




Article

Expression of Two Rye CENH3 Variants and Their Loading into Centromeres

Elena V. Evtushenko ¹, Evgeny A. Elisafenko ^{1,2}, Sima S. Gatzkaya ¹, Veit Schubert ³ , Andreas Houben ³ 
and Alexander V. Vershinin ^{1,*} 

¹ Institute of Molecular and Cellular Biology, SB RAS, Acad. Lavrentiev Ave. 8/2, 630090 Novosibirsk, Russia; evt@mcb.nsc.ru (E.V.E.); kanopus@ngs.ru (E.A.E.); jait@mail.ru (S.S.G.)

² Institute of Cytology and Genetics, SB RAS, Acad. Lavrentiev Ave. 10, 630090 Novosibirsk, Russia

³ Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstr. 3, 06466 Seeland, Germany; schubertv@ipk-gatersleben.de (V.S.); houben@ipk-gatersleben.de (A.H.)

* Correspondence: avershin@mcb.nsc.ru

Abstract: Gene duplication and the preservation of both copies during evolution is an intriguing evolutionary phenomenon. Their preservation is related to 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. Therefore, they represent a good model for a comparative study of the functional activity of the duplicated *CENH3* genes and their protein products. We determined the organization of the *CENH3* locus in rye (*Secale cereale* L.) and identified the functional motifs in the vicinity of the *CENH3* genes. We compared the expression of these genes at different stages of plant development and the loading of their products, the CENH3 proteins, into nucleosomes during mitosis and meiosis. Using extended chromatin fibers, we revealed patterns of loading CENH3 proteins into polynucleosomal domains in centromeric chromatin. Our results indicate no sign of neofunctionalization, subfunctionalization or specialization in the gene copies. The influence of negative selection on the coding part of the genes led them to preserve their conserved function. The advantage of having two functional genes appears as the gene-dosage effect.

Keywords: CENH3 variants; gene duplication; *Secale cereale*; functional DNA motifs; *CENH3* expression; subdomain organization of nucleosomes; CENH3 loading



Citation: Evtushenko, E.V.; Elisafenko, E.A.; Gatzkaya, S.S.; Schubert, V.; Houben, A.; Vershinin, A.V. Expression of Two Rye CENH3 Variants and Their Loading into Centromeres. *Plants* **2021**, *10*, 2043. <https://doi.org/10.3390/plants10102043>

Academic Editor: Tomás Naranjo

Received: 31 August 2021

Accepted: 24 September 2021

Published: 28 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gene duplications, which occur as a result of polyploidization following whole-genome duplications (WGDs) or as a result of local duplications of genes and duplication of small genomic regions (small-scale duplications (SSDs)), are thought to have an essential role in generating evolutionary novelty and adaptation [1]. They are common among angiosperms [2], but the mechanisms and evolutionary processes that ultimately determine the fate of duplicated genes are far from being entirely understood. In the simplest case, an SSD gives birth to two copies of a previously single-copy gene. Previous studies identified several possible molecular mechanisms that contribute to the preservation of the second, “daughter” copy in the genome during evolution [3]: (i) *conservation*, in which the function of the ancestral copy is preserved in the daughter due to negative selection aimed at maintaining an increased dose of the gene; (ii) *neofunctionalization*, in which one of the copies acquires a new function as a result of positive selection, while the other retains its ancestral function; (iii) *subfunctionalization*, in which the multiple functions of the parental gene are distributed between two copies [4], is observed either under positive selection, which optimizes the subfunctions of each copy [5], or under neutral selection in the case of unfavorable mutations that weaken subfunctions of each copy [4]; and (iv) *specialization*, in which neofunctionalization is followed by a rapid

subfunctionalization, which leads both copies to acquire new functions that are other than the function of the ancestral gene [6]. Subfunctionalization may be accompanied by changes in the subcellular localization of the corresponding protein products [7]. For example, as a result of alternative splicing, the proteins produced by each copy of a duplicated gene can be localized only in one of the cellular structures that they originally occupied (this phenomenon is called *sublocalization*) [7].

The first studies seeking to identify centromeric variants of H3 histone (CENH3) in cereals revealed one form of this protein in maize (*Zea mays*) [8] and rice (*Oryza sativa*) [9]. These findings led to the conclusion that the genomes of these species each contain one copy of the corresponding *CENH3* gene. Later studies found that some cereal species in the tribe Triticeae, namely barley, wheat and rye, have two different variants of CENH3 [10–12]. The highest heterogeneity was observed in the N-terminal tail (NTT) of CENH3s, which is rapidly evolving to adapt to lineage-specific centromeric constraints [13]. The C-terminus of the CENH3 protein contains a histone fold domain (HFD), which is more conserved and structurally similar to the canonical histone H3. This domain is required for nucleosome assembly and CENH3 targeting to centromeres [14].

The increase in the number of sequenced cereal genomes and advances in their assembly in recent years made it possible to determine the localization of the *CENH3* genes and to trace the evolution of their structure and their surrounding regions in the genomes of various tribes. It was found that around 50MYA, the common ancestor of the subfamilies Bambusoideae, Oryzoideae and Pooideae developed the *CENH3* locus composed of the syntenic genes *CDPK2-1*, *CENH3* and *bZIP*. A duplication event occurred within the *CENH3* locus in Stipeae and Brachypodieae species around 35–40 MYA and produced two genes, α *CENH3* and β *CENH3*. The distance between these genes (the intergenic spacer) is 2.16 kb in *Brachypodium distachyon* and 5.9 kb in *Stipa sibirica* and is filled with relatively short fragments of retrotransposons and tracks of simple repeats [15]. Later in the evolution of the *CENH3* locus in Triticeae, the *CDPK2-1* gene located to the left of β *CENH3* was replaced with the *LHCB-1* gene, which encodes the chlorophyll-binding protein 3C. An increase in the distance between the β *CENH3* and α *CENH3* genes occurred mainly due to massive insertion of the elements of the LTR-containing retrotransposon superfamilies *gypsy* and *copia*. At the same time, the changes in the composition and contribution of individual *gypsy* and *copia* families to the molecular structure of the *CENH3* locus display a significant diversity even in evolutionarily close species [15].

The functionality of the α *CENH3* and β *CENH3* genes in Triticeae was studied in the cultivated barley (*Hordeum vulgare*) [16] and diploid and tetraploid wheat species (*Triticum* spp.) [11]. It was demonstrated by immunostaining that in young barley embryos, both variants of the CENH3 protein are loaded into most interphase centromeres. In contrast, in the cells of the root meristem, β *CENH3* is loaded in the centromeres less efficiently than α *CENH3*, and most of the β *CENH3* molecules are dispersed in the nucleoplasm. However, in differentiated tissues, β *CENH3* transcription levels were higher than those of α *CENH3* [16]. Differences in α *CENH3* and β *CENH3* expression were also observed between cultivated and wild wheats [11]. In wild tetraploids, the expression levels of β *CENH3* were significantly lower than those of α *CENH3*; however, in cultivated tetraploids, β *CENH3* transcription levels would increase to nearly reach α *CENH3* levels.

Similarly to other Triticeae species, two forms of the protein, α *CENH3* and β *CENH3*, were revealed in rye (*Secale cereale* L.) [12]. However, the structure of the entire *CENH3* locus, the expression levels of the genes encoding the CENH3 proteins and the efficiency of these proteins loading into chromosomes remain unknown. The sequencing of two rye accessions [17,18] makes it possible to thoroughly explore the coding and regulatory parts of the α *CENH3* and β *CENH3* genes. In addition, it becomes possible to address how their features correlate with the expression levels of these genes and the localization of the proteins in chromosomes and nuclei at different stages of plant development. In this work, we investigated the organization of the *CENH3* locus in the *S. cereale* genome to identify functionally important sites in the vicinity of the α *CENH3* and β *CENH3* genes. The

expression levels of these genes at different stages of plant development; the subsequent loading of the products of the expression of these genes, the α CENH3 and β CENH3 proteins, into nucleosomes; and the patterns of their organization in linear centromeric chromatin using elongated DNA fibers were analyzed.

2. Results

2.1. Molecular Organization of the CENH3 Locus in Rye

The availability of the sequenced rye genome (https://www.ncbi.nlm.nih.gov/assembly/GCA_902687465.1, accessed on 9 December 2020) made it possible to describe the structure of the CENH3 locus and to compare the structure of this locus in closely related species, barley and wheat. The size of the CENH3 locus in rye line Lo7 [17] from the left-border (upstream) gene, *LHCB3-1* (*LHCB3-like*), which contains the domain of chlorophyll a-b binding protein 3C, to the right-border (downstream) gene, *bZIP*, which is a transcription factor gene, is 218.3 kb (Figure 1).

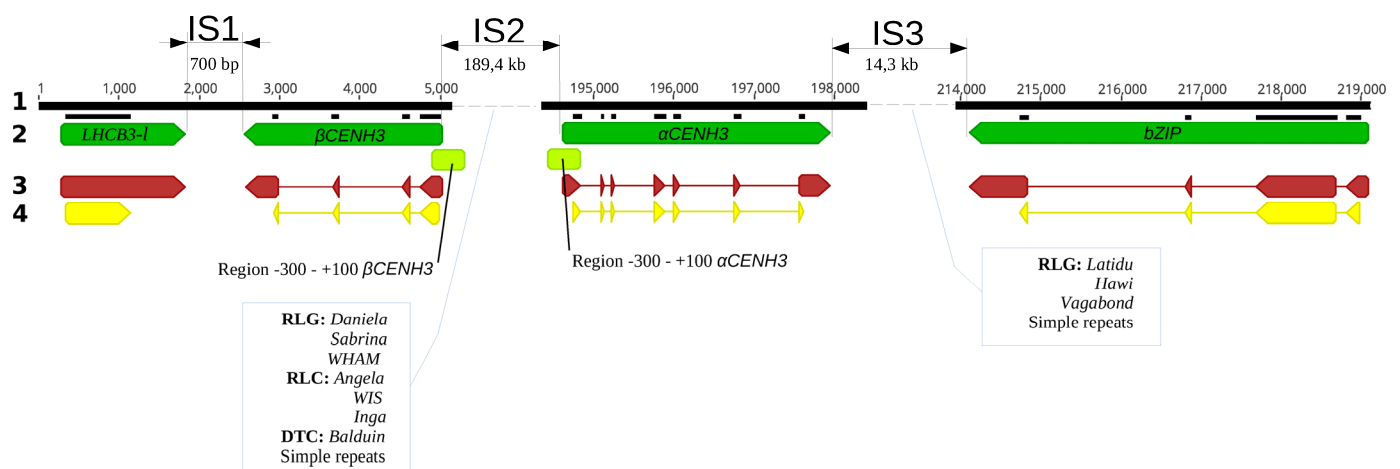


Figure 1. The structure of the CENH3 locus in rye (*S. cereale*). **1:** size in kb; **2:** syntenic genes appear as dark-green fingerpost arrows; **3:** exon–intron structure of genes at mRNA level, where exons appear as fingerpost arrows; **4:** coding regions of genes appear as yellow fingerpost arrows. Light-green rectangles depict -300 and $+100$ regions around the TSSs of α CENH3 and β CENH3, where functional motifs were searched for.

The intron–exon structure of the paralogs was determined by analysis of genomic and transcriptome libraries. The α CENH3 gene has seven exons encoding a protein 166 amino acids in length and is separated by six introns. The β CENH3 gene has four exons and produces a protein 151 amino acids in length, which is shorter due to an extensive deletion in the NTT domain of the parental gene. Thus, the structure of the CENH3 locus and the intron–exon structure of the CENH3 paralogs in rye are similar to those in other Triticeae species [15] but differ from those in CENH3 paralogs in legumes [19], *Mimulus* [20] and *Arabidopsis* [21].

At the CENH3 locus, the left-border and right-border genes are separated from the β CENH3 and α CENH3 genes by intervening spacers. A 189.4 kb intergenic spacer separates β CENH3 and α CENH3 (IS2, Figure 1). Most of IS2 is made up of class I transposable elements (TEs) within two retrotransposon superfamilies: *gypsy*-like and *cop*ia-like, abbreviated RLG and RLC, respectively, in the notation from unified classification system [22]. Together, their DNA sequences account for 98.2% of all sequences in IS2 identified using the *Viridiplantae RepeatMasker* program. The RLG-to-RLC ratio is 2.92, which is only slightly different from the mean value for the rye genome, 3.26. However, the prevalence of separate RLG and RLC families in IS2 and in the entire rye genome differs more noticeably. For example, the most prevalent families in IS2 (*Daniela*, RLG and *WIS*, RLC) are not the most prevalent in rye (*Sabrina*, RLG and *Angela*, RLC) (Table 1). Class I TEs (DTX, transposons) are represented by *Balduin*, a family in the CACTA superfamily, which accounts for only 1.3% of IS2, although the most prevalent DTX family in rye is *Jorge*.

Table 1. Contribution of the most abundant class I and class II families to the IS2 and genome of rye.

Family	<i>Gypsy-Like (RLG) *</i>					<i>Copia-Like (RLC) *</i>					
	% in IS2	% in Super-family (in IS2)	Family	% in Super-family (in Genome)	% in Genome	Family	% in IS2	% in Super-family (in IS2)	Family	% in Super-family (in Genome)	% in Genome
<i>Daniela</i>	22.5	32.3	<i>Sabrina</i>	18.2	8.8	<i>WIS</i>	10.4	44.8	<i>Angela</i>	37.38	5.37
<i>Sabrina</i>	16.5	24.5	<i>Daniela</i>	10.3	4.9	<i>Inga</i>	6.2	26.8	<i>WIS</i>	22.65	3.25
<i>WHAM</i>	9.3	13.8	<i>Erika</i>	6.2	3.0	<i>Angela</i>	5.1	22.1	<i>Barbara</i>	14.41	2.07
<i>Erika</i>	5.3	7.9	<i>Laura</i>	6.0	2.9				<i>Inga</i>	5.15	0.74
<i>Romani</i>	5.2	7.7	<i>Sabine</i>	5.8	2.8				<i>Eugene</i>	2.79	0.40

* In brackets: codes in accordance with the classification by Wicker et al. [22].

In the 14.3 kb long IS3, *RepeatMasker* identified 79% of DNA sequences, and, with the exception of a very small part composed of simple repeats, the rest of the DNA was found to be made up of three *gypsy*-like families: *Latidu* (56.8%), *Hawi* (36.3%) and *Vagabond* (6.2%). The representation of these families in the entire rye genome is negligible: the respective shares of the most common of them, *Hawi* and *Latidu*, are 1.796 and 1.696 in the *RLG* superfamily. The short IS1, 0.7 kb in length, is devoid of transposon elements, and only one 55 bp sequence appearing as a tandem dimer was identified in it. An exciting feature of the organization of intergenic spacers in the *CENH3* locus is that the coding genes *CENH3* and TEs never occur in adjacency to each other. The insertion sites of TEs begin and end several kilobases away from these genes. These peculiar buffer zones are largely filled with short tracks of micro- and minisatellites and AT-enriched regions. This gradient of different classes of TEs in the vicinity of coding genes is a common rule, at least for Triticeae, because it is characteristic of the entire barley genome [23].

2.2. Regulatory Regions of the *CENH3* Genes

The observed changes in the structure of the *CENH3* locus and the coding regions of the α *CENH3* and β *CENH3* genes, which took place during the shaping of the rye genome, raise the following questions: To what extent have duplications and related changes affected the structure of the regulatory regions that surround these genes in the course of evolution? How are these regions involved in the regulation of the expression of the *CENH3* genes? Promoters have a complex, gene-specific structure. They may include several short DNA motifs, which serve to bind transcription factors (TFs), and each promoter has a unique composition of these transcription factor binding sites [24]. The core promoter is defined as the shortest DNA motif capable of initiating basal transcription. Typically, it is located from -60 to $+40$ bp relative to the transcription start site (TSS). The main functions of core promoters are (1) to support the assembly of the preinitiating complex (PIC), which is composed of polymerase II and general transcription factors (GTFs), and (2) to initiate transcription from an exact position and at an exact level [25]. The region from -200 to -300 is normally assigned to the proximal promoter, which contains multiple regulatory elements responsible for the specific regulation of transcription [26].

We analyzed the primary structure of 300 bp regions (-300) upstream of the TSS and 100 bp ($+100$) downstream of the start of the transcription start sites of α *CENH3* and β *CENH3* (Figure 1). The presence of whole-genome libraries of close-to-rye Triticeae species makes it possible to compare these regions (Figures 2 and 3) in different species using the TSSPlant software program [27]. This program predicts the presence of TATA-containing and TATA-less promoters in the genomic sequences of a wide range of plant species. In all species, six regulatory motifs for the α *CENH3* gene were found in -300 bp regions, including the TATA-box sequence located at position 38 (in rye) from the TSS. The -300 bp regulatory region of the β *CENH3* gene reveals five regulatory elements, among which

TSSPlant identifies two pairs of promoters. One pair consists of a TATA box sequence, which is farther from the TSS here than in α CENH3 (−87 in rye) and presumably functions together with another core promoter, INR (initiator element), located 16 bp downstream of the TATA-box in rye. The TATA-box partially overlaps with the sequence of another downstream promoter element (DPE), which is normally located downstream of the TSS. Presumably, DPE functions in conjunction with another INR located above the TATA-box. Thus, the regulatory region of the β CENH3 gene has a complex, specific architecture of promoters, given that the promoter elements DPE and INR are characteristic of TATA-less promoters [27].

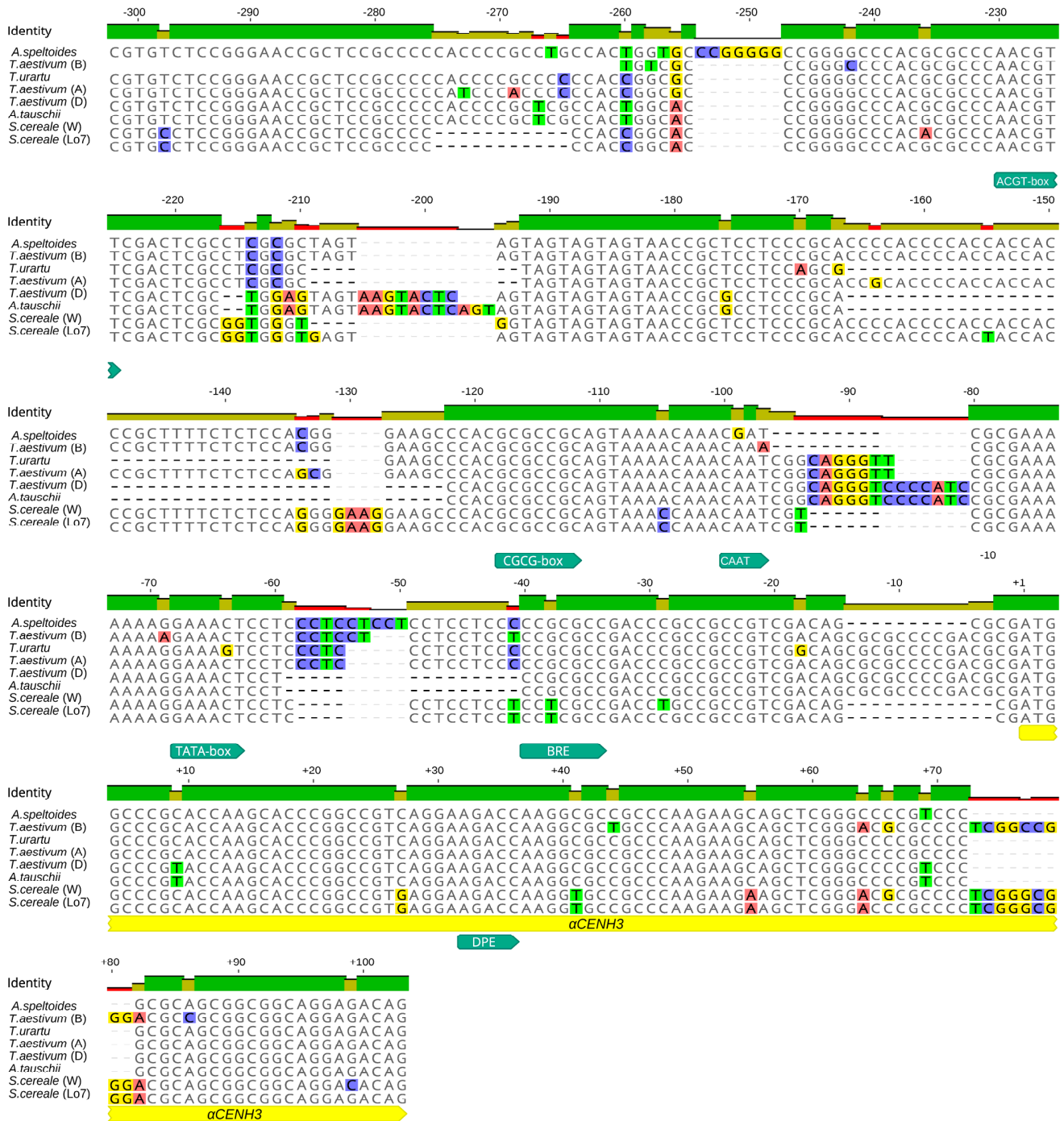


Figure 2. Alignment of nucleotide sequences in −300 and +100 regions around the TSSs of α CENH3, where functional motifs were searched for. Functional motifs appear as turquoise fingerpost arrows. The gene transcriptional regions appear as yellow fingerpost arrows.

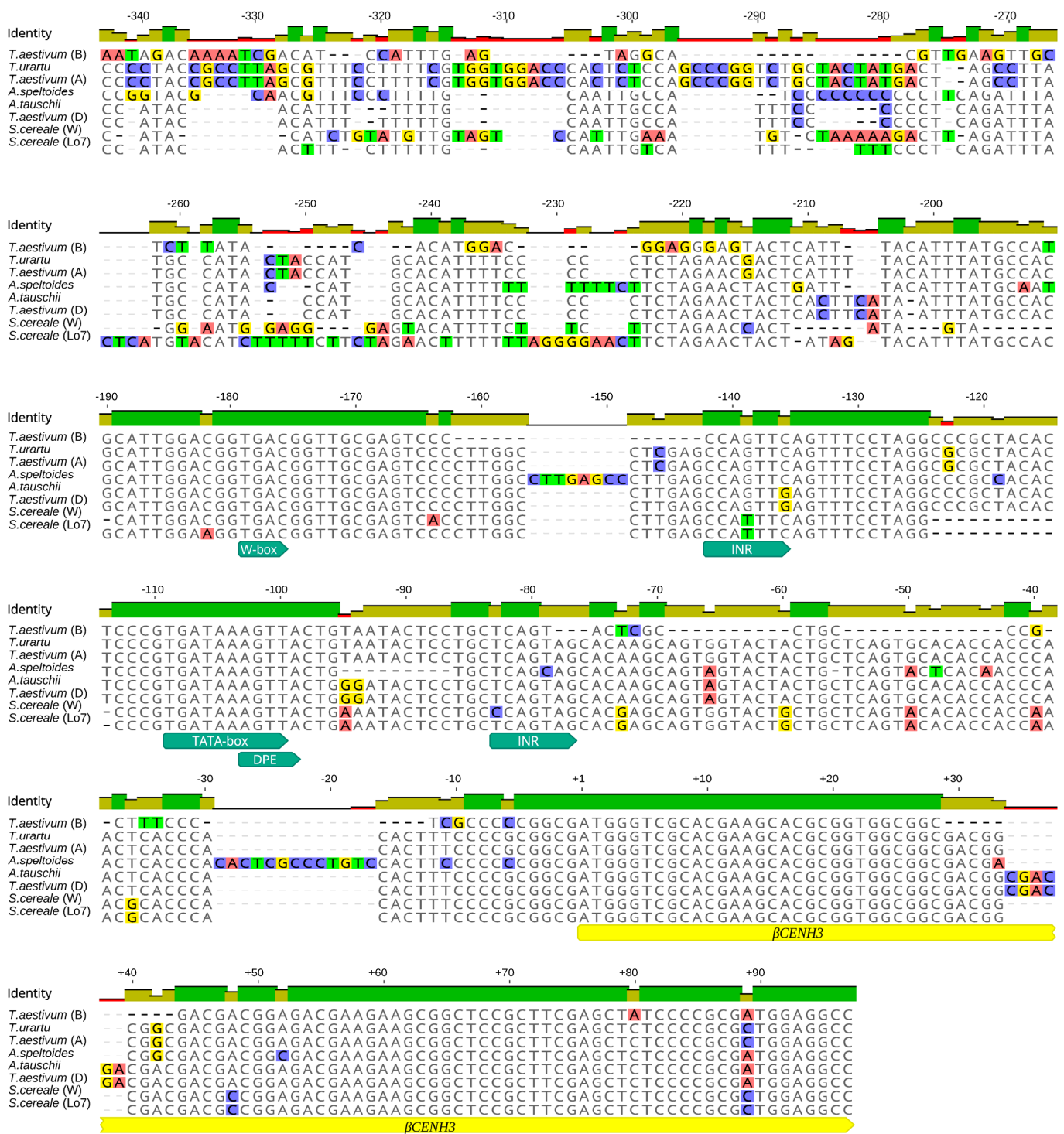


Figure 3. Alignment of nucleotide sequences in -300 and $+100$ regions around the TSSs of β CENH3, where functional motifs were searched for. Functional motifs appear as turquoise fingerpost arrows. The gene transcriptional regions appear as yellow fingerpost arrows.

In addition to two pairs of promoter sequences, the regulatory region of the β CENH3 gene in all Triticeae species contains a TGAC motif (at position -139 in rye), which is known as a W-box or DNA motif for specific binding of WRKY proteins, a TF superfamily [28]. Members of this superfamily participate in the regulation of various plant-specific physiological processes [28]. The TGAC motif is absent in the regulatory region of the α CENH3 gene, in which, in addition to the TATA-box, five more regulatory elements have been found. The way these regulatory elements are located around α CENH3 is more con-

sistent with a common pattern than the way they occur around β CENH3. For example, the core DPE promoter is located at position +30 downstream of the TSS. In addition to the most common core promoter motifs, TATA and DPE, the α CENH3 gene contains a recognition element, BRE, to recognize the general transcription factor TFIIB. The entire –300 bp region in the α CENH3 gene is rich in GC. TSSPlant identifies there a CGCG box upstream of the TATA-box. CG-rich regions serve as a template for epigenetic mechanisms in many promoters and the presence of such promoters is associated with histone modifications [29]. This CG enrichment is characteristic of TATA-less promoters, which suggests that the α CENH3 gene has in its regulatory region rather a complex architecture consisting of several promoters—as does β CENH3. Another functionally important motif in the regulatory region of the α CENH3 gene is the ACGT box, which is farthest from the TSS (–180 bp in rye). This motif, together with the coupling element (CE), forms part of the ABRC of promoter complexes responding to the action of abscisic acid (ABA), a hormone that inhibits plant growth and development [30].

2.3. Transcription of the CENH3 Genes

The presence of promoters and regulatory motifs around the CENH3 genes raises the question as to what effect they might have on the behavior of both variants in different tissues during the development of plants. To find out, we compared their transcription levels by qPCR in the rye variety ‘Imperial’ (*S. cereale* subsp. *ceriale*) and ‘Korotkostebel’naya69’ (*S. cereale* subsp. *ceriale*, K69 throughout) carrying the dominant dwarfing gene *Ddw1*. The results presented in Figure 4 show two general trends. One indicates that the transcription of both variants CENH3 reaches its highest in the generative tissue. The most remarkable difference is observed between transcription in pistils and transcription in leaves and stems (quantitative data are given in Table S1). It is about 30-fold for α CENH3 and β CENH3 in ‘Imperial’ and about 60-fold for α CENH3 and 30-fold for β CENH3 in ‘K69’. The other suggests (with one exception) that in both varieties, transcription levels of α CENH3 are higher than those of β CENH3 at all developmental stages studied except in stems (Figure 4). The largest difference between the variants is observed in coleoptiles: 4.7-fold in ‘Imperial’ (Table S1) and 3.0-fold in ‘K69’. However, in stems, where transcription is most downregulated, the transcription level of β CENH3 is 2.4 times higher in ‘Imperial’ and 3.2 times higher in ‘K69’. Thus, it should be noted that both tendencies are characteristic of both varieties, which differ significantly in some phenotypic traits.

2.4. Subdomain Organization of Nucleosomes Containing Different CENH3 Variants in Centromeric Chromatin

To find out whether proteins are synthesized from the α CENH3 and β CENH3 variants, whether these proteins are loaded into nucleosomes 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. That the number of immunostaining signals in flow-sorted interphase nuclei isolated from rye seedlings (Figure 5a) was equal to the chromosome number indicates that both α CENH3 and β CENH3 are indeed loaded into rye centromeres. The colocalized CENH3 ultrastructures visualized by SIM are intermingled (Figure 5b, Video S1).

Thus, it is likely that rye centromeres contain subdomains of CENH3-free nucleosomes. Analysis of the α CENH3 and β CENH3 signals at individual stages of mitosis (Figure S1) demonstrated their presence on chromosomes throughout the entire process of cell division. This suggests a synchronized synthesis and dynamics of incorporation of both variants into nucleosomes.

A similar trend was observed in the analysis of the dynamics of the α CENH3 and β CENH3 signals when young anthers are progressing through meiosis. Figure 6 shows that chromosomes tend to cluster during meiosis. In interphase I of meiosis, CENH3 signals varying in size—3–4 large, composed of associated centromeres, and 3–5 singletons—are located on one side of the nucleus. Putting chromosomes together into a small number of groups at the nuclear periphery is a commonly observed phenomenon, which is centromere

coupling involving nonhomologous, pairwise associations [31]. In the zygotene stage, green signals (β CENH3) slightly outnumber red signals (α CENH3), while in pachytene, when the chromosomes appear more extended, red signals begin to prevail. Because the quantitative fluctuations in the intensity of signals at this level of resolution depend on many factors, and so it is not feasible to assess their influence, the best we can do is confirm that the dynamics of the loading of both CENH3 variants into meiotic chromosomes corresponds to that observed for mitotic chromosomes.

To understand the subcentromeric organization of both CENH3 proteins, we used extended chromatin fibers prepared from interphase nuclei isolated from 5-day-old seedlings. The fibers were immunolabeled and analyzed by confocal laser scanning microscopy. Confocal image stacks collected from immunostained clusters for the α CENH3 and β CENH3 proteins with different lengths were measured in nanometers and converted to units of length for DNA, a diameter of 11 nm per nucleosome [32] and a mean nucleosome size of 180 bp [33]. We measured and quantitatively analyzed the following: (1) the size of clusters consisting only of α CENH3 signals (red) (Figure 7a), (2) the size of clusters consisting only of β CENH3 signals (green) (Figure 7b), (3) clusters consisting of alternating signals of both proteins (red, green and yellow signals) (Figure 7c,d) and (4) the sizes of the gaps between the signal tracks. Such gaps were found between the clusters of α CENH3 signals and tracks consisting of signals of both proteins, with α CENH3 signals prevailing. They are likely represented by regions of centromeric chromatin, where nucleosomes include the canonical histone H3.

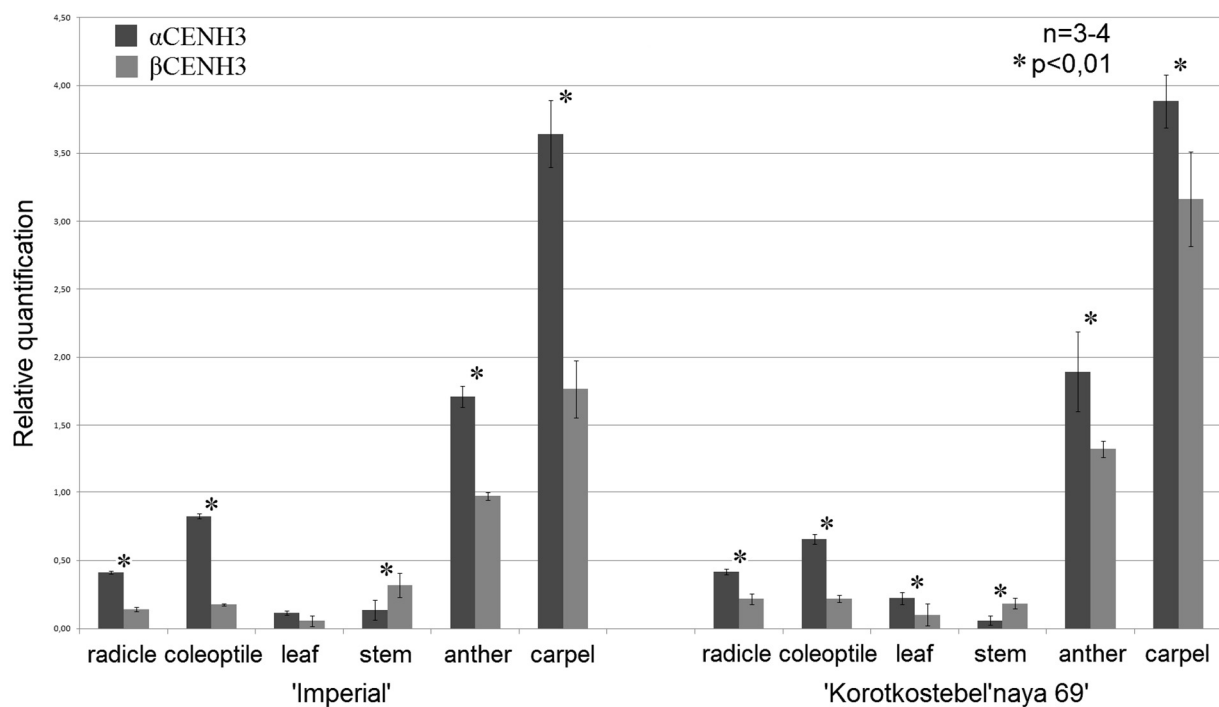


Figure 4. Transcription levels of the α CENH3 and β CENH3 genes in different tissues of two rye (*S. cereale*) varieties. Transcription levels normalized to the geometric mean of two references. Each value represents the mean of target/reference ratios with SD bar (linear bar chart). Significant differences between genes for each tissue type are asterisked (*), $p < 0.01$ (Mann–Whitney U test).

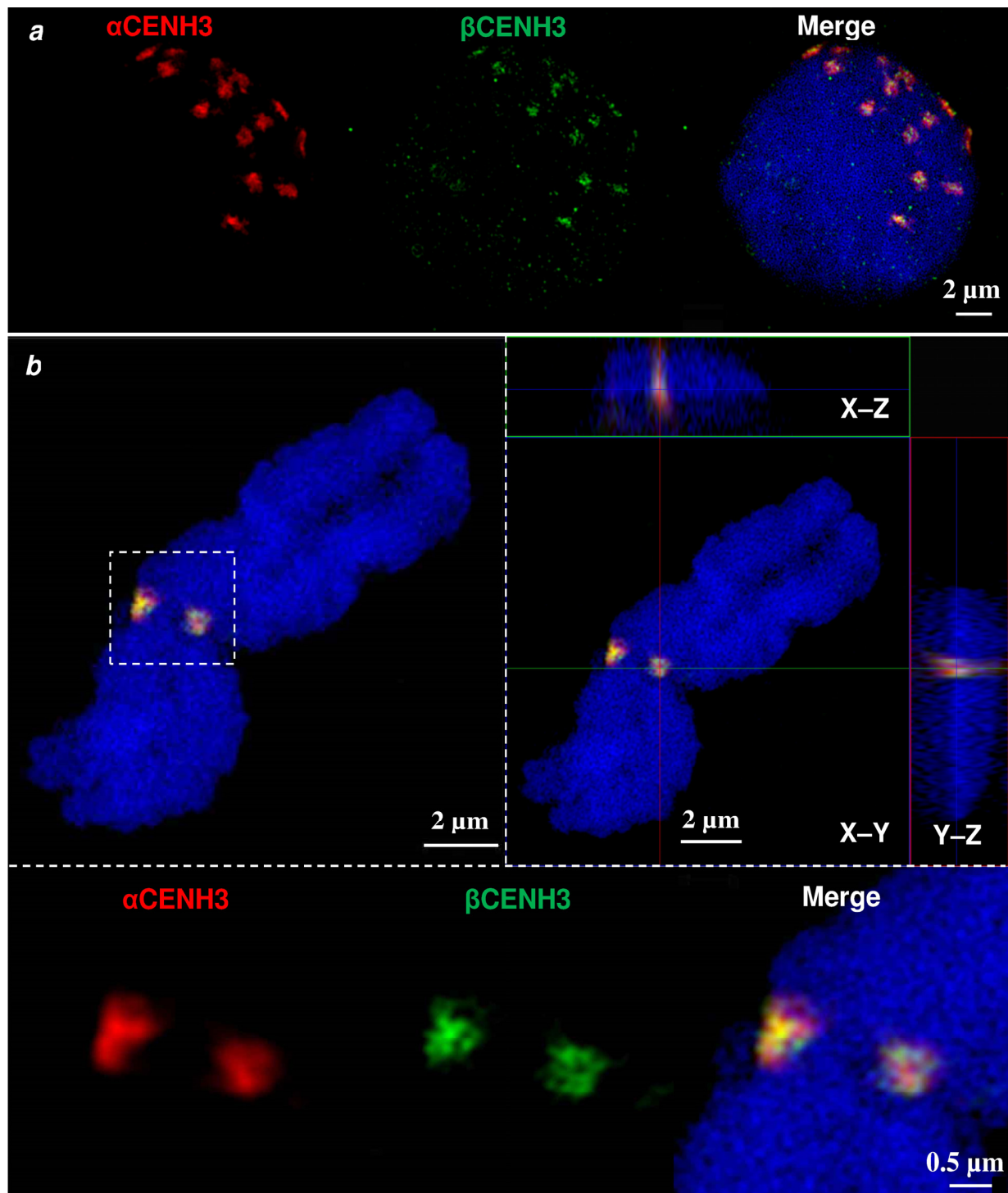


Figure 5. α CENH3 and β CENH3 localization in the centromeres of interphase nuclei (a) and metaphase chromosomes (b) of rye. The increased resolution achieved via SIM reveals that both CENH3 variants colocalize and intermingle. (a) The centromere concentration at the top-right clearly visualizes the Rab1 orientation in the nucleus. (b) The orthoview (right) indicates the spatial distribution of CENH3 within the centromere. Below, the centromeric region (dashed rectangle) is shown enlarged.

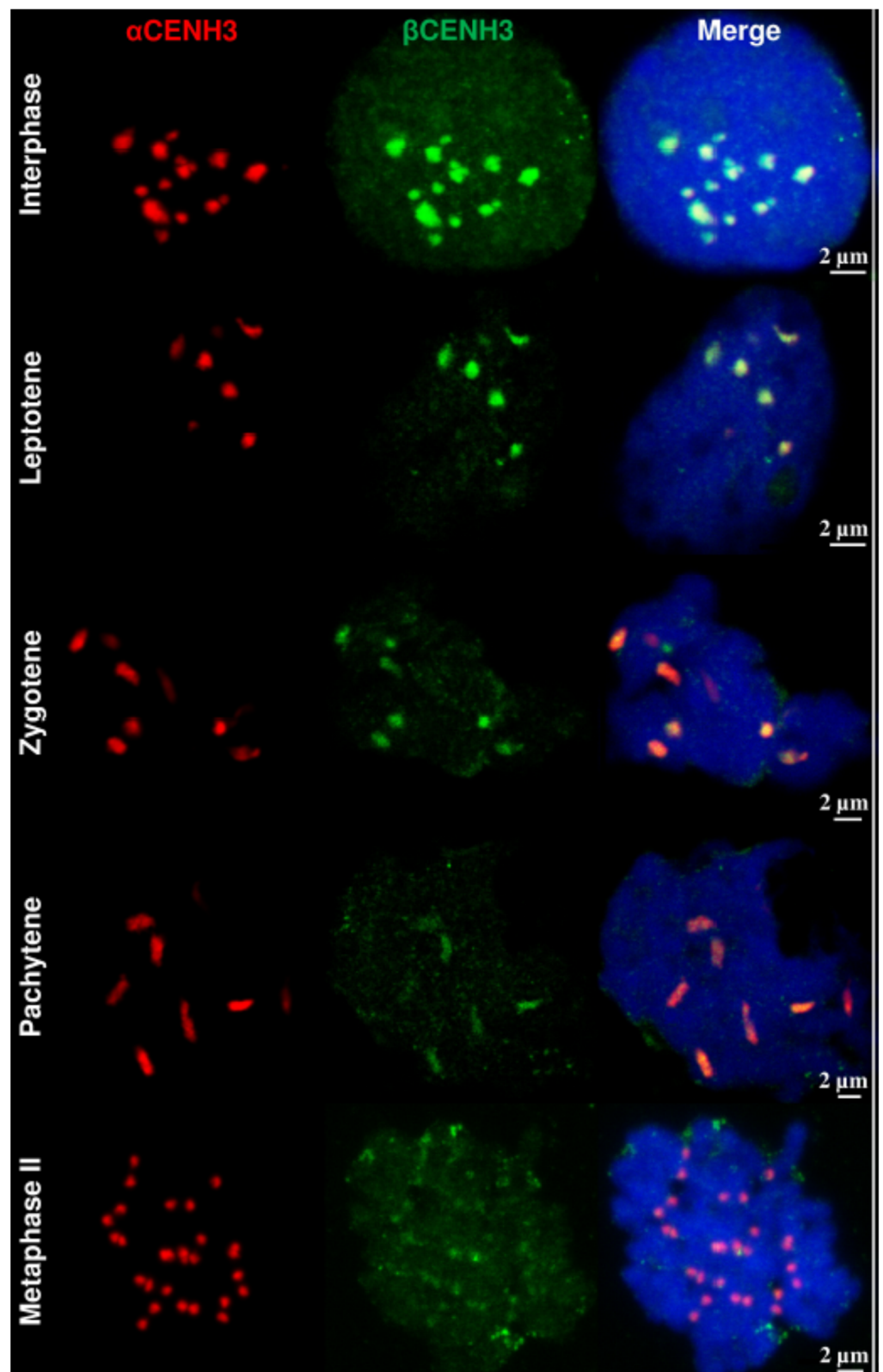


Figure 6. Loading the α CENH3 and β CENH3 proteins into rye chromosomes in meiosis. CENH3 variant-specific antibodies in combination with laser confocal scanning microscopy were used for analysis of meiotic stages in young anthers.

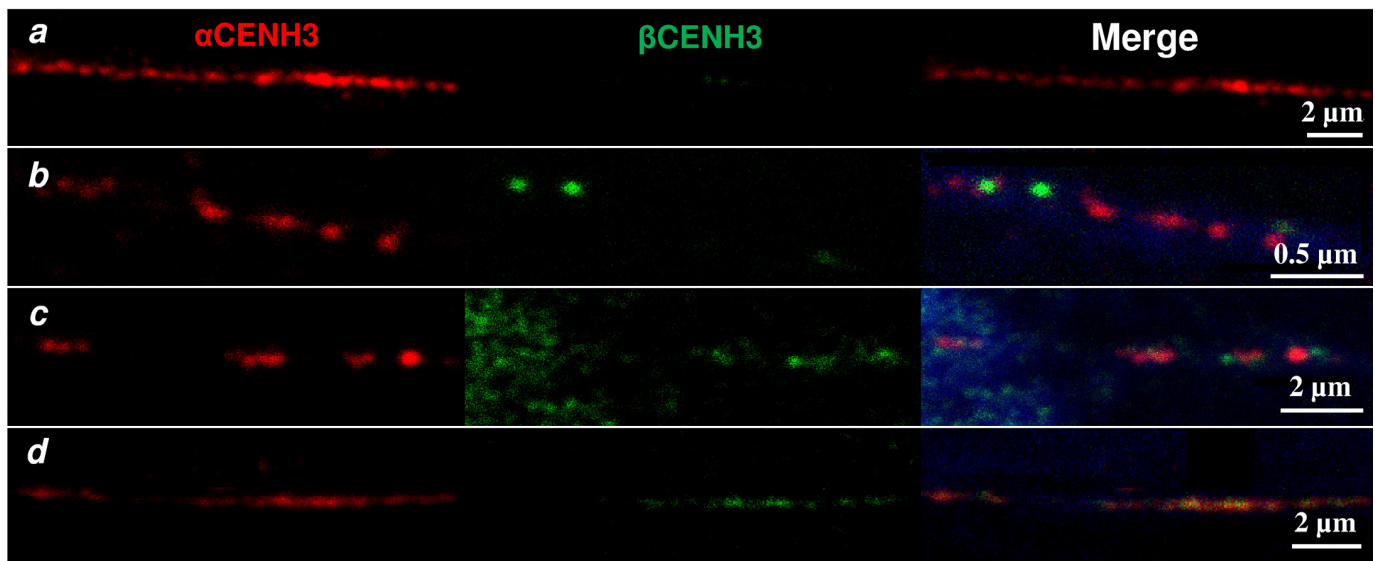


Figure 7. Different types of nucleosome clusters visible on extended chromatin fibers. (a) Clusters containing nucleosomes with α CENH3-only signals. (b) Clusters containing nucleosomes with β CENH3-only signals. (c,d) Examples of clusters containing nucleosomes with α CENH3 and β CENH3 signals colocalized. The colocalization values are: c = β CENH3/ α CENH3—25–75%, α CENH3/ β CENH3—25–35%; d = β CENH3/ α CENH3—56–90%, α CENH3/ β CENH3—22–62%. Clusters with CENH3 signals are interrupted by gaps with nucleosomes containing the canonical histone H3.

Table 2 shows the results of the analysis, which reveals two clear-cut tendencies:

Table 2. Characteristics of chromatin fibers with CENH3 variants in the rye centromeric chromatin.

Types of Clusters with Variants of CENH3	Clusters Measured	Size, nm	Size, kb *
Clusters containing nucleosomes with α CENH3-only signals	97	278–2588	4.5–42.3
Clusters containing nucleosomes with β CENH3-only signals	27	202–714	3.3–11.7
Clusters containing nucleosomes with α CENH3 and β CENH3 signals colocalized	48	293–7205	4.8–117.9
Size of gaps between clusters with CENH3 signals	72	206–5399	3.4–88.3

* To calculate cluster sizes in kilobases, we assumed that the chromatin fibers have approximately the same and evenly distributed degree of stretching. The diameter was set at 11 nm per nucleosome [31], and the mean size of nucleosomes was set at 180 bp [32].

(1) β CENH3 signals form the shortest clusters, the maximum size of which is about 12 kb, while the maximum size of the clusters of α CENH3-only signals reaches about 40 kb.

(2) The greatest range of values is observed in the clusters consisting of signals from both α CENH3 and β CENH3 proteins (red, green and yellow) and is from about 5 to 120 kb. We think that such clusters are indeed formed by chromatin regions consisting of alternating nucleosomes with α CENH3 and β CENH3, which is confirmed when using the “Colocalization” option of the software ZEN (Zeiss). Quantitative estimates of colocalization are given in the Figure 7 caption. As the size distribution plots show (Figure S2), the increase in the length of the clusters occurs gradually. Most of them are relatively small in size. The majority of clusters with α CENH3 signals are found within 1000 nm or approximately 15–16 kb, and most of the clusters with alternating signals of both proteins as well as gaps in signals are found within 1500 nm or 25 kb. A smooth appearance of the plots is disrupted by a sharp increase in the size of the clusters with the signals of both proteins within 2000–7000 nm (or 30–120 kb) in a small number of measurements made (six). Even though the sizes of the extended chromatin fibers and, consequently, the number of clusters with signals can be influenced by various factors, including mechanical ones, the large number of measurements that we have performed allows us to claim that the tendencies noted above reflect the likely picture of the nucleosomal organization of centromeric chromatin in rye.

3. Discussion

Gene duplication and the preservation of both copies, parental and daughter, during evolution is an intriguing evolutionary phenomenon, especially if we look at closely related taxonomic entities (tribes, genera and even species) that have been getting by with a single copy for millions of years. Although an important role is attributed to gene duplications in generating evolutionary novelty and adaptation [2], this role is often not so evident. The four possible molecular mechanisms supporting the preservation of the daughter copy in the genome during evolution [3] are not equally widespread across taxa. A comparison of the expression profiles between the ancestral single-copy gene in one species and the parent and child copies arising from an SSD event in a closely related sister species demonstrated that 65% of duplicated genes in *Drosophila* have undergone *neofunctionalization* [34]. By contrast, the use of this approach with eight mammalian species suggests that only 33% of the duplicated genes have been preserved by *neofunctionalization*, while the rest (i.e., most) carry on due to the *conservation* of functions [35]. This tendency appears to be also characteristic of plant species, which is confirmed by an impressive analysis performed by Jiang and Assis [3]. From three cereal species, the authors formed 272 SSD-derived gene pairs in *Brachypodium distachyon*, 289 pairs in *Oryza sativa japonica* and 340 pairs in *Sorghum bicolor*. It was demonstrated that duplicated genes in all these species traverse similar evolutionary paths after SSD and 60.6% of these duplicated genes were performing the same function as the ancestral copy (that is, they were conservative), 23.8% had acquired a new function (that is, they became *neofunctionalized*), 0.4% had become *subfunctionalized* and 15.2% had become specialized [3]. Thus, *conservation* is the most preferred mechanism with these species; it preserves the duplicated copy in the genome during evolution, which leads to an increase in gene dosage [35]. Interestingly, the *CENH3* locus consisting of the syntenic genes *LHCB3-l*, *CENH3* and *bZIP* is also present in *B. distachyon* and *O. sativa* [15]. However, in *O. sativa*, the *CENH3* gene occurs in the locus as a single copy, while in *B. distachyon*, it occurs as two copies—just as in rye (Figure 1)—strongly suggesting that whether or not a particular gene would become duplicated could rather be a chance circumstance.

No matter what molecular mechanism of duplication, DNA-mediated or RNA-mediated, the copies find themselves in a new genomic context outside their ancestral regulatory elements, hence the expectation that the copies will become *neofunctional* or *subfunctional*. That is why it was somewhat surprising that, according to the above publications, the copies do not live up to expectations. A comparison of the molecular compositions of DNA sequences in the *CENH3* locus between rye and Stipeae and Brachypodieae species that are the first cereals to have evolved two variants of the *CENH3* gene (about 35–40 MYA) shows a general tendency towards the expansion of the *CENH3* locus. An increase in the size of the locus and, consequently, in the distance between β *CENH3* and α *CENH3* genes was due to the massive introduction of various families of TEs, mainly of the main LTR-containing retrotransposon superfamilies, *gypsy* and *copia*. A comparison of the composition and prevalence of these families between this locus and the genomes of different Triticeae species revealed substantial differences [15]. In this respect, the *CENH3* locus in rye is different from the rest of the genome too (Table 1). However, hardly could these differences exert any effect on the expression levels of the β *CENH3* and α *CENH3* genes: the insertion sites of TEs begin and end several kilobases away from the genes, and these peculiar buffer zones are largely filled with short tracks of micro- and minisatellites, AT-enriched regions and unidentified sequences.

The alignment of the primary structure of –300 and +100 regions relative to the TSS, that is, the regions where TF-binding DNA motifs [24] and regulatory elements responsible for specific transcription regulation [25] normally reside, showed that they are highly conserved in both the α *CENH3* and β *CENH3* genes of different Triticeae species (Figures 2 and 3). However, a comparison of these regions with their counterparts in *Brachypodium* species, *B. sylvaticum* and *B. distachyon*, revealed a very high level of heterogeneity, which prevented their alignment with Triticeae 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 α CENH3 and β CENH3. These sets contain both common regulatory motifs and those specific for each gene. A common feature of both genes is the complex architecture of their promoters; along with TATA boxes identifiable by TSSPlant [27], it contained the DPE and INR motifs characteristic of TATA-less promoters. It is possible that the extensive sets of functional sites make the regulatory regions of the α CENH3 and β CENH3 genes so universal. Comparing the expression levels of both genes between different developmental stages of two phenotypically different rye varieties, 'Imperial' and 'K69', revealed a clear-cut correlation between expression levels and cell division intensity (Figure 4 and Table S1). The transcription level of α CENH3 prevailed over that of β CENH3 in all studied stages of plant development, except in stems. The importance of a higher level of α CENH3 expression is confirmed by the result of an amino acid exchange leading to impaired centromere loading of β CENH3, which, however, did not cause phenotypic changes in barley [36].

Cell division in stems is lowest, and the expression of cell cycle-related genes probably reaches its minimum basal level. This phase of plant development was the only one at which the transcription level of the β CENH3 gene was higher than that of α CENH3. This fluctuation might be due to the presence of the functionally important motif ACGT box in the regulatory region of the α CENH3 gene. This motif is part of ABRC promoter complexes responding to the action of abscisic acid (ABA), a hormone that inhibits plant growth and development [30]. Because this motif is not found in the vicinity of the β CENH3 gene, the inhibitory effect of the hormone is probably exerted on α CENH3 to a higher degree. Together with our comparison of the transcription levels of the α CENH3 and β CENH3 genes at different stages of plant development, the presence of the proteins synthesized from these transcripts in chromosomes at different stages of mitosis and meiosis reveals no sign of neofunctionalization, subfunctionalization or specialization in the copies. The influence of negative selection (purifying selective pressure) on the coding part of the β CENH3 gene [15] led this gene to preserve its conserved function, and the benefit of the presence of two genes normally appears as the gene-dosage effect. Another advantage could be associated with the fact that each copy encodes the N-terminal tail with big differences in amino acid sequences, which is the case for the α CENH3 and β CENH3 proteins in rye [12], and may form the basis for different protein–protein interactions [37]. It is possible that new approaches to studying and assessing functional divergence will help reveal the above functions in cereals—in addition to conservation. The study of centromeric functions of the CENH3 variants in cowpea through the CRISPR/Cas9-based inactivation of one of the variants [38] is an example of an informative approach.

The question of the linear organization of nucleosomes in centromeric chromatin in species expressing more than one variant of the centromere histone H3 is a special question. Assuming that separate clusters of signals correspond to particular nucleosomal subdomains, we identified four main patterns of the co-organization of the CENH3 variants, which correspond to the main polynucleosomal subdomains in centromeric chromatin. It should be noted that not all CENH3 is loaded into chromatin in many species. For example, about 66% of CENH3 was found to be loaded into chromatin outside centromeres in human retinal pigment epithelium cells [39] and about 30% in chicken cells [40]. Not aware of any similar quantitative assessments in plants, we propose that the four patterns of polynucleosomal CENH3 subdomains form part of centromeres. A similar subdomain organization of centromeric chromatin with alternating nucleosomal clusters containing both CENH3 variants and nucleosomes containing the canonical histone H3 was described for centromeres in barley [16], legumes [38,41], humans and *Drosophila* [42,43]. Thus, the organization of chromatin with polynucleosomal subdomains containing a unique set of histone variants is evolutionarily conserved for centromeres and makes centromeric chromatin different from the neighboring noncentromeric regions [41–43]. A high diversity of the sizes of the four patterns suggests that each chromosome of rye has, apparently, a specific organization of polynucleosomal subdomains. This diversity reveals several most

extended clusters largely containing signals from both CENH3 variants likely interrupted by small nucleosomal subdomains with the canonical H3 (Figure S2). We propose that such extended subdomains may appear to be a centromeric core surrounded peripherally by shorter subdomains with diverse combinations of histone variants and post-translational histone modifications.

4. Materials and Methods

4.1. Identification of the CENH3 Locus in the Rye Genome; Analysis of Its DNA Sequences

The rye genome assembly was downloaded from GenBank (https://www.ncbi.nlm.nih.gov/assembly/GCA_902687465.1, accessed on 9 December 2020). The search for CENH3 sequences in the genome was performed using the TBLASTN program from the AB-BLAST package [44]. The amino acid sequence of rye CENH3 obtained by us earlier (ACC AUN88458) was used as a query [15]. The composition of repeated DNA sequences in the CENH3 locus was determined using the RepeatMasker version open-4.0.5, sensitive mode, and the Viridiplantae database (RepeatMasker-RepBase Sequence Database RELEASE 20170127) [45]. Data were visualized using the Geneious 11.0.2 software (<http://www.geneious.com>, accessed on 9 December 2020) [46]. For comparative analysis, the genomic sequences of CENH3 in other Triticeae species obtained from previously sequenced genomes were used [15].

4.2. Identification of Functional Sites in the Vicinity of the α CENH3 and β CENH3

Multiple alignment of regulatory regions in the vicinity of the α CENH3 and β CENH3 genes was performed using MUSCLE, v3.8.31 [47]. Identification of functional sites was performed using TSSPlant and NSITE-PL [48,49] on <http://www.softberry.com>, accessed on 9 December 2020. Core promoter elements were identified using the online tool YAPP Eukaryotic Core Promoter Predictor (<http://www.bioinformatics.org/yapp/cgi-bin/yapp.cgi>, accessed on 16 December 2020).

4.3. Plant Material

Transcription levels of α CENH3 and β CENH3 were determined in vernalized rye plants of two cultivars, *S. cereale* subsp. *cereale*, cv. 'Imperial' and *S. cereale* subsp. *cereale*, cv. 'Korotkostebel'naya69'. The seeds were obtained from the gene banks of the N.I. Vavilov Institute of Plant Genetic Resources, St.Petersburg, Russia ('Imperial', accession No. 9368, K69 accession No. 10892), and Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany ('Imperial', accession No. R1771). Plants were grown in a greenhouse (a 16/8 h day/night cycle at a 25 °C/20 °C day/night temperature). The growth stages at which the samples were collected for analysis were characterized by the Zadoks scale (Zadoks 2-digit code [50]): (1) primary roots, 2–3 cm (Z06); (2) coleoptiles, 1.5–2 cm (Z09); (3) leaves at tillering (Z21–Z24); (4) stems at the 2-node stage (Z32); (5) anthers at the stage between meiosis and development of mature pollen (Z42–44); (6) nonpollinated pistils (Z50–52). Tissues of the same type from five different plants were combined into one biological sample.

4.4. RNA Extraction and cDNA Synthesis; RT-qPCR

Total RNA was extracted from the biological samples using the TRIzol method (TRI Reagent, MRC, Inc., OH, USA) and treated with DNase using a DNA-free DNA Removal Kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). First-strand cDNA was synthesized using FireScript Reverse Transcriptase (R04-50, BiolabMix, Novosibirsk, Russia) in a volume of 20 μ L from 3.6 μ g of DNase-treated total RNA. Each PCR reaction was carried out in a volume of 25 μ L and contained 12.5 μ L BioMaster HS-qPCR SYBR Blue (2 \times) (BiolabMix, Novosibirsk, Russia), 8 μ L of first-strand cDNA diluted 25 times (which corresponds to 0.32 μ L of the initial cDNA per reaction) and 0.3 mM of forward and reverse primers. The optimum concentration and temperature for primer annealing had preliminarily been set for each target. Reactions were performed in duplicate for the reference

genes and in triplicate for α CENH3 and β CENH3 using a LightCycler 480 Instrument II (Roche, Basel, Switzerland) according to the following program: at 95 °C for 5 min, followed by 45 cycles at 95 °C for 15 s, at the annealing temperature of 62 °C for 20 s, and at 72 °C for 30 s.

Because the literature had nothing about reference genes suitable for comparative transcription analysis in the rye genome, the reference genes used for the normalization of the transcription levels of the rye genes α CENH3 and β CENH3 were Ta2776 (RNase L inhibitor-like protein) and Ta53967 (Vacuolar ATP synthase 16 kDa proteolipid subunit), selected following the search for the best reference genes for various tissues of *T. aestivum* [51]. For amplification of the transcripts of the rye reference genes, primers were selected taking into account the differences between wheat and rye sequences of the reference genes, for which purpose we used contigs of *S. cereale* orthologous genes (BioSample: SAMEA3928734BioProject: PRJEB13501, <https://www.ncbi.nlm.nih.gov/biosample/SAMEA3928734/>, accessed on 20 September 2020). Primers used in the study are listed in Table S2. The amplification products of the reference genes were checked by electrophoresis and were found to be consistent with the expected lengths of the amplicons.

The primers for amplification of the α CENH3 and β CENH3 transcripts (Table S2) were designed using an IDT PrimerQuestTool (<https://eu.idtdna.com/pages/tools/primerquest/>, accessed on 18 November 2020) at the exon boundaries in order to exclude the influence of genomic DNA [12]. Consistency between the reaction products and expected products was checked by sequencing. The relative quantification was carried out using the standard curve method [52]. External standard curves for each gene were constructed in a series of six five-time dilutions. The results obtained were analyzed using the LC480 software (section 1.5.1.62) supplied with the LightCycler 480. The primer efficiency ranged from 96.6% for β CENH3 to 102.6% for α CENH3 and 100% for the reference genes. The comparison of the transcription levels of α CENH3 and β CENH3 was performed using the “Relative quantification analysis with high confidence 2nd derivative max method” (LC480 software 1.5.1.62).

4.5. Synthesis of Antibodies

The following peptides were used to obtain polyclonal antibodies: rye α CENH3, KKLGTTPSGGTQRRQDTDGAGTSATPRRAGR, in rabbits; rye β CENH3, TATTPEKKKRLRFELSPRWR, in mice. Life-Tein (LLC, NJ, USA, www.lifetein.com, accessed on 20 March 2017) and Biosan Company (Novosibirsk, Russia, <https://biosan-nsk.ru>, accessed on 12 March 2019) performed the peptide synthesis, immunization of rabbits and mice and peptide affinity purification of antisera.

4.6. Slide Preparation and Indirect Immunostaining

Seedlings of the rye variety ‘Imperial’ with about 2 cm long main roots were incubated for 20 h at 4 °C in ice water. The nuclei were sorted according to [53]. Mitotic preparations were made from root meristems fixed in Tris buffer containing freshly prepared 3% paraformaldehyde (PFA) under vacuum at 4 °C for 5 min followed by a 20 min fixation at 4 °C without vacuum [54]. After washing for 5 min three times in ice-cold Tris buffer, the meristems of about 20 root tips were chopped up with a sharp razor blade in 1 mL of ice-cold LB01 lysis buffer composed of 15 mM Tris-HCl, 80 mM KCl, 20 mM NaCl, 2 mM disodium EDTA, 0.5 mM spermine, 0.1% (*v/v*) Triton X-100 and 15 mM mercaptoethanol, pH 7.5 [55]. This suspension of released chromosomes and nuclei was filtered through a 50 μ m pore size filter of Cell Strainer cap (Corning Inc., NY, USA) and kept on ice. Filtrated suspension was diluted 1:10 in LB01 lysis buffer, and 100 μ L was centrifuged onto slides by using a Shandon Cytospin 4 (Thermo Fisher Scientific, Waltham, MA, USA) at 2000 rpm for 10 min. The loaded slides were treated with 0.5% Triton X-100 for 5 min at 4 °C, fixed in 4% PFA for 10 min at room temperature (RT) and rinsed for 5 min three times in 1 \times phosphate-buffered saline (PBS) [56]. Extended chromatin fibers from young etiolated leaf nuclei were prepared as described by Houben et al. [57]. For the analysis of pollen mother cells, meiotic chromosome slides were prepared according to Hesse et al. [58].

Immunostaining was performed according to Jasencakova et al. [54] with minor modifications. The nuclei were postfixed in 4% (w/v) PFA in PBS for 10 min, washed three times in PBS and blocked for 1 h at RT in 2xSSC containing 3% BSA and 0.1% Triton X-100. Immunostaining was performed at 4 °C for 48 h with primary antibodies diluted in AK buffer (2xSSC containing 1% BSA and 0.1% Triton X-100). Slides were washed for 5 min three times in 2xSSC, and secondary antibodies diluted in AK buffer were applied for 1 h at 37 °C. Slides were washed for 5 min three times in 2xSSC and afterward mounted and counterstained with 4',6-diamidino-2'-phenylindole dihydrochloride (DAPI, 1 mg/mL) in Vectashield (Vector Laboratories Inc., Burlingame, CA, USA). The following primary antibodies were used: rabbit rye anti- α CENH3 (diluted 1:500) and mouse rye anti- β CENH3 (diluted 1:75). As secondary antibodies, a goat anti-rabbit rhodamine (1:200; Jackson Immuno Research Laboratories, West Grove, PA, USA) and goat anti-mouse Alexa 488 (1:400; Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) were used.

4.7. Microscopy

To analyze the ultrastructure of FISH signals and chromatin beyond the classical Abbe/Raleigh limit, we applied spatial structured illumination microscopy (3D-SIM) at a lateral resolution of ~120 nm (super-resolution, achieved with a 488 nm laser), using a 63 \times /1.4 oil plan-apochromat objective of an Elyra PS.1 microscope system with the software ZEN Black (Carl Zeiss GmbH, Jena, Germany). Images were captured separately for each fluorochrome using the 405, 488 and 561 nm laser lines for excitation and appropriate emission filters [59]. A SIM image stacks were used to produce the 3D movie with the Imaris 8.0 (Bitplane AG, Zurich, Switzerland) software. Maximum intensity projections were calculated via the ZENBlack software. The male meiosis and extended chromatin fiber images were captured by a confocal laser scanning microscope with a 100 \times /1.46 oil plan-apochromat objective (LSM 780, Carl Zeiss GmbH, Jena, Germany), with the multitrack configuration for detecting rhodamine (excitation at 561 nm, emission collected by a bandpass of 566–685 nm) and Alexa 488 (excitation at 488 nm, emission collected by a bandpass of 497–558 nm). Laser intensity and gain were set at similar levels for all experiments. Maximum intensity projections were done with the ZEN2010B software. Measurements and colocalization analysis were done by the Line measurement tool and Colocalization function of the ZEN2010B software.

Supplementary Materials: The following are available on line at <https://www.mdpi.com/article/10.3390/plants10102043/s1>, Figure S1: Loading the α CENH3 and β CENH3 proteins into rye chromosomes at different mitotic stages. CENH3 variant-specific antibodies in combination with laser confocal scanning microscopy were used for analysis of mitotic rye cells at early prophase, late prophase, metaphases, anaphase and telophase (half of the telophase is shown). Figure S2: Size distribution of different types of nucleosome clusters in centromeric chromatin in rye (data from Table 2). On the x-axis: number of measurements made; on the y-axis: size of clusters in nm. Red curve: distribution of clusters with α CENH3 signals; green curve: distribution of clusters with β CENH3; yellow curve: distribution of clusters with α CENH3 and β CENH3 signals; dark-blue curve: distribution of gaps between clusters with signals telling that the nucleosomes contain the canonical histone H3. Table S1: Normalized transcription levels of α CENH3 and β CENH3 in different tissues of rye (*S. cereale*). Table S2: List of primers used in this study. Video S1: A spinning rye metaphase chromosome demonstrating the loading and centromeric localization of α CENH3 (red) and β CENH3 (green).

Author Contributions: Conceptualization, A.V.V. and A.H.; methodology, E.V.E., E.A.E. and V.S.; formal analysis, E.A.E. and S.S.G.; investigation, E.V.E., E.A.E. and S.S.G.; writing—original draft preparation, A.V.V.; writing—review and editing, all authors; visualization, E.V.E. and V.S.; supervision, A.V.V. and A.H.; project administration, A.V.V. and A.H.; funding acquisition, A.V.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Russian Fundamental Scientific Research Program (project No.0246-2021-0014); molecular analysis was supported by a grant from the Russian Foundation for Basic Research (project No.20-04-00699), and cytogenetic analysis was funded by the Russian Science Foundation (project No.19-14-00051).

Acknowledgments: We thank Katrin Kumke (IPK, Gatersleben) and E.Grishko (IMCB, Novosibirsk) for the technical help, Jörg Fuchs (IPK, Gatersleben) for flow sorting of nuclei and A.I. Shevchenko (Institute of Cytology and Genetics (ICG SB RAS, Novosibirsk)) for valuable advice in cytogenetic analysis. We thank the Microscopic Center of the ICG SB RAS (Novosibirsk, headed by S. I. Bayborodin) for access to microscopic equipment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Van de Peer, Y.; Mizrachi, E.; Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **2017**, *18*, 411–424. [[CrossRef](#)]
2. Flagel, L.E.; Wendel, J.F. Gene duplication and evolutionary novelty in plants. *New Phytol.* **2009**, *183*, 357–364. [[CrossRef](#)] [[PubMed](#)]
3. Jiang, X.; Assis, R. Rapid functional divergence after small-scale gene duplication in grasses. *BMC Evol. Biol.* **2019**, *19*, 97. [[CrossRef](#)] [[PubMed](#)]
4. Force, A.; Lynch, M.; Pickett, F.B.; Amores, A.; Yan, Y.-I.; Postlethwait, J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **1999**, *151*, 1531–1545. [[CrossRef](#)] [[PubMed](#)]
5. Hittinger, C.T.; Carroll, S.B. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **2007**, *449*, 677–681. [[CrossRef](#)]
6. Rastogi, S.; Liberles, D.A. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol. Biol.* **2005**, *5*, 28. [[CrossRef](#)]
7. Qiu, Y.; Tay, Y.V.; Ruan, Y.; Adams, K.L. Divergence of duplicated genes by repeated partitioning of splice forms and subcellular localization. *New Phytol.* **2020**, *225*, 1011–1022. [[CrossRef](#)]
8. Zhong, C.X.; Marshall, J.B.; Topp, C.; Mroczek, R.; Kato, A.; Nagaki, K.; Birchler, J.A.; Jiang, J.M.; Dawe, K.R. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* **2002**, *14*, 2825–2836. [[CrossRef](#)]
9. Nagaki, K.; Cheng, Z.K.; Yang, S.O.; Talbert, P.B.; Kim, M.; Jones, K.M.; Henikoff, S.; Buell, C.R.; Jiang, J.M. Sequencing of a rice centromere uncovers active genes. *Nat Genet.* **2004**, *36*, 138–145. [[CrossRef](#)]
10. Sanei, M.; Pickering, R.; Kumke, K.; Nasuda, S.; Houben, A. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proc. Nat. Acad. Sci. USA* **2011**, *108*, 498–505. [[CrossRef](#)]
11. Yuan, J.; Guo, X.; Hu, J.; Lv, Z.; Han, F. Characterization of two CENH3 genes and their roles in wheat evolution. *New Phytol.* **2015**, *206*, 839–851. [[CrossRef](#)]
12. Evtushenko, E.V.; Elisafenko, E.A.; Gatzkaya, S.S.; Lipikhina, Y.A.; Houben, A.; Vershinin, A.V. Conserved molecular structure of the centromeric histone CENH3 in *Secale* and its phylogenetic relationships. *Sci. Rep.* **2017**, *7*, 17628. [[CrossRef](#)]
13. Maheshwari, S.; Tan, E.H.; West, A.; Franklin, F.C.; Comai, L.; Chan, S.W. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet.* **2015**, *11*, e1004970. [[CrossRef](#)] [[PubMed](#)]
14. Black, B.E.; Foltz, D.R.; Chakravarthy, S.; Luger, K.; Woods, V.L.; Cleveland, D.W. Structural determinants for generating centromeric chromatin. *Nature* **2004**, *430*, 578–582. [[CrossRef](#)]
15. Elisafenko, E.A.; Evtushenko, E.V.; Vershinin, A.V. The Origin and Evolution of a Two-Component System of Paralogous Genes Encoding the Centromeric Histone CENH3 in Cereals. *Res. Sq.* **2021**. Available online: <https://www.researchsquare.com/article/rs-409648/v1> (accessed on 20 March 2021). [[CrossRef](#)]
16. Ishii, T.; Karimi-Ashtiyani, R.; Banaei-Moghaddam, A.M.; Schubert, V.; Fuchs, J.; Houben, A. The differential loading of two barley CENH3 variants into distinct centromeric substructures is cell type- and development-specific. *Chromosome Res.* **2015**, *23*, 277–284. [[CrossRef](#)] [[PubMed](#)]
17. Rabanus-Wallace, M.T.; Hackauf, B.; Mascher, M.; Lux, T.; Wicker, T.; Gundlach, H.; Baez, M.; Houben, A.; Mayer, K.F.X.; Guo, L.; et al. Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nat. Genet.* **2021**, *53*, 564–573. [[CrossRef](#)]
18. Li, G.; Wang, L.; Yang, J.; He, H.; Jin, H.; Li, X.; Ren, T.; Ren, Z.; Li, F.; Han, X.; et al. A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nat. Genet.* **2021**, *53*, 574–584. [[CrossRef](#)]
19. Neumann, P.; Pavlíková, Z.; Koblížková, A.; Fuková, I.; Jedličková, V.; Novák, P.; Macas, J. Centromeres off the hook: Massive changes in centromere size and structure following duplication of *CenH3* gene in *Fabaeae* species. *Mol. Biol. Evol.* **2015**, *32*, 1862–1879. [[CrossRef](#)]
20. Finseth, F.R.; Dong, Y.; Saunders, A.; Fishman, L. Duplication and adaptive evolution of a key centromeric protein in *Mimulus*, a genus with female meiotic drive. *Mol. Biol. Evol.* **2015**, *32*, 2694–2706. [[CrossRef](#)]
21. Kawabe, A.; Nasuda, S.; Charlesworth, D. Duplication of centromeric histone H3 (HTR12) gene in *Arabidopsis halleri* and *A. lyrata*, plant species with multiple centromeric satellite sequences. *Genetics* **2006**, *174*, 2021–2032. [[CrossRef](#)] [[PubMed](#)]

22. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [[CrossRef](#)] [[PubMed](#)]
23. Wicker, T.; Gundlach, H.; Schulman, A.H. The Repetitive Landscape of the Barley Genome. In *The Barley Genome*; Stein, N., Muehlbauer, J.G., Eds.; Springer: Cham, Switzerland, 2018; pp. 123–138.
24. Hernandez-Garcia, C.M.; Finer, J.J. Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* **2014**, *217–218*, 109–119.
25. Haberer, V.; Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 621–637. [[CrossRef](#)]
26. Roy, A.L.; Singer, D.S. Core promoters in transcription: Old problem, new insights. *Trends Biochem. Sci.* **2015**, *40*, 165–171. [[CrossRef](#)]
27. Shahmuradov, I.A.; Umarov, R.K.; Solovyev, V.V. TSSRlant: A new tool for prediction of plant Pol II promoters. *Nucl. Acids Res.* **2017**, *45*, e65.
28. Eulgem, T.; Rushton, P.J.; Somssich, I.E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **2000**, *5*, 199–206. [[CrossRef](#)]
29. Muller, F.; Tora, L. Chromatin and DNA sequences in defining promoters for transcription initiation. *Biochim. Biophys. Acta* **2014**, *1839*, 118–128. [[CrossRef](#)]
30. Shen, Q.J.; Casaretto, J.A.; Zhang, P.; Ho, T.-H.D. Functional definition of ABA-response complexes: The promoter units necessary and sufficient for ABA induction of gene expression in barley (*Hordeum vulgare* L.). *Plant Mol. Biol.* **2004**, *54*, 111–124. [[CrossRef](#)]
31. Prieto, P.; Santos, A.P.; Moore, G.; Shaw, P. Chromosomes associate premeiotically and in xylem vessel cells via their telomeres and centromeres in diploid rice (*Oryza sativa*). *Chromosoma* **2004**, *112*, 300–307. [[CrossRef](#)]
32. Luger, K.; Mader, A.W.; Richmond, R.K.; Sargent, D.F.; Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature* **1997**, *389*, 251–260. [[CrossRef](#)] [[PubMed](#)]
33. Vershinin, A.V.; Heslop-Harrison, J.S. Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. *Plant Mol. Biol.* **1998**, *36*, 149–161. [[CrossRef](#)] [[PubMed](#)]
34. Assis, R.; Bachtrog, D. Neofunctionalization of young duplicate genes in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17409–17414. [[CrossRef](#)] [[PubMed](#)]
35. Assis, R.; Bachtrog, D. Rapid divergence and diversification of mammalian duplicate gene function. *BMC Evol. Biol.* **2015**, *15*, 138. [[CrossRef](#)] [[PubMed](#)]
36. Karimi-Ashtiyani, R.; Ishii, T.; Niessen, M.; Stein, N.; Heckmann, S.; Gurushidze, M.; Banaei-Moghaddam, A.M.; Fuchs, J.; Schubert, V.; Koch, K.; et al. Point mutation impairs centromeric CENH3 loading and induces haploid plants. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 11211–11216. [[CrossRef](#)]
37. Kursel, L.E.; Malik, H.S. Recurrent gene duplication leads to diverse repertoires of centromeric histones in *Drosophila* species. *Mol. Biol. Evol.* **2017**, *34*, 1445–1462. [[CrossRef](#)]
38. Ishii, T.; Juranić, M.; Maheshwari, S.; Bustamante, F.O.; Vogt, M.; Salinas-Gamboa, R.; Dreissig, S.; Gursansky, N.; How, T.; Demidov, D.; et al. Unequal contribution of two paralogous CENH3 variants in cowpea centromere function. *Commun. Biol.* **2020**, *3*, 775. [[CrossRef](#)]
39. Bodor, D.L.; Mata, J.F.; Sergeev, M.; David, A.F.; Sallmian, K.J.; Panchenko, T.; Cleveland, D.W.; Black, B.E.; Shah, J.V.; Jansen, L.E. The quantitative architecture of centromeric chromatin. *eLife* **2014**, *3*, e02137. [[CrossRef](#)]
40. Shang, W.H.; Hori, T.; Martins, N.M.; Toyoda, A.; Misu, S.; Monma, N.; Hiratani, I.; Maeshima, K.; Ikeo, K.; Fujiyama, A.; et al. Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. *Dev. Cell* **2013**, *24*, 635–648. [[CrossRef](#)]
41. Neumann, P.; Schubert, V.; Fuková, I.; Manning, J.E.; Houben, A.; Macas, J. Epigenetic histone marks of extended meta-polycentric centromeres of *Lathyrus* and *Pisum* chromosomes. *Front. Plant Sci.* **2016**, *7*, 234. [[CrossRef](#)]
42. Blower, M.D.; Sullivan, D.A.; Karpen, G.H. Conserved organization of centromeric chromatin in flies and human. *Dev. Cell* **2002**, *2*, 319–330. [[CrossRef](#)]
43. Sullivan, B.A.; Karpen, G.H. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1076–1083. [[CrossRef](#)] [[PubMed](#)]
44. Gish, W. Advanced Biocomputing 1996–2019. Available online: <https://blast.advbcomp.com> (accessed on 24 August 2020).
45. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2013–2015. Available online: <http://www.repeatmasker.org> (accessed on 5 February 2014).
46. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)] [[PubMed](#)]
47. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
48. Solovyev, V.V.; Shahmuradov, I.A.; Salamov, A.A. Identification of Promoter Regions and Regulatory Sites. In *Computational Biology of Transcription Factor Binding*; Part of the Methods in Molecular Biology Book Series; Ladunga, I., Ed.; Springer: New York, NY, USA; Dordrecht, The Netherlands; Heidelberg, Germany; London, UK, 2010; Volume 674, pp. 57–83.
49. Shahmuradov, I.A.; Solovyev, V.V. Nsite, NsiteH and NsiteM computer tools for studying transcription regulatory elements. *Bioinformatics* **2015**, *31*, 3544–3545. [[CrossRef](#)]

50. Zadoks, J.C.; Chang, T.T.; Konzak, C.F. A decimal code for the growth stages of cereals. *Weed Res.* **1974**, *14*, 415–421. [[CrossRef](#)]
51. Paolacci, A.R.; Tanzarella, O.A.; Porceddu, E.; Ciaffi, M. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Mol. Biol.* **2009**, *10*, 11. [[CrossRef](#)]
52. Larionov, A.; Krause, A.; Miller, W. A standard curve based method for relative real time PCR data processing. *BMC Bioinform.* **2005**, *6*, 62. [[CrossRef](#)]
53. Wu, D.; Ruban, A.; Fuchs, J.; Macas, J.; Novák, P.; Vaio, M.; Zhou, Y.; Houben, A. Nondisjunction and unequal spindle organization accompany the drive of *Aegilops speltoides* B chromosomes. *New Phytol.* **2019**, *223*, 1340–1352. [[CrossRef](#)]
54. Jasencakova, Z.; Meister, A.; Walter, J.; Turner, B.M.; Schubert, I. Histone H4 acetylation of euchromatin and heterochromatin is cell cycle dependent and correlated with replication rather than with transcription. *Plant Cell* **2000**, *12*, 2087–2100. [[CrossRef](#)]
55. Loureiro, J.; Rodriguez, E.; Doležel, J.; Santos, C. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. *Ann. Bot.* **2006**, *98*, 679–689. [[CrossRef](#)] [[PubMed](#)]
56. Vaskova, E.A.; Dementyeva, E.V.; Shevchenko, A.I.; Pavlova, S.V.; Grigor'eva, E.V.; Zhelezova, A.I.; VandeBerg, J.L.; Zakian, S.M. Dynamics of the Two Heterochromatin Types during Imprinted X Chromosome Inactivation in Vole *Microtus levis*. *PLoS ONE* **2014**, *9*, e88256.
57. Houben, A.; Schroeder-Reiter, E.; Nagaki, K.; Nasuda, S.; Wanner, G.; Murata, M.; Endo, T.R. CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley. *Chromosoma* **2007**, *116*, 275–283. [[CrossRef](#)] [[PubMed](#)]
58. Hesse, S.; Zelkowski, M.; Mikhailova, E.I.; Keijzer, C.J.; Houben, A.; Schubert, V. Ultrastructure and dynamics of synaptonemal complex components during meiotic pairing and synapsis of standard (A) and accessory (B) rye chromosomes. *Front. Plant Sci.* **2019**, *10*, 773. [[CrossRef](#)] [[PubMed](#)]
59. Weisshart, K.; Fuchs, J.; Schubert, V. Structured Illumination Microscopy (SIM) and Photoactivated Localization Microscopy (PALM) to Analyze the Abundance and Distribution of RNA Polymerase II Molecules on Flow-sorted *Arabidopsis* Nuclei. *Bio-protocol* **2016**, *6*, e1725. [[CrossRef](#)]