

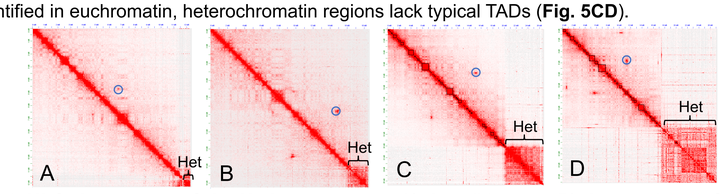


Fig. 5. Hi-C heat maps of the X chromosomes. A) *An. albimanus*. B) *An. stephensi*. C) *An. coluzzii*. D) *An. atroparvus*. Blue circles show 5-6 Mb long interactions. Black squares indicate TADs. Het – pericentromeric heterochromatin.

Using the Hi-C data, we have identified the known 16 Mb-inversion 2Rb in *An. stephensi* and discovered a novel polymorphic 11 Mb-inversion 2La in *An. atroparvus* (**Fig. 6**).

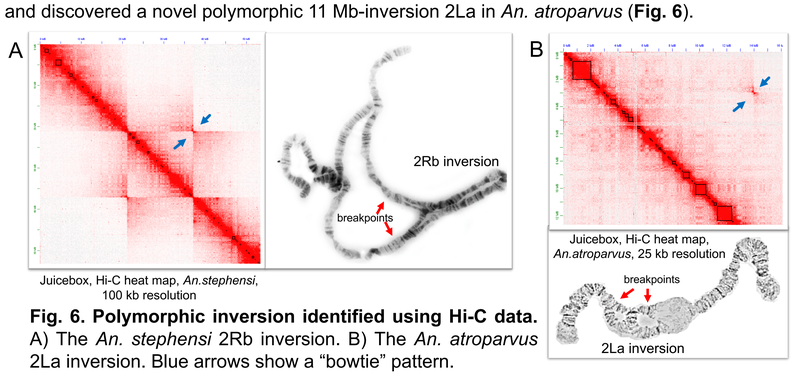


Fig. 6. Polymorphic inversion identified using Hi-C data. A) The *An. stephensi* 2Rb inversion. B) The *An. atroparvus* 2La inversion. Blue arrows show a "bowtie" pattern.

CONCLUSIONS

Our study provides insights into general principles of 3D genome organization in *Anopheles* mosquitoes revealed by the Hi-C approach.

1. Using the Hi-C data, we have obtained improved chromosome-level genome assemblies for *An. albimanus*, *An. atroparvus*, *An. coluzzii*, and *An. stephensi*.
2. Rabl-like conformation exists at the interphase stage in different cell types of *Anopheles*.
3. The mosquito genomes are partitioned into A/B compartments, TADs and loops.
4. Conserved long-distance chromatin interactions are observed across species of the *Anopheles* genus.
5. Heterochromatin lacks typical TADs and has random interactions across the entire region.
6. Hi-C is a robust tool for visualization and discovery of chromosomal inversions.

REFERENCES:

1) Rao, Suhas SP, et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." Cell 159.7 (2014): 1665-1680;
2) Comet, Itys, et al. "A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber." Proceedings of the National Academy of Sciences 108.6 (2011): 2294-2299
3) Servant, Nicolas, et al. "HiC-Pro: an optimized and flexible pipeline for Hi-C data processing." Genome biology 16.1 (2015): 259;
4) Sexton, Tom, et al. "Three-dimensional folding and functional organization principles of the Drosophila genome." Cell 148.3 (2012): 458-472;

FUNDING:

This work was supported by the NSF grant MCB-1715207 and NIH grant R21AI135298 to IVS.