DOI 10.18699/SBB-2022-340

CENH3 gene duplication leads to heterogeneous organization of centromeric chromatin in rye

Evtushenko E.[1]*, Elisafenko E.[1, 2], Gatskaya S.[1], Vershinin A.[1]

[1] *Institute of Molecular and Cellular Biology, SB RAS, Novosibirsk, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

* *evt@mcb.nsc.ru*

Key words: CENH3 variants, gene duplication, *Secale cereale*, functional DNA motifs, *CENH3* expresssion, subdomain organization of nucleosomes, CENH3 loading

Motivation and Aim: Gene duplication and the preservation of both copies during evolution is an intriguing evolutionary phenomenon. Although an important role is attributed to gene duplications in generating evolutionary novelty and adaptation [1], this role is often not so evident. Their preservation depends on the importance of the function they perform. The central component of centromere specification and function is the centromere-specific histone H3 (CENH3). Some cereal species (maize, rice) have one copy of the gene encoding this protein, while some (wheat, barley, rye) have two, and so they represent a good model for a comparative study of the functional activity of the duplicated *CENH3* genes and their protein products. We investigated the organization of the *CENH3* locus in the rye (*Secale cereale* L.) genome to identify functionally important sites in the vicinity of the *αCENH3* and *βCENH3* genes, the expression levels of these genes at different stages of plant development, the subsequent loading of their products, the αCENH3 and βCENH3 proteins, into nucleosomes and patterns of their organization in linear centromeric chromatin using elongated DNA fibers.

Methods and Algorithms: The rye genome assembly was downloaded from GenBank (https://www.ncbi.nlm.nih.gov/assembly/GCA_902687465.1). The search for *CENH3* sequences in the genome was performed using the *TBLASTN* program from the AB-BLAST package [2]. Multiple alignment of regulatory regions in the vicinity of the *αCENH3* and *βCENH3* genes was performed using *MUSCLE*. Identification of functional sites was performed using TSSPlant and NSITE-PL [3, 4] on http://www.softberry.com. Plants were grown in a greenhouse. The relative expression levels of both functional CENH3 variants were examined in different rye tissue types (meiotic anther, carpel, leaf, root, stem) using quantitative real-time PCR. To find out whether proteins are synthesized from the *αCENH3* and *βCENH3* variants, whether these proteins are loaded into nucleosomes of extended chromatin fibers from young etiolated leaf nuclei and how they are organized within the centromeric chromatin, we employed CENH3 variant-specific antibodies in combination with super-resolution spatial structured illumination microscopy (3D-SIM) and confocal microscopy.

Results: The alignment of the primary structure of -300 and +100 regions relative to the TSS (transcription start site), where regulatory elements responsible for specific transcription regulation normally reside, showed that they are highly conserved both the *αCENH3* and *βCENH3* genes of different cereal species. Thus, the amount of pressure exerted by negative selection on these regions during the evolution of the *CENH3* locus was rather low. This statement is supported by a comparison of the sets of functional

sites between *αCENH3* and *βCENH3*. A common feature of both genes is the complex architecture of their promoters; along with TATA boxes identifiable by TSSPlant. These sets contain both common regulatory motifs, TATA boxes and the DPE and INR motives characteristic of TATA-less promoters, and those specific for each gene. These extensive sets of functional sites make the regulatory regions of the *αCENH3* and *βCENH3* genes universal. A comparison of the expression levels of both genes between different developmental stages of two genotypically different rye varieties, 'Imperial' and 'K69', revealed a clear-cut correlation between expression levels and cell division intensity. Assuming that separate clusters of signals on DNA fibers correspond to particular nucleosomal subdomains, we identified main patterns of the co-organization of the CENH3 variants, which correspond to the main polynucleosomal subdomains in centromeric chromatin: 1) clusters consisting only of αCENH3, the maximum size of which is about 40 kb (Fig. 1a), or βCENH3 signals, the maximum size of which is about 12 kb (Fig. 1b); 2) clusters consisting of alternating signals of both proteins, their size is from about kb to 120 kb (Fig. 1c, d). Clusters with CENH3 signals are interrupted by gaps with nucleosomes containing the canonical histone H3 (Fig. 1, b, c). We propose that these main patterns of polynucleosomal CENH3 subdomains form part of centromeres.

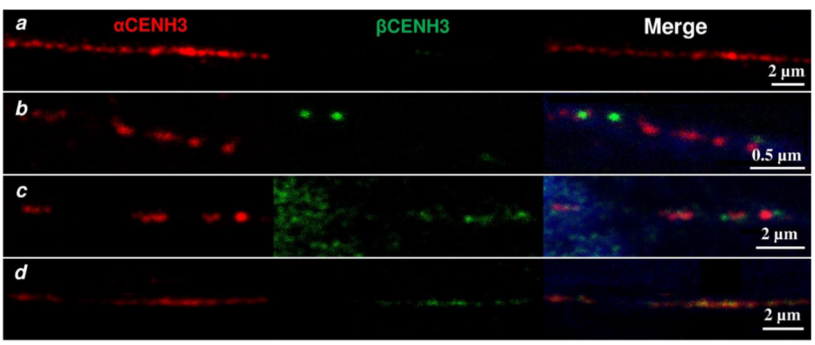

Fig. 1. Different types of nucleosome clusters visible on extended chromatin fibers

Conclusion: We determined the organization of the *CENH3* locus in rye and identified the functional motifs in the vicinity of the *CENH3* genes. Both CENH3 variants are transcribed and the corresponding proteins are centromere-incorporated in a tissue specific manner. Using extended chromatin fibers, we revealed main patterns of loading CENH3 proteins into polynucleosomal domains in centromeric chromatin where alternating nucleosomal clusters contain both CENH3 variants and nucleosomes containing the canonical histone H3.

Acknowledgements: The study is supported by grant of Russian Foundation for Basic Research (project No. 20–04-00699).

References

1. Flagel L.E., Wendel J.F. Gene duplication and evolutionary novelty in plants. *New Phytol*. 2009;183:357-364.
2. Gish W. 1996-2019. https://blast.advbiocomp.com.
3. Solovyev V.V., Shahmuradov I.A., Salamov A.A. Identification of Promoter Regions and Regulatory Sites. In: Ladunga I. (Ed.). Computational Biology of Transcription Factor Binding. Methods in Molecular Biology, vol 674. Humana Press, Totowa, NJ. 2010.
4. Shahmuradov I.A., Solovyev V.V. Nsite, NsiteH and NsiteM computer tools for studying transcription regulatory elements. *Bioinformatics.* 2015;31:3544-3545.