

CTCF and the evolution of genome spatial organization

Nurislamov A.^{1, 2*}, Popov A.^{1, 2}, Gridina M.², Shadskiy A.^{1, 2}, Fishman V.^{1, 2}

¹ Novosibirsk State University, Novosibirsk, Russia

² Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

* a.nurislamov@g.nsu.ru

Key words: chromatin; 3D genomics; epigenetics; evolution

Motivation and Aim: The CCCTC-binding factor (CTCF) is a crucial insulator protein in bilaterian animals. While the structure of CTCF is conserved, its functions can vary across different taxa. In vertebrates, CTCF is involved in forming chromatin loops via a cohesin/CTCF-dependent extrusion mechanism, which is a primary factor in the formation of topologically associated domains (TADs) in these animals. This mechanism has not been observed in other bilaterian groups despite the conserved structure of the protein. Therefore, it is of great interest to identify factors that determine the presence or absence of cohesin/CTCF-dependent loop extrusion in bilaterian animals.

Methods and Algorithms: We conducted an evolutionary analysis of CTCF amino acid sequences, focusing particularly on the N-terminal domain. This domain exhibits the most variable structure among bilaterian animals and is critically important for chromatin loop formation in vertebrates. To compare the emergence time in evolution of crucial amino acid motifs for cohesin interaction, we employed multiple protein alignment techniques. By aligning CTCF sequences from a wide range of bilaterian taxa, we were able to identify conserved and divergent motifs and determine their evolutionary timelines.

Results: Our analysis identified N-terminal domain sequences specific to jawed vertebrates, which are absent in jawless vertebrates and cephalochordates. We discovered that the KTYQR motif emerged specifically in jawed vertebrates, indicating a potential evolutionary adaptation linked to their chromatin organization mechanisms. Additionally, we found that the YXF motif is present in CTCF across all studied bilaterian groups, except for *Trichinella spiralis*, where CTCF likely loses its function. In vertebrates, the YXF motif in CTCF is essential for chromatin loop formation as it facilitates the interaction between CTCF and the loop extrusion complex. This widespread presence of the YXF motif across bilaterian taxa underscores its fundamental role in CTCF functionality, whereas the emergence of the KTYQR motif in jawed vertebrates suggests additional evolutionary adaptations enhancing CTCF's role in chromatin organization in these species.

Conclusion: Our findings suggest that the ability of vertebrate CTCF to participate in chromatin loop formation is influenced not only by the presence of the YXF motif but also by other factors. This highlights the complexity of the mechanisms underlying chromatin organization and the evolutionary divergence of CTCF functions in bilaterian animals.

Funding: The study is supported by RSF grant No. 22-14-00247.