Josip Juraj Strossmayer University of Osijek
University of Dubrovnik
Rudjer Boskovic Institute
University Postgraduate Interdisciplinary Doctoral Study
Molecular Biosciences

Evelin Despot Slade

Characterization and evolution of centromere in nematodes of the genus *Meloidogyne*

PhD thesis

Osijek, 2022

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište Josipa Jurja Strossmayera u Osijeku **Doktorska disertacije**
Sveučilište u Dubrovniku
Institut Ruđer Bošković
Poslijediplomski interdisciplinarni sveučilišni studij Molekularne bioznanosti

Znanstveno područje: Prirodne znanosti
Znanstveno polje: Biologija

Karakterizacija i evolucija centromere nematoda roda *Meloidogyne*

Evelin Despot Slade

Disertacija je izrađena u: Laboratoriju za strukturu i funkciju heterokromatina i Laboratoriju za nekodirajuće DNA Instituta Ruđer Bošković u Zagrebu

Mentor/i: dr. sc. Nevenka Meštrović Radan

Kratki sažetak doktorske disertacije:
Centromere su kromosomske regije odgovorne za pravilnu segregaciju kromosoma. Unatoč očuvanoj funkciji centromere, centromerni protein (CenH3) i pripadajuća centromerna DNA brzo evoluiraju. Međutim, naša je studija otkrila očuvanost αCenH3 varijante zajedno s konzerviranim motivom od 19-pb brzo evoluirajuće centromerne DNA kod nespolnih nematoda roda *Meloidoygne*. Sličan pristup primijenjen je i na kukcu *Tribolium castaneum* te pokazao metapolicentričnu organizaciju centromera s neobično izduženim centromernim područjima. Nanopore sekvenciranje dugih poteza je optimizirano za daljnje istraživanje centromernih nizova u genomima vrsta *Meloidogyne* i *Tribolium*.

Broj stranica: 229
Broj slika: 46
Broj tablica: 19
Broj literaturnih navoda: 284
Jezik izvornika: engleski
Ključne riječi: centromera, CenH3, centromerna DNA, nematode, *Meloidogyne*, metapolicentromera, kukac, *Tribolium castaneum*, satelitska DNA, Nanopore sekvenciranje

Datum obrane: 26. siječnja 2022.

Stručno povjerenstvo za obranu:
1. dr. sc. Đurđica Ugarković
2. dr. sc. Ivica Rubelj
3. prof.dr.sc. Ivana Majić
4. (zamjena) dr. sc. Snježana Mihaljević

Disertacija je pohranjena u: Nacionalnoj i sveučilišnoj knjižnici Zagreb, Ul. Hrvatske bratske zajednice 4, Zagreb;Gradskoj i sveučilišnoj knjižnici Osijek, Europska avenija 24, Osijek; Sveučilištu Josipa Jurja Strossmayera u Osijeku, Trg sv. Trojstva 3, Osijek

BASIC DOCUMENTATION CARD

Josip Juraj Strossmayer University of Osijek **PhD thesis**
University of Dubrovnik
Ruđer Bošković Institute
University Postgraduate Interdisciplinary Doctoral Study of Molecular biosciences

Scientific Area: Natural sciences
Scientific Field: Biology

Characterization and evolution of centromere in nematodes of the genus *Meloidogyne*

Evelin Despot Slade

Thesis performed at: Laboratory for structure and function of heterochromatin and Laboratory for non-coding DNA, Ruđer Bošković Institute, Zagreb

Supervisor/s: dr. sc. Nevenka Meštrović Radan

Short abstract:
Centromeres are chromosomal regions responsible for the faithful segregation of chromosomes. In contrast to conserved function, the centromere protein (CenH3) and associated repetitive DNA evolve rapidly. However, our study revealed preservation of the αCenH3 variant along with conserved 19-bp box sequence in rapidly evolved centromeric DNA in asexual *Meloidoygne* nematodes. A similar approach was applied to the beetle *Tribolium castaneum* and showed a metapolycentric centromere organization with extended centromere regions. Long-range Nanopore sequencing is optimized for further study of centromeric arrays in the genomes of *Meloidogyne* and *Tribolium* species.

Number of pages: 229
Number of figures: 46
Number of tables: 19
Number of references: 284
Original in: english
Key words: centromere, CenH3, centromeric DNA, nematode, *Meloidogyne*, metapolycentromere, beetle, *Tribolium castaneum*, satellite DNA, Nanopore sequencing

Date of the thesis defense: January 26th 2022.

Reviewers:
1. dr. sc. Đurđica Ugarković
2. dr. sc. Ivica Rubelj
3. prof.dr.sc. Ivana Majić
4.(substitute) dr. sc. Snježana Mihaljević

Thesis deposited in: National and University Library in Zagreb, Ul. Hrvatske bratske zajednice 4, Zagreb; City and University Library of Osijek, Europska avenija 24, Osijek; Josip Juraj Strossmayer University of Osijek, Trg sv. Trojstva 3, Osijek

This doctoral thesis was made in the Laboratory for structure and function of heterochromatin and in Laboratory for non-coding DNA at the Ruđer Bošković Institute in Zagreb, under the supervision of Nevenka Meštrović Radan, PhD. Postgraduate studies were done within the program of University Postgraduate Interdisciplinary Doctoral Study of Molecular biosciences in Osijek. The research has been supported by grant “Invertebrate centromere genomics” (IP-2014-09-3183) and “(Epi)genomics and transcriptomics of euchromatic satellite DNAs in embryogenesis and development” (IP-2019-04-4910) from Croatian Science foundation.

Acknowledgments

First and foremost, I would like to thank Nevenka, who has not only been my mentor, but also a great teacher, friend, and person I could only wish to share this part of my journey with. I have experienced so much joy throughout our analyses and discoveries. I am deeply grateful to her for selflessly sharing her knowledge with me and training me to be a scientist in every sense of the word. She showed understanding of all aspects of my life and pointed me to the true meaning of sunny days.

My special thanks goes to Brankica, who stood by me every step of the way and offered me immense help and thoughtful advices. Her optimism and inspirational stories lifted me up even in the hardest moments. The special support came from the way she showed me that everything can be done by paying attention to even the smallest detail.

I would also like to thank Miro for the warm welcome and the opportunity to work on such a fascinating research topic. He has encouraged me to continue to strive and become stronger.

I would like to extend my sincere thanks to all current and former lab members for all the wonderful moments that I will never forget.

I am deeply grateful to my dear parents for their unconditional support and for always being there for me. I am proud that with their love and understanding I have become the person I am today. My heartfelt appreciation goes out to my mother for every extra book she found for me.

I want to thank K. for always being my true partner. His serenity has allowed me to keep my high spirit and pursue my dreams.

Table of Contents

1. Introduction

1.1. Centromeres

Cell divisions, mitosis and meiosis are essential processes in cell growth, development, regeneration and reproduction of all living organisms. These two types of cell divisions differ; mitosis is a division in somatic cells, while meiosis is a more complex process leading to the formation of haploid gametes that function in species reproduction and ultimately impact evolution of a population or species. These processes ensure the faithful propagation of genetic material to daughter cells and subsequent generations of cells. One of the key players in the exact process of chromosome segregation in mitosis and meiosis is centromere. Centromere generally appears as primary constriction of mitotic chromosomes, first identified cytologically by Walther Flemming in 1882. In the last 40 years, investigation of centromeres in a different eukaryote has revealed that centromeres represent specific chromosomal regions that recruit components of the kinetochore complex and enable accurate chromosome segregation during mitosis and meiosis (Figure 1). High-fidelity segregation is vital for all eukaryotic organisms and centromeric defects lead to chromosome breakage and aneuploidy.

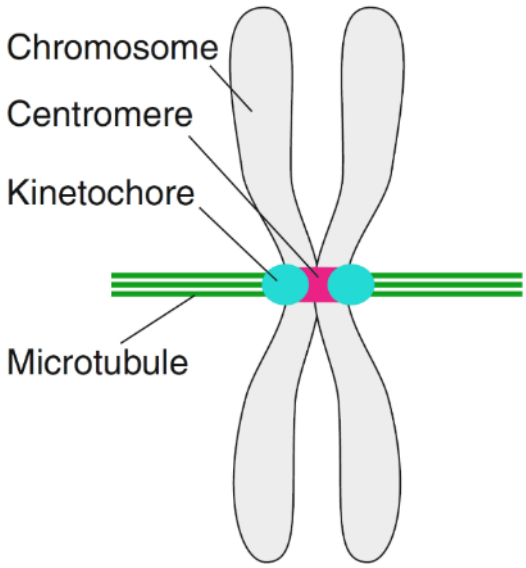


Figure 1. Centromere localisation. Scheme of a typical metacentric vertebrate chromosome with centromere in pink as primary constriction site on which kinetochore (blue) assembles and attachecs to microtubules ensuring proper chromosome segregation (adjusted from Hara & Fukagawa, 2017).

1.1.1. Centromere structure

Centromere is a complex protein/DNA structure responsible for holding two sister chromatids together and accurate segregation during cell division (Figure 1). Highly compact chromatin found at centromeres is called heterochromatin and unlike euchromatin holds very few genes. It is considered that this condensly organized structure is more befitted for transmitting mechanical force and adjusting topology for spindle attachment than open euchromatin (Bloom, 2014). In general, the fundamental subunit of chromatin, the nucleosome, is composed of 147 nucleotides wrap around histone proteins, where two copies of each histone types (H2A, H2B, H3 and H4) form octameric core (Figure 2A). Histones contain two common structural domains: histone fold domain (HFD) and N-terminal tail. The HFD is formed by three α-helices linked with two loops (Figure 2B) that enable specific positioning and heterodimeric interaction between histones on one side and hydrogen bonds with DNA on the other side (Mariño-Ramírez et al., 2005). N-terminal tail extends from DNA-histone core and is subjected to an enormous number of post-translation modification (PTM) where they control site-specific functionality (Figure 2C) consequently altering biological processes (Peterson & Laniel, 2004). This epigenetic signature is known as “histone code” and most observable can be examples of silent heterochromatin which is widely marked by the presence of methylation of lysine 9 of histone H3 (H3K9me), whereas active euchromatin is linked with dimethylation of lysine 4 of histone H3 (H3K4me2).

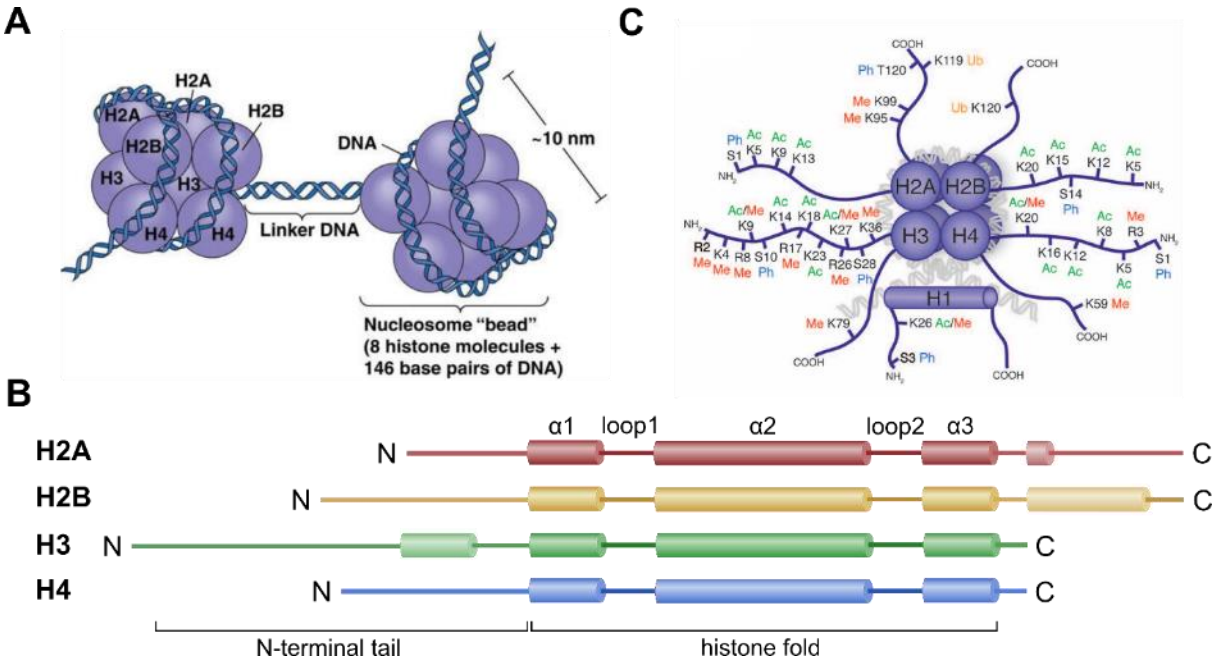


Figure 2. Organization of a nucleosome. A Nucleosome schematics with DNA helix wrapped around histone octamer with 146 bp tightly associated with histon core and linker DNA in between adjacent nucleosome "beads" (from Caputi et al., 2017). **B** Structural organization of core histone proteins with marked N-terminal tail and histone fold domain containing α-helices and loops enabeling proper folding and interaction (according to Alberts et al., 2008). **C** Histone post-translational modifications with highlighted covalent modificationos on N- and C-terminal tails of specific amino residues (Me-methylation, Ac-acetylation, Ub-ubiquitination, Ph-phosphorylation) (from Tollervey & Lunyak, 2012).

Centromeric nucleosomes differs from the conventional by the presence of centromere-specific H3 histone variant called CenH3 that replaces canonical H3 in centromere (Figure 3A). Unlike canonical H3 variants which is highly conserved protein between different lineages, CenH3 rapidly evolve. Despite the conserved role of CenH3 in maintaining centromere integrity, CenH3 demonstrates accelerated evolution which is especially pronounced at its N-terminal tail and loop 1 of the histone-fold domain (HFD) (Malik and Henikoff 2001; Talbert et al. 2004) (Figure 3B).

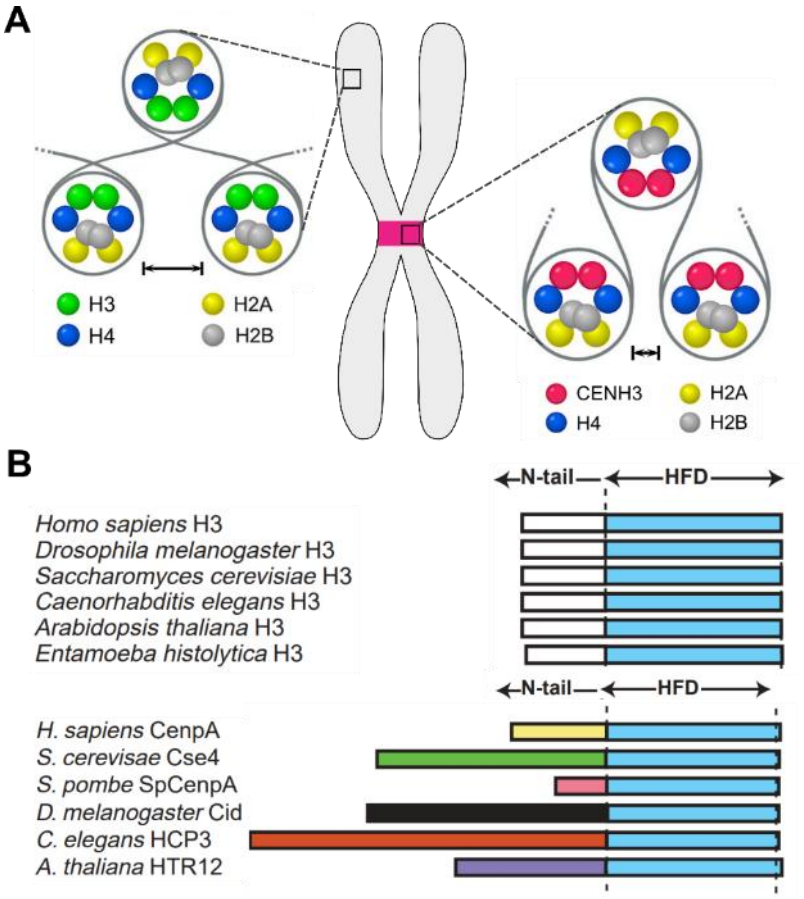


Figure 3. Characteristics of centromere structure and organization. A Difference between H3-containing nucleosomal array from chromosomal arm and CenH3 arrays localized in centromere (adopted from Panchenko et al., 2011). **B** Comparison between conservation of canonical H3 in species

of different lineages and variability of N-terminal tails between CenH3 histone variants (adopted from Malik & Henikoff, 2003).

Specifically, there are two changes within HFD domain that are found to be specific for CenH3 proteins; longer loop 1 region and a lack of conserved glutamine residue in the α1 helix (Malik & Henikoff, 2003). Furthermore, centromere positioning is regulated additionally by the association with the underlying centromeric DNA (cenDNA) that is wrapped around nucleosomes containing CenH3. CenDNA is usually composed of highly repetitive DNA such as satellite DNA (satDNA) and transposable elements (TEs) (reviewed in Biscotti et al., 2015). SatDNAs are composed of highly repetitive short sequences or monomers that repeat tandemly to form large arrays that can be found at more or less same positions within the genome. They are usually AT rich and monomers have been found to be 100-500 bp long in most animals and plants. Centromeric repeats also evolve rapidly, and significantly differ between closely related species (Plohl et al., 2014). Transposable elements, on the other hand, often move within genome with ability to insert at novel locations and thereby shape surrounding coding and non-coding landscape. TEs come in a variety of forms and are divided into two major classes based on their transposition intermediates. Class I elements have an RNA intermediate which is reversely transcribed into DNA and integrated in the genome while class II excises and moves itself to a new location. Based on protein encoding domains and other structural characteristics they are further separated into subclasses and families (reviewed in Wells & Feschotte, 2020)

1.1.2. Centromere organization

Centromeres differ in their sizes and sequence composition across different species lineages. In humans, CenH3 variant is called CENP-A and its cenDNA has been found out to be comprised of α-satellite DNA. Centromeres can be as small as 125 bp as seen in *Saccharomyces cervisiae* (Cottarel et al., 1989), contain specific organized domains, mixed with several repetitive elements but also have high order repeats (HORs) and be megabases long as found in plants and humans as depicted in Figure 4 (Muller et al., 2019).

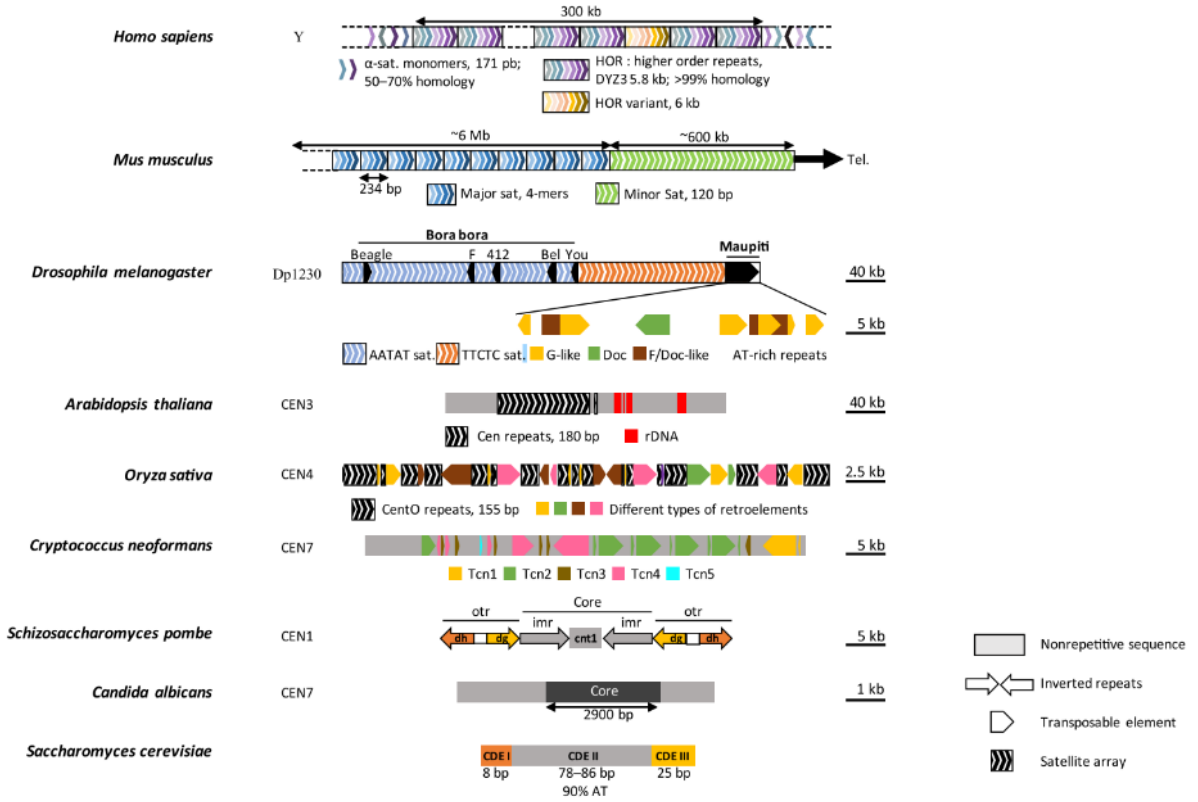


Figure 4. Organization of centromere sequences in Eukaryotes. Typical centromere organization is shown for each species with size comparison and organizational unit indicated. Image adopted from Muller et al., 2019.

Centromere of a *S. cervisiae* is referred as point centromere due to extremely short sequence creating through kinetochore only attachment site for single microtubule. A vast number of organisms contain regional epigenetically determined centromeres that are much larger in sizes. They can range from several kilobases as in *Candida albicans* (4-18 kb) to hundreds kilobases in flys (420 kb) all up to megabase long as seen in maize (0,3-3 Mb) and humans (0,5-5 Mb) (Wong et al., 2020).

Centromere in *Drosophila melanogaster* is formed on islands of complex DNA sequences found enriched in retroelements flanked by large arrays of satellite repeats (Chang et al., 2019). Interestingly, even though centromeres differ in sizes and arrangement, one *G2/Jockey-3* retroelement is the prevalent sequence in CenH3 chromatin and the only one shared among all centromeres and even in sister species *D. simulans*. *Arabidopsis thaliana* has megabase centromeres consisting mainly of 180 bp satellite repeats flanked by various types of transposons and 5S rDNA (Kumekawa et al., 2001). Centromeric region of *Mus Musculus* is represented by 120 bp satDNA known as minor satellite (MiSat) wherease pericentromeric region is characterized by even higher abundant 234 bp major satellite (MaSat) and are found on all telocentric

mouse chromosomes (Komissarov et al., 2011). Human centromeres are determined by the presence of CENP-A (human CenH3) and alpha satellite, 171 bp long monomer organized into HORs from 2 to 35 monomers that are than making long tandem arrays up to 5 Mb in length. In contrast to HOR, monomeric arrays without higher organization of alpha satDNA have been found among other satDNAs in pericentromere regions. Alpha satDNA monomers show 20-50% variation among same HOR array and provide opportunity to disclose fine-scale mapping and investigation of evolutionary homogenization processes (Hartley & O'Neill, 2019).

1.1.3. Centromere function

Mitosis and meiosis are two types of cell divisions and important events in the life of a cell where the centromere plays a crucial role. Somatic cells undergo mitosis in which the main role is cell duplication while maintaining the existing ploidy. The second type of cell division is meiosis, which takes place in the germ cells, where the number of chromosomes gets reduced, so that the resulting cells have a haploid number of chromosomes. The ultimate goal of a centromere is the proper segregation of sister chromatids and chromosomes in mitosis and meiosis and the pairing of homologous chromosomes and recombination in meiosis (Figure 5).

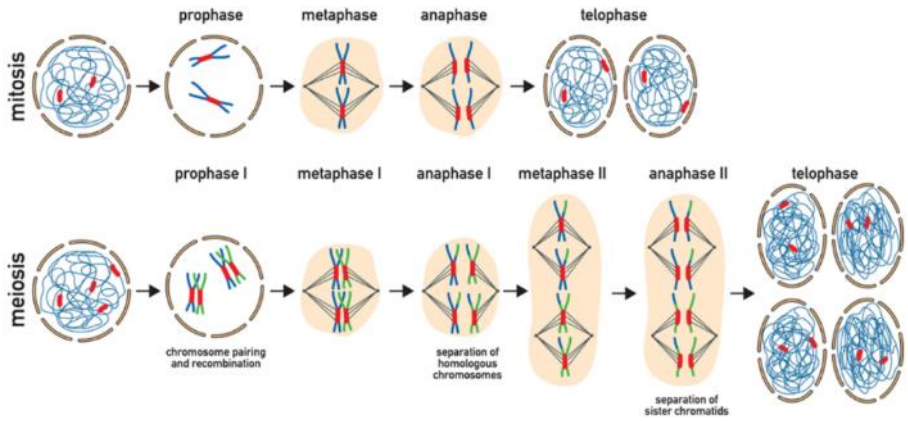


Figure 5. Overview of mitotic and mitotic stages. Centromeres are depicted in red and their role in equal and faithful segregation in mitosis is highlighted. In meiosis they have an additional role in chromosome pairing and enabling DNA recombination (from Prosée et al., 2020).

One of the key prerequisite for cell growth and transmission of genetic material is chromosome segregation which is mainly dependent on proper recruitment of the kinetochore. Kinetochore is a giant multilayered protein structure build on centromeric

heterochromatin region of the chromosome. It has attachment points for microtubules and allows their polymerization and depolymerization in order to control movement of the chromosomes on the dividing spindle.

Protein complex cohesin holds two sister chromatids in close proximity by forming ring-like structure in DNA replication phase. Later in anaphase centromeric cohesion persists unlike arm cohesion suggesting robust regulation at centromeres in order to resist pulling forces in mitosis. It has been speculated the role of kinetochore and pericentromeric heterochromatin in recruiting cohesin specifically to centromere (reviewed in Watanabe, 2005). Furthermore, centromere identity is precisely regulated by deposition of new CenH3s as coordinated activity of several assembly factors but also by presence of permissive marks that enable proper CenH3 localization (reviewed in Mckinley & Cheeseman, 2016). Thus, centromere function is achieved and properly maintained by CenH3 containing nucleosomes positioned on specific underlying DNA and unique chromatin marks ensuring foundation for building multiprotein kinetochore complex that attaches to microtubules of mitotic and meiotic spindles. But to date, there is a dilemma as to how the epigenetically defined centromere and associated DNA participate in the construction of centromere identity and function.

1.1.4. Centromere evolution

A striking centromere feature is its conserved role as opposed to extremely divergent underlying DNA sequences. Furthermore, CenH3 differs in its protein sequence which is primarily seen at the protruding N-terminus part of the protein facing outward of the nucleosome even in fairly closely related species. Still, investigation of CenH3 proteins remain most reliable marker and starting point for centromere research.

One of the main theories that explains the centromere paradox (Henikoff et al., 2001) i.e. the conserved function of the centromere and the rapid evolution of CenH3 and centromeric DNA is the centromere drive (Figure 6A). This model explains the diversity of eukaryotic centromeric DNA and CenH3 proteins as a consequence of the conflict between the rapid evolution of centromeric DNA repeats and CenH3. One of the main reasons is the asymmetry of meiosis in females, where due to rapidly evolving repetitive sequences that tend to spread, unequal deposition of CenH3 and kinetochore proteins occurs, and stronger centromeres are oriented towards the pole from which the oocyte is formed and thus transmitted. In the process of repeat

spreading centromeric protein will occupy newly formed similar sequences in neighboring areas and these centromeres will be more associated with the dividing spindle and not eliminated via polar bodies (Figure 6B).

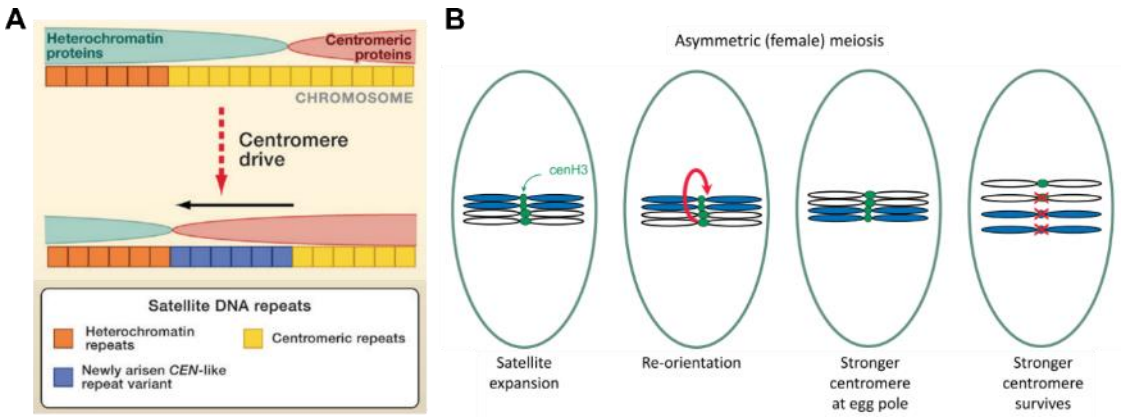


Figure 6. Proposed driving mechanism for centromere evolution. A Spreading of centromeric proteins on newly arisen centromeric-like repeat variants due to centromere drive which can account for rapid evolution of centromeric DNA and CenH3 (adopted from (Malik & Henikoff, 2009). **B** Scheme of influence of satellite expansion in asymmetric female meiosis on formation of stronger centromere that re-orients toward the egg pole and gets transmitted to the next generation (adopted form Talbert & Henikoff, 2020).

Evolutionary, positive selection of CenH3 histone as response to fast evolving cenDNA refers to preferential survival of individuals carrying positive genetic change increasing reproductive success. On the contrary, certain genotypes can cause deleterious impact on species fitness and will be purified through elimination process of negative selection. Thus, in many species undergoing asymmetric meiosis, CenH3 has been observed to evolve under positive selection to compensate for the negative effects of rapid changes in centromeric DNA.

Most diploid genomes possess a single copy of the CenH3 gene while in polyploid species (so far most commonly in plants) multiple copies of CenH3 have been observed that may have different expression profiles and be tissue-specific. However, recent studies on diploid *Drososphila* species also indicate the existence of more Cid (CenH3 gene) copies in animal species as a result of at least four duplication events during its evolution (Kursel & Malik, 2017). These genes encode functional CenH3 paralogs that co-localize on the centromere during cell division with intriguing function speculation for these variants.

1.1.5. Centromere types

Regarding centromere organization chromosomes can be roughly divided to mono- and holocentric with meta-polycentric chromosome as in between state containing characteristics of both groups (Figure 7). Monocentric chromosomes contain a single narrow constriction with one centromere per chromosome. Holocentromeres are on the other hand characterized by their diffused distribution along the length of the chromosome and their identity is still not fully elucidated. Meta-polycentromere is a fairly new type of centromere firstly described in peas that has multiple centromere domains each presented as microtubule binding site but still distinct from each other observed as long primary constriction site (Neumann et al., 2012).

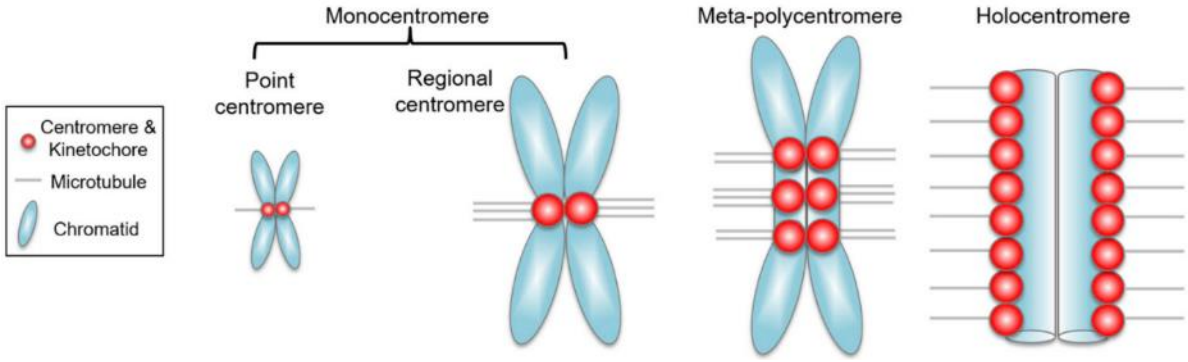


Figure 7. Centromere types. Schematics of centromere architectures (from Wong et al., 2020).

1.1.5.1. Monocentric chromosomes

Monocentric chromosomes are widely found among eukaryotes. Based on its positioning along the chromosome arms metaphase chromosomes can be divided to four types. Metacentric chromosomes have centromere located in the middle equally dividing chromosome length. In submetacentric chromosomes, centromere is visibly off-centered creating one shorter and one longer arm. Lastly, if centromere is located at the near end or at the end of the chromosome they are called acro- and telocentric chromosomes, respectively. Many monocentric species have their centromere described and they are often build on repetitive DNA sequences such as satDNA and transposons (reviewed in Plohl et al., 2014). Centromeres can be looked at based on which type of sequences are they comprised of and how are they distributed (Figure 8).

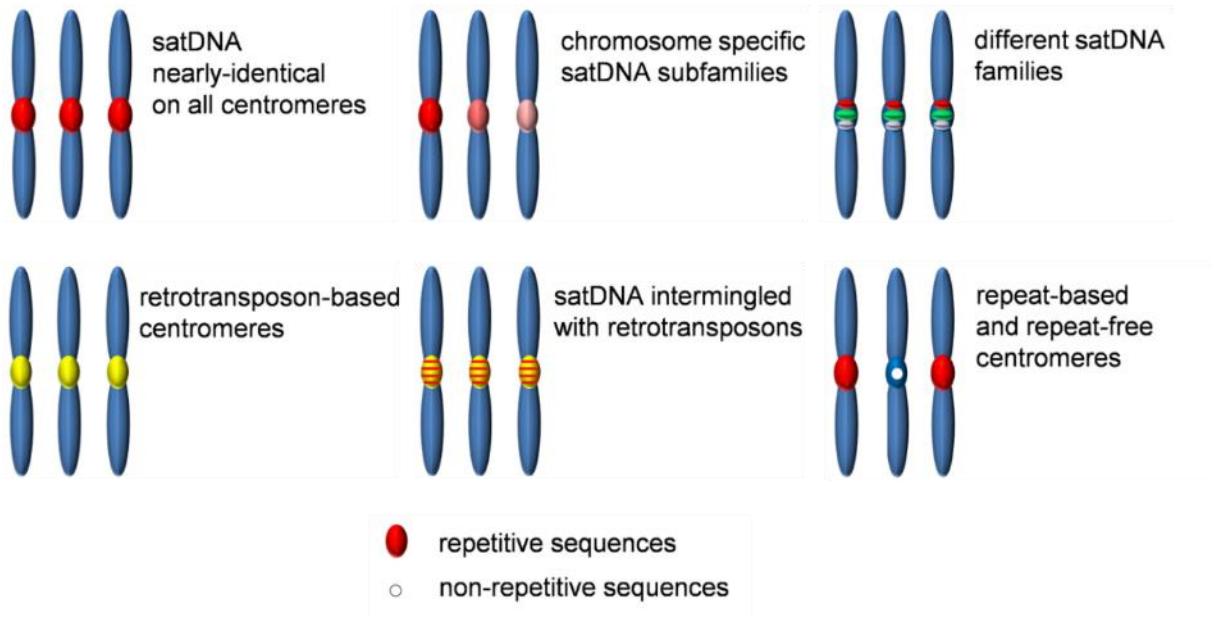


Figure 8. Monocentric chromosome repeat organization. Different centromere types with functional DNA sequences (adopted from Plohl et al., 2014).

Different satDNA families can be specific for certain chromosomes but also they can be mixed and form organizational patterns. If all chromosomes carry nearly identical satDNA on all centromeres it indicates high sequence homogenization and there are no chromosome-specific satDNA variants as seen in pericentromere and centromere regions of mouse major and minor satellite DNA which form distinct functional heterochromatin (Guenatri et al., 2004). Often, there are satellite subfamilies that characterize specific chromosomes as seen in humans, where alpha satellite HORs are found to be chromosome specific and distinguished from each other in variation of the sequence, order of monomers and overall size of HOR (McNulty et al., 2017). Sometimes centromere can be characterized with different satDNA families as found in pea where functional centromere domains are associated with repetitive DNA sequences belonging to 13 distinct families of satellite DNA and one retrotransposon family unevenly distributed between chromosomes (Neumann et al., 2012). Centromere can also be predominately build on retrotransposons as described in maize that also contains smaller amount of satellite CentC arrays (Wolfgruber et al., 2009). It has been found that retrotransposons can target centromeres and functionally replace tandem repeats that could potentially coordinate the evolution of centromeric DNA (Presting, 2018). Human pericentromere also contains insertions of unrelated satDNAs (gamma-satellite and SatIII) but also LINE elements (reviewed in Lopes et

al., 2021). One of the most extreme examples are species with several chromosomes whose centromeres lack in repetitive sequences as found in *Equus* genus (Piras et al., 2010) and chicken (Shang et al., 2010) where centromeres of some chromosomes are devoided of any tandem repeat. Still, extreme diversity of cenDNA even among closely related species offers unique platform for dynamics and evolution of repetitive DNA.

1.1.5.2. Holocentric chromosomes

Holocentric chromosomes have been firstly described in 1935 using cytology which was for very long most reliable way of assessment based on behavior in mitosis (reviewed in Mandrioli & Manicardi, 2020). Development of new technologies, especially chromatin manipulation coupled with sequencing enabled many studies on species that were very hard or impossible to discriminate based on chromosome appearances. So the rise in the number of recognized holocentric species today came as no surprise, as combination of cytological and molecular methods have shown that holocentricity has arisen at least 13 times independently; four times in plants and even nine times in animals (Melters et al., 2012). In these species kinetochore proteins bind along the entire length of the chromosome which are identified by the lack of primary constriction and during division they migrate in parallel to spindle poles. Furthermore, understanding of cenDNA in holocentromere is still limited and mostly comes from studies on vastly investigated model organism nematode *Caenorhabditis elegans* which is holocentric. Research has found out that its centromere is organized as dispersed but discretely localized point centromere that interestingly coincide with binding sites for transcription factors but without one centromere specific sequence (Steiner & Henikoff, 2014). In contrast, holocentromere of *Rhynchospora* plant showed association with interspersed arrays of specific satDNA family and retrotransposons (Marques et al., 2015). Unique pattern of CenH3 distribution is described in holocentric plant *Cuscuta europaea* with two expressed CenH3 variants deposited into one to three regions per chromosome with absence of CenH3 on the rest of the chromatin (Oliveira et al., 2020). Observations based on holocentric species analyzed so far, lead to the conclusion that in contrast to monocentromere, holocentromeres show greater flexibility in the organization at the chromosome level.

1.1.6. Medical aspects of centromere

Due to this extreme repetitive nature of centromeres is it believed that they have a complex DNA topology, forming mechanically interlocked molecular architecture (catenanes) and DNA loops. On one side, this probably offers compensation platform for enduring high degrees of stress generated during chromosome segregation but on the other hand are susceptible for DNA recombination machinery and replication pausing thus making them vulnerable (reviewed in Barra & Fachinetti, 2018). This events could result in repositioning of centromere to a non-centromeric DNA site, rise of dicentric chromosomes and aberrant expression of kinetochore proteins which have all been linked to human diseases (Felekkis & Voskarides, 2015). If it leads to loss or an addition of chromosome, consequence can be aneuploidy, a state of abnormal chromosome number, a common cause of several genetic disorders known to arise from errors in chromosome segregation. Also it has been linked to cancer cells which almost always display a high rate of chromosomal instability (CIN) that leads to aneuploidy (Sansregret & Swanton, 2017). Obscured by complex phenotypes it has been recently found that aneuploidic cancer cells are sensitive to mitotic checkpoint inhibition giving potential opportunity to selectively kill cancer cells (Cohen-Sharir et al., 2021). Rise of new centromeres on atypical genomic region known as neocentromeres has been first described in humans but have since been investigated in several species (reviewed in Scott & Sullivan, 2014). Even though determined in clinical phenotypes, neocentromeres do not necessary lead to disease but could offer an escape mechanism in situations where chromosomal damage would lead to cell death providing support to sequence-independent centromere determination (Fukagawa & Earnshaw, 2014). As they are still a place for kinetochore assembly and enable segregation of such chromosomes there are additional studies required to provide explanations of how they could be regulated and to what extent they contribute to diseased cell stated.

Unlike first beliefs, centromeres are not completely silent regions but are transcribed into long non-coding RNAs whose roles are still widely investigated with proposing roles for accurate centromeric functions (reviewed in Rošić & Erhardt, 2016). Centromere transcription is essential for proper differentiation and development but also miss regulated in cancer cells and offers through quantitative expression analysis potential for clinical targets as cancer biomarkers (Smurova & De Wulf, 2018).

Moreover, recent studies provided the evidence that (peri)centromeric satDNAs overexpression in epithelial tumor tissues contributes to malignant transformation by inducing chromosomal instability and mutations (Ting et al., 2011).

1.2. Model organism-nematode *Meloidogyne*

1.2.1. Characteristics and mode of reproduction

Species from the genus *Meloidogyne spp*. are plant obligate endoparasites and belong to root knot nematodes (RKNs). They reduce plant health and growth by causing changes in root physiology and consequentially present huge economic damages (Figure 6A). There are 98 species of Meloidogyne genus and are all obligate parasites on over 2500 species with growing number of new hosts found each year. The juvenile form (J2) enters the root where it further develops into the adult form, primarily in the sessile female (Figure 6B). Upon taking nutrients J2 induces dedifferentiation od root cells into multinucleate feeding cells. Nematode than goes through three molts (J3, J4 and adult) and becomes sedentary female form (Figure 6C). They further lay eggs that are secreted in egg sacks on the root surface. Males appear only in stressful conditions when they migrate out of the root but are not required for reproduction. Overview of life cycle and morphological differences of nematode stages is presented on Figure 6D.

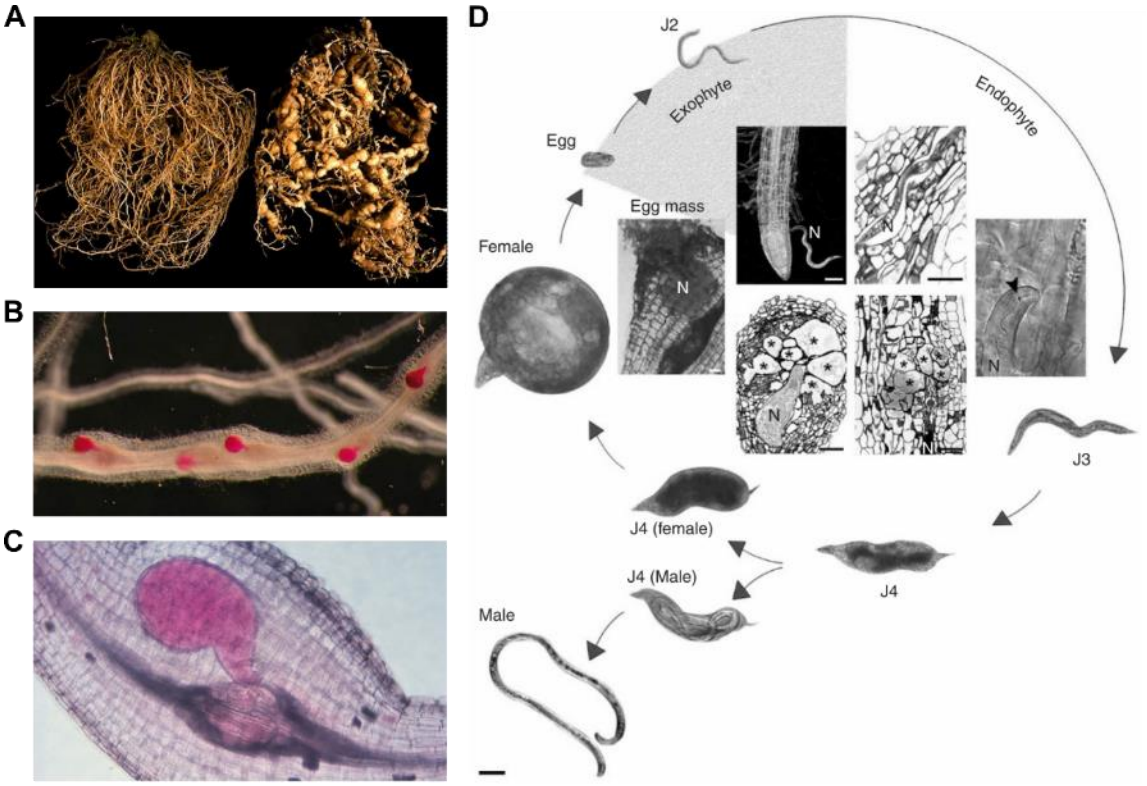


Figure 6. Characteristics of *Meloidogyne* root-knot nematodes. A Damages that species from the Meloidogyne genus cause on roots of a tomato. Left is a picture of a helthy root versus the infected root on the right with massive amounts of galled root system (Jonathan D. Eisenback, Virginia Polytechnic Institute and State University, Bugwood.org). **B** Females of *M. chitwoodi* inside bean roots stained in pink with fuchsin acid forming knots around each one (from Wesemael, 2007). **C** Female sessile stage in root causing the formation of "giant cells" as a feeding site (from Agrios, 2004). **D** Life cycle of a parasitic *M. incognita* nematode (N) with developmental stages depicted; eggs, juvenile stages 2-4 (J2, J3, J4), females and males. Juvenile with stylet connected with esophagus piercing plant cell wall and formation of feeding site is depicted on the pictures in the middle. Scale bar= 50 µm. (from Abad et al., 2008).

Interesting characteristics of RKNs is their diversity of modes of reproduction (reviewed in (P. Castagnone-Sereno, 2006). Only several species are amphimictic and reproduce by classical crossfertilization of male and female gametes. Most of the species reproduce by parthenogenesis, either meiotic (automictic) or mitotic (apomictic), with some of them using two modes of reproduction depending on the presence of males.

1.2.2. Parthenogenesis and evolution of species

Parthenogenesis is asexual reproduction where individual develops from unfertilized egg. It is common for invertebrates and found among some vertebrates. In most animal cases it is referred to female producing parthenogenesis without genetic contribution from males and without regular undergoing through meiosis. Parthenogenetic RKN are diploids, triploids and rare tetraploids with most population having unstable chromosome number due to processes such as aneuploidy, polysomy and structural rearrangement while amphimictic RKN species are only diploid. *M. incognita* is the most prevalent and investigated RKN species. Its populations are considered to be mostly hypotriploid (<3n or 3n-x) with 40-46 chromosomes (Triantaphyllou, 1981). Asexuals have a long-term evolutionary disadvantage due to lack of recombination but are still widespread and this expansion is therefore always a unique process presenting a great challenge in elucidating evolution of not only sex but eukaryots in general (in depth overview of the field in Schön et al., 2009). Recently there were several studies trying to resolve evolution of RKN species by using standardize phylogenomics analysis but also by sequencing and comparative genomics. It is speculated that divergent genome copies in *Meloidogyne incognita* group (MIG) of species have been a consequence of complex series of interspecies hybridization events (Blanc-Mathieu et al., 2017). Such additive hybridizations between parental taxa also explains ploidy levels with noncrossover recombination as additional force changing the genomic landscape of these species (Szitenberg et al., 2017).

1.2.3. Satellite DNAs in *Meloidogyne*

Dominant satDNAs have been characterized in various *Meloidogyne* species. The data revealed different A+T rich satDNA families with 170–300 bp monomers and usually with moderate genome abundance (eg. Philippe Castagnone-Sereno et al., 1998; Meštrović et al., 2013). It has also been shown that they evolved according to the library concept where occurrence of satDNAs profile in related species is a result of amplifications and/or contractions of different satDNAs from a library (Meštrović et al., 1998). In addition, the study of the satDNA library of the three related satDNAs differently amplified in Meloidogyne species indicates selection as a limiting factor in formation and persistence of satDNAs in the library (Meštrović et al., 2006b). The

distribution profile of six different satDNAs, in terms of their presence/absence in related Meloidogyne genomes, has been shown to be informative in phylogenetic studies of these species (Meštrović et al., 2009).

1.3. Model organism-beetle *Tribolium*

1.3.1. Main characteristics and life cycle

The *Tribolium* genus, also known as flour beetles, is comprised of 36 cosmopolitan species. Their original habitat was likely beneath the tree bark in rotting wood but today these small dark bodied insects are global pests of stored food products. As secondary pests of damaged grain and dust they cause damage by feeding but also by contamination with shed skin, secreted liquids and dead bodies resulting in large agricultural losses. Furthermore, *Tribolium* are of foremost interest as a representative of the *Coleoptera* order which is the most species rich taxon on the earth (Hunt et al., 2007). *Tribolium castaneum* (red flour beetle) emerged as an excellent model organism (reviewed in Pointer et al., 2021) due to ease of culture, short generation time and efficiency in genetic manipulations. One female can lay up to 400 eggs with larva emerging after several days. In this active stage they will feed and grow for few months depending on the temperature, humidity and food availability. Eventually, after undergoing molting they transform to inactive pupae. In the next week huge internal changes will take place before adults appear. Due to this complete metamorphosis it is thought that *Tribolium* development is more representative of other insects than *Drosophila*.

1.3.2. Satellite DNAs and (peri)centromere

T. castaneum possesses abundant, species specific satellite DNA (TCAST) whose predicted amount ranges from 17% up to 35% (Ugarković et al., 1996; Feliciello et al., 2011). It is evenly distributed in the regions of large blocks of (peri)centromeric heterochromatin of all 20 chromosomes. Similar observations of single (or two) highly abundant satDNAs with equivalent distribution pattern have been made for other closely related *Tribolium* species (Juan et al., 1993; Mravinac et al., 2004; Mravinac & Plohl, 2010). Additionally, *T. castaneum* has nine satDNA families making 4% of the

genome located primarily in euchromatin (Pavlek et al., 2015). Because TCAST makes only 0,3% of the assemble genome, most conclusions about centromere have been made based on *in situ* hybridizations of satDNAs and cytological experiments of highly stained heterochromatin in regions of chromosome constriction. Regardless of the existence of quantity discrepancies when standardized methods were applied, TCAST satellite still presents major candidate for applying next-generation of sequencing (NGS) in order to examine its competence as centromere specific repeat.

1.4. Nanopore sequencing in investigation of the centromere

It has passed 20 years since initial efforts form International Human Genome Sequencing Consortium to make a first leap towards complete assembly of a human genome (Lander et al., 2001) which is known to have more than 50% of repeated sequences including interspersed transposable elements but also huge region of mostly alpha satellite containing (peri)centromere. Even though they used the best available technique at the time for repeat-rich genome; hierarchical shotgun sequencing of large-insert clones, multi-megabase satellite arrays of centromere, ribosomal DNA arrays and huge regions with segmental duplications remained unresolved. First significant progress was made with the application of Nanopore sequencing where it was shown it holds the potential to bridge even large (>50 kb) scaffold gaps by spanning them with single long read (Jain, Koren, et al., 2018). This was mostly possible thanks to the development of third-generation of sequencing, also known as long-read sequencing that is starting to enhance and accelerate filed of genomic research. Among them, two have stand out as leaders in this field. Pacific Biosciences offers high consensus accuracy (>99,9%) with average read length of 15 kb and is limited by polymerase activity. On the other hand, Oxford Nanopore's technology is based on the detection of current change as a single-stranded DNA passes through an engineered protein Nanopore inserted into a synthetic polymer membrane. Due to this principle it bears no limitation in regard to read length (current record is >4 Mb) with a decent read accuracy (>95%) that can be further increased by higher coverages and read correction steps.

These unprecedented properties of Nanopore sequencing were ultimately applied for assembly of the human Y chromosome and for the first time encompass whole centromeric locus with array mean of 315 kb (Jain, Olsen, et al., 2018). Similar

approach was further applied for telomere-to-telomere assembly of X chromosome (Miga et al., 2020) and complete assembly of chromosome 8 (Logsdon et al., 2021). Furthermore, Nanopore sequencing also allows detection of structural variants, gene expression and methylation profiles with broad applications not only in human genomic but also in wide research areas such as metagenomics, plant and animal studies and population genomics. As developed technology provides robust workflows, main challenge in genome sequencing is the intactness of the starting DNA. Usually, quality of output data will mainly depend on starting material and as such presents critical step for obtaining reads of greatest lengths. For centromere investigations in species with large repetitive regions, isolation of high-molecular weight (HMW) DNA will especially be of importance and needs to be optimized for each organism.

2. Aim of the study

Paradoxically, in spite of essential function, previous studies showed that centromere exhibit remarkably fast evolution at both the DNA and protein levels. One of the most intriguing questions in cell biology is what is the driving force behind this evolution. Previous studies of centromeres have been primarily focused on sexual species where the centromere plays a role in both mitosis and meiosis. The long-held hypothesis of fast centromeric evolution, in spite of its essential function, has been explained as the function of centromeres as a response to the conflict between fast evolving DNA and protein components in meiosis. To test this hypothesis, nematodes of the genus *Meloidogyne* are used as model organism because they offer the unique platform to explore the structure and evolutionary dynamics of centromeric H3 histones (CenH3) and cenDNA components in species which reproduces exclusively through mitosis. Therefore, the main goal of this study is to analyze centromere components and their evolution in parthenogenetic (asexual) *Meloidogyne* species. The first step will be identification and investigation of CenH3 proteins in mitotic and meiotic *Meloidogyne* species based on available protein databases. Then, for the purpose of determining the functional potential, the expression profile of CenH3 transcripts in individual species and developmental stages will be determined using transcriptome bases. Furthermore, CenH3 deposition profile will be determined on chromosomes of *M. incognita* using immunofluorescence methods followed by determination and analysis of centromere-specific DNA sequences using chromatin immunoprecipitation (ChIP). Evolutionary centromere sequence trends will be analyzed in closely related species and its functional significance will be evaluated in mitosis. Also, the structure of the centromere will be viewed in the light of holocentrism and connected with this species polyploidy and complex evolutional origin.

Similar approach will be applied to centromere of beetle *T. castaneum* which challenges structural genome organization studies with presence of highly abundant repetitive fraction. Centromere-specific sequences will be evaluated in regard to high content of a single satellite DNA family and compared to landscape of centromere in *Meloidogyne* species.

Finally, new emerging approach of Nanopore sequencing in centromere research will be investigated. Elevating current obstacles in long stretches of similar sequences hold the potential to assemble centromeres from its start to end. First steps towards achieving these goals will be optimized on isolation of *T. castaneum* genomic DNA suitable for library preparations and obtaining sufficient read lengths.

3. Manuscripts

The Centromere Histone Is Conserved and Associated with Tandem Repeats Sharing a Conserved 19-bp Box in the Holocentromere of *Meloidogyne* Nematodes

Evelin Despot Slade[1], Brankica Mravinac[1], Saša Širca[2], Philippe Castagnone-Sereno[3], Miroslav Plohl[1*], Nevenka Meštrović[1*]

[1]Ruđer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia
[2]Agricultural Institute Slovenia, Ljubljana, Slovenia
[3]INRAE, Université Côte d'Azur, CNRS, Sophia Antipolis, France

* Corresponding authors:
E-mail: nevenka@irb.hr (NM)
plohl@irb.hr (MP)

Molecular Biology and Evolution
May 2021
Volume 38, Issue 5, Pages 1943-1965

Abstract

Although centromeres have conserved function, centromere-specific histone H3 (CenH3) and centromeric DNA evolve rapidly. The centromere drive model explains this phenomenon as a consequence of the conflict between fast-evolving DNA and CenH3, suggesting asymmetry in female meiosis as a crucial factor. We characterized evolution of the CenH3 protein in three closely related, polyploid mitotic parthenogenetic species of the *Meloidogyne incognita* group, and in the distantly related meiotic parthenogen Meloidogyne *hapla*. We identified duplication of the CenH3 gene in a putative sexual ancestral *Meloidogyne*. We found that one CenH3 (αCenH3) remained conserved in all extant species, including in distant *Meloidogyne hapla*, whereas the other evolved rapidly and under positive selection into four different CenH3 variants. This pattern of CenH3 evolution in *Meloidogyne* species suggests the subspecialization of CenH3s in ancestral sexual species. Immunofluorescence performed on mitotic *Meloidogyne incognita* revealed a dominant role of αCenH3 on its centromere, whereas the other CenH3s have lost their function in mitosis. The observed αCenH3 chromosome distribution disclosed cluster-like centromeric organization. The ChIP-Seq analysis revealed that in *M. incognita* αCenH3-associated DNA dominantly comprises tandem repeats, composed of divergent monomers which share a completely conserved 19-bp long box. Conserved αCenH3-associated DNA is also confirmed in the related mitotic *Meloidogyne incognita* group species suggesting preservation of both centromere protein and DNA constituents. We hypothesize that the absence of centromere drive in mitosis might allow for CenH3 and its associated DNA to achieve an equilibrium in which they can persist for long periods of time.

Introduction

Centromeres are specific chromosomal regions that recruit components of the kinetochore complex to enable accurate chromosome segregation during mitosis and meiosis. High-fidelity segregation is vital for all eukaryotic organisms and centromeric defects lead to chromosome breakage and aneuploidy. Regarding centromere architecture, the majority of animal and plant species have monocentric chromosomes characterized by primary constriction with a single regional centromere. In contrast, holocentric or polycentric chromosomes, with the centromere function distributed at multiple sites along the chromosome length, were observed in some nematode, insect, and plant species (Dernburg 2001; Guerra et al. 2010; Melters et al. 2012). In total, approximately 800 species have been reported to possess holocentromeres (Cuacos et al. 2015).

In general, centromere identity is defined by epigenetic determinants. An epigenetic mark of almost all functional centromeres is the specialized histone H3 variant, CenH3, which replaces the canonical H3 in centromeric nucleosomes (Allshire and Karpen 2008). CenH3 is associated with the centromeric DNA (cenDNA) and its incorporation into centromeric nucleosomes is considered a prerequisite for the proper assembly and function of the kinetochore (Blower and Karpen 2001; Talbert et al. 2002; Steiner and Henikoff 2015). Despite the conserved role of CenH3 in maintaining centromere integrity, CenH3 demonstrates accelerated evolution which is especially pronounced at its N-terminal tail and loop 1 of the histone-fold domain (HFD) (Malik and Henikoff 2001; Talbert et al. 2004). In most diploid genomes CenH3 is encoded by a single gene, whereas multiple copies of CenH3 have been common in polyploid plants. Multiple copies usually show a different expression pattern and the efficiency of their incorporation at centromeres can vary among different tissues (Yuan et al. 2015). In addition, recent studies revealed duplication events of CenH3 genes in some diploid plants and in Drosophila species (Sanei et al. 2011; Kursel and Malik 2017). These genes encode for functional CenH3 paralogous that colocalize at centromere during cell division.

Centromere regions of monocentric chromosomes are often enriched in repetitive DNA families, mainly megabase-sized satellite DNAs (satDNAs). Centromeric repeats

usually evolve rapidly, and significantly differ between closely related species (Plohl et al. 2014). Although many organisms possess a single satDNA which dominates in all centromeres (Hartley and O'Neill 2019), recent studies disclosed multiple satDNAs in centromeres, as it has been shown in the plant Pisum and related Fabeae species (Neumann et al. 2012; Ávila Robledillo et al. 2020). Extensive phylogenetic study of cenDNA candidates from 282 animal and plant species revealed astonishing diversity in their sequences which is difficult to associate with their conserved function (Melters et al. 2013). In the context of cenDNA role, evidences from studies of neocentromeres and dicentric chromosomes indicate that cenDNAs are neither necessary nor sufficient for centromere assembly (reviewed in Barra and Fachinetti 2018). On the other hand, studies on human alpha-satDNA show that, in contrast to other sequences, alpha-satDNA has property to facilitate assembly of CENPA (human CenH3) (Dumont and Fachinetti 2017). It has also been shown that existence of alpha-satDNA is necessary for de novo formation of human artificial chromosomes (HACs) (reviewed in McNulty and Sullivan 2018). In support, during the process of maturation, evolutionarily new centromeres rapidly accumulate satDNAs, and their recruitment increases segregation fidelity through binding with specific kinetochore proteins (Piras et al. 2010; Yang et al. 2018). In contrast to studies in monocentric species, the characterization of cenDNA in holocentric organisms is rare. For example, in the nematode Caenorhabditis elegans, the most studied holocentric species, centromere-specific sequences were not identified (Gassmann et al. 2012; Steiner and Henikoff 2014). Similarly, none of the identified high-copy repeats characterized in the holocentric plant Luzula showed colocalization with the centromere (Heckmann et al. 2013). On the other hand, a detailed CenH3-ChIP analysis in the holocentric plant Rhynchospora confirmed one satDNA as the underlying centromere sequence (Marques et al. 2015). However, the dilemma between exclusively epigenetic centromere definition and the role of cenDNAs in mediating centromere identity and function still remains unresolved (Talbert and Henikoff 2020).

The centromere drive model explains diversity of eukaryotic centromeres as a consequence of the conflict between rapidly evolving centromeric repeats and CenH3, suggesting asymmetry in female meiosis as the main factor responsible for rapid evolution of cenDNA and concomitant adaptive evolution of CenH3 (Malik 2009). Multiple reports in various animal and plant species with asymmetric meiosis have

suggested that CenH3 evolves under positive selection to suppress the deleterious effect of rapid changes in cenDNA (Henikoff et al. 2001; Malik and Henikoff 2001; Cooper and Henikoff 2004; Malik and Bayes 2006; Hirsch et al. 2009; Talbert et al. 2009; Schueler et al. 2010; Zedek and Bureš 2012). It was also shown that centromeres with expanded cenDNA and higher amount of CenH3 ("stronger centromeres") are more likely to segregate to the egg and thus be transmitted to offspring (Chmátal et al. 2014; Iwata-Otsubo et al. 2017). In support to this hypothesis, species with symmetric meiosis display a lower frequency of adaptive evolution of CenH3 compared with those with asymmetric meiosis (Zedek and Bureš 2016a). However, information about the possible role of centromere drive in species with holocentric chromosomes is scarce and controversial. The absence of positive selection on CenH3 in holocentric Luzula species suggests that holocentric chromosomes may suppress centromere drive (Zedek and Bureš 2016b). On the other hand, HCP-3 (CenH3) from *C. elegans* is rapidly evolving, even though HCP-3 was not required for oocyte meiotic divisions in this holocentric nematode (Monen et al. 2005).

In this work, we address the evolution and organization of centromere components, CenH3 and cenDNA, using the plant-parasitic nematode *Meloidogyne incognita* and its congeners as a model system. Among them, *M. incognita*, *M. javanica*, and *M. arenaria* (*M. incognita* group—MIG), are closely related species which have been determined as obligatory mitotic parthenogens which do not undergo meiosis and reproduce asexually (Castagnone-Sereno and Danchin 2014). On the contrary, the phylogenetically distant species *M. hapla*, reproduces by both meiotic parthenogenesis and cross fertilization (Castagnone-Sereno et al. 2013). In recent years, whole-genome sequencing of MIG species and *M. hapla*, enabled comparative genome analyses which revealed substantial differences in their genome structure (Abad et al. 2008; Opperman et al. 2008; Blanc-Mathieu et al. 2017). Genome studies indicated that MIG species are polyploids with whole-genome duplications, whereas *M. hapla* is a diploid species with a small and compact genome. Interspecific hybridization has been highlighted as a critical force in the processes of polyploidization in MIG species (Lunt 2008; Blanc-Mathieu et al. 2017). Regarding the organization of the centromere, the absence of primary constriction and holocentric-like mitosis observed by classical cytological approach proclaimed *Meloidogyne* to be holocentric species

(Triantaphyllou 1981). This classification is also supported by karyotype variability between *M. incognita* populations, fluctuating from 40 to 46 chromosomes (Triantaphyllou 1981). In general, the studies of centromere have been performed mostly in diploid sexual species with monocentric centromere organization. Therefore, *Meloidogyne* species offer the unique platform to explore the evolutionary dynamics of CenH3 and cenDNA components in exclusively asexual animal species which also possess holocentromere, the poorly investigated centromere organization. In addition, the fact that the CenH3 evolutionary trends in polyploidization have been explored very limitedly in animal species makes *Meloidogyne* a valuable model to address those studies.

In the present work, we characterized CenH3 proteins in the selected parthenogenetic *Meloidogyne* species and analyzed their evolution considering the complex species history. Our results suggested the duplication of a CenH3 gene in a common sexual ancestor of both mitotic and meiotic *Meloidogyne* species. We found that one CenH3 gene is preserved as nearly identical in all analyzed species, including in the distantly related *M. hapla*, whereas the other evolved rapidly and formed four different CenH3 variants. We further investigated the centromere DNA composition in *M. incognita* using chromatin immunoprecipitation (ChIP) and immunofluoreoscence (IF) techniques and unveiled the unique characteristics of the holocentromeres in an exclusively mitotic species.

Results

Meloidogyne Species Have Multiple and Divergent CenH3 Genes/Proteins

To identify CenH3 candidates in selected *Meloidogyne* species (three obligatory mitotic species; *M. incognita*, *M. arenaria*, and *M. javanica* [MIG] and facultative meiotic *M. hapla*) *C. elegans* CenH3 (HCP-3) was used as query for BLAST analysis against a protein database for each species. After elimination of truncated protein sequences (see Materials and Methods), the 21 CenH3 protein candidates with specific CenH3 features (Vermaak et al. 2002) were detected (supplementary fig. 1). The sources of all CenH3 protein candidates together with abbreviated names are listed in supplementary table 1. Given that N-terminal tails among CenH3 proteins were hypervariable (supplementary fig. 1) an alignment of the more conserved histone-fold domains (HFD) was selected to estimate mutual sequence identities (supplementary table 2) and phylogenetic relationships (fig. 1A, left). Phylogenetic tree showed the branch topology with two distant well-supported clades. Interestingly, although both groups show CenH3-specific features tree branching support their polyphyletic origin (fig. 1A, left). Among the CenH3 candidates, abcCenH3 group of sequences shows low HFD sequence identity to H3 (33–42%), whereas αβγδεCenH3 group shares considerably higher sequence identity to H3 histone (from 50% to 63%) (supplementary table 2). The abcCenH3 group includes highly divergent CenH3 candidates divided in two subgroups (aCenH3s and bcCenH3s) which show the low mutual sequence identity in HFD (34–37%) (supplementary table 2). One subgroup consists of three similar sequences (aCenH3; identity 85–100%) which belong to closely related MIG species whereas the other clade includes two rather divergent CenH3 candidates (bCenH3 and cCenH3, showing mutual HFD identity of 79%) from more distant *M. hapla*.

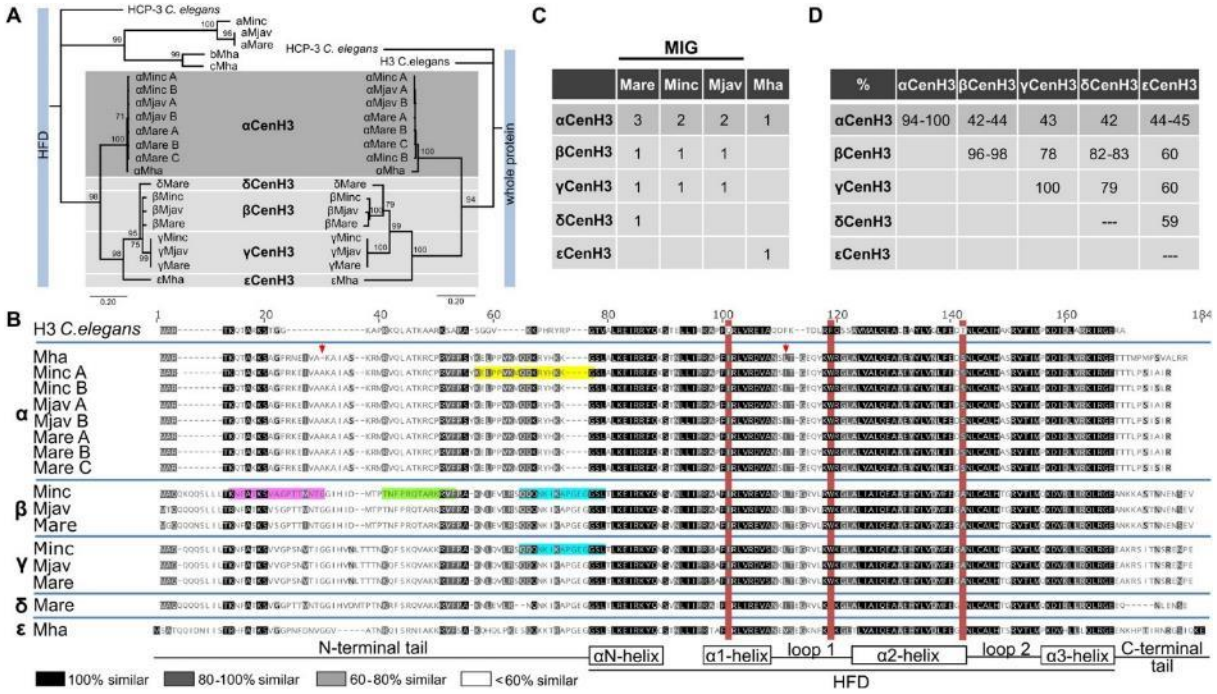


Fig. 1. Identification of CenH3 proteins in *Meloidogyne* species (Minc, *M. incognita*; Mare, *M. arenaria*; Mjav, *M. javanica*; and Mha, *M. hapla*). (A) Phylogenetic analyses using NJ with a protein alignment of the (HFD) (left) and full length of CenH3 sequences (right) of all detected *Meloidogyne* CenH3s. Bootstrap values above 50 are displayed. CenH3 sequences are separated in the subgroups (a, b, c, α, β, γ, δ, and ε). (B) Amino acid alignment of CenH3 proteins separated into α, β, γ, δ, and ε subgroups. The red boxes indicate diagnostic amino acid changes in comparison to H3 from Caenorhabditis elegans. Secondary structure of histone-fold domain (HFD) is depicted below the alignment. The arrows indicate amino acids changes in αCenH3 from *M. hapla* in comparison to other αCenH3s. Highlighted are peptide sequence regions which were used as antigens to produce antibodies to αCenH3 (yellow), βCenH3-1 (green), βCenH3-2 (magenta), and βγCenH3 (cyan). (C) Copy number of CenH3 genes in *Meloidogyne* genomes. (D) Sequence identity matrix of CenH3 protein sequences.

In contrast, αβγδε CenH3 group represents a far more homogeneous CenH3 sequences with mutual HFD identity from 65% to 100%. As expected for CenH3 proteins, most of the sequence divergence was concentrated in the N-terminal tail (fig. 1B). The phylogenetic trees based on a multiple alignment of HFD domains exclusively (fig. 1A, left) and on alignment of complete α, β, γ, δ, and εCenH3 protein sequences (fig. 1A, right) showed the same branching topology with two monophyletic subclades. The α subclade consists of eight αCenH3 proteins. The multiple αCenH3s were detected in MIG species: two in *M. incognita* and in *M. javanica*, and three in *M. arenaria* (fig. 1C). On the other hand, only one αCenH3 was detected in *M. hapla*. All members of αCenH3 subclade are almost completely conserved within MIG species with only one change in the C-terminal tail (fig. 1B). Interestingly, in distant *M. hapla*,

except at the very end of C-terminal tail which is substantial different, only two AA changes were found in comparison to MIG species. These changes include deletion of one AA within the N-tail and one AA change within Loop 1 in HFD region (fig. 1B). The other subclade comprises β, γ, and δCenH3s detected in MIG species. The β and γCenH3s were found in *M. incognita* and in *M. javanica*, whereas β, γ, and δCenH3s were found in *M. arenaria*. The εCenH3 was exclusively found in distant *M. hapla* (fig. 1C). Detected intragroup identity ranges from 96% to 98% for βCenH3s to 100% for γCenH3s (fig. 1D and supplementary table 3). Concerning intergroup identity, αCenH3s share ~43% identity with βγδεCenH3 sequence group, whereas identities between the members of βγδε CenH3 group were considerably higher, ranging from 59% for δCenH3 and εCenH3 to ~83% for βCenH3 and δCenH3 (fig. 1D and supplementary table 3).

To test expression profile of all CenH3 candidates, publicly available raw Illumina transcriptome data from five clearly defined developmental stages including egg, juvenile J2, J3, J4, and female/male in *M. incognita*. Results showed that all copies of the CenH3 genes are actively transcribed (fig. 2A). However, through the development of *M. incognita* expression of αCenH3 was 20- to 50-fold higher in contrast to other CenH3 genes whose transcription proved to be extremely low. The exception was female stage where transcription of αCenH3 is decreased to the level of other CenH3. To test the quality of analyzed *M. incognita* transcriptomes expression of the reference gene disulfide-isomerase (Disu) (Hu and DiGennaro 2019) was performed. The results showed a comparable amount of Disu transcripts in all *M. incognita* developmental stages (fig. 2B) proving the reliability of the used RNA-seq data sets. In the closely related *M. arenaria* and *M. javanica* analyses on available mixed eggs and J2 stages also revealed significantly higher expression of αCenH3 and very low level of transcription of other CenH3 genes. A similar phenomenon was observed in diploid meiotic *M. hapla* where αCenH3 transcripts prevailed in analyzed samples (eggs and J2) in comparison to other CenH3s. In conclusion, expression profile of different CenH3 genes in all developmental stages as well as in different *Meloidogyne* species shows similar pattern characterized by dominant expression of αCenH3. It should be noted that RNA-seq from reproductive stages (females and males) were not available for *M. arenaria*, *M. javanica*, and *M. hapla*.

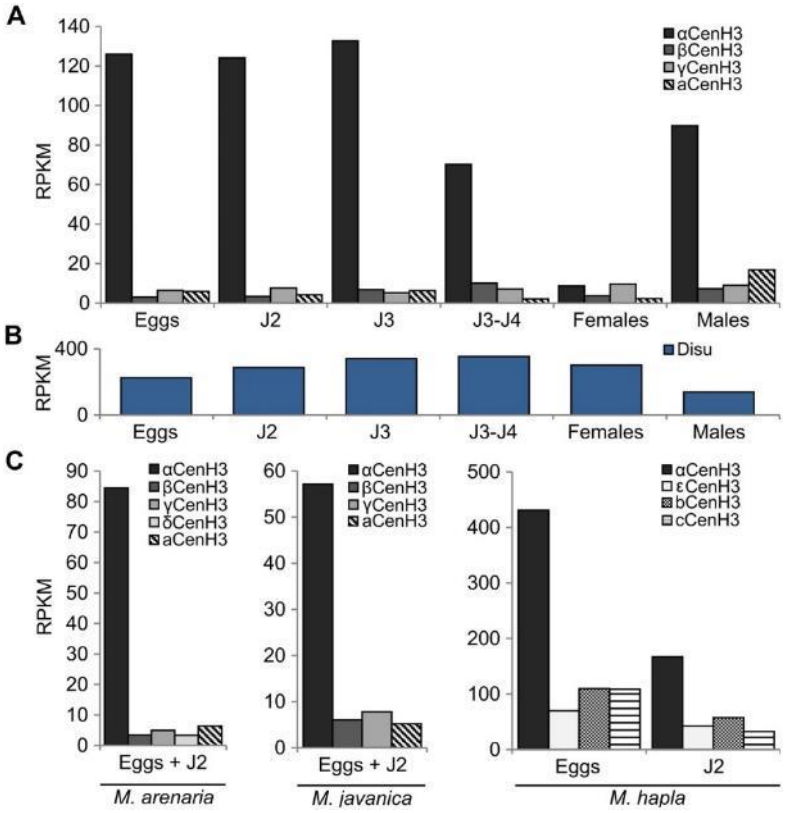


Fig. 2. Expression profile of CenH3 genes in different species and developmental stages. (A) CenH3 expression through life cycle of *Meloidogyne incognita*. (B) Expression of Disu reference gene in *M. incognita*. (C) Expression of CenH3 in three closely related species; *M. arenaria*, *M. javanica*, and *M. hapla*. The relative expressions of CenH3 genes in different samples of RNA-seq data were analyzed using Bowtie2 v.2.3.0 mapper (Langmead and Salzberg 2012). Hits were normalized with RPKM (reads per kilobase of transcript per million mapped reads) method. Expression profile was shown as logarithmic transformation of RPKM values. The developmental stages include eggs, different juvenile stages (J2, J3, J4), females, and males. The RNA-seq data with accession numbers are listed in Materials and Methods section.

Different Evolutionary Dynamics of αβγδε CenH3 Group of Genes

Dominant expression and conservation of αCenH3 gene in all analyzed species, including *M. hapla* provoked us to focus the study on αCenH3 and its closely related CenH3s grouped in the monophyletic αβγδε CenH3 clade. To evaluate how different CenH3 genes evolve, we examined the overall ω value for different gene comparisons (supplementary fig. 2) using the substitution model implemented in MEGA version X (Kumar et al. 2018). First, intragroup comparisons were done on the full-length sequence alignment among α, β, and γ CenH3 variants originating from different species. All intragroup analyses revealed extremely purifying selective pressure where ω ratio varied from 0 to 0.5 (supplementary table 4). Intergroup analyses of CenH3 genes were carried out on HFD and N-termini alignments separately. The results indicated purifying selective pressure on HFD domain in all pair-wise comparisons with ω values ranging from 0.01 to 0.1 (supplementary table 4). In N-termini intergroup analyses, αCenH3 was excluded due to its high divergence to β, γ, δ, and εCenH3s, which resulted in incorrect alignments. N-termini selection analyses in β, γ, δ, and εCenH3 comparisons showed that ω value was higher in comparison with HFD analyses but still <1, which appears to be a signature of stabilizing selection (supplementary table 4). Given that positive selection could act only on a few codons, to identify potential sites under positive selection, we carried out MEME model of codon substitution that allows ω value to vary across both codons and branches. The results are shown in figure 3 and supplementary table 5. These analyses detected the same pattern of codon evolution in all gene comparisons among β,γ,δ, and εCenH3 group (βδ, βγδ, βγε, and βγδε) identifying few statistically significant positively selected sites in the first part of N-terminal tail (fig. 3). No codon under positive selection was found in comparison of αCenH3s from MIG species and distant *M. hapla*. Our results therefore confirm that αCenH3 sequences have evolved under purifying selection, whereas βγδεCenH3 sequences have undergone positive selection in the part of N-terminal domain.

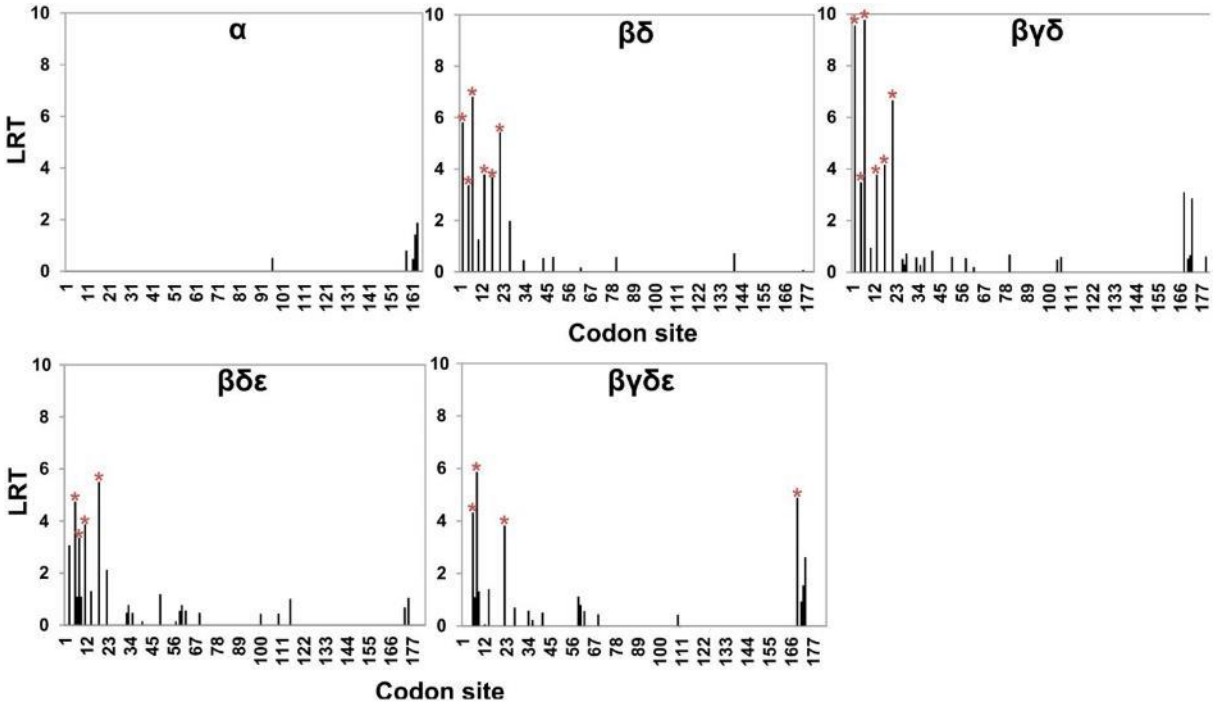


Fig. 3. Codon-specific tests for positive selection of CenH3 genes. Tests were inferred by mixed effects model of evolution (MEME) using the likelihood ratio test (LRT). LRT values with P values <0.1 were considered as codons under positive selection (red stars) (source data are supplied in the supplementary table 5). Multiple alignments of CenH3 genes used for analyses are shown in supplementary figure 2.

Distribution of αCenH3 Centromeres in *M. incognita*

Chromosomal localization of CenH3 proteins were assayed in female gonads of *M. incognita* using polyclonal antibodies against peptide corresponding to the N-terminus of CenH3s (fig. 1B). First analyses were done with αCenH3 because it is dominantly expressed compared with the rest of CenH3s. Western blot with rabbit-raised anti-αCenH3 demonstrated that the antibody recognizes proteins of the predicted molecular weight of 18 kDa in protein isolate of eggs, J2 stage and also in females (supplementary fig. 3A). Although low transcription of αCenH3 gene has been shown in females, somewhat lower but comparable amount of αCenH3 proteins exist in females in comparison to J2 and eggs (supplementary fig. 3A). This unusual phenomenon of low transcription in females could be explained by stability of αCenH3 protein in the process of transition from J4 stage (high αCenH3 transcription) to females (low αCenH3 transcription). In support to this, recent data on Mus musculus

oocytes showed that centromere function of oocyte does not depend on the loading of newly transcribed CenH3 implying the stability of CenH3 protein (Smoak et al. 2016).

Given that *M. incognita* is a mitotic parthenogenetic species, cytosmear preparations from reproductive female tissue (ovaries and uterus) represent exclusively mitotic divisions. Chromosomal distribution of αCenH3 through different mitotic phases evaluated by anti-αCenH3 immunofluorescence (IF) is presented in figure 4A. In interphase nuclei, many αCenH3 signals differing in intensities were found. With the progression of the mitotic cycle when the chromosomes' contours became visible, αCenH3 clusters that differ in intensity and representation become more apparent (fig. 4A, prophase). In addition to high number, *M. incognita* chromosomes are characterized by remarkable diminutives. *Meloidogyne incognita* population analyzed in this work has 46 chromosomes ranging in size from 0.4 to 1.5 μm in metaphase, whereas, for comparison *C. elegans* possess only five pairs of significantly bigger chromosomes (~5 μm in length in metaphase). Immunofluorescence on *M. incognita* metaphase chromosomes revealed unexpected patterns of αCenH3 distribution according to which the chromosomes can be classified roughly into six types according to αCenH3 distribution (fig. 4A metaphase and fig. 4B). The chromosomes with strong αCenH3 signal which seems to occupy the entire chromosome length in the condensed metaphase (fig. 4B, chromosome type 1) are predominant (~20 out of 46). The other group includes the chromosomes with uneven distribution of αCenH3 signal characterized by combining discrete and abundant αCenH3 regions in different chromosome areas (fig. 4B, chromosome types 2–5). Abundant αCenH3 regions may occupy even more than half of the chromosome length (4B, chromosome type 2) or appear as bicentric (fig. 4B, chromosome types 3 and 4) and telocentric clusters (fig. 4B, chromosome type 5). Finally, a few chromosomes show discrete αCenH3 clusters dispersed along the entire chromosome (fig. 4A and B, chromosome type 6). To examine the αCenH3 centromeric chromatin at higher resolution, immunostaining experiments were performed on chromatin fibers. The results show organizational pattern in which interspersed αCenH3 domains are interrupted by αCenH3-free subdomains (fig. 4C).

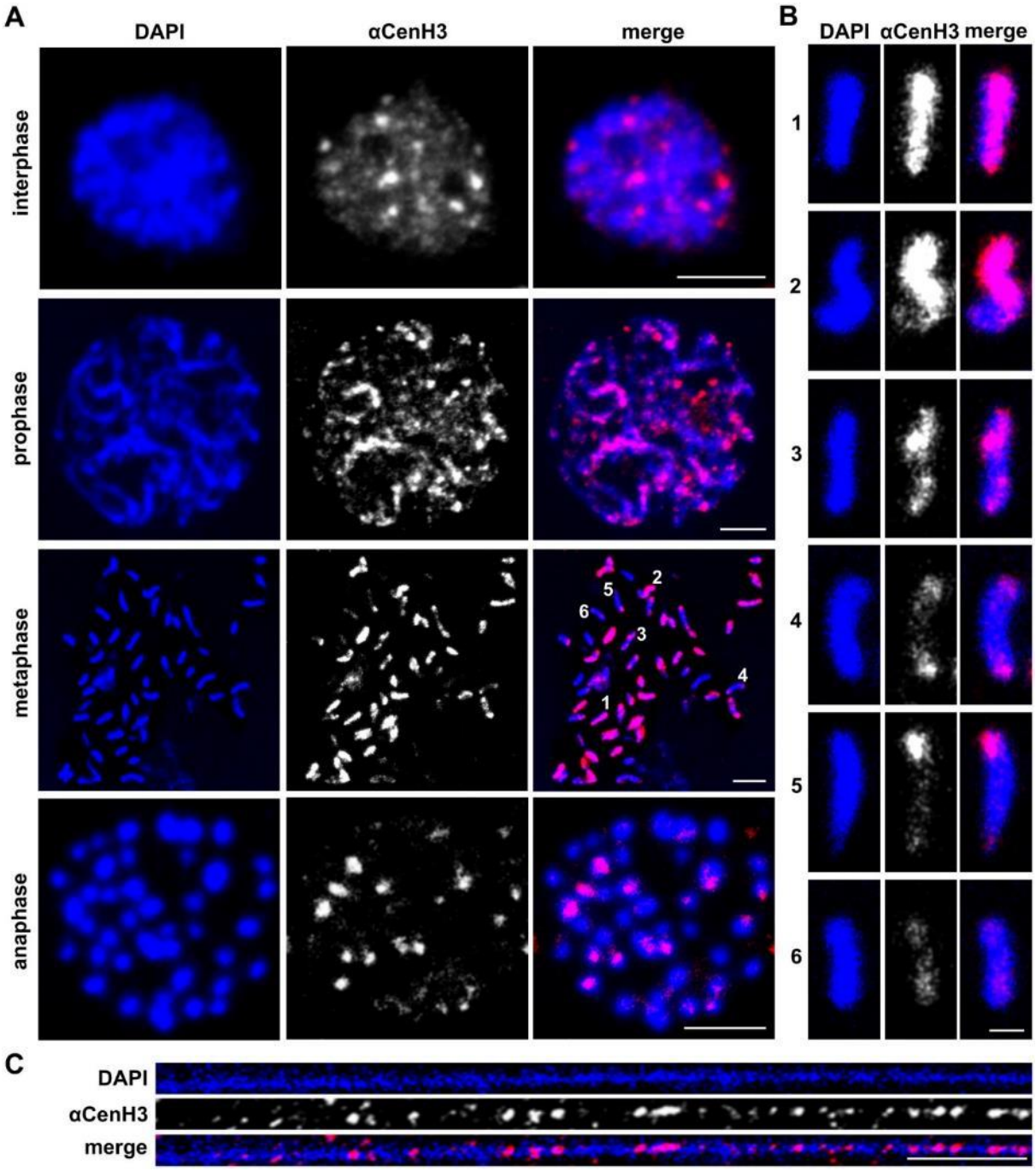


Fig. 4. The organization of αCenH3 centromeres in *Meloidogyne incognita*. Slides were prepared from isolated reproductive tissue of females (ovaries and uterus). (A) Immunofluorescence of αCenH3-containing domains (red) during the mitosis cycle in *M. incognita* using anti-αCenH3 antibodies raised in rabbit 2 (supplementary fig. 3A). Scale bar = 5 μm. (B) Distribution pattern of αCenH3-containing domains along metaphase chromosomes in six different chromosome types. Selected chromosome types were indicated in metaphase spread with numbers. Scale bar = 1 μm. (C) Immunofluorescence of αCenH3-containing domains (red) on chromatin fiber. Scale bar = 5 μm. All chromosomes and fibers were counterstained with DAPI (blue). Images were acquired with confocal microscopy and shown as z-stack projection.

To disclose chromosomal deposition of β and γCenH3s in *M. incognita*, polyclonal antibodies against two epitopes specific for βCenH3 (βCenH3-1 and βCenH3-2) as well as the epitope shared by β and γCenH3 (βγCenH3 epitope) (fig. 1B) were generated in parallel in rabbits and guinea pigs. The peptides selected for immunization encompassed even 55% of βCenH3 N-terminal sequence (41AA of 75AA) and represented all potential N-tail βCenH3-specific epitopes (fig. 1B). Western-blot analysis using anti-βCenH3-1 and anti-βCenH3-2 as well as anti-βγCenH3 did not detect any specific band of the expected molecular weight of 18 kDa (supplementary fig. 3A). Although, Western-blot results were in accordance with very low transcription of βCenH3 and γCenH3 genes in RNA-seq data (fig. 2) it remained doubtful whether antibodies recognized the epitopes and consequently βCenH3 and γCenH3 proteins. Peptide dot blot validation revealed that anti-βCenH3-1, anti-βCenH3-2, and anti-βγCenH3 generated in rabbits were sensitive for a peptide of interest without cross-reactivity with nonspecific peptides, including peptide specific for αCenH3. Among antibodies raised in guinea pig only anti-βCenH3-2 recognized specific peptide (supplementary fig. 3B).

Taking into account that β and γCenH3s could be present at a level below the detection limit of Western-blot analyses, we performed IF experiments in an additional attempt to verify if these CenH3s still participate in female centromeres. The double IF combining anti-αCenH3 (produced in rabbit) and anti-βCenH3-2 (produced in guinea pig) as well as immunostaining with anti-βCenH3-1/-2 or anti-βγCenH3 antibodies raised in rabbits were performed. Although approximately 100 cytological specimens were analyzed, neither of these experiments could confirm the presence of βCenH3 and/or γCenH3 on M incognita chromosomes (data not shown). Although, the participation of β and γCenH3 in *M. incognita* centromere cannot be completely ruled out due to limited methodological approaches in nonmodel organism their low transcription in all developmental stages together with efficient validation of antibodies by indirect assay of peptide dot blot, suggest that unsuccessfully detection of βCenH3 and γCenH3 by Western blot and IF most likely reflects the absence of βCenH3 and γCenH3 protein synthesis and consequently their lack in *M. incognita* centromere. Thus, it is the most probable that αCenH3 is, if not exclusively, then certainly predominantly incorporated into *M. incognita* centromeres.

In addition, to check if microtubules attach to the αCenH3 units, coimmunostaining with antibodies against α-tubulin and anti-αCenH3 was analyzed. The specificity of anti-α-tubulin was confirmed by Western blot on protein crude extract from *M. incognita* eggs (supplementary fig. 4A). Considering that the most of the oocytes present in the ovaries and uteri of *M. incognita* females appear to be in prophase/prometaphase (Triantaphyllou 1981), we were able to visualize α-tubulin and αCenH3 on chromosomes only in prometaphase (supplementary fig. 4B). Although, the results showed that the αCenH3 clusters mainly colocalized with the α-tubulin these images could not offer high-resolution view on the microtubule attachment at sites where αCenH3 was not detected nor detailed examination of α-tubulin distribution in comparison to αCenH3 along the chromosome length.

Composition of αCenH3-Associated DNA in *M. incognita*

To map DNA sequences in αCenH3 centromere of *M. incognita*, native ChIP followed by next-generation sequencing (ChIP-seq) was performed. Three ChIPped DNA (supplementary fig. 5A) and input DNA libraries were generated and sequenced. To evaluate the enrichment of αCenH3-associated DNA in ChIP samples, ChIPped DNA (supplementary fig. 5B) was labeled and combined IF/fluorescence in situ hybridization (FISH) with anti-αCenH3 and labeled ChIPped DNA probe was conducted. The results showed highly abundant overlapping signals of ChIPped DNA and αCenH3 clusters thus confirming the specificity of ChIP experiments (supplementary fig. 5C). These strong FISH signals could likely have derived from repetitive DNA associated with αCenH3, because repeats enriched in ChIP DNA would result in strong hybridization. Mapping of ChIP reads on the reference genome using BWA (Li and Durbin 2009) produced high and unreliable peaks at very end of scaffolds or around placeholders (N stretches) which probable represent loci with repetitive sequences are only partially included in the genome assembly The fact that *M. incognita* genome is assembled into 12,091 scaffolds indicates high genome fragmentation and speak in favor of poor representation of repetitive regions in the assembled genome (Blanc-Mathieu et al. 2017). Following the above observations, we decided to focus our ChIP analysis on the repetitive genome fraction.

Therefore, an alternative approach of graph-based repeat clustering (Novák et al. 2013), which is independent of the assembled reference genome, was used for estimating read enrichment associated with repeat sequence types (fig. 5). The results of ChIP/input ratio for each ChIPped DNA together with values of SD are presented in supplementary table 6. The ChIP/input ratio >1.5 was chosen as a threshold for considering the clusters enriched in the ChIP samples (fig. 5). Detailed analyses of enriched clusters, CL25 and CL16, revealed that they represent complex groups composed of different contigs, which made impossible to define the cluster consensus sequence. For that reason, more in-depth but complementary analyses of ChIP/input ratio on individual contig sequences (at least 0.002% genome abundance) were performed. The results disclosed that two contigs (ctg11 and ctg20) from the enriched cluster CL16 and two contigs (ctg8 and ctg5) from the enriched cluster CL25 show enrichment >1.5 suggesting that these sequences were most likely the αCenH3-associated DNAs (fig. 5 and supplementary table 6). The analyses of enriched contigs by Tandem Repeat Finder (TRF) pipeline revealed that CL25ctg8 and CL25ctg5 represent arrays of tandem repeated 70-bp-long monomers (Cl25m1i) (supplementary fig. 6A). In addition, the same analysis of CL16ctg11 and CL16ctg20 showed that these contigs are composed of three different tandem repeats (TRs) with monomer units of 83, 55, and 45 bp (CL16m1i, CL16m2i, and CL16m3i, respectively) (supplementary fig. 6B). Aligned consensus sequences of monomers extracted from the enriched contigs (CL25m1i, CL16m1i, CL16m2i, and CL16m3i) are presented in figure 6A. Moreover, annotation analyses of all contigs from the cluster CL25 revealed that CL25 mostly comprised TR arrays with CL25m1i monomer (supplementary fig. 6A). In addition, mapping of CL16-specific contigs with CL16m1i-m3i monomers showed that only 1/3 of contigs represent monomeric (monomers belonging to the same family) or mosaic TR arrays composed of monomers from different families (fig. 6D and supplementary fig. 6B). The reason that only two contigs from each cluster (CL25 and CL16) proved to be enriched in contig analysis lies in the fact that these sequences are the longest TR arrays composed of αCenH3-associated monomers in enriched clusters. Interestingly, despite different monomer length (45–83 bp) and relatively low-sequence similarity (55.8–72.7%, fig. 6B), αCenH3-associated monomers show a completely conserved 19-bp sequence box (TCGGGCCTTCGGCCCTCGC, fig. 6C).

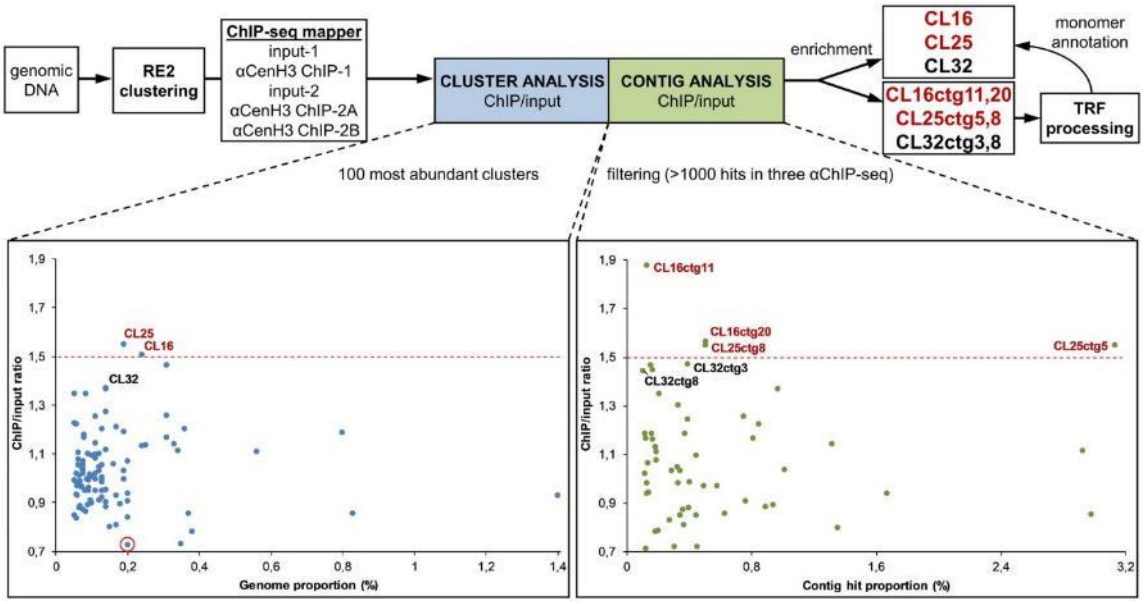


Fig. 5. Identification of αCenH3 ChIP-enriched sequences. Strategies for identifying the most abundant repeat clusters and contigs associated with αCenH3 chromatin in *Meloidogyne incognita* (on the top). Relative enrichments of repeat DNA families in the ChIP-seq data are presented for clusters (left graph) and contigs (right graph) analysis. Clusters/contigs are represented by dots. The y axis is the ratio of the ChIP-seq reads to input-seq reads, representing the enrichment of each corresponding cluster/contig from the ChIP-seq data. The x axis is the genome proportion for each cluster (left graph) or hit proportion for each contig (right graph). A cluster rounded in red was used as a negative control in the IF-FISH experiment. Data with ChIP enrichment analyses of clusters and contigs with SDs are presented in supplementary table 6.

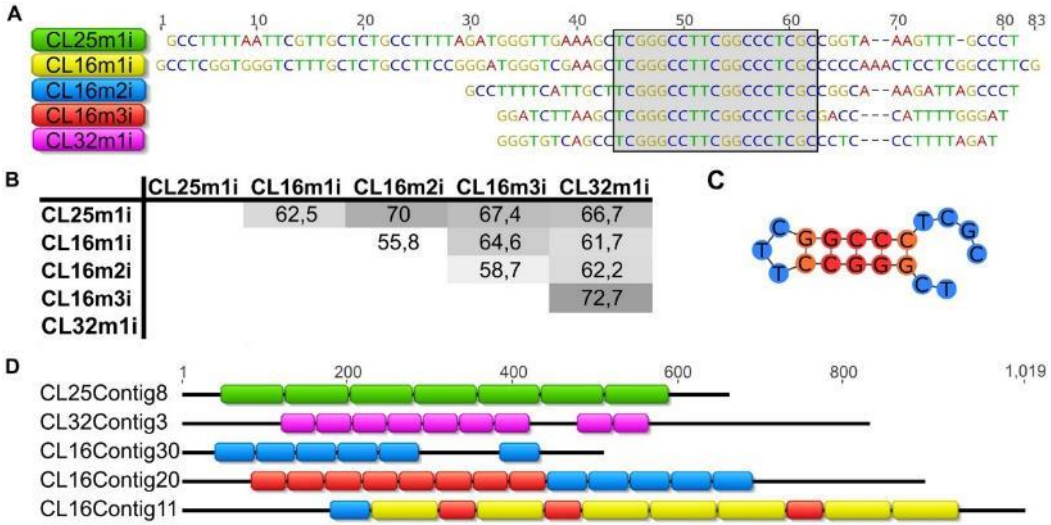


Fig. 6. Candidates for αCenH3-associated sequences in *Meloidogyne incognita*. (A) Alignment of consensus monomer sequences extracted from tandem repeated arrays (TRs) enriched in ChIP-seq analyses. The names of the consensus monomers were derived according to the cluster from which they originated. The conserved 19-bp box is indicated within the gray shaded area. (B) The percentage of identity among consensus monomer sequences. (C) Secondary structure of the conserved 19-bp box

sequence. Folding free energy of the 19-bp fragment is −5.40 kcal/mol. (D) The most prominent examples of array organization of αCenH3-associated sequences (source data are supplied in supplementary fig. 6). The color code of monomer sequences in arrays corresponds to monomer labels at the panel (A).

In order to detect possible additional αCenH3 centromeric candidates in ChIP data, all other clusters/contigs were searched against the conserved 19-bp box. The results revealed that the cluster CL32 together with the contigs CL32ctg8 and CL32ctg3 have tandem repeat organization of monomers which contain the conserved 19-bp box (fig. 6D and supplementary fig. 6C). In support, the results of ChIP/input ratio for these cluster/contigs showed enrichment slightly below the defined threshold (fig. 5). Extraction of monomers from CL32ctg8 and CL32ctg3 contigs disclosed a new TR family, CL32m1i, which shares sequence similarity of 72.7% with CL16m3i monomer sequence (fig. 6A and B). Annotation of CL32m1i monomer to contigs of CL32 cluster showed that only 14 of 60 contigs contain short TR arrays composed of CL32m1i monomers (supplementary fig. 6C). The other contigs from clusters CL25, CL16, and CL32 which are not involved in αCenH3-associated monomers, probably represent different TR surrounding regions (supplementary fig. 6). This assertion is supported by the fact that different non-αCenH3-associated contigs assemble the flanking regions of αCenH3-containing TRs (supplementary fig. 6D). The specific feature of all αCenH3-associated monomers is relatively high GC content (>50%), especially pronounced in the conserved 19-bp box (~80%), in contrast to 70% AT-richness that characterizes the assembled part of the genome (Abad et al. 2008). In addition, sequence analyses of the conserved 19-bp box revealed two sequence segments repeated in an inverted orientation which has a high potential to form an energetically stable dyad structure (fig. 6C).

Validation of the ChIP-Identified αCenH3-Associated Sequences in *M. incognita*

The localization of αCenH3 on the chromosome level showed pattern with highly abundant and discrete αCenH3 clusters. It can be assumed that TRs associated with αCenH3-highly abundant cluster are not represented in the reference genome, although the existence of a genome-wide distribution of αCenH3 discrete clusters implies their presence (at least to some extent) in the genome assembly. Therefore, to further validate our results of ChIP analyzes obtained on reference repeat database, we performed an additional survey on the reference genome. The ChIP background was removed by subtracting input signal from the ChIP data. The reference genome was simultaneously mapped with ChIP sequences and the 19-bp box to assess the enrichment of TRs-containing the 19-bp box in the genome assembly. We identified 1,117 19-bp boxes on the reference genome and as expected from previous ChIP mapping data the most of the detected 19-bp box-containing TRs were located at the ends of the scaffolds thus confirming the underrepresentation of these sequences in the assembly. However, TRs distributed along the scaffolds could be associated with discrete αCenH3 clusters were also found. Some representative examples provided as genome browser views in supplementary figure 7A, clearly showed that distribution of enrichment correlates with TRs containing the 19-bp box in all three ChIP samples, whereas the 19-bp motif-free regions of the scaffolds did not show enrichment for αCenH3-ChIPped sequences.

In addition, we mapped WGS, input, and ChIP data with the 19-bp box to compare abundance of αCenH3-associated TRs among those genome resources and comparing it also to genome assembly (supplementary fig. 7B). WGS and input with ∼0.22% of TRs containing 19-bp box sequences showed about 5× higher abundance related to the assembled genome (0.04%). All three ChIP data showed the same ratio of TRs containing the 19-bp box enrichment (1.55×) in comparison to input data. This is in accordance to average enrichment obtained in ChIP analyses on repetitive database conducted by Repeat Explorer (fig. 5). Based on Repeat Explorer analysis, 6.3% of WGS reads were classified as tandem repeats. TRs containing the 19-bp box with 0.22% abundance of WGS thus make only 3.5% of the detected tandem repetitive fraction in *M. incognita* genome. Considering that the several methodological issues in ChIP and Illumina sequencing workflow (e.g. fragility or resistance to DNA

fragmentation, A/T homopolymers, high %GC) may cause underrepresentation of different types of repetitive sequences in WGS and ChIP data, to estimate experimentally the genome abundance of αCenH3-associated TRs, we performed the dot blot of *M. incognita* genomic DNA with TRs probes (CL25m1, CL16m1i, and CL16m3i/CL32m1i; probe preparation is explained in the section below). Interestingly, the results of dot blots revealed that αCenH3-associated TRs comprise cumulatively about 2.3% of the genome (supplementary fig. 7C), which is almost 10× higher than was estimated in WGS/input data and more than 50× higher than in the assembled genome (supplementary fig. 7B), Since there is no significant discrepancy in abundance between WGS and inputs data related to αCenH3-associated TRs it can be assumed that the Illumina sequencing rather than DNA fragmentation step cause underrepresentation of these repetitive sequences in WGS/input data. It is known that high GC content interfere with PCR-based library amplification, causing a depletion of the GC-rich templates (Aird et al. 2011). Regarding to relatively high GC content (53–68%) of αCenH3-associated TRs, it is most likely that this factor causes a depletion of the corresponding sequences in WGS/input data.

Cytological Confirmation of the ChIP-Identified αCenH3-Centromeric Repeats in *M. incognita*

To test the association of ChIP-enriched TRs with functional αCenH3 centromere domains, IF-FISH using αCenH3-specific antibody and centromere associated monomers as hybridization probes was performed. The labeled DNA probes were generated for CL25m1i, CL16m1i, and CL16m3i (supplementary fig. 8A). Primers specific for CL32m1i were difficult to design due to short sequence and high similarity with CL16m3i. Therefore, FISH was conducted under moderate stringency conditions which allow a single probe to hybridize to both variants, CL32m1i and CL16m3i. Since primers specific for CL16m2i did not produce relevant PCR profile, this probe was excluded from analyses. Given that metaphases are extremely rare in chromosomal preparation of *M. incognita* IF-FISH analyses were presented on interphases. To confirm αCenH3-specific localization and disclose distribution pattern of putative cenDNA sequences on αCenH3 domains, we first performed IF-FISH using CL25m1 and CL16m1i as hybridization probes. The results of combined IF-FISH analysis with CL25m1-i and CL16m1i probes showed that approximately half of the αCenH3

centromeres contain a considerable amount of these TRs (fig. 7A). In addition, IF-FISH with all TRs probes (CL25m1, CL16m1i, and CL16m3i/CL32m1i) showed that these sequences cover the majority of αCenH3 domains resulting in overlapped regions (yellow fluorescence signals at fig. 7B).

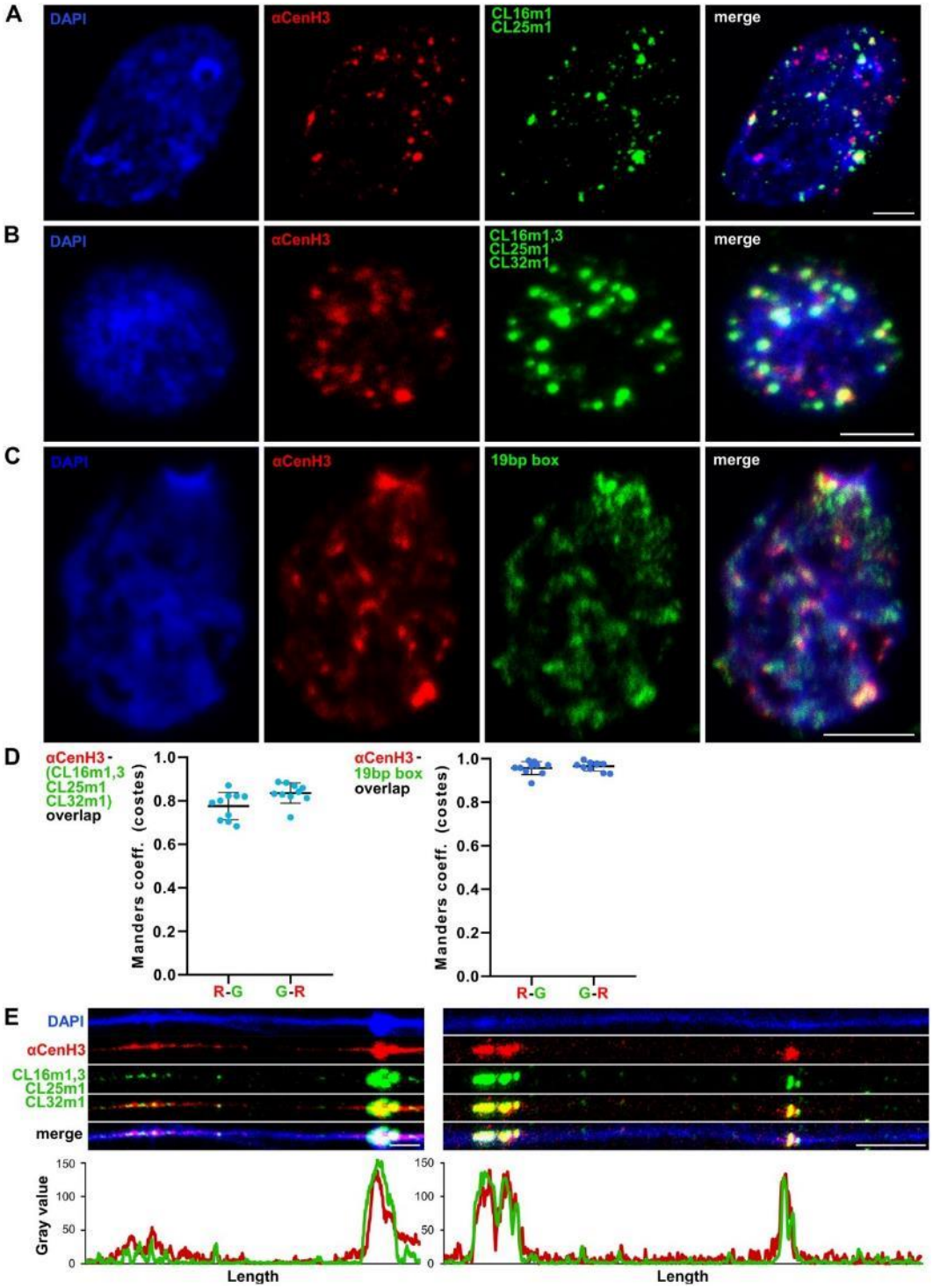


Fig. 7. Simultaneous detection of αCenH3 centromere and αCenH3-associated DNA in *Meloidogyne incognita*. Slides were prepared from isolated reproductive tissue of females (ovaries and uterus). (A)

Combined immunofluorescence with anti-αCenH3 raised in rabbit 2 (red) and FISH with CL16m1i and CL25m1i αCenH3-associated monomers as probes (green). (B) IF-FISH with anti-αCenH3 (red) and mixed probe for αCenH3-associated monomers, CL16m1i, CL16m3i, CL25m1i, and CL32m1i (green). The overlapped IF-FISH signals are yellow. (C) IF-PRINS with anti-αCenH3 raised in rabbit 2 (red) and centromeric 19-bp box sequence (green). (D) Quantification of signal colocalization for ten representative images with calculated Manders coefficients (costes thresholding) seen as high overlapping ratios for both channel pairs (R-G represents overlapping ratio of red vs. green signals; G-R is overlapping ratio of green vs. red signals); αCenH3 with αCenH3-associated monomers, CL16m1i, CL16m3i, CL25m1i, and CL32m1i (left panel) and αCenH3 with 19-bp box regions (right panel). Data are presented as mean ± SD (source data are listed in supplementary table 7). (E) Dual-color fiber-IF/FISH using anti-αCenH3 (red) and αCenH3-associated monomers CL16m1i, CL16m3i, CL25m1i, and CL32m1i (green) as probes. The plots below the images represent intensities of the IF (red) and FISH (green) signals. DNA was counterstained with DAPI (blue). Images were acquired with confocal microscopy and shown as z-stack projection. Scale bar = 5 μm.

To provide a more accurate estimation of αCenH3/TRs colocalization, the quantification of IF (anti-αCenH3) and FISH signals (all TRs probes) on the original confocal images of ten interphases was performed. Based on the quantification, the IF signals overlapped average 78% of FISH signals, whereas FISH signals coincide with 83% of IF signals (fig. 7D, left and supplementary table 7). We suppose that a minor fraction of αCenH3-specific domains, which are not overlapped with FISH signals, originate from αCenH3 domains enriched in monomeric TR type of CL16m2i which was absent from analyses or some other sequence(s)-containing 19-bp box. Thus, in order to confirm our assumption that αCenH3 associated sequences contain the 19-bp box, combined IF with anti-αCenH3 and primed in situ labeling (PRINS) assay using the 19-bp box as primer sequence was conducted. Although PRINS showed lower brightness of signals in contrast to FISH signals, which is probably due to a different methodology, it can be observed that the incidence of signal overlap is higher in comparison to IF/FISH experiment (fig. 7C vs. fig. 7B). This is especially evident in quantification graph of IF/PRINS overlapping, where the αCenH3 IF signals match >95% of the 19-bp box PRINS signals (fig. 7D right panel and supplementary table 7). To examine organization of αCenH3-specific domains and selected TR repeats at higher resolution, sequential detection of αCenH3 and Cl25m1, Cl16m1i and Cl16m3i/Cl32m1i sequences on stretched chromatin fibers was performed. These results proved that αCenH3-associated DNA coincides with αCenH3 domains,

following interspersed organization of αCenH3 centromere interrupted by αCenH3-free domains (fig. 7E).

To exclude the possibility that αCenH3-associated TRs, detected as sequences with relatively low enrichments, could be the result of a centromere inclination to accumulate any repetitive sequences a control test was performed. Satellite DNA of similar genome abundance that was not found to be enriched in the ChIP-seq analysis (marked in fig. 5) was selected for IF/FISH. Simultaneous detection of αCenH3 and nonenriched satDNA revealed that the regions of satellite DNAs did not coincide with αCenH3 clusters (supplementary fig. 9). In addition, it can be observed that in contrast to αCenH3-associated TRs, noncentromeric satellite DNA showed a relatively comparable genome abundance that was estimated in WGS analysis (fig. 5), which further confirms the previous conclusions on abundance discrepancy of αCenH3-associated TRs in WGS/input data.

Prediction of αCenH3 Centromeric Repeats in *M. incognita*-Related Species

Since αCenH3 turned out to be conserved in all analyzed *Meloidogyne* species, we wondered whether the 19-bp box-containing sequences, as αCenH3-associated DNA in *M. incognita* remained conserved in other *Meloidogyne* genomes. We hypothesize that short TRs with conserved 19-bp box-containing monomers in other *Meloidogyne* species could disclose putative cenDNA regions of the corresponding species. For this analysis, we used publicly available Illumina WGS databases for two closely related species, *M. arenaria* and *M. javanica* (Blanc-Mathieu et al. 2017). Using as a criterion short TRs with the conserved 19-bp box, clustering of *M. arenaria* and *M. javanica* WGS reads followed by cluster annotation with conserved the 19-bp box was performed. The results revealed the appearance of the 19-bp box in repetitive form in the majority of contigs in clusters CL8 and CL7 in *M. arenaria*, and *M. javanica*, respectively (supplementary fig. 10A and B). Using TRF pipeline, contigs with repeated organization of the 19-bp box were subjected to monomer unit extraction. Consensus sequences of extracted monomers were compared with αCenH3-associated monomers from *M. incognita*. The alignments show that the 19-bp box-associated monomers from *M. arenaria* and *M. javanica* grouped with all TR families of *M. incognita* (fig. 8A). Moreover, sequence comparison revealed that monomers remained

almost completely conserved (94–100%) among these closely related species (fig. 8B). Monomer variants with the conserved 19-bp box and 45-bp monomer length specific for *M. arenaria* and *M. javanica* only (16/32m1-a and 16/32m1-j) were also identified (fig. 8A). In addition to the conserved sequence features of putative αCenH3-associated monomers, mapping of these monomers to contigs from *M. arenaria* and *M. javanica* revealed organization previously detected in *M. incognita*, with monomeric and mosaic TR arrays embedded in unrelated sequence environment (supplementary fig. 10A and B). To confirm presumption that 19-bp box-containing TRs are associated with αCenH3 in *M. arenaria* and in *M. javanica*, combined IF using anti-αCenH3 and PRINS using the 19-bp box sequences as primer was performed on cytosmears. Although it was even more difficult than in *M. incognita* to obtain cytological preparation in these species due to their tetraploidy, similarly as in *M. incognita* the results showed high coincidence of overlapped αCenH3/19-bp box signals indicating αCenH3 deposition on their chromosomes as well as association of αCenH3 with 19-bp box-containing TRs (supplementary fig. 11).

B

CL25m1	98,7
CL16m1	98,8 - 100
CL16m2	94 - 98
CL16m3	97,8 - 100
CL32m1	95,5 - 97,7
16/32m1	97,8

C

	61bp-h	84bp-h
61bp-h		40,4
84bp-h	40,4	
CL25m1i	39,7	48,8
CL16m1i	42,1	51,1
CL16m2i	41,2	33,3
CL16m3i	47,5	32,1
CL32m1i	45,1	38,0
16/32m1	44,2	33,3

Fig. 8. Candidates for αCenH3 centromeric sequences in *Meloidogyne incognita*-related species. (A) Alignments of consensus monomers sequences (CL25m1, CL16m1, CL16m2, CL32m1, and 16/32m1) extracted from clusters with tandem repeated arrays containing the conserved 19-bp box from *M. arenaria* (a) and *M. javanica* (j) in comparison to *M. incognita* (i). 61 bp-h and 84 bp-h represent monomers from tandem repeats containing the 19-bp box found in *M. hapla* assembled genome. The conserved 19-bp box is indicated within the gray shaded areas. (B) The percentage of sequence identity among consensus monomers in MIG species. (C) The percentage of sequence identity of monomers from *M. hapla* in comparison to monomers from MIG species.

Considering that our analyses of CenH3s showed almost completely conserved αCenH3 protein sequence and highly transcribed αCenH3 gene in nonreproductive stages (eggs and juveniles) of distant meiotic *M. hapla*, we asked whether its genome also comprises putative αCenH3-associated sequences. We did not apply the same clustering strategy done for *M. arenaria* and *M. javanica* because Illumina WGS data for *M. hapla* were not public available. Instead, the 19-bp box was mapped to the assembled *M. hapla* genome, and monomers were extracted from detected TR arrays. Two types of short TRs, composed of 61- and 84-bp long monomers, were associated with the 19-bp box (fig. 8A and supplementary fig. 10C). In contrast to putative αCenH3-associated monomers in *M. arenaria* and *M. javanica* which share high sequence identity (95–100%) with *M. incognita*, *M. hapla* putative αCenH3-associated monomers except conserved 19-bp box exhibit low-sequence identity in comparison to MIG species (32–51%) (fig. 8C).

Discussion

The availability of sequenced genomes and complex species evolution makes *Meloidogyne* an ideal system to study evolution of CenH3 proteins. We identified 21 CenH3 proteins in the three closely related mitotic *M. incognita*, *M. arenaria*, *M. javanica* (MIG) species and in the distantly related meiotic parthenogenetic species M *hapla*. Interestingly, phylogenetic analysis suggests the presence of two polyphyletic groups of CenH3s, and multiple copies of CenH3 genes in all analyzed species. The abcCenH3 group comprises rather divergent sequences with features of H3 variants but with relatively low HFD sequence identity related to the canonical H3. In the contrast, αβγδεCenH3s clade represents more homogenous group of sequences with identity to H3 considered as a common between H3 and CenH3s in many organisms analyzed so far (Malik and Henikoff 2003). These observations suggest independent evolution of two CenH3 groups from H3, indicating complex pattern of CenH3s in *Meloidogyne* species. The existence of multiple CenH3 candidates with polyphyletic origin in *Meloidogyne* species raises the question of their classification and function, The presence of two CenH3 genes is not uncommon in plant genomes (Kawabe et al. 2006; Moraes et al. 2011; Sanei et al. 2011; Finseth et al. 2015; Ishii et al. 2015; Neumann et al. 2015). In contrast, occurrence of paralogs in animals is considered to be a rare event. However, the recent comprehensive study of many high-quality sequenced genomes of Drosophila species revealed multiple copies of Cid histone (Drosophila CenH3) (Kursel and Malik 2017). The similar phenomenon was found in genomes of mosquitoes, where CenH3 paralogs evolve under different selective constraints, and have been coretained for over 150 My (Kursel et al. 2020). Interestingly, among detected CenH3s in *Meloidogyne*, only αCenH3 showed abundant expression in all analyzed species regardless of mode of reproduction (mitotic or meiotic), whereas the other CenH3s were dropped to a relatively low level of transcription. Anticipating that αCenH3 represents centromeric protein in *Meloidogyne* species, we focused our analyses on αCenH3 and on monophyletic group of βγδε CenH3s closely related to αCenH3s.

To understand evolution of αβγδε CenH3s in the selected *Meloidogyne* species, their complex species evolution history should be considered. MIG species have been determined as polyploids formed by recent and multiple interspecific hybridization

events (Castagnone-Sereno et al. 2013). They reproduce exclusively asexually by mitotic parthenogenesis. On the contrary, *M. hapla* is a diploid species, and reproduces asexually by meiotic parthenogenesis, although alternatively can also be sexual (Castagnone-Sereno and Danchin 2014). The most parsimonious scenario of CenH3 evolution in the analyzed species is therefore based on the integration of results obtained in this work together with previous data on species evolution (fig. 9). The phylogenetic analysis of CenH3 sequences strongly suggests that CenH3 gene has undergone one duplication in an ancestral sexual species, a progenitor of MIG species and of *M. hapla*, resulting in an appearance of αCenH3 and (βγδε)CenH3 ancestral genes, (ANC [α] and ANC [βγδε] in fig. 9). Earlier phylogenetic studies based on different mitochondrial and nuclear markers revealed significant distance between MIG species and *M. hapla*, estimating their separation for 17–50 Ma (reviewed in Castagnone-Sereno et al. 2013). Unexpectedly, during this time the αCenH3 gene evolved under strong purifying selection in all analyzed species, resulting in almost completely conserved αCenH3 proteins in mitotic MIG species as well as in the distant meiotic *M. hapla*. So far, the only example of nearly identical protein sequences of CenH3s in related species were found in plant genus Secale (Evtushenko et al. 2017). The other copy of the gene, the ancestral βγδε CenH3, evolved rapidly into four different but related proteins: β, γ, δ, and εCenH3 (fig. 9). β, γ, and δCenH3s are specific for MIG species, whereas εCenH3 was found exclusively in *M. hapla*. In mitotic *M. incognita*, the αCenH3 showed high expression and chromosomal deposition on all chromosomes of the complement, whereas β and γCenH3s exhibit low expression and absence of chromosomal deposition. In support, two other MIG species, *M. arenaria* and *M. javanica* also showed the chromosome deposition of αCenH3 and silencing of β, γ, and δCenH3s. This indicates centromere competence of αCenH3 and most probably the loss of centromere-associated function for β, γ, and δCenH3s in exclusively mitotic MIG species. In support, detected dominant expression of αCenH3 in contrast to εCenH3 in the nonreproductive stages of *M. hapla* where mitosis is expected to occur, speaks in favor of dominant role of αCenH3 in mitotic cells of the meiotic *M. hapla*. The comparative analyses of MIG genomes and ITS markers suggest that polyploid genomes of MIG species result from additive interspecies hybridization between related parental sexual taxa (Hugall et al. 1999; Blanc-Mathieu et al. 2017; Szitenberg et al. 2017). Moreover, studies of mitochondrial DNA and genetic test of allelic sequence divergence suggest that hybridization events included

in formation of MIG species have a recent origin (Giorgi et al. 2002; Lunt 2008; García and Sánchez-Puerta 2015; Blanc-Mathieu et al. 2017). In line with this scenario, MIG species possess multiple copies of αCenH3 and one copy of β, γ, and δCenH3s, whereas diploid *M. hapla* has only one copy of αCenH3 and of εCenH3 (fig. 9). Moreover, the copy number of αCenH3s and the presence of β, γ, and δ CenH3s in MIG species are consistent with the estimated ploidy levels based on protein-coding sequences (CDSs) data mapping in MIG genomes (Blanc-Mathieu et al. 2017). In addition, one copy of both αCenH3 and εCenH3 in *M. hapla* are corroborated by CDS mapping where one single locus was detected for the particular gene in *M. hapla* genome (Blanc-Mathieu et al. 2017). The observed pattern of CenH3s evolution in MIG species is in accordance with proposed species evolution characterized by independent evolution of α, β, γ, and δCenH3s in parental sexual taxa followed by polyploidization as result of recent species hybridization (fig. 9). In the light of the rapid evolution of β, γ, and δCenH3s, the evidence that these CenH3s remain almost completely conserved among MIG species suggests that MIG interspecific hybridization is a relatively recent event. Consequently, it is the most probable that β, γ, and δCenH3s diverged in sexual progenitors and that their redundancy in the nascent MIG coincided with recent species hybridization followed by mitotic mode of reproduction. Otherwise, if they were subjected to a long period of nonfunctionality, accumulation of random mutations and pseudogenization would be expected. The putative role of βγδ and εCenH3s in sexual progenitors has been additionally supported by the fact that they have been retained in both, the extant MIG and *M. hapla* species since their divergence for about 17–50 Ma (Castagnone-Sereno et al. 2013). In contrast to αCenH3 conservation, β, γ, δ, and εCenH3s evolved under the purified selection with the positive evolution trend on the first several amino acids of the N-terminal tail. The similar selection pattern was found in CenH3 proteins of holocentric chromosomes among the related species of the nematode genus Caenorhabditis which reproduces sexually, by asymmetric meiosis (Zedek and Bureš 2012). Rapid evolution of CenH3s with positive selection has been detected so far exclusively in sexual lineages with asymmetric meiosis where centromere drive occurs (Henikoff et al. 2001; Talbert et al. 2004; Hirsch et al. 2009; Schueler et al. 2010; Zedek and Bureš 2012; Finseth et al. 2015).

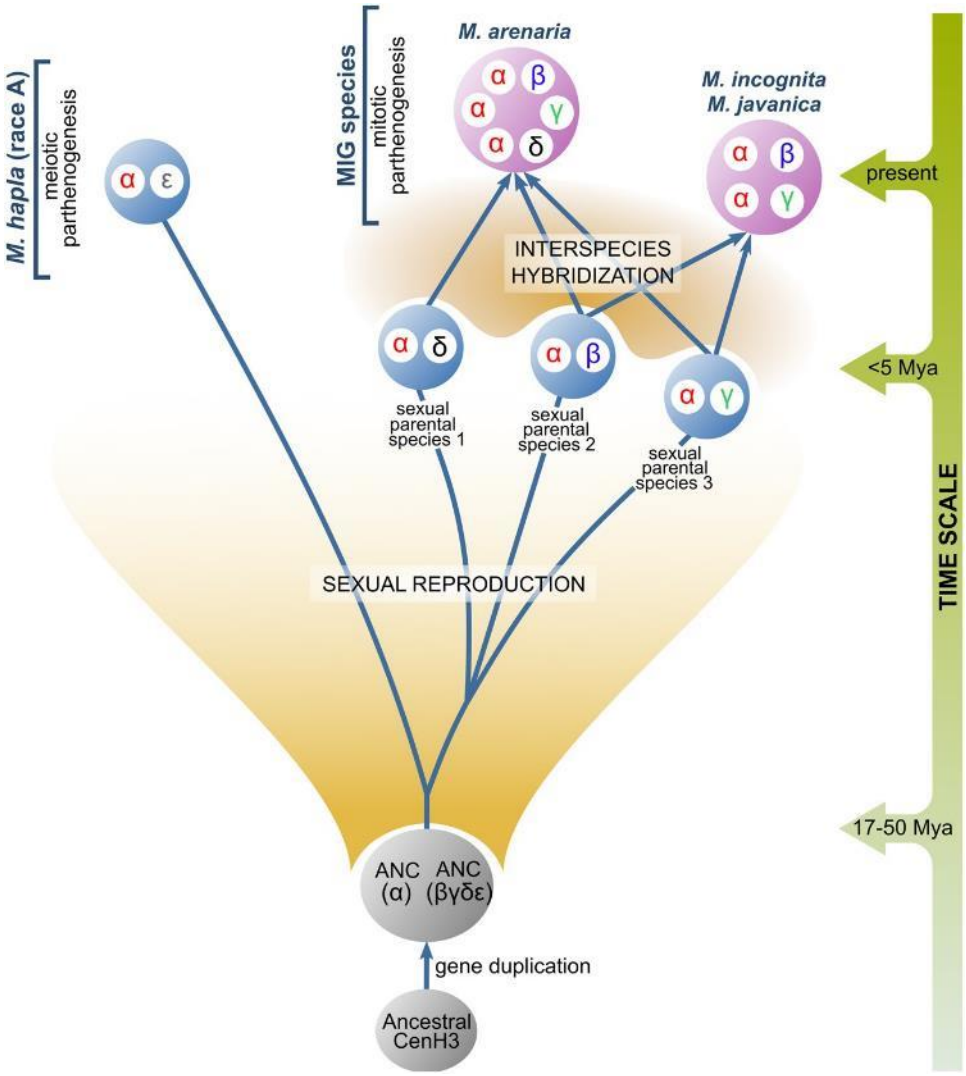


Fig. 9. The most parsimonious evolution of α, β, γ, δ, and εCenH3 variants in *Meloidogyne* species (*M. hapla* [race A], *M. incognita*, *M. arenaria*, *M. javanica*). Presented species evolution with evolutionary time scale was based on different previous studies. Separation between MIG species and *M. hapla* was estimated between 17 and 50 Ma (reviewed in Castagnone-Sereno et al. 2013). Interspecies hybridization between related sexual parental species was proposed in formation of MIG species (Blanc-Mathieu et al. 2017; Szitenberg et al. 2017). Interspecific hybridization is estimated as relatively recent event (<5 Ma) (Giorgi et al. 2002; García and Sánchez-Puerta 2015; Blanc-Mathieu et al. 2017). ANC(α) represents the ancestral αCenH3, ANC(βγδε) represents the ancestral CenH3 of βγδε variants.

The major question raised by our observation is why the two subgroups of CenH3s, αCenH3, and βγδεCenH3s, have completely different evolutionary rate. We propose that different evolutionary dynamics of analyzed *Meloidogyne* CenH3s might be due to distinct requirements posed on αCenH3 in contrast to β, γ, δ, and εCenH3s in *Meloidogyne* centromere. Based on recent comprehensive evolutionary studies of CenH3 duplication in Drosophila and mosquito species the authors suggested that gene duplications of CenH3 could be required for multiple centromeric functions, for example, in mitosis versus meiosis (Kursel and Malik 2017; Kursel and Malik 2019). The observed evolution pattern characterized by purified selection with positive trends of β, γ, δ, and ε CenH3 could be predicted by the centromeric drive model which implies asymmetric meiosis and sexual reproduction. In that case, β, γ, and δCenH3 could have evolved in parental sexual lineages as a consequence of centromere drive during meiosis and become redundant in MIG species due to transition from sexual to mitotic parthenogenesis. It has been considered that transition from sexual reproduction to mitotic parthenogenesis in MIG species correlates with recent species hybridization (Lunt 2008; Blanc-Mathieu et al. 2017).in MIG species. In support, the major sperm protein which is a meiotic-specific gene, shows no increase in evolutionary rate nor change in substitution pattern in the mitotic *Meloidogyne* taxa, indicating that the locus has been maintained by selection (Lunt 2008). In MIG species there are also no morphological abnormalities in the sperm development, and insemination still occurs sporadically but without fertilization (Triantaphyllou 1981). Given that β, γ, and δCenH3s show no signs of pseudogenization and sequence degeneration which could be caused by random mutations, it is likely that these CenH3s, similarly as in the case of the sperm protein represent a “meiotic relict” in exclusively mitotic MIG species. Analogously, the αCenH3 conservation through the same period of the time could be due to the possible subspecialization. The main prediction of centromere drive is that CenH3 coevolves with cenDNA in order to suppress the deleterious effect of rapidly evolving cenDNAs in meiosis. In that case, the possible subspecialization of αCenH3 for mitosis could release αCenH3 from the adaptive conflict imposed in meiosis. This could result in strong amino acid conservation of αCenH3 among distant species. In support to different CenH3 functions in cell divisions, holocentric nematode *C. elegans* harbors two CenH3-related proteins, HCP-3 and CPAR-1, which indicates centromere functional specialization. HCP-3 has been proven to be essential for mitosis but is not required for meiotic kinetochore formation or chromosome segregation (Monen et al.

2005). Although the functional importance of Cpar-1 is not completely understood, its enrichment on meiotic chromosomes was documented (Monen et al. 2015). In addition, recent data on Arabidopsis suggest that in species with a single-copy CenH3 gene, one protein probably must be customized for different centromere functions (Ravi et al. 2011). The experiments showed that impaired CenH3 lost its function in meiotic centromeres of Arabidopsis, whereas the C-terminal region and HFD were sufficient for centromere function during mitosis (Lermontova et al. 2006; Lermontova et al. 2011; Ravi et al. 2011).

Predicted rapid evolution of CenH3s as a response to the deleterious effect of extremely divergent cenDNAs motivated the investigation of the genetic landscape features of *M. incognita* centromere determined by long-term conserved αCenH3. The αCenH3-associated cenDNA found in *M. incognita* is organized in a form of short arrays of tandem repeats (TRs), up to 1 kb in length, composed of five different families based on 50- to 80-bp monomers. Although presence of tandem repeats, in the form of long arrays of satDNAs is a common characteristic found in many monocentric species (Plohl et al. 2014), holocentric organisms investigated so far exhibit different patterns of cenDNAs. The robust ChIP CenH3-based studies on the whole-genome scale showed that *C. elegans* holocentromeres do not coincide with satDNAs but do coincide with nonspecific binding sites for multiple transcription factors (Steiner and Henikoff 2014). The native ChIP-seq of the parasitic nematode Ascaris also suggested absence of centromere-specific DNA sequence (Kang et al. 2016). So far, the only example of holocentric chromosomes which possess satDNAs as the centromere-specific sequence has been found in the plant *Rhynchospora* (Marques et al. 2015; Ribeiro et al. 2018). Comparative analyses of αCenH3-associated centromeric repeats revealed an exceptional feature in form of the 19-bp long-conserved box shared by extremely divergent monomers, suggesting selective pressure imposed on this sequence part regardless of the fast-evolving nature of repetitive DNAs. In addition to high GC content, the 19-bp conserved box exhibits a specific potential to form a stable dyad structure. In spite of enormous diversity in cenDNA detected in many species, the recent study of structural features of centromeric satDNAs from diverse eukaryotes pointed out two major characteristics to be crucial for putative cenDNA: a specific DNA sequence as a binding site for proteins and/or a specific feature of the sequence itself such as DNA secondary structure (Kasinathan and Henikoff 2018). Several studies

have shown that centromeric satDNAs may form various types of non-B-form including single-stranded DNA, hairpins, R-loops, and i-motifs (Garavís et al. 2015; Kabeche et al. 2018; Kasinathan and Henikoff 2018). Consistent with this, if the conserved 19-bp box is a binding site for αCenH3 in *M. incognita*, primary as well as secondary structure of the 19-bp box could be crucial for its binding capacity. In contrast to AT-rich DNA which is common feature in centromeres (Talbert and Henikoff 2020) centromeric TRs in *M. incognita* show the extremely high GC content compared with the high AT composition of the genome (Abad et al. 2008). Therefore, we proposed that these unique sequence features, such as primary and secondary structures of αCenH3-associated DNA in the form of the 19-bp box incorporated into GC-rich short TRs, act in concert to ensure the faithful formation of an αCenH3 centromere in *M. incognita*. Another important finding that arose from this work is the existence of completely preserved αCenH3 centromere associated TRs in terms of sequence and organization in closely related *M. incognita*, *M. javanica*, and *M. arenaria*. The colocalization of αCenH3 and the 19-bp box in *M. arenaria* and *M. javanica*, similar to *M. incognita*, suggested the preservation of the αCenH3 centromere in protein and DNA aspects. Moreover, divergent TR short arrays with monomers containing almost completely conserved 19-bp box were shown in the distant *M. hapla*, implying the functional constraints imposed on this sequence part even in distantly related species. Concerning cenDNA with conserved sequence features, the recent study of early-diverging fungi showed the presence a 41-bp unique DNA motif in all nine core centromeres which has been proposed as a binding site for some kinetochore proteins (Navarro-Mendoza et al. 2019). The most prominent example of cenDNA sequence conservation is the CENP-B box, the conserved 17-bp long-sequence motif specific for alpha-satDNA in humans (Ohzeki et al. 2002) as well as in alphoid repeats in mammalian species (Alkan et al. 2011). This motif proved to be a binding site for centromeric protein CENP-B that is involved in kinetochore formation (Masumoto et al. 2004). Interestingly, interspecifically preserved motifs that probably evolve under functional constraints whose potential role(s) remain elusive were observed in many satDNAs, including in satDNAs of *Meloidogyne* species (Meštrović et al. 2006a, 2006b, 2013).

Regarding to chromosome organization of αCenH3 centromere in *M. incognita* an unusual pattern characterized by uneven distribution of αCenH3 among and along the

chromosomes has been shown. Immunofluorescence on prophase chromosomes and on extended chromatin fibers, revealed discontinuous pattern of αCenH3 domains separated by αCenH3-lacking chromatin. The observed αCenH3 distribution pattern can be defined as cluster-like centromeric organization. In more condensed metaphase chromosomes, αCenH3 encompasses the entire chromosome length in the form of abundant or discrete signals or exhibits extremely uneven distribution with highly abundant domains in different chromosome regions. The α-tubulin was observed to be mostly colocalized with αCenH3 domains thus indicating functional potential of αCenH3 centromere in mitosis. In contrast to the point centromere subunits in *C. elegans* (Steiner and Henikoff 2014) the observed cluster-like organization of *M. incognita* is similar to the nematode *Ascaris* where CenH3 is organized into 1–15 kb domains distributed across the chromosomes (Kang et al. 2016), Recent data on the nematode *C. elegans* (Buchwitz et al. 1999; Moore et al. 1999), and plant *Rhynchospora* (Marques et al. 2015) suggest different organization of CenH3 domains in mitotic and meiotic holocentromeres In addition, holocentromere of the plant *Cuscuta* showed CenH3 restricted only to one to three regions per chromosome, whereas the rest of the chromatin appeared to be devoid of CenH3 (Oliveira et al. 2020). Even more extreme situation has been revealed in holocentric insects characterized by complete loss of CenH3s in at least four lineages (Drinnenberg et al. 2014). Observations based on different holocentric species analyzed so far, including *Meloidogyne*, lead to the conclusion that in contrast to monocentromere, holocentromeres show greater flexibility in the organization of CenH3 domains at the chromosome level.

In conclusion, our study represents the first insight into the centromere evolution and composition in an exclusively mitotic species belonging to higher eukaryotes. By generating and analyzing CenH3s from different *Meloidogyne* species, we have for the first time demonstrated almost complete conservation of one CenH3 protein among distant animal species and hypothesized its subspecialization, presumably associated with mitosis. We confirmed that the TRs arrays with the conserved 19-bp box span almost the entire αCenH3 centromere and represent the underlying DNA sequence of *M. incognita* centromere. Moreover, conserved αCenH3 and αCenH3-associated DNA in the form of 19-bp box was found in related MIG species suggesting preservation of αCenH3 centromere across mitotic *Meloidogyne* species. Our study disclosed for the

first time a long-term conservation of CenH3 and its association with a conserved box regardless to highly evolved centromeric tandem repeats, thus suggesting the state where CenH3 and cenDNA achieved an equilibrium in which they can coexist for a long period of time. An exciting line of future investigation concerning specialization of CenH3s in *Meloidogyne* species would be to address the potential of the αCenH3 and εCenH3 in mitosis in comparison to meiosis in meiotic parthenogenetic *M. hapla*.

Materials and Methods

Nematodes

Meloidogyne incognita was cultivated on tomato (Solanum lycopersicum cultivar Saint Pierre) in greenhouse at 20 °C in laboratories from INRAE (Sophia Antipolis, France) and Agricultural institute of Slovenia (Ljubljana, Slovenia). Plants were inoculated with one to three second-stage juveniles per ml silver sand. Females or egg masses were harvested from roots under stereo microscope (SteREO Discovery.V20, Zeiss) and collected into an isotonic salt solution (M9 buffer). Egg masses were shaken in 15% bleach for 5 min to release eggs and eggs were isolated by successively passing through the sieves.

DNA and Protein Isolation

DNA was isolated from eggs using DNeasy Blood and Tissue Kit (Qiagen) in accordance to manufacturer's protocol and quantification was done by Qubit fluorimeter (Invitrogen). For protein isolation, eggs, J2, or females were transferred in cold RIPA buffer supplemented with 10 mM PMSF and cOmplete (Roche) protease inhibitors and homogenized in Dounce homogenizer with 10–15 strokes. The homogenate was incubated with rotation on 7 rpm for 2 h at 4 °C and centrifuged for 20 min at 12,850 g at 4 °C. Supernatant containing whole cell proteins was collected and stored at −80 °C. Protein concentration was estimated using Bradford assay.

Identification and Sequence Analyses of CenH3 Proteins

To identify CenH3 sequences in four *Meloidogyne* species (*M. incognita*, *M. arenaria*, *M. javanica*, and *M. hapla*) a nonredundant database of protein sequences generated from the automatic annotation of sequenced genomes available at INRA website (http://Meloidogyne.inra.fr/) and WormBase ParaSite (http://parasite.wormbase.org/, Howe et al. 2017) were used. BLAST search for CenH3 proteins was done using *C. elegans* H3 protein sequence (NCBI accession number P08898) as query. Among 23 detected CenH3 candidates two of them were truncated in N-terminal tail (supplementary fig. 1B). Assuming that these truncated copies are the result of

assembly/annotation error or represent CenH3 pseudogenes, we omitted them from the further analysis. The WormBase proteins and genes IDs with the list of CenH3s and corresponding species are available in supplementary table 1. Multiple alignments of CenH3 candidates were generated using MUSCLE with default parameters implemented in Geneious v9.1. The structural features of CenH3 candidates such as histone-fold domain (HFD), N-terminal tail, αhelix, loops, and C-terminus were defined in accordance with Malik and Henikoff (2003). CenH3 candidates were tested for diagnostic features in HFD which include longer loop1 region and absence of glutamine, phenylalanine, and threonine at positions 69, 85, and 118, respectively, in comparison to canonical H3 (Malik and Henikoff 2003). Neighbor-joining trees of CenH3 proteins and pairwise percent identity calculations were generated using Geneious v9.1. Bootstrap values were calculated from at 1,000 replicates. Phylogenetic trees were drawn and edited using the FigTree 1.4.4 software (Rambaut 2018).

Tests for Selective Pressure

CenH3 gene sequences (supplementary fig. 2) related to detected CenH3 protein candidates (supplementary table 1) were generated from full-length transcripts databases from WormBase ParaSite (http://parasite.wormbase.org/, Howe et al. 2017). Multiple alignments of CenH3 nucleotide sequences were done using MUSCLE algorithm with default parameters. Alignments were further refined manually and used for downstream analyses. To determine the selective pressures acting on a CenH3 genes using the nonsynonymous/synonymous substitution rate ratio ($dN/dS = \omega$), distance computation using Nei–Gojobori (Jukes–Cantor) substitution model implemented in MEGA version X was done (Kumar et al. 2018). All obtained comparison values ($dN/dS = \omega$) are shown in supplementary table 4. Generally, $\omega = 1$ indicates neutral selection, $\omega < 1$ purifying selection and $\omega > 1$ positive selection. If purifying selection is relaxed, ω tends to be elevated toward 1. To assess the positive selection at the level of individual codons, Mixed Effects Model of Evolution (MEME) model was used (Murrell et al. 2012). MEME allows ω to vary across both codons and branches and infers selective regimes independently for each codon of a given alignment pooling information over branches. MEME analyses with a significance level cutoff of 0.1, correspondingly were performed through the Datamonkey server

(http://datamonkey.org/). Analysis of the positive selection at individual codons was carried out on αCenH3 variants and also among different CenH3s using likelihood ratio test (LRT) values plotted against each codon site for visualization (supplementary table 5).

Expression Profile of CenH3 Candidates

To compare gene expression among CenH3s in analyzed species or developmental stages RNA-seq data from M. incognita (PRJEB8846, Danchin et al. 2013), *M. arenaria* (PRJEB8845, Blanc-Mathieu et al. 2017), *M. javanica* (PRJEB8843; Blanc-Mathieu et al. 2017), and *M. hapla* (PRJEB14142) were used. The relative expressions of CenH3 genes and reference gene Disu (Hu and DiGennaro 2019) were analyzed using Bowtie2 v.2.3.0 mapper (Langmead and Salzberg 2012). Single-end reads were mapped with parameters -a and –very-sensitive for each transcriptome separately to CenH3 genes. Hits were normalized with RPKM (reads per kilobase of transcript per million mapped reads) method. This approach takes in account different CenH3 variant length and size of RNA-seq libraries dividing CenH3 hits by number of mapped reads per million reads and gene length in kilobase.

Production of CenH3 Antibodies

Polyclonal IgG antibodies against αCenH3 were raised in rabbits using peptide KELPPVKMQQKRYHKKGC. Two another antibodies were raised in rabbits and in guinea pigs using two peptides specific exclusively for βCenH3 (the peptide CTNFPRQTARKRVF specific for βCenH3-1; and the peptide KNFATKSVAGPTTMNTG specific for βCenH3-2) and peptide specific for both, β and γ CenH3 (the peptide QQQNKIKAPGEGGSL specific for βγ CenH3). Selected peptides correspond to region of divergent N-terminal tails of CenH3s (fig. 1B) and meet the parameters (amino acid phosphorylation, glycosylation profile, and secondary structure; Parker et al. 1986) that were prerequisite for suitable antibodies production and specificity. Peptide synthesis, immunization, and peptide affinity purification were performed by Pineda Service (Berlin, Germany). The preimmune sera as well as the sera samples were tested during the immunization process by Western blot monthly, to monitor the immune response. Immunizations were stopped after 90–120 days and

affinity purification of the monospecific IgG fraction of CenH3 antisera was performed. Purified monospecific IgG fraction was concentrated 25× using Amicon Ultra-0.5 centrifugal filter device (Merck), and used in all downstream applications.

Western Blot

For Western blot 20 µg/reaction of whole protein extract from eggs, J2 or females were denatured in 1xLaemmli buffer (50 mM Tris–HCl pH 6.8, 10% glycerol, 2% SDS, 0.005% bromophenol blue) with 0.1 M dithiothreitol (DTT) at 65 °C for 15 min. The protein samples were separated on 4–20% Mini-PROTEAN TGX (Bio-Rad) SDS–PAGE gels at 200 V for 30 min followed with protein transfer for 40 min at 200 mA onto Amersham Protran 0.2-µm nitrocellulose membrane (GE Healthcare Life Sciences). Membranes were simultaneously incubated for 1 h in blocking solution of 5% BSA in TBST buffer (20 mM Tris, 150 mM NaCl, pH 7.6, 0.1% Tween 20) followed by overnight incubation at 4 °C with CenH3 rabbit or guinea pig polyclonal primary antibodies (dilution 1:500). HRP-linked goat antirabbit (Cell Signaling Technology 7074) or antiguinea pig (Invitrogen, A18769) antibodies diluted 1:2,000 were used as secondary antibodies. Dilution of primary and secondary antibodies was performed in TBST buffer with 5% BSA. Signals were detected using the Pierce ECL Western Blotting substrate (Thermo Scientific) and Amersham Hyperfilm ECL X-ray films (GE Healthcare Life Sciences). The α-tubulin mouse monoclonal antibody (Sigma Aldrich, T6199) used for combined α-tubulin and αCenH3 immunofluorescence assay was tested by Western blot as described above. H3K9 polyclonal antibody (Abcam, ab8898) was used as a positive control in Western-blot experiments.

Peptide Dot Blot

The assay was performed using specific antibodies and dilutions of several peptides which were plotted onto nitrocellulose membrane. The specificity of antibodies we produced against CenH3s was tested on specific and nonspecific peptides. Synthetized 14–18mer peptides were dissolved in PBS to 1 mg/ml and then further diluted to concentration of 1, 0.1, and 0.01 µg/µl. The series of peptide dilutions in form of 2 µl spots were applied to the nitrocellulose membrane; the membrane was dried for 30 min and blocked in 5% BSA in TBST buffer for 1 h with mild shaking at RT. Affinity

purified monospecific IgG fractions of anti-CenH3s were used as primary antibodies, diluted 1:500 in blocking buffer. For βCenH3-1, βCenH3-2, βγCenH3 antibody testing, two rabbits or three guinea pigs antibodies were pooled together when testing rabbit of guinea pig antibodies, respectively. After 1 h incubation at RT, membranes were washed 3 × 5 min with TBST and incubated with secondary antibody (HRP-conjugated antirabbit; Cell Signaling Technology, 7074) or antiguinea pig (Invitrogen, A18769) antibody diluted 1:1,000 in blocking buffer for 1 h at RT followed by 3 × 5 min washes with TBST. The final detection was carried out as described for the Western blot.

Microscope Slide Preparation

Slides were prepared from isolated reproductive tissue (ovaries and uterus) of *M. incognita*, *M. arenaria*, and *M. javanica* females using cytospin technique. Samples were collected in 10 µg/ml colcemid (Roche) and pierced using needle. The ovaries and uteri were isolated and incubated for 1 h to overnight at 4 °C. Samples were washed with PBS and then mixed for 30 s with microtube homogenizer. Suspension was transferred to Dounce homogenizer and tissue parts were broken with 30 strokes of pestle A followed by straining through 100- and 40-µm cell strainers. Volume of suspension is adjusted with PBS up to 400 µl for loading into one Cytospin funnel that corresponds to five to ten females per one coated Cytoslide (Shandon, ThermoFisher Scientific). Slides were spun for 10 min at 1,200 rpm by Cytospin 4 cytocentrifuge (Shandon, ThermoFischer Scientific), dried, and fixed by immersing into ice-cold fixative for 20 min, completely dried and stored. Several fixatives were tested and methanol: acetone (1:1) incubation for 20 min at −20 °C showed the best results in IF and IF-FISH analyses. For chromatin fiber preparation the best results were obtained following the protocol described in Frum et al. (2013) with some modifications. Briefly, slides were dried after cytospinning and incubated with 15 µl of freshly prepared mild SDS lysis buffer (0.2% SDS, 200 mM Tris–HCl pH 7.5, 50 mM EDTA, 1 mM PMSF) by covering it with 18 mm square coverslips. Lysis reaction was performed at room temperature for 30 min and coverslip was carefully removed with a blade. Slides were then fixed in methanol:acetone (1:1) for 20 min at −20 °C, completely dried, and stored at −80 °C until use.

DNA Probe Preparation

FISH probes for centromere candidates were obtained by PCR labeling with biotin-16-dUTP (Jena BioScience) using genomic DNA. Primers for centromeric candidates were designed based on monomer sequences enriched in ChIP analyses (Cl25m1i, Cl16m1i, Cl16m2i, and Cl16m3i) using Primer 3 (Rozen and Skaletsky 2000); the primer sequences are listed in supplementary figure 8A. PCR reactions were performed in 25 µl reaction volume containing 0.02 ng of gDNA, 0.1 µM primers, 2.5 mM MgCl2, 1× Green GoTaq Reaction Buffer, 0.1 mM dNTPs, and 0.5 U of GoTaq G2 DNA Polymerase (Promega). Thirty-five amplification cycles (20 s at 95 °C, 20 at 58 °C annealing temperature and 40 s at 72 °C) were run. ChIPped-DNA was labeled using random priming approach with Klenow fragment in accordance to manufacturer's protocol (New England BioLabs). Characteristic ladder-like profile with expected fragment sizes was taken as proof for specificity of hybridization probes (supplementary fig. 8B).

DNA Dot Blot

For genomic DNA, 50, 100, and 200 ng were spotted onto positively charged nylon membrane (Roche). PCR products corresponding to CL25m1, CL16m1, and CL16m3/32m1 fragments were spotted in the amounts of 0.5, 1, 2, 4, and 8 ng. Hybridization was done in 250 mM phosphate buffer pH 7.2, 1 mM EDTA pH 8, 20% SDS, and 0.5% Blocking Reagent (Roche) with 50 ng of biotin-labeled probes at 65 °C with agitation overnight. Posthybridization washes were done in 20 mM Na2HPO4, 1 mM EDTA, and 1% SDS 3 × 20 min at 62 °C. Detection was carried out using streptavidin-AP-conjugate (1:5,000, Roche) followed by chemiluminescence with AP substrate CDP-Star (1:50, Roche). Dot blot intensities were compared using ImageJ with measurement of mean gray values that were inverted and normalized for background.

Immunofluorescence

Affinity purified monospecific IgG fractions of anti-CenH3 were concentrated using Amicon Ultra-0.5 Centrifugal Filter Unit (Merck) with 30 kDa cutoff. Slides with

cytosmear or chromatin fibers were blocked with 2.5% BSA in PBST (PBS, 0.2% Tween 20) and incubated with anti-CenH3 (1:400 dilution) at 37 °C overnight. After 3 × 5 min washes in PBST, slides were incubated with secondary Alexa594 antirabbit (Abcam, ab150080) or Alexa488 antiguinea pig (Abcam, ab150158) antibodies, diluted 1:1,000 dilution in blocking solution for 1 h at 37 °C. After two washes in PBST for 5 min and one wash in PBS for 5 min, the slides were counterstained with DAPI or continued with the FISH protocol. In double immunostaining with two anti-CenH3, the primary as well as secondary antibodies were incubated together.

For combined α-tubulin and αCenH3 immunofluorescence, slides were after O/N incubation with αCenH3 antibody the slides were washed 3 × 5 min with PBST and then incubated with α-tubulin mouse monoclonal antibody (Sigma–Aldrich, T6199) diluted 1:1,000 in PBST with 2.5% BSA for 3 h at 37 °C. After 3 × 5 min washes in PBST and additional blocking for 1 h at 37 °C, slides were consecutively incubated with secondary Alexa594 antirabbit (Abcam, ab150080) and CF488 antimouse (Sigma–Aldrich, SAB4600035) antibodies followed by washes and DAPI staining as described above.

Fluorescence In Situ Hybridization

For combined detection of the αCenH3 and centromeric candidates, immunodetection procedure was followed by FISH. After IF detection and washing, slides were immediately pretreated for FISH washing in 45% acetic acid for 10 min and in 2xSSC for 5 min. After RNase A treatment for 30 min at 37 °C slides were washed 3 × 5 min with PBS and fixed with 1% formaldehyde in PBS with 50 mM MgCl2. After washing 2 × 5 min with PBS, slides were dehydrated in a series of cold ethanol. Denaturation was carried out in 70% formamide in 2xSSC at 70 °C for 2 min and slides were dehydrated and air dried. Lyophilized-specific probe (100 ng/slide) was denatured at 75 °C for 5 min in 15 µl of hybridization buffer (60% formamide in dextran sulfate buffer [20% DeSO4, 4xSSC, 50 mM Na-phosphate pH 7.0]) and chilled on ice. Hybridization was performed at 37 °C overnight. Posthybridization washes were carried out with 50% formamide in 2xSSC for four times during 5 min at 37 °C. Slides were blocked with 5% Blocking Reagent (Roche) in 4xSSC. Immunodetection was performed with fluorescein avidin D and biotinylated antiavidin D system (Vector Laboratories). Slides were

counterstained with DAPI, dried, and embedded in Mowiol 4-88 mounting medium (Sigma–Aldrich). To minimize nonspecific staining slides with chromatin fibers were incubated in Image-iT FX Signal Enhancer (Invitrogen) for 30 min.

Primed In Situ Labeling

For combined detection of the αCenH3 and 19-bp box sequence immunodetection procedure was followed by primed in situ labeling (PRINS). Slides were pretreated and denatured as for FISH experiments. The reaction mixture was prepared in 50 µl containing 2 µM primer (19-bp box; TCGGGCCTTCGGCCCTCGC), 2.5 µM MgCl2, 150 µM each of dATP, dCTP, dGTP, 96 µM dTTP and 54 µM biotin-16-dUTP, 1 U of GoTaq G2 DNA Polymerase (Promega), and 1× Colorless GoTaq Reaction Buffer (Promega). On each prewarmed slide, 25 µl of prepared mixture was applied, covered with coverslip, sealed, and continued to heat at appropriate annealing and elongation temperature of 65 °C for 30 min. Reaction was stopped by washing in 50 mM NaCl, 50 mM EDTA, pH 8 buffer for 5 min at 65 °C followed by 3 × 5 min washes in 4× SSC with 0.05% Tween 20. Immunodetection and DAPI staining were afterward performed in the same manner as for FISH.

Images Processing and Quantification

Microscopic images were recorded using confocal laser scanning microscope Leica TCS SP8 X (Leica Microsystems) equipped with an HC PL APO CS2 63×/1.40 oil objective, 405 nm diode laser, and a supercontinuum excitation laser (Leica Microsystems). Images were acquired as z-stacks with five slices and average step size of 0.5 µm per cytosmear. Each fluorochrome was capture separately and images were merged and analyzed using Image J and Adobe Photoshop software. Images quantification was done using CellProfiler (https://cellprofiler.org) with align and measure colocalization modules. Ten nuclei were selected as separate regions of interests on original images acquired with confocal microscopy where grayscale separated red and green channels were analyzed. Manders coefficients with Costes automated thresholding were calculated for channel interrelationship and all values were shown on graphs using GraphPad Prism version 8.

Native Chromatin Immunoprecipitation Followed by Sequencing (ChIP-Seq)

Native ChIP was performed, with some modifications, according to protocol previously described for nematode *C. elegans* (Steiner and Henikoff 2014). Briefly, approximately 500 mg of frozen *M. incognita* eggs were ground using liquid nitrogen. Suspension was homogenized for 2 min with pestle A and 4 min with pestle B in 3 ml of ice-cold buffer (15 mM Tris–HCl pH 7.5, 2 mM MgCl2, 340 mM sucrose, 0.2 mM spermine, 0.5 mM spermidine, 0.1% Triton X-100, and 0.5 mM PMSF) using Dounce homogenizer. Prolonged homogenization step has previously shown to be crucial for higher chromatin yield in *M. incognita* (Perfus-Barbeoch et al. 2014). Cellular debris was removed by spinning for 2 min at 100 g. Supernatant was centrifuged at 1,000 × g for 10 min and nuclei were gently resuspended in 250 µl of buffer (10 mM Tris–HCl pH 7.5, 2 mM MgCl2, 0.5 mM PMSF) and incubated for 5 min at 37 °C with addition of 2 mM CaCl2. Different MNase (Thermo Scientific, Cat No. 88216) concentrations were tested and optimal digestion with prevalent fraction of mononucleosome ~150 bp was obtained with 0.2 U/µl eggs and 0.4 U/µl eggs for 2 min. The reactions were stopped by adding EDTA (final concentration of 20 mM). Chromatin was solubilized by cavitation using needle extraction (ten times with 26 gauge) and suspension was centrifuged at 1,000 × g for 5 min. The supernatant containing well-digested chromatin was used for ChIP (supplementary fig. 5A).

ChIP was done using Dynabeads Protein A Immunoprecipitation Kit (Invitrogen) in accordance to manufacturer's protocol, with some modifications. Three ChIP experiments with chromatin from two different MNase digestions using two different αCenH3 antibody concentration (10 and 30 µg) were performed (supplementary fig. 5B). Beads were first washed with PBS followed by binding of αCenH3 antibody in 200 µl of Ab Binding and Washing Buffer for 2 h at 4 °C. Beads–antibody complex were washed and separated using a magnetic rack. For input control sample, 10% of starting chromatin was stored at 4 °C until the last step of elution. ChIP reactions were done with 50 µl of beads–antibody complex and 200 µl of isolated chromatin fraction (~6 µg). Chromatin was diluted 1:2 with ChIP dilution buffer (50 mM NaCl, 20 mM Tris–HCl, pH 7.5, 5 mM EDTA, 0.2 mM PMSF, 1× cOmplete protease inhibitor; Roche) as used in Neumann et al. (2012). Antibody bound beads were incubated with diluted chromatin O/N on rotation at 4 °C. Precipitated immunocomplexes were washed three times for

5 min with 200 µl of Washing buffer (Invitrogen) and ChIPped chromatin was eluted two times for 15 min at 65 °C with 100 µl of elution buffer (1% SDS, 0.1 M NaHCO3). After washing, beads were resuspended in TE buffer followed by RNase and Proteinase K treatment to release DNA from the immunoprecipitated nucleosomes. Finally, ChIPped DNA and input were isolated with DNA purification kit (Qiagen) and eluted in 50 µl of 10 mM Tris–Cl, pH 8.5 buffer. About 5 µg of normal rabbit IgG (Cell Signaling Technology, No. 2729) was included as a negative control in each ChIP experiment in order to optimize the experimental conditions. The relatively high ratio of antialpha CenH3 ChIPed DNA versus rabbit IgG ChIPed DNA was used as an indicator of a successful ChIP experiment. Genomic DNA for WGS sequencing was isolated from eggs using DNeasy Blood and Tissue Kit (Qiagen). Library construction (KAPA Hyper Prep kit) and sequencing were done via multiplexing using Illumina HiSeq technology which produced 151-bp paired-end reads (AdmeraHealth). Raw Illumina Input and ChIP-sequencing reads have been deposited to NCBI BioProject database under the study accession number (PRJNA639449).

ChIP-Seq Data Analysis

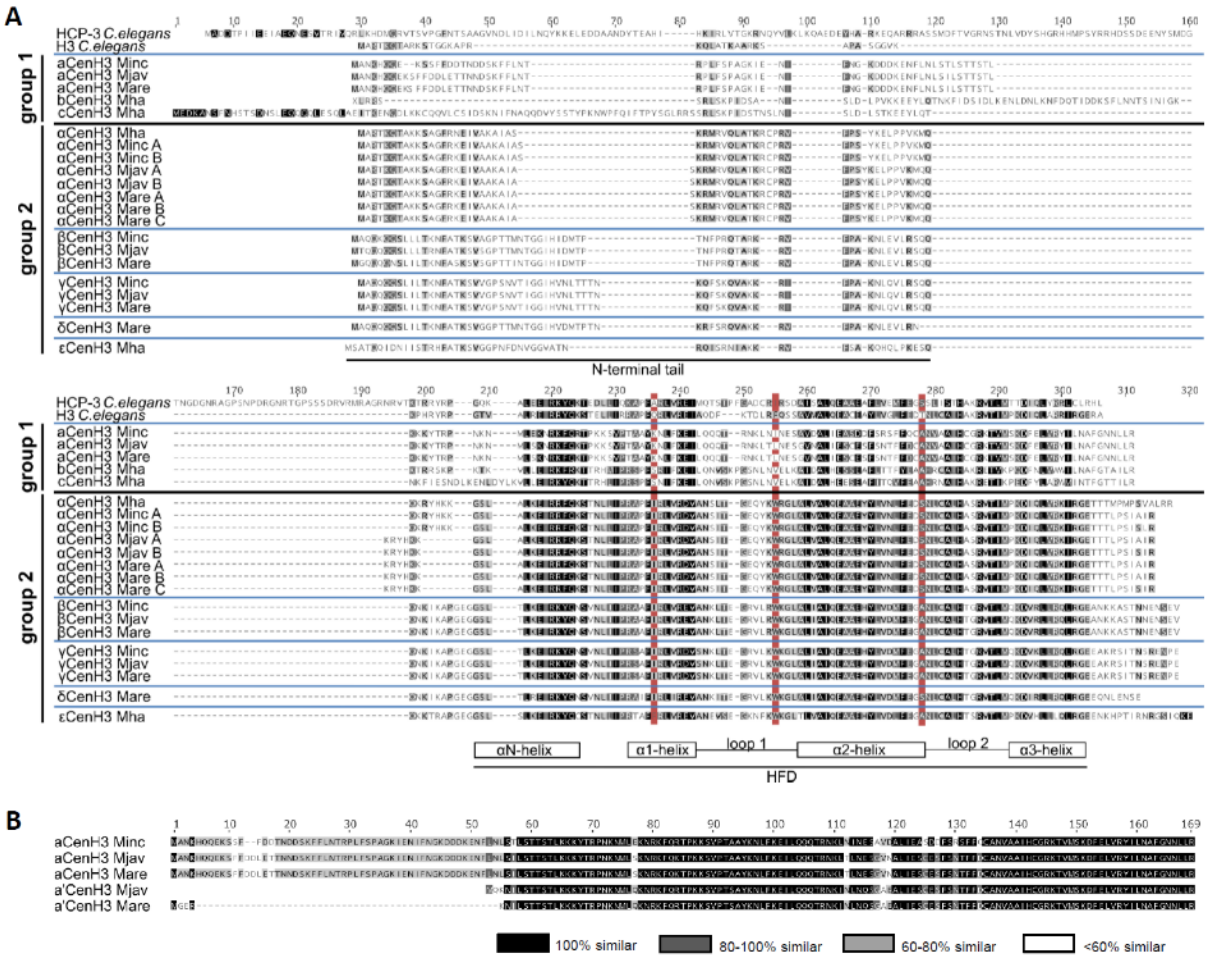
Sequencing data were first tested with FastQC (Andrews 2010) and preprocessed to get high-quality reads. About 2 million of 151-bp pair-end sequence reads were mapped to the *M. incognita* genome reference genome GCA_900182535.1 (Blanc-Mathieu et al. 2017) using the Burrows–Wheeler Aligner (BWA) with default parameters (Li and Durbin 2009).

Repetitive part of the *M. incognita* genome was analyzed using graph-based clustering which utilizes grouping of reads based on their shared similarity. A total of 1 million paired reads (0.8 genome coverage) was used for RepeatExplorer (Novák et al. 2013) clustering. The output of analysis was clusters, composed of contigs with overlapping sequences, where each cluster represents a repetitive element. One million of single-end reads of ChIP and input were mapped to the clusters with ChIP-seq Mapper (Neumann et al. 2012). Repeats enriched in ChIPped DNA were identified by elevated proportions of reads from ChIP versus input data. The threshold for ChIP versus input ratio >1.5 was chosen. This relatively low threshold for ChIP enrichment was also selected previously in studies where CenH3 was associated with many different

sequences (e.g., in the plant Beta; Kowar et al. 2016). Recently, it has been proposed that both the abundance and enrichment have to be taken into account in ChIP analyses to estimate of sequences associated with the centromeres (Talbert et al. 2018). Therefore, 100 most abundant repeat clusters, with at least 0.02% *M. incognita* genome content were presented. In addition to cluster analysis, mapping of ChIP and input on 10,000 contigs (with at least 0.002% genome proportion) using criteria that each read can only be assigned to one contig was done. The analysis was performed for all three ChIP replicates and enriched contigs were determined based on calculated ChIP/input hit ratio (supplementary table 6). The centromeric candidates (from clusters and contigs analyses) were further analyzed for tandem repeats using the Tandem Repeats Finder server (https://tandem.bu.edu/trf/trf.html) with default parameters (Benson 1999). Sequence monomer analyses such as multiple sequence comparison, GC content, pairwise identity, and motif search were done using Genious v9.1, together with implemented DNA-fold Vienna package for secondary structure analysis. Surrounding TR regions were investigated by assembling contigs using de novo assembler. In order to find putative centromere candidates in related *M. incognita* species, the first step was clustering of raw Illumina WGS data of *M. arenaria* (SRR4242477) and *M. javanica* (SRR4242459) raw Illumina WGS data (Szitenberg et al. 2017) using RepeatExplorer. The second step included detection of clusters/contigs which contain TRs-associated with *M. incognita*-specific centromere sequence motif. The selected putative centromeric TR arrays in closely related species were further analyzed as described for *M. incognita*.

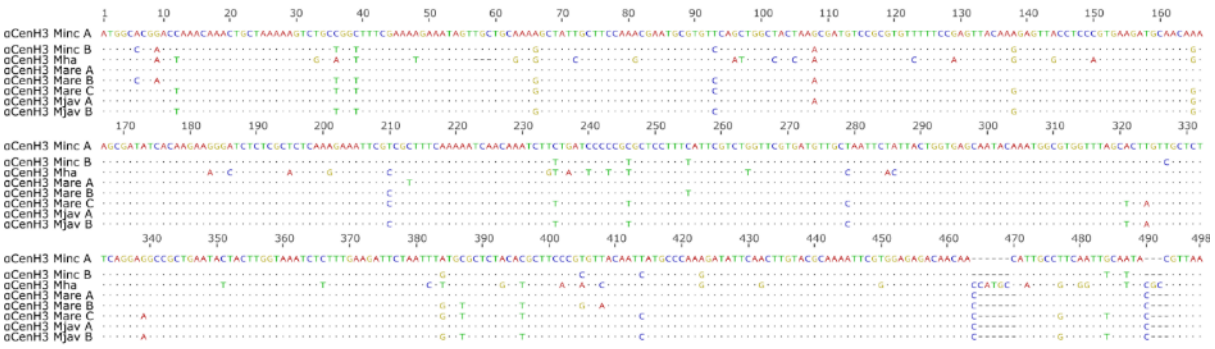
For validation of ChIP enrichment obtained on repetitive part of the *M. incognita* genome preprocessed and subsampled ChIP and input reads for each replicate were mapped to *M. incognita* genome assembly (Blanc-Mathieu et al. 2017) using bowtie2 (Langmead and Salzberg 2012) with –very-sensitive-local and -a option on the Galaxy platform. For normalizing read counts bamCompare (Ramírez et al. 2016) was used with default parameters, where difference was computed by subtracting read number of input from corresponding ChIP sample. Results were visualized using Integrative Genomics Viewer (Robinson et al. 2011).

Supplementary Material

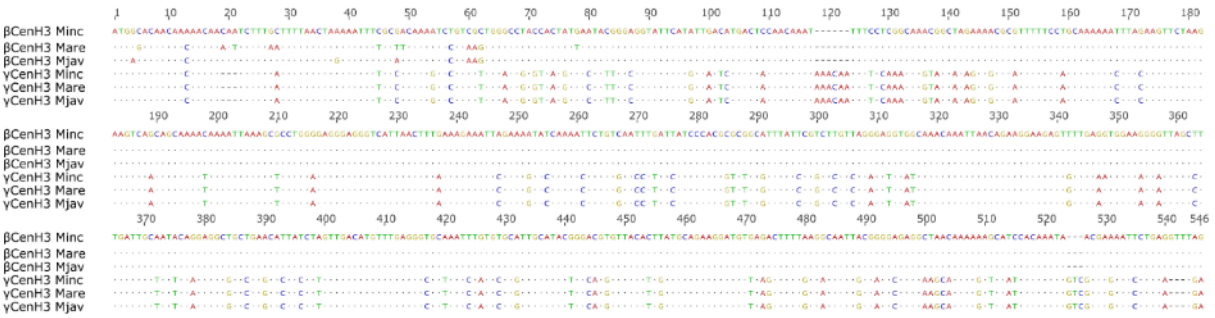


Supplementary Figure 1. A) Alignment of CenH3 candidates from *Meloidogyne* species (Minc, *M. incognita*; Mare, *M. arenaria*; Mjav, *M. javanica* and Mha, *M. hapla*). Secondary structure of HFD is depicted below the alignment. The red boxes indicate diagnostic amino acids changes of HFD in comparison to H3 from *C. elegans*. Group 1 represents CenH3 candidates with low sequence identitiy of HFD (<50%) in comparison to H3 from *C. elegans*. Group 2 contains CenH3 candidates with high HFD identity (≥50%) to H3 from *C. elegans*. The sources of all CenH3 candidates together with abbreviated names are supplied in Supplementary Table 1. Sequence identity matrix of HFDs is provided in Supplementary Table 2. B) Alignment of aCenH3s together with truncated a'CenH3s found in *M. javanica* and *M. arenaria*

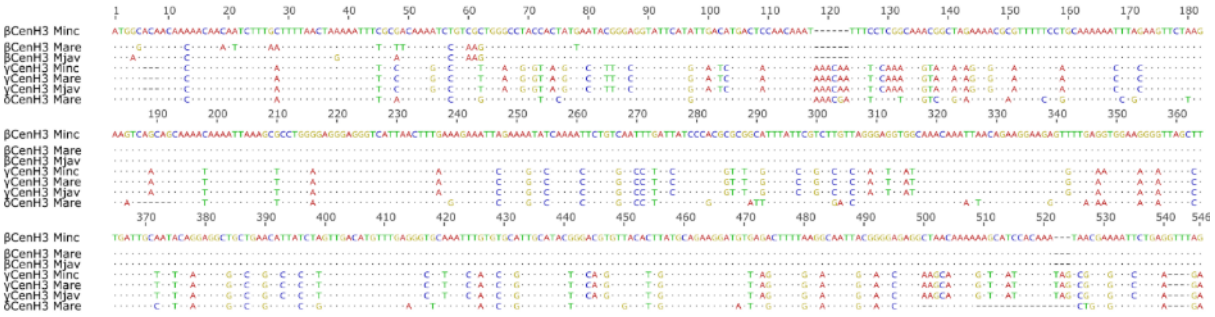
αCenH3



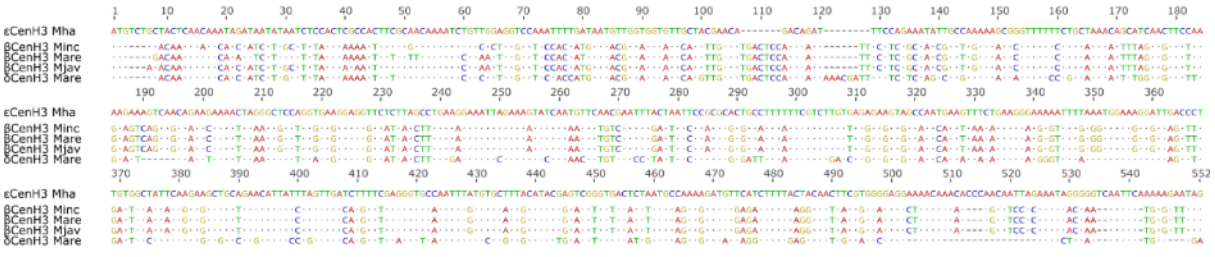
βγCenH3



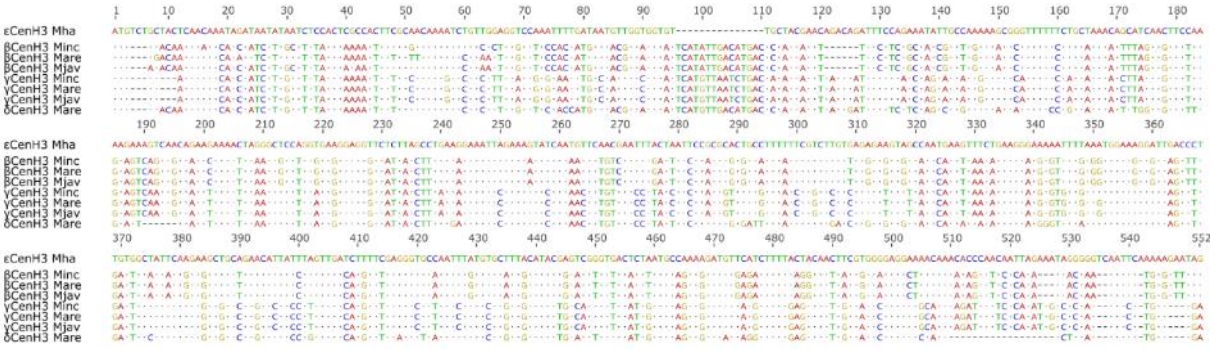
βγδCenH3



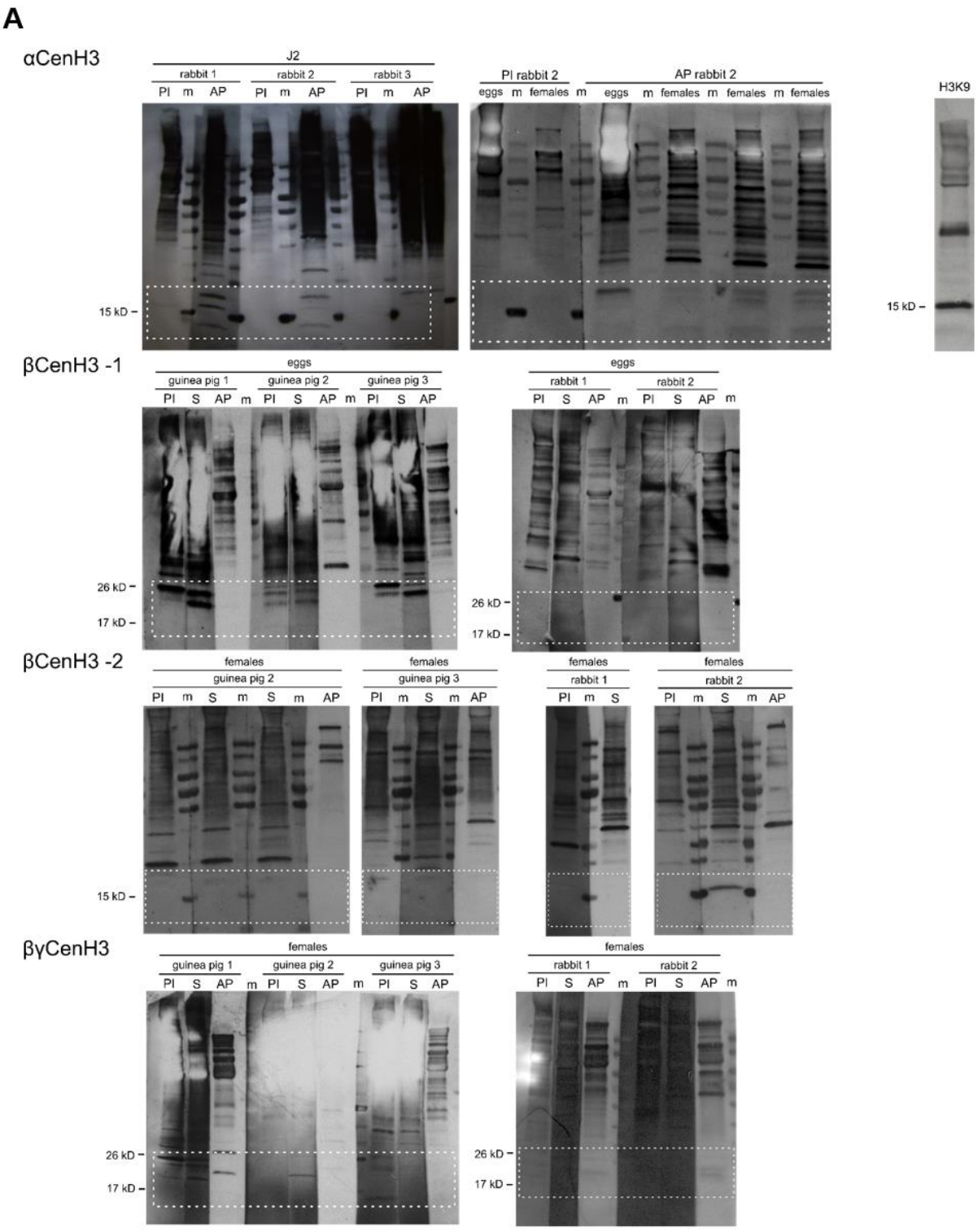
βδεCenH3

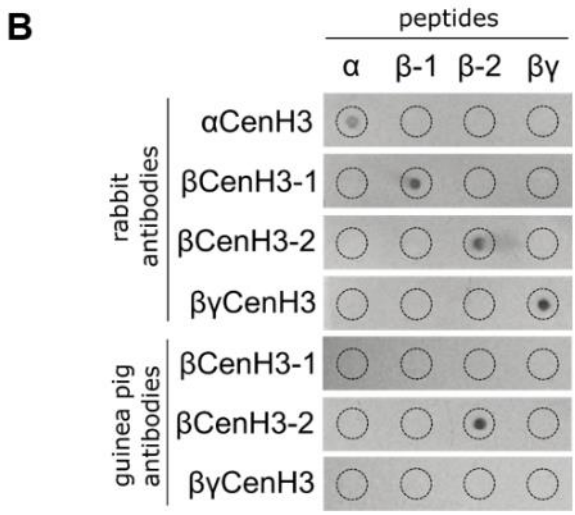


βγδεCenH3

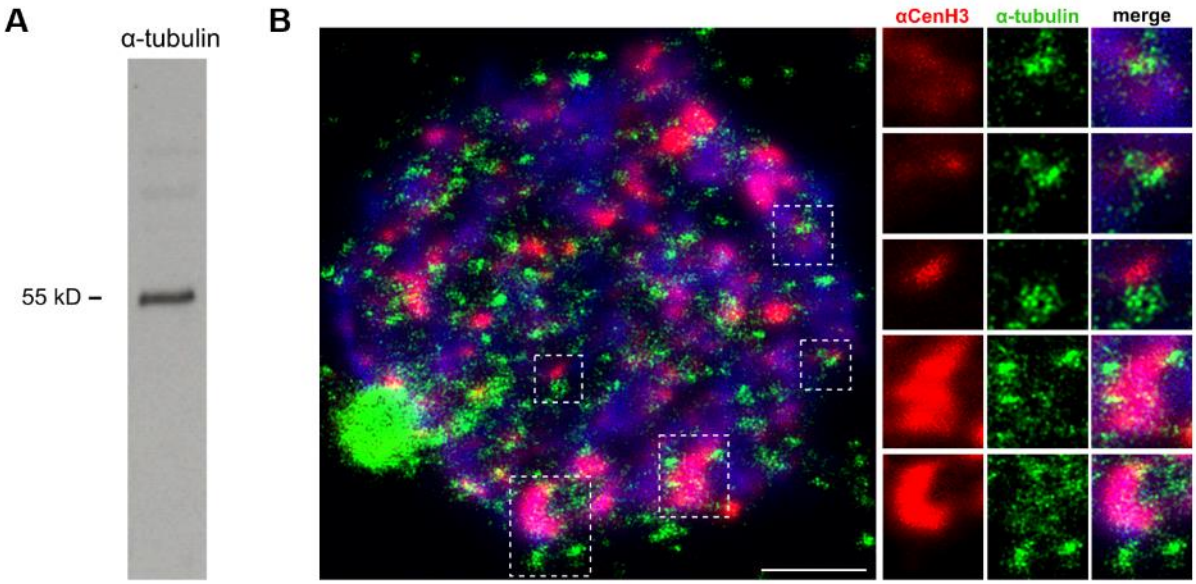


Supplementary Figure 2. Alignments of CenH3 gene sequences from *Meloidogyne* (Minc-*M. incognita*; Mare-*M. arenaria*; Mjav-*M. javanica* and Mha-*M. hapla*) generated from full-length transcripts databases (WormBase ParaSite; https://parasite.wormbase.org/index.html) used for codon positive selection test. The IDs of all CenH3s are supplied in Supplementary Table 1.

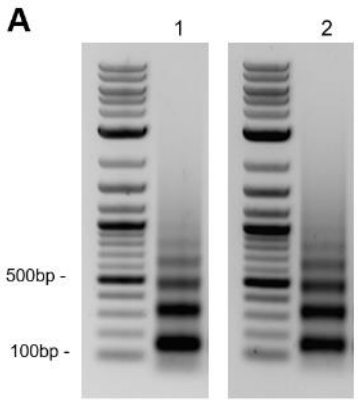




Supplementary Figure 3. A) Western blot of CenH3 antibodies (αCenH3, βCenH3-1, βCenH3-2 and βγCenH3) raised in rabbits or in guinea pigs on *M. incognita* proteins isolated from eggs, J2 and females (ovaries and uterus). The same amount of the whole protein isolate were loaded in all samples (20µg/lane) except in Western blot of □CenH3 on females protein isolate (AP rabbit 2) where the three different protein amounts were applied (20, 30 and 40µg/lane). Western blot with H3K9 polyclonal antibody is also shown (up, right). Outlined are regions of expected CenH3 signals based on calculated heights. (PI = pre-immune serum, AP = affinity purified monospecific IgG fraction after 120 days of immunization with specific peptide). B) Peptide dot blot of all CenH3 antibodies on synthetized peptides corresponding to CenH3 epitopes used for animal immunization. For β and βγCenH3 screening, mixed antibodies of three guinea pigs and two rabbits are used. The peptide sequences are indicated in alignment of Figure 1A.

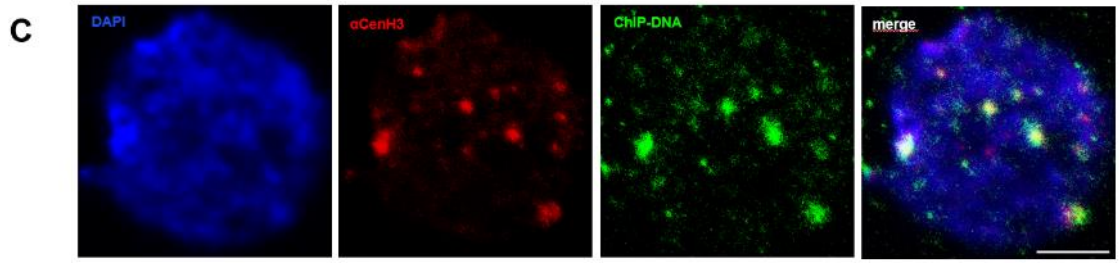


Supplementary Figure 4. A) Western blot of mouse monoclonal α-tubulin antibody on proteins isolated from *M. incognita* eggs B) Double immunofluorescence with anti-αCenH3 (red) and anti-α-tubulin (green) antibody in *M. incognita*. Outlined segments of visible close centromere-tubulin interaction are enlarged and channels are shown separated. DNA was counterstained with DAPI (blue). Scale bar = 3 µm.

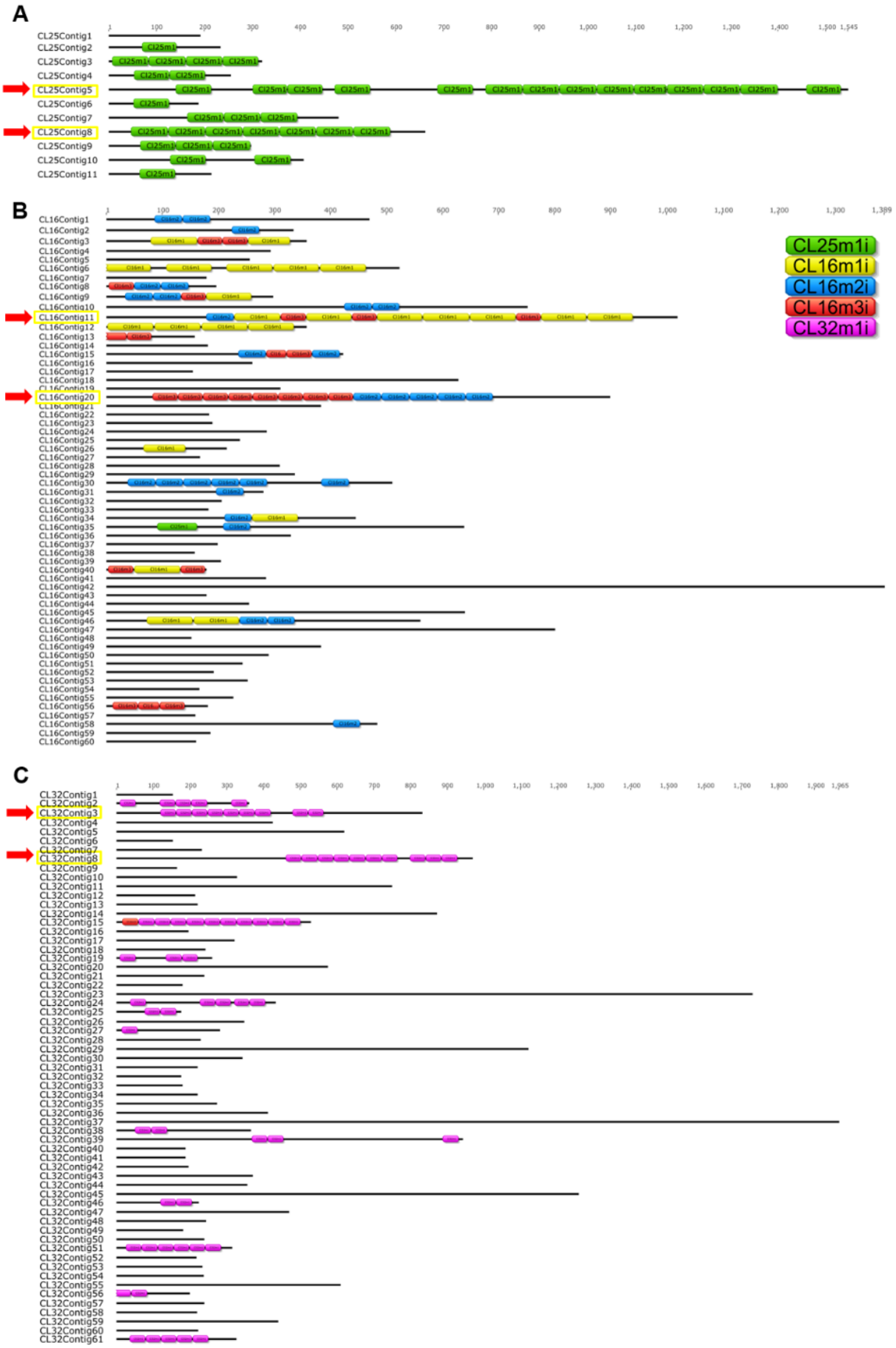


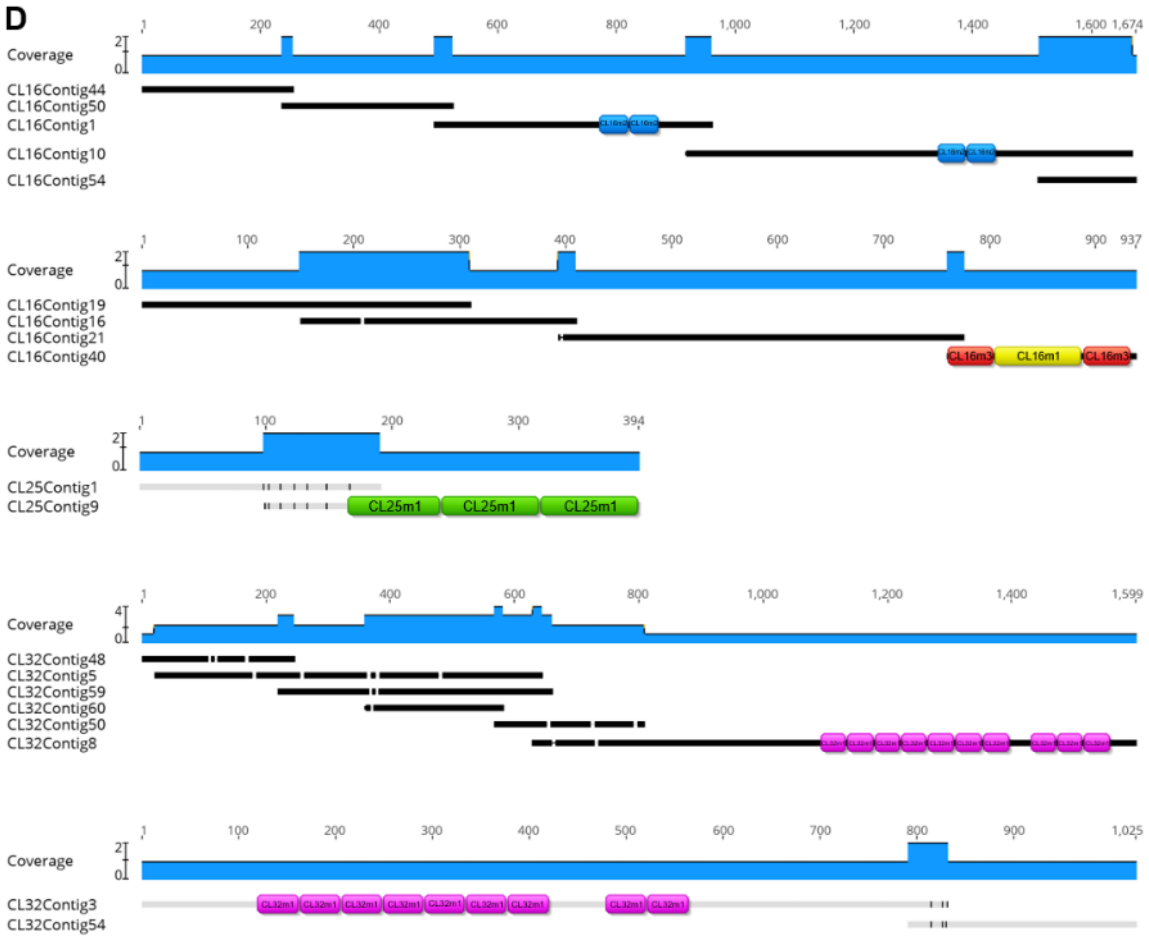
B

αCenH3 ChIP experiment	Native chromatin isolation	Antibody amount (µg)
ChIP-1	1	30
ChIP-2A	2	10
ChIP-2B		30



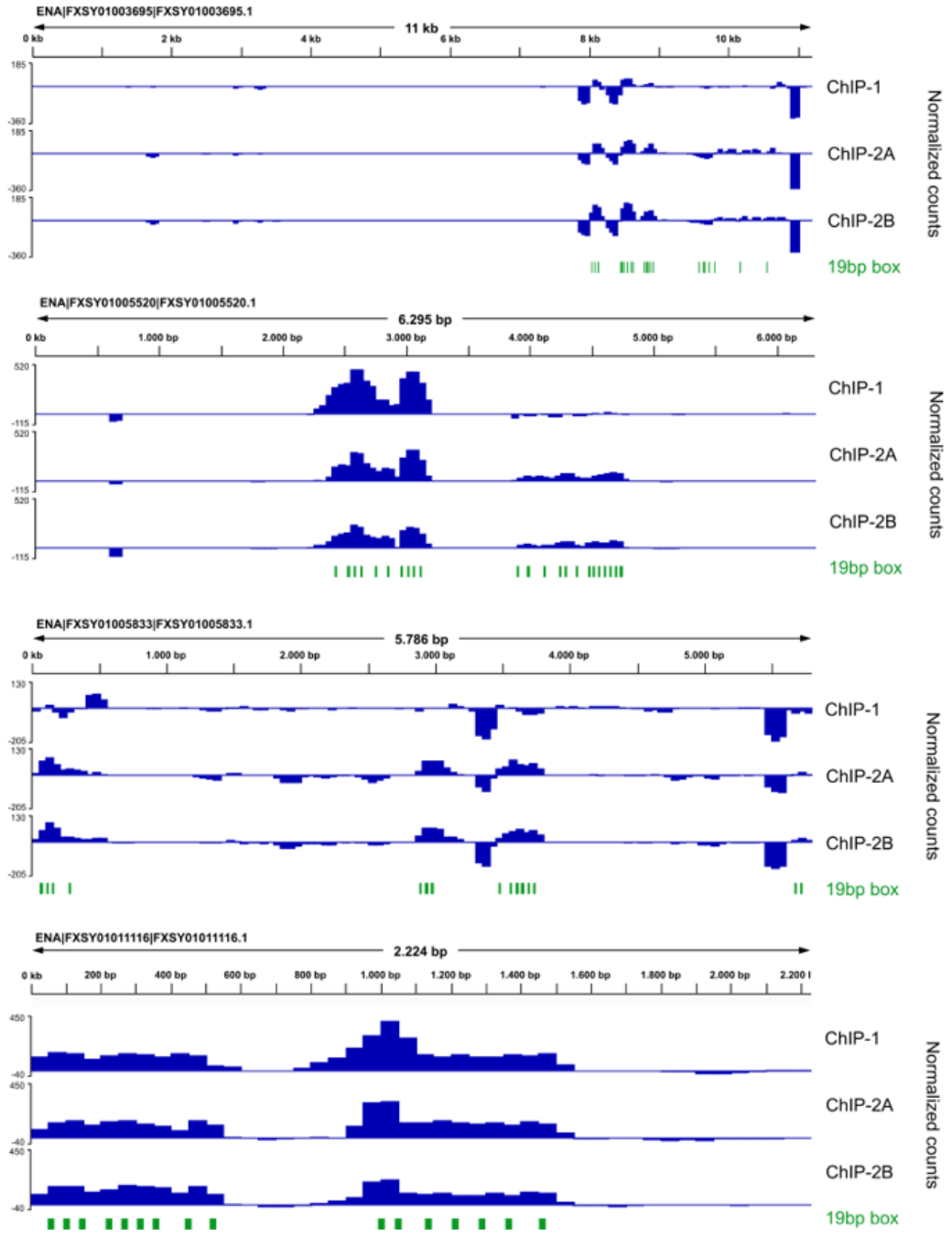
Supplementary Figure 5. A) Agarose gel of DNA isolated from MNase treated chromatin using 0.2U MNase/µL eggs and 0.4U MNase/µL eggs for 2 min (line 1 and 2, respectively). B) Three ChIP experiments using different chromatin isolation and different antibody amount. C) IF-FISH with αCenH3 antibody and ChIPped-DNA as probe on the prophase of *M. incognita*. Scale bar = 5 µm.


A
CL25Contig1
CL25Contig2
CL25Contig3
CL25Contig4
CL25Contig5
CL25Contig6
CL25Contig7
CL25Contig8
CL25Contig9
CL25Contig10
CL25Contig11
B
CL16Contig1
CL16Contig2
CL16Contig3
CL16Contig4
CL16Contig5
CL16Contig6
CL16Contig7
CL16Contig8
CL16Contig9
CL16Contig10
CL16Contig11
CL16Contig12
CL16Contig13
CL16Contig14
CL16Contig15
CL16Contig16
CL16Contig17
CL16Contig18
CL16Contig19
CL16Contig20
CL16Contig21
CL16Contig22
CL16Contig23
CL16Contig24
CL16Contig25
CL16Contig26
CL16Contig27
CL16Contig28
CL16Contig29
CL16Contig30
CL16Contig31
CL16Contig32
CL16Contig33
CL16Contig34
CL16Contig35
CL16Contig36
CL16Contig37
CL16Contig38
CL16Contig39
CL16Contig40
CL16Contig41
CL16Contig42
CL16Contig43
CL16Contig44
CL16Contig45
CL16Contig46
CL16Contig47
CL16Contig48
CL16Contig49
CL16Contig50
CL16Contig51
CL16Contig52
CL16Contig53
CL16Contig54
CL16Contig55
CL16Contig56
CL16Contig57
CL16Contig58
CL16Contig59
CL16Contig60
CL25m1i
CL16m1i
CL16m2i
CL16m3i
CL32m1i
C
CL32Contig1
CL32Contig2
CL32Contig3
CL32Contig4
CL32Contig5
CL32Contig6
CL32Contig7
CL32Contig8
CL32Contig9
CL32Contig10
CL32Contig11
CL32Contig12
CL32Contig13
CL32Contig14
CL32Contig15
CL32Contig16
CL32Contig17
CL32Contig18
CL32Contig19
CL32Contig20
CL32Contig21
CL32Contig22
CL32Contig23
CL32Contig24
CL32Contig25
CL32Contig26
CL32Contig27
CL32Contig28
CL32Contig29
CL32Contig30
CL32Contig31
CL32Contig32
CL32Contig33
CL32Contig34
CL32Contig35
CL32Contig36
CL32Contig37
CL32Contig38
CL32Contig39
CL32Contig40
CL32Contig41
CL32Contig42
CL32Contig43
CL32Contig44
CL32Contig45
CL32Contig46
CL32Contig47
CL32Contig48
CL32Contig49
CL32Contig50
CL32Contig51
CL32Contig52
CL32Contig53
CL32Contig54
CL32Contig55
CL32Contig56
CL32Contig57
CL32Contig58
CL32Contig59
CL32Contig60
CL32Contig61



Supplementary Figure 6. Contigs from αCenH3 ChIP-enriched clusters, A) Cl25, B) CL16 and C) CL32, with mapped monomers on the contigs (CL25m1i, CL16m1i, CL16m2i, CL16m3i and CL32m1i) in *M. incognita*. Red arrows indicate contigs enriched in ChIP analyses. D) Examples of assemblies of αCenH3-associated contigs with contigs which represent their flanking regions originating from the same cluster.

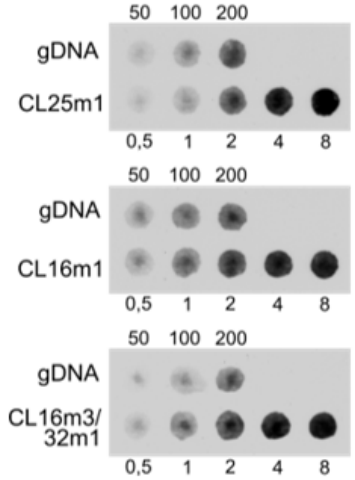
A



B

	WGS	Input-1	Input-2	ChIP-1	ChIP-2A	ChIP-2B	Assembled genome
Hits	6649	5615	5347	8690	8054	7951	1117
Abundance (%)	0,221	0,223	0,212	0,345	0,320	0,316	0,037

C



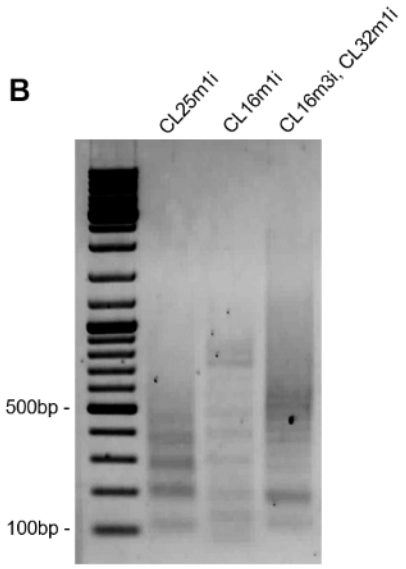
						genome abundance (%)
gDNA	spotted (ng)	**50**	**100**	**200**		**0,78**
	gray value	5,91	18,24	43,13		
CL25m1	spotted (ng)	**0,5**	**1**	**2**	**4**	**8**
	gray value	4,25	12,37	40,88	64,45	77,49
gDNA	spotted (ng)	**50**	**100**	**200**		**1,12**
	gray value	14,66	27,51	41,12		
CL16m1	spotted (ng)	**0,5**	**1**	**2**	**4**	**8**
	gray value	13,79	33,10	50,83	60,18	63,32
gDNA	spotted (ng)	**50**	**100**	**200**		**0,41**
	gray value	2,94	12,07	30,17		
CL16m3/ 32m1	spotted (ng)	**0,5**	**1**	**2**	**4**	**8**
	gray value	6,49	24,94	50,58	65,31	65,75

Supplementary Figure 7. A) αCenH3-ChIP visualization with genome browser on *M. incognita* genome assembly with three ChIP replicates normalized to input. The prominent examples of scaffolds with αCenH3 peaks are shown together with 19bp box below (green). B) Abundance of ChIP specific sequences in the assembled genome, whole genome sequencing, input and ChIP samples calculated based on genome size of ~180 Mb and average monomer length of 60bp in 106 sequenced reads. C) Dot blots hybridization analysis of *M. incognita* genomic DNA (gDNA) using probes specific for CL25m1, CL16m1 and CL16m3/32m1 monomers, respectively. The amounts of DNAs blotted were 50, 100 and 200ng. Positive control were monomers CL25m1, CL16m1 and CL16m3/32m1 in the amounts of 0.5, 1, 2, 4 and 8 ng (left panel). The calculation of relative abundance of CL25m1, CL16m1 and CL16m3/32m1 were calculated based on densitometry of dot blots (right panel).

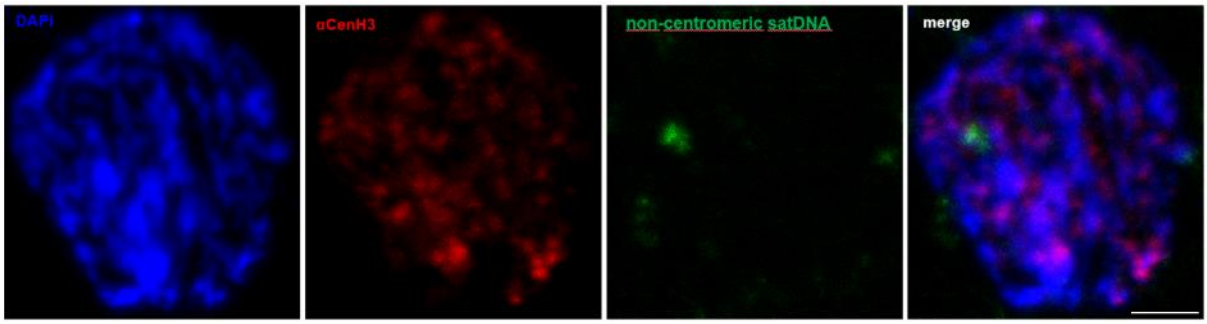
A

FISH hybridization probes	Primer sequence (5'-3')	
CL25m1i	forward	TCGGGCCTTCGGCCCTCGCC
	reverse	CAACCCATCTAAAAGGCAGA
CL16m1i	forward	CGTTTTAGATGTAATGAAAGC
	reverse	GAGCAATAGTCAAAGGCAG
CL16m3i, CL32m1i	forward	CGCGAGGGCCGAAGGC
	reverse	GGATGGATCTTAAGCTCG

B

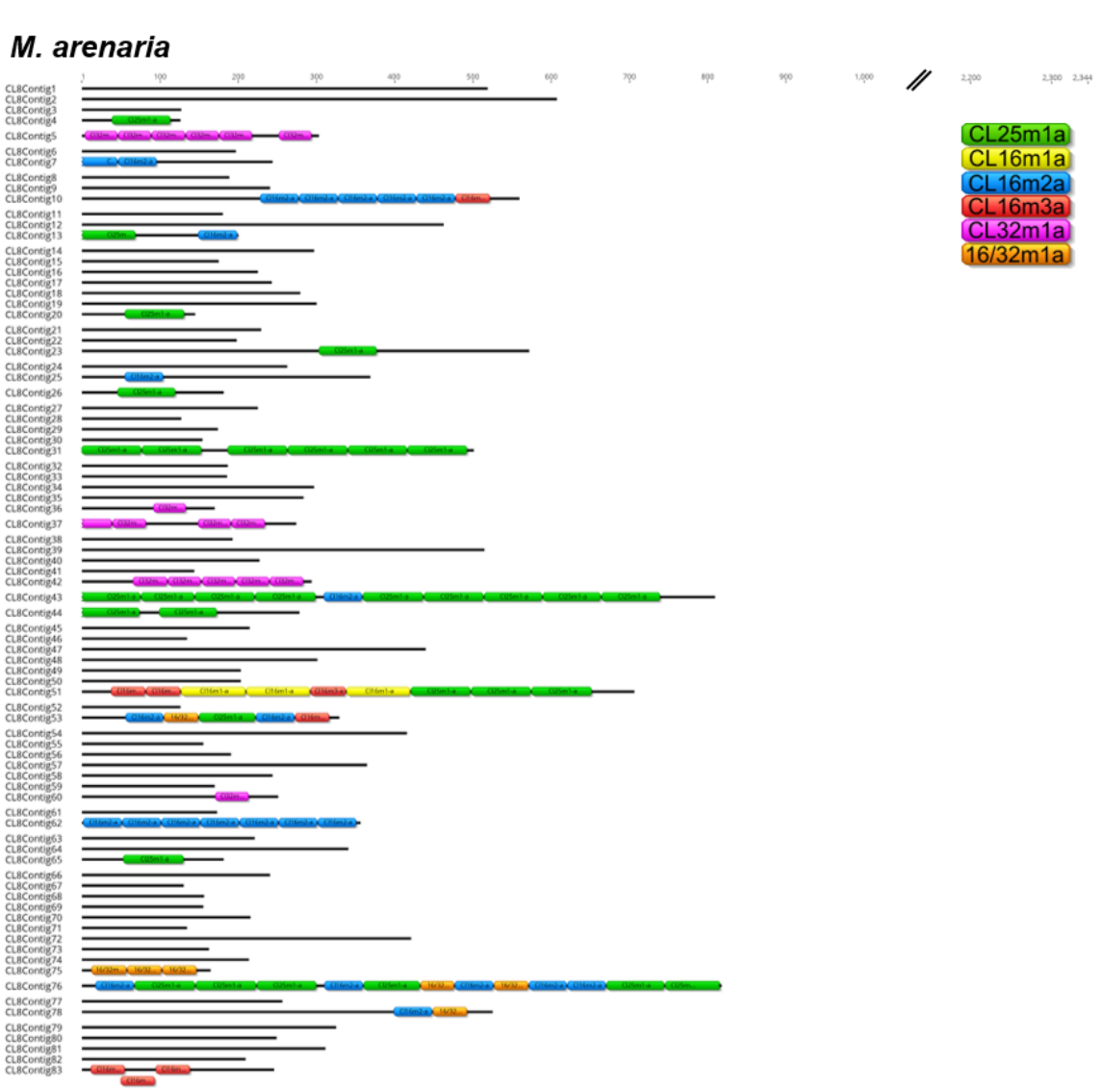


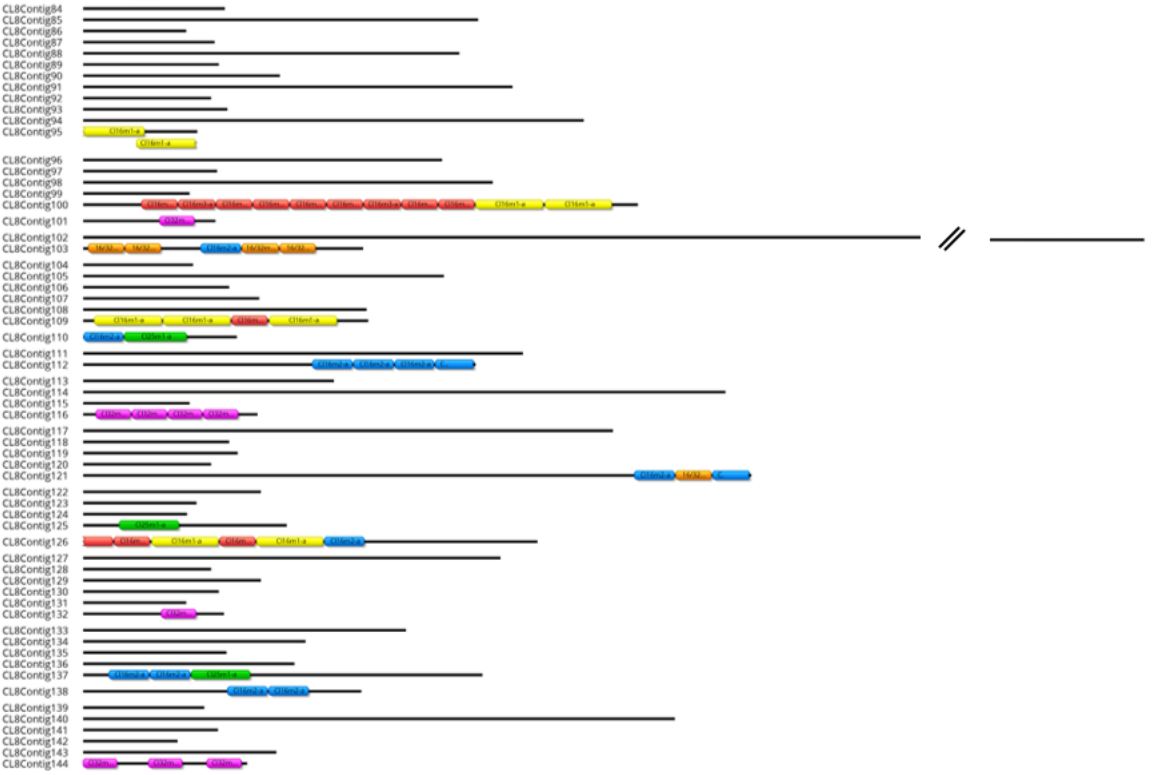
Supplementary Figure 8. A) Primers used for labelling of αCenH3-associated monomers (CL25m1i, CL16m1i, CL16m3i and CL32m1i) which were used in FISH and dot blot hybridizations as probe. B) Gel electrophoresis of hybridization probes.



Supplementary Figure 9. Combined immunofluorescence and fluorescence in situ labeling with anti-αCenH (red) and αCenH3 non-enriched satellite DNA (green) on *M. incognita* cytosmear. DNA was counterstained with DAPI (blue). Scale bar = 3 µm.

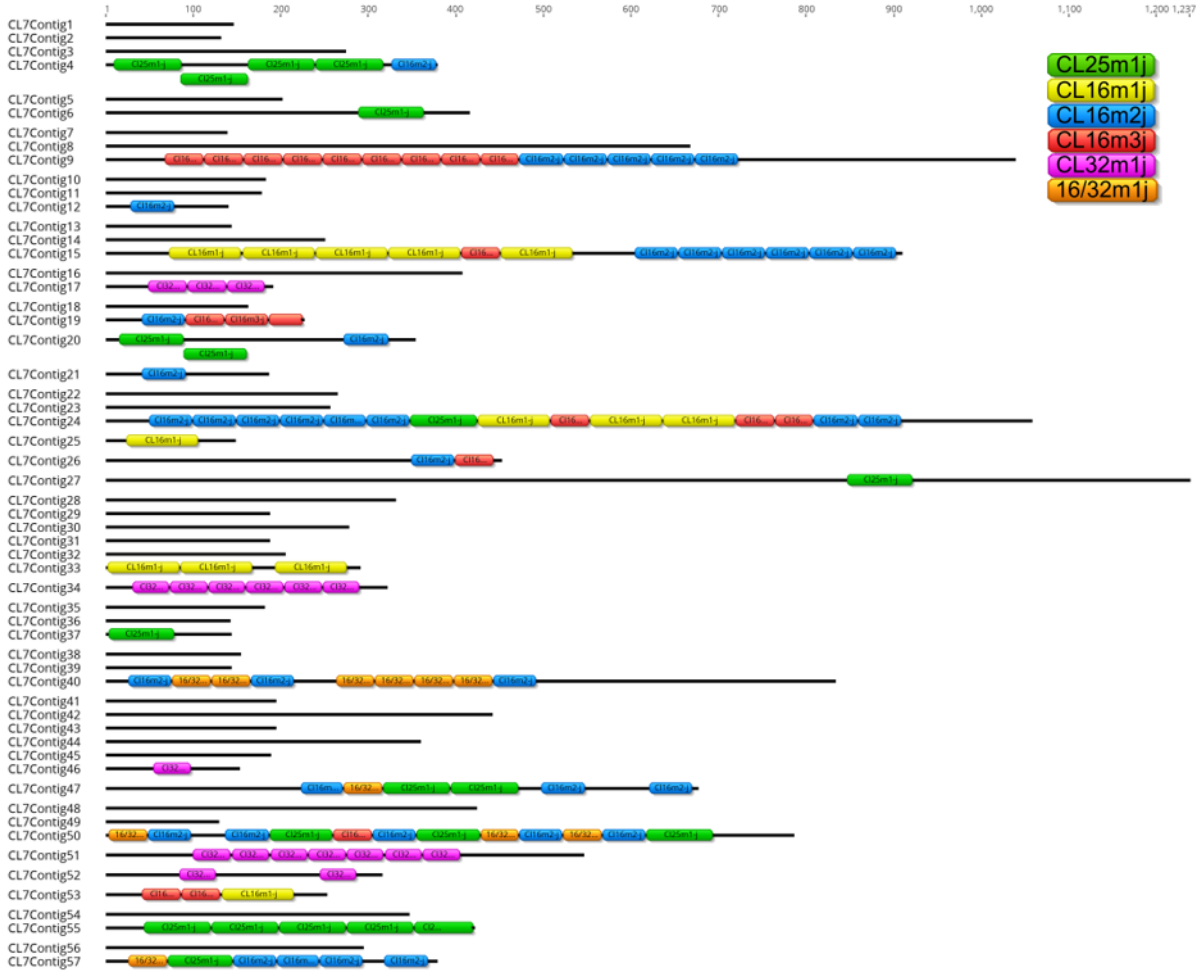
A

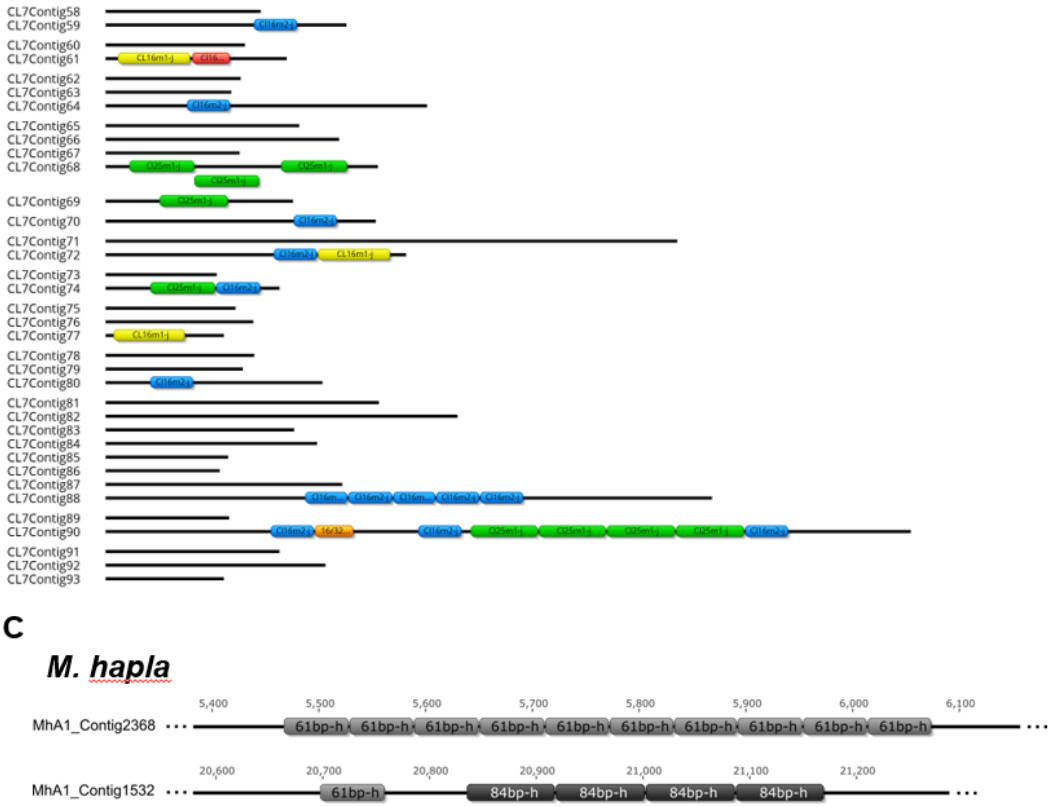




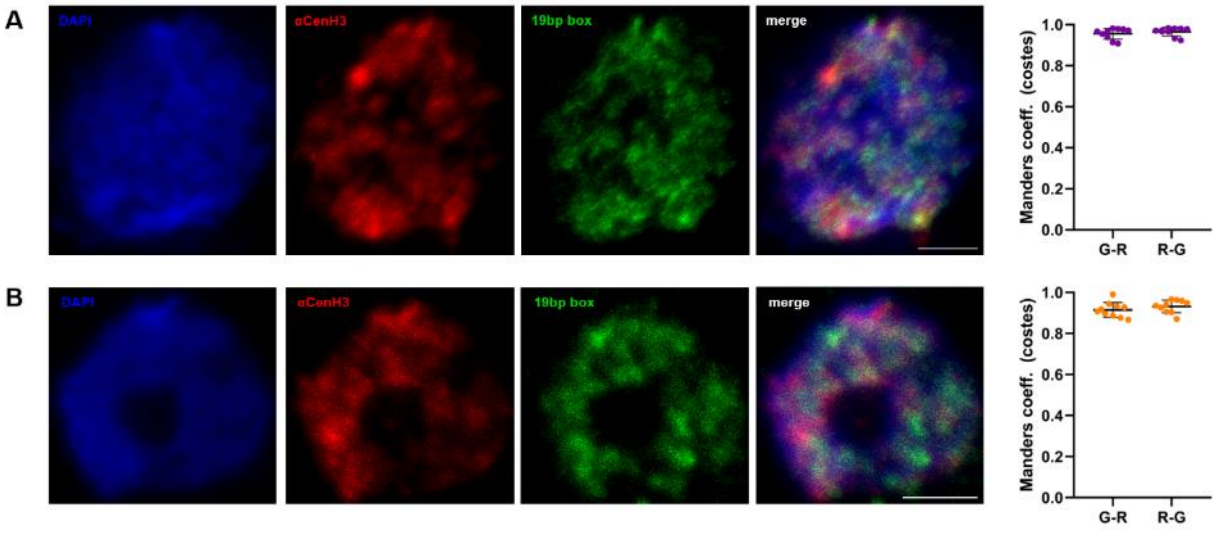
B

M. javanica





Supplementary Figure 10. Clusters obtained by clustering of Illumina WGS reads followed by mapping with αCenH3-associated monomers (Figure 8A) in A) *M. arenaria* (CL8) and B) *M. javanica* (CL7). C) Representatives of contigs with tandemly repeated monomers-containing 19 bp box (Figure 8A) from *M. hapla* assembled genome (Opperman et al, 2008).



Supplementary Figure 11. Combined immunofluorescence and primed in situ labeling with anti-αCenH (red) and centromeric 19 bp box (green) together with graph of signal colocalization quantification for 10 representative images with calculated Manders coefficients (costes thresholding) in A) *M. arenaria* and B) *M. javanica*. Data on graph are presented as mean ± SD (supplementary table 7). DNA was counterstained with DAPI (blue). Scale bar = 3 µm.

Supplementary Table 1. List of CenH3 candidates with abbreviations and IDs from WormBase ParaSite database (https://parasite.wormbase.org/index.html) from *Meloidogyne* species (Minc-*M. incognita*; Mare-*M. arenaria*; Mjav-*M. javanica* and Mha-*M. hapla*). Grouping of sequences was done according to their mutual HFD sequence similarity (Supplementary Table 2).

	CenH3 abbreviations	Gene/protein IDs in WormBase ParaSite database
group 1	aCenH3 Minc	Minc3s02591g30793
	aCenH3 Mjav	M.Javanica_Scaff7130g046866
	aCenH3 Mare	M.Arenaria_Scaff7091g061788
	bCenH3 Mha	MhA1_Contig881.frz3.gene1
	cCenH3 Mha	MhA1_Contig773.frz3.gene16
group 2	αCenH3 Mha	MhA1_Contig20.frz3.gene12
	αCenH3 Minc A	Minc3s00059g03025
	αCenH3 Minc B	Minc3s00931g19001
	αCenH3 Mjav A	scaffold14337_cov303.g17394
	αCenH3 Mjav B	scaffold7103_cov296.g11640
	αCenH3 Mare A	tig00001842.g73055.t1
	αCenH3 Mare B	tig00002608.g35385.t1
	αCenH3 Mare C	tig00002642.g7848.t1
	βCenH3 Minc	Minc3s02448g30096
	βCenH3 Mjav	M.Javanica_Scaff2g000047
	βCenH3 Mare	M.Arenaria_Scaff2725g035701
	γCenH3 Minc	Minc3s00059g03056
	γCenH3 Mjav	M.Javanica_Scaff219g003590
	γCenH3 Mare	M.Arenaria_Scaff81g002502
	δCenH3 Mare	M.Arenaria_Scaff17829g091072
	εCenH3 Mha	MhA1_Contig20.frz3.gene42

a'CenH3 Mjav M.Javanica_Scaff8474g052097

a'CenH3 Mare M.Arenaria_Scaff14550g084642

Supplementary Table 2. Identity matrix of HFD protein regions of all CenH3 candidates from *Meloidogyne* species (Minc, *M. incognita*; Mare, *M. arenaria*; Mjav, *M. javanica* and Mha, *M. hapla*) together with HCP-3 and H3 from *C. elegans*

	HCP-3 *C.el*	H3 *C.el*	aMinc	aMjav	aMare	bMha	cMha	αMha	αMinc A	αMinc B	αMjav A	αMjav B	αMare A	αMare B	αMare C	βMinc	βMjav	βMare	γMinc	γMjav	γMare	δMare	εMha
HCP-3 *C.el*		54,8	29	26,9	26,9	35,5	36,1	49,5	49,5	49,5	49,5	49,5	49,5	49,5	49,5	47,3	47,3	47,3	44,1	44,1	44,1	47,3	41,9
H3 *C.el*			31,9	33	33	38,7	39,2	64,8	64,8	64,8	64,8	64,8	64,8	64,8	64,8	52,2	52,2	52,2	47,8	47,8	47,8	52,2	51,1
aMinc				85,7	85,7	39,8	35,1	28,3	28,3	28,3	28,3	28,3	28,3	28,3	28,3	30,4	30,4	30,4	28,3	28,3	28,3	28,3	23,9
aMjav					100	39,8	33	28,3	28,3	28,3	28,3	28,3	28,3	28,3	28,3	29,3	29,3	29,3	27,2	27,2	27,2	27,2	23,9
aMare						39,8	33	28,3	28,3	28,3	28,3	28,3	28,3	28,3	28,3	29,3	29,3	29,3	27,2	27,2	27,2	27,2	23,9
bMha							74,2	34,4	34,4	34,4	34,4	34,4	34,4	34,4	34,4	29	29	29	29	29	29	28	32,3
cMha								33	33	33	33	33	33	33	33	29,9	29,9	29,9	29,9	29,9	29,9	28,9	33
αMha									98,9	98,9	98,9	98,9	98,9	98,9	98,9	66,3	66,3	66,3	66,3	66,3	66,3	66,3	67,4
αMinc A										100	100	100	100	100	100	65,2	65,2	65,2	65,2	65,2	65,2	67,4	67,4
αMinc B											100	100	100	100	100	65,2	65,2	65,2	65,2	65,2	65,2	67,4	67,4
αMjav A												100	100	100	100	65,2	65,2	65,2	65,2	65,2	65,2	67,4	67,4
αMjav B													100	100	100	65,2	65,2	65,2	65,2	65,2	65,2	67,4	67,4
αMare A														100	100	65,2	65,2	65,2	65,2	65,2	65,2	67,4	67,4
αMare B															100	65,2	65,2	65,2	65,2	65,2	65,2	67,4	67,4
αMare C																65,2	65,2	65,2	65,2	65,2	65,2	67,4	67,4
βMinc																	100	100	94,6	94,6	94,6	92,4	78,3
βMjav																		100	94,6	94,6	94,6	92,4	78,3
βMare																			94,6	94,6	94,6	92,4	78,3
γMinc																				100	100	89,1	77,2
γMjav																					100	89,1	77,2
γMare																						89,1	77,2
δMare																							73,9
εMha																							

Supplementary Table 3. Identity martix of whole CenH3 protein sequences (α, β, δ, δ and εCenH3) from *Meloidogyne* species (Minc, *M. incognita*; Mare, *M. arenaria*; Mjav, *M. javanica* and Mha, *M. hapla*) together with H3 from *C. elegans*

	H3 *C.el*	αMha	αMinc B	αMinc A	αMjav A	αMjav B	αMare A	αMare B	αMare C	βMinc	βMjav	βMare	γMinc	γMjav	γMare	δMare	εMha
H3 *C.el*		51,3	51	51	51	51	51	51	51	36,7	35,5	36,1	34,1	34,1	34,1	36,7	35,6
αMha			95,1	94,4	94,4	95,1	94,4	94,4	95,1	43,3	42,1	43,3	42,5	42,5	42,5	42,2	43,3
αMinc B				98,8	98,8	99,4	98,8	98,8	99,4	44	42,9	44	43,2	43,2	43,2	44,1	44,6
αMinc A					100	99,4	100	100	99,4	44	42,9	44	43,2	43,2	43,2	43,5	44,6
αMjav A						99,4	100	100	99,4	44	42,9	44	43,2	43,2	43,2	43,5	44,6
αMjav B							99,4	99,4	100	44	42,9	44	43,2	43,2	43,2	43,5	44,6
αMare A								100	99,4	44	42,9	44	43,2	43,2	43,2	43,5	44,6
αMare B									99,4	44	42,9	44	43,2	43,2	43,2	43,5	44,6
αMare C										44	42,9	44	43,2	43,2	43,2	43,5	44,6
βMinc											97,8	96,1	78,3	78,3	78,3	82,7	60,1
βMjav												96,6	77,8	77,8	77,8	82,1	60,7
βMare													77,8	77,8	77,8	81,6	60,1
γMinc														100	100	78,8	60
γMjav															100	78,8	60
γMare																78,8	60
δMare																	59,2
εMha																	

Supplementary Table 4. Non-synonymous/synonymous substitution rate ratio (dN/dS=ω) in α, β and γCenH3 variants among species as well as dN/dS ratio using HFD or N-tail of different CenH3s

αCenH3								
	αMinc A	αMinc B	αMha	αMare A	αMare B	αMare C	αMjav A	αMjav B
αMinc A		0,0353	0,0302	0,0000	0,0000	0,0156	0,0000	0,0156
αMinc B			0,0342	0,0331	0,0518	0,0188	0,0405	0,0188
αMha				0,0292	0,0323	0,0414	0,0322	0,0414
αMare A					0,0000	0,0147	0,0000	0,0147
αMare B						0,0177	0,0000	0,0177
αMare C							0,0156	#DIV/0!
αMjav A								0,0156
αMjav B								

βCenH3			
	βMinc	βMare	βMjav
βMinc		0,5236	0,4878
βMare			0,7830
βMjav			

γCenH3			
	γMinc	γMare	γMjav
γMinc		0,0000	0,0000
γMare			#DIV/0!
γMjav			

HFD αβγδε																
	εMha	δMare	βMinc	βMare	βMjav	γMinc	γMare	γMjav	αMinc A	αMinc B	αMha	αMare A	αMare B	αMare C	αMjav A	αMjav B
εMha		/	/	/	/	/	/	/	0,0918	/	0,1270	0,0918	0,1359	0,0895	0,0918	0,0895
δMare			0,0430	0,0430	0,0430	0,1305	0,1228	0,1228	/	/	/	/	/	/	/	/
βMinc				#DIV/0!	#DIV/0!	/	0,0140	0,0140	/	/	/	/	0,0970	/	/	/
βMare					#DIV/0!	/	0,0140	0,0140	/	/	/	/	0,0970	/	/	/
βMjav						/	0,0140	0,0140	/	/	/	/	0,0970	/	/	/
γMinc							0,0000	0,0000	/	/	/	/	/	/	/	/
γMare								#DIV/0!	/	/	/	/	/	/	/	/
γMjav									/	/	/	/	/	/	/	/
αMinc A										0,0000	0,0212	0,0000	0,0000	0,0000	#DIV/0!	0,0000
αMinc B											0,0211	0,0000	0,0000	0,0000	0,0000	0,0000
αMha												0,0201	0,0213	0,0212	0,0212	0,0212
αMare A													0,0000	0,0000	0,0000	0,0000
αMare B														0,0000	0,0000	0,0000
αMare C															0,0000	#DIV/0!
αMjav A																0,0000
αMjav B																

N-tail βγδ							
	γMare	γMjav	γMinc	δMare	βMare	βMinc	βMjav
γMare		#DIV/0!	#DIV/0!	0,1531	0,2209	0,2142	0,2371
γMjav			#DIV/0!	0,1531	0,2209	0,2142	0,2371
γMinc				0,1531	0,2209	0,2142	0,2371
δMare					0,1449	0,1308	0,1591
βMare						0,3904	0,5726
βMinc							0,4774
βMjav							

N-tail βγδε								
	εMha	γMare	γMjav	γMinc	δMare	βMare	βMinc	βMjav
εMha		0,2407	0,2407	0,2407	0,1528	0,2244	0,2730	0,2628
γMare			#DIV/0!	#DIV/0!	0,1354	0,2181	0,1873	0,2107
γMjav				#DIV/0!	0,1354	0,2181	0,1873	0,2107
γMinc					0,1354	0,2181	0,1873	0,2107
δMare						0,1603	0,1162	0,1435
βMare							0,5235	0,7979
βMinc								0,4937
βMjav								

N-tail ββε					
	εMha	δMare	βMare	βMinc	βMjav
εMha		0,2940	0,3806	0,4393	0,4252
δMare			0,1723	0,1288	0,1569
βMare				0,5062	0,7710
βMinc					0,4774
βMjav					

N-tail γε				
	εMha	γMare	γMjav	γMinc
εMha		0,2332	0,2332	0,2332
γMare			#DIV/0!	#DIV/0!
γMjav				#DIV/0!
γMinc				

Supplementary Table 5. Analysis of the positive selection at individual codons on αCenH3s and also among different CenH3s using likelihood ratio test (LRT) values. LTR values with P-values <0.1 were considered as codons under positive selection (highligted in yellow)

αCenH3										βδCenH3									
Site	α	β^-	p^-	β^+	p^+	LRT	p-value	# branches under selection	Total branch length	Site	α	β^-	p^-	β^+	p^+	LRT	p-value	# branches under selection	Total branch length
1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0
2	10,41	0	0,01	0	0,99	0	0,67	0	1,84	2	0	0	0,46	25,68	0,54	5,82	0,02	0	7,39
3	2,09	0	0,02	0	0,98	0	0,67	0	0,37	3	0	0	1	0	0	0	1	0	0
4	1,52	0	0,05	0	0,95	0	0,67	0	0,27	4	0	0	1	0	0	0	1	0	0
5	0	0	1	0	0	0	1	0	0	5	0	0	0,71	25,51	0,29	3,37	0,09	1	4
6	0	0	1	0	0	0	1	0	0	6	0	0	1	0	0	0	1	0	0
7	0	0	1	0	0	0	1	0	0	7	0	0	0,79	760,7	0,21	6,82	0,01	1	84,75
8	0	0	1	0	0	0	1	0	0	8	0	0	1	0	0	0	1	0	0
9	0	0	1	0	0	0	1	0	0	9	8,63	0	0,18	0	0,82	0	0,67	0	1,48
10	0	0	1	0	0	0	1	0	0	10	0	0	0	2,84	1	1,27	0,27	1	1,53
11	3,41	0	0,2	0	0,8	0	0,67	0	0,6	11	0	0	1	0	0	0	1	0	0
12	9,34	0	0,02	0	0,98	0	0,67	0	1,65	12	0	0	1	0	0	0	1	0	0
13	5,65	0	0,2	0	0,8	0	0,67	0	1	13	0	0	0,72	32,44	0,28	3,79	0,07	0	4,87
14	0	0	1	0	0	0	1	0	0	14	0	0	1	0	0	0	1	0	0
15	0	0	1	0	0	0	1	0	0	15	20,77	0	0,05	0	0,95	0	0,67	0	3,57
16	0,73	0,73	1	0,8	0	0	0,67	0	0,46	16	96,21	0	0,02	0	0,98	0	0,67	0	16,53
17	0	0	1	0	0	0	1	0	0	17	0	0	0,64	14,08	0,36	3,69	0,07	1	2,75
18	0	0	1	0	0	0	1	0	0	18	0	0	1	0	0	0	1	0	0
19	0	0	1	0	0	0	1	0	0	19	1,33	0	0,15	0	0,85	0	0,67	0	0,23
20	0	0	1	0	0	0	1	0	0	20	1841,84	0	0,05	0	0,95	0	0,67	0	316,39
21	1,84	0	0,01	0	0,99	0	0,67	0	0,33	21	0,48	0,48	0,73	78,14	0,27	5,44	0,03	0	11,39
22	16,24	0	0,02	0	0,98	0	0,67	0	2,88	22	0	0	1	0	0	0	1	0	0
23	0	0	1	0	0	0	1	0	0	23	0	0	1	0	0	0	1	0	0
24	2,59	0	0,15	0	0,85	0	0,67	0	0,46	24	1,64	0	0,25	0	0,75	0	0,67	0	0,28
25	0	0	1	0	0	0	1	0	0	25	2,06	0	0,09	0	0,91	0	0,67	0	0,35
26	0	0	1	0	0	0	1	0	0	26	0	0	0,58	5,58	0,42	1,99	0,18	1	1,25
27	3,31	0	0,12	0	0,88	0	0,67	0	0,59	27	0	0	1	0	0	0	1	0	0
28	0	0	1	0	0	0	1	0	0	28	0	0	1	0	0	0	1	0	0
29	0	0	1	0	0	0	1	0	0	29	0	0	1	0	0	0	1	0	0
30	0	0	1	0	0	0	1	0	0	30	0	0	1	0	0	0	1	0	0
31	5,39	0	0,2	0	0,8	0	0,67	0	0,95	31	0	0	1	0	0	0	1	0	0
32	2,41	0	0	0	1	0	0,67	0	0,43	32	0	0	1	0	0	0	1	0	0
33	1	0	0,02	0	0,98	0	0,67	0	0,18	33	0	0	0	0,79	1	0,46	0,44	1	0,42
34	2,1	0	0,15	0	0,85	0	0,67	0	0,37	34	0	0	1	0	0	0	1	0	0
35	2,1	0	0,15	0	0,85	0	0,67	0	0,37	35	0	0	1	0	0	0	1	0	0
36	87,66	0	0,01	0	0,99	0	0,67	0	15,52	36	0	0	1	0	0	0	1	0	0
37	0	0	1	0	0	0	1	0	0	37	0	0	1	0	0	0	1	0	0
38	0	0	1	0	0	0	1	0	0	38	0	0	1	0	0	0	1	0	0
39	2,45	0	0,15	0	0,85	0	0,67	0	0,43	39	0	0	1	0	0	0	1	0	0
40	0	0	1	0	0	0	1	0	0	40	0	0	1	0	0	0	1	0	0
41	3,95	0	0,2	0	0,8	0	0,67	0	0,7	41	0	0	1	0	0	0	1	0	0
42	0	0	1	0	0	0	1	0	0	42	0	0	1	0	0	0	1	0	0
43	1,54	0	0,04	0	0,96	0	0,67	0	0,27	43	0	0	0	0,68	1	0,54	0,42	1	0,36
44	0	0	1	0	0	0	1	0	0	44	1,64	0	0,1	0	0,9	0	0,67	0	0,28
45	0	0	1	0	0	0	1	0	0	45	0	0	1	0	0	0	1	0	0
46	16,24	0	0,02	0	0,98	0	0,67	0	2,88	46	2,42	1,96	1	1,94	0	0	0,67	0	1,47
47	0	0	1	0	0	0	1	0	0	47	3,33	0	0,11	0	0,89	0	0,67	0	0,57
48	1,32	0	0,03	0	0,97	0	0,67	0	0,23	48	0	0	0	0,84	1	0,58	0,41	1	0,45
49	0	0	1	0	0	0	1	0	0	49	0	0	1	0	0	0	1	0	0
50	1,93	0	0,05	0	0,95	0	0,67	0	0,34	50	1,59	0	0,19	0	0,81	0	0,67	0	0,27
51	0	0	1	0	0	0	1	0	0	51	0	0	1	0	0	0	1	0	0
52	0	0	1	0	0	0	1	0	0	52	5,98	0	0,24	0	0,76	0	0,67	0	1,03
53	0	0	1	0	0	0	1	0	0	53	3,33	0	0,05	0	0,95	0	0,67	0	0,57
54	0	0	1	0	0	0	1	0	0	54	0	0	1	0	0	0	1	0	0
55	3,83	0	0,15	0	0,85	0	0,67	0	0,68	55	0	0	1	0	0	0	1	0	0
56	0	0	1	0	0	0	1	0	0	56	0	0	1	0	0	0	1	0	0
57	0	0	1	0	0	0	1	0	0	57	3,8	0	0,01	0	0,99	0	0,67	0	0,65
58	0	0	1	0	0	0	1	0	0	58	0	0	1	0	0	0	1	0	0
59	0	0	1	0	0	0	1	0	0	59	0	0	1	0	0	0	1	0	0
60	0	0	1	0	0	0	1	0	0	60	2,9	0	0,1	0	0,9	0	0,67	0	0,5
61	2,41	0	0,2	0	0,8	0	0,67	0	0,43	61	0	0	1	0	0	0	1	0	0
62	2,79	0	0,2	0	0,8	0	0,67	0	0,49	62	0	0	0	0,67	1	0,17	0,54	1	0,36
63	0	0	1	0	0	0	1	0	0	63	0	0	1	0	0	0	1	0	0
64	0	0	1	0	0	0	1	0	0	64	0	0	1	0	0	0	1	0	0
65	4,25	0	0,15	0	0,85	0	0,67	0	0,75	65	0	0	1	0	0	0	1	0	0
66	0	0	1	0	0	0	1	0	0	66	3,66	0	0,17	0	0,83	0	0,67	0	0,63
67	3,31	0	0,12	0	0,88	0	0,67	0	0,59	67	0	0	1	0	0	0	1	0	0
68	0	0	1	0	0	0	1	0	0	68	0	0	1	0	0	0	1	0	0
69	0	0	1	0	0	0	1	0	0	69	0	0	1	0	0	0	1	0	0
70	3,68	0	0,2	0	0,8	0	0,67	0	0,65	70	2,47	0	0,14	0	0,86	0	0,67	0	0,42
71	1,34	0	0,05	0	0,95	0	0,67	0	0,24	71	0	0	1	0	0	0	1	0	0
72	0	0	1	0	0	0	1	0	0	72	1,81	0	0,09	0	0,91	0	0,67	0	0,31
73	0	0	1	0	0	0	1	0	0	73	0	0	1	0	0	0	1	0	0
74	0	0	1	0	0	0	1	0	0	74	0	0	1	0	0	0	1	0	0
75	0	0	1	0	0	0	1	0	0	75	0	0	1	0	0	0	1	0	0
76	0	0	1	0	0	0	1	0	0	76	0	0	1	0	0	0	1	0	0
77	0	0	1	0	0	0	1	0	0	77	0	0	1	0	0	0	1	0	0
78	2,7	0	0,01	0	0,99	0	0,67	0	0,48	78	0	0	1	0	0	0	1	0	0
79	3,04	0	0,02	0	0,98	0	0,67	0	0,54	79	0	0	1	0	0	0	1	0	0
80	1,61	0	0,02	0	0,98	0	0,67	0	0,29	80	0	0	0	0,87	1	0,57	0,41	1	0,47
81	1,52	0	0,04	0	0,96	0	0,67	0	0,27	81	0	0	1	0	0	0	1	0	0
82	5	0	0,2	0	0,8	0	0,67	0	0,89	82	0	0	1	0	0	0	1	0	0

83	0	0	1	0	0	0	1	0	0
84	0	0	1	0	0	0	1	0	0
85	2,36	0	0,19	0	0,81	0	0,67	0	0,42
86	0	0	1	0	0	0	1	0	0
87	0	0	1	0	0	0	1	0	0
88	1,41	0	0,02	0	0,98	0	0,67	0	0,25
89	0	0	1	0	0	0	1	0	0
90	0	0	1	0	0	0	1	0	0
91	0	0	1	0	0	0	1	0	0
92	0	0	1	0	0	0	1	0	0
93	2,1	0	0,15	0	0,85	0	0,67	0	0,37
94	0	0	1	0	0	0	1	0	0
95	4,25	0	0,15	0	0,85	0	0,67	0	0,75
96	0	0	0	0,83	1	0,52	0,43	1	0,37
97	0	0	1	0	0	0	1	0	0
98	0	0	1	0	0	0	1	0	0
99	0	0	1	0	0	0	1	0	0
100	0	0	1	0	0	0	1	0	0
101	0	0	1	0	0	0	1	0	0
102	0	0	1	0	0	0	1	0	0
103	0	0	1	0	0	0	1	0	0
104	10,26	0	0,05	0	0,95	0	0,67	0	1,82
105	0	0	1	0	0	0	1	0	0
106	0	0	1	0	0	0	1	0	0
107	0	0	1	0	0	0	1	0	0
108	0	0	1	0	0	0	1	0	0
109	5,39	0	0,2	0	0,8	0	0,67	0	0,95
110	0	0	1	0	0	0	1	0	0
111	0	0	1	0	0	0	1	0	0
112	0	0	1	0	0	0	1	0	0
113	0	0	1	0	0	0	1	0	0
114	0	0	1	0	0	0	1	0	0
115	0	0	1	0	0	0	1	0	0
116	0	0	1	0	0	0	1	0	0
117	2,15	0	0,01	0	0,99	0	0,67	0	0,38
118	5,36	0	0,04	0	0,96	0	0,67	0	0,95
119	0	0	1	0	0	0	1	0	0
120	0	0	1	0	0	0	1	0	0
121	17,26	0	0,05	0	0,95	0	0,67	0	3,06
122	1,51	0	0,01	0	0,99	0	0,67	0	0,27
123	0	0	1	0	0	0	1	0	0
124	0	0	1	0	0	0	1	0	0
125	0	0	1	0	0	0	1	0	0
126	0	0	1	0	0	0	1	0	0
127	0	0	1	0	0	0	1	0	0
128	8,59	0	0,01	0	0,99	0	0,67	0	1,52
129	5,36	0	0,19	0	0,81	0	0,67	0	0,95
130	0	0	1	0	0	0	1	0	0
131	1,89	0	0,05	0	0,95	0	0,67	0	0,34
132	6,75	0	0,2	0	0,8	0	0,67	0	1,19
133	0	0	1	0	0	0	1	0	0
134	1,93	0	0,14	0	0,86	0	0,67	0	0,34
135	8,51	0	0,05	0	0,95	0	0,67	0	1,51
136	2,1	0	0,01	0	0,99	0	0,67	0	0,37
137	0	0	1	0	0	0	1	0	0
138	7,65	0	0,14	0	0,86	0	0,67	0	1,35
139	0	0	1	0	0	0	1	0	0
140	0	0	1	0	0	0	1	0	0
141	16,24	0	0,02	0	0,98	0	0,67	0	2,88
142	0	0	1	0	0	0	1	0	0
143	0	0	1	0	0	0	1	0	0
144	3,31	0	0,02	0	0,98	0	0,67	0	0,59
145	0	0	1	0	0	0	1	0	0
146	0	0	1	0	0	0	1	0	0
147	0	0	1	0	0	0	1	0	0
148	0	0	1	0	0	0	1	0	0
149	0	0	1	0	0	0	1	0	0
150	2,81	0	0,15	0	0,85	0	0,67	0	0,5
151	0	0	1	0	0	0	1	0	0
152	0	0	1	0	0	0	1	0	0
153	0	0	1	0	0	0	1	0	0
154	0	0	1	0	0	0	1	0	0
155	0	0	1	0	0	0	1	0	0
156	0	0	1	0	0	0	1	0	0
157	0	0	1	0	0	0	1	0	0
158	0	0	0	1,06	1	0,81	0,35	1	0,47
159	15,64	0	0,02	0	0,98	0	0,67	0	2,77
160	1,84	0	0,02	0	0,98	0	0,67	0	0,33
161	0	0	0	0,76	1	0,48	0,44	1	0,34
162	0	0	0,65	5,57	0,35	1,43	0,25	1	0,87
163	0	0	0,7	9,33	0,3	1,88	0,19	1	1,24
164	0	0	1	0	0	0	1	0	0
165	0	0	1	0	0	0	1	0	0

83	3,24	0	0,09	0	0,91	0	0,67	0	0,56
84	3,6	0	0,1	0	0,9	0	0,67	0	0,62
85	5,98	0	0,18	0	0,82	0	0,67	0	1,03
86	0	0	1	0	0	0	1	0	0
87	5,98	0	0,18	0	0,82	0	0,67	0	1,03
88	0	0	1	0	0	0	1	0	0
89	2,43	0	0,18	0	0,82	0	0,67	0	0,42
90	5,98	0	0,18	0	0,82	0	0,67	0	1,03
91	5,62	0	0,05	0	0,95	0	0,67	0	0,96
92	0	0	1	0	0	0	1	0	0
93	0	0	1	0	0	0	1	0	0
94	2,13	0	0,05	0	0,95	0	0,67	0	0,37
95	0	0	1	0	0	0	1	0	0
96	0	0	1	0	0	0	1	0	0
97	4,76	1,83	0	1,83	1	0	0,67	0	1,8
98	0	0	1	0	0	0	1	0	0
99	0	0	1	0	0	0	1	0	0
100	0	0	1	0	0	0	1	0	0
101	2,29	0	0,1	0	0,9	0	0,67	0	0,39
102	2,23	0,83	1	0,83	0	0	0,67	0	0,83
103	0	0	1	0	0	0	1	0	0
104	0	0	1	0	0	0	1	0	0
105	0	0	1	0	0	0	1	0	0
106	0	0	1	0	0	0	1	0	0
107	0	0	1	0	0	0	1	0	0
108	0	0	1	0	0	0	1	0	0
109	4,62	2,1	1	2,08	0	0	0,67	0	1,92
110	0	0	1	0	0	0	1	0	0
111	0	0	1	0	0	0	1	0	0
112	0	0	1	0	0	0	1	0	0
113	2,23	0	0,05	0	0,95	0	0,67	0	0,38
114	0	0	1	0	0	0	1	0	0
115	1,69	0	0,01	0	0,99	0	0,67	0	0,29
116	2,31	0,8	0	0,8	1	0	0,67	0	0,83
117	0	0	1	0	0	0	1	0	0
118	3,04	0	0,1	0	0,9	0	0,67	0	0,52
119	1,81	0	0,09	0	0,91	0	0,67	0	0,31
120	0	0	1	0	0	0	1	0	0
121	2,06	0	0,11	0	0,89	0	0,67	0	0,35
122	0	0	1	0	0	0	1	0	0
123	0	0	1	0	0	0	1	0	0
124	3,16	0	0,05	0	0,95	0	0,67	0	0,54
125	9,2	0	0,05	0	0,95	0	0,67	0	1,58
126	3,03	0	0,09	0	0,91	0	0,67	0	0,52
127	0	0	1	0	0	0	1	0	0
128	3,33	0	0,11	0	0,89	0	0,67	0	0,57
129	2,06	0	0,11	0	0,89	0	0,67	0	0,35
130	3,59	0	0,07	0	0,93	0	0,67	0	0,62
131	0	0	1	0	0	0	1	0	0
132	5,98	0	0,18	0	0,82	0	0,67	0	1,03
133	1,27	0	0,05	0	0,95	0	0,67	0	0,22
134	0	0	1	0	0	0	1	0	0
135	0	0	1	0	0	0	1	0	0
136	0	0	1	0	0	0	1	0	0
137	0	0	1	0	0	0	1	0	0
138	3,03	0	0,07	0	0,93	0	0,67	0	0,52
139	0	0	1	0	0	0	1	0	0
140	0	0	0	0,89	1	0,73	0,37	1	0,48
141	0	0	1	0	0	0	1	0	0
142	1,69	0	0,01	0	0,99	0	0,67	0	0,29
143	5,99	0	0,17	0	0,83	0	0,67	0	1,03
144	2,12	0	0,05	0	0,95	0	0,67	0	0,36
145	0	0	1	0	0	0	1	0	0
146	0	0	1	0	0	0	1	0	0
147	2,47	0	0,05	0	0,95	0	0,67	0	0,42
148	0	0	1	0	0	0	1	0	0
149	0	0	1	0	0	0	1	0	0
150	3,33	0	0,05	0	0,95	0	0,67	0	0,57
151	0	0	1	0	0	0	1	0	0
152	7,26	0	0,05	0	0,95	0	0,67	0	1,25
153	0	0	1	0	0	0	1	0	0
154	0	0	1	0	0	0	1	0	0
155	0	0	1	0	0	0	1	0	0
156	0	0	1	0	0	0	1	0	0
157	2,34	1,05	0	1,05	1	0	0,67	0	0,97
158	2,23	0	0,05	0	0,95	0	0,67	0	0,38
159	0	0	1	0	0	0	1	0	0
160	1,82	0	0,07	0	0,93	0	0,67	0	0,31
161	2,04	0	0,09	0	0,91	0	0,67	0	0,35
162	0	0	1	0	0	0	1	0	0
163	1,82	0	0,07	0	0,93	0	0,67	0	0,31
164	1,13	0	0,05	0	0,95	0	0,67	0	0,19
165	3,16	0	0,05	0	0,95	0	0,67	0	0,54
166	0	0	1	0	0	0	1	0	0
167	0	0	1	0	0	0	1	0	0
168	0	0	1	0	0	0	1	0	0
169	0	0	1	0	0	0	1	0	0
170	0	0	1	0	0	0	1	0	0
171	0	0	1	0	0	0	1	0	0
172	0	0	1	0	0	0	1	0	0
173	0	0	1	0	0	0	1	0	0
174	0	0	1	0	0	0	1	0	0
175	1,79	1,79	0	3,57	1	0,09	0,58	0	2,22
176	3,59	0	0,07	0	0,93	0	0,67	0	0,62
177	0	0	1	0	0	0	1	0	0
178	0	0	1	0	0	0	1	0	0
179	3,03	0	0,07	0	0,93	0	0,67	0	0,52
180	0	0	1	0	0	0	1	0	0

βCenH3, γCenH3 and δCenH3									
Site	α	β^-	p^-	β^+	p^+	LRT	p-value	# branches under selection	Total branch length
1	0	0	1	0	0	0	1	0	0
2	0	0	0,67	30,88	0,33	9,56	0	0	8,53
3	0	0	1	0	0	0	1	0	0
4	0	0	1	0	0	0	1	0	0
5	0	0	0,83	32,42	0,17	3,48	0,08	1	4,71
6	0	0	1	0	0	0	1	0	0
7	0	0	0,87	930,23	0,13	9,77	0	1	101,65
8	0	0	1	0	0	0	1	0	0
9	0,93	0	0,05	0	0,95	0	0,67	0	0,25
10	0	0	0	1,17	1	0,95	0,33	1	0,99
11	0	0	1	0	0	0	1	0	0
12	0	0	1	0	0	0	1	0	0
13	0	0	0,83	35,2	0,17	3,77	0,07	1	5,15
14	0	0	1	0	0	0	1	0	0
15	26,11	0	0,05	0	0,95	0	0,67	0	7,04
16	45,92	0	0,05	0	0,95	0	0,67	0	12,38
17	0	0	0,81	19,66	0,19	4,16	0,06	1	3,21
18	3,36	0	0,09	0	0,91	0	0,67	0	0,91
19	0,71	0	0,02	0	0,98	0	0,67	0	0,19
20	0,81	0	0,02	0	0,98	0	0,67	0	0,22
21	0,66	0,66	0,83	107,38	0,17	6,66	0,02	0	16,07
22	0,91	0	0,02	0	0,98	0	0,67	0	0,25
23	1,42	0	0,12	0	0,88	0	0,67	0	0,38
24	1,48	0,54	1	0,31	0	0	0,67	0	0,85
25	2,32	0,49	0,29	0,48	0,71	0	0,67	0	1,03
26	0	0	0	0,76	1	0,51	0,43	1	0,64
27	0	0	0	0,47	1	0,31	0,49	0	0,4
28	0	0	0,57	2,19	0,43	0,73	0,37	1	0,8
29	1,78	0	0,1	0	0,9	0	0,67	0	0,48
30	0	0	1	0	0	0	1	0	0
31	0	0	1	0	0	0	1	0	0
32	0	0	1	0	0	0	1	0	0
33	0	0	0	0,47	1	0,57	0,41	0	0,4
34	2,25	0,37	1	0,23	0	0	0,67	0	0,92
35	0	0	0	0,35	1	0,27	0,5	0	0,3
36	0	0	1	0	0	0	1	0	0
37	0	0	0	0,41	1	0,57	0,41	0	0,35
38	0	0	1	0	0	0	1	0	0
39	0	0	1	0	0	0	1	0	0
40	0	0	1	0	0	0	1	0	0
41	0	0	0	1,18	1	0,83	0,35	0	1
42	0	0	1	0	0	0	1	0	0
43	1,44	0,43	1	0,3	0	0	0,67	0	0,75
44	3,03	3,03	0,37	0	0,63	0	0,67	0	1,75
45	0	0	1	0	0	0	1	0	0
46	3,11	2,97	0,4	0	0,6	0	0,67	0	1,83
47	3,75	0	0,02	0	0,98	0	0,67	0	1,01
48	2,36	0,51	1	0,37	0	0	0,67	0	1,06
49	3,36	0	0,09	0	0,91	0	0,67	0	0,91
50	0,84	0	0,05	0	0,95	0	0,67	0	0,23
51	0	0	0	0,47	1	0,59	0,41	0	0,4
52	7,83	0	0,02	0	0,98	0	0,67	0	2,11
53	3,75	0	0,01	0	0,99	0	0,67	0	1,01
54	0	0	1	0	0	0	1	0	0
55	0	0	1	0	0	0	1	0	0
56	6,08	0	0	0	1	0	0,67	0	1,64
57	4,57	0	0,15	0	0,85	0	0,67	0	1,23
58	0	0	0	0,63	1	0,55	0,42	0	0,53
59	0	0	1	0	0	0	1	0	0
60	3,7	0	0,09	0	0,91	0	0,67	0	1
61	0	0	1	0	0	0	1	0	0
62	0	0	0,55	1,14	0,45	0,2	0,53	0	0,44
63	2,55	0	0,18	0	0,82	0	0,67	0	0,69
64	0	0	1	0	0	0	1	0	0
65	0	0	1	0	0	0	1	0	0
66	4,39	0	0,05	0	0,95	0	0,67	0	1,18
67	0	0	1	0	0	0	1	0	0
68	0	0	1	0	0	0	1	0	0
69	0	0	1	0	0	0	1	0	0
70	1,29	0	0,12	0	0,88	0	0,67	0	0,35
71	0	0	1	0	0	0	1	0	0
72	1,33	0	0,19	0	0,81	0	0,67	0	0,36
73	0	0	1	0	0	0	1	0	0
74	0	0	1	0	0	0	1	0	0
75	0	0	1	0	0	0	1	0	0
76	0	0	1	0	0	0	1	0	0
77	0	0	1	0	0	0	1	0	0
78	0	0	1	0	0	0	1	0	0
79	1,04	0	0,05	0	0,95	0	0,67	0	0,28
80	0	0	0	0,73	1	0,68	0,38	0	0,61
81	0	0	1	0	0	0	1	0	0
82	0	0	1	0	0	0	1	0	0
83	1,06	0	0,01	0	0,99	0	0,67	0	0,29
84	2,36	0	0,07	0	0,93	0	0,67	0	0,64
85	2,62	0	0,18	0	0,82	0	0,67	0	0,71
86	0	0	1	0	0	0	1	0	0
87	2,62	0	0,02	0	0,98	0	0,67	0	0,71
88	0	0	1	0	0	0	1	0	0
89	1,41	0	0,19	0	0,81	0	0,67	0	0,38
90	2,62	0	0,02	0	0,98	0	0,67	0	0,71

βCenH3, γCenH3, δCenH3 and εCenH3									
Site	α	β^-	p^-	β^+	p^+	LRT	p-value	# branches under selection	Total branch length
1	0	0	1	0	0	0	1	0	0
2	0	0	1	0	0	0	1	0	0
3	12,46	12,46	0,29	0	0,71	0	0,67	0	20,29
4	0,83	0,56	1	0,53	0	0	0,67	0	2,2
5	0	0	1	0	0	0	1	0	0
6	0	0	0,85	17,04	0,15	4,32	0,05	1	6,63
7	0	0	0,27	0,57	0,73	1,1	0,3	1	1,11
8	0	0	0,72	29,3	0,28	5,87	0,02	2	21,63
9	0	0	0,31	0,52	0,69	1,31	0,27	1	0,96
10	1,07	0,24	1	0,24	0	0	0,67	0	1,56
11	1,24	1,24	0,33	0	0,67	0	0,67	0	2,17
12	0,13	0,13	0	0,32	1	0,06	0,6	0	0,98
13	0	0	1	0	0	0	1	0	0
14	0	0	0,72	23,42	0,28	1,41	0,25	1	17,73
15	1,25	0,13	0	0,13	1	0	0,67	0	1,42
16	24,73	0	0,19	0	0,81	0	0,67	0	21,21
17	41,26	0	0,02	0	0,98	0	0,67	0	35,4
18	0,1	0,1	1	0	0	0	0,67	0	0,36
19	3,41	0	0,1	0	0,9	0	0,67	0	2,92
20	0,69	0	0,2	0	0,8	0	0,67	0	0,59
21	20,8	0	0,18	0	0,82	0	0,67	0	17,85
22	0,8	0,32	0,84	50,13	0,16	3,82	0,07	1	22,39
23	1,08	0	0,02	0	0,98	0	0,67	0	0,93
24	1,8	0	0,02	0	0,98	0	0,67	0	1,55
25	0,97	0,4	1	0,33	0	0	0,67	0	1,91
26	1,92	0,73	1	0,68	0	0	0,67	0	3,6
27	0	0	0	1,06	1	0,7	0,38	1	2,84
28	0,34	0,34	0,52	0	0,48	0	0,67	0	0,77
29	0,71	0,71	0,61	0,76	0,39	0	0,66	0	2,56
30	1,9	0	0,02	0	0,98	0	0,67	0	1,63
31	0	0	1	0	0	0	1	0	0
32	0	0	1	0	0	0	1	0	0
33	0	0	1	0	0	0	1	0	0
34	0	0	0	0,45	1	0,58	0,41	0	1,19
35	2,19	0,34	1	0,21	0	0	0,67	0	2,79
36	0	0	0	0,32	1	0,23	0,52	0	0,86
37	0	0	1	0	0	0	1	0	0
38	0,7	0,39	1	0,37	0	0	0,67	0	1,64
39	0,38	0	0,02	0	0,98	0	0,67	0	0,32
40	1,25	0	0,02	0	0,98	0	0,67	0	1,07
41	0	0	0	0,19	1	0,52	0,42	0	0,51
42	1,45	1,45	0,42	0	0,58	0	0,67	0	2,88
43	0,16	0,16	1	0,14	0	0	0,67	0	0,57
44	1,31	0,65	0,41	0	0,59	0	0,67	0	1,84
45	3,19	3,19	0,24	0	0,76	0	0,67	0	4,81
46	0,36	0,35	0,97	0,35	0,03	0	0,67	0	1,25
47	2,95	2,85	0,33	0,21	0,67	0	0,67	0	5,45
48	3,54	0	0,02	0	0,98	0	0,67	0	3,04
49	2,35	0,77	0,4	0	0,6	0	0,67	0	2,85
50	3,41	0	0,05	0	0,95	0	0,67	0	2,92
51	0,78	0	0,19	0	0,81	0	0,67	0	0,67
52	0,11	0,11	1	0	0	0	0,67	0	0,39
53	7,27	0	0,06	0	0,94	0	0,67	0	6,23
54	3,62	0,14	1	0,14	0	0	0,67	0	3,48
55	0,7	0	0,18	0	0,82	0	0,67	0	0,6
56	0	0	1	0	0	0	1	0	0
57	6,19	0,32	0,35	0,32	0,65	0	0,67	0	6,16
58	4,54	0,3	1	0,21	0	0	0,67	0	4,68
59	0	0	0	0,51	1	1,12	0,3	0	1,37
60	0	0	0	0,2	1	0,8	0,36	1	0,54
61	2,68	0,17	0	0,17	1	0	0,67	0	2,74
62	0	0	0	0,14	1	0,56	0,41	1	0,39
63	0,95	0,6	1	0,57	0	0	0,67	0	2,43
64	2,78	0,47	0,57	0,47	0,43	0	0,67	0	3,64
65	0,59	0	0,2	0	0,8	0	0,67	0	0,51
66	0,68	0	0,18	0	0,82	0	0,67	0	0,58
67	3,87	0,14	1	0,14	0	0	0,67	0	3,68
68	0	0	1	0	0	0	1	0	0
69	0	0	0	0,15	1	0,45	0,44	1	0,39
70	0,56	0,15	1	0,14	0	0	0,67	0	0,88
71	0,67	0	0,07	0	0,93	0	0,67	0	0,58
72	0,71	0	0,25	0	0,75	0	0,67	0	0,61
73	1,46	0	0,02	0	0,98	0	0,67	0	1,25
74	0,49	0	0,04	0	0,96	0	0,67	0	0,42
75	0	0	1	0	0	0	1	0	0
76	0,33	0	0,02	0	0,98	0	0,67	0	0,28
77	0,7	0	0,19	0	0,81	0	0,67	0	0,6
78	1,23	0	0,2	0	0,8	0	0,67	0	1,05
79	0,48	0,18	1	0,17	0	0	0,67	0	0,89
80	1,07	0	0,1	0	0,9	0	0,67	0	0,92
81	0,94	0,94	0,36	0	0,64	0	0,67	0	1,71
82	0	0	1	0	0	0	1	0	0
83	0	0	1	0	0	0	1	0	0
84	1,02	0	0,09	0	0,91	0	0,67	0	0,88
85	1,92	0	0,05	0	0,95	0	0,67	0	1,64
86	2,44	0	0,03	0	0,97	0	0,67	0	2,09
87	0	0	1	0	0	0	1	0	0
88	2,29	0,41	0,77	0	0,23	0	0,67	0	2,81
89	0,71	0	0,19	0	0,81	0	0,67	0	0,61
90	0,59	0,33	1	0,33	0	0	0,67	0	1,38

91	3,57	0	0,14	0	0,86	0	0,67	0	0,96
92	2,45	0	0,07	0	0,93	0	0,67	0	0,66
93	0	0	1	0	0	0	1	0	0
94	2,19	0	0,06	0	0,94	0	0,67	0	0,59
95	0,91	0	0,05	0	0,95	0	0,67	0	0,24
96	0,93	0,41	1	0,33	0	0	0,67	0	0,6
97	5,24	1,81	0,89	1,8	0,11	0	0,67	0	2,94
98	0	0	1	0	0	0	1	0	0
99	2,45	0	0,07	0	0,93	0	0,67	0	0,66
100	1,17	0	0,2	0	0,8	0	0,67	0	0,32
101	2,52	0	0,19	0	0,81	0	0,67	0	0,68
102	0,91	0,63	1	0,49	0	0	0,67	0	0,78
103	1,21	0	0,19	0	0,81	0	0,67	0	0,32
104	0	0	0	0,41	1	0,48	0,43	0	0,35
105	0,91	0	0,02	0	0,98	0	0,67	0	0,25
106	0	0	0	0,44	1	0,59	0,4	0	0,37
107	0	0	1	0	0	0	1	0	0
108	0	0	1	0	0	0	1	0	0
109	6,06	6,06	0,35	0	0,65	0	0,67	0	3,44
110	0	0	1	0	0	0	1	0	0
111	0	0	1	0	0	0	1	0	0
112	0	0	1	0	0	0	1	0	0
113	2,38	0	0,07	0	0,93	0	0,67	0	0,64
114	1,42	0	0,19	0	0,81	0	0,67	0	0,38
115	1,58	0	0,2	0	0,8	0	0,67	0	0,43
116	9,52	0,49	1	0,43	0	0	0,67	0	2,98
117	0	0	1	0	0	0	1	0	0
118	2,92	0	0,09	0	0,91	0	0,67	0	0,79
119	1,33	0	0,19	0	0,81	0	0,67	0	0,36
120	0	0	1	0	0	0	1	0	0
121	0,87	0	0,05	0	0,95	0	0,67	0	0,23
122	0	0	1	0	0	0	1	0	0
123	0	0	1	0	0	0	1	0	0
124	3,99	0	0,02	0	0,98	0	0,67	0	1,08
125	8,11	0	0,07	0	0,93	0	0,67	0	2,19
126	2,92	0	0,07	0	0,93	0	0,67	0	0,79
127	0	0	1	0	0	0	1	0	0
128	1,01	0	0,15	0	0,85	0	0,67	0	0,27
129	0,87	0	0,05	0	0,95	0	0,67	0	0,23
130	2,36	0	0,06	0	0,94	0	0,67	0	0,64
131	6,08	0	0,05	0	0,95	0	0,67	0	1,64
132	2,62	0	0,18	0	0,82	0	0,67	0	0,71
133	2,22	0	0,1	0	0,9	0	0,67	0	0,6
134	0	0	1	0	0	0	1	0	0
135	0	0	1	0	0	0	1	0	0
136	0	0	1	0	0	0	1	0	0
137	0	0	1	0	0	0	1	0	0
138	3,71	0	0,01	0	0,99	0	0,67	0	1
139	1,42	0	0,18	0	0,82	0	0,67	0	0,38
140	3,94	1,3	0,5	0	0,5	0	0,67	0	1,61
141	0	0	1	0	0	0	1	0	0
142	2,94	0	0,04	0	0,96	0	0,67	0	0,79
143	2,61	0	0,05	0	0,95	0	0,67	0	0,7
144	1	0	0,07	0	0,93	0	0,67	0	0,27
145	0	0	1	0	0	0	1	0	0
146	0	0	1	0	0	0	1	0	0
147	1,29	0	0,13	0	0,87	0	0,67	0	0,35
148	1,78	0	0,1	0	0,9	0	0,67	0	0,48
149	4,01	0	0,07	0	0,93	0	0,67	0	1,08
150	2,15	0	0,17	0	0,83	0	0,67	0	0,58
151	0	0	1	0	0	0	1	0	0
152	5,67	0	0,14	0	0,86	0	0,67	0	1,53
153	0	0	1	0	0	0	1	0	0
154	0	0	1	0	0	0	1	0	0
155	0	0	1	0	0	0	1	0	0
156	0	0	1	0	0	0	1	0	0
157	1,21	0,78	0,89	0	0,11	0	0,67	0	0,91
158	1,41	0,4	1	0,32	0	0	0,67	0	0,72
159	0	0	1	0	0	0	1	0	0
160	1,09	0	0,02	0	0,98	0	0,67	0	0,29
161	1,55	0	0,18	0	0,82	0	0,67	0	0,42
162	0	0	1	0	0	0	1	0	0
163	1,09	0	0,02	0	0,98	0	0,67	0	0,29
164	0,75	0	0,01	0	0,99	0	0,67	0	0,2
165	0,9	0	0,02	0	0,98	0	0,67	0	0,24
166	0	0	1	0	0	0	1	0	0
167	4,55	0,57	0,71	0,58	0,29	0	0,67	0	1,71
168	0	0	0,75	7,77	0,25	3,1	0,1	1	1,65
169	0	0	1	0	0	0	1	0	0
170	0	0	0	0,58	1	0,53	0,42	0	0,49
171	0	0	0	0,64	1	0,66	0,39	0	0,54
172	0	0	0	1,95	1	2,86	0,11	0	1,64
173	0	0	1	0	0	0	1	0	0
174	0	0	1	0	0	0	1	0	0
175	0	0	1	0	0	0	1	0	0
176	2,39	2,39	0,37	2,77	0,63	0	0,65	0	2,86
177	2,36	0	0,06	0	0,94	0	0,67	0	0,64
178	0	0	1	0	0	0	1	0	0
179	0	0	0	0,43	1	0,61	0,4	0	0,36
180	2,92	0	0,09	0	0,91	0	0,67	0	0,79
181	0	0	1	0	0	0	1	0	0

91	2,43	0	0,02	0	0,98	0	0,67	0	2,09
92	3,91	0	0,1	0	0,9	0	0,67	0	3,35
93	2,27	0,18	1	0,15	0	0	0,67	0	2,44
94	0,35	0	0,05	0	0,95	0	0,67	0	0,3
95	1,87	0	0,01	0	0,99	0	0,67	0	1,6
96	0,24	0	0,04	0	0,96	0	0,67	0	0,21
97	1,12	0,31	1	0,31	0	0	0,67	0	1,79
98	5,48	5,48	0,2	0	0,8	0	0,67	0	7,57
99	0	0	1	0	0	0	1	0	0
100	1,55	0,2	0	0,2	1	0	0,67	0	1,86
101	0,51	0	0,02	0	0,98	0	0,67	0	0,44
102	2	0	0,02	0	0,98	0	0,67	0	1,72
103	1,17	1,17	0,33	0	0,67	0	0,67	0	2,02
104	0,96	0	0,1	0	0,9	0	0,67	0	0,82
105	0,59	0,59	0,43	0	0,57	0	0,67	0	1,18
106	0,88	0	0,18	0	0,82	0	0,67	0	0,76
107	0,58	0,58	0,41	0	0,59	0	0,67	0	1,13
108	0,67	0	0,18	0	0,82	0	0,67	0	0,57
109	0	0	0	0,15	1	0,43	0,45	1	0,41
110	1,22	0,31	0,76	11,24	0,24	-0,2	0,67	0	8,77
111	0,71	0,2	1	0,19	0	0	0,67	0	1,13
112	0	0	1	0	0	0	1	0	0
113	0,38	0	0,01	0	0,99	0	0,67	0	0,32
114	1,64	0,15	1	0,15	0	0	0,67	0	1,8
115	0,93	0,42	1	0,4	0	0	0,67	0	1,92
116	1,1	0,17	0	0,17	1	0	0,67	0	1,39
117	898,78	0,86	0,39	0	0,61	0	0,67	0	771,99
118	0	0	1	0	0	0	1	0	0
119	2,56	0	0,08	0	0,92	0	0,67	0	2,2
120	0,8	0	0,01	0	0,99	0	0,67	0	0,68
121	0,35	0	0,01	0	0,99	0	0,67	0	0,3
122	0,24	0,13	1	0,13	0	0	0,67	0	0,56
123	0,81	0	0,02	0	0,98	0	0,67	0	0,69
124	0,67	0,18	1	0,18	0	0	0,67	0	1,05
125	3,47	0	0,01	0	0,99	0	0,67	0	2,98
126	5,09	0	0,07	0	0,93	0	0,67	0	4,37
127	2,56	0	0,09	0	0,91	0	0,67	0	2,2
128	0,49	0	0,04	0	0,96	0	0,67	0	0,42
129	1,09	0	0,02	0	0,98	0	0,67	0	0,94
130	0,97	0	0,07	0	0,93	0	0,67	0	0,83
131	2,02	0	0,02	0	0,98	0	0,67	0	1,73
132	6,09	0	0,02	0	0,98	0	0,67	0	5,23
133	2,44	0	0,03	0	0,97	0	0,67	0	2,09
134	2,45	0	0,07	0	0,93	0	0,67	0	2,1
135	0	0	1	0	0	0	1	0	0
136	0,67	0	0,18	0	0,82	0	0,67	0	0,57
137	49,8	0,14	1	0,01	0	0	0,67	0	43,11
138	1,25	0	0,02	0	0,98	0	0,67	0	1,07
139	3,27	0	0,05	0	0,95	0	0,67	0	2,81
140	0,68	0	0,1	0	0,9	0	0,67	0	0,58
141	3,58	2,31	0,24	0	0,76	0	0,67	0	4,57
142	0	0	1	0	0	0	1	0	0
143	2,26	0	0,02	0	0,98	0	0,67	0	1,94
144	2,44	0	0,1	0	0,9	0	0,67	0	2,09
145	1,11	0	0,02	0	0,98	0	0,67	0	0,95
146	0,3	0	0,01	0	0,99	0	0,67	0	0,26
147	0	0	1	0	0	0	1	0	0
148	1,4	0	0,01	0	0,99	0	0,67	0	1,21
149	2,29	0,13	1	0,13	0	0	0,67	0	2,32
150	4,28	0	0	0	1	0	0,67	0	3,67
151	2,44	0	0,05	0	0,95	0	0,67	0	2,09
152	0,7	0	0,18	0	0,82	0	0,67	0	0,6
153	5,08	0	0,09	0	0,91	0	0,67	0	4,36
154	0	0	1	0	0	0	1	0	0
155	0,42	0,15	0	0,15	1	0	0,67	0	0,76
156	0,49	0	0,05	0	0,95	0	0,67	0	0,42
157	0	0	1	0	0	0	1	0	0
158	1,21	1,21	0,32	0	0,68	0	0,67	0	2,08
159	1,72	0,54	1	0,54	0	0	0,67	0	2,92
160	0	0	1	0	0	0	1	0	0
161	0,79	0	0,14	0	0,86	0	0,67	0	0,68
162	1,41	0,22	1	0,22	0	0	0,67	0	1,78
163	0	0	1	0	0	0	1	0	0
164	1,74	0	0,18	0	0,82	0	0,67	0	1,5
165	0,81	0	0,19	0	0,81	0	0,67	0	0,7
166	1,01	0	0,07	0	0,93	0	0,67	0	0,87
167	0	0	1	0	0	0	1	0	0
168	5,06	1,09	0,35	0	0,65	0	0,67	0	5,37
169	0	0	0,63	8,91	0,37	4,88	0,04	0	8,79
170	0	0	1	0	0	0	1	0	0
171	0	0	0	0,82	1	0,93	0,33	1	2,18
172	0	0	0	0,44	1	1,55	0,23	0	1,19
173	0	0	0	2,29	1	2,62	0,13	1	6,11
174	0,95	0,17	1	0,13	0	0	0,67	0	1,27
175	2,11	0,38	1	0,41	0	0	0,67	0	2,83
176	0,14	0,14	1	0,13	0	0	0,67	0	0,5
177	2,17	2,17	0,08	2,26	0,92	0	0,66	0	7,89
178	1,59	0,15	1	0,15	0	0	0,67	0	1,76
179	0	0	1	0	0	0	1	0	0
180	1,4	0	0,04	0	0,96	0	0,67	0	1,2
181	0	0	1	0	0	0	1	0	0
182	2,52	0,15	0	0,15	1	0	0,67	0	2,56
183	1,57	0,83	1	0,81	0	0	0,67	0	3,56

βCenH3, δCenH3 and εCenH3									
Site	α	β^-	p^-	β^+	p^+	LRT	p-value	# branches under selection	Total branch length
1	0	0	1	0	0	0	1	0	0
2	0	0	1	0	0	0	1	0	0
3	0,81	0	0,64	25,43	0,36	3,06	0,1	1	23,22
4	1	0,56	1	0,58	0	0	0,67	0	2,18
5	0	0	1	0	0	0	1	0	0
6	0	0	0,8	16,23	0,2	4,74	0,04	1	7,94
7	0	0	0	0,53	1	1,09	0,3	1	1,31
8	0	0	0,69	1655	0,31	3,36	0,09	0	1255,47
9	0	0	0	0,49	1	1,09	0,3	1	1,22
10	7,23	0,27	0	0,27	1	0	0,67	0	6,37
11	0,66	0	0,84	84,68	0,16	3,88	0,07	1	32,91
12	0,31	0,31	1	0,35	0	0	0,67	0	1,01
13	0	0	1	0	0	0	1	0	0
14	0	0	0,61	23,54	0,39	1,31	0,27	0	22,75
15	1,56	0,15	0	0,15	1	0	0,67	0	1,6
16	23,15	0	0,2	0	0,8	0	0,67	0	18,31
17	71,82	0	0,05	0	0,95	0	0,67	0	56,79
18	0	0	0,8	17,28	0,2	5,5	0,03	1	8,42
19	0	0	1	0	0	0	1	0	0
20	1,05	0	0,05	0	0,95	0	0,67	0	0,83
21	1216,65	0	0,05	0	0,95	0	0,67	0	962,04
22	1,11	0,16	0,8	50,8	0,2	2,12	0,17	1	26,45
23	0,55	0	0,01	0	0,99	0	0,67	0	0,43
24	0,87	0	0,05	0	0,95	0	0,67	0	0,69
25	1,31	0,17	0	0,18	1	0	0,67	0	1,47
26	1,85	0,51	1	0,51	0	0	0,67	0	2,73
27	1306,58	6,96	0,22	0,49	0,78	0	0,67	0	1037,85
28	0	0	1	0	0	0	1	0	0
29	0,49	0,41	1	0,4	0	0	0,67	0	1,39
30	0,81	0	0,05	0	0,95	0	0,67	0	0,64
31	0	0	1	0	0	0	1	0	0
32	0	0	0	0,17	1	0,48	0,43	0	0,41
33	0	0	0	0,45	1	0,78	0,36	1	1,1
34	0,68	0,68	0	0,78	1	0	0,65	0	2,47
35	0	0	0	0,12	1	0,47	0,44	1	0,3
36	0	0	1	0	0	0	1	0	0
37	0	0	1	0	0	0	1	0	0
38	0	0	1	0	0	0	1	0	0
39	0	0	1	0	0	0	1	0	0
40	0	0	0	0,14	1	0,15	0,55	0	0,35
41	0	0	1	0	0	0	1	0	0
42	0	0	1	0	0	0	1	0	0
43	0	0	1	0	0	0	1	0	0
44	0,68	0,68	0,49	0	0,51	0	0,67	0	1,35
45	2,33	0	0,05	0	0,95	0	0,67	0	1,85
46	0,29	0,29	0	0,41	1	0,01	0,65	0	1,25
47	2,68	1,8	0	1,81	1	0	0,67	0	6,57
48	3,74	0	0,05	0	0,95	0	0,67	0	2,96
49	0	0	0,55	1,05	0,45	1,19	0,29	0	1,17
50	0,75	0	0,09	0	0,91	0	0,67	0	0,59
51	1,22	0	0,05	0	0,95	0	0,67	0	0,96
52	0	0	1	0	0	0	1	0	0
53	6,43	0	0,13	0	0,87	0	0,67	0	5,09
54	3,89	0,17	0	0,17	1	0	0,67	0	3,49
55	0,81	0	0,05	0	0,95	0	0,67	0	0,64
56	0	0	1	0	0	0	1	0	0
57	0	0	0	0,37	1	0,15	0,55	1	0,9
58	4,25	0,34	0	0,34	1	0	0,67	0	4,18
59	0	0	0	0,22	1	0,55	0,42	1	0,53
60	0	0	0	0,24	1	0,77	0,36	1	0,59
61	2,17	0,19	0	0,19	1	0	0,67	0	2,19
62	0	0	0	0,16	1	0,56	0,41	1	0,39
63	0,91	0,71	1	0,68	0	0	0,67	0	2,46
64	0,9	0,9	1	0,94	0	0	0,67	0	2,94
65	0,97	0	0,05	0	0,95	0	0,67	0	0,77
66	0,75	0	0,09	0	0,91	0	0,67	0	0,59
67	3,91	0,15	1	0,15	0	0	0,67	0	3,47
68	0	0	1	0	0	0	1	0	0
69	0	0	0	0,17	1	0,48	0,44	1	0,42
70	0,59	0,16	0	0,16	1	0	0,67	0	0,87
71	2,79	0	0,09	0	0,91	0	0,67	0	2,2
72	0,87	0	0,05	0	0,95	0	0,67	0	0,69
73	1,8	0	0,05	0	0,95	0	0,67	0	1,42
74	0,57	0	0,19	0	0,81	0	0,67	0	0,45
75	0	0	1	0	0	0	1	0	0
76	0,55	0	0,01	0	0,99	0	0,67	0	0,43
77	0,81	0	0,09	0	0,91	0	0,67	0	0,64
78	1,62	0	0,18	0	0,82	0	0,67	0	1,28
79	0,6	0,21	1	0,2	0	0	0,67	0	0,98
80	0,28	0	0,05	0	0,95	0	0,67	0	0,22
81	0,84	0,84	0,49	0	0,51	0	0,67	0	1,67
82	0	0	1	0	0	0	1	0	0
83	0	0	1	0	0	0	1	0	0
84	2,46	0	0,09	0	0,91	0	0,67	0	1,94
85	3,26	0	0,1	0	0,9	0	0,67	0	2,58
86	6,43	0	0,01	0	0,99	0	0,67	0	5,09
87	0	0	1	0	0	0	1	0	0
88	6,16	0,38	0,37	0,38	0,63	0	0,67	0	5,82
89	0,87	0	0,01	0	0,99	0	0,67	0	0,69
90	2,61	0,4	1	0,25	0	0	0,67	0	3,06
91	6,44	0	0,08	0	0,92	0	0,67	0	5,09
92	5,68	0	0,19	0	0,81	0	0,67	0	4,49
93	0,98	0,21	0	0,21	1	0	0,67	0	1,3
94	0,43	0	0,05	0	0,95	0	0,67	0	0,34
95	1,7	0	0,05	0	0,95	0	0,67	0	1,34
96	0	0	1	0	0	0	1	0	0
97	0,55	0,15	0	0,15	1	0	0,67	0	0,8
98	3,49	0	0,7	4,26	0,3	0,01	0,64	0	5,96
99	0	0	1	0	0	0	1	0	0
100	0	0	0	0,22	1	0,43	0,45	1	0,54
101	0	0	1	0	0	0	1	0	0
102	1,46	0	0,1	0	0,9	0	0,67	0	1,16
103	2,31	1,17	0,46	0	0,54	0	0,67	0	3,15
104	0,38	0	0,04	0	0,96	0	0,67	0	0,3
105	0,57	0	0,19	0	0,81	0	0,67	0	0,45
106	0,35	0	0,02	0	0,98	0	0,67	0	0,28
107	0,55	0	0,11	0	0,89	0	0,67	0	0,44
108	0,8	0	0,04	0	0,96	0	0,67	0	0,63
109	0	0	0	0,16	1	0,44	0,45	0	0,4
110	4,26	2,09	0,59	0,32	0,41	0	0,67	0	6,73
111	0,82	0,22	1	0,2	0	0	0,67	0	1,18
112	0	0	1	0	0	0	1	0	0
113	0,41	0	0,05	0	0,95	0	0,67	0	0,32
114	1,47	0,16	1	0,17	0	0	0,67	0	1,57
115	0	0	0	0,45	1	1,02	0,31	1	1,12
116	1,39	0,2	1	0,33	0	0	0,67	0	1,6
117	2,1	1,12	0,44	0	0,56	0	0,67	0	2,87
118	0	0	1	0	0	0	1	0	0
119	2,7	0	0,09	0	0,91	0	0,67	0	2,13
120	1,48	0	0,1	0	0,9	0	0,67	0	1,17
121	0,38	0	0,02	0	0,98	0	0,67	0	0,3
122	0,42	0,15	1	0,15	0	0	0,67	0	0,7
123	1,24	0	0,19	0	0,81	0	0,67	0	0,98
124	1,14	0,19	0,29	0,19	0,71	0	0,67	0	1,37
125	2,65	0	0,1	0	0,9	0	0,67	0	2,1
126	6,15	0	0,05	0	0,95	0	0,67	0	4,86
127	2,7	0	0,18	0	0,82	0	0,67	0	2,13
128	0,57	0	0,19	0	0,81	0	0,67	0	0,45
129	3,74	0	0,02	0	0,98	0	0,67	0	2,96
130	2,27	0	0,14	0	0,86	0	0,67	0	1,79
131	3,26	0	0,02	0	0,98	0	0,67	0	2,58
132	0	0	1	0	0	0	1	0	0
133	6,43	0	0,01	0	0,99	0	0,67	0	5,09
134	1,43	0	0,1	0	0,9	0	0,67	0	1,13
135	0	0	1	0	0	0	1	0	0
136	0,8	0	0,05	0	0,95	0	0,67	0	0,63
137	1654,55	0,17	0	0,17	1	0	0,67	0	1308,72
138	1,55	0	0,18	0	0,82	0	0,67	0	1,23
139	2,7	0	0,05	0	0,95	0	0,67	0	2,13
140	0	0	1	0	0	0	1	0	0
141	0,9	0,9	0,46	0	0,54	0	0,67	0	1,73
142	0	0	1	0	0	0	1	0	0
143	0,92	0	0,1	0	0,9	0	0,67	0	0,73
144	6,43	0	0,05	0	0,95	0	0,67	0	5,09
145	1,97	0	0,11	0	0,89	0	0,67	0	1,55
146	0,34	0	0,02	0	0,98	0	0,67	0	0,27
147	0	0	1	0	0	0	1	0	0
148	2,8	0	0,14	0	0,86	0	0,67	0	2,21
149	0,91	0,17	0	0,17	1	0	0,67	0	1,13
150	0,54	0	0,19	0	0,81	0	0,67	0	0,42
151	3,74	0	0,07	0	0,93	0	0,67	0	2,96
152	0,81	0	0,02	0	0,98	0	0,67	0	0,64
153	7,39	0	0,18	0	0,82	0	0,67	0	5,84
154	0	0	1	0	0	0	1	0	0
155	0,46	0,17	0	0,17	1	0	0,67	0	0,79
156	0,57	0	0,19	0	0,81	0	0,67	0	0,45
157	0	0	1	0	0	0	1	0	0
158	2,82	1,33	0,43	0	0,57	0	0,67	0	3,65
159	3,65	0,19	0	0,19	1	0	0,67	0	3,36
160	0	0	1	0	0	0	1	0	0
161	0,97	0	0,09	0	0,91	0	0,67	0	0,76
162	1,97	0,34	0,41	0,34	0,59	0	0,67	0	2,39
163	0	0	1	0	0	0	1	0	0
164	2,47	0	0,18	0	0,82	0	0,67	0	1,95
165	1	0	0,09	0	0,91	0	0,67	0	0,79
166	2,63	0	0,09	0	0,91	0	0,67	0	2,08
167	0	0	1	0	0	0	1	0	0
168	2,14	0,23	0	0,23	1	0	0,67	0	2,26
169	0	0	1	0	0	0	1	0	0
170	0	0	1	0	0	0	1	0	0
171	0	0	1	0	0	0	1	0	0
172	0	0	1	0	0	0	1	0	0
173	0	0	0	0,18	1	0,69	0,38	0	0,45
174	0,73	0,73	0	0,91	1	0,01	0,64	0	2,81
175	0	0	0	0,33	1	1,05	0,31	0	0,81
176	0	0	1	0	0	0	1	0	0
177	2,32	2,32	0	3,02	1	0,01	0,64	0	9,28
178	2,77	0,16	1	0,16	0	0	0,67	0	2,6
179	0	0	1	0	0	0	1	0	0
180	0	0	1	0	0	0	1	0	0
181	0	0	1	0	0	0	1	0	0
182	2,67	0,16	1	0,16	0	0	0,67	0	2,5
183	1,67	0,62	1	0,62	0	0	0,67	0	2,85

Supplementary Table 6. ChIP/input ratio of ChIP experiments in *M. incognita* with standard deviation values in A) first 100 most abundant clusters in cluster analysis and B) filtered contigs with >1000 hits in all three αCenH3 ChIP-seq reactions in contig analysis. Most prominent ChIP enriched sequences are highlighted in yellow in accordance with Figure 5.

A

	αCenH3 ChIP-1			αCenH3 ChIP-2A			αCenH3 ChIP-2B					
Cluster	Chip Hits	Input Hits	Ratio Chip/Input	Chip Hits	Input Hits	Ratio Chip/Input	Chip Hits	Input Hits	Ratio Chip/Input	Average ratio	stdev	cluster genome proportion (%)
1	40005	44802	0,89292889	39974	42509	0,9403656	40305	42509	0,9481522	0,927149	0,02989	1,4
2	23326	27719	0,84151665	23194	26656	0,870123	22636	26656	0,8491897	0,85361	0,014807	0,83
3	12539	9841	1,27415913	10044	8972	1,1194828	10446	8972	1,1642889	1,185977	0,079586	0,8
4	50732	44933	1,12905882	43485	39992	1,0873425	44437	39992	1,1111472	1,109183	0,020927	0,56
5	785	829	0,946924	705	1023	0,6891496	725	1023	0,7086999	0,781591	0,143516	0,38
6	30008	33836	0,88686606	31683	37348	0,8483185	31148	37348	0,8339938	0,856393	0,027345	0,37
7	5746	4671	1,23014344	5748	4845	1,1863777	5734	4845	1,1834881	1,200003	0,026142	0,36
8	1049	1160	0,90431034	1073	1495	0,7177258	854	1495	0,5712375	0,731091	0,166938	0,35
9	29474	28470	1,03526519	30616	26636	1,1494218	30716	26636	1,1531762	1,112621	0,067018	0,34
10	7099	6234	1,13875521	7062	6054	1,1665015	6768	6054	1,1179386	1,141065	0,024364	0,33
11	3916	3333	1,17491749	3958	3331	1,1882318	3782	3331	1,1353948	1,166181	0,02748	0,31
12	4059	3159	1,28490028	3877	3155	1,2288431	3946	3155	1,2507132	1,254819	0,028253	0,31
13	4317	2935	1,47086882	4218	2895	1,4569948	4240	2895	1,4645941	1,464153	0,006948	0,31
14	438	321	1,36448598	393	384	1,0234375	393	384	1,0234375	1,13712	0,196904	0,25
15	3085	2678	1,15197909	2920	2634	1,1085801	2991	2634	1,1355353	1,132032	0,021911	0,24
16	14116	8843	1,59629085	12714	8581	1,4816455	12330	8581	1,4368955	1,504944	0,082212	0,24
17	539	582	0,92611684	705	734	0,9604905	608	734	0,8283379	0,904982	0,068565	0,2
18	6548	10909	0,60023834	7787	9565	0,814114	7351	9565	0,7685311	0,727628	0,112652	0,2
19	10948	10295	1,06342885	11585	10594	1,0935435	11188	10594	1,0560695	1,071014	0,019855	0,2
20	17206	17725	0,97071932	16897	18448	0,9159258	17119	18448	0,9279597	0,938202	0,028797	0,2
21	522	572	0,91258741	469	665	0,7052632	595	665	0,8947368	0,837529	0,114893	0,2
22	2757	2225	1,23910112	2579	2194	1,1754786	2539	2194	1,157247	1,190609	0,042973	0,19
23	6517	6602	0,98712511	6825	6361	1,0729445	6564	6361	1,0319132	1,030661	0,042923	0,19
24	7463	7236	1,03137092	7206	7337	0,9821453	7097	7337	0,9672891	0,993602	0,033542	0,19
25	42162	27397	1,53892762	40423	25923	1,5593488	40014	25923	1,5435713	1,547283	0,010705	0,19
26	3400	3875	0,87741935	3675	3970	0,9256927	3505	3970	0,8828715	0,895328	0,026438	0,18
27	3461	4669	0,74127222	3761	4611	0,8156582	4009	4611	0,8694426	0,808791	0,064361	0,17
28	3643	4064	0,89640748	3945	4069	0,9695257	3726	4069	0,9157041	0,927212	0,037893	0,17
29	3264	2552	1,27899687	2972	2524	1,177496	2953	2524	1,1699683	1,20882	0,060891	0,17
30	1962	1845	1,06341463	1964	1844	1,0650759	1933	1844	1,0482646	1,058918	0,009264	0,16
31	13783	17320	0,79578522	14196	17610	0,8061329	13991	17610	0,7944918	0,798803	0,00638	0,15
32	7658	5413	1,41474229	7322	5451	1,3432398	7377	5451	1,3533297	1,370437	0,0387	0,14
33	160	144	1,11111111	160	193	0,8290155	148	193	0,7668394	0,902322	0,18347	0,14
34	1356	1047	1,29512894	1180	944	1,25	1203	944	1,2743644	1,273164	0,022588	0,14
35	1597	1372	1,16399417	1463	1366	1,0710102	1525	1366	1,1163982	1,117134	0,046496	0,14
36	10753	8208	1,31006335	10656	7747	1,3755002	10948	7747	1,4131922	1,366252	0,052183	0,14
37	7999	9645	0,82934163	9860	10962	0,8994709	10057	10962	0,9174421	0,882085	0,046552	0,14
38	1351	1505	0,89767442	1554	1584	0,9810606	1543	1584	0,9741162	0,95095	0,046269	0,14
39	2486	1840	1,35108696	2152	1932	1,1138716	2193	1932	1,1350932	1,200017	0,13126	0,13
40	3685	3824	0,96365063	3757	3983	0,9432589	3754	3983	0,9425056	0,949805	0,011997	0,13
41	7117	7660	0,92911227	7547	7652	0,9862781	7899	7652	1,0322791	0,982557	0,051684	0,13
42	1771	1772	0,99943567	1560	1809	0,8623549	1631	1809	0,9016031	0,921131	0,070596	0,13
43	3812	3871	0,98475846	3788	3490	1,0853868	3799	3490	1,0885387	1,052895	0,059029	0,13
44	3220	3413	0,94345151	3308	3244	1,0197287	3279	3244	1,0107891	0,991323	0,041698	0,13
45	547	523	1,0458891	584	590	0,9898305	576	590	0,9762712	1,003997	0,036908	0,13
46	3930	3735	1,05220884	3348	3289	1,0179386	3526	3289	1,0720584	1,047402	0,027378	0,12
47	4158	3889	1,06916945	3955	3606	1,0967831	4093	3606	1,1350527	1,100335	0,033085	0,12
48	3531	3448	1,02407193	3653	3457	1,0566966	3486	3457	1,0083888	1,029719	0,024644	0,12
49	11276	11926	0,94549723	11075	12803	0,8650316	11687	12803	0,9128329	0,907787	0,040469	0,12
50	3776	3782	0,99841354	3632	3643	0,9969805	3626	3643	0,9953335	0,996909	0,001541	0,11
51	1956	2210	0,88506787	2226	2263	0,98365	2211	2263	0,9770217	0,94858	0,055103	0,11
52	5666	5439	1,04173561	8807	6400	1,3760938	8592	6400	1,3425	1,253443	0,184112	0,11
53	1198	1073	1,11649581	1160	1109	1,0459874	1194	1109	1,0766456	1,07971	0,035354	0,11
54	3933	4029	0,97617275	4204	4007	1,049164	4112	4007	1,0262041	1,01718	0,037323	0,11
55	3759	3604	1,04300777	3936	3437	1,1451848	3787	3437	1,101833	1,096675	0,051283	0,11
56	13975	11978	1,16672232	13361	11948	1,1182625	13718	11948	1,1481419	1,144376	0,024448	0,11
57	2895	2686	1,07781087	2745	2591	1,0594365	2945	2591	1,1366268	1,091291	0,040322	0,1
58	9031	9746	0,92663657	9788	11198	0,8740847	9930	11198	0,8867655	0,895829	0,027423	0,1
59	137	132	1,03787879	118	122	0,9672131	124	122	1,0163934	1,007162	0,036226	0,099
60	7668	8144	0,94155206	8120	8885	0,9138998	7702	8885	0,8668542	0,907435	0,037766	0,098
61	2216	2127	1,04184297	2192	2253	0,972925	2309	2253	1,0248557	1,013208	0,035905	0,096
62	2218	2440	0,90901639	2424	2375	1,0206316	2178	2375	0,9170526	0,9489	0,062251	0,093
63	2241	2330	0,96180258	2223	2356	0,9435484	2297	2356	0,9749576	0,960103	0,015773	0,092
64	3118	3445	0,90507983	3639	3548	1,0256483	3679	3548	1,0369222	0,989217	0,073082	0,092
65	1041	913	1,14019715	981	906	1,0827815	976	906	1,0772627	1,10008	0,034852	0,09
66	124	130	0,95384615	133	149	0,8926174	122	149	0,8187919	0,888419	0,067625	0,089
67	2897	2850	1,01649123	3064	2984	1,0268097	2804	2984	0,9396783	0,994326	0,047607	0,086
68	2351	1777	1,32301632	2211	1618	1,3665019	2189	1618	1,3529048	1,347474	0,022246	0,085
69	5677	6992	0,81192792	7086	7770	0,9119691	6692	7770	0,8612613	0,861719	0,050022	0,085
70	2423	2468	0,98176661	2469	2638	0,9359363	2447	2638	0,9275967	0,948433	0,029167	0,081
71	2866	2429	1,17990943	2745	2422	1,1333609	2947	2422	1,216763	1,176678	0,041795	0,08

72	1040	888	1,17117117	984	833	1,1812725	954	833	1,1452581	1,165901	0,018577	0,078
73	10269	9521	1,07856318	10149	10278	0,9874489	10752	10278	1,0461179	1,037377	0,046182	0,077
74	2175	2190	0,99315068	2193	2150	1,02	2243	2150	1,0432558	1,018802	0,025074	0,077
75	1997	2692	0,74182764	2486	2513	0,9892559	2213	2513	0,8806208	0,870568	0,12402	0,076
76	2161	2082	1,03794428	2066	1942	1,0638517	2062	1942	1,061792	1,054529	0,0144	0,076
77	2207	2247	0,98219849	2492	2269	1,0982812	2289	2269	1,0088145	1,029765	0,060811	0,076
78	1479	1473	1,00407332	1562	1442	1,0832178	1623	1442	1,1255201	1,070937	0,061648	0,075
79	1996	2098	0,95138227	2133	1970	1,0827411	2139	1970	1,0857868	1,03997	0,076734	0,073
80	1975	1986	0,99446123	1933	2079	0,9297739	2013	2079	0,968254	0,964163	0,032537	0,073
81	1552	1564	0,99232737	1503	1648	0,9120146	1625	1648	0,9860437	0,963462	0,044665	0,07
82	1495	1609	0,92914854	1571	1524	1,0308399	1532	1524	1,0052493	0,988413	0,052895	0,068
83	3228	3882	0,83153014	4379	4864	0,9002878	4492	4864	0,9235197	0,885113	0,047836	0,067
84	1386	1420	0,97605634	1381	1412	0,9780453	1377	1412	0,9752125	0,976438	0,001454	0,067
85	1404	1602	0,87640449	1765	1667	1,0587882	1832	1667	1,0989802	1,011391	0,118616	0,066
86	1174	1341	0,87546607	1263	1412	0,8944759	1215	1412	0,8604816	0,876808	0,017037	0,065
87	2298	2050	1,12097561	2443	2225	1,0979775	2433	2225	1,0934831	1,104145	0,014748	0,063
88	1642	1644	0,99878345	1775	1562	1,1363636	1705	1562	1,0915493	1,075565	0,070169	0,063
89	1208	1006	1,20079523	836	894	0,935123	912	894	1,0201342	1,052018	0,135675	0,062
90	9080	9500	0,95578947	9289	10401	0,8930872	11025	10401	1,0599942	0,969624	0,084309	0,06
91	1249	1367	0,91367959	1323	1380	0,9586957	1256	1380	0,9101449	0,927507	0,027068	0,059
92	2684	3130	0,85750799	2811	3516	0,7994881	2984	3516	0,8486917	0,835229	0,031265	0,058
93	790	580	1,36206897	537	436	1,2316514	466	436	1,0688073	1,220843	0,146929	0,057
94	4946	4174	1,18495448	4497	4804	0,9360949	4475	4804	0,9315154	1,017522	0,145019	0,057
95	3220	3113	1,03437199	3190	3570	0,8935574	3096	3570	0,8672269	0,931719	0,08987	0,057
96	1837	1938	0,94788442	1893	1835	1,0316076	1803	1835	0,9825613	0,987351	0,042067	0,054
97	2129	1526	1,39515072	2179	1655	1,3166163	2186	1655	1,3208459	1,344204	0,044172	0,052
98	9508	6911	1,37577775	8585	7331	1,1710544	8307	7331	1,1331333	1,226655	0,130528	0,052
99	1041	1131	0,9204244	1085	1020	1,0637255	1005	1020	0,9852941	0,989815	0,071757	0,052
100	4269	4718	0,90483256	4915	5896	0,833616	4714	5896	0,7995251	0,845991	0,053733	0,052

B

	αCenH3 ChIP-1			αCenH3 ChIP-2A			αCenH3 ChIP-2B					
Contig name	Input-1	Chip-1	Ratio-1	Input-2	Chip-2A	Ratio-2A	Input-2	ChIP-2B	Ratio-2B	Average ratio	stdev	Average hit number
CL2Contig6	2618	1796	0,68602	2193	1871	0,853169	2193	1770	0,807114	0,782101	0,086336	1812
CL2Contig10	6357	4408	0,693409	6288	4702	0,747774	6288	4527	0,719943	0,720375	0,027185	4546
CL3Contig110	3246	4438	1,367221	3045	3606	1,184236	3045	3595	1,180624	1,244027	0,106704	3880
CL4Contig8	3390	4065	1,199115	3008	3458	1,149601	3008	3625	1,20512	1,184612	0,030469	3716
CL4Contig15	1682	1536	0,913199	1221	1170	0,958231	1221	1152	0,943489	0,938306	0,022959	1286
CL4Contig39	1417	1644	1,160198	1217	1273	1,046015	1217	1203	0,988496	1,064903	0,087395	1373
CL4Contg64	7160	8683	1,212709	6849	7587	1,107753	6849	8052	1,175646	1,16537	0,053228	8107
CL4Contig74	6368	5812	0,912688	5774	5949	1,030308	5774	5596	0,969172	0,970723	0,058825	5786
CL4Contig77	2116	1571	0,742439	1553	1082	0,696716	1553	1082	0,696716	0,711957	0,026398	1245
CL4Contig79	1015	1729	1,703448	1051	1343	1,277831	1051	1486	1,413892	1,465057	0,217373	1519
CL6Contig27_1_sc_0.664164_l_87	32638	28928	0,886329	36016	30499	0,846818	36016	29966	0,832019	0,855055	0,028076	29798
CL7Contig95	1296	1631	1,258488	1385	1611	1,163177	1385	1571	1,134296	1,18532	0,064989	1604
CL9Contig27_1_sc_0.544156_l_30	27530	28439	1,033019	25661	29655	1,155645	25661	29679	1,15658	1,115081	0,07107	29258
CL10Contig10	1025	1240	1,209756	989	1177	1,190091	989	1149	1,16178	1,187209	0,024118	1189
CL10Contig12	1030	1297	1,259223	1100	1360	1,236364	1100	1107	1,006364	1,167317	0,139857	1255
CL12Contig8	2589	3432	1,325608	2517	3215	1,277314	2517	3299	1,310687	1,304537	0,024728	3315
CL13Contig10	2672	3931	1,471183	2625	3867	1,473143	2625	3854	1,46819	1,470839	0,002494	3884
CL16Contig11	747	1435	1,921017	656	1361	2,074695	656	1070	1,631098	1,875603	0,225259	1289
CL16Cotig20	3361	5532	1,645939	3189	4951	1,552524	3189	4780	1,498902	1,565788	0,07441	5088
CL18Contig27_2_sc_0.642742_l_168	4906	2704	0,551162	4054	3254	0,802664	4054	3278	0,808584	0,720803	0,146944	3079
CL18Contig27_1_sc_0.614693_l_167	2962	1813	0,612086	2460	2078	0,844715	2460	2218	0,901626	0,786143	0,153399	2036
CL19Contig11_1_sc_0.696255_l_183	4006	4236	1,057414	4064	4634	1,140256	4064	4435	1,091289	1,09632	0,041649	4435
CL19Contig27_2_sc_0.495346_l_184	2703	2752	1,018128	2834	3021	1,065984	2834	2861	1,009527	1,031213	0,030418	2878
CL19Contig27_1_sc_0.384703_l_183	2994	3271	1,092518	3174	3248	1,023314	3174	3278	1,032766	1,049533	0,037525	3266
CL20Contig23_1_sc_0.588571_l_103	17342	16820	0,9699	18041	16524	0,915914	18041	16751	0,928496	0,938103	0,028246	16698
CL22Contig14	1460	1760	1,205479	1422	1630	1,146273	1422	1607	1,130098	1,160617	0,039685	1666
CL25Contig5	20901	32523	1,55605	19868	30854	1,552949	19868	30597	1,540014	1,549671	0,008506	31325
CL25Contig8	3330	4979	1,495195	3249	5238	1,612188	3249	4987	1,534934	1,547439	0,059491	5068
CL27Contig27_1_sc_0.774208_l_184	4550	3374	0,741538	4463	3669	0,822093	4463	3879	0,869146	0,810926	0,064533	3641
CL31Contig27_1_sc_0.738448_l_83	16854	13360	0,79269	17074	13757	0,805728	17074	13559	0,794131	0,797517	0,007148	13559
CL32Contig3	1159	1597	1,377912	1150	1701	1,47913	1150	1714	1,490435	1,449159	0,06196	1671
CL32Contig8	734	1069	1,456403	740	1125	1,52027	740	1005	1,358108	1,444927	0,081688	1066
CL36Contig8	7524	9628	1,279638	6894	9513	1,379896	6894	9975	1,44691	1,368815	0,084185	9705
CL37Contig27_1_sc_0.782873_l_89	9163	7657	0,835643	10429	9402	0,901525	10429	9555	0,916195	0,884454	0,042903	8871
CL41Contig27_1_sc_0.708498_l_134	3124	3063	0,980474	3510	3324	0,947009	3510	3574	1,018234	0,981905	0,035634	3320
CL41Contig19_1_sc_0.686608_l_134	4346	3906	0,898757	3968	4057	1,022429	3968	4113	1,036542	0,98591	0,075805	4025
CL49Contig19_2_sc_0.474972_l_74	3425	3628	1,05927	3229	3227	0,999381	3229	3345	1,035924	1,031525	0,030186	3400
CL49Contig27_1_sc_0.456271_l_74	3970	3567	0,898489	4220	3634	0,861137	4220	3645	0,863744	0,874457	0,020853	3615
CL49Contig23_1_sc_0.24808_l_72	3590	3106	0,865181	4233	3353	0,79211	4233	3791	0,895582	0,850958	0,053183	3417
CL52Contig27_1_sc_0.787934_l_76	5283	5530	1,046754	6220	8578	1,3791	6220	8361	1,344212	1,256689	0,182644	7490
CL55Contig71	1736	1858	1,070276	1571	1879	1,196053	1571	1772	1,127944	1,131425	0,062961	1836
CL56Contig27_1_sc_0.682824_l_56	11533	13416	1,163271	11504	12815	1,11396	11504	13175	1,145254	1,140828	0,024951	13135
CL58Contig27_1_sc_0.766317_l_80	9605	8862	0,922644	11015	9633	0,874535	11015	9749	0,885066	0,894082	0,02529	9415
CL60Contig27_1_sc_0.690143_l_76	7965	7484	0,939611	8653	7920	0,915289	8653	7504	0,867214	0,907371	0,036842	7636
CL69Contig27_1_sc_0.554202_l_83	6819	5508	0,807743	7565	6845	0,904825	7565	6496	0,858691	0,857086	0,048561	6283
CL73Contig27_1_sc_0.868693_l_63	9288	9998	1,076443	10012	9877	0,986516	10012	10473	1,046045	1,036335	0,045743	10116
CL83Contig27_1_sc_0.647019_l_97	3826	3156	0,824882	4793	4298	0,896724	4793	4412	0,920509	0,880705	0,049785	3955
CL87Contig2	1604	1731	1,079177	1779	1969	1,106802	1779	2035	1,143901	1,10996	0,032477	1912
CL90Contig27_9_sc_0.894693_l_153	4655	4567	0,981096	5351	4568	0,853672	5351	5782	1,080546	0,971771	0,113724	4972
CL90Contig27_8_sc_0.842709_l_123	1203	1152	0,957606	1325	1250	0,943396	1325	1390	1,049057	0,983353	0,057343	1264
CL92Contig11_1_sc_0.72508_l_94	2998	2588	0,863242	3421	2698	0,788658	3421	2883	0,842736	0,831545	0,038531	2723
CL94Contig23_1_sc_0.649269_l_95	1437	1465	1,019485	1579	1404	0,88917	1579	1450	0,918303	0,942319	0,068396	1440
CL94Contig27_1_sc_0.63172_l_95	1516	2000	1,319261	1904	1865	0,979517	1904	1764	0,926471	1,075083	0,213121	1876
CL94Contig15_1_sc_0.495727_l_95	1105	1302	1,178281	1176	1069	0,909014	1176	1146	0,97449	1,020595	0,140429	1172
CL97Contig27_1_sc_0.829814_l_88	1467	2069	1,410361	1604	2117	1,319825	1604	2124	1,32419	1,351459	0,051058	2103
CL98Contig23_1_sc_0.641447_l_55	6637	9138	1,376827	7064	8253	1,168318	7064	7986	1,130521	1,225222	0,132647	8459
CL100Contig27_1_sc_0.727401_l_69	4509	4102	0,909736	5666	4718	0,832686	5666	4534	0,800212	0,847545	0,056254	4451

Supplementary Figure 7. Quantification of αCenH3 colocalization with centromeric probes (CL25m1, CL16m1 and CL16m3/32m1) or 19bp box as shown in Figure 7B-D and Supplementary Figure 9 for three species; *M. incognita*, *M. arenaria* and *M. javanica*. The quantification was done on ten interphase iamges using CellProfiler.

	M. incognita				***M. arenaria***		***M. javanica***	
	FISH centromeric probes		19bp box		19bp box		19bp box	
	R-G Manders coefficient (costes)	G-R Manders coefficient (costes)	R-G Manders coefficient (costes)	G-R Manders coefficient (costes)	R-G Manders coefficient (costes)	G-R Manders coefficient (costes)	R-G Manders coefficient (costes)	G-R Manders coefficient (costes)
1	0,823	0,883	0,943	0,978	0,924	0,972	0,87	0,887
2	0,8	0,724	0,97	0,954	0,978	0,955	0,964	0,917
3	0,683	0,869	0,969	0,971	0,98	0,979	0,942	0,99
4	0,734	0,886	0,958	0,976	0,972	0,967	0,905	0,896
5	0,71	0,832	0,971	0,982	0,965	0,914	0,904	0,928
6	0,82	0,841	0,959	0,996	0,933	0,982	0,949	0,944
7	0,822	0,859	0,991	0,961	0,971	0,969	0,926	0,876
8	0,791	0,812	0,934	0,931	0,966	0,977	0,96	0,908
9	0,704	0,829	0,887	0,973	0,984	0,939	0,966	0,933
10	0,87	0,82	0,987	0,934	0,981	0,995	0,935	0,941
average	0,7757	0,8355	0,9569	0,9656	0,9654	0,9649	0,9321	0,922

Acknowledgments

We thank M. Pavlek for performing the preliminary Western blot experiments and L. Horvat for assistance in high-resolution microscopy. This work has been supported by grant IP-2014-09-3183 from Croatian Science Foundation (https://www.hrzz.hr/default.aspx?id=47) to M.P. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

References

Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M.-C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V, Hamamouch, N., … Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, *26*(8), 909–915. https://doi.org/10.1038/nbt.1482

Agrios, G. (2004). Plant pathology: Fifth edition. In *Plant Pathology: Fifth Edition* (Vol. 9780080473). https://doi.org/10.1016/C2009-0-02037-6

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). *Molecular Biology of the Cell*. New York: Garland Science.

Aldrup-MacDonald, M. E., Kuo, M. E., Sullivan, L. L., Chew, K., & Sullivan, B. A. (2016). Genomic variation within alpha satellite DNA influences centromere location on human chromosomes with metastable epialleles. *Genome Research*, *26*(10), 1301–1311. https://doi.org/10.1101/gr.206706.116

Barra, V., & Fachinetti, D. (2018). The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nature Communications*, *9*(1), 4340. https://doi.org/10.1038/s41467-018-06545-y

Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (Pat. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, *23*(3), 415–420. https://doi.org/10.1007/s10577-015-9499-z

Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J. M., Da Rocha, M., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J. F., Kozlowski, D. K., Cazareth, J., Couloux, A., Da Silva, C., Guy, J., Kim-Jo, Y. J., Rancurel, C., Schiex, T., Abad, P., … Danchin, E. G. J. (2017). Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genetics*, *13*(6), e1006777. https://doi.org/10.1371/journal.pgen.1006777

Bloom, K. S. (2014). Centromeric heterochromatin: The primordial segregation machine. *Annual Review of Genetics*, *48*(September), 457–484. https://doi.org/10.1146/annurev-genet-120213-092033

Brown, S. J., & Coleman, M. (2019). Isolation of high molecular weight DNA from insects. *Methods in Molecular Biology*, *1858*, 27–32. https://doi.org/10.1007/978-1-4939-8775-7_3

Caputi, F. F., Candeletti, S., & Romualdi, P. (2017). Epigenetic Approaches in Neuroblastoma Disease Pathogenesis. In *Neuroblastoma - Current State and Recent Updates*. https://doi.org/10.5772/intechopen.69566

Castagnone-Sereno, P. (2006). Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity*, *96*(4), 282–289. https://doi.org/10.1038/sj.hdy.6800794

Castagnone-Sereno, P, & Danchin, E. G. J. (2014). Parasitic success without sex – the nematode experience. *Journal of Evolutionary Biology*, *27*(7), 1323–1333. https://doi.org/10.1111/jeb.12337

Castagnone-Sereno, Philippe, Danchin, E. G. J., Perfus-Barbeoch, L., & Abad, P. (2013). Diversity and Evolution of Root-Knot Nematodes, Genus *Meloidogyne*: New Insights from the Genomic

Era. *Annual Review of Phytopathology*, *51*(1), 203–220. https://doi.org/10.1146/annurev-phyto-082712-102300

Castagnone-Sereno, Philippe, Semblat, J. P., Leroy, F., & Abad, P. (1998). A New AluI Satellite DNA in the Root-Knot Nematode *Meloidogyne fallax*: Relationships with Satellites from the Sympatric Species *M. hapla* and *M. chitwoodi*. *Molecular Biology and Evolution*, *15*(9), 1115–1122. https://doi.org/10.1093/oxfordjournals.molbev.a026019

Chang, C. H., Chavan, A., Palladino, J., Wei, X., Martins, N. M. C., Santinello, B., Chen, C. C., Erceg, J., Beliveau, B. J., Wu, C. T., Larracuente, A. M., & Mellone, B. G. (2019). Islands of retroelements are major components of *Drosophila* centromeres. In *PLoS Biology* (Vol. 17, Issue 5). https://doi.org/10.1371/journal.pbio.3000241

Cohen-Sharir, Y., McFarland, J. M., Abdusamad, M., Marquis, C., Bernhard, S. V., Kazachkova, M., Tang, H., Ippolito, M. R., Laue, K., Zerbib, J., Malaby, H. L. H., Jones, A., Stautmeister, L. M., Bockaj, I., Wardenaar, R., Lyons, N., Nagaraja, A., Bass, A. J., Spierings, D. C. J., … Ben-David, U. (2021). Aneuploidy renders cancer cells vulnerable to mitotic checkpoint inhibition. *Nature*, *590*(7846), 486–491. https://doi.org/10.1038/s41586-020-03114-6

Cottarel, G., Shero, J. H., Hieter, P., & Hegemann, J. H. (1989). A 125 bp CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae*. *Trends in Genetics*, *5*(C), 322–324. https://doi.org/10.1016/0168-9525(89)90119-4

Drinnenberg, I. A., deYoung, D., Henikoff, S., & Malik, H. S. ing. (2014). Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *ELife*, *3*, 1–19. https://doi.org/10.7554/eLife.03676

Felekkis, K., & Voskarides, K. (2015). Genomic elements in health, disease and evolution: Junk DNA. *Genomic Elements in Health, Disease and Evolution: Junk DNA*, 1–311. https://doi.org/10.1007/978-1-4939-3070-8

Feliciello, I., Chinali, G., & Ugarković, D. (2011). Structure and population dynamics of the major satellite DNA in the red flour beetle *Tribolium castaneum*. *Genetica*, *139*(8), 999–1008. https://doi.org/10.1007/s10709-011-9601-1

Fukagawa, T., & Earnshaw, W. C. (2014). Neocentromeres. *Current Biology*, *24*(19), R946–R947. https://doi.org/10.1016/j.cub.2014.08.032

Gamba, R., & Fachinetti, D. (2020). From evolution to function: Two sides of the same CENP-B coin? *Experimental Cell Research*, *390*(2). https://doi.org/10.1016/j.yexcr.2020.111959

Guenatri, M., Bailly, D., Maison, C., & Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *Journal of Cell Biology*, *166*(4), 493–505. https://doi.org/10.1083/jcb.200403109

Hara, M., & Fukagawa, T. (2017). Critical Foundation of the Kinetochore: The Constitutive Centromere-Associated Network (CCAN). *Progress in Molecular and Subcellular Biology*, *56*, 29–57. https://doi.org/10.1007/978-3-319-58592-5_2

Hartley, G., & O'Neill, R. J. (2019). Centromere repeats: Hidden gems of the genome. *Genes*, *10*(3), 223. https://doi.org/10.3390/genes10030223

He, L., Liu, J., Torres, G. A., Zhang, H., Jiang, J., & Xie, C. (2013). Interstitial telomeric repeats are

enriched in the centromeres of chromosomes in *Solanum* species. *Chromosome Research*, *21*(1), 5–13. https://doi.org/10.1007/s10577-012-9332-x

Henikoff, S., Ahmad, K., & Malik, H. S. (2001). The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science*, *293*(5532), 1098–1102. https://doi.org/10.1126/science.1062939

Herndon, N., Shelton, J., Gerischer, L., Ioannidis, P., Ninova, M., Dönitz, J., Waterhouse, R. M., Liang, C., Damm, C., Siemanowski, J., Kitzmann, P., Ulrich, J., Dippel, S., Oberhofer, G., Hu, Y., Schwirz, J., Schacht, M., Lehmann, S., Montino, A., … Bucher, G. (2020). Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. *BMC Genomics*, *21*(1), 1–13. https://doi.org/10.1186/s12864-019-6394-6

Huang, Y. C., Lee, C. C., Kao, C. Y., Chang, N. C., Lin, C. C., Shoemaker, D., & Wang, J. (2016). Evolution of long centromeres in fire ants. *BMC Evolutionary Biology*, *16*(1), 1–14. https://doi.org/10.1186/s12862-016-0760-7

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., St. John, O., Wild, R., Hammond, P. M., Ahrens, D., Balke, M., Caterino, M. S., Gómez-Zurita, J., Ribera, I., Barraclough, T. G., Bocakova, M., Bocak, L., & Vogler, A. P. (2007). A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science*, *318*(5858), 1913–1916. https://doi.org/10.1126/science.1146954

Ishii, T., Juranić, M., Maheshwari, S., Bustamante, F. de O., Vogt, M., Salinas-Gamboa, R., Dreissig, S., Gursanscky, N., How, T., Demidov, D., Fuchs, J., Schubert, V., Spriggs, A., Vielle-Calzada, J. P., Comai, L., Koltunow, A. M. G., & Houben, A. (2020). Unequal contribution of two paralogous CENH3 variants in cowpea centromere function. *Communications Biology*, *3*(1). https://doi.org/10.1038/s42003-020-01507-x

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., … Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, *36*(4), 338–345. https://doi.org/10.1038/nbt.4060

Jain, M., Olsen, H. E., Turner, D. J., Stoddart, D., Bulazel, K. V., Paten, B., Haussler, D., Willard, H. F., Akeson, M., & Miga, K. H. (2018). Linear assembly of a human centromere on the Y chromosome. *Nature Biotechnology*, *36*(4), 321–323. https://doi.org/10.1038/nbt.4109

Juan, C., Vazquez, P., Rubio, J. M., Petitpierre, E., & Hewitt, G. M. (1993). Presence of highly repetitive DNA sequences in *Tribolium* flour-beetles. *Heredity*, *70*(1), 1–8. https://doi.org/10.1038/hdy.1993.1

Kang, Y., Wang, J., Neff, A., Kratzer, S., Kimura, H., & Davis, R. E. (2016). Differential Chromosomal Localization of Centromeric Histone CENP-A Contributes to Nematode Programmed DNA Elimination. *Cell Reports*, *16*(9), 2308–2316. https://doi.org/10.1016/j.celrep.2016.07.079

Kasinathan, S., & Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. *Molecular Biology and Evolution*, *35*(4), 949–962. https://doi.org/10.1093/molbev/msy010

Komissarov, A. S., Gavrilova, E. V., Demin, S. J., Ishov, A. M., & Podgornaya, O. I. (2011). Tandemly repeated DNA families in the mouse genome. *BMC Genomics*, *12*. https://doi.org/10.1186/1471-

2164-12-531

Kumekawa, N., Hosouchi, T., Tsuruoka, H., & Kotani, H. (2001). The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 4. *DNA Research*, *8*(6), 285–290. https://doi.org/10.1093/dnares/8.6.285

Kursel, L. E., & Malik, H. S. (2017). Recurrent gene duplication leads to diverse repertoires of centromeric histones in *Drosophila* species. *Molecular Biology and Evolution*, *34*(6), 1445–1462. https://doi.org/10.1093/molbev/msx091

Kursel, L. E., Welsh, F. C., & Malik, H. S. (2020). Ancient Coretention of Paralogs of *Cid* Centromeric Histones and *Cal1* Chaperones in Mosquito Species. *Molecular Biology and Evolution*, *37*(7), 1949–1963. https://doi.org/10.1093/molbev/msaa056

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Logsdon, G. A., Vollger, M. R., Hsieh, P. H., Mao, Y., Liskovykh, M. A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P. C., Rhie, A., de Lima, L. G., Dvorkina, T., Porubsky, D., Harvey, W. T., Mikheenko, A., Bzikadze, A. V., Kremitzki, M., Graves-Lindsay, T. A., Jain, C., … Eichler, E. E. (2021). The structure, function and evolution of a complete human chromosome 8. *Nature*, *593*(7857), 101–107. https://doi.org/10.1038/s41586-021-03420-7

Lopes, M., Louzada, S., Gama-Carvalho, M., & Chaves, R. (2021). Genomic tackling of human satellite dna: Breaking barriers through time. *International Journal of Molecular Sciences*, *22*(9). https://doi.org/10.3390/ijms22094707

Lunt, D. H. (2008). Genetic tests of ancient asexuality in Root Knot Nematodes reveal recent hybrid origins. *BMC Evolutionary Biology*, *16*, 1–16. https://doi.org/10.1186/1471-2148-8-194

Macas, J., Navrátilová, A., & Mészáros, T. (2003). Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on *Vicia sativa* chromosomes. *Chromosoma*, *112*(3), 152–158. https://doi.org/10.1007/s00412-003-0255-3

Malik, H. S., & Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nature Structural Biology*, *10*(11), 882–891. https://doi.org/10.1038/nsb996

Malik, H. S., & Henikoff, S. (2009). Major Evolutionary Transitions in Centromere Complexity. *Cell*, *138*(6), 1067–1082. https://doi.org/10.1016/j.cell.2009.08.036

Mandrioli, M., & Manicardi, G. C. (2020). Holocentric chromosomes. *PLoS Genetics*, *16*(7), 1–11. https://doi.org/10.1371/journal.pgen.1008918

Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A., & Landsman, D. (2005). Histone structure and nucleosome stability. *Expert Review of Proteomics*, *2*(5), 719–729. https://doi.org/10.1586/14789450.2.5.719

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., Pellino, M., Fuchs, J., Ma, W., Kuhlmann, M., Brandt, R., Vanzela, A. L. L., Beseda, T., Šimková, H., Pedrosa-Harand, A., & Houben, A. (2015). Holocentromeres in *Rhynchospora* are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. *Proceedings of the National*

Academy of Sciences of the United States of America, *112*(44), 13633–13638. https://doi.org/10.1073/pnas.1512255112

Mckinley, K. L., & Cheeseman, I. M. (2016). The molecular basis for centromere identity and function. *Nature Reviews Molecular Cell Biology*, *17*(1), 16–29. https://doi.org/10.1038/nrm.2015.5

McNulty, S. M., Sullivan, L. L., & Sullivan, B. A. (2017). Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. *Developmental Cell*, *42*(3), 226-240.e6. https://doi.org/10.1016/j.devcel.2017.07.001

Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J. G., Sebra, R., Peluso, P., Eid, J., Rank, D., Garcia, J. F., Derisi, J. L., Smith, T., Tobias, C., Ross-Ibarra, J., Korf, I., & Chan, S. W. L. (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology*, *14*(1), 1–20.

Melters, D. P., Paliulis, L. V., Korf, I. F., & Chan, S. W. L. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Research*, *20*(5), 579–593. https://doi.org/10.1007/s10577-012-9292-1

Meštrović, N., Castagnone-Sereno, P., & Plohl, M. (2006a). High conservation of the differentially amplified MPA2 satellite DNA family in parthenogenetic root-knot nematodes. *Gene*, *376*(2), 260–267. https://doi.org/10.1016/j.gene.2006.04.008

Meštrović, N., Castagnone-Sereno, P., & Plohl, M. (2006b). Interplay of Selective Pressure and Stochastic Events Directs Evolution of the MEL172 Satellite DNA Library in Root-Knot Nematodes. *Molecular Biology and Evolution*, *23*(12), 2316–2325. https://doi.org/10.1093/molbev/msl119

Meštrović, N., Pavlek, M., Car, A., Castagnone-Sereno, P., Abad, P., & Plohl, M. (2013). Conserved DNA Motifs, Including the CENP-B Box-like, Are Possible Promoters of Satellite DNA Array Rearrangements in Nematodes. *PLoS ONE*, *8*(6), e67328. https://doi.org/10.1371/journal.pone.0067328

Meštrović, N., Plohl, M., & Castagnone-Sereno, P. (2009). Relevance of satellite DNA genomic distribution in phylogenetic analysis: A case study with root-knot nematodes of the genus *Meloidogyne*. *Molecular Phylogenetics and Evolution*, *50*(1), 204–208. https://doi.org/10.1016/j.ympev.2008.10.013

Meštrović, N., Plohl, M., Mravinac, B., & Ugarković, Đ. (1998). Evolution of Satellite DNAs from the genus *Palorus*-Experimental Evidence for the “Library” Hypothesis. *Molecular Biology and Evolution*, *15*(8), 1062–1068. https://doi.org/10.1093/oxfordjournals.molbev.a026005

Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G. A., Schneider, V. A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., … Phillippy, A. M. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, *585*(7823), 79–84. https://doi.org/10.1038/s41586-020-2547-7

Monen, J., Hattersley, N., Muroyama, A., Stevens, D., Oegema, K., & Desai, A. (2015). Separase Cleaves the N-Tail of the CENP-A Related Protein CPAR-1 at the Meiosis I Metaphase-Anaphase Transition in *C . elegans*. *PloS One*, *10*(4), e0125382.

https://doi.org/10.1371/journal.pone.0125382

Monen, J., Maddox, P. S., Hyndman, F., Oegema, K., & Desai, A. (2005). Differential role of CENP-A in the segregation of holocentric *C. elegans* chromosomes during meiosis and mitosis. *Nature Cell Biology*, *7*(12), 1248–1255. https://doi.org/10.1038/ncb1331

Mravinac, B., & Plohl, M. (2010). Parallelism in Evolution of Highly Repetitive DNAs in Sibling Species. *Molecular Biology and Evolution*, *27*(8), 1857–1867. https://doi.org/10.1093/molbev/msq068

Mravinac, B., Plohl, M., & Ugarković, D. (2004). Conserved patterns in the evolution of *Tribolium* satellite DNAs. *Gene*, *332*(1–2), 169–177. https://doi.org/10.1016/j.gene.2004.02.055

Muller, H., Gil, J., & Drinnenberg, I. A. (2019). The Impact of Centromeres on Spatial Genome Architecture. *Trends in Genetics*, *35*(8), 565–578. https://doi.org/10.1016/j.tig.2019.05.003

Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., Novák, P., Wanner, G., & Macas, J. (2012). Stretching the Rules: Monocentric Chromosomes with Multiple Centromere Domains. *PLoS Genetics*, *8*(6), e1002777. https://doi.org/10.1371/journal.pgen.1002777

Neumann, P., Pavlíková, Z., Koblížková, A., Fuková, I., Jedličková, V., Novák, P., & Macas, J. (2015). Centromeres Off the Hook: Massive Changes in Centromere Size and Structure Following Duplication of CenH3 Gene in *Fabeae* Species. *Molecular Biology and Evolution*, *32*(7), 1862–1879. https://doi.org/10.1093/molbev/msv070

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., … Phillippy, A. M. (2021). The complete sequence of a human genome. *BioRxiv*. https://doi.org/10.1101/2021.05.26.445798

Oliveira, L., Neumann, P., Jang, T. S., Klemme, S., Schubert, V., Koblížková, A., Houben, A., & Macas, J. (2020). Mitotic Spindle Attachment to the Holocentric Chromosomes of *Cuscuta europaea* Does Not Correlate With the Distribution of CENH3 Chromatin. *Frontiers in Plant Science*, *10*(January), 1–11. https://doi.org/10.3389/fpls.2019.01799

Panchenko, T., Sorensen, T. C., Woodcock, C. L., Kan, Z. Y., Wood, S., Resch, M. G., Luger, K., Englander, S. W., Hansen, J. C., & Black, B. E. (2011). Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(40), 16588–16593. https://doi.org/10.1073/pnas.1113621108

Pavlek, M., Gelfand, Y., Plohl, M., & Meštrović, N. (2015). Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. *DNA Research*, *22*(6), 387–401. https://doi.org/10.1093/dnares/dsv021

Peterson, C. L., & Laniel, M. A. (2004). Histones and histone modifications. *Current Biology : CB*, *14*(14), 546–551. https://doi.org/10.1016/j.cub.2004.07.007

Piras, F. M., Nergadze, S. G., Magnani, E., Bertoni, L., Attolini, C., Khoriauli, L., Raimondi, E., & Giulotto, E. (2010). Uncoupling of Satellite DNA and Centromeric Function in the Genus *Equus*.

PLoS Genetics, *6*(2), e1000845. https://doi.org/10.1371/journal.pgen.1000845

Plohl, M., Meštrović, N., & Mravinac, B. (2014). Centromere identity from the DNA point of view. *Chromosoma*, *123*(4), 313–325. https://doi.org/10.1007/s00412-014-0462-0

Pointer, M. D., Gage, M. J. G., & Spurgin, L. G. (2021). *Tribolium* beetles as a model system in evolution and ecology. *Heredity*, *126*(6), 869–883. https://doi.org/10.1038/s41437-021-00420-1

Presting, G. G. (2018). Centromeric retrotransposons and centromere function. *Current Opinion in Genetics and Development*, *49*, 79–84. https://doi.org/10.1016/j.gde.2018.03.004

Prosée, R. F., Wenda, J. M., & Steiner, F. A. (2020). Adaptations for centromere function in meiosis. *Essays in Biochemistry*, *64*(2), 193–203. https://doi.org/10.1042/ebc20190076

Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S., Denell, R., Beeman, R. W., Gibbs, R., Bucher, G., Friedrich, M., Grimmelikhuijzen, C. J. P., Klingler, M., Lorenzen, M., Roth, S., Schröder, R., Tautz, D., Zdobnov, E. M., Muzny, D., Attaway, T., Bell, S., … Grossmann, D. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, *452*(7190), 949–955. https://doi.org/10.1038/nature06784

Rošić, S., & Erhardt, S. (2016). No longer a nuisance: Long non-coding RNAs join CENP-A in epigenetic centromere regulation. *Cellular and Molecular Life Sciences*, *73*(7), 1387–1398. https://doi.org/10.1007/s00018-015-2124-7

Sansregret, L., & Swanton, C. (2017). The role of aneuploidy in cancer evolution. *Cold Spring Harbor Perspectives in Medicine*, *7*(1), 1–17. https://doi.org/10.1101/cshperspect.a028373

Schön, I., Martens, K., & Dijk, P. (2009). Lost Sex. In I. Schön, K. Martens, & P. Dijk (Eds.), *Lost Sex*. Springer Netherlands. https://doi.org/10.1007/978-90-481-2770-2

Schwenk, K., Brede, N., & Streit, B. (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1505), 2805–2811. https://doi.org/10.1098/rstb.2008.0055

Scott, K. C., & Sullivan, B. A. (2014). Neocentromeres: A place for everything and everything in its place. *Trends in Genetics*, *30*(2), 66–74. https://doi.org/10.1016/j.tig.2013.11.003

Shang, W. H., Hori, T., Toyoda, A., Kato, J., Popendorf, K., Sakakibara, Y., Fujiyama, A., & Fukagawa, T. (2010). Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Research*, *20*(9), 1219–1228. https://doi.org/10.1101/gr.106245.110

Smurova, K., & De Wulf, P. (2018). Centromere and Pericentromere Transcription: Roles and Regulation … in Sickness and in Health. *Frontiers in Genetics*, *9*(December), 1–26. https://doi.org/10.3389/fgene.2018.00674

Steiner, F. A., & Henikoff, S. (2014). Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *ELife*, *3*(e02025), 1–22. https://doi.org/10.7554/eLife.02025

Szitenberg, A., Salazar-Jaramillo, L., Blok, V. C., Laetsch, D. R., Joseph, S., Williamson, V. M., Blaxter, M. L., & Lunt, D. H. (2017). Comparative Genomics of Apomictic Root-Knot Nematodes: Hybridization, Ploidy, and Dynamic Genome Change. *Genome Biology and Evolution*, *9*(10), 2844–2861. https://doi.org/10.1093/gbe/evx201

Talbert, P. B., & Henikoff, S. (2020). What makes a centromere? *Experimental Cell Research*, *389*(2),

111895. https://doi.org/10.1016/j.yexcr.2020.111895

Ting, D. T., Lipson, D., Paul, S., Brannigan, B. W., Akhavanfard, S., Coffman, E. J., Contino, G., Deshpande, V., Iafrate, A. J., Letovsky, S., Rivera, M. N., Bardeesy, N., Maheswaran, S., & Haber, D. A. (2011). Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. *Science*, *331*(6017), 593–596. https://doi.org/10.1126/science.1200801

Tollervey, J. R., & Lunyak, V. V. (2012). Epigenetics: judge, jury and executioner of stem cell fate. *Epigenetics*, *7*(8), 823–840. https://doi.org/10.4161/epi.21141

Triantaphyllou, A. C. (1981). Oogenesis and the Chromosomes of the Parthenogenetic Root-knot Nematode *Meloidogyne incognita*. *Journal of Nematology*, *13*(2), 95–104.

Tyson, J. (2020). Rocky Mountain adventures in Genomic DNA sample preparation , ligation protocol optimisation / simplification and Ultra long read generation. *Protocols.Io*, 1–5. https://doi.org/dx.doi.org/10.17504/protocols.io.7euhjew

Ugarković, D., Podnar, M., & Plohl, M. (1996). Satellite DNA of the Red Flour Beetle *Tribolium castaneum* - Comparative Study of Satellites from the Genus *Tribolium*. *Molecular Biology and Evolution*, *13*(8), 1059–1066. https://doi.org/10.1093/oxfordjournals.molbev.a025668

Vlahović, I., Gluncić, M., Rosandić, M., Ugarković, D. I., & Paar, V. (2017). Regular higher order repeat structures in beetle *Tribolium castaneum* genome. *Genome Biology and Evolution*, *9*(10), 2668–2680. https://doi.org/10.1093/gbe/evw174

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, *7*. https://doi.org/10.1038/s41587-021-01108-x

Watanabe, Y. (2005). Sister chromatid cohesion along arms and at centromeres. *Trends in Genetics*, *21*(7), 405–412. https://doi.org/10.1016/j.tig.2005.05.009

Wells, J. N., & Feschotte, C. (2020). A Field Guide to Eukaryotic Transposable Elements. *Annual Review of Genetics*, *54*, 539–561. https://doi.org/10.1146/annurev-genet-040620-022145

Wesemael, W. M. L. (2007). *Biology and management of the root-knot nematode* Meloidogyne chitwoodi *in field vegetable crops*. Ghent University.

Wolfgruber, T. K., Sharma, A., Schneider, K. L., Albert, P. S., Koo, D. H., Shi, J., Gao, Z., Han, F., Lee, H., Xu, R., Allison, J., Birchler, J. A., Jiang, J., Dawe, R. K., & Presting, G. G. (2009). Maize centromere structure and evolution: Sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. *PLoS Genetics*, *5*(11), 13–16. https://doi.org/10.1371/journal.pgen.1000743

Wong, C. Y. Y., Ling, Y. H., Mak, J. K. H., Zhu, J., & Yuen, K. W. Y. (2020). “Lessons from the extremes: Epigenetic and genetic regulation in point monocentromere and holocentromere establishment on artificial chromosomes.” *Experimental Cell Research*, *390*(2), 111974. https://doi.org/10.1016/j.yexcr.2020.111974

Zhao, J., Bacolla, A., Wang, G., & Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences*, *67*(1), 43–62. https://doi.org/10.1007/s00018-009-0131-2

CenH3 distribution reveals extended centromeres in the model beetle *Tribolium castaneum*

Tena Gržan☯, Evelin Despot-Slade☯, Nevenka Meštrović, Miroslav Plohl*, Brankica Mravinac*

Division of Molecular Biology, Ruđer Bošković Institute, Zagreb, Croatia

☯ These authors contributed equally to this work.

* plohl@irb.hr (MP); Brankica.Mravinac@irb.hr (BM)

PLOS Genetics

October, 2020

16(10): e1009115

Abstract

Centromeres are chromosomal domains essential for kinetochore assembly and correct chromosome segregation. Inconsistent in their underlying DNA sequences, centromeres are defined epigenetically by the presence of the centromere-specific histone H3 variant CenH3. Most of the analyzed eukaryotes have monocentric chromosomes in which CenH3 proteins deposit into a single, primary constriction visible at metaphase chromosomes. Contrary to monocentrics, evolutionary sporadic holocentric chromosomes lack a primary constriction and have kinetochore activity distributed along the entire chromosome length. In this work, we identified cCENH3 protein, the centromeric H3 histone of the coleopteran model beetle *Tribolium* castaneum. By ChIP-seq analysis we disclosed that cCENH3 chromatin assembles upon a repertoire of repetitive DNAs. cCENH3 in situ mapping revealed unusually elongated T. castaneum centromeres that comprise approximately 40% of the chromosome length. Being the longest insect regional centromeres evidenced so far, T. castaneum centromeres are characterized by metapolycentric structure composed of several individual cCENH3-containing domains. We suggest that the model beetle T. castaneum with its metapolycentromeres could represent an excellent model for further studies of non-canonical centromeres in insects.

Author summary

Centromeres are specialized chromosome regions that mark a chromosomal “attaching site” from which the microtubules of the mitotic spindle pull the sister chromatids apart, ensuring accurate segregation of genetic material during cell division. Eukaryotic species studied to date mostly possess monocentric chromosomes with a functional centromere sited at a single locus, but there are also species with holocentric (polycentric) chromosomes whose centromeres stretch throughout the entire length of the chromosome. By investigating centromere-specific variant of the histone protein H3 and its associated centromeric DNA in the model beetle *Tribolium castaneum*, we discovered that *T. castaneum* centromeres comprise several functional domains merged into one remarkably extended region classified as a metapolycentromere. *T. castaneum* metapolycentromeres occupy approximately 40% of chromosome length, and emerge as the longest centromeres documented in insects to date. We believe that *T. castaneum*, as the representative species of the most numerous eukaryotic order Coleoptera could be a valuable model for future studies addressing atypical centromere structures.

Introduction

Centromeres are chromosomal loci essential for chromosome pairing and accurate chromosome segregation during cell division. They represent a base upon which the kinetochore, a proteinaceous complex and the attachment site for spindle microtubules, is assembled. Despite centromere crucial role in chromosome inheritance and genome stability, organization of centromere domains varies among organisms [1,2]. From structural point of view, eukaryotic chromosomes are mainly monocentric, having functional centromere located within a single, cytologically distinct primary constriction. Regions of primary constrictions are often made up of highly repetitive DNA sequences such as satellite DNAs and mobile elements [3]. Contrary to monocentrics, holocentric chromosomes lack a primary constriction and have kinetochore activity distributed along almost the entire chromosome length. Holocentricity is less commonly found in eukaryotes studied so far, but it is assumed that it has arisen in different plant and animal lineages in multiple independent events [4]. In animals, holocentric chromosomes have been evidenced in nematodes [5] and arthropods [6] to date. A novel, possibly intermediate type of centromere structure has recently been identified in the plant genera Pisum [7] and Lathyrus [8], whose metapolycentric chromosomes comprise several centromere domains in a single, but remarkably elongated primary constriction.

Striking diversity of centromeric DNA sequences, not only between evolutionary distant species but also between closely related ones, led to a conclusion that there is no conserved DNA sequence that would be requisite or sufficient for centromere function. Neocentromeres, formed in ectopic sites upon anonymous sequences [9], as well as the new generation of the human artificial chromosomes that lack repetitive centromeric DNA [10], additionally challenge the idea of DNA sequence relevance. In the absence of a DNA cornerstone, the currently accepted definition of centromere identity is formulated epigenetically and relies on the presence of CenH3 proteins, centromere-specific histone H3 variants [11]. The human CENP-A was the first detected CenH3 protein [12], and afterwards its homologues have been identified in a wide range of eukaryotic organisms, from yeast to different animal and plant species [13]. Unlike the canonical histone H3, that is evolutionary well-conserved across eukaryotes, CenH3 proteins show considerable sequence variability at different

taxonomic levels. The N-terminal tail is the most variable part of the amino acid sequence, making CenH3s often species-specific. On the other hand, the C-terminal part with a histone fold domain (HFD) is more conserved. In addition to the highly divergent N-terminal tail, certain changes in the HFD region are declared to be CenH3-specific, so they discriminate CenH3 variants from the canonical H3 and set bioinformatic criteria for identifying putative CenH3 in silico [14]. Although there have been evidenced species that lack CenH3 [6,15,16], CenH3 proteins still represent the most reliable markers of active centromeres. Nevertheless, great interspecies CenH3 varieties, corroborated by extremely divergent underlying DNA sequences, provoke the question of the centromere paradox where the function of the centromere is evolutionarily preserved while its DNA and protein constituents evolve rapidly [17]. The centromere drive hypothesis postulates that the rapid evolution of centromeric components could be triggered by centromeric sequences behaving as selfish genetic elements which drive non-Mendelian chromosome transmission during meiosis, thus stimulating the concurrent evolution of centromeric proteins to reconstitute fair segregation [17,18].

The red flour beetle *Tribolium castaneum* is an important world-wide pest of stored grain and grain products [19]. It is also the representative species of the Coleoptera, the most numerous and most diverse eukaryotic order with approximately 400000 species and 25% of all animal species described [20]. Due to easy rearing, plentiful offspring, relatively short life cycle, lots of mutants, and availability and proliferation of tools for its genetic analysis and manipulation, *T. castaneum* has been used in laboratory research for nearly 50 years, what makes it the most studied insect model system after Drosophila [21,22]. *T. castaneum* is also the first beetle whose genome has been sequenced [23]. In spite of that, 20% of the genome (44 Mb out of 204 Mb), primarily from the (peri)centromere regions, remained unmapped to genome assembly. The genome assembly with the contigs assembled into 10 chromosome groups was recently upgraded to the improved version Tcas5.2 [24], but the centromeric gaps are still unpatched. It was experimentally estimated that 17% of the *T. castaneum* genome is built of the highly abundant major satellite DNA TCAST that, according to cytological evidence, resides mainly in pericentromeres and centromeres of all *T. castaneum* chromosomes [25]. Since TCAST makes only 0.3% of the assembled genome [26], it can be assumed that this satellite is the major candidate

for a centromere-eligible DNA sequence. As determined by reassociation kinetics, the repetitive DNA content in the *T. castaneum* genome is about 40% [27], and is mainly composed of tandemly repeated sequences and transposable elements residing for the most part within the chromosomal regions of low recombination [26]. Surprisingly, it was shown that at least 4% of the total genome sequence is comprised of nine satellite DNA families that are located principally in euchromatic portion of the *T. castaneum* genome [28]. Despite several elaborated studies on repetitive DNA content and distribution patterns in *T. castaneum*, its centromere regions as well as centromeric proteins have not been explored thus far.

In this work, we identified cCENH3 protein, a histone H3 centromeric variant of the beetle *T. castaneum*. By cCENH3 in situ mapping we provided evidence that cCENH3 occupies approximately 40% of *T. castaneum* chromosomes' length, suggesting extraordinarily extended centromeres. By chromatin immunoprecipitation followed by high-throughput sequencing of cCENH3-associated DNA, we determined centromere-competent DNA sequences, among which the major satellite DNA dominates. This work represents the survey of centromeric regions in a species with unusually high content of a single repetitive DNA family, but also the pioneering experimental study of CenH3 proteins in the most species-rich eukaryotic order of Coleoptera in general.

Results

Identification of *Tribolium castaneum* cCENH3 coding sequence

In order to determine the CenH3 protein in *T. castaneum*, we performed an initial BLASTP search against the *Tribolium castaneum* OGS3 (Official Gene Set) database (http://beetlebase.org/blast/blast.html) using histone H3 protein sequence (NCBI Reference Sequence XP_966487.1) as a query. BLASTP search resulted in 13 reported hits (S1 Appendix). In addition to the five matches of canonical H3 to itself, BLASTP search revealed the gene ID TC012577 match that shared 42% identity (E-value 4e-19) with the H3 sequence. TC012577 has been annotated as a "histone H3-like protein" coding sequence according to the unreviewed computer-annotated UniProtKB/TrEMBL section. The remaining seven BLASTP matches showed significantly higher E-values from 0.8 to 8.8, covering the H3 sequence query only partially (S1 Appendix). From the 13 hits, TC012577 emanated as the most promising *T. castaneum* CenH3 candidate, and we named it cCENH3, as abbreviation for "castaneum CENH3".

The protein sequence alignment of H3 and cCENH3 demonstrated that the major difference between the two proteins is the length and composition of N-terminal domains, which share only 18% identity in amino acid sequence (Fig 1). The histone fold domains (HFD) are more conserved showing 51% identity in pairwise comparison. In the HFD domain the cCENH3 protein shows the differences reputed to be characteristic for CenH3 variants: 1) a longer loop 1 region, and 2) absence of specific amino acids including glutamine, phenylalanine, and threonine at positions 69, 85, 108, respectively as compared to the canonical H3 (Fig 1).

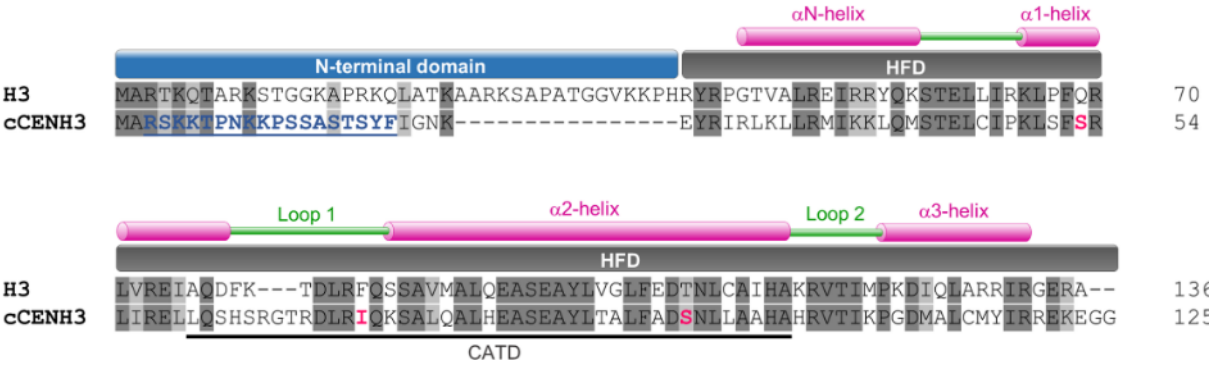


Fig 1. Protein sequence alignment of canonical histone H3 and its centromeric variant cCENH3 from *Tribolium castaneum*. The N-terminal domain and the histone fold domain (HFD) with its secondary structure are indicated above the alignment. The putative centromere-targeting domain (CATD) is marked under the cCENH3 sequence. The identical residues are highlighted in dark grey, while the similar residues are marked by light gray. The amino acid residues used for raising a peptide antibody against cCENH3 are highlighted and underlined in blue, and CenH3-characteristic amino acid changes are highlighted in red.

We performed TBLASTN search using cCENH3 protein sequence to query the latest *T. castaneum* genome assembly Tcas5.2 [24] in order to examine whether multiple copies of the gene are present in the genome. The TBLASTN search mapped the cCENH3 gene with 100% identity (E-value 1e-67) through entire length of the sequence only to one location at chromosome 9 (positions LG9:11651877–11652254) spanning a 378 bp long sequence. All the other matches, that showed remarkably lower significance (E-value >1.3e-12) covering the gene just partially (<62%), clearly suggest that the *T. castaneum* genome encodes a single copy cCENH3 gene. By using cCENH3 specific primers, we amplified and sequenced the cCENH3 sequence from genomic DNAs isolated from three different *T. castaneum* strains, including GA2 that was used in *T. castaneum* genome sequencing project [23]. With the exception of a single synonymous substitution in one of the strains, the alignment of the sequenced PCR products showed no difference in DNA sequence between the three strains (S1 Fig), disclosing high conservation of the CENH3 coding sequence between the strains.

To confirm that the cCENH3 gene is actively transcribed, reverse transcription was performed. RT-PCR reaction resulted in a unique band corresponding to ~380 bp fragments (S2 Fig). PCR products from amplification of cDNA template were cloned and sequenced. The cCENH3 transcript sequence completely matched the coding sequence (S1 Fig), proving that the cCENH3 protein is encoded by a 378 bp long gene

with no introns. We examined the expression of cCENH3 in *T. castaneum* embryo, germline and somatic tissues by analyzing publicly available transcriptome datasets from Khan et al. [29]. We found that the cCENH3 expression was the highest during embryonic development, and the expression profiles of the two embryo stages (0–5 hr and 6–11 hr) were similar (Fig 2, S2 Appendix). Compared to embryos, the ovaries and testes showed 2.4–5.4x lower expression, while the expression is drastically decreased in adult female (without ovaries) and male (without testes) carcasses (Fig 2, S2 Appendix). Such expression pattern is concordant with the significantly higher cell division activity in embryogenesis and in germline compared to somatic tissues.

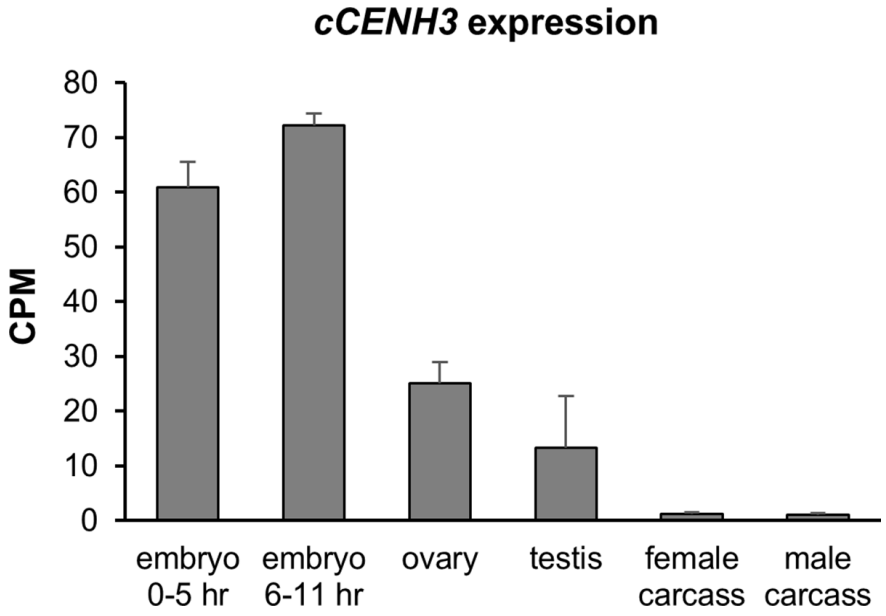


Fig 2. Expression of cCENH3 gene. The expression profiles of cCENH3 gene in embryos (0–5 hr and 6–11 hr), germline (ovary, testis) and somatic tissues (female and male carcasses without gonads) were obtained from the original RNA-seq datasets from Khan et al. [29]. Transcript hits were normalized using CPM method. Error bars represent standard deviation calculated from two (testes) or three (all the other samples) biological replicates. Data on cCENH3 transcript hits are provided in S2 Appendix.

Chromosomal localization of the cCENH3 protein

We further produced an antibody specific for the N-terminal domain of cCENH3 (Fig 1). Western blot on the whole protein extract from *T. castaneum* revealed that the rabbit-raised antibody recognizes a protein of around 15 kDa (S3 Fig), consistent with cCENH3 predicted molecular weight of 14.23 kDa.

Using the cCENH3 antibody, we conducted immunofluorescence (IF) experiments on *T. castaneum* chromosome spreads. *T. castaneum* (2n = 20) has a karyotype based on 18 autosomes and a sex chromosome pair, which is composed of two relatively large X chromosome in females, while males have a Xyp parachute-like association based on the X and a minute yp chromosome [30]. Immunodetection of cCENH3 on metaphase spreads revealed paired signals at primary constriction of all chromosomes (Fig 3A), thus confirming that cCENH3 is indeed a centromeric histone. *T. castaneum* chromosomes differ in their size and form, and therefore the position of cCENH3 signals varies between chromosomes reflecting their metacentric, submetacentric or acrocentric architecture.

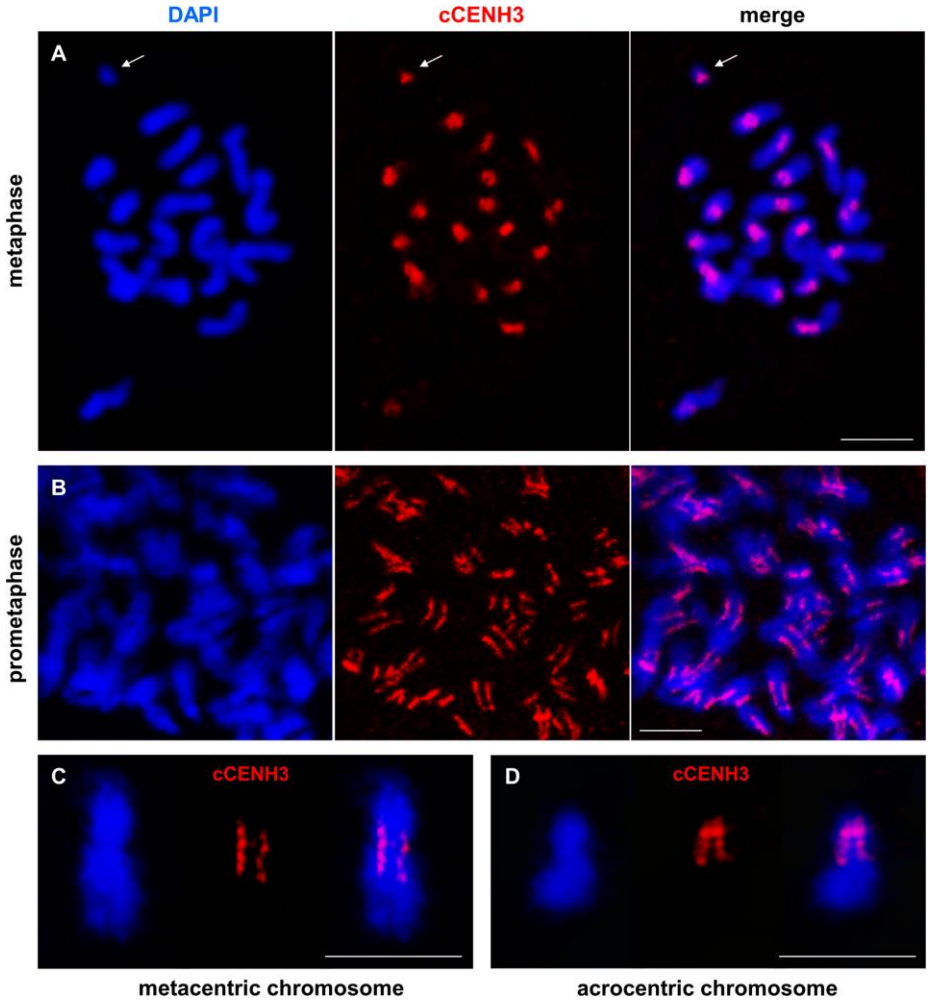


Fig 3. Localization of cCENH3 protein on *Tribolium castaneum* chromosomes. The position of cCENH3 protein (shown in red) was determined on *T. castaneum* chromosomes (shown in blue) by immunodetection using cCENH3 antibody. (A) On metaphase chromosomes cCENH3 localizes to the primary constrictions of all chromosomes including the minute yp chromosome (marked by an arrow). (B) Prometaphase chromosomes show extended cCENH3-containing regions with bead-like signal patterns composed of multiple cCENH3-containing domains. Enlarged view of metapolycentromeres with cCENH3-containing domains separated by cCENH3-lacking segments is shown at metacentic (C) and acrocentric (D) chromosome. The chromosomes are counterstained with DAPI. Scale bars = 5 μm.

Interestingly, cCENH3 distribution at prometaphase chromosomes did not demonstrate a typical centromeric dot-like signal. Instead, elongated cCENH3 signals disclosed extended centromeric regions (Fig 3B), which are best apparent at the metacentric (Fig 3C) and acrocentric (Fig 3D) chromosomes. Moreover, prometaphase chromosomes revealed the bead-like cCENH3 signals composed of several, in most instances four, individual cCENH3-containing domains (Fig 3B–3D). This form of "metapolycentric" signal has been evidenced irrespectively to the centromere position on the chromosome (Fig 3C and 3D). To estimate the relative size of the extended *T.*

castaneum centromeres, the distances between the two outermost cCENH3-containing domains were measured. Based on measurement of 240 prometaphase chromosomes, the centromere regions average 43.5% of the chromosome length (Fig 4, S3 Appendix). Due to relative small size (1–5 μm in the most condensed form), *T. castaneum* chromosomes cannot be easily distinguished [30]. However, some of them can be recognized as the longest metacentric autosome ch3 (LG3), the second largest metacentric chromosome ch2 (LG2), the largest acrocentric chromosome 4 (LG4), and the smallest chromosome yp (Fig 4). According to the measurements performed on these four chromosomes, the cCENH3 signals occupy 40.6–46.5% of their length (Fig 4, S1 Table, S3 Appendix). Although the exact calculation is not possible due to different level of chromatin condensation in euchromatic and heterochromatic regions, we estimate that the longest centromere of approximate 15.8 Mb belongs to the longest chromosome ch3 (LG3), while the minute yp chromosome has the smallest centromere of about 2.3 Mb. It has to be stressed that in contrast to the other *T. castaneum* chromosomes that exhibit bead-like cCENH3 distribution, the yp shows a single, dot-like cCENH3 signal (Fig 4). Notwithstanding the bead-like or dot-like patterns, the cCENH3 signals are always located at the poleward surfaces of the primary constrictions (Figs 3 and 4), as expected for a centromeric histone H3 variant.

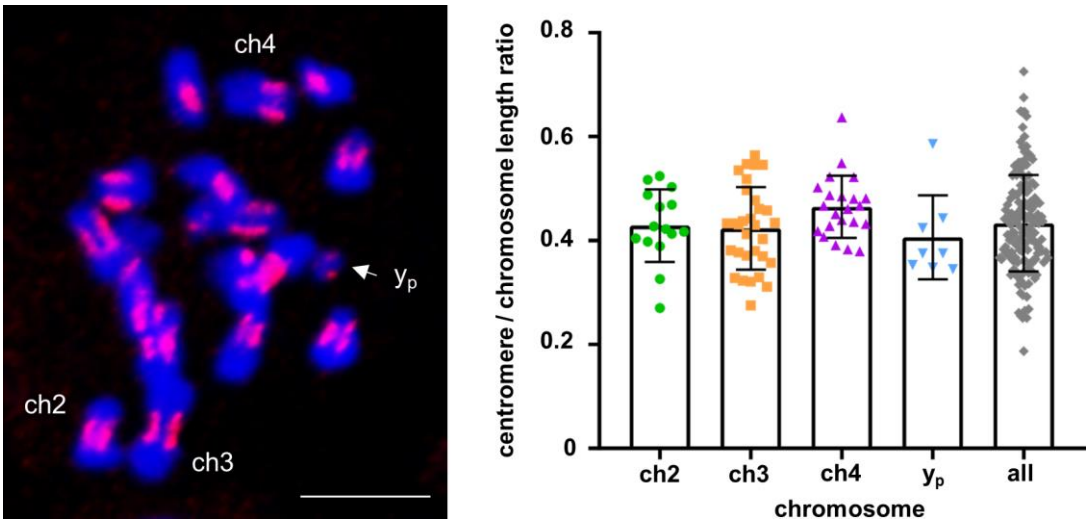


Fig 4. *Tribolium castaneum* relative centromere length estimation. Centromere length relative to chromosome length was calculated from 240 measured chromosomes. The left panel shows a representative chromosome spread with DAPI-stained chromosomes (shown in blue) with red-stained cCENH3 regions. Centromere length was defined as the distance between the two outermost cCENH3-containing domains at each chromosome. The scatter plot displays values for all measured chromosomes, and for the individual chromosomes ch2, ch3, ch4 and yp that can be distinguished in

the chromosome complement (marked in the left panel, scale bar = 5 μm). Error bars represent standard deviations of the underlying measurement data supplied in the S3 Appendix.

Detection of cCENH3-associated DNA sequences

In order to elucidate DNA sequences associated with cCENH3 at *T. castaneum* centromeres, we applied a native chromatin immunoprecipitation (ChIP) approach using the antibody directed against cCENH3. DNA extracted from cCENH3-immunoprecipitated chromatin was fluorophore-labeled and fluorescence in situ hybridization (FISH) experiment was performed. Although the intensity and the size of FISH signals were not uniform, the cCENH3-ChIPped DNA probe hybridized to centromeric regions of all *T. castaneum* chromosomes (S4 Fig). The cCENH3-ChIPped DNA signals cover on average 45.8% of the chromosome length (S4 Appendix), which is in agreement with the estimated average size of cCENH3 regions (Fig 4). The position of the cCENH3-ChIPped DNA signals corresponds to the acrocentric and submetacentic morphology previously presumed for the majority of *T. castaneum* chromosomes [30]. This result provided cytological evidence for centromere specificity of cCENH3-ChIPped DNA material and suggested a large amount of repetitive DNA in *T. castaneum* centromeres.

Next, to investigate the identity of cCENH3-ChIPped DNA sequences, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq). Because of the assumed high DNA repetitiveness of centromeric regions, in the ChIP-seq data analysis we applied the strategy introduced by Neumann et al. [7]. In this approach, the repetitive DNA content of the genome is first classified into clusters according to the graph-based repeat clustering analysis [31], and the output of the annotated clusters serves as a reference for similarity-based mapping of ChIP-Seq reads. The advantage of this approach is that it does not depend on a genome assembly, and is optimized for analyzing next-generation sequence reads, as the algorithm uses short sequences randomly sampled from the low pass genome sequencing data. The workflow of the cCENH3-ChIP-seq experiments and data analysis is presented in S5 Fig. The *T. castaneum* genome has been assembled into ten chromosome/linkage groups, but the current Tcas5.2 assembly lacks 20% of the genome sequence attributed to the heterochromatic (peri)centromeric regions [23,24,26], so it could not

serve as a reliable repeat database reference. Therefore, we re-sequenced the *T. castaneum* genomic DNA using the Illumina HiSeq platform. From the low pass genome sequencing data and graph-based sequence clustering obtained by RepeatExplorer2 analysis, we generated the repeat database directly from the unassembled Illumina WGS reads. The repeat reference database was constructed from 270200 randomly selected, 151 nt long WGS reads that ensured 0.2x genome coverage, determined as optimal (as explained in the Materials and Methods section). Based on the RepeatExplorer2 analysis, 51.44% of the WGS reads were classified into 21564 clusters, while the rest (48.56%) represents ungrouped, singleton reads (Fig 5). As the clusters' order reflects their genome abundance, the first 1000 clusters were used for subsequent similarity-based mapping of ChIP-seq reads, while the rest were omitted from the analysis due to their low representation. For ChIP-Seq Mapper analysis, one million reads were randomly selected for the cCENH3-ChIPped sample together with one million reads randomly selected for the input sample (DNA isolated from an aliquot of native chromatin prior to ChIP). The ratio of ChIP to input read hits was calculated for the 1000 WGS reference clusters. The ChIP-Seq Mapper analysis set the mean ratio between ChIP and input hits to 2 (S6A Fig), propounding it as a ChIP enrichment threshold that 37 out of 1000 analyzed clusters exceeded (S6B Fig). However, as the enrichment threshold was set relatively low, when judging the centromere competence of a certain cluster, the number of supporting ChIP hits has to be considered. For that reason, we proceeded to further analysis only with the >2-fold enriched clusters that were supported by at least 0.01% of ChIP hits (i.e. 100 ChIP hits) (Table 1). We also included into analyses additional seven clusters showing ChIP enrichment >1, but being represented with >10000 ChIP hits (>1% of the analyzed reads). The clusters that met mentioned criteria (Table 1) were mapped to the *T. castaneum* Tcas5.2 genome assembly. Strikingly, we mapped them largely outside the assembled regions, in unplaced scaffolds and singletons, or in the arrays adjoining the unassembled regions. The clusters listed in the Table 1 were also BLAST-searched against Repbase and GenBank databases to reveal the possible identity or similarity with annotated sequences. According to similarity-search analysis, the clusters enriched in cCENH3-ChIP-seq data can be classified into four groups (Table 1): tandemly repeated sequences or satellite DNAs, transposable elements, rDNA-like sequences, and anonymous sequences.

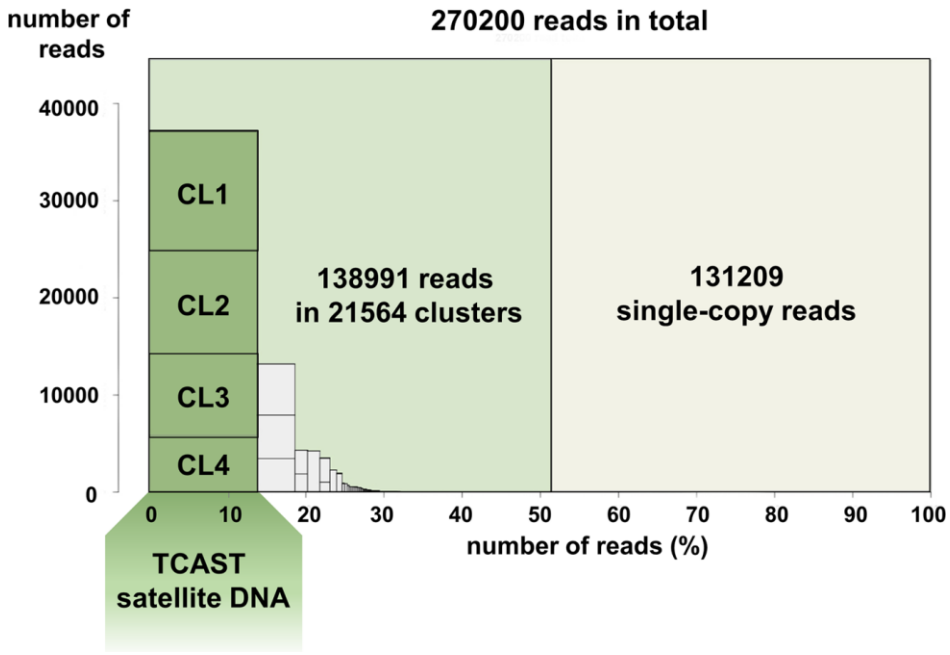


Fig 5. Summary of the *Tribolium castaneum* genome clustering. 270220 Illumina reads corresponding to 0.2x genome coverage were analyzed using RepeatExplorer2 pipeline [31]. 138991 reads were assorted into 21564 clusters representing repetitive fraction of genomic DNA, while 131209 reads were classified as single-copy reads. The four most highly-repetitive clusters (CL1-CL4, marked by dark green columns) belong to the supercluster of *T. castaneum* major satellite DNA TCAST, and cumulatively comprise 13.8% of the genome.

Table 1. Characterization of cCENH3-associated repetitive clusters.

Criteria	Cluster[1]	ChIP hits[2]	Input hits[2]	ChIP/ Input ratio	Repeat type[3]
>10000 ChIP hits (>1%) AND ChIP/Input ratio >1	CL3	42004	38140	1.10	TCAST satellite DNA[4]
	CL4	31220	27988	1.12	TCAST satellite DNA[4]
	CL10	23779	14796	1.61	28S rDNA-like
	CL12	28617	19621	1.46	*T. castaneum* LINE-1 element ORF2 protein (LOC658088)
	CL13	42765	26017	1.64	18S rDNA-like
	CL19	12573	9903	1.27	Cast6 satellite DNA[5]
	CL20	10166	7557	1.35	DNA transposon (Helitron)-like
	CL48	10565	5052	2.09	tandem repeat
>100 ChIP hits (>0.01%) AND ChIP/Input ratio >2	CL141	488	158	3.09	*T. castaneum* non-LTR retrotransposon SARTTc1
	CL143	460	192	2.40	*T. castaneum* non-LTR retrotransposon SARTTc1
	CL169	495	238	2.08	*T. castaneum* non-LTR retrotransposon SARTTc3
	CL198	210	99	2.12	DNA transposon (Polinton)-like
	CL234	402	158	2.54	LTR retrotransposon (Copia)-like
	CL253	288	110	2.62	*T. castaneum* non-LTR retrotransposon SARTTc3
	CL264	337	145	2.32	*T. castaneum* non-LTR retrotransposon SARTTc1
	CL270	732	342	2.14	5S rDNA
	CL298	296	137	2.16	non-LTR retrotransposon (Jockey)-like
	CL336	177	70	2.53	*T. castaneum* non-LTR retrotransposon SARTTc3
	CL391	1591	602	2.64	unknown
	CL394	442	85	5.20	*T. castaneum* uncharacterized LOC107399056. partial mRNA
	CL399	202	61	3.31	*T. castaneum* non-LTR retrotransposon SARTTc1
	CL469	145	53	2.74	*T. castaneum* non-LTR retrotransposon SARTTc1
	CL611	216	78	2.77	*T. castaneum* tigger element-derived protein 4 (LOC103314401)
	CL669	118	58	2.03	unknown
	CL711	174	78	2.23	*T. castaneum* non-LTR retrotransposon SARTTc3
	CL797	115	41	2.80	*T. castaneum* mobile element jockey-like (LOC103314666)
	CL942	133	39	3.41	unknown
	CL992	118	34	3.47	*T. castaneum* non-LTR R2 retrotransposon

[1]Clustering of 270200 randomly selected WGS Illumina reads, corresponding to 0.2x genome coverage, was performed using RepeatExplorer2 pipeline [31].

[2] ChIP and Input hits values were generated using one million cCENH3-ChIP and Input reads in ChIP-seq Mapper analysis [7]. ChIP enrichment was calculated using the top 1000 most repetitive WGS clusters obtained by RepeatExplorer2 clustering.

[3] Repeat type was determined by BLAST searching against GIRI Repbase and NCBI GenBank database.

[4] Ugarković *et al.* [25]

[5] Pavlek *et al.* [28]

Repeat classification of cCENH3-associated DNA sequences

Among cCENH3-associated tandem repeats, the clusters CL3 and CL4 dominate in terms of ChIP hits numbers (Table 1). These two clusters belong to the *T. castaneum* major satellite DNA TCAST, based on 360 bp long monomers [25]. In addition to being highly abundant, TCAST is a very heterogeneous satellite family composed of five subfamilies mutually divergent up to 30% [28]. The RepeatExplorer2 analysis assorted the TCAST satellite into four clusters, CL1-CL4, which cumulatively comprise 13.8% of the genome (Fig 5, Table 2). According to ChIP-Seq Mapper analysis, the clusters CL3 and CL4 were found enriched 10% and 12%, respectively, while the clusters CL1 and CL2 showed the ChIP/Input ratios of 0.84 and 0.90, respectively (Table 2). We analyzed the proportion of individual TCAST subfamilies defined by Pavlek et al. [28] in the CL1-CL4 clusters (Table 2), and we found that none of the four clusters

corresponds specifically to any of the TCAST subfamilies. For instance, the cluster CL3 includes mainly the reads associated to the subfamilies 1 and 4, but it also comprises the reads associated with subfamilies 2, 3 and 5 (Table 2). Similarly, CL4 is mostly based on the subfamily 3 reads, but it also harbors the reads sharing similarity with other subfamilies (Table 2). It is possible that the clusters cannot be attributed to the individual subfamilies because they were generated from the reads of submonomeric length (151 bp of ~360 bp). Since we did not find either a motif or any specific trait that would strictly differentiate CL3 and CL4 reads in comparison to CL1 and CL2 reads, we presume that all five TCAST subfamilies participate in the centromeric chromatin.

Table 2. Contribution of the TCAST satellite subfamilies to four most abundant WGS clusters.

Cluster[1]	Genome proportion[1]	cCENH3-ChIP enrichment[2]	TCAST subfamilies[3] proportions in the cluster				
			Subfamily 1	Subfamily 2	Subfamily 3	Subfamily 4	Subfamily 5
CL1	4.5%	0.84	7.18%	0.18%	15.27%	35.50%	41.25%
CL2	4%	0.90	71.32%	22.33%	0.14%	5.47%	0.67%
CL3	3.1%	1.10	47.12%	7.58%	2.20%	39.17%	3.88%
CL4	2.2%	1.12	0.03%	0.58%	82.12%	9.06%	8.14%

[1] Clustering of WGS Illumina reads and their estimated genome proportions were obtained by RepeatExplorer2 analysis [31].

[2] ChIP enrichment for cCENH3 protein was calculated by performing ChIP-seq Mapper analysis [7]. The values represent the ratio of cCENH3-ChIP hits and Input hits.

[3] Satellite DNA TCAST classification into five subfamilies is based on Pavlek *et al.* [28]. The subfamily consensus sequences were used in RepeatExplorer2 clustering to determine the proportions of satellite subfamilies within each cluster.

Although the reads belonging to TCAST satellite dominate by number in the ChIP sample, based on previously established TCAST localization in the heterochromatic (peri)centromeric blocks of *T. castaneum* chromosomes [25], we expected that TCAST would show higher enrichment for cCENH3 than it did in our ChIP experiment. Therefore, to explore in situ interrelation between cCENH3 and TCAST satellite DNA, we performed combined IF-FISH using the mixed DNA probe representative of all five TCAST subfamilies. First, IF-FISH experiments were conducted on *T. castaneum* metaphase chromosomes (Fig 6). In spite of the fact that the intensity and the extent of TCAST signals on metaphase chromosomes oscillate from very strong at majority of autosomes to discrete at the yp chromosome, IF-FISH confirmed that cCENH3 and TCAST signals coincide at centromeric regions of all chromosomes. The expanse of the FISH signals indicates that TCAST arrays outspread beyond cCENH3 domains, suggesting their presence also in the pericentromeric regions. Additionally, the TCAST signal outspread beyond the cCENH3 domains could potentially be influenced to some

degree by the methodology itself as the use of acetic acid in IF-FISH chromosome preparation reduces the cCENH3 signal strength, while the FISH signal amplification procedure could result in the TCAST signal overestimation. Since resolution of highly condensed chromatin at metaphase chromosomes does not provide insight into linear organization of centromeric chromatin constituents, we applied IF-FISH to extended chromatin fibers. Optical mapping on extended chromatin fibers revealed long arrays of TCAST satellite that were also largely occupied by cCENH3 protein (Fig 7A). The quantification of the 15 chromatin fibers with high presence of both TCAST and cCENH3 signals showed that TCAST/cCENH3 signal overlaps cover between 62% and 88% of the measured fibers' length (S5 Appendix). We observed, nevertheless, that at some continuous TCAST arrays cCENH3 is loosely distributed (Fig 7B). cCENH3-poor regions could be "linkers" between cCENH3-rich domains observed as bead-like patterns at prometaphase chromosomes (Fig 3C and 3D). Alternatively, they might come from pericentromere with potentially sparse cCENH3 domains. Additionally, in some segments of chromatin fibers we also noticed sporadic distribution of cCENH3 signals that do not co-localize with TCAST signals, but reside in the gaps within TCAST arrays (Fig 7C). It can be assumed that at these scattered spots cCENH3 nucleosomes are associated with other sequence candidates listed in Table 1.

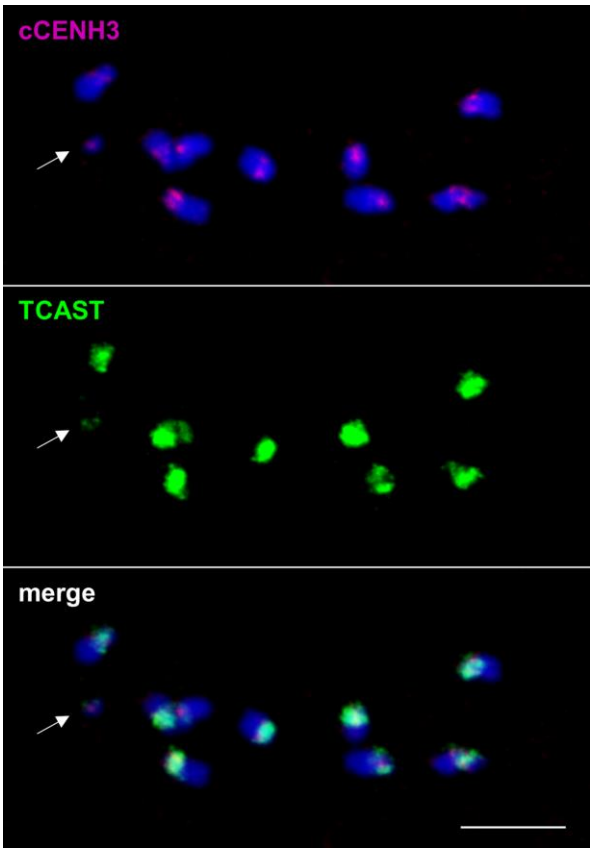


Fig 6. Co-localization of cCENH3 protein and TCAST satellite DNA on *Tribolium castaneum* meiotic chromosomes. Combined IF-FISH analysis of in situ localization of cCENH3 protein (shown in magenta) and TCAST satellite DNA (shown in green) demonstrated their co-localization at centromeric regions of all *T. castaneum* chromosomes. An arrow points to the yp chromosome. Scale bar = 5 μm.

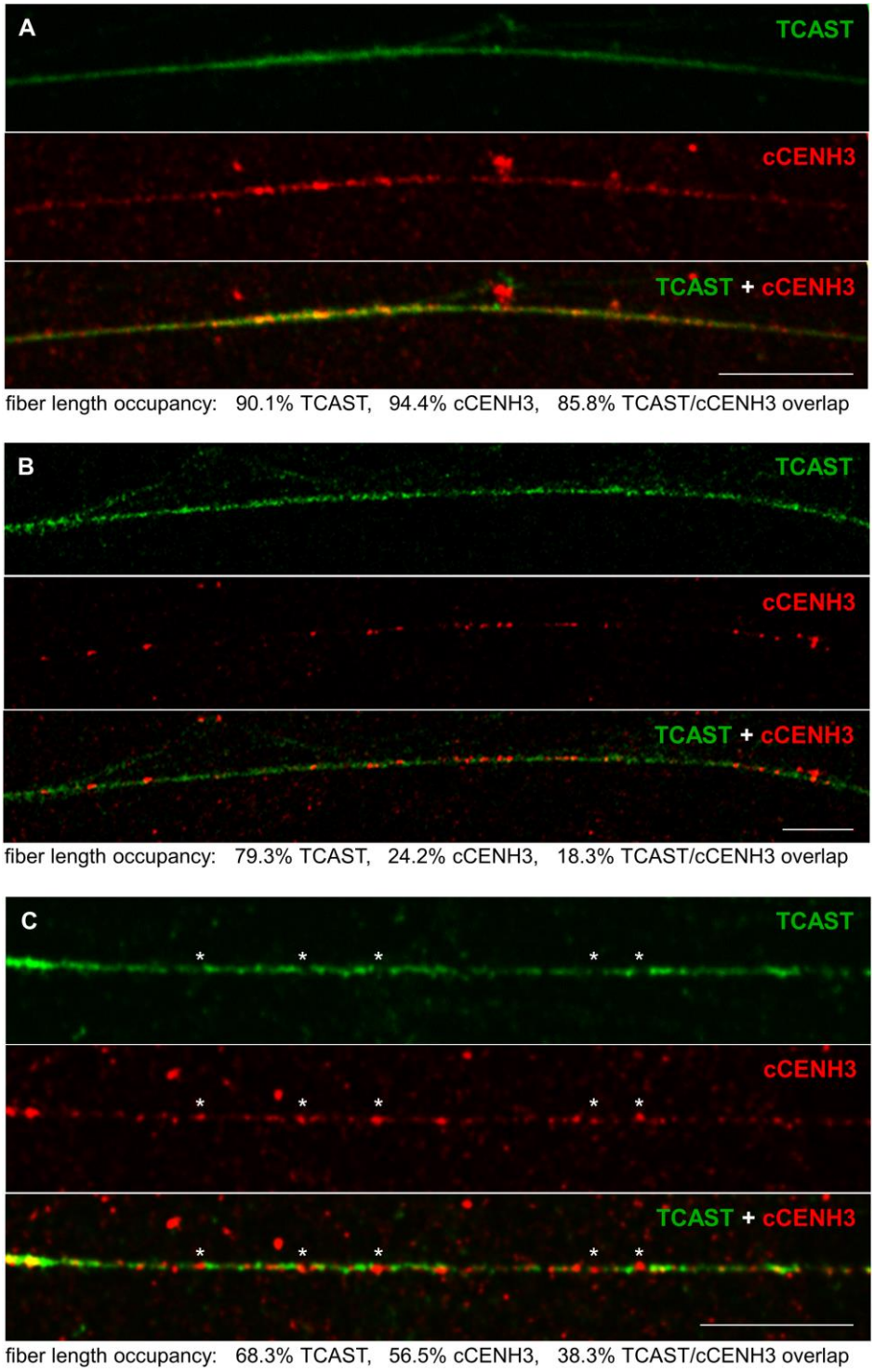


Fig 7. Co-localization of cCENH3 protein and TCAST satellite DNA on *Tribolium castaneum* extended chromatin fibers. By combined IF-FISH analysis, position of TCAST satellite DNA (green signals) was revealed by FISH, while the position of cCENH3 protein (red) was detected by IF. The examples of a fiber with prominent TCAST-cCENH3 overlapping (A), a fiber mostly deprived of cCENH3 signal (B), and a fiber with asterisks-mark positions of cCENH3 signals which do not coincide with TCAST satellite signals (C) are shown. The percentages of the chromatin fiber length occupied by TCAST and cCENH3 signals are indicated under the panels. Scale bar = 10 μm.

Within tandem repeat class of sequences, in addition to TCAST, ChIP-seq data revealed enriched cluster CL19 (Table 1) which corresponds to significantly less abundant satellite DNA Cast6. Cast6, being based on a ~180 bp monomer unit and making up to 0.5% of the genome, is one of the *T. castaneum* satellites that were proclaimed to be present primarily in euchromatic regions [28]. However, Cast6 occurrence in the centromeric heterochromatin blocks of at least one chromosome [28] suggests that some of its arrays contribute to the centromere chromatin. It has to be stressed that among nine "euchromatic" satellites that have been described in *T. castaneum* [28], only Cast6 was cCENH3-enriched, while the others were underrepresented in ChIPped data set showing enrichment factors between 0.44–0.73. CL48 is another tandem repeat enriched in cCENH3 reads (enrichment factor of 2.09, Table 1), that comprises 0.05% of the genome. Composed of average 73 bp long monomers and predominantly present in unplaced singletons, this repeat represents a novel satellite DNA not described in the *T. castaneum* genome so far.

In the group of cCENH3-enriched clusters that comprise mobile elements, there are predominant nine clusters which share similarity with *T. castaneum* non-long terminal repeat (non-LTR) retrotransposons from the SARTTc family. The SARTTc family includes seven types of retrotransposons (SARTTc1-SARTTc7) with target sequence preference to telomeric TCAGG motif [32]. According to our analysis, five clusters CL141, CL143, CL264, CL399, and CL469, enriched 2.3–3.3 times, are related to the non-LTR retrotransposon SARTTc1 sequence segments, while the clusters CL169, CL253, CL336, and CL711, with enrichment factors 2.1–2.6, are associated with the SARTTc3 sequence (Table 1). The presence of SARTTc sequences in the centromeric chromatin could explain the weak telomeric FISH signals on some of the *T. castaneum* chromosomes obtained by cCENH3-ChIPped DNA probe (S4 Fig).

Among rDNA-related clusters, the cluster CL270 shows the highest ChIP/Input ratio of 2.14 (Table 1). Interestingly, this cluster is identical to the 5S rDNA, whose 192 bp long repeat unit consists of a 119 bp long 5S rRNA gene and 73 bp long non-transcribed spacer (NTS) region. Although centromeric regions are not typical rDNA locations, ChIP enrichment suggests that a portion of the 5S rDNA arrays is drafted into centromere function. Besides tandem repeats, mobile elements and rDNA-like sequences, ChIP-seq analysis implies that other sequence candidates, whose status

we have not been able to determine, are likely to be centromeric (Table 1). Even though their contribution might be less prominent in terms of sequence abundance, they manifest the high diversity of DNA sequence composition at *T. castaneum* centromeres.

Discussion

The red flour beetle *T. castaneum* with unassembled 20% of the genome corresponding to (peri)centromeric regions [23] certainly represents a challenging and compelling model organism for centromere biology research. In this work we ascertained cCENH3, the centromere-specific histone H3 of *T. castaneum*, and utilized it to determine architecture and characterize DNA composition of *T. castaneum* centromeres.

We identified the gene ID TC012577 from the current *T. castaneum* genome assembly Tcas5.2 [24] as the CenH3 homologue, and its coding sequence is identical to a candidate computationally predicted by Drinnenberg and coauthors [6]. By demonstrating cCENH3 chromosomal position, we validated cCENH3 experimentally as an authentic centromeric variant of histone H3. In this way it was confirmed that the prediction of CenH3 protein using defined bioinformatics criteria [14] was accurate and reliable in the *T. castaneum* case. CenH3 proteins differ from the canonical H3 foremost in N-tail domain, and CenH3 N-tail domain often extends over 100 amino-acid residues, as in *Drosophila melanogaster* CID [33], *Caenorhabditis elegans* HCP-3 [34] or *Saccharomyces cerevisiae* CSE4 [35]. *T. castaneum* cCENH3 is a rather small protein of 125 aa, having N-tail domain even shorter than the canonical H3 (24 aa versus 40 aa). Despite the shortness of NH2-terminal tail, we successfully produced an antibody that specifically recognizes cCENH3 and discriminates it from the canonical H3.

The repertoire of *T. castaneum* centromeric DNA sequences, which we obtained by cCENH3-ChIP-seq, includes a variety of repetitive DNAs. Notably, all of these sequences, regardless of a repeat type they belong to, are predominantly found in the unplaced scaffolds and singletons related to the unassembled part of the current Tcas5.2 genome assembly. Association with the unassembled genome fraction might indicate that these sequences belong to centromeric regions that have not yet been delineated for any of *T. castaneum* chromosomes. Not surprisingly, the major centromeric DNA constituent is the TCAST satellite that comprises one-sixth of the genome [25]. According to optical mapping on metaphase chromosomes and extended chromatin fibers, TCAST constitutes very long arrays that also include cCENH3

domains. Cytological experiments showed, however, that TCAST is not limited to the centromeres, and its arrays stretch beyond the cCENH3 signals. TCAST presence in the pericentromere could partly be a reason why its enrichment in cCENH3-ChIP experiments was lower than expected. Furthermore, it has been known for a long time that at centromeres with underlying highly abundant repeats, CenH3 nucleosomes interact generally only with a fraction of those repeats [36,37], and such a partial occupancy can diminish their relative ChIP enrichment. In the recent study of *D. melanogaster* centromeres, its most copious centromeric repeat Prodsat revealed lower CenH3-enrichment in comparison to AATAG repeat, although Prodsat was seven times more abundant in the IP fraction [38], and the authors stressed that not only enrichment, but also the abundance of a sequence has to be regarded when estimating its centromere potential. In the model plant species Arabidopsis thaliana functional centromeres assemble on the most homogeneous and evolutionary youngest specific subset of CEN180 repeats, mostly excluded from the genome assembly [39]. Similarly, in human centromeric chromatin dominate two dimeric alpha-satellite DNA units, mutually highly diverged, but both being highly homogenized and younger than the rest of the alpha-satellite variants [40]. *T. castaneum* TCAST has been described as a very divergent satellite DNA compiled from five subfamilies [28], but in our study we have not found an evidence that cCENH3-containing nucleosomes have any preference for a unique subfamily or a subset of TCAST monomers. According to FISH analysis performed by Pavlek and coauthors [28], all five TCAST subfamilies coincide at the centromeric regions of all chromosomes. Computational analysis of higher order repeat (HOR) structures in *T. castaneum* disclosed TCAST-incorporating HORs found solely in unplaced scaffolds and singletons; particularly interesting is the fact that the TCAST monomers in those HORs are very heterogeneous, come from different subfamilies and are combined with extraneous sequence elements [41]. Our ChIP results corroborate previous cytogenetic TCAST localization [28] and bioinformatic HOR analysis [41], and suggest that functional *T. castaneum* centromeres are not built upon an exclusive fraction of TCAST satellite. Instead, it is more likely that they comprise different variants/subfamilies of the major satellite DNA intermingled with other DNA sequences.

Satellite DNAs are among the most frequent DNA constituents of functional centromeres [42], and therefore it is not unexpected that in cCENH3-ChIP data set,

beside the major satellite DNA TCAST, we also found enriched other satellite DNAs. However, those satellites are less abundant by an order of magnitude or even two in comparison to TCAST. Thirteen satellite DNA families whose genome abundance differ ~10–1000 times have been recorded in the pea centromeres [7], confirming that tandem organization is favorable, while abundance of a tandem repeat could be less determinative criteria for centromeric assortment. In addition to satellites, retroelements can be equally important or even dominant centromere components [43], and contribution of retrotransposons to functional centromeres have been documented in different animal, plant and fungal lineages [43–47]. In *T. castaneum*, in the group of different cCENH3-associated transposable elements we found prevalent non-LTR retrotransposons SARTTc. Notwithstanding SARTTc elements were reported as telomeric retrotransposons that intermingle with TCAGG repeats in *T. castaneum* "composite" telomeres [32], we mapped them in unassembled scaffolds in the vicinity of (TCAGG)n arrays that are relatively short (100–400 bp) and might represent heterochromatic interstitial telomeric sequences (ITS) within centromeric regions. It has been assumed that ITSs at (peri)centromeres are mostly remnants of telomere-mediated DNA repairs and chromosomal rearrangements [48], but for different Solanum species including tomato, potato and eggplant, it was revealed that some ITS subfamilies expanded into functional centromeres [49]. Besides satellite repeats and retrotransposons, in the *T. castaneum* centromeric chromatin we also detected rDNA-associated sequences, and the most prominent of them is 5S rDNA with its NTS region. rDNA tracts can serve as an origin of satellite repeats [50,51], and centromerically located satellite DNA partially derived from 5S rDNA was described, for instance, in the frog Physalaemus cuvieri [52]. Remarkably, in the switchgrass Panicum virgatum it was recently shown that one of its centromeres propagated the exact 5S rDNA unit with its completely conserved 5S rRNA gene sequence and NTS region into functional centromeric DNA, without obstructing the fundamental function of the 5S rRNA genes [53]. We believe a similar scenario could be assumed in *T. castaneum*. Its centromeres, compiling different kinds of repetitive sequences, confirm that centromeric regions can be quite permissive regarding their DNA content, and that the presence of one highly abundant and dominant satellite DNA does not exclude or prevent other repetitive sequences to contribute. It can be even speculated that a widely outspread, abundant satellite DNA allows and/or fosters intrusion of different sequences whose locations are not primarily or exclusively centromeric. It is

questionable, though, whether these intrusive elements are indeed epigenetically favored or simply tolerated given their scarcity in the centromeric regions.

What makes *T. castaneum* centromeres exceptional is their unusual extent. We estimated that *T. castaneum* centromeres, ranging in size from 2.3 to 15.8 Mb, comprise over 40% of the chromosome length. In comparison, the centromeric regions in *D. melanogaster*, the insect of comparable genome size (180 Mb vs. 200 Mb), range between 100 and 170 kb [43]. To the best of our knowledge, *T. castaneum* centromeres are the longest insect regional centromeres described so far. Interestingly, in the red imported fire ant Solenopsis invicta unusually long centromeres, on average spanning one third of chromosomes' length (3.6 Mb), have also been observed [54]. In addition to the remarkable length, *T. castaneum* centromeres are marked by their metapolycentric structure, characterized by multiple CenH3-domains stretching along a large chromosome region that forms an extended, yet single primary constriction. Neumann and colleagues [7] introduced the term of metapolycentricity to describe such an unusual type of centromeres evidenced in the two closely related plant genera Pisum and Lathyrus [7,8]. In comparison to *T. castaneum*, the peas' genomes are 20–30 times bigger and their chromosomes are remarkably longer, so it does not surprise that the pea metapolycentric centromeres are notably larger ranging in size from 69 to 107 Mbp [7]. Notwithstanding the significant difference in centromere size, the peas' and *T. castaneum* centromeres share the bead-like pattern, a hallmark of metapolycentric structure.

The presence of metapolycentricity in very distant eukaryotic species suggests its convergent evolution, but the foundation of metapolycentricity remains an open question. In the pea species with metapolycentric chromosomes, CenH3 proteins are encoded by two paralogous genes that are equally expressed, and the two CenH3 proteins co-localize at the same domains of extended centromeres [7,8]. Nevertheless, Neumann and coauthors [8] warn that the potential dosage effect of CenH3 duplication cannot be solely responsible for such a drastic centromere expansion. *T. castaneum* cCENH3 protein is encoded by a single gene, and we found no evidence of additional CenH3 copies, which makes us conclude that metapolycentricity in *T. castaneum* arose through mechanisms unrelated to CenH3 duplication. Different chromosomal rearrangements, especially Robertsonian centric fusions, can lead to the enlarged

centromere size. One of the most illustrative examples is an unusually long, bead-like kinetochore region of chromosome C3X in Indian muntjac, the mammal whose small karyotype (2n = 6♀,7♂) evolved through numerous linear centromere-telomere fusions from the acrocentric 2n = 70 ancestral karyotype [55,56]. As the *T. castaneum* karyotype is based on the Coleoptera prevalent diploid number 2n = 20, it is not very plausible that its metapolycentromeres were generated via chromosomal fusions. In our opinion, the expansion of *T. castaneum* centromeres could be most convincingly explained by centromere drive hypothesis [17,18,57]. According to the centromere drive model, “stronger” centromeres, i.e. those that segregate more successfully to the egg during female meiosis, provide more microtubule attachment sites by recruiting more centromeric proteins, and enhanced recruitment of centromeric nucleosomes could be driven by centromeric satellite expansion [17]. In a mouse model system it was shown that centromeres with more centromere-competent satellite repeats indeed engage more CENP-A, thus gaining a transmission advantage in the female germline [58]. We hypothesize that the highly repetitive *T. castaneum* metapolycentric regions can be evolutionary favored due to increased capacity for promoting cCENH3 chromatin expansion, which might ensure a selective advantage in the race to reach the oocyte. Similarly, long primary constrictions in the Solenopsis fire ants coincide with the position of the major, high-copy satellite DNA according to FISH analysis, and centromere drive was proposed to propel centromere elongation [54]. Although CenH3 localization was not examined so there is no exact data on fire ants’ centrochromatin structure, it is possible that metapolycentricity among insects, especially those with highly abundant centromeric satellites, is not unique to *Tribolium*.

Morphology of metapolycentric chromosomes suggests an intermediate state between monocentric and holocentric chromosome architecture [59]. It has been proposed that holocentricity has evolved at least 13 times among eukaryotes [4]; nevertheless, the evolutionary relationship between holocentricity and monocentricity remains an enigma. Assuming monocentricity as the initial state, Drinnenberg and coauthors [6] revealed independent transitions to holocentricity in at least four lineages of insects. Significantly, 13 of the 15 analyzed holocentric insect species have lost CenH3, and it was theorized that CenH3 loss was preceded by monocentricity-to-polycentricity architectural changes [6,60]. It is tempting to speculate that *T. castaneum*, which engages CenH3 in its extended metapolycentromeres, could represent an

intermediate between monocentricity and polycentricity in insects. However, further extensive analyses are required to test this speculative hypothesis.

The beetle *T. castaneum* with its metapolycentric chromosomes whose centromeres are founded on one, genome-prevailing satellite DNA is a dare par excellence for the correct centromere assembly. For complete understanding how kinetochore sites are designated and maintained in diverse eukaryotic species, the studies of kinetochore assembly in organisms with unusual centromere anatomies are crucial. We believe that *T. castaneum*, as the coleopteran model organism, could also be an excellent model to address these questions.

Materials and methods

Insect material

All experiments in this work were performed using the highly inbred Georgia 2 (GA2) strain of the red flour beetle *T. castaneum*. The initial GA2 strain stock was obtained from USDA-ARS (Manhattan, Kansas, USA), and subsequently maintained as a laboratory culture. The insects were reared in whole wheat flour at 28°C, and sub-cultured every four weeks. In addition to GA2 strain, genomic DNAs isolated from the two additional *T. castaneum* strains DE and ES (obtained from Germany and Spain, respectively) were used for checking the nucleotide sequence of the cCENH3 gene.

DNA extraction and PCR amplification

Total DNA was isolated from ~50 mg of adult insects (25 individuals) by using DNeasy Blood and Tissue Kit (Qiagen). DNA quantification was done by Qubit fluorometer (Invitrogen) using Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific). Based on the cCENH3 coding sequence and its flanking regions, the primers cCENH3pr0 (5'-AAGGTAGGATAATGCCGA-3'), cCENH3pr1 (5'-ATGGCCCGTTCTAAG-3'), cCENH3pr3 (5'-TTAACCACCTTCTTTTTCC-3’), and cCENH3pr5 (5'-TTAACCACCTTCTTTTTCCCTCC-3’) were designed to amplify the entire cCENH3 gene sequence. PCR reaction was performed in 50-µl volume containing 1× Green GoTaq Reaction Buffer (Promega), 2.5 mM MgCl2, 0.1 mM of each dNTP, 0.1 µM of each primer and 2 units of GoTaq G2 DNA Polymerase (Promega). PCR program included predenaturation at 94°C for 3 min, 35 cycles od amplification (denaturation at 94°C for 20 s, annealing at 61°C for 20 s, and extension at 72°C for 30 s), and final extension at 72°C for 7 min. PCR products were analyzed in agarose gel electrophoresis, extracted by QIAquick Gel Extraction Kit (Qiagen), and Sanger-sequenced in Macrogen Europe Laboratory (Amsterdam, The Netherlands).

RNA isolation and reverse transcription

Total RNA was purified from ~55 mg of larvae (20 individuals) by using RNeasy Mini Kit (Qiagen), and additionally treated with RQ1 RNase-Free DNase (Promega). RNA was quality-checked on agarose gel, and quantified by BioSpec-nano Micro-volume UV-Vis Spectrophotometer (Shimadzu). The coding region of the cCENH3 gene was obtained by RT-PCR amplification using OneStep RT-PCR Kit (Qiagen) and the gene-specific primers cCENH3pr1 and cCENH3pr3. cDNA was synthesized from 250 ng of RNA in 25 µl RT reaction at 50°C for 30 min. Initial PCR activation step at 95°C for 15 min was followed by 35 cycles of amplification (denaturation at 95°C for 30 s, annealing at 61°C for 30 s, and extension at 72°C for 1 min), and final extension at 72°C for 10 min using HotStarTaq DNA Polymerase. Three negative controls were included in the RT-PCR experiment: 1) a control reaction in which reverse-transcriptase activity was inhibited by keeping the reaction on ice and placing it in the thermal cycler only after it had reached 95°C for the HotStarTaq DNA Polymerase activation step, 2) a control reaction in which RNA template was added during the HotStarTaq DNA Polymerase activation step, 3) a blank control without template RNA. RT-PCR cCENH-3 products were purified using PCR Purification Kit (Qiagen), and Sanger-sequenced with the primers above.

cCENH3 expression analysis

To examine the expression pattern of the cCENH3 gene, we used publicly available RNA-seq datasets from Khan et al. [29] deposited at the Gene Expression Omnibus (GEO) repository under accession number GSE119739. Sets of transcriptome reads corresponding to embryo 0–5 hr, embryo 5–11 hr, ovary, testis, female carcasses (lacking ovaries), and male carcasses (lacking testes), were separately mapped to the cCENH3 gene sequence using Bowtie2 with—local and—very-sensitive option implemented in Geneious R11.1.4 (Biomatters, Ltd.). All obtained hits were normalized with CPM method ((counts per million reads mapped) = (hit number/library size)*106) and average values were calculated for different runs and biological replicates (four runs for each replicate; two replicates for testes, three replicates for all the other samples).

Antibody design and production

Polyclonal IgG antibodies were generated against a synthetic peptide NH2-RSKKTPNKKPSSASTSYF-CONH2, corresponding to amino acids 3–20 of the N-terminal end of the *T. castaneum* cCENH3 protein (Fig 1). Peptide synthesis and immunization of two rabbits were performed by Pineda Antikörper-Service (Berlin, Germany). To monitor the development of the immune response, we tested by Western blotting the preimmune sera as well as the sera samples on the monthly basis during the immunization period. Although the preimmune sera showed certain background of non-specific bands corresponding to higher molecular weight proteins in the Western blot testing (S3 Fig), the polyclonal antibodies to the targeted protein of ~15 kDa were first detected after the 2-month immunization period. After 120 days, the immunization was stopped, and affinity purification of the monospecific IgG fraction from cCENH3-antisera on sepharose columns was performed. Purified monospecific IgG fraction was concentrated 25x using Amicon Ultra-0.5 centrifugal filter device (Merck), and used in all downstream applications.

Protein extraction and Western blot

125 mg of snap-frozen *T. castaneum* adults was ground to a fine powder using mortar, pestle and liquid nitrogen. Immediately after grinding, tissue powder was transferred to 5 ml of ice-cold RIPA buffer (150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl, pH 8.0) supplemented with 2 mM PMSF and cOmplete Mini EDTA-free protease inhibitor cocktail (Roche). The suspension was homogenized in an ice-cold glass dounce homogenizer, and the cells were disrupted with 15–20 strokes of the pestle. The sample was rotated on an orbital shaker (7 rpm) for 2 h at 4°C. The homogenate was centrifuged for 20 min at 13500 g at 4°C, and the supernatant containing the protein extract was collected, aliquoted to ice-chilled tubes, and stored at -80°C until use. The protein concentration was estimated using the Bradford protein assay. For Western blot, 20 µg of proteins per reaction was resuspended in 15 µl total volume of 1xLaemmli buffer (50 mM TrisHCl pH 6.8, 10% glycerol, 2% SDS, 0.005% bromophenol blue) supplemented with 0.1 M dithiothreitol (DTT). The samples were heated at 37°C for 30 min, and then run on a 4–20% Mini-PROTEAN TGX precast protein gel (Bio-Rad) and transferred onto Amersham Protran

Supported 0.2 µm nitrocellulose membrane (GE Healthcare Life Sciences). For probing, the rabbit preimmune serum or the primary anti-cCENH3 antibody, and the horseradish peroxidase (HRP)-linked goat anti-rabbit IgG antibody (Cell Signaling Technology), were used at 1:500 and 1:2000 dilution, respectively, in TBST buffer (20 mM Tris, 150 mM NaCl, pH 7.6, 0.1% Tween 20) supplemented with 5% BSA. The blots were developed using the Pierce ECL Western Blotting substrate (Thermo Scientific) and Amersham Hyperfilm ECL X-ray films (GE Healthcare Life Sciences).

Cytological preparations

Immunofluorescence (IF), fluorescence in situ hybridization (FISH), and IF-FISH experiments were done on metaphase spreads, interphase nuclei and chromatin fibers isolated from the gonads of male pupae. We optimized, however, different protocols for IF, FISH and IF-FISH chromosome and fiber preparations in order to preserve chromosome/chromatin architecture and the antibody's epitopes, but at the same time allowing the penetration of the DNA probes when FISH performed.

For IF experiments, freshly isolated male gonads were incubated in 10 µg/ml colcemid solution for 1–2 h, then transferred to 0.5% sodium citrate for 5 min. After hypotonic shock, the testes were incubated in pre-warmed 2% paraformaldehyde (PFA) at 37°C for 10 min, then transferred to phosphate-buffered saline (PBS) supplemented with 30 mM glycine for 10 min at RT. After 3x5 min washing in PBS, the tissue was manually disaggregated on a slide in a drop of 0.05 M NaOH, covered with a coverslips and squashed. Slides were then frozen briefly (30–60 s) in liquid nitrogen. Coverslips were removed immediately, and slides were air-dried and stored at -80°C until use. For FISH experiments, freshly dissected testes were incubated in 10 µg/ml colcemid solution (Roche Diagnostics) for 1–2 h at RT. The tissue was subjected to hypotonic shock in 75 mM KCl for 15 min, and then fixed in Clarke's solution (absolute ethanol:acetic acid, 3:1) for at least 30 min. The testes were macerated on a slide in a drop of 45% acetic acid, squashed and frozen as described for IF slide preparation. Air-dried slides were stored at -20°C until use.

For combined IF-FISH analysis, dissected male gonads were incubated in 10 µg/ml colcemid solution for 1–2 h, and subjected to hypotonic shock in 75 mM KCl for 15 min.

The fixation was done in freshly prepared solution of 3.7% PFA, 5% acetic acid and 0.9% sodium chloride at RT for 18 min. After 10 min washing in PBS, the gonads were disaggregated and squashed on a slide in a drop of 0.05 M NaOH, frozen and processed as described for IF slide preparation, being kept at -80°C until use. In all aforementioned cytological preparations, the poly-D-lysine coated slides were used.

The extended chromatin fibers were prepared from dissected testes isolated and washed in PBS supplemented with 1 mM PMSF. Approximately 20 gonads resuspended in 100 µl PBS + 1 mM PMSF were mixed by microtube homogenizer for 30 s, supplemented with 700 µl PBS + 1 mM PMSF, and strained through 100 µm cell strainer. Using Cytospin 4 cytocentrifuge (Shandon, ThermoFisher Scientific), 400 µl of suspension was cytospun for 10 min at 1200 rpm onto coated Shandon Cytoslides (ThermoFisher). After short air-drying, 15 µl of a freshly prepared mild SDS lysis buffer (200 mM Tris-HCl, pH 7.4, 50 mM EDTA, 0.2% SDS, 1 mM PMSF) was added per slide, that was immediately covered with a 18x18 mm square coverslip. After 4–8 min lysis reaction at room temperature, a coverslip was lifted slowly. Air-dried slides were fixed in 2% formaldehyde diluted in PBS for 10 min, and then washed in PBS for 3x5 min. Four drops of Image-iT FX signal enhancer (ThermoFisher Scientific) were applied per slide, and slides were covered by 20x40 mm coverslips and incubated for 30 min at room temperature in a humid chamber. After that, coverslips were removed, and slides were thoroughly rinsed with PBS for 3x5 min, and directly subjected to IF-FISH procedure.

Immunofluorescence (IF) and fluorescence in situ hybridization (FISH)

For IF, slides were rinsed in PBS for 2 min, and then incubated in PBS containing 1% Triton X-100 for 25 min on ice. After 3x5 min washing in PBST (PBS, 0.2% Tween 20), slides were blocked in PBST supplemented with 2.5% BSA and 0.3 M glycine in a moist chamber at 37°C for 1 h. Anti-cCENH3 antibody was diluted 1:400 (for chromosome spreads) or 1:200 (for chromatin fibers) in blocking solution (PBST containing 2.5% BSA), and added to slides which were incubated in a humid chamber overnight at 37°C. Slides were washed 5x5 min with PBST at RT, and then incubated for 1 h at 37°C with Alexa Fluor 594-labeled or Alexa Fluor 488-labeled goat anti-rabbit antibodies (Abcam), diluted 1:1000 in blocking solution. Final washes were performed

4x5 min in PBST, and 1x5 min in PBS at RT. Slides were counterstained in 4′,6-diamidino-2-phenylindole (DAPI) solution for 15 min, then rinsed briefly with distilled water and air-dried. Slides were finally embedded in Mowiol 4–88 mounting medium (Sigma-Aldrich).

For FISH, cCENH3-ChIPped DNA was labeled with Cy3-dCTP (GE Healthcare Life Sciences) using random priming assay. TCAST satellite probe was prepared from the mix of plasmid clones CT8 and CT19 (available upon request) and labelled with biotin-16-dUTP (Roche) by PCR using the primers Tcastan1 (5'-TGTAGGACTAACCATAAGCG-3') and Tcastan2 (5'-CAATGTTTGAGACGAAGACG3'). Slides were pretreated with RNase A and pepsin, post-fixed by 1% formaldehyde and dehydrated in an ice-cold ethanol series. Slides were denatured in 70% formamide at 70°C for 2 min and hybridized with denatured probes (150 ng of a labeled DNA probe in 15 µl of 60% deionized formamide, 8% dextran sulfate, 1.6xSSC, 20 mM sodium phosphate pH 7.0) overnight at 37°C. Posthybridization washes were done at 37°C in 50% formamide. While the Cy3-labeled cCENH3-ChIPped DNA probe was detected directly after posthybridization washes, the TCAST biotin-labeled probe was visualized with fluorescein avidin D and biotinylated anti-avidin D system (Vector Laboratories) by signal amplification through three layers of fluorophore conjugates and antibodies using the following dilutions: 1:500 fluorescein avidin D, 1:100 biotinylated anti-avidin D, 1:2000 fluorescein avidin D. Preparations were counterstained and embedded as described above.

When IF was followed by FISH, RNase A and pepsin pretreatments were omitted.

Microscopy and image analyses

Image acquisition was done using a confocal laser scanning microscope Leica TCS SP8 X (Leica Microsystems) equipped with a HC PL APO CS2 63×/1.40 oil objective, 405 nm diode laser and a supercontinuum excitation laser (Leica Microsystems). Images were captured separately for each fluorochrome and processed with ImageJ [61] and Adobe Photoshop (CS5).

In order to estimate the relative size of *T. castaneum* centromeres, the extent of cCENH3 signals was measured. For quantification of the signal lengths, maximum intensity projections of z-stacked images were analyzed using “Measure” tool in ImageJ with normalization of scale length across different images. The lengths of the chromosomes and associated centromeres were measured separately for each color channel that was converted to grayscale. Due to different degree of chromosome condensation, centromere sizes were examined using centromere to chromosome length ratios. The chromosomes ch2, ch3, ch4, and yp, which can be differentiated from the other chromosomes by their size and centromere position, have been singled out for individual analysis. More than 30 chromosome spreads with accompanied non-overlapping chromosomes were analyzed. Measurement of FISH signals obtained by cCENH3-ChIPped DNA probe was performed on 10 spreads in the same manner described above for the cCENH3 measurement. The average value was calculated from ratios of cCENH3-ChIPped DNA signal length to chromosome length obtained for all chromosomes.

Quantification of cCENH3 and TCAST signal interrelation on extended chromatin fibers was done by examining 15 selected regions where cCENH3 and TCAST showed evident co-localization. The z-stacked images were first adjusted for regions of interest in order to minimize background interference. The same default threshold was then applied in ImageJ on grayscale channels for each image. Manual spline fitted lines were drawn on chromatin fibers and “Plot profile” tool was used to obtain values of pixel intensities along the selected line. Presence or absence of each signal was evaluated by extracting values that were higher or equal to 0, respectively, followed by comparison of the degree of two signals’ overlap. The scatter plots were done in GraphPad Prism version 8.

Native chromatin isolation and chromatin immunoprecipitation (ChIP)

750 mg of snap-frozen adult insects were ground to a fine powder and resuspended in 10 ml of ice-cold chromatin isolation buffer, CIB (15 mM Tris HCl pH = 7.5, 60 mM KCl, 15 mM NaCl, 0.34 M sucrose, 0.15 mM spermine, 0.5 mM spermidine) supplemented with cOmplete Mini EDTA-free protease inhibitor cocktail (Roche), 2 mM PMSF and 0.5% Triton X-100. The suspension was homogenized in an ice-cold glass dounce

homogenizer and strained through 40 µm cell strainer. The isolated nuclei were centrifuged at 2500 g for 5 min at 4°C, rewashed in CIB, centrifuged again, and resuspended in 1.5 ml CIB supplemented with 1 mM CaCl2 and 4 mM MgCl2. The chromatin was digested with 200 U MNase (Thermo Scientific) for 30 min at 37°C, and the reaction was stopped by adding EDTA to a final concentration of 10 mM. After centrifugation at 8000 g for 5 min at 4°C, the pellet containing chromatin was resuspended in 1.2 ml of PBS supplemented with 0.5 M NaCl. After rotation on an orbital shaker (7 rpm) for 2 h at 4°C, the sample was centrifuged for 15 min at 15000 g at 4°C. The supernatant containing digested chromatin of preferentially mononucleosomal length was saved, and kept at -80°C until use.

For ChIP assay, first we tested the two commercial ChIP kits, Pierce Magnetic ChIP Kit (Thermo Scientific) and Dynabeads Protein A Immunoprecipitation Kit (Invitrogen), but with them we obtained unspecific immunoprecipitation presumably due to insufficiently stringent washing with commercial washing buffers. Therefore, using Dynabeads Protein A, we modified the protocol as follows. In short, Dynabeads were pre-blocked with BSA, and soluble chromatin was pre-cleared with pre-blocked Dynabeads in ChIP buffer (16.7 mM Tris-HCl pH 8.1, 1.2 mM EDTA, 0.167 M NaCl, 0.01% SDS, 1.1% Triton X-100) supplemented with cOmplete Mini EDTA-free protease inhibitor cocktail (Roche) and 2 mM PMSF, by rotation on an orbital shaker (7 rpm) for 1 h at 4°C. Pre-cleared chromatin was separated on a magnetic rack, and 5% was reserved as an input control sample. In ChIP experiment we tested two different anti-cCENH3 antibody amounts per reaction: 5 µg (as a typical amount) and 8 µg (as a Dynabeads Protein A maximum binding capacity). 30 µl of pre-blocked Dynabeads per reaction were mixed with 5 µg / 8 µg anti-cCENH3 antibody diluted in 120 µl Ab Binding & Washing Buffer (Invitrogen), and antibody binding was done by rotation for 30 min at 4°C. As negative controls, 5 µg of normal rabbit IgG (Cell Signaling Technology) and a mock control (a sample without antibody) were included in each ChIP experiment. The Dynabeads-antibody complexes were separated on a magnetic rack, and 230 µl of pre-cleared chromatin (~4 µg) was added. After 2-hour incubation at 4°C with rotation, the precipitated immunocomplexes were washed in a series of buffers (5 min each wash): 3 times low salt buffer (0.15 M NaCl, 20 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0, 0.1% SDS, 1% Triton X-100), 3 times high salt buffer (0.5 M NaCl, 20 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0, 0.1% SDS, 1% Triton X-

100), once LiCl buffer (0.25 M LiCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 1% sodium deoxycholate, 0.1% Triton X-100), twice TE buffer pH 8.0. Chromatin was eluted from the Dynabeads with 2x100 µl of elution buffer (1%SDS, 0.1M NaHCO3) after 2x15 min incubation at 55°C. After RNase A / proteinase K treatment, DNA from ChIP and input samples was extracted using QIAquick PCR Purification Kit (Qiagen). In S6 Appendix we showed the comparison of the subsequent ChIP-Seq analyses using 5 µg and 8 µg anti-cCENH3 antibody in ChIP experiments, as well as the comparison between ChIP experiments performed by using commercial washing buffers and prepared low/high salt buffers. Unlike washing with prepared low/high salt buffers, using commercial washing buffers did not yield any difference between cCEN3-ChIP and input samples, and we concluded that the washing with commercial buffers was not stringent enough. ChIP experiments using different amounts of anti-cCENH3 antibody produced congruent results, and the data for the ChIP experiment with 5 µg anti-cCENH3 were used in the following analysis.

ChIP-seq data analysis

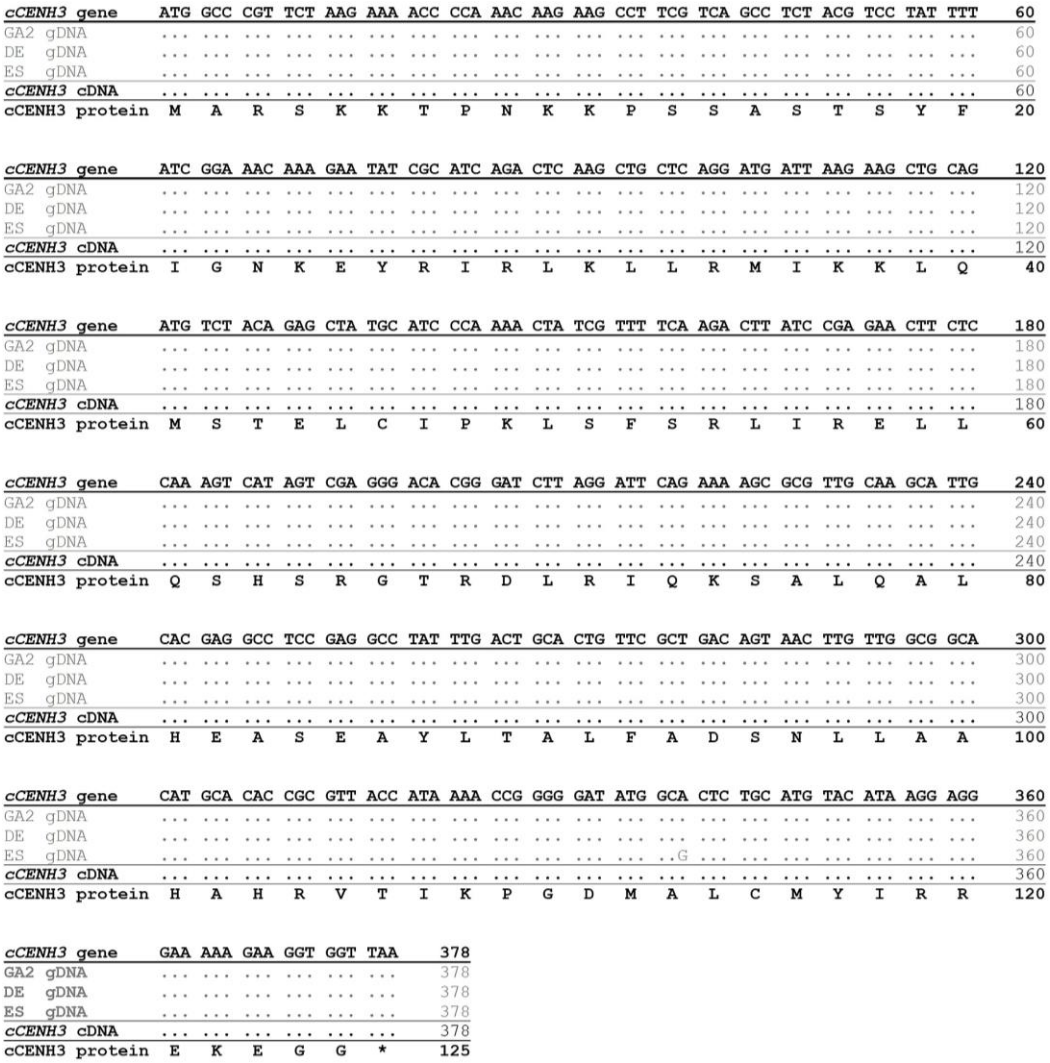
Following the strategy described by Neumann and coauthors [7], we evaluated the enrichment of repetitive sequences in the ChIP-seq data by using the ChIP-Seq Mapper tool at the Galaxy web-server (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/). The workflow of the ChIP-Seq analysis is shown in S5 Fig. The ChIP-Seq Mapper performs BLASTN similarity search of ChIP reads and control input reads (DNA isolated from the chromatin aliquot prior to ChIP experiment) against repeats' database used as a reference. Repetitive DNA reference database is obtained by the RepeatExplorer2 tool that performs similarity-based clustering and groups short unassembled WGS reads into repeat clusters [31]. The ChIP enrichment for individual WGS repeat clusters is calculated based on ChIP/Input hit ratio. Sequencing of cCENH3-ChIPped DNA and input DNA control, as well as *T. castaneum* whole genome sequencing (WGS) was performed on the Illumina HiSeq2500 platform (Admera Health, USA) resulting in 10904694, 9232324 and 9664142 paired-end reads, respectively. By RepeatExplorer2 tools available at Galaxy web-server (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/), FASTQ reads were preprocessed using cutadapt filtering, quality-filtering (95% of bases equal to or higher than the quality cutoff value of 10), trimming to 151 nt, interlacing, and random sampling

(different random number generator seeds tested) to scale down large data sets. To generate repetitive DNA reference database, from 9290920 preprocessed WGS reads, 270200 paired-end reads were randomly selected to ensure low genome coverage (0.2x). Repeat identification by graph-based sequence clustering was performed using RepeatExplorer2 pipeline [31] that recommends low-pass genome sequencing corresponding to 0.01–0.50x genome coverage because in genome skimming only the reads derived from the repetitive regions can produce multiple similarity hits and group into the clusters of frequently overlapping sequences. Also, the repetitiveness of the analyzed genome limits the number of reads that can be analyzed with this pipeline. For highly repetitive genome of *T. castaneum*, 0.2x genome coverage by 270200 randomly selected reads was determined as an optimal sample size to be processed. RepeatExplorer2 analysis resulted in 21564 WGS clusters (Fig 5), ranked according their genome proportion. Although the genome proportions for cluster CL167 and onwards were estimated lower than 0.01%, in the ChIP-Seq Mapper analysis we examined the first 1000 clusters. From 6623300 and 3828990 preprocessed reads obtained for cCENH3-ChIPped DNA and input DNA, respectively, for each data set three random sampling subsets were generated comprising 1000000, 500000 or 250000 reads. We performed ChIP-Seq Mapper analysis for the top 1000 WGS repeat clusters testing different bit score thresholds (30, 90, 150). As the results of ChIP-Seq Mapper analysis from three different randomly sampled subsets of ChIP/Input reads were consistent (S7 Appendix), only the data for the 1000000 reads analysis were presented. Using Geneious R11.1.4 software (Biomatters, Ltd.), the sequences of cCENH3-ChIP-enriched clusters were mapped against the current *T. castaneum* genome assembly Tcas5.2 [24], comprising ten chromosome/linkage groups (LG) and the unassembled sequence represented as 305 unplaced scaffolds (GenBank accessions DS497665-DS497969) and 1848 unplaced singletons (GenBank accessions GG694051-GG695898). Enriched clusters were also subjected to BLAST search against NCBI GenBank database [62], using different BLAST algorithms (megablast, discontiguous megablast, and blastn). CENSOR tool was used to search GIRI Repbase [63].

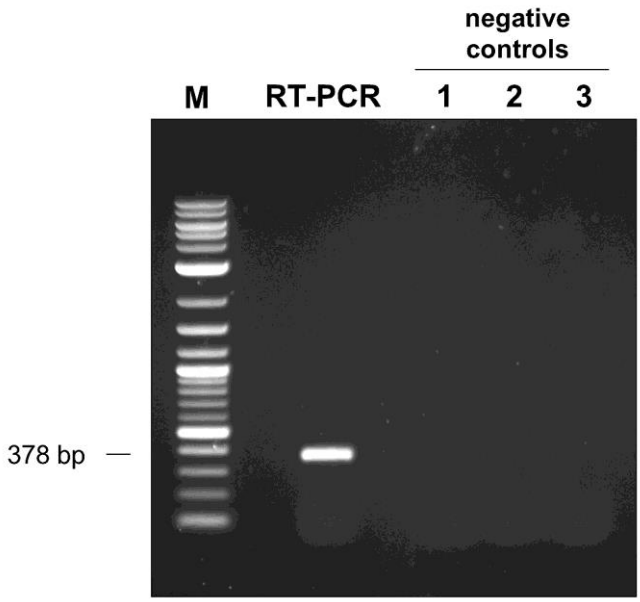
Accession codes

The cCENH3 nucleotide sequence has been deposited in NCBI GenBank under the accession number MT043459. Raw Illumina reads from ChIP-seq experiment have been deposited in the Sequence Read Archive under the study accession number PRJNA606031.

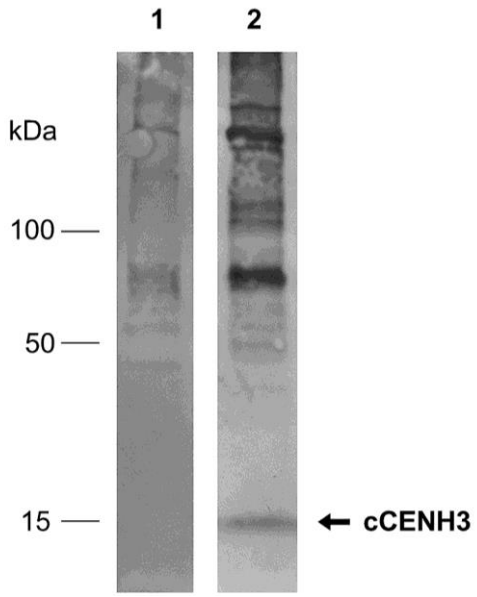
Supporting information



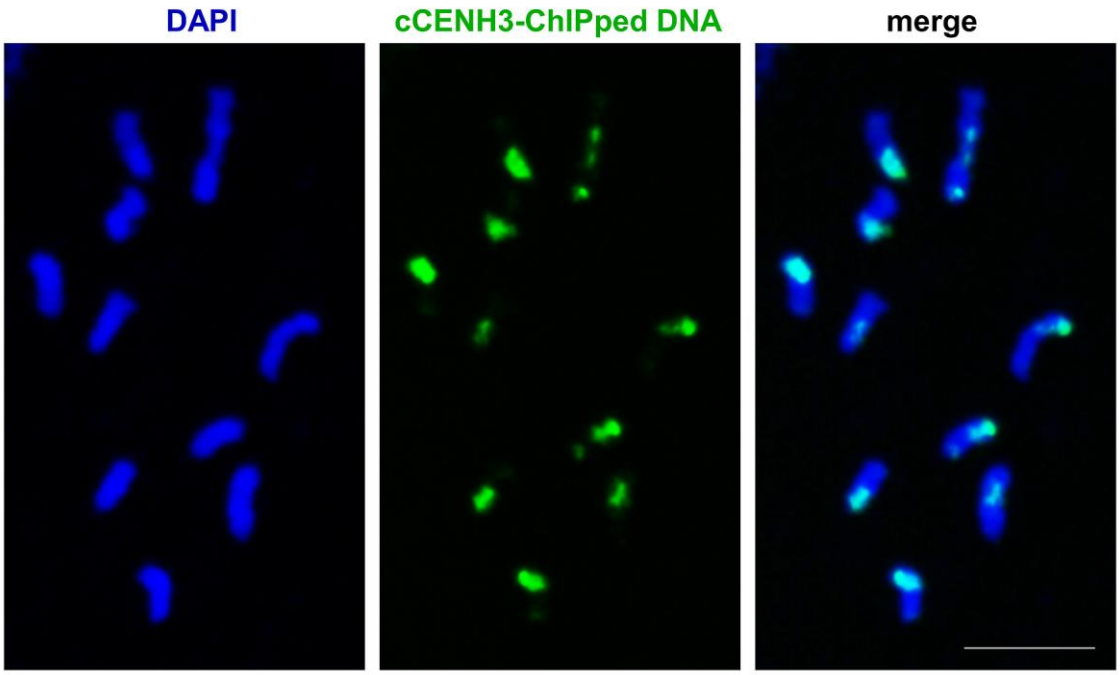
S1 Fig. Alignment of cCENH3 coding sequence and cCENH3 cDNA. 378 bp long cCENH3 coding sequence (gDNA) was amplified from genomic DNAs isolated from three different *T. castaneum* strains (GA2, DE, and ES). Complementary DNA (cDNA) was synthesized from cCENH3 transcript by RT-PCR using primers specific for cCENH3 gene. Identical nucleotides are indicated by a dot, and only one synonymous substitution is present in ES strain. The cCENH3 protein translation is presented below the nucleotide alignment.



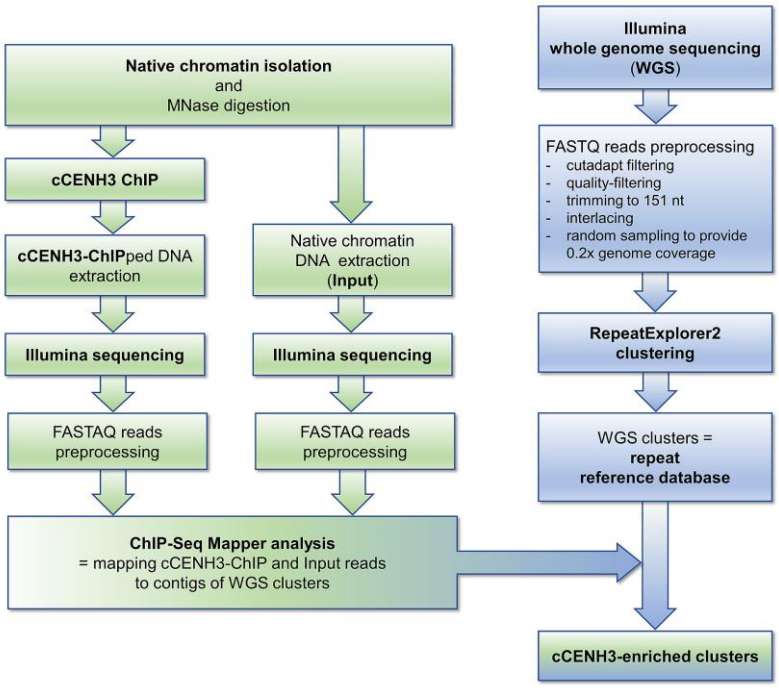
S2 Fig. Agarose gel electrophoresis of RT-PCR amplification of cCENH3 mRNA. Reverse transcription PCR amplification was done on total RNA isolated from *T. castaneum* larvae by using the primers specific for the cCENH3 gene. The three negative controls included reactions: (1) without RT step, (2) with template RNA added after RT step, (3) RT-PCR without template RNA. M lane represents a 100-bp-size marker ladder.



S3 Fig. Western blot of the cCENH3 antibody on *Tribolium castaneum* protein lysate. *T. castaneum* whole protein extract was fractionated by SDS-PAGE under denaturing conditions and transferred to nitrocellulose membrane. The membrane was subjected to Western blot analysis. (1) Testing of the rabbit preimmune serum. (2) Testing of the monospecific IgG fraction purified from the rabbit immunoserum after 120 days of immunization with the cCENH3-specific peptide (NH2-RSKKTPNKKPSSASTSYF-CONH2) revealed a ~15 kDa signal which is consistent with the expected molecular weight of cCENH3 protein. Molecular weights of protein sizes are indicated in kDa.



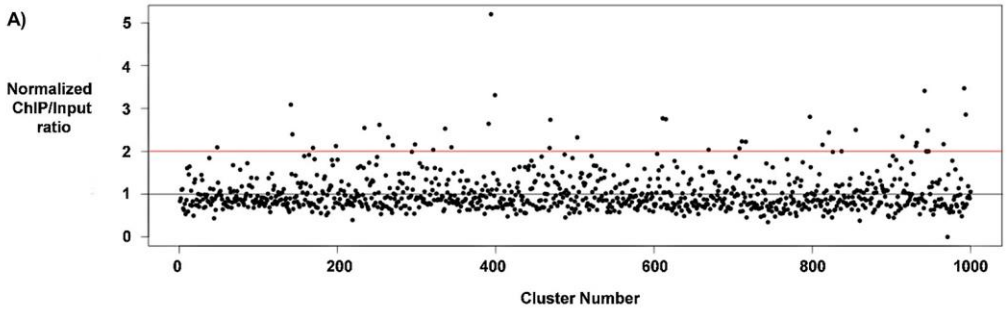
S4 Fig. Fluorescence in situ hybridization with cCENH3-ChIPped DNA. DNA immunoprecipitated by using cCENH3 antibody (cCENH3-ChIPped DNA) was Cy3-labelled and hybridized to *T. castaneum* chromosomes. cCENH3-ChIPped DNA (pseudocolored in green) hybridizes to the centromeric regions of all chromosomes (counterstained in DAPI). Scale bar = 5 μm.



S5 Fig. The workflow of cCENH3-ChIP-seq analysis. DNA sequences enriched for cCENH3 were identified following the strategy introduced by Neumann et al. [7]. First, the repetitive DNA reference database is formed by low-pass Illumina sequencing and RepeatExplorer2 similarity-based clustering of WGS unassembled reads. Chromatin immunoprecipitation (ChIP) was performed using the cCENH3

antibody. Immunoprecipitated DNA was Illumina sequenced, as well as DNA obtained from the chromatin preparation prior to ChIP (Input). ChIP and Input reads were mapped to the WGS clusters, and the cCENH3-enriched clusters were determined based on ChIP/Input reads elevated ratio.

A)



B)

Cluster	ChIP Hits	Input Hits	Normalized ChIP/Input ratio
48	10565	5052	2.1
141	488	158	3.1
143	460	192	2.4
169	495	238	2.1
198	210	99	2.1
234	402	158	2.5
253	288	110	2.6
264	337	145	2.3
270	732	342	2.1
298	296	137	2.2
321	61	30	2.0
336	177	70	2.5
344	67	32	2.1
391	1591	602	2.6
394	442	85	5.2
399	202	61	3.3
468	54	26	2.1
469	145	53	2.7
503	79	34	2.3
611	216	78	2.8
615	44	16	2.8
669	118	58	2.0
708	31	15	2.1
711	174	78	2.2
716	60	27	2.2
797	115	41	2.8
813	99	46	2.2
821	61	25	2.4
855	10	4	2.5
914	75	32	2.3
931	17	8	2.1
932	66	30	2.2
942	133	39	3.4
946	97	39	2.5
966	13	6	2.2
992	118	34	3.5
994	20	7	2.9

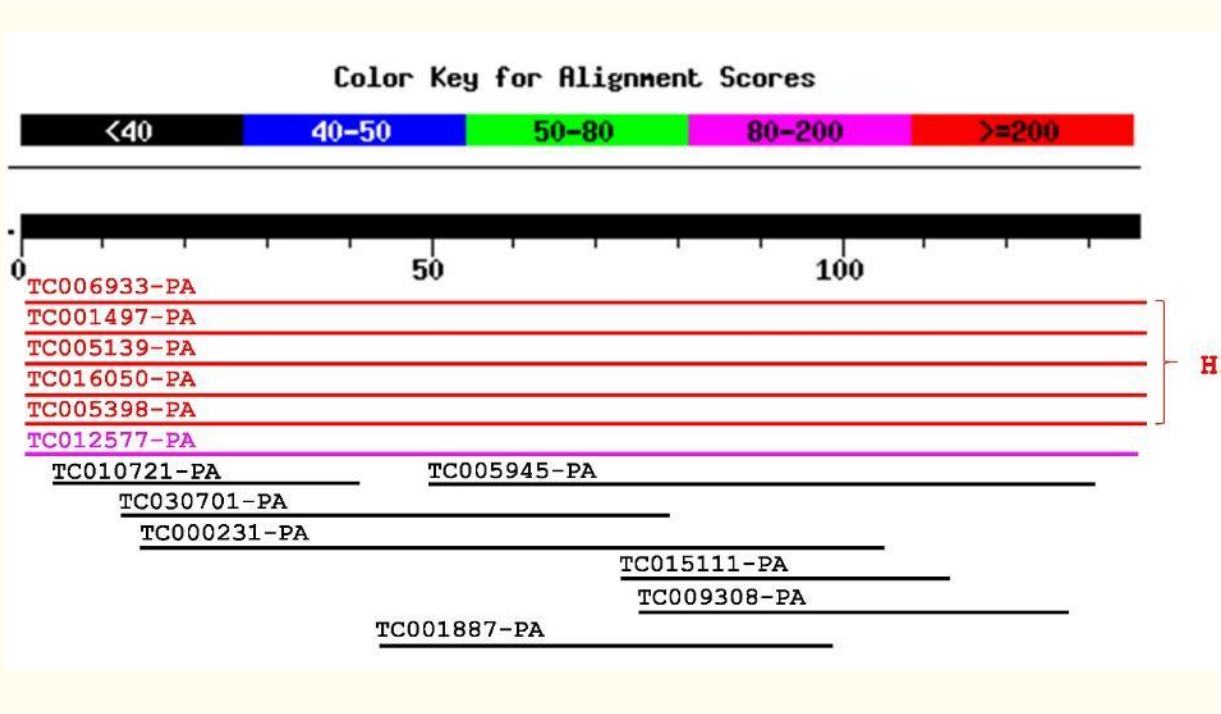
S6 Fig. cCENH3-ChIP-Seq Mapper analysis. ChIP-Seq Mapper analysis based on one million cCENH3-ChIP and one million Input Illumina reads mapped to the top 1000 WGS *T. castaneum* repeat clusters obtained by RepeatExplorer2 analysis. (A) ChIP-Seq Mapper plot for the top 1000 WGS *T. castaneum* clusters analyzed for cCENH3 enrichment. The red line marks the mean ratio between ChIP and Input hits, and the clusters above red line show >2-fold enrichment for cCENH3. (B) List of 37 out of the top 1000 *T. castaneum* WGS clusters showing cCENH3-ChIP/Input ratio >2.

S1 Table. Estimation of *Tribolium castaneum* centromere sizes.

Chromosome	Chromosome size [Mbp][1]	Number of measured chromosomes	Centromere proportion [%][2*]	Estimated centromere size [Mbp]*
all	204	240	43.47 (±8.68)	88.68 (±17.71)
ch2	22.42	15	42.89 (±6.97)	9.62 (±1.56)
ch3	37.41	31	42.34 (±7.92)	15.84 (±2.96)
ch4	22.52	22	46.48 (±5.99)	10.47 (±1.35)
y_p	5.71	8	40.60 (±8.09)	2.31 (±0.46)

[1] The chromosome sizes were calculated from the relative length of the individual chromosomes [30] and haploid genome size of 204 Mb [23].
[2] Centromere proportions were estimated from the ratio of the cCENH3 signal length demarcated by the outermost cCENH3-containing regions to the chromosome length. The quantifications are presented in the S3 Appendix.
* Standard deviations are indicated in brackets.

S1 Appendix. BLASTP search against the *Tribolium castaneum* OGS3 database using histone H3 protein sequence as a query. Distribution of 13 Blast hits on the H3 query sequence is shown by the graphic alignment. The sequences in red represent H3 matches, while the potential CenH3 candidate (TC012577-PA) is marked in magenta. The matches in black represent partial and less significant hits. The alignments between H3 and all 13 matches are listed below, including the alignment score and expected values, and percentage of query/subject identities and positives.



```
Sequences producing alignments:                    Score(bits)  E Value

TC006933-PA (Unknown:14570956..14571366)              264         7e-72
TC001497-PA (Unknown:29373457..29373867)              264         7e-72
TC005139-PA (Unknown:23334197..23334607)              264         7e-72
TC016050-PA (Unknown:36741348..36741758)              264         7e-72
TC005398-PA (ChLG8:17239546..17240958)                258         4e-70
TC012577-PA (ChLG9:16066895..16067272)                 89         4e-19
TC005945-PA (ChLG8:5643896..5650497)                   28         0.80
TC030701-PA (ChLG3:35759566..35834262)                 26         4.0
TC009308-PA (ChLG7:1817276..1819114)                   25         6.7
TC015111-PA (ChLG6:5039228..5049650)                   25         6.7
TC000231-PA (ChLG2:18781787..18783630)                 25         6.7
TC010721-PA (Unknown:11584977..11586401)               25         8.8
TC001887-PA (Unknown:10737145..10747799)               25         8.8


>TC006933-PA (Unknown:14570956..14571366)
          Length = 136

 Score =  264 bits (675), Expect = 7e-72
 Identities = 136/136 (100%), Positives = 136/136 (100%)

[GBrowse][Subject FASTA]
Query: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
           MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE
Sbjct: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
Query: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
           LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI
Sbjct: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
Query: 121 MPKDIQLARRIRGERA 136
           MPKDIQLARRIRGERA
Sbjct: 121 MPKDIQLARRIRGERA 136


>TC001497-PA (Unknown:29373457..29373867)
          Length = 136

 Score =  264 bits (675), Expect = 7e-72
 Identities = 136/136 (100%), Positives = 136/136 (100%)

[GBrowse][Subject FASTA]
Query: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
           MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE
Sbjct: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
Query: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
           LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI
Sbjct: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
Query: 121 MPKDIQLARRIRGERA 136
           MPKDIQLARRIRGERA
Sbjct: 121 MPKDIQLARRIRGERA 136


>TC005139-PA (Unknown:23334197..23334607)
          Length = 136

 Score =  264 bits (675), Expect = 7e-72
 Identities = 136/136 (100%), Positives = 136/136 (100%)

[GBrowse][Subject FASTA]
Query: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
           MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE
Sbjct: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
Query: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
           LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI
Sbjct: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
Query: 121 MPKDIQLARRIRGERA 136
```

```
           MPKDIQLARRIRGERA
Sbjct: 121 MPKDIQLARRIRGERA 136


>TC016050-PA (Unknown:36741348..36741758)
          Length = 136

 Score =  264 bits (675), Expect = 7e-72
 Identities = 136/136 (100%), Positives = 136/136 (100%)

[GBrowse][Subject FASTA]
Query: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
           MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE
Sbjct: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
Query: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
           LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI
Sbjct: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
Query: 121 MPKDIQLARRIRGERA 136
           MPKDIQLARRIRGERA
Sbjct: 121 MPKDIQLARRIRGERA 136


>TC005398-PA (ChLG8:17239546..17240958)
          Length = 136

 Score =  258 bits (660), Expect = 4e-70
 Identities = 132/136 (97%), Positives = 135/136 (99%)

[GBrowse][Subject FASTA]
Query: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
           MARTKQTARKSTGGKAPRKQLATKAARKSAP+TGGVKKPHRYRPGTVALREIRRYQKSTE
Sbjct: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYQKSTE 60
Query: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
           LLIRKLPFQRLVREIAQDFKTDLRFQS+A+ ALQEASEAYLVGLFEDTNLCAIHAKRVTI
Sbjct: 61  LLIRKLPFQRLVREIAQDFKTDLRFQSAAIGALQEASEAYLVGLFEDTNLCAIHAKRVTI 120
Query: 121 MPKDIQLARRIRGERA 136
           MPKDIQLARRIRGERA
Sbjct: 121 MPKDIQLARRIRGERA 136


>TC012577-PA (ChLG9:16066895..16067272)
          Length = 125

 Score = 89.4 bits (220), Expect = 4e-19
 Identities = 59/138 (42%), Positives = 71/138 (51%), Gaps = 19/138 (13%)

[GBrowse][Subject FASTA]
Query: 1   MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE 60
           MAR+K+T  K     +    +  K                 YR     LR I++ Q STE
Sbjct: 1   MARSKKTPNKKPSSASTSYFIGNK----------------EYRIRLKLLRMIKKLQMSTE 44
Query: 61  LLIRKLPFQRLVREIAQDFKT---DLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKR 117
           L I KL F RL+RE+ Q       DLR Q SA+ AL EASEAYL  LF D+NL A HA R
Sbjct: 45  LCIPKLSFSRLIRELLQSHSRGTRDLRIQKSALQALHEASEAYLTALFADSNLLAAHAHR 104
Query: 118 VTIMPKDIQLARRIRGER 135
           VTI P D+ L   IR E+
Sbjct: 105 VTIKPGDMALCMYIRREK 122


>TC005945-PA (ChLG8:5643896..5650497)
          Length = 769

 Score = 28.5 bits (62), Expect = 0.80
 Identities = 23/90 (25%), Positives = 44/90 (48%), Gaps = 14/90 (15%)
```

```
[GBrowse][Subject FASTA]
Query: 51  EIRRYQKSTELLIRKLPFQRLVREIAQDFKTDL--------RFQSSAVMALQEASEAYLV 102
           E + Y K+ E+L  K+PF+R+ + + QD    +        RF     + +Q  +  + +
Sbjct: 210 EEKNYVKALEVL--KIPFKRIPKRVTQDLVNMMLELLLLTDRFTECLDIFIQHCNFTFEI 267
Query: 103 GLFEDTNLCAIHAKRVTIMPKDIQLARRIR 132
            + ED N  AI +    +MP  IQ+  +++
Sbjct: 268 TVNED-NTIAIDS---YVMPPSIQIDLKVK 293


>TC030701-PA (ChLG3:35759566..35834262)
          Length = 18024

 Score = 26.2 bits (56), Expect = 4.0
 Identities = 16/68 (23%), Positives = 32/68 (47%)

[GBrowse][Subject FASTA]
Query: 14    GKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVR 73
              GK P+K    + A K   A   VK+       T  ++EI   ++  ++   K  F++  +
Sbjct: 15978 GKVPKKPEEEEEAVKLKKAPERVKEEELPEAKTPKVKEIEEKEEPEKIEFEKPEFEKPEK 16037
Query: 74    EIAQDFKT 81
              E+ ++ K+
Sbjct: 16038 EVPEEEKS 16045


>TC009308-PA (ChLG7:1817276..1819114)
          Length = 518

 Score = 25.4 bits (54), Expect = 6.7
 Identities = 19/55 (34%), Positives = 27/55 (49%), Gaps = 1/55 (1%)

[GBrowse][Subject FASTA]
Query: 82  DLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLA-RRIRGER 135
            DLR +   V  LQ+  E+  V L ++  L    +K +     DIQL  +RIR  R
Sbjct: 326 DLRQKDQLVKHLQQEIESLKVELMKEKCLSENKSKELLEKSDDIQLLNKRIRSSR 380


>TC015111-PA (ChLG6:5039228..5049650)
          Length = 509

 Score = 25.4 bits (54), Expect = 6.7
 Identities = 13/42 (30%), Positives = 21/42 (50%)

[GBrowse][Subject FASTA]
Query: 71  LVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCA 112
            +VR+   D    +   S  V   QE +EA++  L+  T +CA
Sbjct: 223 IVRQDGLDLPIGIYRPSIVVSTYQEPTEAWINNLYGPTGVCA 264


>TC000231-PA (ChLG2:18781787..18783630)
          Length = 424

 Score = 25.4 bits (54), Expect = 6.7
 Identities = 22/92 (23%), Positives = 37/92 (40%), Gaps = 5/92 (5%)

[GBrowse][Subject FASTA]
Query: 15  KAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVRE 74
           + P+   A      S  A G   KP   RP  V L+ + +Y     +  ++LP+   V
Sbjct: 43  RQPKASTAHMVKSFSRSAAGKQIKPETLRPPQVLLKTV-KYLLCDVINTKRLPYWHSV-- 99
Query: 75  IAQDFKTDLRFQSSAVMALQEASEAYLVGLFE 106
              DF TD        + +Q  S+A  + + +
Sbjct: 100 --YDFITDRLLAVRQDLVVQNVSKAESITILQ 129
```

```
>TC010721-PA (Unknown:11584977..11586401)
          Length = 451

 Score = 25.0 bits (53), Expect = 8.8
 Identities = 12/38 (31%), Positives = 18/38 (47%)

[GBrowse][Subject FASTA]
Query: 3   RTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPH 40
           RT  T+ ++T  K P +Q A   +  + P T     PH
Sbjct: 389 RTPLTSPQTTSLKIPSRQTAISPSTPATPTTSRTPFPH 426


>TC001887-PA (Unknown:10737145..10747799)
          Length = 1264

 Score = 25.0 bits (53), Expect = 8.8
 Identities = 12/57 (21%), Positives = 30/57 (52%)

[GBrowse][Subject FASTA]
Query: 42  YRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASE 98
           Y   +  L+E++R      + + +     L+R++A++ +  + F+  AV  + E++E
Sbjct: 56  YGKMSTRLKELQRKYHDENIKVARKDGLLLIRDMAKEVQNMMSFKIEAVRRITESAE 112


  Database: Official_Gene_Proteins.fa
    Posted date:  Jun 15, 2011  2:12 PM
  Number of letters in database: 7,426,099
  Number of sequences in database:  16,531

Lambda     K      H
   0.321    0.131    0.361

Gapped
Lambda     K      H
   0.267   0.0410    0.140

Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 1,669,841
Number of Sequences: 16531
Number of extensions: 57156
Number of successful extensions: 177
Number of sequences better than 10.0: 13
Number of HSP's better than 10.0 without gapping: 10
Number of HSP's successfully gapped in prelim test: 3
Number of HSP's that attempted gapping in prelim test: 162
Number of HSP's gapped (non-prelim): 17
length of query: 136
length of database: 7,426,099
effective HSP length: 86
effective length of query: 50
effective length of database: 6,004,433
effective search space: 300221650
effective search space used: 300221650
T: 11
A: 40
X1: 16 ( 7.4 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.9 bits)
S2: 53 (25.0 bits)
```

S2 Appendix. RNA-seq data on cCENH3 expression in different life stages and tissues of *Tribolium castaneum*. Normalized expression of cCENH3 gene in *T. castaneum* embryos (0-5 hr and 6-11 hr), ovary, testis, female and male carcasses is based on publicly available RNA-seq datasets from Khan et al. [29]. Hits were normalized with CPM method [(counts per million reads mapped) = (hit number/library size)*106]. Average values were calculated for separate runs and replicates.

Run accession	Read number	Life stage	Replicate	Run	cCENH3 hits	CPM (counts per million)
SRR7814724	8212693	embryo 0-5 hr	S1	1	541	65,87
SRR7814725	7818713	embryo 0-5 hr	S1	2	508	64,97
SRR7814726	8334917	embryo 0-5 hr	S1	3	519	62,27
SRR7814727	7919481	embryo 0-5 hr	S1	4	480	60,61
SRR7814728	8586948	embryo 0-5 hr	S2	1	510	59,39
SRR7814729	8164411	embryo 0-5 hr	S2	2	512	62,71
SRR7814730	8716601	embryo 0-5 hr	S2	3	584	67,00
SRR7814731	8276024	embryo 0-5 hr	S2	4	545	65,85
SRR7814732	8734144	embryo 0-5 hr	S3	1	467	53,47
SRR7814733	8322886	embryo 0-5 hr	S3	2	468	56,23
SRR7814734	8874840	embryo 0-5 hr	S3	3	488	54,99
SRR7814735	8415091	embryo 0-5 hr	S3	4	490	58,23
SRR7814736	7931572	embryo 6-11 hr	S4	1	539	67,96
SRR7814737	7564299	embryo 6-11 hr	S4	2	533	70,46
SRR7814738	8057008	embryo 6-11 hr	S4	3	563	69,88
SRR7814739	7664197	embryo 6-11 hr	S4	4	559	72,94
SRR7814740	8221210	embryo 6-11 hr	S5	1	604	73,47
SRR7814741	7829229	embryo 6-11 hr	S5	2	564	72,04
SRR7814742	8346068	embryo 6-11 hr	S5	3	605	72,49
SRR7814743	7937478	embryo 6-11 hr	S5	4	556	70,05
SRR7814744	5998859	embryo 6-11 hr	S6	1	462	77,01
SRR7814745	5693470	embryo 6-11 hr	S6	2	425	74,65
SRR7814746	6087320	embryo 6-11 hr	S6	3	436	71,62
SRR7814747	5772177	embryo 6-11 hr	S6	4	431	74,67
SRR7814748	8094525	female carcass	S7	1	9	1,11
SRR7814749	7688288	female carcass	S7	2	13	1,69
SRR7814750	8204746	female carcass	S7	3	10	1,22
SRR7814751	7776066	female carcass	S7	4	16	2,06
SRR7814752	8096717	female carcass	S8	1	5	0,62
SRR7814753	7705939	female carcass	S8	2	9	1,17
SRR7814754	8206133	female carcass	S8	3	13	1,58
SRR7814755	7781236	female carcass	S8	4	11	1,41
SRR7814756	8303056	female carcass	S9	1	9	1,08
SRR7814757	7896102	female carcass	S9	2	8	1,01
SRR7814758	8429008	female carcass	S9	3	14	1,66
SRR7814759	7996299	female carcass	S9	4	5	0,63
SRR7814760	8089681	male carcass	S10	1	9	1,11
SRR7814761	7687479	male carcass	S10	2	10	1,30
SRR7814762	8204191	male carcass	S10	3	8	0,98
SRR7814763	7781451	male carcass	S10	4	10	1,29
SRR7814764	8417967	male carcass	S11	1	11	1,31
SRR7814765	7981745	male carcass	S11	2	11	1,38
SRR7814766	8526966	male carcass	S11	3	8	0,94
SRR7814767	8075638	male carcass	S11	4	18	2,23
SRR7814768	8157086	male carcass	S12	1	8	0,98
SRR7814769	7766807	male carcass	S12	2	6	0,77
SRR7814770	8287846	male carcass	S12	3	9	1,09
SRR7814771	7860727	male carcass	S12	4	4	0,51
SRR7814772	8417428	ovary	S13	1	247	29,34
SRR7814773	7992090	ovary	S13	2	224	28,03
SRR7814774	8533917	ovary	S13	3	228	26,72
SRR7814775	8089904	ovary	S13	4	243	30,04
SRR7814776	10118373	ovary	S14	1	276	27,28
SRR7814777	9611048	ovary	S14	2	239	24,87
SRR7814778	10270646	ovary	S14	3	277	26,97
SRR7814779	9721794	ovary	S14	4	242	24,89
SRR7814780	7762183	ovary	S15	1	169	21,77
SRR7814781	7365673	ovary	S15	2	148	20,09

SRR7814782	7878986	ovary	S15	3	161	20,43
SRR7814783	7459026	ovary	S15	4	154	20,65
SRR7814784	7380882	testis	S16	1	123	16,66
SRR7814785	7005021	testis	S16	2	148	21,13
SRR7814786	7474868	testis	S16	3	156	20,87
SRR7814787	7086993	testis	S16	4	152	21,45
SRR7814788	7590852	testis	S17	1	39	5,14
SRR7814789	7224330	testis	S17	2	44	6,09
SRR7814790	7659461	testis	S17	3	48	6,27
SRR7814791	7295975	testis	S17	4	64	8,77

Life stage	Replicate	Run average CPM
embryo 0-5 hr	S1	63,43
	S2	63,74
	S3	55,73
embryo 6-11 hr	S4	70,31
	S5	72,01
	S6	74,49
female carcass	S7	1,52
	S8	1,20
	S9	1,10
male carcass	S10	1,17
	S11	1,46
	S12	0,84
ovary	S13	28,53
	S14	26,00
	S15	20,74
testis	S16	20,03
	S17	6,57

	embryo 0-5 hr	embryo 6-11 hr	ovary	testis	female carcass	male carcass
Replicate average CPM	60,97	72,27	25,09	13,30	1,27	1,16
stdev	4,54	2,10	3,98	9,52	0,22	0,31

S3 Appendix. Related to Fig 4. Source data for *Tribolium castaneum* centromere size estimation.

	all chromosomes measured		
chromosome number	chromosome length (µm)	cCENH3 region length (µm)	cCENH3/chr length
1	4,588	2,206	0,4808
2	7,380	3,262	0,4420
3	4,630	2,233	0,4823
4	3,409	1,789	0,5248
5	5,309	1,746	0,3289
6	3,610	2,095	0,5803
7	4,337	2,272	0,5239
8	4,492	2,465	0,5488
9	4,131	2,079	0,5033
10	4,102	1,850	0,4510
11	1,768	0,877	0,4960
12	8,509	3,683	0,4328
13	2,593	1,879	0,7246
14	3,855	2,493	0,6467
15	7,232	2,766	0,3825
16	5,270	2,272	0,4311
17	4,235	2,207	0,5211
18	5,214	2,328	0,4465
19	5,325	2,280	0,4282
20	2,824	0,912	0,3229
21	6,101	2,316	0,3796
22	3,576	1,648	0,4609
23	4,177	1,836	0,4395
24	6,396	2,413	0,3773
25	5,829	2,353	0,4037
26	4,353	1,788	0,4108
27	6,455	2,956	0,4579
28	5,084	3,140	0,6176
29	2,325	1,508	0,6486
30	4,878	2,548	0,5223
31	3,983	2,001	0,5024
32	5,433	2,986	0,5496
33	9,034	4,933	0,5460
34	2,866	1,690	0,5897
35	2,208	1,170	0,5299
36	6,360	3,285	0,5165
37	3,812	1,405	0,3686
38	4,408	1,511	0,3428
39	3,755	1,278	0,3403
40	3,457	1,416	0,4096
41	3,392	1,884	0,5554
42	5,016	1,795	0,3579
43	4,579	0,855	0,1867
44	4,521	1,859	0,4112
45	2,102	1,143	0,5438
46	4,161	1,777	0,4271
47	3,607	1,614	0,4475
48	1,563	0,551	0,3525
49	1,841	1,243	0,6752
50	2,244	0,881	0,3926
51	3,577	1,908	0,5334
52	3,178	1,616	0,5085
53	6,991	1,924	0,2752
54	7,413	3,209	0,4329
55	6,063	2,558	0,4219
56	4,282	1,977	0,4617
57	4,537	1,814	0,3998
58	6,761	3,223	0,4767
59	4,068	2,306	0,5669
60	4,713	1,878	0,3985
61	3,941	1,874	0,4755
62	4,233	1,977	0,4670
63	5,574	3,144	0,5640
64	5,578	3,049	0,5466
65	1,282	0,568	0,4431
66	5,756	2,802	0,4868
67	4,888	2,153	0,4405
68	3,553	1,951	0,5491
69	1,617	0,558	0,3451
70	5,721	2,637	0,4609
71	3,377	1,552	0,4596
72	4,001	1,726	0,4314
73	1,793	0,760	0,4239
74	4,931	1,793	0,3636
75	4,246	1,419	0,3342
76	4,728	1,232	0,2606
77	2,772	1,232	0,4444
78	5,027	2,071	0,4120
79	3,159	1,253	0,3966
80	4,409	1,108	0,2513
81	2,274	1,281	0,5633
82	3,403	1,532	0,4502
83	2,998	1,576	0,5257
84	3,563	1,432	0,4019
85	3,164	1,453	0,4592
86	1,287	0,483	0,3753
87	2,695	1,206	0,4475
88	3,403	1,162	0,3415
89	3,037	1,282	0,4221
90	4,648	1,706	0,3670
91	3,223	1,631	0,5061
92	2,701	1,274	0,4717
93	3,098	1,447	0,4671
94	3,188	1,623	0,5091
95	3,425	1,606	0,4689
96	3,885	1,415	0,3642
97	4,016	1,696	0,4223
98	4,037	1,480	0,3666
99	1,136	0,395	0,3477
100	5,433	1,361	0,2505
101	3,517	1,730	0,4919
102	3,397	1,501	0,4419
103	3,076	1,277	0,4151
104	4,320	1,654	0,3829
105	2,424	1,007	0,4154
106	4,094	2,092	0,5110
107	6,186	2,490	0,4025
108	1,654	0,782	0,4728
109	4,721	1,745	0,3696
110	2,806	1,063	0,3788
111	2,707	1,070	0,3953
112	3,280	1,369	0,4174
113	2,780	0,924	0,3324
114	4,354	2,166	0,4975
115	3,828	1,781	0,4653
116	5,278	2,881	0,5459
117	1,834	1,008	0,5496
118	3,117	1,870	0,5999
119	1,617	0,940	0,5813
120	3,311	1,743	0,5264

121	2,085	1,036	0,4969
122	4,475	2,426	0,5421
123	5,627	2,943	0,5230
124	5,218	2,363	0,4529
125	9,122	3,001	0,3290
126	4,542	1,808	0,3981
127	6,503	2,860	0,4398
128	4,299	1,668	0,3880
129	4,775	2,028	0,4247
130	6,791	3,268	0,4812
131	4,667	1,804	0,3865
132	7,733	4,140	0,5354
133	7,175	2,562	0,3571
134	4,870	2,031	0,4170
135	2,710	1,269	0,4683
136	5,180	2,625	0,5068
137	4,070	1,835	0,4509
138	3,667	1,828	0,4985
139	2,394	1,331	0,5560
140	5,401	2,593	0,4801
141	3,499	1,544	0,4413
142	4,808	1,838	0,3823
143	9,201	2,968	0,3226
144	6,614	2,890	0,4370
145	5,504	2,670	0,4851
146	3,321	1,472	0,4432
147	4,088	1,847	0,4518
148	3,824	1,811	0,4736
149	4,634	1,912	0,4126
150	4,289	2,013	0,4693
151	3,431	1,501	0,4375
152	6,155	2,652	0,4309
153	3,969	2,268	0,5714
154	4,369	2,133	0,4882
155	4,680	1,957	0,4182
156	4,129	1,568	0,3798
157	3,584	2,321	0,6476
158	9,324	4,044	0,4337
159	3,820	1,654	0,4330
160	6,416	3,461	0,5394
161	6,598	2,216	0,3359
162	4,778	1,810	0,3788
163	9,063	4,333	0,4781
164	5,843	2,153	0,3685
165	7,408	3,148	0,4249
166	6,157	2,216	0,3599
167	0,917	0,537	0,5856
168	4,307	1,710	0,3970
169	4,668	2,216	0,4747
170	3,568	1,308	0,3666
171	4,428	1,445	0,3263
172	4,165	1,302	0,3126
173	3,854	1,795	0,4657
174	3,858	1,713	0,4440
175	2,126	0,939	0,4417
176	4,113	1,814	0,4410
177	5,776	2,813	0,4870
178	3,621	1,474	0,4071
179	4,349	1,798	0,4134
180	4,768	2,348	0,4924
181	6,744	2,209	0,3276
182	4,866	1,837	0,3775
183	2,916	1,862	0,6385
184	3,204	1,163	0,3630
185	4,158	1,476	0,3550
186	2,158	1,064	0,4930
187	5,110	2,061	0,4033
188	4,005	1,197	0,2989
189	3,459	1,882	0,5441
190	4,270	1,670	0,3911
191	3,720	1,818	0,4887
192	3,308	1,327	0,4011
193	8,129	3,017	0,3711
194	3,852	1,391	0,3611
195	7,458	3,206	0,4299
196	5,406	2,104	0,3892
197	3,509	1,343	0,3827
198	4,292	1,763	0,4108
199	1,485	0,557	0,3751
200	2,014	0,761	0,3779
201	3,154	0,845	0,2679
202	2,538	1,617	0,6371
203	2,149	1,001	0,4658
204	2,883	1,084	0,3760
205	3,168	0,838	0,2645
206	4,686	1,505	0,3212
207	2,513	1,128	0,4489
208	2,685	1,167	0,4346
209	2,208	0,790	0,3578
210	4,256	2,203	0,5176
211	2,587	0,877	0,3390
212	4,032	1,177	0,2919
213	3,598	1,567	0,4355
214	3,088	1,358	0,4398
215	2,323	0,872	0,3754
216	2,686	1,199	0,4464
217	2,520	0,841	0,3337
218	2,541	0,904	0,3558
219	2,617	0,884	0,3378
220	3,836	1,380	0,3597
221	2,846	1,180	0,4146
222	4,058	1,097	0,2703
223	4,527	1,424	0,3146
224	3,587	1,958	0,5459
225	3,112	1,063	0,3416
226	3,761	1,439	0,3826
227	3,492	1,108	0,3173
228	4,026	1,533	0,3808
229	3,857	1,665	0,4317
230	3,682	1,751	0,4756
231	2,362	0,697	0,2951
232	2,889	1,115	0,3859
233	3,366	0,858	0,2549
234	3,489	1,197	0,3431
235	5,129	1,939	0,3780
236	5,812	1,808	0,3111
237	4,397	2,308	0,5249
238	2,522	0,825	0,3271
239	5,181	2,079	0,4013
240	3,666	1,447	0,3947
		average	0,4347
		stdev	0,0868

ch2		
chromosome length (µm)	cCENH3 region length (µm)	cCENH3/chromosome length
4,337	2,272	0,5239
4,131	2,079	0,5033
5,829	2,353	0,4037
6,360	3,285	0,5165
4,161	1,777	0,4271
6,063	2,558	0,4219
4,713	1,878	0,3985
3,425	1,606	0,4689
3,280	1,369	0,4174
3,828	1,781	0,4653
4,634	1,912	0,4126
4,369	2,133	0,4882
4,428	1,445	0,3263
5,406	2,104	0,3892
4,058	1,097	0,2703
	average	**0,4289**
	stdev	0,0697

ch4		
chromosome length (µm)	cCENH3 region length (µm)	cCENH3/chromosome length
4,588	2,206	0,4808
4,492	2,465	0,5488
4,102	1,85	0,4510
5,325	2,28	0,4282
3,576	1,648	0,4609
4,878	2,548	0,5223
3,983	2,001	0,5024
4,233	1,977	0,4670
4,888	2,153	0,4405
3,098	1,447	0,4671
2,806	1,063	0,3788
5,627	2,943	0,5230
5,401	2,593	0,4801
5,504	2,670	0,4851
4,680	1,957	0,4182
5,776	2,813	0,4870
3,621	1,474	0,4071
4,270	1,670	0,3911
2,538	1,617	0,6371
3,598	1,567	0,4355
3,761	1,439	0,3826
3,857	1,665	0,4317
	average	**0,4648**
	stdev	0,0599

ch3		
chromosome length (µm)	cCENH3 region length (µm)	cCENH3/chromosome length
7,38	3,262	0,4420
8,509	3,683	0,4328
6,101	2,316	0,3796
6,396	2,413	0,3773
6,455	2,956	0,4579
9,034	4,933	0,5460
6,991	1,924	0,2752
7,413	3,209	0,4329
6,761	3,223	0,4767
5,574	3,144	0,5640
5,578	3,049	0,5466
5,721	2,637	0,4609
5,027	2,071	0,4120
4,721	1,745	0,3696
4,354	2,166	0,4975
5,278	2,881	0,5459
9,122	3,001	0,3290
7,733	4,140	0,5354
7,175	2,562	0,3571
9,201	2,968	0,3226
6,614	2,890	0,4370
6,155	2,652	0,4309
9,324	4,044	0,4337
6,744	2,209	0,3276
5,110	2,061	0,4033
8,129	3,017	0,3711
7,458	3,206	0,4299
4,686	1,505	0,3212
4,256	2,203	0,5176
4,026	1,533	0,3808
5,812	1,808	0,3111
	average	**0,4234**
	stdev	0,0792

y_p		
chromosome length (µm)	cCENH3 region length (µm)	cCENH3/chromosome length
1,563	0,551	0,3525
1,282	0,568	0,4431
1,617	0,558	0,3451
1,793	0,760	0,4239
1,287	0,483	0,3753
1,136	0,395	0,3477
0,917	0,537	0,5856
1,485	0,557	0,3751
	average	**0,4060**
	stdev	0,0809

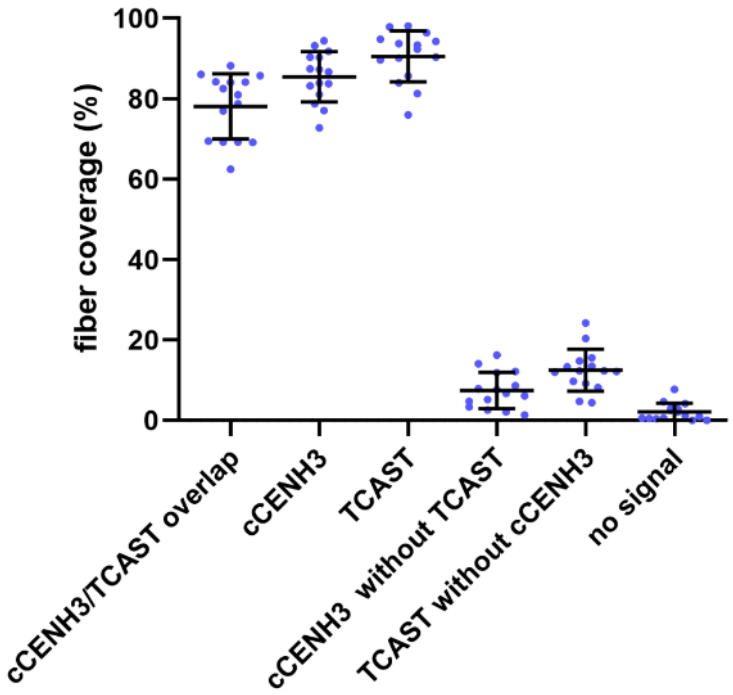
S4 Appendix. cCENH3-ChIPped DNA FISH signal quantification on *Tribolium castaneum* chromosomes. DNA immunoprecipitated by using cCENH3 antibody (cCENH3-ChIPped DNA) was fluorochrome-labeled and in situ hybridized to *T. castaneum* chromosomes. cCENH3-ChIPped DNA signal proportion on *T. castaneum* metaphase chromosomes was calculated as a ratio of the cCENH3-ChIPped DNA signal length to the corresponding chromosome length. 122 chromosomes were analyzed.

chromosome number	chromosome length (μm)	cCENH3-ChIPped DNA FISH signal length (μm)	cCENH3-ChIPped DNA signal length / chromosome length
1	2,489	1,359	0,546
2	2,417	1,002	0,415
3	2,413	1,096	0,454
4	2,869	0,677	0,236
5	2,107	1,184	0,562
6	2,449	0,978	0,399
7	3,064	0,906	0,296
8	3,190	1,224	0,384
9	2,799	1,217	0,435
10	3,095	1,398	0,452
11	3,052	1,207	0,395
12	3,014	1,552	0,515
13	1,149	0,439	0,382
14	2,752	1,428	0,519
15	2,617	1,342	0,513
16	2,781	0,847	0,305
17	1,793	1,045	0,583
18	2,875	1,596	0,555
19	3,638	1,527	0,420
20	3,052	1,810	0,593
21	2,446	1,121	0,458
22	2,460	1,501	0,610
23	3,349	1,618	0,483
24	2,803	1,616	0,577
25	2,668	1,474	0,552
26	2,916	1,164	0,399
27	2,624	1,259	0,480
28	3,012	1,722	0,572
29	3,051	1,541	0,505
30	3,058	1,542	0,504
31	4,145	2,113	0,510
32	3,320	1,910	0,575
33	4,600	1,847	0,402
34	3,450	1,550	0,449
35	3,489	1,720	0,493
36	3,377	1,552	0,460
37	3,773	1,985	0,526
38	6,001	2,367	0,394
39	3,095	1,318	0,426
40	3,991	1,571	0,394
41	2,773	1,731	0,624
42	2,688	1,328	0,494
43	3,053	1,252	0,410
44	2,422	1,177	0,486
45	2,406	1,200	0,499
46	2,278	1,236	0,543
47	2,786	0,663	0,238
48	3,030	0,952	0,314
49	2,969	1,310	0,441
50	4,289	1,496	0,349
51	5,710	1,195	0,209
52	3,467	1,291	0,372
53	4,248	1,375	0,324
54	3,299	1,612	0,489
55	4,313	1,487	0,345
56	2,641	1,411	0,534
57	3,623	1,962	0,542
58	4,010	2,058	0,513
59	2,526	1,243	0,492
60	3,111	1,452	0,467
61	3,296	1,424	0,432
62	2,332	1,297	0,556
63	3,265	1,448	0,443
64	2,913	1,530	0,525
65	2,706	1,349	0,499
66	3,024	1,087	0,359
67	2,738	1,035	0,378
68	3,790	1,107	0,292
69	3,052	1,624	0,532
70	2,697	1,477	0,548
71	2,983	1,736	0,582
72	2,976	1,310	0,440
73	3,220	1,289	0,400
74	3,127	1,461	0,467
75	2,841	1,832	0,645
76	2,628	1,739	0,662
77	3,142	1,168	0,372
78	3,526	1,786	0,507
79	2,842	1,200	0,422
80	2,714	1,306	0,481
81	3,326	1,446	0,435
82	2,617	1,247	0,476
83	3,333	1,019	0,306
84	3,830	1,898	0,496
85	2,292	1,156	0,504
86	3,105	1,837	0,592
87	2,931	1,663	0,567
88	3,161	1,400	0,443
89	3,362	1,454	0,432
90	3,068	1,717	0,560
91	2,670	1,465	0,549
92	3,640	1,576	0,433
93	5,112	1,385	0,271
94	2,837	1,537	0,542
95	3,161	1,028	0,325
96	2,644	1,428	0,540
97	3,578	1,740	0,486
98	2,150	1,005	0,467
99	2,959	1,173	0,396
100	2,637	1,337	0,507
101	2,874	1,392	0,484
102	3,002	1,244	0,414
103	3,363	1,166	0,347
104	2,516	0,880	0,350
105	3,213	1,337	0,416
106	2,936	1,147	0,391
107	2,494	0,708	0,284
108	2,583	1,195	0,463
109	1,878	1,184	0,630
110	2,638	1,090	0,413
111	3,724	1,198	0,322
112	4,050	1,556	0,384
113	3,173	1,199	0,378
114	3,107	1,433	0,461
115	4,877	1,412	0,290
116	3,641	1,925	0,529
117	2,319	1,067	0,460
118	3,933	2,027	0,515
119	3,848	2,068	0,537
120	3,082	1,425	0,462
121	3,154	1,859	0,589
122	3,896	1,876	0,482

Chromosomes analyzed	Average cCENH3-ChIPped DNA signal length / chromosome length ratio (%)	stdev (%)
122	45,81	9,3366

S5 Appendix. Quantification analysis of chromatin fiber regions with overlapping cCENH3 and TCAST signals. Selected regions of 15 extended chromatin fibers with prominent cCENH3 and TCAST co-localization were quantified. Pixel intensities along the spline fitted lines were measured, and presence or absence of each signal was evaluated by extracting values that were higher or equal to 0, respectively. The degree of overlapping was calculated based on coincident presence of cCENH3 and TCAST signals.

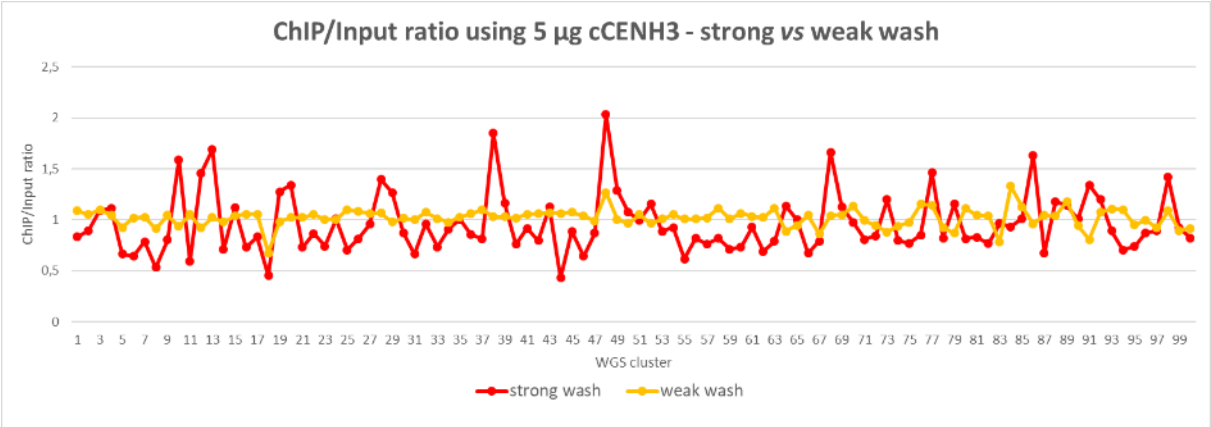
fiber	cCENH3/TCAST overlap fiber coverage	cCENH3 fiber coverage	TCAST fiber coverage	cCENH3 without TCAST fiber coverage	TCAST without cCENH3 fiber coverage	no signal fiber coverage
1	86,07%	87,41%	98,11%	1,34%	12,04%	0,55%
2	78,75%	83,92%	94,28%	5,18%	15,53%	0,54%
3	69,22%	77,06%	89,64%	7,85%	20,42%	2,52%
4	69,14%	83,23%	81,29%	14,09%	12,15%	4,62%
5	69,22%	81,07%	83,98%	11,84%	14,76%	4,17%
6	85,76%	94,41%	90,14%	8,65%	4,38%	1,20%
7	80,99%	93,19%	85,68%	12,21%	4,69%	2,11%
8	84,24%	90,36%	93,37%	6,12%	9,14%	0,50%
9	82,51%	87,26%	94,82%	4,75%	12,31%	0,43%
10	69,49%	72,79%	93,75%	3,31%	24,26%	2,94%
11	84,09%	86,69%	96,43%	2,60%	12,34%	0,97%
12	88,21%	90,29%	97,92%	2,08%	9,71%	0,00%
13	62,49%	78,76%	75,98%	16,27%	13,49%	7,75%
14	77,01%	83,76%	90,35%	6,75%	13,34%	2,89%
15	84,14%	91,81%	92,33%	7,67%	8,19%	0,00%
average	**78,09%**	**85,47%**	**90,54%**	**7,38%**	**12,45%**	**2,08%**
std	8,11%	6,22%	6,36%	4,53%	5,22%	2,16%

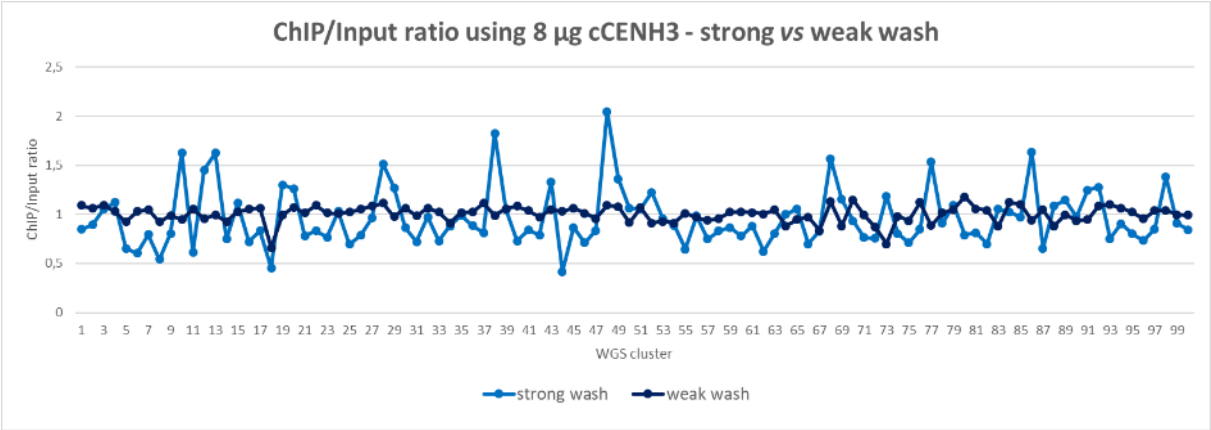


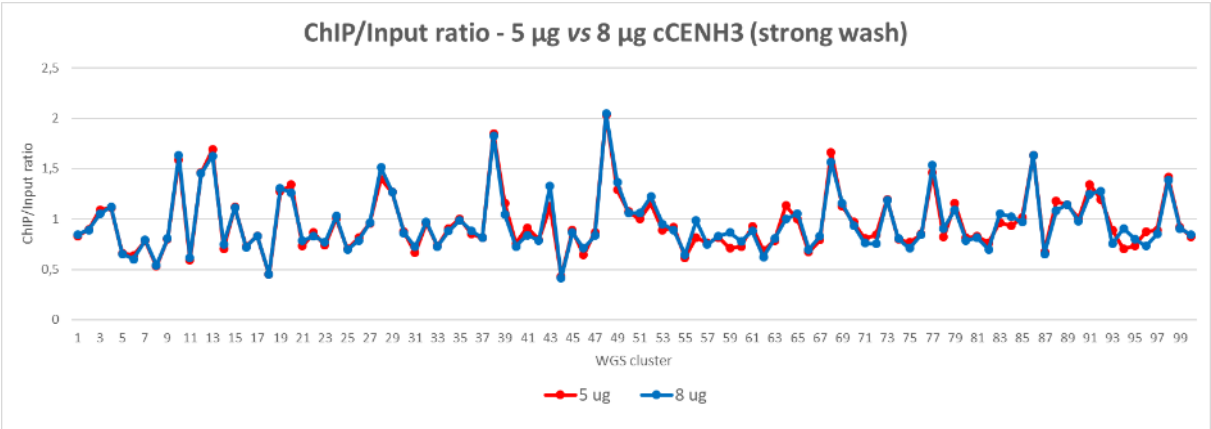
S6 Appendix. Comparison of cCENH3-ChIP experiments using different Ab amounts and washing conditions. ChIP experiments were performed using different amounts of cCENH3 antibody (5 and 8 µg) and different washing conditions (weak wash with commercial washing buffers and strong wash with low/high salt buffers). The ChIP/Input ratio for the top 100 WGS clusters was compared between different ChIP experiments.

	ChIP 1	ChIP 2	ChIP 3	ChIP 4
Ab amount	5 µg cCENH3	8 µg cCENH3	5 µg cCENH3	8 µg cCENH3
washing conditions	strong wash	strong wash	weak wash	weak wash
WGS cluster	ChIP/ Input ratio			
1	0,831181518	0,848227805	1,090083196	1,089969229
2	0,894365255	0,893160677	1,050759553	1,059626581
3	1,090875721	1,052621919	1,098321972	1,093969586
4	1,109582678	1,120730313	1,043697299	1,030548807
5	0,663164439	0,654173143	0,925917003	0,926002487
6	0,640644561	0,603677076	1,018803877	1,030343027
7	0,786124349	0,792707525	1,026177546	1,050958661
8	0,530925312	0,541790773	0,913593713	0,923116723
9	0,801932367	0,806073154	1,044858523	0,990338164
10	1,589416058	1,630508245	0,933630711	0,946066504
11	0,590654914	0,616023758	1,051717057	1,05187502
12	1,458845115	1,452270526	0,920646246	0,958259008
13	1,689318523	1,625629396	1,022792789	0,993734866
14	0,708230655	0,750616067	0,979793001	0,927550517
15	1,116978989	1,11641113	1,038046564	1,025553663
16	0,729319988	0,720783044	1,054607006	1,05195761
17	0,834174969	0,831581821	1,055547973	1,064146308
18	0,451270114	0,454711062	0,674425665	0,660155855
19	1,274361305	1,302534585	0,982328587	0,992628496
20	1,340743681	1,262935027	1,022628027	1,073971153
21	0,730811706	0,784925456	1,022087245	1,015184981
22	0,866280061	0,831144465	1,052532833	1,091079652
23	0,740589865	0,769594264	1,005214274	1,014176308
24	1,0084332	1,030625832	1,005326232	1,00665779
25	0,702854671	0,695717993	1,101211073	1,022491349
26	0,814334731	0,786885246	1,080060999	1,055280213
27	0,958490566	0,962264151	1,063522013	1,085534591
28	1,395411887	1,513242961	1,066944734	1,11637122
29	1,266509434	1,270636792	0,980542453	0,978183962
30	0,872917786	0,862170876	1,013970983	1,063406771
31	0,666156202	0,723583461	1,00229709	0,985451761
32	0,96013289	0,975498339	1,074750831	1,065199336
33	0,734034268	0,728971963	1,013629283	1,024922118
34	0,907832107	0,880787538	0,97490264	0,912591952
35	1,005128205	0,987179487	1,025641026	1,019230769
36	0,85387132	0,884405671	1,058887677	1,021810251
37	0,815312683	0,812974868	1,099649328	1,113676213
38	1,84947839	1,827868852	1,031296572	0,989567809
39	1,161290323	1,049694856	1,029642546	1,056669573
40	0,758179959	0,729550102	1,016359918	1,082310838
41	0,915195867	0,838570814	1,055101162	1,038743005
42	0,796664019	0,789515488	1,059571088	0,973788721
43	1,124855491	1,330635838	1,069364162	1,050867052
44	0,43052476	0,417775314	1,064301552	1,031042129
45	0,888272583	0,867670365	1,073692552	1,064580032
46	0,641928721	0,710272537	1,038155136	1,010062893
47	0,870111732	0,836592179	0,991620112	0,958100559
48	2,032462391	2,046912114	1,269002375	1,09263658
49	1,289828983	1,362286229	1,005850585	1,081008101
50	1,075102041	1,062040816	0,964897959	0,916734694

51	0,998979592	1,058163265	1,052040816	1,067346939
52	1,156069364	1,223781998	0,965317919	0,914120562
53	0,886912325	0,952986023	1,008894536	0,927573062
54	0,922389726	0,888888889	1,050809604	0,914014517
55	0,615686275	0,647058824	1,009803922	1,010457516
56	0,816831683	0,98349835	1,008250825	0,968646865
57	0,761706556	0,747138398	1,018730489	0,939646202
58	0,818577649	0,833091437	1,114658926	0,959361393
59	0,711724138	0,867586207	1,009655172	1,024827586
60	0,728985507	0,778985507	1,063768116	1,02826087
61	0,92712294	0,881495564	1,035487959	1,020912548
62	0,689314636	0,623740865	1,022516295	1,003752716
63	0,789719626	0,806853583	1,115264798	1,046728972
64	1,135025381	0,998984772	0,887309645	0,883248731
65	1	1,056951424	0,94639866	0,948073702
66	0,673174873	0,697792869	1,044991511	0,975382003
67	0,791304348	0,827536232	0,860869565	0,837681159
68	1,664948454	1,56921944	1,04197349	1,130338733
69	1,127240143	1,154121864	1,048387097	0,88172043
70	0,97173913	0,936956522	1,132608696	1,15
71	0,805471125	0,762917933	0,993920973	0,992907801
72	0,845108696	0,755434783	0,942934783	0,869565217
73	1,197452229	1,18895966	0,881104034	0,700636943
74	0,797752809	0,806179775	0,938202247	0,980337079
75	0,770114943	0,709770115	0,977011494	0,936781609
76	0,852022059	0,848345588	1,15625	1,120404412
77	1,460176991	1,533185841	1,14380531	0,887168142
78	0,821875	0,9078125	0,9234375	1,0140625
79	1,154696133	1,093922652	0,870165746	1,049723757
80	0,815351195	0,789283128	1,113685735	1,180304127
81	0,830472103	0,815450644	1,045779685	1,058655222
82	0,766922095	0,700510856	1,041507024	1,041507024
83	0,966165414	1,052631579	0,785714286	0,879699248
84	0,933078394	1,026768642	1,328871893	1,122370937
85	1,013392857	0,970982143	1,129464286	1,102678571
86	1,634259259	1,635802469	0,959876543	0,938271605
87	0,677137871	0,654450262	1,046538685	1,046538685
88	1,179710145	1,086956522	1,037681159	0,88115942
89	1,140142518	1,144893112	1,180522565	0,997624703
90	1,008281573	0,979296066	0,944099379	0,933747412
91	1,342857143	1,248979592	0,808163265	0,946938776
92	1,196911197	1,27992278	1,073359073	1,088803089
93	0,892307692	0,753846154	1,107692308	1,101098901
94	0,704280156	0,906614786	1,101167315	1,06614786
95	0,737634409	0,802150538	0,948387097	1,023655914
96	0,872727273	0,732467532	0,997402597	0,953246753
97	0,892857143	0,850840336	0,922268908	1,037815126
98	1,416370107	1,384341637	1,088967972	1,042704626
99	0,922261484	0,908127208	0,893992933	0,996466431
100	0,820547945	0,843835616	0,915068493	0,993150685


ChIP/Input ratio using 5 µg cCENH3 - strong vs weak wash
ChIP/Input ratio
2,5
2
1,5
1
0,5
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99
WGS cluster
strong wash
weak wash


ChIP/Input ratio using 8 µg cCENH3 - strong vs weak wash
ChIP/Input ratio
2,5
2
1,5
1
0,5
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99
WGS cluster
strong wash
weak wash


ChIP/Input ratio - 5 µg vs 8 µg cCENH3 (strong wash)
ChIP/Input ratio
2,5
2
1,5
1
0,5
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99
WGS cluster
5 ug
8 ug

S7 Appendix. Comparison of ChIP-Seq Mapper outcome using different randomly sampled subsets of cCENH3-ChIP and Input reads. Following cCENH3-ChIP experiment and subsequent sequencing of ChIPped DNA fragments, ChIP-Seq Mapper analysis was performed using three different subsets of ChIP and Input randomly subsampled reads: 250000, 500000, and 1000000. ChIP and Input reads were mapped to the top 1000 WGS repeat clusters. Based on the ChIP/Input ratio, WGS clusters enriched for cCENH3 were determined.

1.000.000 subsampled reads analyzed			
WGS cluster	ChIP_ Hits	Input_ Hits	ChIP/Input ratio
1	51342	61421	0,83590
2	80589	89658	0,89885
3	42004	38140	1,10131
4	31220	27988	1,11548
5	85663	128680	0,66571
6	2E+05	279570	0,62656
7	34893	45115	0,77342
8	38185	71511	0,53397
9	1249	1449	0,86197
10	23779	14796	1,60712
11	18830	31653	0,59489
12	28617	19621	1,45849
13	42765	26017	1,64373
14	1514	2029	0,74618
15	3801	3522	1,07922
16	4977	6794	0,73256
17	6085	7327	0,83049
18	4903	9881	0,49620
19	12573	9903	1,26962
20	10166	7557	1,34524
21	2691	3622	0,74296
22	5102	5863	0,87020
23	4541	6137	0,73994
24	2262	2253	1,00399
25	3256	4624	0,70415
26	2171	2623	0,82768
27	1565	1590	0,98428
28	6942	4795	1,44776
29	2254	1696	1,32901
30	3269	3722	0,87829
31	889	1306	0,68070
32	2378	2408	0,98754
33	1800	2568	0,70093
34	4278	4622	0,92557
35	765	780	0,98077
36	769	917	0,83860
37	2797	3422	0,81736
38	2471	1342	1,84128
39	1294	1147	1,12816
40	1469	1956	0,75102
41	2102	2323	0,90486
42	1053	1259	0,83638
43	964	865	1,11445
44	2358	5412	0,43570
45	2240	2524	0,88748
46	1505	2385	0,63103
47	685	716	0,95670
48	10565	5052	2,09125
49	2816	2222	1,26733
50	1224	1225	0,99918
51	1022	980	1,04286
52	1300	1211	1,07349
53	762	787	0,96823
54	1575	1791	0,87940
55	991	1530	0,64771
56	468	606	0,77228
57	741	961	0,77107
58	559	689	0,81132
59	539	725	0,74345
60	997	1380	0,72246
61	1415	1578	0,89670
62	3537	5063	0,69860
63	535	642	0,83333
64	1111	985	1,12792
65	625	597	1,04690
66	804	1178	0,68251
67	556	690	0,80580
68	4551	2716	1,67563
69	597	558	1,06989
70	410	460	0,89130
71	817	987	0,82776
72	349	368	0,94837
73	503	471	1,06794
74	311	356	0,87360
75	524	696	0,75287
76	956	1088	0,87868
77	710	452	1,57080
78	557	640	0,87031
79	427	362	1,17956
80	1147	1381	0,83056
81	1118	1398	0,79971
82	1165	1566	0,74393
83	235	266	0,88346
84	500	523	0,95602
85	451	448	1,00670
86	1115	648	1,72068
87	1202	1719	0,69924
88	415	345	1,20290
89	508	421	1,20665
90	464	483	0,96066
91	1016	735	1,38231
92	602	518	1,16216
93	379	455	0,83297
94	175	257	0,68093
95	730	930	0,78495
96	327	385	0,84935
97	417	476	0,87605
98	385	281	1,37011
99	271	283	0,95760
100	587	730	0,80411
101	626	862	0,72622
102	772	1007	0,76663
103	218	260	0,83846
104	486	528	0,92045
105	186	231	0,80519
106	624	813	0,76753
107	221	242	0,91322
108	133	184	0,72283
109	142	181	0,78453
110	289	325	0,88923
111	256	300	0,85333
112	344	341	1,00880
113	515	375	1,37333
114	462	669	0,69058
115	250	430	0,58140
116	267	292	0,91438
117	607	410	1,48049
118	239	310	0,77097
119	301	272	1,10662
120	375	314	1,19427
121	266	318	0,83648
122	161	171	0,94152
123	354	463	0,76458
124	193	238	0,81092
125	190	182	1,04396
126	225	227	0,99119
127	265	310	0,85484
128	124	122	1,01639
129	166	190	0,87368
130	2079	2007	1,03587
131	157	208	0,75481
132	191	171	1,11696
133	54	71	0,76056
134	262	303	0,86469
135	93	103	0,90291
136	637	906	0,70309
137	486	620	0,78387
138	158	248	0,63710
139	131	224	0,58482
140	255	284	0,89789
141	488	158	3,08861
142	2741	1869	1,46656
143	460	192	2,39583
144	882	1356	0,65044
145	703	869	0,80898
146	281	306	0,91830
147	1007	701	1,43652
148	247	260	0,95000
149	144	175	0,82286
150	175	139	1,25899
151	189	248	0,76210
152	107	152	0,70395
153	534	461	1,15835
154	348	282	1,23404
155	221	174	1,27011
156	638	615	1,03740
157	196	260	0,75385
158	285	151	1,88742
159	342	395	0,86582
160	1120	900	1,24444
161	194	195	0,99487
162	389	378	1,02910
163	157	196	0,80102
164	255	133	1,91729
165	199	273	0,72894
166	163	138	1,18116
167	374	474	0,78903
168	409	700	0,58429
169	495	238	2,07983
170	111	146	0,76027
171	356	196	1,81633
172	125	123	1,01626
173	350	504	0,69444
174	165	114	1,44737
175	455	668	0,68114
176	390	452	0,86283
177	264	319	0,82759
178	903	1073	0,84157
179	179	322	0,55590
180	126	142	0,88732
181	114	107	1,06542
182	142	140	1,01429
183	307	226	1,35841
184	202	143	1,41259
185	56	81	0,69136
186	81	159	0,50943
187	195	199	0,97990
188	80	142	0,56338
189	252	231	1,09091
190	144	127	1,13386
191	138	115	1,20000
192	64	54	1,18519
193	275	153	1,79739
194	221	132	1,67424
195	81	118	0,68644
196	164	182	0,90110
197	513	647	0,79289
198	210	99	2,12121
199	63	79	0,79747
200	202	112	1,80357
201	198	195	1,01538
202	262	338	0,77515
203	200	226	0,88496
204	113	103	1,09709
205	118	91	1,29670
206	58	81	0,71605
207	76	114	0,66667
208	120	112	1,07143
209	126	129	0,97674
210	159	164	0,96951
211	87	115	0,75652
212	57	73	0,78082
213	87	102	0,85294
214	236	194	1,21649
215	76	93	0,81720
216	412	547	0,75320
217	75	80	0,93750
218	319	406	0,78571
219	251	635	0,39528
220	591	813	0,72694
221	200	222	0,90090
222	171	201	0,85075
223	81	75	1,08000
224	179	190	0,94211
225	351	348	1,00862
226	207	128	1,61719
227	101	105	0,96190
228	165	201	0,82090
229	380	230	1,65217
230	965	1242	0,77697
231	110	154	0,71429
232	100	102	0,98039
233	71	44	1,61364
234	402	158	2,54430
235	100	113	0,88496
236	108	139	0,77698
237	115	85	1,35294
238	91	129	0,70543
239	256	396	0,64646
240	194	112	1,73214
241	729	604	1,20695
242	225	178	1,26404
243	107	118	0,90678
244	107	81	1,32099
245	1053	813	1,29520
246	119	162	0,73457
247	110	109	1,00917
248	110	138	0,79710
249	170	101	1,68317
250	84	45	1,86667
251	55	79	0,69620
252	177	171	1,03509
253	288	110	2,61818
254	126	157	0,80255
255	47	88	0,53409
256	87	131	0,66412
257	114	151	0,75497
258	90	123	0,73171
259	155	216	0,71759
260	124	136	0,91176
261	92	64	1,43750
262	93	131	0,70992
263	80	85	0,94118
264	337	145	2,32414
265	158	238	0,66387
266	153	134	1,14179
267	49	54	0,90741
268	67	106	0,63208
269	100	178	0,56180
270	732	342	2,14035
271	84	83	1,01205
272	155	174	0,89080
273	122	124	0,98387
274	305	423	0,72104
275	262	326	0,80368
276	57	50	1,14000
277	350	425	0,82353
278	143	97	1,47423
279	92	71	1,29577
280	532	793	0,67087
281	297	331	0,89728
282	139	111	1,25225
283	72	69	1,04348
284	245	338	0,72485
285	92	62	1,48387
286	241	333	0,72372
287	956	1772	0,53950
288	226	199	1,13568
289	169	114	1,48246
290	202	220	0,91818
291	69	104	0,66346
292	68	73	0,93151
293	76	79	0,96203
294	157	79	1,98734
295	95	86	1,10465
296	90	65	1,38462
297	37	68	0,54412
298	296	137	2,16058
299	152	174	0,87356
300	240	260	0,92308
301	57	63	0,90476
302	199	218	0,91284
303	43	62	0,69355
304	301	175	1,72000
305	94	100	0,94000
306	80	88	0,90909
307	206	126	1,63492
308	46	56	0,82143
309	154	174	0,88506
310	177	167	1,05988
311	305	385	0,79221
312	138	103	1,33981
313	88	55	1,60000
314	165	218	0,75688
315	191	247	0,77328
316	105	133	0,78947
317	128	130	0,98462
318	498	651	0,76498
319	52	41	1,26829
320	105	96	1,09375
321	61	30	2,03333
322	125	145	0,86207
323	116	72	1,61111
324	138	101	1,36634
325	78	77	1,01299
326	114	133	0,85714
327	507	575	0,88174
328	88	72	1,22222
329	121	139	0,87050
330	55	35	1,57143
331	369	425	0,86824
332	172	181	0,95028
333	44	47	0,93617
334	124	206	0,60194
335	137	212	0,64623
336	177	70	2,52857
337	277	329	0,84195
338	60	49	1,22449
339	197	190	1,03684
340	64	48	1,33333
341	57	72	0,79167
342	80	89	0,89888
343	65	67	0,97015
344	67	32	2,09375
345	297	351	0,84615
346	110	138	0,79710
347	134	90	1,48889
348	57	59	0,96610
349	183	231	0,79221
350	122	147	0,82993
351	34	45	0,75556
352	58	66	0,87879
353	40	51	0,78431
354	50	83	0,60241
355	96	73	1,31507
356	297	416	0,71394
357	75	49	1,53061
358	136	166	0,81928
359	114	111	1,02703
360	240	208	1,15385

361	57	67	0,85075
362	391	441	0,88662
363	112	85	1,31765
364	35	55	0,63636
365	66	50	1,32000
366	168	166	1,01205
367	187	295	0,63390
368	36	46	0,78261
369	312	287	1,08711
370	230	282	0,81560
371	138	192	0,71875
372	90	100	0,90000
373	80	77	1,03896
374	374	408	0,91667
375	75	115	0,65217
376	69	55	1,25455
377	262	329	0,79635
378	666	1015	0,65616
379	115	125	0,92000
380	84	99	0,84848
381	189	193	0,97927
382	113	155	0,72903
383	57	52	1,09615
384	126	131	0,96183
385	80	57	1,40351
386	62	115	0,53913
387	96	80	1,20000
388	91	98	0,92857
389	56	82	0,68293
390	54	50	1,08000
391	1591	602	2,64286
392	224	319	0,70219
393	138	138	1,00000
394	442	85	5,20000
395	53	48	1,10417
396	74	110	0,67273
397	82	55	1,49091
398	49	51	0,96078
399	202	61	3,31148
400	150	216	0,69444
401	114	103	1,10680
402	89	67	1,32836
403	45	65	0,69231
404	48	39	1,23077
405	109	132	0,82576
406	56	71	0,78873
407	58	64	0,90625
408	62	87	0,71264
409	76	115	0,66087
410	212	241	0,87967
411	145	136	1,06618
412	33	44	0,75000
413	57	75	0,76000
414	43	56	0,76786
415	169	353	0,47875
416	155	204	0,75980
417	48	40	1,20000
418	71	81	0,87654
419	192	244	0,78689
420	96	91	1,05495
421	242	327	0,74006
422	149	174	0,85632
423	79	50	1,58000
424	170	264	0,64394
425	88	90	0,97778
426	133	249	0,53414
427	135	117	1,15385
428	76	60	1,26667
429	115	109	1,05505
430	103	63	1,63492
431	62	101	0,61386
432	48	50	0,96000
433	85	99	0,85859
434	91	97	0,93814
435	70	92	0,76087
436	74	131	0,56489
437	535	328	1,63110
438	61	110	0,55455
439	45	55	0,81818
440	249	292	0,85274
441	66	42	1,57143
442	73	44	1,65909
443	106	136	0,77941
444	47	42	1,11905
445	121	68	1,77941
446	68	60	1,13333
447	76	102	0,74510
448	349	211	1,65403
449	51	40	1,27500
450	60	46	1,30435
451	206	292	0,70548
452	33	51	0,64706
453	39	60	0,65000
454	89	61	1,45902
455	257	305	0,84262
456	205	307	0,66775
457	42	49	0,85714
458	45	24	1,87500
459	123	177	0,69492
460	84	59	1,42373
461	155	225	0,68889
462	112	111	1,00901
463	87	102	0,85294
464	154	124	1,24194
465	73	73	1,00000
466	69	57	1,21053
467	87	140	0,62143
468	54	26	2,07692
469	145	53	2,73585
470	81	115	0,70435
471	296	182	1,62637
472	104	73	1,42466
473	18	20	0,90000
474	33	42	0,78571
475	119	93	1,27957
476	88	91	0,96703
477	19	15	1,26667
478	140	193	0,72539
479	266	218	1,22018
480	162	199	0,81407
481	62	55	1,12727
482	36	36	1,00000
483	54	67	0,80597
484	212	140	1,51429
485	91	99	0,91919
486	41	40	1,02500
487	104	54	1,92593
488	76	167	0,45509
489	61	56	1,08929
490	91	94	0,96809
491	50	70	0,71429
492	51	50	1,02000
493	50	91	0,54945
494	44	31	1,41935
495	286	328	0,87195
496	151	228	0,66228
497	46	25	1,84000
498	79	68	1,16176
499	69	119	0,57983
500	137	165	0,83030
501	177	271	0,65314
502	15	16	0,93750
503	79	34	2,32353
504	63	84	0,75000
505	41	39	1,05128
506	46	81	0,56790
507	116	84	1,38095
508	164	110	1,49091
509	40	27	1,48148
510	86	100	0,86000
511	48	38	1,26316
512	77	64	1,20313
513	80	64	1,25000
514	38	42	0,90476
515	36	50	0,72000
516	55	70	0,78571
517	81	57	1,42105
518	120	189	0,63492
519	58	40	1,45000
520	49	66	0,74242
521	168	89	1,88764
522	44	42	1,04762
523	643	902	0,71286
524	90	54	1,66667
525	100	109	0,91743
526	45	40	1,12500
527	83	50	1,66000
528	36	57	0,63158
529	119	91	1,30769
530	182	214	0,85047
531	53	60	0,88333
532	60	38	1,57895
533	125	92	1,35870
534	336	352	0,95455
535	164	194	0,84536
536	12	14	0,85714
537	54	39	1,38462
538	105	68	1,54412
539	60	59	1,01695
540	54	63	0,85714
541	18	31	0,58065
542	49	71	0,69014
543	80	134	0,59701
544	50	65	0,76923
545	200	344	0,58140
546	70	67	1,04478
547	9	13	0,69231
548	79	94	0,84043
549	78	54	1,44444
550	66	108	0,61111
551	40	35	1,14286
552	21	18	1,16667
553	187	206	0,90777
554	418	461	0,90672
555	220	272	0,80882
556	26	45	0,57778
557	116	94	1,23404
558	150	201	0,74627
559	569	789	0,72117
560	158	153	1,03268
561	171	191	0,89529
562	78	99	0,78788
563	79	63	1,25397
564	121	156	0,77564
565	242	293	0,82594
566	35	27	1,29630
567	164	98	1,67347
568	127	143	0,88811
569	300	410	0,73171
570	99	103	0,96117
571	45	48	0,93750
572	53	65	0,81538
573	137	126	1,08730
574	81	87	0,93103
575	352	320	1,10000
576	43	65	0,66154
577	177	186	0,95161
578	47	35	1,34286
579	64	102	0,62745
580	32	45	0,71111
581	180	213	0,84507
582	36	48	0,75000
583	71	89	0,79775
584	63	64	0,98438
585	65	64	1,01563
586	48	58	0,82759
587	21	25	0,84000
588	35	22	1,59091
589	57	83	0,68675
590	229	247	0,92713
591	27	32	0,84375
592	24	24	1,00000
593	51	43	1,18605
594	41	66	0,62121
595	73	54	1,35185
596	136	153	0,88889
597	107	105	1,01905
598	44	47	0,93617
599	106	179	0,59218
600	61	80	0,76250
601	23	32	0,71875
602	40	52	0,76923
603	63	61	1,03279
604	97	50	1,94000
605	30	22	1,36364
606	28	18	1,55556
607	145	123	1,17886
608	39	44	0,88636
609	39	56	0,69643
610	51	31	1,64516
611	216	78	2,76923
612	57	55	1,03636
613	28	37	0,75676
614	81	78	1,03846
615	44	16	2,75000
616	28	42	0,66667
617	84	113	0,74336
618	84	62	1,35484
619	68	65	1,04615
620	39	59	0,66102
621	71	70	1,01429
622	75	78	0,96154
623	8	13	0,61538
624	56	71	0,78873
625	107	156	0,68590
626	55	31	1,77419
627	25	33	0,75758
628	39	72	0,54167
629	132	81	1,62963
630	116	95	1,22105
631	74	92	0,80435
632	375	535	0,70093
633	54	66	0,81818
634	83	92	0,90217
635	27	44	0,61364
636	166	163	1,01840
637	95	63	1,50794
638	73	150	0,48667
639	266	292	0,91096
640	47	65	0,72308
641	119	157	0,75796
642	87	118	0,73729
643	27	42	0,64286
644	31	25	1,24000
645	32	54	0,59259
646	56	42	1,33333
647	50	37	1,35135
648	161	201	0,80100
649	38	38	1,00000
650	45	62	0,72581
651	228	211	1,08057
652	122	143	0,85315
653	98	121	0,80992
654	16	24	0,66667
655	170	236	0,72034
656	15	28	0,53571
657	36	48	0,75000
658	118	121	0,97521
659	78	48	1,62500
660	35	31	1,12903
661	139	195	0,71282
662	56	39	1,43590
663	50	57	0,87719
664	60	88	0,68182
665	51	70	0,72857
666	37	48	0,77083
667	26	30	0,86667
668	16	21	0,76190
669	118	58	2,03448
670	100	66	1,51515
671	112	172	0,65116
672	60	63	0,95238
673	18	19	0,94737
674	53	35	1,51429
675	30	33	0,90909
676	111	150	0,74000
677	64	51	1,25490
678	44	34	1,29412
679	56	66	0,84848
680	198	202	0,98020
681	29	27	1,07407
682	321	269	1,19331
683	333	371	0,89757
684	167	252	0,66270
685	43	61	0,70492
686	35	42	0,83333
687	71	86	0,82558
688	280	367	0,76294
689	42	47	0,89362
690	100	96	1,04167
691	135	225	0,60000
692	81	75	1,08000
693	73	106	0,68868
694	57	81	0,70370
695	57	38	1,50000
696	61	93	0,65591
697	86	130	0,66154
698	53	52	1,01923
699	81	102	0,79412
700	30	26	1,15385
701	79	71	1,11268
702	35	37	0,94595
703	58	31	1,87097
704	22	16	1,37500
705	35	35	1,00000
706	41	42	0,97619
707	60	42	1,42857
708	31	15	2,06667
709	35	59	0,59322
710	119	159	0,74843
711	174	78	2,23077
712	32	53	0,60377
713	54	53	1,01887
714	170	207	0,82126
715	40	43	0,93023
716	60	27	2,22222
717	60	132	0,45455
718	39	52	0,75000
719	138	180	0,76667
720	125	95	1,31579
721	189	242	0,78099
722	182	190	0,95789
723	78	120	0,65000
724	89	111	0,80180
725	145	124	1,16935
726	95	135	0,70370
727	291	368	0,79076
728	23	34	0,67647
729	62	47	1,31915
730	61	82	0,74390
731	26	26	1,00000
732	87	132	0,65909
733	61	92	0,66304
734	51	39	1,30769
735	33	70	0,47143
736	72	105	0,68571
737	11	14	0,78571
738	42	39	1,07692
739	97	195	0,49744
740	37	36	1,02778
741	51	59	0,86441
742	115	67	1,71642
743	148	207	0,71498
744	37	107	0,34579
745	171	272	0,62868
746	101	79	1,27848
747	226	289	0,78201
748	52	74	0,70270
749	37	62	0,59677
750	78	48	1,62500
751	36	45	0,80000
752	38	30	1,26667
753	61	128	0,47656
754	85	125	0,68000
755	161	252	0,63889
756	20	24	0,83333
757	78	92	0,84783
758	143	176	0,81250
759	46	82	0,56098
760	49	56	0,87500
761	336	617	0,54457
762	38	37	1,02703
763	10	18	0,55556
764	64	96	0,66667
765	188	303	0,62046
766	38	50	0,76000
767	70	77	0,90909
768	78	43	1,81395
769	172	269	0,63941
770	48	33	1,45455
771	49	40	1,22500
772	91	167	0,54491
773	36	38	0,94737
774	61	78	0,78205
775	151	144	1,04861
776	7	8	0,87500
777	79	96	0,82292
778	23	18	1,27778
779	29	51	0,56863
780	22	45	0,48889
781	26	19	1,36842
782	50	68	0,73529
783	29	36	0,80556
784	98	133	0,73684
785	38	32	1,18750
786	29	22	1,31818
787	75	117	0,64103
788	61	35	1,74286
789	70	91	0,76923
790	17	24	0,70833
791	97	101	0,96040
792	41	59	0,69492
793	38	44	0,86364
794	37	37	1,00000
795	14	17	0,82353
796	48	74	0,64865
797	115	41	2,80488
798	70	43	1,62791
799	92	90	1,02222
800	67	81	0,82716
801	3	3	1,00000
802	93	107	0,86916
803	262	268	0,97761
804	320	449	0,71269
805	166	155	1,07097

806	32	37	0,86486
807	6	5	1,20000
808	118	137	0,86131
809	63	43	1,46512
810	308	401	0,76808
811	77	107	0,71963
812	31	37	0,83784
813	99	46	2,15217
814	244	506	0,48221
815	72	70	1,02857
816	111	177	0,62712
817	57	65	0,87692
818	39	36	1,08333
819	51	66	0,77273
820	130	228	0,57018
821	61	25	2,44000
822	54	89	0,60674
823	114	131	0,87023
824	66	71	0,92958
825	40	88	0,45455
826	133	67	1,98507
827	222	248	0,89516
828	93	149	0,62416
829	30	47	0,63830
830	98	81	1,20988
831	34	47	0,72340
832	66	63	1,04762
833	67	83	0,80723
834	65	55	1,18182
835	39	45	0,86667
836	88	134	0,65672
837	20	10	2,00000
838	77	63	1,22222
839	46	56	0,82143
840	85	88	0,96591
841	46	50	0,92000
842	360	456	0,78947
843	55	70	0,78571
844	74	101	0,73267
845	63	84	0,75000
846	41	44	0,93182
847	79	65	1,21538
848	30	22	1,36364
849	26	33	0,78788
850	36	56	0,64286
851	42	45	0,93333
852	144	177	0,81356
853	152	146	1,04110
854	96	128	0,75000
855	10	4	2,50000
856	27	25	1,08000
857	13	12	1,08333
858	37	59	0,62712
859	97	104	0,93269
860	11	29	0,37931
861	46	43	1,06977
862	46	52	0,88462
863	40	27	1,48148
864	23	31	0,74194
865	32	31	1,03226
866	24	27	0,88889
867	17	25	0,68000
868	18	21	0,85714
869	57	52	1,09615
870	34	27	1,25926
871	179	145	1,23448
872	47	34	1,38235
873	74	86	0,86047
874	345	442	0,78054
875	124	137	0,90511
876	31	25	1,24000
877	33	58	0,56897
878	38	37	1,02703
879	21	33	0,63636
880	57	50	1,14000
881	26	25	1,04000
882	40	26	1,53846
883	62	86	0,72093
884	44	59	0,74576
885	221	273	0,80952
886	21	16	1,31250
887	160	200	0,80000
888	18	15	1,20000
889	35	31	1,12903
890	16	15	1,06667
891	20	19	1,05263
892	64	68	0,94118
893	62	57	1,08772
894	58	77	0,75325
895	13	12	1,08333
896	8	16	0,50000
897	38	81	0,46914
898	40	24	1,66667
899	40	56	0,71429
900	114	155	0,73548
901	265	220	1,20455
902	17	9	1,88889
903	30	67	0,44776
904	27	20	1,35000
905	16	29	0,55172
906	106	59	1,79661
907	73	94	0,77660
908	153	163	0,93865
909	125	204	0,61275
910	8	11	0,72727
911	50	37	1,35135
912	46	69	0,66667
913	221	138	1,60145
914	75	32	2,34375
915	56	54	1,03704
916	55	51	1,07843
917	27	37	0,72973
918	281	273	1,02930
919	88	72	1,22222
920	89	117	0,76068
921	41	50	0,82000
922	68	67	1,01493
923	33	38	0,86842
924	1057	1128	0,93706
925	35	20	1,75000
926	18	18	1,00000
927	166	220	0,75455
928	53	36	1,47222
929	41	55	0,74545
930	219	250	0,87600
931	17	8	2,12500
932	66	30	2,20000
933	14	16	0,87500
934	24	30	0,80000
935	73	82	0,89024
936	17	32	0,53125
937	81	90	0,90000
938	20	22	0,90909
939	14	14	1,00000
940	47	55	0,85455
941	74	110	0,67273
942	133	39	3,41026
943	18	22	0,81818
944	30	15	2,00000
945	2	1	2,00000
946	97	39	2,48718
947	6	3	2,00000
948	50	43	1,16279
949	6	6	1,00000
950	145	231	0,62771
951	48	41	1,17073
952	93	73	1,27397
953	23	44	0,52273
954	30	47	0,63830
955	70	85	0,82353
956	48	66	0,72727
957	61	48	1,27083
958	32	44	0,72727
959	76	61	1,24590
960	7	8	0,87500
961	28	21	1,33333
962	68	96	0,70833
963	39	44	0,88636
964	93	125	0,74400
965	84	110	0,76364
966	13	6	2,16667
967	103	134	0,76866
968	25	34	0,73529
969	29	26	1,11538
970	14	21	0,66667
971	0	5	0,00000
972	51	35	1,45714
973	94	161	0,58385
974	12	21	0,57143
975	260	391	0,66496
976	21	29	0,72414
977	178	100	1,78000
978	14	24	0,58333
979	27	31	0,87097
980	31	33	0,93939
981	41	26	1,57692
982	143	189	0,75661
983	63	88	0,71591
984	28	29	0,96552
985	49	43	1,13953
986	20	34	0,58824
987	228	231	0,98701
988	42	31	1,35484
989	14	29	0,48276
990	29	37	0,78378
991	83	122	0,68033
992	118	34	3,47059
993	58	83	0,69880
994	20	7	2,85714
995	48	63	0,76190
996	77	84	0,91667
997	103	87	1,18391
998	233	243	0,95885
999	28	31	0,90323
1000	58	55	1,05455

500.000 subsampled reads analyzed			
WGS cluster	ChIP_ Hits	Input_ Hits	ChIP/Input ratio
1	25329	30577	0,82837
2	40103	44689	0,89738
3	20776	18969	1,09526
4	15582	13938	1,11795
5	42039	65219	0,64458
6	94394	130644	0,72253
7	17914	22838	0,78439
8	19436	34950	0,55611
9	636	744	0,85484
10	12000	7050	1,70213
11	9487	15349	0,61809
12	14242	10344	1,37684
13	21587	13280	1,62553
14	800	993	0,80564
15	2036	1867	1,09052
16	2511	3487	0,72010
17	3194	3759	0,84969
18	2268	5240	0,43282
19	6440	4934	1,30523
20	5238	3585	1,46109
21	1356	1804	0,75166
22	2449	2977	0,82264
23	2382	3106	0,76690
24	1154	1169	0,98717
25	1561	2316	0,67401
26	1093	1257	0,86953
27	786	804	0,97761
28	3305	2399	1,37766
29	1040	880	1,18182
30	1580	1814	0,87100
31	476	612	0,77778
32	1189	1256	0,94666
33	1009	1359	0,74246
34	2055	2410	0,85270
35	348	389	0,89460
36	395	506	0,78063
37	1396	1795	0,77772
38	1199	698	1,71777
39	573	639	0,89671
40	727	1025	0,70927
41	1019	1082	0,94177
42	546	569	0,95958
43	472	484	0,97521
44	1203	2773	0,43383
45	1122	1229	0,91294
46	706	1218	0,57964
47	345	367	0,94005
48	5138	2808	1,82977
49	1433	1118	1,28175
50	616	556	1,10791
51	483	505	0,95644
52	705	653	1,07963
53	393	397	0,98992
54	838	810	1,03457
55	483	775	0,62323
56	261	302	0,86424
57	435	490	0,88776
58	269	365	0,73699
59	262	336	0,77976
60	509	750	0,67867
61	699	783	0,89272
62	1560	2311	0,67503
63	273	284	0,96127
64	468	463	1,01080
65	271	308	0,87987
66	400	580	0,68966
67	267	327	0,81651
68	2525	1309	1,92895
69	349	302	1,15563
70	197	227	0,86784
71	394	469	0,84009
72	161	166	0,96988
73	361	251	1,43825
74	158	179	0,88268
75	279	370	0,75405
76	463	515	0,89903
77	306	241	1,26971
78	260	309	0,84142
79	185	179	1,03352
80	601	732	0,82104
81	518	657	0,78843
82	557	786	0,70865
83	126	132	0,95455
84	222	297	0,74747
85	228	247	0,92308
86	580	320	1,81250
87	515	782	0,65857
88	191	182	1,04945
89	250	221	1,13122
90	232	256	0,90625
91	501	357	1,40336
92	379	284	1,33451
93	210	260	0,80769
94	92	97	0,94845
95	372	433	0,85912
96	166	219	0,75799
97	229	259	0,88417
98	231	156	1,48077
99	135	128	1,05469
100	292	354	0,82486
101	301	402	0,74876
102	426	555	0,76757
103	123	122	1,00820
104	233	289	0,80623
105	94	106	0,88679
106	310	373	0,83110
107	122	124	0,98387
108	65	82	0,79268
109	86	98	0,87755
110	142	179	0,79330
111	179	151	1,18543
112	178	162	1,09877
113	252	187	1,34759
114	244	323	0,75542
115	143	184	0,77717
116	189	148	1,27703
117	335	179	1,87151
118	126	142	0,88732
119	147	117	1,25641
120	141	124	1,13710
121	129	178	0,72472
122	89	77	1,15584
123	158	239	0,66109
124	76	131	0,58015
125	116	82	1,41463
126	96	112	0,85714
127	153	157	0,97452
128	54	65	0,83077
129	99	119	0,83193
130	1047	998	1,04910
131	90	107	0,84112
132	112	74	1,51351
133	30	32	0,93750
134	132	124	1,06452
135	33	34	0,97059
136	356	502	0,70916
137	221	294	0,75170
138	68	129	0,52713
139	97	101	0,96040
140	139	155	0,89677
141	222	69	3,21739
142	1174	1010	1,16238
143	231	99	2,33333
144	466	725	0,64276
145	386	508	0,75984
146	126	117	1,07692
147	463	336	1,37798
148	114	152	0,75000
149	81	89	0,91011
150	116	90	1,28889
151	80	117	0,68376
152	40	72	0,55556
153	265	303	0,87459
154	175	144	1,21528
155	130	91	1,42857
156	299	242	1,23554
157	89	171	0,52047
158	132	70	1,88571
159	187	244	0,76639
160	459	399	1,15038
161	81	106	0,76415
162	228	186	1,22581
163	63	94	0,67021
164	130	66	1,96970
165	150	159	0,94340
166	73	68	1,07353
167	220	329	0,66869
168	213	293	0,72696
169	300	99	3,03030
170	72	87	0,82759
171	145	74	1,95946
172	52	80	0,65000

173	198	225	0,88000
174	123	62	1,98387
175	257	284	0,90493
176	185	204	0,90686
177	111	128	0,86719
178	451	502	0,89841
179	108	146	0,73973
180	62	61	1,01639
181	53	62	0,85484
182	62	73	0,84932
183	169	118	1,43220
184	102	72	1,41667
185	21	27	0,77778
186	56	61	0,91803
187	90	113	0,79646
188	57	85	0,67059
189	115	102	1,12745
190	64	84	0,76190
191	51	54	0,94444
192	18	30	0,60000
193	128	79	1,62025
194	103	52	1,98077
195	34	54	0,62963
196	67	107	0,62617
197	229	320	0,71563
198	101	50	2,02000
199	36	42	0,85714
200	77	50	1,54000
201	117	115	1,01739
202	123	158	0,77848
203	92	112	0,82143
204	54	48	1,12500
205	55	35	1,57143
206	33	50	0,66000
207	60	75	0,80000
208	51	59	0,86441
209	49	65	0,75385
210	72	104	0,69231
211	29	41	0,70732
212	25	38	0,65789
213	49	61	0,80328
214	101	110	0,91818
215	52	57	0,91228
216	192	285	0,67368
217	26	40	0,65000
218	188	190	0,98947
219	160	320	0,50000
220	282	464	0,60776
221	105	129	0,81395
222	85	96	0,88542
223	34	41	0,82927
224	80	97	0,82474
225	136	183	0,74317
226	99	50	1,98000
227	33	68	0,48529
228	110	86	1,27907
229	169	85	1,98824
230	462	625	0,73920
231	39	84	0,46429
232	54	67	0,80597
233	49	25	1,96000
234	203	82	2,47561
235	26	61	0,42623
236	56	80	0,70000
237	57	31	1,83871
238	51	63	0,80952
239	121	198	0,61111
240	60	31	1,93548
241	337	288	1,17014
242	121	84	1,44048
243	39	54	0,72222
244	51	39	1,30769
245	445	454	0,98018
246	74	67	1,10448
247	48	49	0,97959
248	46	60	0,76667
249	72	82	0,87805
250	49	21	2,33333
251	32	44	0,72727
252	85	80	1,06250
253	160	58	2,75862
254	46	72	0,63889
255	25	29	0,86207
256	42	48	0,87500
257	53	61	0,86885
258	47	92	0,51087
259	85	97	0,87629
260	51	63	0,80952
261	35	35	1,00000
262	52	65	0,80000
263	33	36	0,91667
264	180	70	2,57143
265	72	112	0,64286
266	79	71	1,11268
267	31	34	0,91176
268	34	46	0,73913
269	61	55	1,10909
270	447	145	3,08276
271	34	47	0,72340
272	63	74	0,85135
273	43	48	0,89583
274	147	221	0,66516
275	109	149	0,73154
276	24	38	0,63158
277	169	251	0,67331
278	70	68	1,02941
279	57	37	1,54054
280	296	329	0,89970
281	158	155	1,01935
282	68	60	1,13333
283	28	42	0,66667
284	89	165	0,53939
285	51	38	1,34211
286	99	140	0,70714
287	494	836	0,59091
288	90	112	0,80357
289	62	42	1,47619
290	96	92	1,04348
291	43	42	1,02381
292	31	42	0,73810
293	24	48	0,50000
294	67	34	1,97059
295	55	42	1,30952
296	51	34	1,50000
297	10	41	0,24390
298	148	68	2,17647
299	71	82	0,86585
300	102	129	0,79070
301	35	37	0,94595
302	95	117	0,81197
303	15	44	0,34091
304	159	88	1,80682
305	39	62	0,62903
306	38	45	0,84444
307	80	41	1,95122
308	25	26	0,96154
309	78	94	0,82979
310	62	60	1,03333
311	135	189	0,71429
312	56	61	0,91803
313	43	25	1,72000
314	92	79	1,16456
315	102	128	0,79688
316	42	66	0,63636
317	67	54	1,24074
318	242	315	0,76825
319	21	29	0,72414
320	67	77	0,87013
321	34	15	2,26667
322	55	65	0,84615
323	61	52	1,17308
324	59	38	1,55263
325	44	38	1,15789
326	62	60	1,03333
327	235	271	0,86716
328	33	32	1,03125
329	49	80	0,61250
330	30	11	2,72727
331	139	213	0,65258
332	76	75	1,01333
333	19	23	0,82609
334	66	98	0,67347
335	69	116	0,59483
336	110	49	2,24490
337	146	169	0,86391
338	30	22	1,36364
339	72	100	0,72000
340	30	35	0,85714
341	26	32	0,81250
342	35	37	0,94595
343	36	41	0,87805
344	42	18	2,33333
345	139	168	0,82738
346	78	67	1,16418
347	57	56	1,01786
348	41	30	1,36667
349	95	105	0,90476
350	74	80	0,92500
351	23	17	1,35294
352	24	45	0,53333
353	29	26	1,11538
354	31	36	0,86111
355	36	37	0,97297
356	155	190	0,81579
357	38	34	1,11765
358	60	54	1,11111
359	36	57	0,63158
360	86	96	0,89583
361	27	43	0,62791
362	195	245	0,79592
363	60	42	1,42857
364	16	33	0,48485
365	40	36	1,11111
366	71	76	0,93421
367	71	121	0,58678
368	15	12	1,25000
369	159	138	1,15217
370	106	126	0,84127
371	87	135	0,64444
372	44	49	0,89796
373	22	42	0,52381
374	163	175	0,93143
375	43	71	0,60563
376	24	19	1,26316
377	123	172	0,71512
378	290	494	0,58704
379	55	75	0,73333
380	55	44	1,25000
381	97	115	0,84348
382	64	74	0,86486
383	19	32	0,59375
384	66	79	0,83544
385	55	29	1,89655
386	40	59	0,67797
387	50	42	1,19048
388	40	45	0,88889
389	31	34	0,91176
390	40	21	1,90476
391	756	356	2,12360
392	120	143	0,83916
393	60	44	1,36364
394	251	73	3,43836
395	23	31	0,74194
396	33	60	0,55000
397	44	31	1,41935
398	18	31	0,58065
399	104	48	2,16667
400	77	104	0,74038
401	59	37	1,59459
402	39	33	1,18182
403	28	46	0,60870
404	15	32	0,46875
405	39	67	0,58209
406	25	35	0,71429
407	25	31	0,80645
408	43	38	1,13158
409	42	64	0,65625
410	136	133	1,02256
411	60	56	1,07143
412	18	17	1,05882
413	24	36	0,66667
414	17	29	0,58621
415	73	190	0,38421
416	65	105	0,61905
417	29	23	1,26087
418	26	39	0,66667
419	103	131	0,78626
420	28	50	0,56000
421	126	138	0,91304
422	68	94	0,72340
423	21	40	0,52500
424	89	124	0,71774
425	49	34	1,44118
426	81	134	0,60448
427	62	59	1,05085
428	44	27	1,62963
429	48	45	1,06667
430	42	28	1,50000
431	29	48	0,60417
432	16	23	0,69565
433	54	53	1,01887
434	54	61	0,88525
435	34	41	0,82927
436	47	75	0,62667
437	245	148	1,65541
438	37	72	0,51389
439	23	44	0,52273
440	115	134	0,85821
441	28	34	0,82353
442	28	20	1,40000
443	45	72	0,62500
444	21	22	0,95455
445	57	29	1,96552
446	29	29	1,00000
447	33	42	0,78571
448	118	118	1,00000
449	25	33	0,75758
450	35	27	1,29630
451	123	155	0,79355
452	18	37	0,48649
453	19	30	0,63333
454	41	52	0,78846
455	113	148	0,76351
456	109	135	0,80741
457	14	30	0,46667
458	22	6	3,66667
459	56	88	0,63636
460	39	20	1,95000
461	71	107	0,66355
462	49	59	0,83051
463	47	42	1,11905
464	95	48	1,97917
465	30	15	2,00000
466	45	23	1,95652
467	48	70	0,68571
468	26	7	3,71429
469	81	22	3,68182
470	51	61	0,83607
471	109	64	1,70313
472	61	37	1,64865
473	13	12	1,08333
474	18	19	0,94737
475	49	60	0,81667
476	33	52	0,63462
477	9	12	0,75000
478	88	101	0,87129
479	156	124	1,25806
480	72	95	0,75789
481	30	27	1,11111
482	10	22	0,45455
483	36	25	1,44000
484	96	59	1,62712
485	32	57	0,56140
486	18	17	1,05882
487	55	34	1,61765
488	57	101	0,56436
489	34	23	1,47826
490	41	51	0,80392
491	22	26	0,84615
492	17	22	0,77273
493	28	38	0,73684
494	45	16	2,81250
495	158	164	0,96341
496	65	93	0,69892
497	19	8	2,37500
498	53	30	1,76667
499	33	42	0,78571
500	47	105	0,44762
501	91	160	0,56875
502	5	11	0,45455
503	28	7	4,00000
504	22	37	0,59459
505	20	34	0,58824
506	23	36	0,63889
507	69	35	1,97143
508	72	51	1,41176
509	21	12	1,75000
510	45	49	0,91837
511	27	16	1,68750
512	35	28	1,25000
513	21	44	0,47727
514	14	14	1,00000
515	27	13	2,07692
516	20	26	0,76923
517	41	27	1,51852
518	73	93	0,78495
519	25	32	0,78125
520	17	33	0,51515
521	77	38	2,02632
522	25	14	1,78571
523	296	368	0,80435
524	47	27	1,74074
525	61	69	0,88406
526	18	37	0,48649
527	38	45	0,84444
528	16	26	0,61538
529	59	52	1,13462
530	79	107	0,73832
531	35	40	0,87500
532	17	22	0,77273
533	62	49	1,26531
534	161	171	0,94152
535	62	79	0,78481
536	12	5	2,40000
537	11	28	0,39286
538	48	24	2,00000
539	31	14	2,21429
540	20	15	1,33333
541	6	18	0,33333
542	35	34	1,02941
543	59	59	1,00000
544	21	33	0,63636
545	80	172	0,46512
546	29	28	1,03571
547	5	2	2,50000
548	52	35	1,48571
549	32	18	1,77778
550	24	60	0,40000
551	18	26	0,69231
552	16	8	2,00000
553	85	111	0,76577
554	225	240	0,93750
555	94	117	0,80342
556	16	20	0,80000
557	67	54	1,24074
558	83	113	0,73451
559	290	396	0,73232
560	83	64	1,29688
561	89	79	1,12658
562	46	45	1,02222
563	27	26	1,03846
564	66	93	0,70968
565	119	143	0,83217
566	12	13	0,92308
567	93	58	1,60345
568	46	62	0,74194
569	126	202	0,62376
570	58	33	1,75758
571	15	14	1,07143
572	28	37	0,75676
573	65	51	1,27451
574	45	46	0,97826
575	182	193	0,94301
576	22	24	0,91667
577	85	74	1,14865
578	27	13	2,07692
579	34	38	0,89474
580	11	20	0,55000
581	74	87	0,85057
582	19	19	1,00000
583	27	31	0,87097
584	27	18	1,50000
585	41	26	1,57692
586	28	25	1,12000
587	5	23	0,21739
588	14	10	1,40000
589	37	57	0,64912
590	94	112	0,83929
591	12	15	0,80000
592	15	10	1,50000
593	16	22	0,72727
594	24	26	0,92308
595	35	22	1,59091
596	78	85	0,91765
597	63	42	1,50000
598	34	32	1,06250
599	57	88	0,64773
600	39	39	1,00000
601	10	11	0,90909
602	27	20	1,35000
603	20	24	0,83333
604	58	20	2,90000
605	23	15	1,53333
606	22	8	2,75000
607	64	69	0,92754
608	27	25	1,08000
609	16	23	0,69565
610	43	13	3,30769
611	120	44	2,72727
612	28	23	1,21739
613	16	17	0,94118
614	67	34	1,97059
615	17	12	1,41667
616	8	25	0,32000
617	52	56	0,92857

618	51	26	1,96154
619	26	27	0,96296
620	18	35	0,51429
621	30	34	0,88235
622	26	26	1,00000
623	5	2	2,50000
624	29	16	1,81250
625	54	86	0,62791
626	25	10	2,50000
627	11	26	0,42308
628	24	35	0,68571
629	75	47	1,59574
630	34	53	0,64151
631	36	38	0,94737
632	196	279	0,70251
633	28	45	0,62222
634	54	35	1,54286
635	17	18	0,94444
636	86	86	1,00000
637	45	23	1,95652
638	42	67	0,62687
639	125	119	1,05042
640	27	29	0,93103
641	54	76	0,71053
642	40	55	0,72727
643	14	20	0,70000
644	16	6	2,66667
645	22	32	0,68750
646	25	28	0,89286
647	27	28	0,96429
648	89	118	0,75424
649	19	12	1,58333
650	20	43	0,46512
651	126	139	0,90647
652	63	71	0,88732
653	39	56	0,69643
654	15	12	1,25000
655	80	108	0,74074
656	9	5	1,80000
657	23	23	1,00000
658	59	40	1,47500
659	32	28	1,14286
660	16	22	0,72727
661	60	97	0,61856
662	23	26	0,88462
663	30	30	1,00000
664	31	33	0,93939
665	27	24	1,12500
666	16	26	0,61538
667	12	22	0,54545
668	7	9	0,77778
669	77	35	2,20000
670	28	42	0,66667
671	58	82	0,70732
672	36	32	1,12500
673	12	11	1,09091
674	15	8	1,87500
675	9	18	0,50000
676	39	75	0,52000
677	36	45	0,80000
678	20	14	1,42857
679	21	27	0,77778
680	119	117	1,01709
681	16	16	1,00000
682	190	152	1,25000
683	169	227	0,74449
684	80	148	0,54054
685	22	20	1,10000
686	16	22	0,72727
687	20	34	0,58824
688	135	157	0,85987
689	9	18	0,50000
690	51	54	0,94444
691	71	102	0,69608
692	40	38	1,05263
693	46	43	1,06977
694	31	40	0,77500
695	29	30	0,96667
696	36	48	0,75000
697	47	51	0,92157
698	28	26	1,07692
699	30	51	0,58824
700	8	15	0,53333
701	35	47	0,74468
702	15	18	0,83333
703	24	18	1,33333
704	13	6	2,16667
705	14	18	0,77778
706	18	13	1,38462
707	23	16	1,43750
708	14	11	1,27273
709	15	28	0,53571
710	62	70	0,88571
711	106	31	3,41935
712	16	28	0,57143
713	19	28	0,67857
714	77	105	0,73333
715	12	32	0,37500
716	20	16	1,25000
717	35	49	0,71429
718	15	20	0,75000
719	58	100	0,58000
720	53	51	1,03922
721	104	112	0,92857
722	86	92	0,93478
723	39	59	0,66102
724	43	56	0,76786
725	66	53	1,24528
726	46	61	0,75410
727	141	205	0,68780
728	12	19	0,63158
729	33	33	1,00000
730	30	41	0,73171
731	17	9	1,88889
732	36	46	0,78261
733	29	23	1,26087
734	22	25	0,88000
735	26	26	1,00000
736	31	43	0,72093
737	1	7	0,14286
738	10	13	0,76923
739	49	71	0,69014
740	16	19	0,84211
741	30	27	1,11111
742	41	47	0,87234
743	61	93	0,65591
744	24	48	0,50000
745	66	135	0,48889
746	38	32	1,18750
747	79	118	0,66949
748	20	28	0,71429
749	11	29	0,37931
750	33	21	1,57143
751	22	20	1,10000
752	17	18	0,94444
753	53	59	0,89831
754	27	54	0,50000
755	67	103	0,65049
756	14	11	1,27273
757	32	59	0,54237
758	59	82	0,71951
759	28	31	0,90323
760	27	34	0,79412
761	172	268	0,64179
762	27	21	1,28571
763	8	11	0,72727
764	33	47	0,70213
765	87	133	0,65414
766	22	26	0,84615
767	27	28	0,96429
768	44	29	1,51724
769	84	160	0,52500
770	35	19	1,84211
771	26	24	1,08333
772	47	84	0,55952
773	22	34	0,64706
774	32	36	0,88889
775	63	69	0,91304
776	4	0	#DIV/0!
777	51	26	1,96154
778	16	8	2,00000
779	24	27	0,88889
780	13	16	0,81250
781	11	8	1,37500
782	33	35	0,94286
783	18	19	0,94737
784	31	67	0,46269
785	16	20	0,80000
786	9	16	0,56250
787	30	59	0,50847
788	26	20	1,30000
789	33	38	0,86842
790	16	18	0,88889
791	29	64	0,45313
792	22	25	0,88000
793	28	15	1,86667
794	12	13	0,92308
795	7	6	1,16667
796	23	35	0,65714
797	43	7	6,14286
798	32	27	1,18519
799	64	43	1,48837
800	28	38	0,73684
801	1	0	#DIV/0!
802	49	57	0,85965
803	100	117	0,85470
804	192	178	1,07865
805	83	76	1,09211
806	22	23	0,95652
807	5	1	5,00000
808	44	67	0,65672
809	43	27	1,59259
810	134	195	0,68718
811	39	38	1,02632
812	24	25	0,96000
813	48	18	2,66667
814	115	207	0,55556
815	49	34	1,44118
816	46	86	0,53488
817	20	29	0,68966
818	17	19	0,89474
819	24	24	1,00000
820	72	100	0,72000
821	17	23	0,73913
822	36	51	0,70588
823	56	75	0,74667
824	52	32	1,62500
825	28	34	0,82353
826	77	44	1,75000
827	102	102	1,00000
828	45	84	0,53571
829	15	35	0,42857
830	34	41	0,82927
831	13	26	0,50000
832	21	35	0,60000
833	25	37	0,67568
834	37	30	1,23333
835	13	17	0,76471
836	54	71	0,76056
837	24	17	1,41176
838	36	26	1,38462
839	12	32	0,37500
840	33	48	0,68750
841	19	23	0,82609
842	174	240	0,72500
843	20	16	1,25000
844	42	56	0,75000
845	26	47	0,55319
846	29	28	1,03571
847	29	33	0,87879
848	22	17	1,29412
849	11	20	0,55000
850	16	25	0,64000
851	33	22	1,50000
852	76	87	0,87356
853	64	51	1,25490
854	55	54	1,01852
855	6	0	#DIV/0!
856	15	13	1,15385
857	6	10	0,60000
858	21	24	0,87500
859	31	35	0,88571
860	6	5	1,20000
861	24	24	1,00000
862	16	33	0,48485
863	29	16	1,81250
864	12	13	0,92308
865	8	26	0,30769
866	10	12	0,83333
867	6	10	0,60000
868	10	10	1,00000
869	18	28	0,64286
870	20	14	1,42857
871	71	75	0,94667
872	21	19	1,10526
873	36	32	1,12500
874	201	241	0,83402
875	64	71	0,90141
876	22	19	1,15789
877	13	23	0,56522
878	18	11	1,63636
879	19	13	1,46154
880	14	25	0,56000
881	18	10	1,80000
882	24	15	1,60000
883	31	55	0,56364
884	19	33	0,57576
885	126	150	0,84000
886	5	2	2,50000
887	73	101	0,72277
888	17	13	1,30769
889	20	20	1,00000
890	16	10	1,60000
891	9	12	0,75000
892	27	54	0,50000
893	19	27	0,70370
894	19	40	0,47500
895	9	8	1,12500
896	3	6	0,50000
897	22	39	0,56410
898	14	12	1,16667
899	21	28	0,75000
900	50	65	0,76923
901	102	118	0,86441
902	3	8	0,37500
903	17	29	0,58621
904	15	8	1,87500
905	6	11	0,54545
906	50	32	1,56250
907	30	52	0,57692
908	85	87	0,97701
909	70	74	0,94595
910	2	1	2,00000
911	29	16	1,81250
912	34	31	1,09677
913	138	90	1,53333
914	31	18	1,72222
915	16	24	0,66667
916	36	22	1,63636
917	10	14	0,71429
918	126	160	0,78750
919	44	38	1,15789
920	43	57	0,75439
921	20	24	0,83333
922	35	47	0,74468
923	19	15	1,26667
924	543	594	0,91414
925	12	9	1,33333
926	12	5	2,40000
927	68	99	0,68687
928	20	10	2,00000
929	9	28	0,32143
930	105	103	1,01942
931	8	7	1,14286
932	23	18	1,27778
933	11	10	1,10000
934	7	21	0,33333
935	25	52	0,48077
936	17	19	0,89474
937	27	47	0,57447
938	10	16	0,62500
939	2	5	0,40000
940	27	25	1,08000
941	40	49	0,81633
942	55	8	6,87500
943	8	12	0,66667
944	12	11	1,09091
945	1	1	1,00000
946	37	10	3,70000
947	6	2	3,00000
948	17	32	0,53125
949	2	3	0,66667
950	81	123	0,65854
951	26	27	0,96296
952	62	36	1,72222
953	8	14	0,57143
954	14	15	0,93333
955	31	50	0,62000
956	23	35	0,65714
957	32	27	1,18519
958	15	15	1,00000
959	35	32	1,09375
960	4	10	0,40000
961	7	14	0,50000
962	27	37	0,72973
963	24	26	0,92308
964	31	53	0,58491
965	47	48	0,97917
966	2	3	0,66667
967	59	60	0,98333
968	9	21	0,42857
969	10	10	1,00000
970	16	7	2,28571
971	3	3	1,00000
972	36	16	2,25000
973	50	78	0,64103
974	13	7	1,85714
975	169	191	0,88482
976	7	3	2,33333
977	61	32	1,90625
978	11	8	1,37500
979	14	19	0,73684
980	13	25	0,52000
981	33	13	2,53846
982	57	101	0,56436
983	44	29	1,51724
984	7	13	0,53846
985	12	16	0,75000
986	20	21	0,95238
987	106	82	1,29268
988	22	12	1,83333
989	14	13	1,07692
990	33	9	3,66667
991	53	53	1,00000
992	42	18	2,33333
993	26	35	0,74286
994	12	5	2,40000
995	25	31	0,80645
996	24	33	0,72727
997	53	46	1,15217
998	94	129	0,72868
999	10	17	0,58824
1000	22	39	0,56410

250.000 subsampled reads analyzed			
WGS cluster	ChIP_ Hits	Input_ Hits	ChIP/Input ratio
1	12534	15455	0,81100
2	20118	22257	0,90390
3	10346	9680	1,06880
4	7983	6835	1,16796
5	20407	32957	0,61920
6	46708	65030	0,71825
7	8836	11152	0,79232
8	9604	17866	0,53756
9	327	370	0,88378
10	5990	3547	1,68875
11	4932	7659	0,64395
12	7507	5047	1,48742
13	11127	6433	1,72968
14	401	472	0,84958
15	1010	945	1,06878
16	1287	1745	0,73754
17	1633	1908	0,85587
18	1250	2645	0,47259
19	3115	2454	1,26936
20	2763	1849	1,49432
21	717	911	0,78705
22	1282	1437	0,89214
23	1238	1514	0,81770
24	560	544	1,02941
25	726	1137	0,63852
26	508	645	0,78760
27	375	391	0,95908
28	1631	1138	1,43322
29	534	431	1,23898
30	775	952	0,81408
31	237	303	0,78218
32	621	573	1,08377
33	516	684	0,75439
34	1025	1205	0,85062
35	169	195	0,86667
36	198	249	0,79518
37	671	949	0,70706
38	633	377	1,67905
39	272	290	0,93793
40	347	504	0,68849
41	488	559	0,87299
42	254	303	0,83828
43	259	236	1,09746
44	626	1411	0,44366
45	589	579	1,01727
46	379	627	0,60447
47	170	183	0,92896
48	2526	1332	1,89640
49	651	581	1,12048
50	293	293	1,00000
51	238	264	0,90152
52	330	312	1,05769
53	177	214	0,82710
54	413	371	1,11321
55	244	376	0,64894
56	125	150	0,83333
57	204	268	0,76119
58	159	174	0,91379
59	138	170	0,81176
60	274	364	0,75275
61	329	368	0,89402
62	700	1090	0,64220
63	135	144	0,93750
64	256	244	1,04918
65	113	134	0,84328
66	194	319	0,60815
67	127	148	0,85811
68	1245	628	1,98248
69	189	164	1,15244
70	93	101	0,92079
71	202	239	0,84519
72	86	84	1,02381
73	150	137	1,09489
74	81	88	0,92045
75	129	190	0,67895
76	228	244	0,93443
77	160	138	1,15942
78	134	154	0,87013
79	98	95	1,03158
80	302	364	0,82967
81	257	306	0,83987
82	264	386	0,68394
83	65	59	1,10169
84	95	147	0,64626
85	109	127	0,85827
86	309	186	1,66129
87	280	417	0,67146
88	88	96	0,91667
89	141	102	1,38235
90	114	140	0,81429
91	236	165	1,43030
92	204	103	1,98058
93	101	131	0,77099
94	39	42	0,92857
95	185	228	0,81140
96	76	93	0,81720
97	136	126	1,07937
98	125	73	1,71233
99	67	72	0,93056
100	156	185	0,84324
101	148	209	0,70813
102	191	292	0,65411
103	72	56	1,28571
104	116	163	0,71166
105	40	49	0,81633
106	158	211	0,74882
107	72	77	0,93506
108	30	40	0,75000
109	55	38	1,44737
110	57	92	0,61957
111	89	79	1,12658
112	92	75	1,22667
113	132	72	1,83333
114	122	177	0,68927
115	70	108	0,64815
116	104	80	1,30000
117	178	87	2,04598
118	52	67	0,77612
119	70	42	1,66667
120	62	70	0,88571
121	64	93	0,68817
122	38	39	0,97436
123	84	110	0,76364
124	43	61	0,70492
125	53	41	1,29268
126	53	66	0,80303
127	65	75	0,86667
128	30	37	0,81081
129	49	53	0,92453
130	529	502	1,05378
131	49	49	1,00000
132	55	41	1,34146
133	11	18	0,61111
134	79	58	1,36207
135	15	18	0,83333
136	162	274	0,59124
137	100	133	0,75188
138	24	75	0,32000
139	50	52	0,96154
140	77	71	1,08451
141	127	42	3,02381
142	657	389	1,68895
143	119	54	2,20370
144	226	323	0,69969
145	132	314	0,42038
146	55	54	1,01852
147	271	187	1,44920
148	45	81	0,55556
149	43	48	0,89583
150	72	39	1,84615
151	43	60	0,71667
152	22	28	0,78571
153	152	138	1,10145
154	88	87	1,01149
155	62	34	1,82353
156	160	105	1,52381
157	45	93	0,48387
158	45	23	1,95652
159	85	72	1,18056
160	227	207	1,09662
161	39	53	0,73585
162	123	95	1,29474
163	35	36	0,97222
164	75	41	1,82927
165	89	53	1,67925
166	20	40	0,50000
167	111	169	0,65680
168	94	129	0,72868
169	139	57	2,43860
170	27	35	0,77143
171	69	41	1,68293
172	17	39	0,43590
173	97	132	0,73485
174	53	27	1,96296
175	130	165	0,78788
176	96	107	0,89720
177	53	51	1,03922
178	228	254	0,89764
179	59	81	0,72840
180	27	26	1,03846
181	28	28	1,00000
182	26	41	0,63415
183	83	71	1,16901
184	62	28	2,21429
185	7	13	0,53846
186	29	31	0,93548
187	46	60	0,76667
188	33	49	0,67347
189	63	42	1,50000
190	28	54	0,51852
191	28	37	0,75676
192	15	23	0,65217
193	58	40	1,45000
194	45	23	1,95652
195	22	32	0,68750
196	34	60	0,56667
197	143	133	1,07519
198	52	26	2,00000
199	16	28	0,57143
200	46	26	1,76923
201	59	61	0,96721
202	67	67	1,00000
203	49	66	0,74242
204	28	29	0,96552
205	25	13	1,92308
206	22	15	1,46667
207	18	40	0,45000
208	29	38	0,76316
209	22	30	0,73333
210	32	57	0,56140
211	15	20	0,75000
212	9	21	0,42857
213	27	31	0,87097
214	55	59	0,93220
215	18	34	0,52941
216	97	134	0,72388
217	13	26	0,50000
218	87	95	0,91579
219	50	136	0,36765
220	148	251	0,58964
221	56	66	0,84848
222	32	42	0,76190
223	20	30	0,66667
224	51	49	1,04082
225	64	108	0,59259
226	53	27	1,96296
227	14	29	0,48276
228	67	40	1,67500
229	81	41	1,97561
230	214	295	0,72542
231	14	49	0,28571
232	29	28	1,03571
233	24	10	2,40000
234	104	40	2,60000
235	16	34	0,47059
236	32	46	0,69565
237	27	16	1,68750
238	31	26	1,19231
239	63	101	0,62376
240	41	21	1,95238
241	204	135	1,51111
242	63	44	1,43182
243	17	23	0,73913
244	27	21	1,28571
245	224	225	0,99556
246	45	31	1,45161
247	27	18	1,50000
248	22	40	0,55000
249	36	48	0,75000
250	15	7	2,14286
251	18	23	0,78261
252	35	31	1,12903
253	85	38	2,23684
254	22	30	0,73333
255	10	13	0,76923
256	15	25	0,60000
257	29	26	1,11538
258	17	19	0,89474
259	39	48	0,81250
260	25	27	0,92593
261	18	13	1,38462
262	15	31	0,48387
263	14	20	0,70000
264	89	36	2,47222
265	31	57	0,54386
266	34	40	0,85000
267	18	14	1,28571
268	20	21	0,95238
269	28	32	0,87500
270	183	83	2,20482
271	18	21	0,85714
272	28	32	0,87500
273	19	21	0,90476
274	76	115	0,66087
275	66	85	0,77647
276	10	20	0,50000
277	95	127	0,74803
278	27	15	1,80000
279	29	15	1,93333
280	158	153	1,03268
281	76	66	1,15152
282	40	26	1,53846
283	10	15	0,66667
284	45	85	0,52941
285	21	10	2,10000
286	52	61	0,85246
287	245	409	0,59902
288	40	59	0,67797
289	29	15	1,93333
290	47	48	0,97917
291	17	22	0,77273
292	16	18	0,88889
293	5	22	0,22727
294	29	15	1,93333
295	21	11	1,90909
296	25	13	1,92308
297	10	15	0,66667
298	77	38	2,02632
299	30	38	0,78947
300	45	78	0,57692
301	11	19	0,57895
302	49	52	0,94231
303	6	14	0,42857
304	67	34	1,97059
305	20	26	0,76923
306	12	17	0,70588
307	53	27	1,96296
308	14	16	0,87500
309	37	53	0,69811
310	30	26	1,15385
311	63	96	0,65625
312	24	45	0,53333
313	21	13	1,61538
314	47	39	1,20513
315	51	74	0,68919
316	19	36	0,52778
317	29	21	1,38095
318	124	157	0,78981
319	8	10	0,80000
320	30	31	0,96774
321	21	8	2,62500
322	26	36	0,72222
323	26	28	0,92857
324	25	21	1,19048
325	27	17	1,58824
326	23	29	0,79310
327	119	124	0,95968
328	16	15	1,06667
329	27	42	0,64286
330	18	5	3,60000
331	61	94	0,64894
332	34	42	0,80952
333	13	11	1,18182
334	34	53	0,64151
335	31	62	0,50000
336	55	24	2,29167
337	66	81	0,81481
338	11	15	0,73333
339	44	56	0,78571
340	11	14	0,78571
341	14	19	0,73684
342	18	24	0,75000
343	16	23	0,69565
344	15	9	1,66667
345	58	78	0,74359
346	39	31	1,25806
347	26	27	0,96296
348	17	17	1,00000
349	44	61	0,72131
350	35	35	1,00000
351	10	10	1,00000
352	16	18	0,88889
353	16	18	0,88889
354	19	15	1,26667
355	20	21	0,95238
356	76	96	0,79167
357	21	24	0,87500
358	37	26	1,42308
359	21	32	0,65625
360	42	37	1,13514
361	13	15	0,86667
362	82	122	0,67213
363	36	23	1,56522
364	9	17	0,52941
365	17	17	1,00000
366	38	38	1,00000
367	44	63	0,69841
368	9	9	1,00000
369	85	63	1,34921
370	50	54	0,92593
371	36	56	0,64286
372	28	34	0,82353
373	9	19	0,47368
374	92	78	1,17949
375	23	38	0,60526
376	10	15	0,66667
377	64	88	0,72727
378	153	226	0,67699
379	34	35	0,97143
380	28	28	1,00000
381	56	52	1,07692
382	33	32	1,03125
383	10	20	0,50000
384	33	45	0,73333
385	28	20	1,40000
386	20	34	0,58824
387	25	23	1,08696
388	13	27	0,48148
389	16	14	1,14286
390	21	9	2,33333
391	363	156	2,32692
392	63	71	0,88732
393	21	19	1,10526
394	144	29	4,96552
395	14	10	1,40000
396	13	26	0,50000
397	18	20	0,90000
398	9	10	0,90000
399	46	21	2,19048
400	39	51	0,76471
401	36	24	1,50000
402	26	19	1,36842
403	14	30	0,46667
404	8	12	0,66667
405	18	27	0,66667
406	17	20	0,85000
407	10	16	0,62500
408	32	17	1,88235
409	23	37	0,62162
410	56	74	0,75676
411	36	27	1,33333
412	4	11	0,36364
413	16	16	1,00000
414	11	8	1,37500
415	32	102	0,31373
416	36	62	0,58065
417	12	13	0,92308
418	17	18	0,94444
419	51	57	0,89474
420	14	27	0,51852
421	66	80	0,82500
422	39	44	0,88636
423	13	19	0,68421
424	48	69	0,69565
425	24	20	1,20000
426	35	71	0,49296
427	32	29	1,10345
428	24	19	1,26316
429	31	24	1,29167
430	26	14	1,85714
431	23	25	0,92000
432	6	16	0,37500
433	26	39	0,66667
434	26	23	1,13043
435	21	24	0,87500
436	30	34	0,88235
437	89	45	1,97778

438	16	40	0,40000
439	10	25	0,40000
440	51	71	0,71831
441	13	22	0,59091
442	12	11	1,09091
443	28	37	0,75676
444	10	12	0,83333
445	29	15	1,93333
446	12	23	0,52174
447	16	18	0,88889
448	68	55	1,23636
449	14	11	1,27273
450	14	14	1,00000
451	66	80	0,82500
452	12	22	0,54545
453	12	14	0,85714
454	17	24	0,70833
455	58	73	0,79452
456	58	56	1,03571
457	9	16	0,56250
458	13	3	4,33333
459	28	46	0,60870
460	23	12	1,91667
461	38	63	0,60317
462	21	34	0,61765
463	24	18	1,33333
464	39	23	1,69565
465	16	8	2,00000
466	19	10	1,90000
467	21	42	0,50000
468	11	5	2,20000
469	44	5	8,80000
470	28	22	1,27273
471	42	28	1,50000
472	25	15	1,66667
473	7	9	0,77778
474	8	6	1,33333
475	21	26	0,80769
476	20	12	1,66667
477	4	9	0,44444
478	41	44	0,93182
479	61	66	0,92424
480	30	56	0,53571
481	15	11	1,36364
482	3	13	0,23077
483	21	12	1,75000
484	47	25	1,88000
485	20	21	0,95238
486	9	7	1,28571
487	11	6	1,83333
488	17	35	0,48571
489	16	18	0,88889
490	24	22	1,09091
491	7	12	0,58333
492	6	10	0,60000
493	14	12	1,16667
494	21	8	2,62500
495	55	72	0,76389
496	40	43	0,93023
497	10	6	1,66667
498	21	11	1,90909
499	17	29	0,58621
500	25	46	0,54348
501	39	80	0,48750
502	4	7	0,57143
503	10	3	3,33333
504	9	20	0,45000
505	5	17	0,29412
506	5	21	0,23810
507	27	14	1,92857
508	39	26	1,50000
509	11	4	2,75000
510	23	19	1,21053
511	13	8	1,62500
512	16	12	1,33333
513	12	24	0,50000
514	6	7	0,85714
515	13	2	6,50000
516	13	17	0,76471
517	23	9	2,55556
518	35	34	1,02941
519	12	19	0,63158
520	8	19	0,42105
521	40	21	1,90476
522	13	4	3,25000
523	150	158	0,94937
524	23	12	1,91667
525	32	24	1,33333
526	11	17	0,64706
527	9	18	0,50000
528	10	12	0,83333
529	44	33	1,33333
530	46	67	0,68657
531	17	16	1,06250
532	11	5	2,20000
533	29	32	0,90625
534	62	86	0,72093
535	26	36	0,72222
536	6	3	2,00000
537	7	6	1,16667
538	19	7	2,71429
539	8	8	1,00000
540	10	5	2,00000
541	3	11	0,27273
542	20	13	1,53846
543	36	32	1,12500
544	13	19	0,68421
545	39	83	0,46988
546	14	21	0,66667
547	2	1	2,00000
548	24	19	1,26316
549	19	8	2,37500
550	15	40	0,37500
551	11	18	0,61111
552	11	2	5,50000
553	38	46	0,82609
554	106	122	0,86885
555	38	57	0,66667
556	9	12	0,75000
557	39	24	1,62500
558	44	57	0,77193
559	143	197	0,72589
560	36	35	1,02857
561	37	43	0,86047
562	17	17	1,00000
563	14	9	1,55556
564	25	44	0,56818
565	65	59	1,10169
566	6	5	1,20000
567	54	18	3,00000
568	27	26	1,03846
569	61	89	0,68539
570	31	16	1,93750
571	10	9	1,11111
572	14	20	0,70000
573	30	26	1,15385
574	20	24	0,83333
575	75	101	0,74257
576	10	19	0,52632
577	44	41	1,07317
578	20	11	1,81818
579	10	26	0,38462
580	10	14	0,71429
581	35	43	0,81395
582	12	8	1,50000
583	15	18	0,83333
584	15	5	3,00000
585	21	11	1,90909
586	11	9	1,22222
587	4	7	0,57143
588	7	2	3,50000
589	15	28	0,53571
590	59	55	1,07273
591	5	6	0,83333
592	7	6	1,16667
593	8	17	0,47059
594	10	15	0,66667
595	12	7	1,71429
596	39	53	0,73585
597	31	18	1,72222
598	23	13	1,76923
599	33	61	0,54098
600	21	18	1,16667
601	3	7	0,42857
602	14	6	2,33333
603	8	11	0,72727
604	23	3	7,66667
605	11	2	5,50000
606	8	4	2,00000
607	35	43	0,81395
608	13	17	0,76471
609	4	8	0,50000
610	16	6	2,66667
611	66	22	3,00000
612	21	16	1,31250
613	8	11	0,72727
614	34	22	1,54545
615	7	4	1,75000
616	3	17	0,17647
617	23	23	1,00000
618	30	16	1,87500
619	12	16	0,75000
620	8	18	0,44444
621	10	14	0,71429
622	16	9	1,77778
623	1	0	#DIV/0!
624	11	9	1,22222
625	20	39	0,51282
626	13	8	1,62500
627	6	10	0,60000
628	11	19	0,57895
629	21	29	0,72414
630	13	38	0,34211
631	18	12	1,50000
632	109	137	0,79562
633	12	24	0,50000
634	26	18	1,44444
635	9	6	1,50000
636	44	47	0,93617
637	25	13	1,92308
638	25	22	1,13636
639	60	60	1,00000
640	9	13	0,69231
641	15	36	0,41667
642	23	29	0,79310
643	9	11	0,81818
644	1	2	0,50000
645	9	21	0,42857
646	14	12	1,16667
647	13	16	0,81250
648	41	64	0,64063
649	13	9	1,44444
650	8	21	0,38095
651	56	81	0,69136
652	29	43	0,67442
653	18	28	0,64286
654	10	3	3,33333
655	28	59	0,47458
656	2	3	0,66667
657	13	14	0,92857
658	29	25	1,16000
659	10	20	0,50000
660	9	12	0,75000
661	35	44	0,79545
662	11	15	0,73333
663	15	12	1,25000
664	15	13	1,15385
665	10	15	0,66667
666	7	13	0,53846
667	8	5	1,60000
668	2	7	0,28571
669	47	21	2,23810
670	15	21	0,71429
671	34	51	0,66667
672	20	27	0,74074
673	10	1	10,00000
674	6	2	3,00000
675	4	11	0,36364
676	27	37	0,72973
677	16	27	0,59259
678	5	10	0,50000
679	11	10	1,10000
680	67	50	1,34000
681	9	6	1,50000
682	96	58	1,65517
683	68	119	0,57143
684	46	75	0,61333
685	15	10	1,50000
686	10	13	0,76923
687	7	15	0,46667
688	69	90	0,76667
689	4	5	0,80000
690	20	40	0,50000
691	36	71	0,50704
692	28	17	1,64706
693	18	24	0,75000
694	15	22	0,68182
695	9	19	0,47368
696	18	26	0,69231
697	25	34	0,73529
698	17	10	1,70000
699	12	24	0,50000
700	3	11	0,27273
701	16	21	0,76190
702	11	8	1,37500
703	15	8	1,87500
704	7	4	1,75000
705	9	8	1,12500
706	12	8	1,50000
707	3	8	0,37500
708	5	1	5,00000
709	6	14	0,42857
710	26	32	0,81250
711	60	14	4,28571
712	7	10	0,70000
713	10	16	0,62500
714	45	50	0,90000
715	8	22	0,36364
716	8	3	2,66667
717	12	21	0,57143
718	8	7	1,14286
719	30	44	0,68182
720	34	25	1,36000
721	43	50	0,86000
722	37	52	0,71154
723	17	33	0,51515
724	23	29	0,79310
725	28	27	1,03704
726	26	26	1,00000
727	79	89	0,88764
728	4	11	0,36364
729	15	13	1,15385
730	16	24	0,66667
731	8	0	#DIV/0!
732	12	24	0,50000
733	15	15	1,00000
734	13	15	0,86667
735	10	14	0,71429
736	20	23	0,86957
737	0	0	#DIV/0!
738	4	6	0,66667
739	23	41	0,56098
740	8	11	0,72727
741	9	19	0,47368
742	19	23	0,82609
743	24	48	0,50000
744	11	19	0,57895
745	29	67	0,43284
746	19	16	1,18750
747	38	62	0,61290
748	12	14	0,85714
749	5	13	0,38462
750	16	12	1,33333
751	8	5	1,60000
752	11	8	1,37500
753	26	33	0,78788
754	16	28	0,57143
755	29	53	0,54717
756	2	8	0,25000
757	14	25	0,56000
758	20	35	0,57143
759	16	11	1,45455
760	6	16	0,37500
761	81	121	0,66942
762	11	9	1,22222
763	3	4	0,75000
764	18	25	0,72000
765	37	71	0,52113
766	10	17	0,58824
767	15	13	1,15385
768	24	14	1,71429
769	35	80	0,43750
770	13	13	1,00000
771	8	15	0,53333
772	24	44	0,54545
773	15	17	0,88235
774	14	16	0,87500
775	38	41	0,92683
776	1	0	#DIV/0!
777	23	8	2,87500
778	9	2	4,50000
779	8	15	0,53333
780	8	6	1,33333
781	5	5	1,00000
782	20	19	1,05263
783	10	12	0,83333
784	11	31	0,35484
785	9	7	1,28571
786	2	6	0,33333
787	18	29	0,62069
788	11	10	1,10000
789	13	18	0,72222
790	9	7	1,28571
791	20	35	0,57143
792	10	16	0,62500
793	14	7	2,00000
794	4	5	0,80000
795	4	3	1,33333
796	10	14	0,71429
797	26	3	8,66667
798	21	11	1,90909
799	27	14	1,92857
800	12	21	0,57143
801	1	0	#DIV/0!
802	16	24	0,66667
803	50	55	0,90909
804	87	85	1,02353
805	49	41	1,19512
806	15	9	1,66667
807	4	0	#DIV/0!
808	27	26	1,03846
809	23	12	1,91667
810	66	83	0,79518
811	18	18	1,00000
812	6	14	0,42857
813	23	8	2,87500
814	65	108	0,60185
815	24	19	1,26316
816	16	35	0,45714
817	9	12	0,75000
818	9	10	0,90000
819	14	10	1,40000
820	43	56	0,76786
821	5	11	0,45455
822	19	29	0,65517
823	24	27	0,88889
824	27	18	1,50000
825	13	17	0,76471
826	33	26	1,26923
827	51	52	0,98077
828	27	40	0,67500
829	5	18	0,27778
830	19	15	1,26667
831	3	13	0,23077
832	11	19	0,57895
833	18	18	1,00000
834	17	18	0,94444
835	7	8	0,87500
836	25	45	0,55556
837	8	7	1,14286
838	17	13	1,30769
839	9	23	0,39130
840	19	21	0,90476
841	10	15	0,66667
842	93	128	0,72656
843	13	10	1,30000
844	22	21	1,04762
845	11	21	0,52381
846	10	11	0,90909
847	10	15	0,66667
848	14	5	2,80000
849	4	7	0,57143
850	8	12	0,66667
851	12	9	1,33333
852	40	49	0,81633
853	30	23	1,30435
854	32	28	1,14286
855	5	0	#DIV/0!
856	6	8	0,75000
857	3	8	0,37500
858	12	9	1,33333
859	15	17	0,88235
860	3	3	1,00000
861	18	15	1,20000
862	6	10	0,60000
863	17	12	1,41667
864	8	6	1,33333
865	7	16	0,43750
866	9	7	1,28571
867	4	4	1,00000
868	10	0	#DIV/0!
869	8	13	0,61538

870	8	9	0,88889
871	28	35	0,80000
872	12	10	1,20000
873	18	15	1,20000
874	103	108	0,95370
875	34	21	1,61905
876	16	10	1,60000
877	10	15	0,66667
878	11	7	1,57143
879	7	10	0,70000
880	10	11	0,90909
881	10	5	2,00000
882	10	7	1,42857
883	15	24	0,62500
884	10	13	0,76923
885	63	59	1,06780
886	5	0	#DIV/0!
887	33	54	0,61111
888	3	8	0,37500
889	9	18	0,50000
890	5	7	0,71429
891	4	8	0,50000
892	10	23	0,43478
893	10	8	1,25000
894	12	14	0,85714
895	3	2	1,50000
896	2	3	0,66667
897	11	19	0,57895
898	4	4	1,00000
899	10	13	0,76923
900	23	36	0,63889
901	69	55	1,25455
902	2	0	#DIV/0!
903	8	8	1,00000
904	7	5	1,40000
905	2	2	1,00000
906	23	12	1,91667
907	19	37	0,51351
908	32	36	0,88889
909	26	33	0,78788
910	2	0	#DIV/0!
911	9	3	3,00000
912	15	24	0,62500
913	56	49	1,14286
914	16	6	2,66667
915	12	20	0,60000
916	21	18	1,16667
917	4	6	0,66667
918	62	66	0,93939
919	11	22	0,50000
920	19	35	0,54286
921	16	15	1,06667
922	14	30	0,46667
923	9	6	1,50000
924	268	288	0,93056
925	5	7	0,71429
926	4	3	1,33333
927	40	43	0,93023
928	12	1	12,00000
929	4	17	0,23529
930	48	50	0,96000
931	4	5	0,80000
932	6	9	0,66667
933	5	6	0,83333
934	4	10	0,40000
935	13	28	0,46429
936	13	7	1,85714
937	15	24	0,62500
938	5	10	0,50000
939	0	4	0,00000
940	18	9	2,00000
941	13	22	0,59091
942	54	6	9,00000
943	1	10	0,10000
944	5	6	0,83333
945	0	0	#DIV/0!
946	25	7	3,57143
947	1	0	#DIV/0!
948	10	12	0,83333
949	2	1	2,00000
950	47	66	0,71212
951	12	11	1,09091
952	37	14	2,64286
953	5	7	0,71429
954	4	3	1,33333
955	15	33	0,45455
956	8	22	0,36364
957	16	12	1,33333
958	2	10	0,20000
959	17	10	1,70000
960	2	6	0,33333
961	5	4	1,25000
962	10	13	0,76923
963	15	12	1,25000
964	15	31	0,48387
965	27	30	0,90000
966	0	2	0,00000
967	31	27	1,14815
968	7	14	0,50000
969	6	6	1,00000
970	10	6	1,66667
971	1	0	#DIV/0!
972	21	7	3,00000
973	24	38	0,63158
974	7	4	1,75000
975	98	90	1,08889
976	5	2	2,50000
977	41	26	1,57692
978	6	3	2,00000
979	5	9	0,55556
980	10	13	0,76923
981	15	5	3,00000
982	34	43	0,79070
983	17	13	1,30769
984	5	6	0,83333
985	6	6	1,00000
986	13	14	0,92857
987	54	40	1,35000
988	11	4	2,75000
989	7	6	1,16667
990	14	2	7,00000
991	25	32	0,78125
992	23	11	2,09091
993	13	18	0,72222
994	5	5	1,00000
995	16	14	1,14286
996	13	15	0,86667
997	21	29	0,72414
998	48	60	0,80000
999	9	3	3,00000
1000	6	15	0,40000

Criteria for cCENH3 ChIP enrichment:
ChIP/Input ratio >1 AND >1% analyzed reads in ChIP hits
ChIP/Input ratio >2 AND >0.01% analyzed reads in ChIP hits

Acknowledgments

We sincerely thank Jiří Macas for valuable discussions on ChIP strategies and helpful comments on the manuscript. We are truly grateful to Merrilee Susan Haas (USDA, ARS) for kindly providing the *Tribolium castaneum* starter culture. We are indebted to Lucija Horvat for the confocal microscopy assistance.

References

1. McKinley KL, Cheeseman IM. The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol. 2016;17: 16–29. doi:10.1038/nrm.2015.5
2. Muller H, Gil J, Drinnenberg IA. The Impact of Centromeres on Spatial Genome Architecture. Trends Genet. 2019;35: 565–578. doi:10.1016/j.tig.2019.05.003
3. Plohl M, Meštrović N, Mravinac B. Centromere identity from the DNA point of view. Chromosoma. 2014;123: 313–325. doi:10.1007/s00412-014-0462-0
4. Melters DP, Paliulis L V., Korf IF, Chan SWL. Holocentric chromosomes: Convergent evolution, meiotic adaptations, and genomic analysis. Chromosom Res. 2012;20: 579–593. doi:10.1007/s10577-012-9292-1
5. Schvarzstein M, Wignall SM, Villeneuve AM. Coordinating cohesion, co-orientation, and congression during meiosis: Lessons from holocentric chromosomes. Genes Dev. 2010;24: 219–228. doi:10.1101/gad.1863610
6. Drinnenberg IA, DeYoung D, Henikoff S, Malik HS. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. Elife. 2014;3: e03676. doi:10.7554/eLife.03676
7. Neumann P, Navrátilová A, Schroeder-Reiter E, Koblížková A, Steinbauerová V, Chocholová E, et al. Stretching the rules: Monocentric chromosomes with multiple centromere domains. PLoS Genet. 2012;8: e1002777. doi:10.1371/journal.pgen.1002777
8. Neumann P, Pavlíková Z, Koblížková A, Fuková I, Jedličková V, Novák P, et al. Centromeres off the hook: Massive changes in centromere size and structure following duplication of cenh3 gene in fabeae species. Mol Biol Evol. 2015;32: 1862–1879. doi:10.1093/molbev/msv070
9. Naughton C, Gilbert N. Centromere chromatin structure - Lessons from neocentromeres. Exp Cell Res. 2020;389: 111899. doi:10.1016/j.yexcr.2020.111899
10. Logsdon GA, Gambogi CW, Liskovykh MA, Barrey EJ, Larionov V, Miga KH, et al. Human Artificial Chromosomes that Bypass Centromeric DNA. Cell. 2019;178: 624-639.e19. doi:https://doi.org/10.1016/j.cell.2019.06.006
11. Gambogi CW, Black BE. The nucleosomes that mark centromere location on chromosomes old and new. Essays Biochem. 2019;63: 15–27. doi:10.1042/EBC20180060

12. Earnshaw WC, Rothfield N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma. 1985;91: 313–321. doi:10.1007/BF00328227
13. Henikoff S, Thakur J, Kasinathan S, Talbert PB. Remarkable Evolutionary Plasticity of Centromeric Chromatin. Cold Spring Harb Symp Quant Biol. 2017;82: 71–82. doi:10.1101/sqb.2017.82.033605
14. Malik HS, Henikoff S. Phylogenomics of the nucleosome. Nat Struct Biol. 2003;10: 882–891. doi:10.1038/nsb996
15. Akiyoshi B, Gull K. Discovery of unconventional kinetochores in kinetoplastids. Cell. 2014;156: 1247–1258. doi:10.1016/j.cell.2014.01.049
16. Navarro-Mendoza MI, Pérez-Arques C, Panchal S, Nicolás FE, Mondo SJ, Ganguly P, et al. Early Diverging Fungus Mucor circinelloides Lacks Centromeric Histone CENP-A and Displays a Mosaic of Point and Regional Centromeres. Curr Biol. 2019;29: 3791-3802.e6. doi:10.1016/j.cub.2019.09.024
17. Henikoff S, Ahmad K, Malik HS. The centromere paradox: Stable inheritance with rapidly evolving DNA. Science (80-). 2001;293: 1098–1102. doi:10.1126/science.1062939
18. Malik HS. The centromere-drive hypothesis: a simple basis for centromere complexity. Prog Mol Subcell Biol. 2009;48: 33–52. doi:10.1007/978-3-642-00182-6_2
19. Sokoloff A. The Biology of *Tribolium*, With Special Emphasis on Genetic Aspects, Vol. 1. Clarendon Press, Oxford; 1972.
20. Grove SJ, Stork NE. An inordinate fondness for beetles. Invertebr Syst. 2000;14: 733–739.
21. Brown SJ, Shippy TD, Miller S, Bolognesi R, Beeman RW, Lorenzen MD, et al. The red flour beetle, *Tribolium castaneum* (Coleoptera): A model for studies of development and pest biology. Cold Spring Harb Protoc. 2009;4. doi:10.1101/pdb.emo126
22. Adamski Z, Bufo SA, Chowański S, Falabella P, Lubawy J, Marciniak P, et al. Beetles as model organisms in physiological, biomedical and environmental studies - A review. Front Physiol. 2019;10: 1–22. doi:10.3389/fphys.2019.00319
23. Richards S, Gibbs RA, Weinstock GM, Brown S, Denell R, Beeman RW, et al. The genome of the model beetle and pest *Tribolium castaneum*. Nature. 2008;452: 949–955. doi:10.1038/nature06784
24. Herndon N, Shelton J, Gerischer L, Ioannidis P, Ninova M, Dönitz J, et al. Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. BMC Genomics. 2020;21: 47. doi:10.1186/s12864-019-6394-6
25. Ugarković D, Podnar M, Plohl M. Satellite DNA of the red flour beetle *Tribolium castaneum* - Comparative study of satellites from the genus *Tribolium*. Mol Biol Evol. 1996;13: 1059–1066. doi:10.1093/oxfordjournals.molbev.a025668
26. Wang S, Lorenzen MD, Beeman RW, Brown SJ. Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. Genome Biol. 2008;9: 1–14. doi:10.1186/gb-2008-9-3-r61

27. Brown SJ, Henry JK, Black IV WC, Denell RE. Molecular genetic manipulation of the red flour beetle: Genome organization and cloning of a ribosomal protein gene. Insect Biochem. 1990;20: 185–193. doi:10.1016/0020-1790(90)90011-I
28. Pavlek M, Gelfand Y, Plohl M, Meštrović N. Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. DNA Res. 2015;22: 387–401. doi:10.1093/dnares/dsv021
29. Khan SA, Eggleston H, Myles KM, Adelman ZN. Differentially and Co-expressed Genes in Embryo, Germ-Line and Somatic Tissues of *Tribolium castaneum*. G3 Genes|Genomes|Genetics. 2019;9: 2363–2373. doi:10.1534/g3.119.400340
30. Stuart JJ, Mocelin G. Cytogenetics of chromosome rearrangements in *Tribolium castaneum*. Genome. 1995;38: 673–680. doi:10.1139/g95-085
31. Novák P, Neumann P, Pech J, Steinhaisl J, MacAs J. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013;29: 792–793. doi:10.1093/bioinformatics/btt054
32. Osanai-Futahashi M, Fujiwara H. Coevolution of telomeric repeats and telomeric repeat-specific non-LTR retrotransposons in insects. Mol Biol Evol. 2011;28: 2983–2986. doi:10.1093/molbev/msr135
33. Malik HS, Henikoff S. Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. Genetics. 2001;157: 1293–1298.
34. Buchwitz BJ, Ahmad K, Moore LL, Roth MB, Henikoff S. A histone-H3-like protein in *C. elegans*. Nature. 1999;401: 547–548. doi:10.1038/44062
35. Stoler S, Keith KC, Curnick KE, Fitzgerald-Hayes M. A mutation in CSE4, an essential gene encoding a novel chromatin- associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev. 1995;9: 573–586. doi:10.1101/gad.9.5.573
36. Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang J. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres. Genetics. 2003;163: 1221–1225.
37. Lam AL, Boivin CD, Bonney CF, Rudd MK, Sullivan BA. Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. Proc Natl Acad Sci U S A. 2006;103: 4186–4191. doi:10.1073/pnas.0507947103
38. Talbert PB, Sivakanthan K, Henikoff S. Simple and Complex Centromeric Satellites in *Drosophila* Sibling Species. Genetics. 2018;208: 977–990. doi:10.1534/genetics.117.300620/-/DC1.1
39. Maheshwari S, Ishii T, Brown CT, Houben A, Comai L. Centromere location in Arabidopsis is unaltered by extreme divergence in CENH3 protein sequence. Genome Res. 2017;27: 471–478. doi:10.1101/gr.214619.116
40. Henikoff JG, Thakur J, Kasinathan S, Henikoff S. A unique chromatin complex occupies young a-satellite arrays of human centromeres. Sci Adv. 2015;1: 1–12. doi:10.1126/sciadv.1400234

41. Vlahović I, Gluncić M, Rosandić M, Ugarković Đ, Paar V. Regular higher order repeat structures in beetle *Tribolium castaneum* genome. Genome Biol Evol. 2017;9: 2668–2680. doi:10.1093/gbe/evw174
42. Hartley G, O'Neill RJ. Centromere repeats: Hidden gems of the genome. Genes (Basel). 2019;10: 223. doi:10.3390/genes10030223
43. Chang CH, Chavan A, Palladino J, Wei X, Martins NMC, Santinello B, et al. Islands of retroelements are major components of *Drosophila* centromeres. PLoS Biol. 2019;17: 1–40. doi:10.1371/journal.pbio.3000241
44. de Sotero-Caio CG, Cabral-de-Mello DC, Calixto M da S, Valente GT, Martins C, Loreto V, et al. Centromeric enrichment of LINE-1 retrotransposons and its significance for the chromosome evolution of Phyllostomid bats. Chromosom Res. 2017;25: 313–325. doi:10.1007/s10577-017-9565-9
45. Presting GG. Centromeric retrotransposons and centromere function. Curr Opin Genet Dev. 2018;49: 79–84. doi:10.1016/j.gde.2018.03.004
46. Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, et al. RNAi is a critical determinant of centromere evolution in closely related fungi. Proc Natl Acad Sci U S A. 2018;115: 3108–3113. doi:10.1073/pnas.1713725115
47. Vondrak T, Avila Robledillo L, Novak P, Koblizkova A, Neumann P, Macas J. Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. Plant J. 2020;101: 484–500. doi:10.1111/tpj.14546
48. Bolzán AD. Interstitial telomeric sequences in vertebrate chromosomes: Origin, function, instability and evolution. Mutat Res. 2017;773: 51–65. doi:10.1016/j.mrrev.2017.04.002
49. He L, Liu J, Torres GA, Zhang H, Jiang J, Xie C. Interstitial telomeric repeats are enriched in the centromeres of chromosomes in Solanum species. Chromosom Res. 2013;21: 5–13. doi:10.1007/s10577-012-9332-x
50. Macas J, Navrátilová A, Mészáros T. Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on Vicia sativa chromosomes. Chromosoma. 2003;112: 152–158. doi:10.1007/s00412-003-0255-3
51. Kumke K, Macas J, Fuchs J, Altschmied L, Kour J, Dhar MK, et al. Plantago lagopus B Chromosome is enriched in 5S rDNA-derived satellite DNA. Cytogenet Genome Res. 2016;148: 68–73. doi:10.1159/000444873
52. Vittorazzi SE, Lourenço LB, Del-Grande ML, Recco-Pimentel SM. Satellite DNA derived from 5S rDNA in Physalaemus cuvieri (Anura, Leiuperidae). Cytogenet Genome Res. 2011;134: 101–107. doi:10.1159/000325540
53. Yang X, Zhao H, Zhang T, Zeng Z, Zhang P, Zhu B, et al. Amplification and adaptation of centromeric repeats in polyploid switchgrass species. New Phytol. 2018;218: 1645–1657. doi:10.1111/nph.15098
54. Huang YC, Lee CC, Kao CY, Chang NC, Lin CC, Shoemaker D, et al. Evolution of long centromeres in fire ants. BMC Evol Biol. 2016;16: 1–14. doi:10.1186/s12862-016-0760-7

55. Brinkley BR, Valdivia MM, Tousson A, Brenner SL. Compound kinetochores of the Indian muntjac. Chromosoma. 1984;91: 1–11. doi:10.1007/bf00286479
56. Chi JX, Huang L, Nie W, Wang J, Su B, Yang F. Defining the orientation of the tandem fusions that occurred during the evolution of Indian muntjac chromosomes by BAC mapping. Chromosoma. 2005;114: 167–172. doi:10.1007/s00412-005-0004-x
57. Kursel LE, Malik HS. The cellular mechanisms and consequences of centromere drive. Curr Opin Cell Biol. 2018;52: 58–65. doi:10.1016/j.ceb.2018.01.011
58. Iwata-Otsubo A, Dawicki-McKenna JM, Akera T, Falk SJ, Chmátal L, Yang K, et al. Expanded Satellite Repeats Amplify a Discrete CENP-A Nucleosome Assembly Site on Chromosomes that Drive in Female Meiosis. Curr Biol. 2017;27: 2365-2373.e8. doi:10.1016/j.cub.2017.06.069
59. Talbert PB, Henikoff S. What makes a centromere? Exp Cell Res. 2020;389: 111895. doi:10.1016/j.yexcr.2020.111895
60. Cortes-Silva N, Ulmer J, Kiuchi T, Hsieh E, Cornilleau G, Ladid I, et al. CenH3-Independent Kinetochore Assembly in Lepidoptera Requires CCAN, Including CENP-T. Curr Biol. 2020;30: 561–572. doi:10.1016/j.cub.2019.12.014
61. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9: 671–675. doi:10.1038/nmeth.2089
62. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2016;44: D67–D72. doi:10.1093/nar/gkv1276
63. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6: 4–9. doi:10.1186/s13100-015-0041-9

Isolation of High Molecular Weight DNA from the Model Beetle *Tribolium* for Nanopore Sequencing

Marin Volarić, Damira Veseljak, Brankica Mravinac, Nevenka Meštrović and Evelin Despot-Slade*

Division of Molecular Biology, Ruder Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia

* Correspondence: evelin.despot.slade@irb.hr

Genes

July, 2020

12(8), 1114

Abstract

The long-read Nanopore sequencing has been recently applied for assembly of complex genomes and analysis of linear genome organization. The most critical factor for successful long-read sequencing is extraction of high molecular weight (HMW) DNA of sufficient purity and quantity. The challenges associated with input DNA quality are further amplified when working with extremely small insects with hard exoskeletons. Here, we optimized the isolation of HMW DNA from the model beetle *Tribolium* and tested for use in Nanopore sequencing. We succeeded in overcoming all the difficulties in HMW handling and library preparation that were encountered when using published protocols and commercial kits. Isolation of nuclei and subsequent purification of DNA on an anion-exchange chromatography column resulted in genomic HMW DNA that was efficiently relaxed, of optimal quality and in sufficient quantity for Nanopore MinION sequencing. DNA shearing increased average N50 read values up to 26 kb and allowed us to use a single flow cell in multiple library loads for a total output of more than 13 Gb. Although our focus was on *T. castaneum* and closely related species, we expect that this protocol, with appropriate modifications, could be extended to other insects, particularly beetles.

1. Introduction

The beetle *Tribolium castaneum* has become one of the most important models in the field of evolution, physiology and development because its development is more representative for insects compared to *Drosophila* [1]. It is a worldwide pest of stored products and represents the most species-rich animal order on Earth, the coleopterans. *T. castaneum* genome has been sequenced, annotated and a reference genome is available [2,3]. The estimated genome size of 204 Mb is 44 Mb larger than the assembled genome sequence [4] suggesting that almost 25% of the genome remains unassembled. It is not unexpected taking into account that structural genome analyses revealed large amounts of different classes of repetitive DNA especially in abundant (peri) centromeric regions [2,5] which are therefore underrepresented or even absent in the genome assemblies [6] which results in assembly gaps. An optimistic perspective for analysis of linear genome organization of *T. castaneum* genome using the Nanopore approach is found in a recently published high-quality assembly of a human genome where Nanopore sequencing strategy spans hundreds of kilobases of highly repetitive DNA [7,8,9]. Moreover, Nanopore sequencing enables to directly detect different methylation states of bases in DNA [10,11], thus opening possibilities to explore *Tribolium* epigenome from a single sequencing run.

Nanopore sequencing by ONT surpasses potentially unlimited read length and a decent sequencing read accuracy (>95%) as it is based on novel principle of detecting the change in ionic current as a single-stranded DNA passes through a protein nanopore embedded in an electrically resistant polymer membrane. Because of these unique properties, one of the main factors for successful sequencing is the quality of the starting DNA. There are three main criteria that isolated genomic DNA should meet. First, it must be devoid of impurities and contaminants, which can be measured with a spectrophotometer and evaluated by absorbance ratios. Additionally, its concentration must be correctly quantified and quantifiably stable throughout measurements, since it is known that HMW DNA tends to be poorly soluble, which can be assisted by longer relaxation times. Finally, to achieve long reads in sequencing, the molecular weight of DNA should be in the desired range meeting the end goal, which is measured by pulsed field gel electrophoresis (PFGE) analysis.

With the development of Nanopore sequencing technologies, the need for optimized protocols for HMW DNA extraction suitable for library preparation has

increased. There are a variety of methods describing the isolation of HMW DNA from cells and specific tissues [12,13]. Tissue isolation usually requires optimization, depending on the condition and availability of the starting material. Insects further represent a challenging material as they are the most diverse group of animals. Although some of DNA extractions have been evaluated for beetles [14] and are capable of producing DNA of moderately high molecular weight (~40 kb), none of them have been tested for use in Nanopore sequencing. In addition, it is crucial to have a method that allows modifications to tailor it to specific scientific questions needed to be answered. Several commercial kits have been developed to isolate HMW genomic DNA suitable for long read sequencing. These kits are based on one of the main principles for isolating nucleic acids: silica-gel based membranes, anion-exchange chromatography columns, magnetic beads and disks with innovative silica-based chemistry, a glass bead-based approach, or the salting-out technique. However, there are also more conventional methods such as grinding in liquid nitrogen and using phenol-chloroform or another extraction buffer. There is also a way to isolate and embed chromosome-sized HMW DNA from nuclei into agarose plugs and the protocol has recently been developed in insects for BAC library construction [15].

Here, we developed a HMW DNA extraction protocol that proved to be optimal for library preparation and Oxford Nanopore MinION sequencing using the beetle *T. castaneum* and closely related species as a model system.

2. Materials and Methods

2.1. Materials

- Insects: Laboratory culture of the red flour beetle *Tribolium castaneum*; highly inbred Georgia 2 (GA2) strain, *T. freemani* and *T. confusum* routinely reared in whole wheat flour at 28 °C and 40% relative humidity in the dark. For the collection of a larger amount of different life stages, 0.71 mm sieve was used for sifting and individual beetles were picked with tweezers.
- Liquid nitrogen.
- Sterilized mortar with pestle and metal spatula.
- 100 µm cell strainer (Thermo Fisher Scientific, Cat. No. 22363549, Waltham, MA, USA).
- 1000 µL wide bore filtered pipette tips (Thermo Fisher Scientific, Cat. No. 2079G, Waltham, MA, USA).
- 1.5 mL DNA LoBind tubes (Eppendorf, Cat. No. 0030108051, Hamburg, Germany).
- Glass rod (Pasteur pipette heated at the end to form thin hook).
- NIB buffer: 10 mM Tris pH 9.4, 60 mM NaCl, 10 mM EDTA pH 8.0, 0.15 mM spermidine, 0.15 mM spermine, 0.5% v/v Triton X-100, 0.1% v/v β-mercaptoethanol.
- Blood and Cell Culture DNA Midi Kit (Qiagen, Cat. No. 13343, Hilden, Germany) or Genomic-Tip 100/G columns (Qiagen, Cat. No. 10243) with prepared G2 (800 mM guanidine HCl, 30 mM Tris-Cl pH 8.0, 30 mM EDTA pH 8.0; 5% Tween-20, 0.5% Triton X-100), QBT (750 mM NaCl, 50 mM MOPS pH 7.0, 15% isopropanol, 0.15% triton X-100), QC (1 M NaCl, 50 mM MOPS pH 7.0, 15% isopropanol) and QF (1.25 M NaCl, 50 mM Tris-Cl pH 8.5, 15% isopropanol) buffers according to recipes of manufacturer's kit handbook.
- RNase A solution (100 mg/mL, Qiagen, Cat. No. 1007885).
- Proteinase K from Tritirachium album (Sigma-Aldrich, Cat. No. SRE0047, St. Louis, MO, USA) or Protease (7.5 AU, Qiagen, Cat. No. 19155), prepared as 20 mg/mL solution in miliQ water.
- Isopropyl alcohol.

- TE buffer pH 8.0: 10 mM Tris, 1 mM EDTA.
- Nuclease-free water (Invitrogen, Cat. No. AM9937, Waltham, MA, USA).
- Pulsed Field Certified Agarose (BioRad, Cat. No. 1620137, Hercules, CA, USA).
- 0.5× TBE buffer: 45 mM Tris-borate, 1 mM EDTA.
- Quick-Load 1 kb Extend DNA Ladder (New England Biolabs, Cat. No. N3239S, Ipswich, MA, USA).
- Lambda PFG Ladder (New England Biolabs, Cat.No. N0341S).
- Syringe and G30 needle.
- Short Read Eliminator XS Kit (Circulomics, Cat. No. SKU SS-100-121-01, Baltimore, MD, USA).
- Qubit dsDNA BR Assay kit (Invitrogen, Cat. No. Q32850).
- Agencourt AMPure XP beads (Beckman Coulter, Cat. No. A63880, Brea, CA, USA).
- Ligation Sequencing Kit (Oxford Nanopore Technologies, Cat. No. LSK-110, Oxford, UK).
- NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (New England BioLabs, Cat. No. E7180S).
- MinION flow cell (Oxford Nanopore Technologies, Cat. No. FLO-MIN111).

2.2. Equipment

2.2.1. Necessary Equipment

- Fume hood.
- Centrifuge with cooling option (Centrifuge 5424 R, Eppendorf).
- Thermomixer (ThermoMixer C with SmartBlock 1.5 mL and ThermoTop, Eppendorf).
- Shaker (Vibramax 100, Heidolph, Schwabach, Germany).
- Qubit 4 Fluorometer (Invitrogen).
- Spectrophotometer (BioSpec-nano, Shimadzu, Kyoto, Japan).

2.2.2. Optional Equipment (for DNA Length Assessment, Library Preparation and Sequencing)

- Magnetic separator for 1.5 mL tubes (MagnaRack Magnetic Separation Rack, Invitrogen).
- Pulsed field gel electrophoresis system (CHEF-DR III system, Bio-Rad).
- Thermal cycler (2720 Thermal cycler, Thermo Fisher Scientific).
- Rotator mixer (Programmable Rotator Multi Bio RS-24, Biosan, Riga, Latvia).
- MinION device (Oxford Nanopore Technologies).

2.3. Procedure

2.3.1. Nuclei Isolation

The nuclei were isolated according to the protocol of Brown and Coleman [15] with a few modifications. Main modifications include the mortar and spatula being precooled in liquid nitrogen rather than the −80 °C, preparation of fresh NIB buffer immediately before use, an additional washing step of the isolated nuclei, as well as modification of the centrifuge times and usage of standard plastic tubes.

Immediately before starting, 20 mL of fresh NIB buffer per reaction was prepared and chilled on ice. The mortar was filled twice with liquid nitrogen, in order to sufficiently precool both the mortar and the spatula. During the second nitrogen evaporation, the sample in amounts according to Table 1 and Supplementary Table S1 was added and grounded in a circular motion. The grinding pressure and the speed were slowly increased in order to produce a fine powder. Using the spatula, all of the powder was scraped into a 50 mL tube containing 8 mL of chilled NIB. In order to achieve maximum efficiency the tube was gently swirled or, if residue remained on the tube walls a wide bore tip was used to gently flush the residue. It is important not to shake the tube. The suspension was transferred through a 100 µm cell strainer into a new chilled 50 mL tube.

Table 1. Overview of isolated HMW DNA samples from *T. castaneum* beetles. Amount of starting material is listed in optimal weights for successful isolation from three developmental stages (larvae, pupae, adults). Accompanying DNA concentrations, final quantities and absorbance ratios at 260 and 280 nm for quality assessment are indicated (see details in Supplementary Table S1).

	Starting Material (mg)	DNA Concentration (ng/µL)	DNA Yield (µg)	$A_{220/260}$	$A_{260/280}$
Larvae	1100	512	51.2	1.87	2.35
Pupae	200	172	17.2	1.87	2.25
Adults	1000	643	64.3	1.85	2.14

The solution was divided into six chilled 1.5 mL tubes and spun at 100× g for 30 s at 4 °C. The supernatant was transferred into six new 1.5 mL tubes using a 1 mL wide bore tip carefully without disturbing the loosely adhering cell debris pellet. The tubes were once again placed in the centrifuge for 3 min at 1800× g and 4 °C to pellet the nuclei. The supernatant was discarded and as much liquid as possible was removed using a regular 1 mL tip. The pellet of compact nuclei was resuspended completely in 1 mL of cold NIB buffer by pipetting with a wide bore tip, but carefully as to avoid introducing air bubbles into the mixture. The previous centrifugation step was repeated. The lysis buffer was prepared by adding 500 µL Protease or 95 µL of Proteinase K and 10 µL of RNase A to 5 mL of G2 buffer. The supernatant after the last centrifugation step was removed and the pellet resuspended in 800 µL of the prepared G2 buffer. It is important to completely resuspend the pellet by pipetting with a wide bore tip, but again, carefully as to not introduce air bubbles. The tubes were incubated at 50 °C for 1 h at 300 rpm in a thermomixer. During incubation, it is recommended either to invert the tubes several times or gently pipette with wide bore tips in order to ensure complete digestion. The resuspended and properly digested nuclei had a milky, stringy texture and were visible to the naked eye. If hard clumps of nuclei remained, they could have been broken by additional pipetting.

2.3.2. Genomic Tip Purification

Genomic Tip manipulation was performed according to the manufacturer's protocol with few modifications, mainly the added pressure during all steps and mandatory prewarming of the QF buffer.

First, a Genomic Tip 100/G column was equilibrated with 4 mL of QBT buffer. After the digestion step, all of the prepared solutions were applied to the column. The column was washed twice with 7.5 mL of QC buffer. The final elution was performed with 5 mL of QF buffer prewarmed to 50 °C. For all steps, it is necessary to apply some pressure with a syringe plunger or a suitable rubber pipette to ensure sufficient flow rate.

The isolated DNA was precipitated by gently adding 3.5 mL of isopropanol at room temperature. The tube was left to stand for 30 s to allow the phases to separate completely. The upper phase should turn a whitish color. The tube was turned over at least 20 times and white DNA strands that stick together with each turn should appear in the solution. The white strands of DNA form a sticky DNA "jelly" which was then picked up with the tip of a thin glass rod. The DNA should stick to the glass rod and not fall off easily. The DNA was spooled and transferred to a 1.5 mL DNA LoBind tube in 100 µL elution buffer of choice (e.g., TE pH 8.0, 10 mM Tris-HCl pH 8.5). The DNA was incubated at 50 °C for up to 2 h to accelerate the homogenization process. At the end of the incubation, the DNA is usually almost completely dissolved, except for single filamentous clumps. The isolated DNA was left overnight at room temperature with gentle horizontal shaking (150 rpm) to achieve final relaxation. At the end, the DNA solution should be completely clear, with increased viscosity, which can be observed by flicking the tube. Subsequent storage was at 4 °C with no noticeable decrease in DNA quality and length for up to several months.

2.3.3. DNA Shearing and Size Selection

The solution of homogenized DNA was sheared 10–30× with a 30-gauge needle. The concentration was then measured in triplicates and adjusted to 150 ng/µL with TE buffer or water. For size selection, the Short Read Eliminator (SRE) XS kit was used according to the manufacturer's protocol. The final resuspension was done in 50 µL EB buffer from SRE kit and the concentration was measured twice to check concentration measurement reproducibility.

2.3.4. Assessment of DNA Quality and Length

DNA concentration was always tested by fluorometric and spectrophotometric methods. Quality was checked with a spectrophotometer and acceptable values are: ~1.8 for A260/280 and ~2.2 for A260/230, according to the official recommendations of ONT. Pulsed field gel electrophoresis was used to assess the length of isolated DNA, sheared DNA and prepared library. DNA fragments were separated by PFGE on a 1% agarose gel run in 0.5× TBE buffer at 6 V/cm, 14 °C, included angle of 120°, switch time 1–10 s for 14 h using a Bio-Rad CHEF-DR III PFGE system. The gel was then stained in 1µg/mL ethidium bromide solution at RT on a shaker for 30 min.

2.3.5. Nanopore Sequencing

For library preparation, 3–3.5 µg of DNA was used for a single reaction and all incubation times were extended as suggested in the Beads free library preparation protocol [16]. A MinION flow cell was loaded with 400–600 ng of DNA and run for 4–10 h before pausing and washing the cell. A prepared library was usually divided into two to three runs with up to five successful runs on one flow cell.

3. Results

To optimize the DNA extraction procedure, we used *T. castaneum* from different developmental stages (larvae, pupae, adults) together with two other *Tribolium* species (*T. freemani*, *T. confusum*). We first tested the E.Z.N.A kit (Omega BioTek, Norcross, GA, USA), the Monarch HMW DNA extraction kit for tissue (New England Biolabs), the Blood and Cell Culture DNA Mini and Midi kit (Qiagen) and phenol-chloroform extraction. Observed strengths and limitations of these approaches are elaborated in detail in Supplementary Table S2. In summary, limitation for the use of DNA obtained with all tested commercial kits and phenol-chloroform extraction in library preparation and manipulation for Nanopore sequencing is primarily due to insoluble DNA pellets that also had impurities, as evident from poor absorbance ratios. Subsequently, when DNA introduced into the standard library preparation protocols of ONT, it resulted in clumping of magnetic beads, large losses after each step of library preparation. Even though some kits were able to produce DNA of considerable length, the quality of subsequent libraries was poor, with insufficient read lengths in Nanopore sequencing and rapid pore death (Supplementary Table S2). Therefore, we developed a HMW extraction protocol from cell nuclei with a purification step using commercially available Genomic Tip columns followed by DNA shearing and size selection (Figure 1).

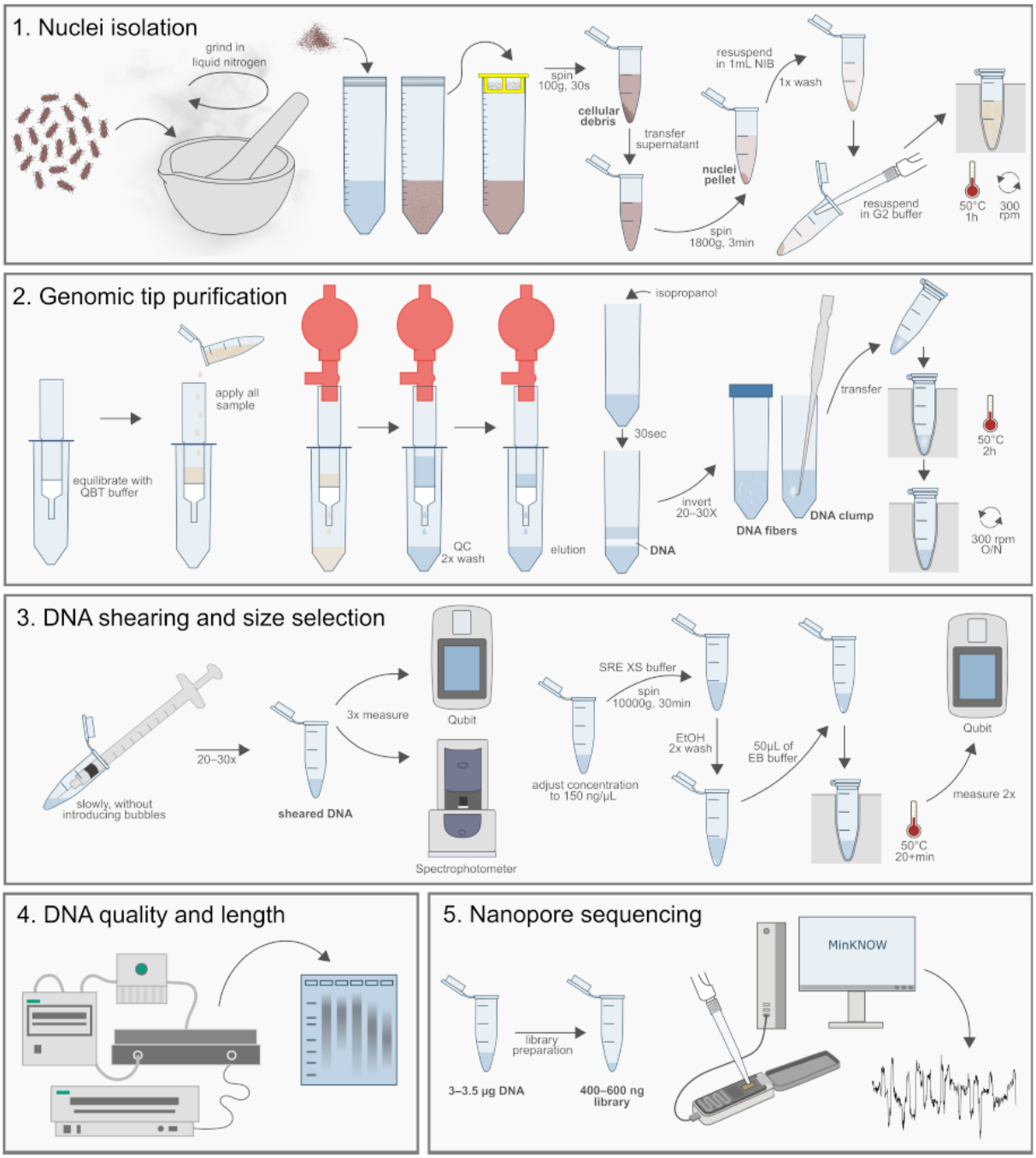


Figure 1. Workflow scheme of optimized HMW DNA isolation in *Tribolium* beetles, quality assessment and preparing for Nanopore sequencing. Steps are explained in detail in the Procedure subsection.

The developed procedure, in combination with the slightly modified standard library preparation protocol (described in Materials and Methods section), showed the highest degree of reproducibility, optimal HMW DNA for Nanopore sequencing and always yielded DNA of sufficient quality and quantity for multiple rounds of sequencing. Results of Nanopore sequencing performed on HMW DNA isolated by using three

different approaches are depicted on Figure 2. The last two plots have the highest N50 values and were obtained from DNA isolated from nuclei and purified on Genomic-Tip.

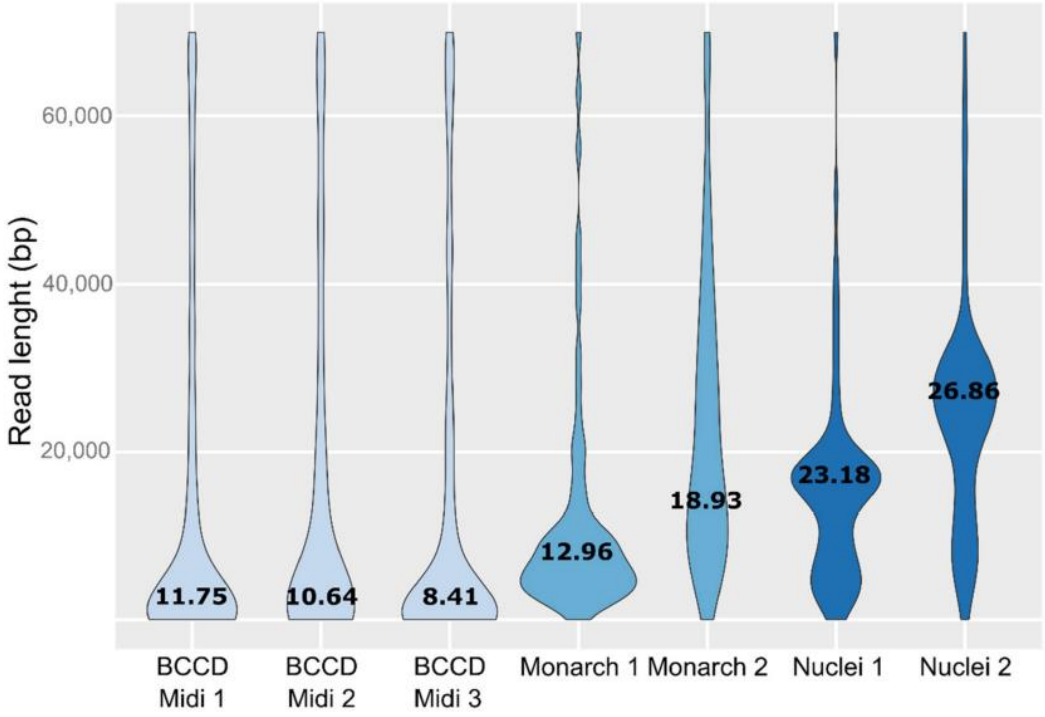


Figure 2. Violin plots of read length distribution of the Nanopore sequencing data derived from six different HMW isolations using Blood and Cell Culture DNA Midi kit (BCCD Midi), Monarch HMW DNA extraction kit for tissues (Monarch) and our developed protocol (Nuclei). Appended numbers represent performed experimental replicates (See also Supplementary Table S2). The width of each violin indicates the size of dataset and the number represent N50 values (kb).

The optimized procedure was tested on all *T. castaneum* developmental stages (larvae, pupae, adults) and the DNA obtained had absorption ratios in the proposed range for sequencing (Table 1). DNA from pupae was isolated from approximately 200 mg of tissue and the obtained DNA yield was sufficient and in concentration adequate for size selection and library preparation. In adults and larvae, in order to achieve sufficient amount of DNA, more than 600 mg of starting tissue is required. This is due to the large amount of non-cellular material found in these developmental stages, mainly chitin in the form of the beetle cuticle in adults and vast amounts of fat and gut in the larvae.

The HMW DNA size distribution was examined by PFGE. The extracted DNA from *T. castaneum* showed most gDNA fragments distributed between 50 and 150 kb (Figure 3a). The DNA isolated from the pupal stage even showed an additional band

at 200 kb. In addition, the same HMW DNA isolation procedure was tested for two congeneric species, *T. confusum* and *T. freemani*, which also had yielded gDNA up to 100 kb in length (Figure 3b).

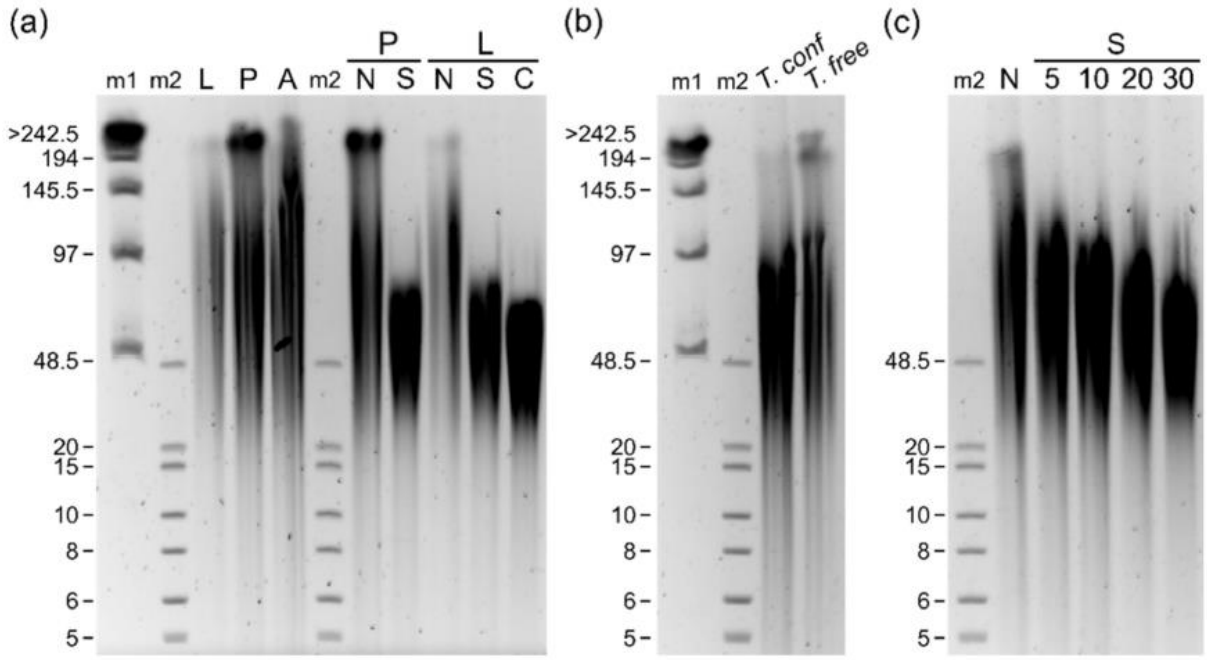


Figure 3. Pulsed field gel electrophoresis of isolated genomic DNA together with sheared and cleaned fractions from different developmental stages of *T. castaneum*, *T. confusum* (T. conf) and *T. freemani* (T. free) beetles. Lambda DNA (m1) and Extend DNA ladder (m2) were used as molecular weight markers. For all samples around 1 µg of DNA was mixed with loading dye and loaded per single well. Electrophoresis was run on CHEF-DR III PFGE system with 1% agarose gel with settings listed in procedure section. (a) Genomic DNA isolated from *T. castaneum* developmental stages (L-larvae, P-pupae, A-adults) together with sheared (30×) and size selected fraction. N-non-sheared fraction, S-sheared fraction with G30 needle, C-size selected DNA with Short Read Eliminator Kit XS (Circulomics). (b) Genomic DNA isolated from *T. confusum* pupae and *T. freemani* adults. (c) Testing of needle shears with increasing amount of passes through needle (indicated with numbers) on *T. castaneum* adult gDNA.

In order to obtain more efficient sequencing, library preparations from sheared DNA were tested. Both pupal and larval DNA after shearing showed a decrease in the ultra-long DNA fraction with a majority of DNA in the 30–80 kb range. This is further tested with shear intensity assessment, where 30 passes through the G30 needle yielded the most compact band, with the majority still above 48 kb (Figure 3c). Interestingly and of great importance for subsequent sequencing, there was no clear increase in the abundance of shorter fragments in gel electrophoresis. Size selection on sheared DNA showed only a slight negative effect on DNA length in PFGE,

attributed to an additional step of centrifugation and manipulation, which is indicated by a slight downward shift. Nonetheless, this step resulted in a reduction of short reads during sequencing, leading to longer pore lifetimes and greater overall yield. Because ultra-long DNA in library preparation did not increase N50 in sequencing, we decided to perform DNA needle shearing, which was shown to be beneficial [17]. Interestingly, we were able to increase N50 from about 11 kb to 20 kb after 20 passes through the G30 needle, and up to 26 kb when 30 shear passes were performed (Figure 4a–c).

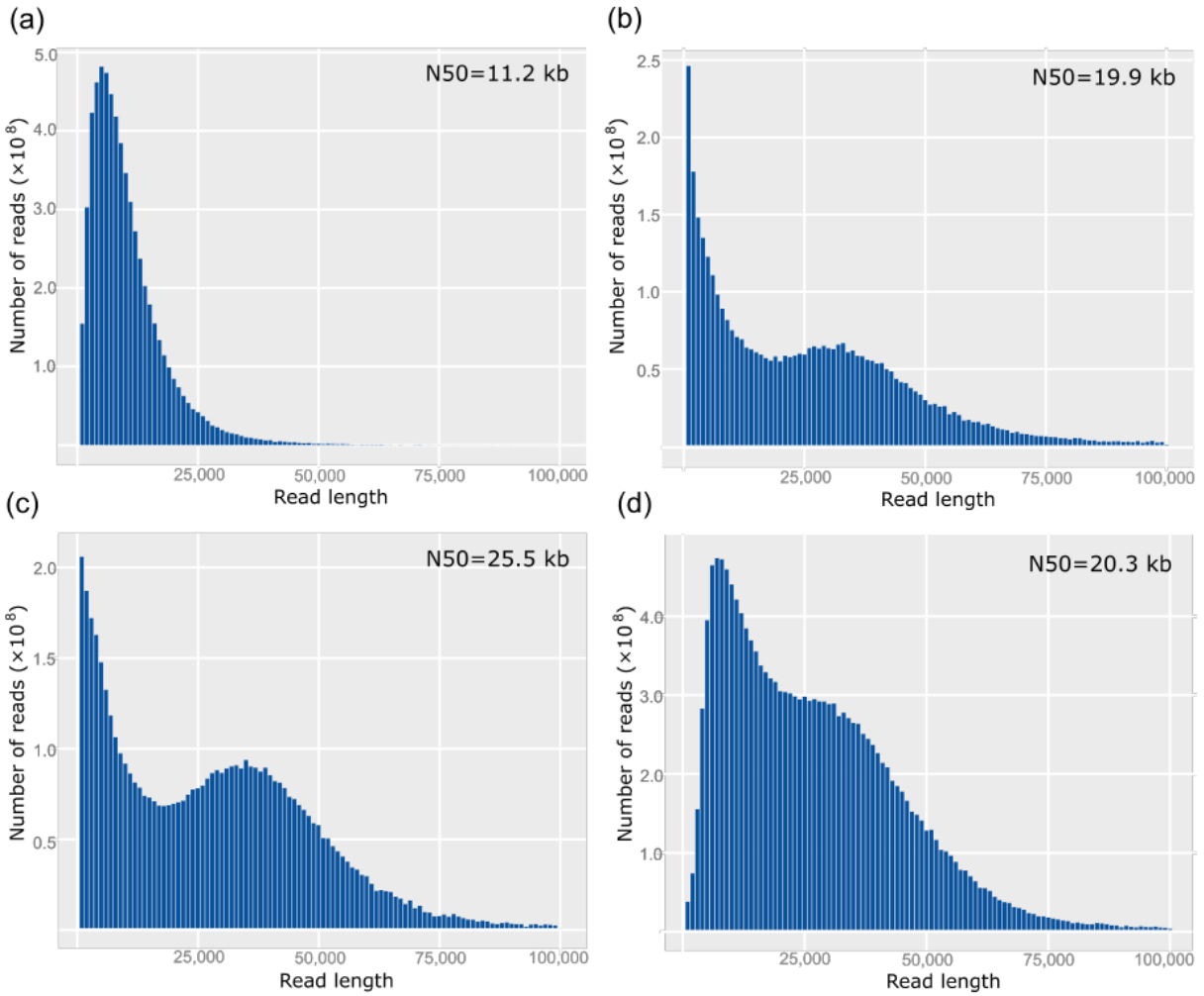


Figure 4. Length distribution graphs in correlation with gDNA shearing and size selection in Nanopore sequencing. Corresponding N50 values are indicated in top right corner of each graph. (a) DNA that has not been sheared. (b) Shearing with 20 passes through G30 needle. (c) Shearing with 30 passes through G30 needle. (d) Shearing with 30 passes through G30 needle and size selection with Short Read Eliminator kit (Circulomics).

To remove smaller DNA fragments, we used centrifugation-based size selection, which efficiently removes the majority of reads below 10 kb, which is not apparently visible on the PFGE but rather makes an immense difference in the

sequencing run. This is best seen in the length histogram by the absence of the leftmost shortest read peak (Figure 4d).

Nanopore sequencing of the prepared library on the MinION cell with five consecutive loads within 48 h yielded 13.17 Gb of data (Figure 5a). The distribution of the Phred quality score (Q) shows that most reads have a quality above Q20, which means less than 1% error rate (Figure 5b). Moreover, quality correlates positively with read length, with the majority of longest reads having 99.6% accuracy (Q24).

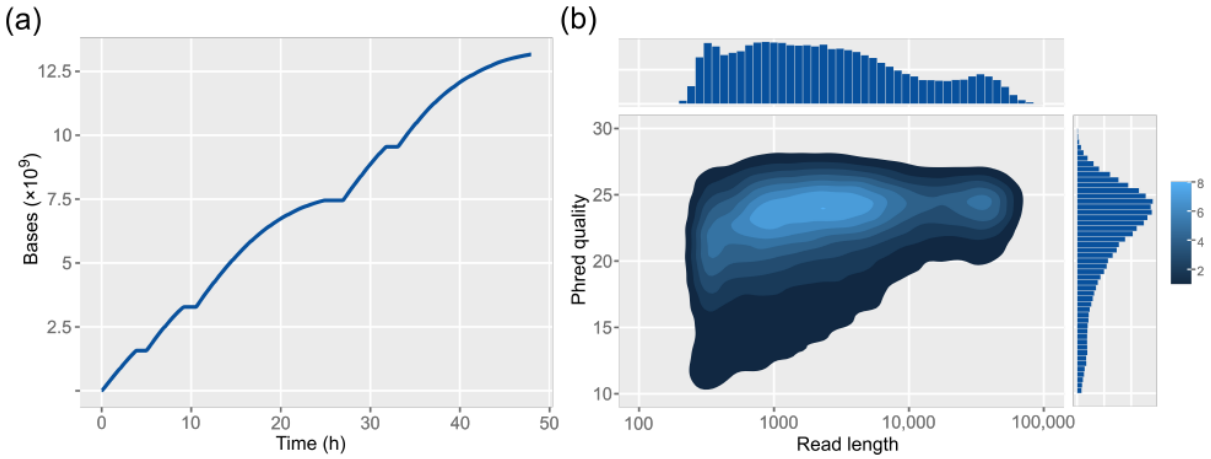


Figure 5. Summary of Nanopore sequencing run output. (a) Cumulative bases output after 48 h run on MinION flow cell with five consecutive library loads and four intermediate DNase washes in the duration of 60 min. (b) 2D density plot of overall read Phred quality scores and read lengths. Color shades represent higher cumulative fractions of reads with a given length and quality.

4. Discussion

Following procedures for commercial kits, especially those made for smaller inputs and even tested for *Tribolium castaneum* [18] as well as phenol-chloroform protocol, did not yield DNA of sufficient quality and quantity for Nanopore sequencing. The Brown and Coleman's protocol [15] proved to be the best starting point; however, although claimed to be suitable for long-range sequencing and optical mapping, this protocol does not elaborate on how to isolate HMW from plugs. Although it is likely to yield DNA of higher molecular weight, there is no information on exact lengths and qualities. Here, we have succeeded in isolating DNA and validating its use for library preparation and Nanopore sequencing.

All commercial kits tested provided DNA of sufficient length but had one major drawback; the DNA pellets were compact and difficult to dissolve posing a major problem during the various DNA clean-up and centrifugation steps required in library preparation. This difficulty was successfully overcome by using liquid nitrogen pulverization as the initial step, as this resulted in superior DNA quality, yield and reproducibility.

We found that the amount of input material is critical and, for beetles, depends on which developmental stage is used. For example, larvae and adults require more starting material due to the high proportion of non-cellular material (fat and chitin). This is most likely the main reason for the nonlinear dependence of input weight and final DNA concentration.

In the method we described, there are some steps that we think need special attention and can lead to unsuccessful isolation if not performed properly. Nuclear pellets should be completely resuspended before incubation with protease and RNase A. During subsequent incubation, additional pipetting can further aid in homogenization of the solution. The use of increased input material almost always resulted in poor flow rates in the purification step of the Genomic Tip, but had no negative effect on the final DNA and was resolved by the use of significant positive pressure using a propipette or syringe. A great indicator of the higher DNA concentrations and likely length, is the viscosity of the DNA eluting from the columns as falling droplets are left behind and expand. When precipitating DNA with isopropanol, it is advisable to keep the two phases separate before inverting so that the upper phase becomes whitish, which facilitates the formation of DNA fibers. Sometimes up to 30 inversions are necessary

for compact DNA aggregates to form. We have found that spooling is key to obtaining DNA with adequate properties, as centrifugation and ethanol washes tend to yield much less concentrated DNA. Spooled DNA in EB buffer quickly forms a “jelly-like” mass, which is the best sign of efficient relaxation of HMW DNA. Moreover, this observed dynamic will ensure its homogenization in all subsequent steps and this DNA requires acceptable and reproducible periods of up to one day to be fully relaxed and ready for library preparation. Genomic DNA isolated by using the procedure described here showed optimal absorbance ratios and had the most reliable concentration values obtained using Qubit after light shearing. Repeated measurements were always within less than 10 % of each other. It is also worth mentioning that when isolated DNA was stored at 4°C for three months and even after freezing/thawing process, we did not observe any decrease in DNA stability compared to freshly isolated DNA.

Although this method produces extremely long DNA fragments, regarding Nanopore sequencing we find that controlled mechanical shear of HMW DNA helps to produce even longer reads. This can possibly be explained by pore blocking of entangled DNA fragments, less efficient adapter ligation but could also be genome specific. This may be particularly true for *Tribolium* genomes, which are known to be A-T rich (>60%) and highly repetitive, with individual satellite DNAs accounting for up to 40% of the genome [19]. For this reason, each species and its DNA will most likely require some degree of optimization in shearing to obtain higher N50 values. Nevertheless, this protocol succeeded in isolating HMW DNA from three closely related beetles and due to its excellent performance, especially in terms of DNA concentrations and qualities, we propose that it provides a suitable starting point for isolation of HMW DNA from other Coleoptera species.

Supplementary Materials

Supplementary Table S1. Results of the DNA isolation using developed protocol performed on different *Tribolium* species as well as on their various developmental stages.

Species	Stage	Starting material (mg)	DNA concentration (ng/µL)	DNA yield (µg)	$A_{220/260}$	$A_{260/280}$
T. castaneum	Pupae	200	172	17,2	1,89	1,95
		200	130	13,0	1,85	2,14
		200	138	13,8	1,84	2,25
		200	154	15,4	1,88	1,78
	Larvae	1100	512	51,2	1,87	2,35
	Adults	1000	643	64,3	1,85	2,14
T. freemani	Adults + Larvae	1050	327	32,7	1,83	2,24
		920	540	54,0	1,88	2,05
	Adults	620	78	4,7	1,94	2,00
T. confusum	Pupae	340	213	12,8	1,87	2,41

Supplementary Table S2. DNA extraction and sequencing methods tested for the isolation of HMW DNA from *Tribolium castaneum*. Main features are described for all used commercial kits and procedures. Obtained DNA is evaluated based on its total amount from a single isolation reaction and molecular weight distribution. All kit isolations are done according to manufacturers' protocol. Observed strengths and potential limitations for the use is library preparation and DNA manipulation for Nanopore sequencing are elaborated in detail.

		Extraction				Sequencing	
Kit/procedure for HMW DNA isolation	Main features	DNA amount	Molecular weight	Strengths	Limitations	Performed	Observations
E.Z.N.A kit (Omega BioTek)	silica-gel based membrane	6 µg	20-50 kb	-isolation from lower amount of input material -reproducible	- DNA rarely above 50 kb - lower DNA yield due to membrane capacities	No	- insufficient DNA for an efficient library preparation
Monarch HMW DNA extraction kit for tissue (New England Biolabs)	glass bead-based approach	0-10 µg	20-200 kb	- DNA of high molecular weight -lower DNA input	- low absorbance ratios - extremely nonreproducible (DNA on glass beads can fall off)	Yes (Figure 2)	- library protocol not reproducible - clumping of beads during library preparation - low pore occupancy
Blood and Cell Culture DNA Mini and Midi kit (Qiagen)	anion-exchange chromatography columns	3 µg (Mini kit) 30 µg (Midi kit)	20-50 kb	- possibility to use greater amounts of starting material - high DNA yield - reproducible	- cell lysate causes high degree of column blockage - DNA pellet difficult to dissolve -low absorbance ratios - isolated DNA causes problems in library preparation (bead clumping, loss of DNA)	Yes (Figure 2)	- problems with bead clumping - low N50 value - low final library concentration
phenol-chloroform extraction	salting-out principle	5 µg (from 30 pupae)	30-60 kb	- possibility to use greater amounts of starting material	- high DNA losses - lower DNA quality - laborious	No	- visible non-dissolved DNA - poor DNA quality that implies poor library output

References

1. Klingler, M. *Tribolium*. Curr. Biol. 2004, 14, R639–R640, doi:10.1016/j.cub.2004.08.004.
2. Richards, S.; Gibbs, R.A.; Weinstock, G.M.; Brown, S.; Denell, R.; Beeman, R.W.; Gibbs, R.; Bucher, G.; Friedrich, M.; Grimmelikhuijzen, C.J.P.; et al. The Genome of the Model Beetle and Pest *Tribolium castaneum*. Nature 2008, 452, 949–955, doi:10.1038/nature06784.
3. Herndon, N.; Shelton, J.; Gerischer, L.; Ioannidis, P.; Ninova, M.; Dönitz, J.; Waterhouse, R.M.; Liang, C.; Damm, C.; Siemanowski, J.; et al. Enhanced Genome Assembly and a New Official Gene Set for *Tribolium castaneum*. BMC Genomics 2020, 21, 1–13, doi:10.1186/s12864-019-6394-6.
4. Wang, S.; Lorenzen, M.D.; Beeman, R.W.; Brown, S.J. Analysis of Repetitive DNA Distribution Patterns in the *Tribolium castaneum* Genome. Genome Biol. 2008, 9, 1–14, doi:10.1186/gb-2008-9-3-r61.
5. Ugarković, D.; Podnar, M.; Plohl, M. Satellite DNA of the Red Flour Beetle *Tribolium castaneum* - Comparative Study of Satellites from the Genus *Tribolium*. Mol. Biol. Evol. 1996, 13, 1059–1066, doi:10.1093/oxfordjournals.molbev.a025668.
6. Pavlek, M.; Gelfand, Y.; Plohl, M.; Meštrović, N. Genome-Wide Analysis of Tandem Repeats in *Tribolium castaneum* Genome Reveals Abundant and Highly Dynamic Tandem Repeat Families with Satellite DNA Features in Euchromatic Chromosomal Arms. DNA Res. 2015, 22, 387–401, doi:10.1093/dnares/dsv021.
7. Miga, K.H.; Koren, S.; Rhie, A.; Vollger, M.R.; Gershman, A.; Bzikadze, A.; Brooks, S.; Howe, E.; Porubsky, D.; Logsdon, G.A.; et al. Telomere-to-Telomere Assembly of a Complete Human X Chromosome. Nature 2020, 585, 79–84, doi:10.1038/s41586-020-2547-7.
8. Jain, M.; Olsen, H.E.; Turner, D.J.; Stoddart, D.; Bulazel, K. V.; Paten, B.; Haussler, D.; Willard, H.F.; Akeson, M.; Miga, K.H. Linear Assembly of a Human Centromere on the Y Chromosome. Nat. Biotechnol. 2018, 36, 321–323, doi:10.1038/nbt.4109.
9. Logsdon, G.A.; Vollger, M.R.; Hsieh, P.H.; Mao, Y.; Liskovykh, M.A.; Koren, S.; Nurk, S.; Mercuri, L.; Dishuck, P.C.; Rhie, A.; et al. The Structure, Function and Evolution of a Complete Human Chromosome 8. Nature 2021, 593, 101–107, doi:10.1038/s41586-021-03420-7.
10. Rang, F.J.; Kloosterman, W.P.; de Ridder, J. From Squiggle to Basepair: Computational Approaches for Improving Nanopore Sequencing Read Accuracy. Genome Biol. 2018, 19, 1–11, doi:10.1186/s13059-018-1462-9.
11. Liu, Y.; Rosikiewicz, W.; Pan, Z.; Jillette, N.; Taghbalout, A.; Foox, J.; Mason, C.; Carroll, M.; Cheng, A.; Li, S. DNA Methylation Calling Tools for Oxford Nanopore Sequencing: A Survey and Human Epigenome-Wide Evaluation. bioRxiv 2021, doi:10.1101/2021.05.05.442849.
12. Mayjonade, B.; Gouzy, J.; Donnadieu, C.; Pouilly, N.; Marande, W.; Callot, C.; Langlade, N.; Muños, S. Extraction of High-Molecular-Weight Genomic DNA for Long-Read Sequencing of Single Molecules. Biotechniques 2016, 61, 203–205, doi:10.2144/000114460.
13. Angthong, P.; Uengwetwanit, T.; Pootakham, W.; Sittikankaew, K.; Sonthirod, C.; Sangsrakru, D.; Yoocha, T.; Nookaew, I.; Wongsurawat, T.; Jenjaroenpun, P.; et al. Optimization of High

Molecular Weight DNA Extraction Methods in Shrimp for a Long-Read Sequencing Platform. PeerJ 2020, 8, 1–18, doi:10.7717/peerj.10340.

14. Chen, H.; Rangasamy, M.; Tan, S.Y.; Wang, H.; Siegfried, B.D. Evaluation of Five Methods for Total DNA Extraction from Western Corn Rootworm Beetles. PLoS One 2010, 5, doi:10.1371/journal.pone.0011963.
15. Brown, S.J.; Coleman, M. Isolation of High Molecular Weight DNA from Insects. Methods Mol. Biol. 2019, 1858, 27–32, doi:10.1007/978-1-4939-8775-7_3.
16. Rai, S. Beads Free ONT Ligation Kit Library Preparation for Ultra-Long Read Sequencing Available online: https://gih.uq.edu.au/research/long-read-sequencing/beads-free-ont-ligation-kit-library-preparation-ultra-long-read-sequencing.
17. Tyson, J. Rocky Mountain Adventures in Genomic DNA Sample Preparation , Ligation Protocol Optimisation / Simplification and Ultra Long Read Generation. 2020, 1–5, doi:dx.doi.org/10.17504/protocols.io.7euhjew High.
18. Oppert, B.; Stoss, S.; Monk, A.; Smith, T. Optimized Extraction of Insect Genomic DNA for Long-Read Sequencing. Methods Protoc. 2019, 2, 1–7, doi:10.3390/mps2040089.
19. Mravinac, B.; Plohl, M. Parallelism in Evolution of Highly Repetitive DNAs in Sibling Species. Mol. Biol. Evol. 2010, 27, 1857–1867, doi:10.1093/molbev/msq068.

4. Discussion

Characterization and evolution of CenH3 in *Meloidogyne* species

Centromeres are determined by epigenetic mark, specific centromeric histone H3 variant, CenH3. Based on numerous studies of sexual animal and plant species, including humans, CenH3 is known to be necessary for centromere function in all organisms, but despite its conserved function, the CenH3 protein sequence evolves rapidly. It has been hypothesized that CenH3 evolve rapidly to suppress the deleterious effects of centromere drive in meiosis.

In this study, CenH3 was investigated in three closely related mitotic *M. incognita*, *M. arenaria* and *M. javanica* together with distantly related meiotic parthenogenetic *M. hapla* species. A total of 21 CenH3 proteins were discovered separated in two polyphyletic groups with each species having more than one copy of CenH3 gene. The abcCenH3 group contained divergent set of proteins having relatively low similarity of HFD region with canonical H3 variants. On the other hand, homogenous αβγδεCenH3 group showed higher degree of identity to H3 in the range and with diagnostic characteristics expected for CenH3 (Malik & Henikoff, 2003). Most probably these two groups had independent evolution and their roles for centromere function can be further evaluated. Paralogous genes mostly result from gene duplication after which genes can keep the same function but also could diverge and acquire different functions with possibility for one copy to be lost over time. For example, in *Drosophila* species it has been recently found that some have multiple copies of *Cid*, centromeric histone gene that also perform nonredundant centromeric roles probably as a result of subfunctionalization (Kursel & Malik, 2017). Similarly, mosquito species encode two *mosqCid* paralogs that have functionally diverged and been coretained for over 150 My (Kursel et al., 2020). Several diploid plants have multiple copies of CenH3 with cowpea as an interesting example where both functional variant are transcribed and knockout experiments showed one necessary for development and reproduction whereas other is still likely in process of subfunctionalization (Ishii et al., 2020). In investigated *Meloidogyne* species, αCenH3 variant showed dominant expression compared to other CenH3s and maintaining its high expression in almost all phases of life cycle suggested that αCenH3 represent the most promising centromere candidate. Consequently, further analysis was focused on αCenH3 and closely related monophyletic βγδεCenH3 group. MIG species possess multiple copies of αCenH3 and one copy of variant from βγδCenH3 group whereas diploid *M. hapla* has one copy of

αCenH3 and εCenH3. Observations of CenH3 copy numbers are in consistence with estimated ploidy level based on analysis of protein-coding sequences (CDSs) in MIG species (Blanc-Mathieu et al., 2017). Investigation of gene/protein evolution is necessary to be performed based on sequence changes that can in case of centromere be revealed by following presence and conservation of CenH3s. The evolution of detected CenH3s in *Meloidogyne* is analyzed by integration of knowledge from previous phylogenetic species studies together with results from our studies.

Two independent studies have recently sequenced several *Meloidogyne* species genome in order to resolve complex evolutionary origin of these species (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017). MIG species have been confirmed as polyploids formed by interspecific hybridization events. Formation of hybrids is a result of different species interbreeding, resulting in the combination of genetic material from preceding gene pools which has become primary source of data for speciation and adaptation studies (reviewed in Schwenk et al., 2008). Their genomes show duplicated gene region with high within-species nucleotide divergence suggesting they could even potentially have originated from multiple hybridization events making their evolutionary history more complex to resolve (Blanc-Mathieu et al., 2017). MIG species reproduce only asexually by mitotic parthenogenesis while *M. hapla* is a diploid species and reproduces asexually by meiotic parthenogenesis but can also be se sexual. Asexuality is more efficient it the resource terms as individuals need less energy for finding mates and reproduction itself. This offers them opportunity to increase population number faster and transfer all of their genes to offspring. On the other hand, asexual reproduction doesn't offer genetic variations and results in decreased diversity which is less suitable for adapting to new environmental conditions. However, MIG species are remarkably widespread and their thriving as clonal parasites presents evolutionary paradox in the terms of sex benefit (reviewed in Castagnone-Sereno & Danchin, 2014). Research into related taxa or species that do not have obligatory asexuality, but have retained meiosis or adaptive potential of sexual reproduction offers unique insight into species evolution. Moreover, several studies point out that aforesaid hybridization events involved in formation of MIG species have a recent origin.

Taking into account this complex species evolutionary history the most parsimonious scenario of CenH3 evolution in the analyzed species is presented in Figure 7. CenH3 sequence phylogenetic comparison strongly suggests that *CenH3* gene has

undergone one duplication event in an ancestral species that is progenitor of MIG species and *M. hapla*. This event resulted in appearance of αCenH3 and βγδεCenH3 ancestral genes. Previous studies done on nuclear and mtDNA genes polymorphism markers suggest separation of these two species clades (mitotic and meiotic) occurred 17-50 My ago (reviewed in Philippe Castagnone-Sereno et al., 2013). Persistence of completely conserved αCenH3 in all investigated species suggests its evolution under strong purifying selection. It resulted in stabilizing selection by removal of potential genetic polymorphism that could arise through random mutations and would have negative or deleterious effect. The other ancestral βγδεCenH3 gene during this time period evolved into four different but related proteins (β, γ, δ, εCenH3). Characteristics of εCenH3 is that it is only found in *M. hapla*.

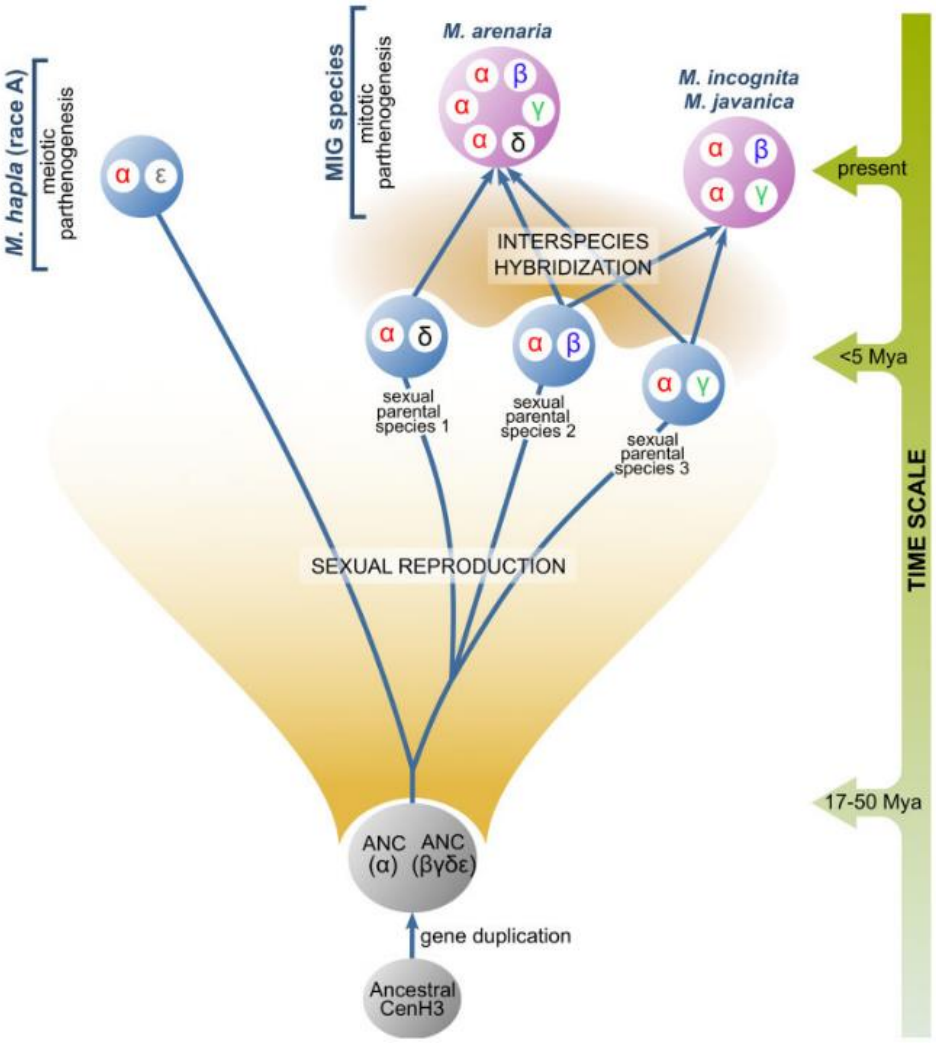


Figure 7. Proposed scenario of CenH3 variants (α, β, γ, δ, ε) evolution in *Meloidogyne* species (*M. hapla* [race A], *M. incognita*, *M. javanica*, *M. arenaria*). ANC(α) - ancestral αCenH3, ANC(βγδε) – ancestral CenH3 of βγδε variants.

Independent evolution of α, β, γ and δCenH3 in sexual taxa is followed by polyploidization as a result of species hybridization. Almost complete conservation of β, γ and δCenH3 among MIG species confirms interspecific hybridization as a relatively recent event. It is most probable they existed for longer period of times in sexual progenitors where they had potential functional role or otherwise there would be evidence of random mutations accumulation and pseudogenization. After interspecies hybridization, existence of CenH3 variants corresponds to their joining in MIG species resulting in redundancy detected in nascent species. Presence of εCenH3 in *M. hapla* additionally supports the putative role of βγδεCenH3 group as it has been retained for 17-50 My since the divergence from MIG species.

Unlike αCenH3, β, γ, δ and εCenH3 evolved under purified selection with positive evolution trend visible on the first several amino acids on the protein N-terminal tail. This different evolutionary dynamic of analyzed *Meloidogyne* CenH3s might be a result of distinct requirements posed on αCenH3 versus ones acting on βγδεCenH3 in centromere. One of possible explanations comes from CenH3 evolutionary studies in *Drosophila* and mosquito (Kursel & Malik, 2017; Kursel et al., 2020) where it has been shown that *CenH3* gene duplications could be necessary for multiple centromeric functions such as mitosis and meiosis. In *Meloidogyne*, observed evolutionary pattern of βγδεCenH3 could be predicted by centromere drive model which resides on asymmetric female meiosis and sexual reproduction. That could impact β, γ and δCenH3 evolutionary scenario in parental sexual lineages as a consequence of centromere drive in meiosis and then becoming redundant in MIG species after transition to asexual mitotic parthenogens which correlates with recent interspecific hybridization event (Blanc-Mathieu et al., 2017; Lunt, 2008). In support, αCenH3, unlike β, γ and δCenH3 shows high expression and localization on all chromosomes of mitotic *M. incognita* but also in two other mitotic MIG species. This speaks also in favor of αCenH3 centromere competence and most probably of loss for β, γ and δCenH3 centromere-associated function in mitotic species. Furthermore, investigation of nonreproductive stages of *M. hapla* where mitosis is preferentially happening also revealed dominant αCenH3 expression. The aforementioned observations all point towards the dominant role of αCenH3 in mitosis.

Additional endorsement that obligatory mitotic parthenogenesis in these species appeared recently comes from the fact that major sperm protein involved in the formation of sperm is found in mitotic *Meloidogyne* taxa with no sign of increased

evolutionary rate despite no apparent functional requirement (Lunt, 2008). Similarly, β, γ and δCenH3 show no evidence of pseudogenization or sequence degeneration caused by random mutations and therefore, like sperm proteins, probably represent "meiotic relict" in exclusively mitotic MIG species. On the other hand, αCenH3 with strong amino acid conservation most probably subspecialized for the role in mitosis. In the context of centromere drive, where CenH3 protein evolves with cenDNA to suppress the deleterious effect of rapidly evolving cenDNA in meiosis, αCenH3 would by subspecialization release itself from adaptive conflicts in meiosis. A similar example can be seen in the well-studied nematode model of *C. elegans* whose holocentric chromosomes harbor two CenH3 related proteins; HCP-3 and CPAR-1. There are evidences supporting role of HCP-3 in mitosis while not required for meiotic chromosome segregation with CPAR-1 loaded only onto meiotic chromosomes with no centromere-localization in mitosis strongly indicating their separate subfunctionalized roles (Monen et al., 2005, 2015).

Structure and organization of centromere in *Meloidogyne* species

Further analyzes were focused on the long-term conserved αCenH3 in order to disclose genetic component of active centromeres. The αCenH3-associated cenDNA in *M. incognita* is formed of short tandem repeat (TR) arrays up to 1 kb in length, composed of five distinct families build up from 50-80 bp monomers. Tandem repeats are known to be highly prevalent at centromeres of both animals and plants, but at the same time show overall lack of sequence conservation as revealed by huge comparative study of 282 species (Melters et al., 2013). More specifically, monocentrics are often characterized by the presence of long arrays of species-specific satDNA (reviewed in Plohl et al., 2014) whereas holocentric organism investigated so far show different patterns of cenDNA. In *C. elegans* there is no specific repetitive sequence underlying centromere but rather it coincides with transcription factor hotspots without high binding affinity for any of them specifically (Steiner & Henikoff, 2014). Another parasitic nematode, *Ascaris suum* did not show any strong, discrete point-like CenH3 peaks in native ChIP-seq analysis but rather its holocentric organization is more diffusely distributed without centromere-specific DNA sequence (Kang et al., 2016). First report of specific sequences residing in centromere are reported for holocentric plant *Rhynchospora pubera*. Its CenH3 interacts with one

satellite DNA family and centromeric retrotransposons forming 3-16 kb arrays intermingled with coding sequences and transposons (Marques et al., 2015).

In *M. incognita* αCenH3-associated pool of divergent tandem repeats (TRs) showed exceptional feature of 19-bp long-conserved box with high GC ratio and potential to form stable dyad structure. This suggests presence of selective pressure acting on this sequence part regardless of the high evolving nature of detected centromere repeats. Even though centromere is mainly determined epigenetically there are still uncertainties as how certain loci are selected and maintained. DNA component could guide localization either by recruiting sequence-specific DNA binding proteins or by intrinsic feature of sequence such as secondary structures. Some DNA motifs are capable to fold into conformations that differ from canonical B-DNA helix and are thus called non-B DNA. Structures such as single-stranded DNA, hairpins, R-loops, i-motifs and triplexes have been primarily linked with genetic instability (Zhao et al., 2010) but recently also found to be enriched at centromeres (Kasinathan & Henikoff, 2018). Thus, conserved 19-bp box could be a binding site for αCenH3 in *M. incognita* and its secondary structure could provide favorable binding capacity. Aforementioned unique sequence characteristics are additionally fortified by the extremely high GC content of centromere 19-bp core that is immersed in high AT composition of the genome (Abad et al., 2008). All these features provide further layer of ensuring proper formation of αCenH3 centromere in *M. incognita*.

Investigation into other two closely related species of *M. incognita*, *M. arenaria* and *M. javanica* revealed presence of completely preserved αCenH3 centromere associated TRs in both sequence and organization aspects. Furthermore, αCenH3 showed colocalization with 19-bp box suggesting preservation of protein and DNA component of centromere in these related species. Conservation of box region has been also found in distant *M. hapla* implying strong functional constraints in species even farther separated in phylogenetic tree. Among the centromeric-specific preserved sequence motifs found so far the most prominent example is CENP-B box. This 17-bp long motif found in alpha-satDNA monomers of humans and alphoid repeats of wide range of mammals presents binding site for centromere protein B (CENP-B). Main roles include mediating kinetochore assembly and influencing centromeric domain architecture by ability to bind two separate CENP-B boxes with speculation it has evolved over a pre-existent centromeric function in order to provide additional stability in the competition occurring during meiotic divisions (reviewed in Gamba & Fachinetti, 2020). Regarding

Meloidogyne, preserved interspecifically motifs have been previously found in many satDNAs, implying they evolved under functional constraints with potential roles that still need to be elucidated (Meštrović et al., 2006a, 2006b, 2013).

Chromosomal spreads of MIG species suitable for immunostainings have been successfully performed for the first time. Knowledge of the chromosomes of these species has so far come mainly from cytological work on *M. incognita*, which revealed a variable chromosome number of 40-46 in oogonial divisions (Triantaphyllou, 1981). In our study, centromere analysis performed on several different stages of chromosomal condensation revealed unusual pattern of uneven αCenH3 distribution among different chromosomes and along the length of individual ones. Moreover, immunofluorescence on prophase chromosomes and especially extended chromatin fibers that enables precise localization assessment, showed discontinuous patterns of αCenH3 domains separated by αCenH3-lacking chromatin. This pattern of αCenH3 distribution can be defined as novel cluster-like centromeric organization of holocentromere. Chromosomes are mainly evaluated for centromere positioning in their fully compact appearance in metaphase as it offers insight in their behavior during divisions. In *M. incognita* one part of chromosomes has αCenH3 encompassing the entire length with either abundant or more discrete signal while the rest is showing extremely uneven distribution with highly abundant regions positioned at different sites relative to the ends of the chromosomes. Functional potential was evaluated by interrelation of centromere protein with α-tubulin as microtubule building subunit involved in chromosome positioning during divisions, known to attach kinetochore assembled on each centromere. Confirmed colocalization of α-tubulin with most αCenH3 domains indicated functionality of αCenH3 centromere in mitosis. Compared to other nematodes, *Meloidogyne* centromere shows most similarity to *Ascaris*, whose 1-15 kb centromeric domains are distributed across the chromosomes rather than to point centromere subunits of *C. elegans*. Also, some chromosomal regions that showed no associated αCenH3 resembled in part to organizational pattern of *Cuscuta* holocentromere that possesses CenH3 restricted to one to three regions per chromosome (Oliveira et al., 2020). All information obtained from chromosome observations in holocentric species provides insight, that holocentromere in contrast to monocentromere offers greater flexibility for organization of CenH3 domains.

CenH3 and centromere organization in *T. castaneum*

In contrast to nematodes which are usually characterized as holocentrics, insects have a vast number of described monocentric species with at least four confirmed independent transitions to holocentrism accompanied by complete loss of CenH3 (Drinnenberg et al., 2014). Same study computationally predicted and compared CenH3 sequence candidates among several insect species, including *T. castaneum*. This premise was experimentally validated in our study and proved cCenH3 as an authentic centromeric variant of histone H3 in *T. castaneum*. Protein sequence of cCenH3 showed all diagnostic criteria of protein sequence suggested to be a predictive mean when assessing centromeric candidates (Malik & Henikoff, 2003). Analysis of *T. castaneum* genome assembly revealed that CenH3 is encoded by a single copy gene which is a common trait for diploid species. One of interesting properties of cCenH3 protein is its exceptionally short N-terminal tail (24 aa), even shorter than canonical H3 counterpart (40 aa). Other species have often N-terminus that extends over 100 aa residues as seen in *D. melanogaster* CID and *C. elegans* HCP-3 while αCenH3 in *Meloidogyne* has around 60 aa protruding from nucleosomal core.

On chromosomes, CenH3 is localized at the primary constrictions with unusual feature of their extraordinary extent. By comparing CenH3 region to full chromosomal size it is estimated they comprise over 40% of the chromosome length with sizes up to 15,8 Mb. For comparison, *D. melanogaster* which has a genome of similar sizes, has centromere regions that span 100-170 kb (Chang et al., 2019), making *T. castaneum* longest insect regional centromeres described so far. Another interesting centromere example are fire ants *Solenopsis invicta* (Huang et al., 2016) whose long centromeres span even one third of the chromosome length (3,6 Mb). The fact that closely related tropical ants *S. geminata* show typical monocentric morphology with small primary constrictions speaks in favor of the exceptional centromere length variability even among related species. Additionally, *T. castaneum* centromeres are characterized by their metapolycentric structure marked by the presence of multiple CenH3-domains along the large region of extended chromosome constriction. This centromere type is firstly described in pea (Neumann et al., 2012) and later also confirmed in closely related legume genera *Lathyrus* (Neumann et al., 2015). As genome of the pea is much larger than of *T. castaneum* and has slightly fewer chromosomes (2n=14 in contrast to 2n=20) it does not surprise its centromeres are much larger in absolute values and can

span up to 107 Mb. Further, *T. castaneum* chromosomes revealed bead-like pattern of cCenH3 signals composed of several, in most cases four, individual cCenH3-contaiing domains. This observation is a hallmark of metapolycentrics and is much similar to pea, where long region of centromere contains three to five explicit CenH3 containing regions. The instructive parallel can be drawn by comparing the metapolycentromere of *T. castaneum* and the holocentromere of *M. incognita*. While *T. castaneum* shows separated bead-like CenH3 organization observable at prometaphase, in *M. incognita* smaller CenH3 components appears at extended chromatin level where it shows cluster-like arrangement. Basic centromeric unit of wide range of different structural types could be individual point centromeric sites as shown in *C. elegans* (Steiner & Henikoff, 2014) which could bring the holocentromere closer to a polycentric view of its organization. The difference then ultimately arises from the very distribution of CenH3 regions, whether localized in a smaller area, regional or dispersed along the entire length, and crucially on the existence of primary constriction on which the movement of chromosomes during the division process later depends. The presence of metapolycentricity in very distant species raises questions about their origin and evolution. The existence of large centromeres in *T. castaneum* could be explained by centromere drive hypothesis where centromeres that provide more microtubule attachment by recruiting more centromeric proteins become “stronger” centromeres and are more successful segregated towards the egg in female meiosis. Enhanced recruitment of centromeric nucleosomes can be led by underlying sequence expansion which is further investigated in *T. castaneum*.

Analyses of results obtained by cCenH3 ChIP-seq in *T. castaneum* revealed a repertoire of centromeric DNA sequences including a variety of repetitive DNAs. As suspected, major component of centromeric DNA is the highly abundant TCAST satellite that has shown to comprise 17% of the *T. castaneum* genome (Ugarković et al., 1996). Chromatin fiber staining that enables fine mapping, disclosed very long TCAST arrays that include cCenH3 domains but also that they are not limited only to centromere but rather stretch beyond the cCenH3 region, into the pericentromere. The best-known example where the same satDNA inhabits centromeric as well as pericentromeric regions comes from human centromeres. The highly abundant high-order alpha satellite interacts with CenH3 nucleosomes as centromeric-specific DNA, but can also be scattered as individual monomers in pericentromere regions (reviewed in Hartley & O’Neill, 2019), with some HOR arrays of specific sequence variants

compositions becoming functionally less competent (Aldrup-MacDonald et al., 2016). Species with lower repetitive content, such as *Meloidogyne*, showed centromeric domain dependence exclusively on different repeats sharing a fully conserved 19-bp box showing how the genomic landscape can potentially affect the selectivity and control of CenH3 for a particular sequence. In *T. castaneum*, TCAST has been previously described as a very divergent satellite DNA composed of five subfamilies (Pavlek et al., 2015). CenH3 enrichment analysis did not show any preferential binding of specific TCAST subfamily which is probably due to knowledge that HOR regions are very heterogeneous and combined with extraneous sequence elements which was shown in study before (Vlahović et al., 2017). In addition to satellites, *T. castaneum* centromeres showed presence of non-LTR retrotransposons which beside residing in telomere, can exist as interstitial telomeric sequence (ITS) within centromere, as shown in different *Solanum* species (He et al., 2013). Also, centromeric chromatin in *T. castaneum* is enriched in rDNA-associated sequences, with 5S rDNA as most prominent one, which can be explained with the observation that rDNA tracts serve as an origin of satellite repeats (Macas et al., 2003) and can therefore be propagated in satellite rich environment of centromere. Taken together, *T. castaneum* centromeres are not build on exclusive fraction of TCAST satellite but rather comprised of different subfamilies intermingled with other DNA sequences. Such scenario mostly arose from permissive nature of centromeres together with possibility that one dominant satellite DNA can even foster intrusion of different sequences that are not primarily centromeric.

Nanopore sequencing in centromere studies

The greatest breakthroughs in the field of centromere organization have recently been made by research groups studying the human centromere using long-read sequencing. Indeed, the centromere region is one of the main reasons for the incompleteness of assemblies, which was overcome for the first time this year with the complete sequence of the human genome (Nurk et al., 2021). Prior to their work, 8% of the genome was unfinished or erroneous, corresponding to regions of the centromeric satellite array or short arms of five acrocentric chromosomes that are now truly complete. The development of new technologies, in particular Nanopore sequencing, made it possible to span complex regions with few or even single read, as it depends only on the size of the DNA molecules in the sequencing libraries. In

addition to its use in de-novo genome assemblies or their improvement Nanopore sequencing can be used to detect full-length transcripts and modified bases, offering the possibility of a wide range of downstream data analyzes (reviewed in Wang et al., 2021). Thus, our findings in centromere research of the nematode *M. incognita* and the beetle *T. castaneum* could be further broadened by applying Nanopore sequencing to disclose long-range centromere organization.

Because root-knot nematodes are of great economic importance, several *Meloidogyne* species were sequenced using a whole genome shotgun approach that resulted in scaffold level assembly. The model organism used for our centromere research, *M. incognita*, had its first reference genome published in 2008 (Abad et al., 2008) and was subsequently revised (Blanc-Mathieu et al., 2017). Given that the genome assembly of *M. incognita* is still fragmented, it is not surprising that our analysis of the mapping of centromere-specific sequences have shown that they are largely not contained in published contigs. Therefore, our primary goal is to use Nanopore sequencing in *M. incognita* to investigate the long-range centromere organization. The genome of *T. castaneum* has been sequenced by Sanger sequencing of prepared BAC libraries (Richards et al., 2008) and recently enhanced by incorporating results from Illumina sequencing of long-insert jumping libraries and RNA-seq data (Herndon et al., 2020). The last enhanced genome assembly (Tcas5.2) is comprised of 2148 scaffolds placed into 10 linkage or chromosome groups. Based on the predicted genome size of 204 Mb, the current assembly (166 Mb) is still missing about 20% of the genome. Together with knowledge of centromere specific sequences in both species, these data provide a solid basis for the information needed to prepare the library prior to sequencing. In *T. castaneum,* most of the larger satellite arrays are not includes in the final assembly due to their complexity. In addition, native sequencing (without PCR amplification bias) requires a larger amount of input DNA for library preparation and thus more starting material. This is much more feasible in the case of free-living, easy propagated *T. castaneum* cultures maintained in the laboratory, which has been the commence point for further experiments.

The first premise in studying the organization of long-range organization centromere is to achieve sufficiently long reads obtained by Nanopore. This is solely dependent on the quality of isolated high molecular weight DNA. Using commercial kits and following suggested procedures, as well as phenol-chloroform protocol did not yield adequate DNA for Nanopore sequencing. Best starting point was the Brown and Coleman's

protocol (Brown & Coleman, 2019) that was missing steps of final HMW recover and was not validated for Nanopore or offered information for obtaining exact lengths. Our developed method succeeded in isolating HMW DNA and was validated for the use in library preparation and Nanopore sequencing. Use of liquid nitrogen pulverization overcome difficulties after encountering hard to dissolve DNA pellets. Further, amount of input material is critical as certain input weight is necessary in order to obtain expected DNA quantities. This is probably due to high proportion of non-cellular material (such as fat and chitin) in larvae and adult stages that needs to be accounted for beforehand. Beside increased input material, critical steps include complete resuspension of nuclear pellets, additional gentle pipetting during incubations and spooling of obtained DNA. This way it is ensured DNA will be of high concentrations which is indicated by viscous droplets during elution. Also, spooled HMW DNA forms “jelly-like” mass that is the best sign of efficient relaxation process. Isolated DNA showed optimal absorbance ratios and had most reliable concentration values. Although this method can produce extremely long DNA fragments, it can have negative impact on several steps in library preparation such as adapter ligation which can be less efficient due to entangled DNA. Also, it can later cause bore blockage during sequencing which decreases final output and causes flow cells to have shorter lifespan. This has been suggested to be solvable by applying controlled mechanical shearing of HMW DNA (Tyson, 2020). And indeed, light needle shear increased average N50 values up to 26 kb and allowed multiple library loads for a total output of more than 13 Gb. We speculate these steps to be genome specific, as repetitive fraction of genome known to form secondary structures, can cause difficulties in enzymatic reaction but also when passing through a nanopore. Nevertheless, developed method of HMW isolation was additionally tested in two closely related beetles (*T. confusum* and *T. freemani*) where it also showed excellent performance. The acquired knowledge and the developed protocol can be the basis for application in more distant species, such as *Meloidogyne*. Their soft bodies and— smaller proportion of repetitive DNA in the genome should present mitigating properties for obtaining HMW DNA.

5. Conclusions

Our study of root-knot nematode of the *Meloidogyne* genus represents the first centromere research on species that reproduce exclusively by mitotic parthenogenesis. In contrast to previous findings that centromeric specific histone H3 (CenH3) evolves rapidly, we have demonstrated the almost complete conservation of αCenH3 even in distantly related animal species. This is hypothesized as a possible consequence of subspecialization, primarily for function in mitosis.

In addition to the conserved αCenH3 with dominant expression, we have also found other different CenH3 variants that have evolved rapidly and lost their centromere-associated function. It is likely that these CenH3s represent a “meiotic relict” from sexually reproducing ancestral species.

The underlying DNA sequence of the αCenH3 centromere in *M. incognita* has shown extreme conservation in the form of a preserved 19-bp box occurring within different 50-80 bp tandem repeat arrays. Moreover, conserved cenDNA was found even in related MIG species, suggesting preservation of the αCenH3 centromere in mitotic *Meloidogyne* species.

Chromosomal localization of αCenH3 confirmed holocentrism and disclosed a novel cluster-like organization with domains of markedly different CenH3 density.

Overall, the long-term conservation of CenH3 and the associated 19-bp box within highly evolved tandem repeats suggest that CenH3 and cenDNA can achieve an equilibrium in mitosis where they can coexist for a long period of time.

The red flour beetle *T. castaneum* showed an unexpected metapolycentric centromere organization marked by the presence of multiple cCenH3-domains. Accounting for approximately 40% of chromosome length, their centromeres are the longest identified in insects. Furthermore, the TCAST satellite DNA, which comprises one-sixth of the genome was found to be the major cenDNA constituent. In addition, several other types of repetitive DNA were found to be mixed with TCAST variants, including retrotransposons and rDNA-associated sequences.
As one of the most recently discovered centromere types, our study of *T. castaneum* metapolycentromere provides valuable new insight, particularly by challenging the

long-prevailing holocentric and monocentric views of centromeres. The holocentric *M. incognita* and the metapolycentric *T. castaneum* display some never before seen characteristics of the centromere. Both are a true example of the incredible diversity in terms of centromere structure and organization despite its conserved function.

Due to repetitive nature of the centromere, the future of its research depends on technological advances such as long-read Nanopore sequencing. It can span arrays hundreds of thousands kilobases long and enable more complete genome assemblies. We have developed a protocol for reliable Nanopore sequencing in *T. castaneum* that will give us a better grasp over centromere's long-range organization.

6. References

Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M.-C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V, Hamamouch, N., … Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, *26*(8), 909–915. https://doi.org/10.1038/nbt.1482

Agrios, G. (2004). Plant pathology: Fifth edition. In *Plant Pathology: Fifth Edition* (Vol. 9780080473). https://doi.org/10.1016/C2009-0-02037-6

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). *Molecular Biology of the Cell*. New York: Garland Science.

Aldrup-MacDonald, M. E., Kuo, M. E., Sullivan, L. L., Chew, K., & Sullivan, B. A. (2016). Genomic variation within alpha satellite DNA influences centromere location on human chromosomes with metastable epialleles. *Genome Research*, *26*(10), 1301–1311. https://doi.org/10.1101/gr.206706.116

Barra, V., & Fachinetti, D. (2018). The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nature Communications*, *9*(1), 4340. https://doi.org/10.1038/s41467-018-06545-y

Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (Pat. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, *23*(3), 415–420. https://doi.org/10.1007/s10577-015-9499-z

Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J. M., Da Rocha, M., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J. F., Kozlowski, D. K., Cazareth, J., Couloux, A., Da Silva, C., Guy, J., Kim-Jo, Y. J., Rancurel, C., Schiex, T., Abad, P., … Danchin, E. G. J. (2017). Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genetics*, *13*(6), e1006777. https://doi.org/10.1371/journal.pgen.1006777

Bloom, K. S. (2014). Centromeric heterochromatin: The primordial segregation machine. *Annual Review of Genetics*, *48*(September), 457–484. https://doi.org/10.1146/annurev-genet-120213-092033

Brown, S. J., & Coleman, M. (2019). Isolation of high molecular weight DNA from insects. *Methods in Molecular Biology*, *1858*, 27–32. https://doi.org/10.1007/978-1-4939-8775-7_3

Caputi, F. F., Candeletti, S., & Romualdi, P. (2017). Epigenetic Approaches in Neuroblastoma Disease Pathogenesis. In *Neuroblastoma - Current State and Recent Updates*. https://doi.org/10.5772/intechopen.69566

Castagnone-Sereno, P. (2006). Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity*, *96*(4), 282–289. https://doi.org/10.1038/sj.hdy.6800794

Castagnone-Sereno, P, & Danchin, E. G. J. (2014). Parasitic success without sex – the nematode experience. *Journal of Evolutionary Biology*, *27*(7), 1323–1333. https://doi.org/10.1111/jeb.12337

Castagnone-Sereno, Philippe, Danchin, E. G. J., Perfus-Barbeoch, L., & Abad, P. (2013). Diversity and Evolution of Root-Knot Nematodes, Genus *Meloidogyne*: New Insights from the Genomic Era. *Annual Review of Phytopathology*, *51*(1), 203–220. https://doi.org/10.1146/annurev-phyto-082712-102300

Castagnone-Sereno, Philippe, Semblat, J. P., Leroy, F., & Abad, P. (1998). A New AluI Satellite DNA in the Root-Knot Nematode *Meloidogyne fallax*: Relationships with Satellites from the Sympatric Species *M. hapla* and *M. chitwoodi*. *Molecular Biology and Evolution*, *15*(9), 1115–1122. https://doi.org/10.1093/oxfordjournals.molbev.a026019

Chang, C. H., Chavan, A., Palladino, J., Wei, X., Martins, N. M. C., Santinello, B., Chen, C. C., Erceg, J., Beliveau, B. J., Wu, C. T., Larracuente, A. M., & Mellone, B. G. (2019). Islands of retroelements are major components of *Drosophila* centromeres. In *PLoS Biology* (Vol. 17, Issue 5). https://doi.org/10.1371/journal.pbio.3000241

Cohen-Sharir, Y., McFarland, J. M., Abdusamad, M., Marquis, C., Bernhard, S. V., Kazachkova, M., Tang, H., Ippolito, M. R., Laue, K., Zerbib, J., Malaby, H. L. H., Jones, A., Stautmeister, L. M., Bockaj, I., Wardenaar, R., Lyons, N., Nagaraja, A., Bass, A. J., Spierings, D. C. J., … Ben-David, U. (2021). Aneuploidy renders cancer cells vulnerable to mitotic checkpoint inhibition. *Nature*, *590*(7846), 486–491. https://doi.org/10.1038/s41586-020-03114-6

Cottarel, G., Shero, J. H., Hieter, P., & Hegemann, J. H. (1989). A 125 bp CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in *Saccharomyces cerevisiae*. *Trends in Genetics*, *5*(C), 322–324. https://doi.org/10.1016/0168-9525(89)90119-4

Drinnenberg, I. A., deYoung, D., Henikoff, S., & Malik, H. S. ing. (2014). Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *ELife*, *3*, 1–19. https://doi.org/10.7554/eLife.03676

Felekkis, K., & Voskarides, K. (2015). Genomic elements in health, disease and evolution: Junk DNA. *Genomic Elements in Health, Disease and Evolution: Junk DNA*, 1–311. https://doi.org/10.1007/978-1-4939-3070-8

Feliciello, I., Chinali, G., & Ugarković, D. (2011). Structure and population dynamics of the major satellite DNA in the red flour beetle *Tribolium castaneum*. *Genetica*, *139*(8), 999–1008. https://doi.org/10.1007/s10709-011-9601-1

Fukagawa, T., & Earnshaw, W. C. (2014). Neocentromeres. *Current Biology*, *24*(19), R946–R947. https://doi.org/10.1016/j.cub.2014.08.032

Gamba, R., & Fachinetti, D. (2020). From evolution to function: Two sides of the same CENP-B coin? *Experimental Cell Research*, *390*(2). https://doi.org/10.1016/j.yexcr.2020.111959

Guenatri, M., Bailly, D., Maison, C., & Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *Journal of Cell Biology*, *166*(4), 493–505. https://doi.org/10.1083/jcb.200403109

Hara, M., & Fukagawa, T. (2017). Critical Foundation of the Kinetochore: The Constitutive Centromere-Associated Network (CCAN). *Progress in Molecular and Subcellular Biology*, *56*, 29–57. https://doi.org/10.1007/978-3-319-58592-5_2

Hartley, G., & O'Neill, R. J. (2019). Centromere repeats: Hidden gems of the genome. *Genes*, *10*(3), 223. https://doi.org/10.3390/genes10030223

He, L., Liu, J., Torres, G. A., Zhang, H., Jiang, J., & Xie, C. (2013). Interstitial telomeric repeats are enriched in the centromeres of chromosomes in *Solanum* species. *Chromosome Research*, *21*(1), 5–13. https://doi.org/10.1007/s10577-012-9332-x

Henikoff, S., Ahmad, K., & Malik, H. S. (2001). The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science*, *293*(5532), 1098–1102. https://doi.org/10.1126/science.1062939

Herndon, N., Shelton, J., Gerischer, L., Ioannidis, P., Ninova, M., Dönitz, J., Waterhouse, R. M., Liang, C., Damm, C., Siemanowski, J., Kitzmann, P., Ulrich, J., Dippel, S., Oberhofer, G., Hu, Y., Schwirz, J., Schacht, M., Lehmann, S., Montino, A., … Bucher, G. (2020). Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. *BMC Genomics*, *21*(1), 1–13. https://doi.org/10.1186/s12864-019-6394-6

Huang, Y. C., Lee, C. C., Kao, C. Y., Chang, N. C., Lin, C. C., Shoemaker, D., & Wang, J. (2016). Evolution of long centromeres in fire ants. *BMC Evolutionary Biology*, *16*(1), 1–14. https://doi.org/10.1186/s12862-016-0760-7

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., St. John, O., Wild, R., Hammond, P. M., Ahrens, D., Balke, M., Caterino, M. S., Gómez-Zurita, J., Ribera, I., Barraclough, T. G., Bocakova, M., Bocak, L., & Vogler, A. P. (2007). A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science*, *318*(5858), 1913–1916. https://doi.org/10.1126/science.1146954

Ishii, T., Juranić, M., Maheshwari, S., Bustamante, F. de O., Vogt, M., Salinas-Gamboa, R., Dreissig, S., Gursanscky, N., How, T., Demidov, D., Fuchs, J., Schubert, V., Spriggs, A., Vielle-Calzada, J. P., Comai, L., Koltunow, A. M. G., & Houben, A. (2020). Unequal contribution of two paralogous CENH3 variants in cowpea centromere function. *Communications Biology*, *3*(1). https://doi.org/10.1038/s42003-020-01507-x

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., … Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, *36*(4), 338–345. https://doi.org/10.1038/nbt.4060

Jain, M., Olsen, H. E., Turner, D. J., Stoddart, D., Bulazel, K. V., Paten, B., Haussler, D., Willard, H. F., Akeson, M., & Miga, K. H. (2018). Linear assembly of a human centromere on the Y chromosome. *Nature Biotechnology*, *36*(4), 321–323. https://doi.org/10.1038/nbt.4109

Juan, C., Vazquez, P., Rubio, J. M., Petitpierre, E., & Hewitt, G. M. (1993). Presence of highly repetitive DNA sequences in *Tribolium* flour-beetles. *Heredity*, *70*(1), 1–8. https://doi.org/10.1038/hdy.1993.1

Kang, Y., Wang, J., Neff, A., Kratzer, S., Kimura, H., & Davis, R. E. (2016). Differential Chromosomal Localization of Centromeric Histone CENP-A Contributes to Nematode Programmed DNA Elimination. *Cell Reports*, *16*(9), 2308–2316. https://doi.org/10.1016/j.celrep.2016.07.079

Kasinathan, S., & Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. *Molecular Biology and Evolution*, *35*(4), 949–962. https://doi.org/10.1093/molbev/msy010

Komissarov, A. S., Gavrilova, E. V., Demin, S. J., Ishov, A. M., & Podgornaya, O. I. (2011). Tandemly repeated DNA families in the mouse genome. *BMC Genomics*, *12*. https://doi.org/10.1186/1471-2164-12-531

Kumekawa, N., Hosouchi, T., Tsuruoka, H., & Kotani, H. (2001). The size and sequence organization

of the centromeric region of *Arabidopsis thaliana* chromosome 4. *DNA Research*, *8*(6), 285–290. https://doi.org/10.1093/dnares/8.6.285

Kursel, L. E., & Malik, H. S. (2017). Recurrent gene duplication leads to diverse repertoires of centromeric histones in *Drosophila* species. *Molecular Biology and Evolution*, *34*(6), 1445–1462. https://doi.org/10.1093/molbev/msx091

Kursel, L. E., Welsh, F. C., & Malik, H. S. (2020). Ancient Coretention of Paralogs of *Cid* Centromeric Histones and *Cal1* Chaperones in Mosquito Species. *Molecular Biology and Evolution*, *37*(7), 1949–1963. https://doi.org/10.1093/molbev/msaa056

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Logsdon, G. A., Vollger, M. R., Hsieh, P. H., Mao, Y., Liskovykh, M. A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P. C., Rhie, A., de Lima, L. G., Dvorkina, T., Porubsky, D., Harvey, W. T., Mikheenko, A., Bzikadze, A. V., Kremitzki, M., Graves-Lindsay, T. A., Jain, C., … Eichler, E. E. (2021). The structure, function and evolution of a complete human chromosome 8. *Nature*, *593*(7857), 101–107. https://doi.org/10.1038/s41586-021-03420-7

Lopes, M., Louzada, S., Gama-Carvalho, M., & Chaves, R. (2021). Genomic tackling of human satellite dna: Breaking barriers through time. *International Journal of Molecular Sciences*, *22*(9). https://doi.org/10.3390/ijms22094707

Lunt, D. H. (2008). Genetic tests of ancient asexuality in Root Knot Nematodes reveal recent hybrid origins. *BMC Evolutionary Biology*, *16*, 1–16. https://doi.org/10.1186/1471-2148-8-194

Macas, J., Navrátilová, A., & Mészáros, T. (2003). Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on *Vicia sativa* chromosomes. *Chromosoma*, *112*(3), 152–158. https://doi.org/10.1007/s00412-003-0255-3

Malik, H. S., & Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nature Structural Biology*, *10*(11), 882–891. https://doi.org/10.1038/nsb996

Malik, H. S., & Henikoff, S. (2009). Major Evolutionary Transitions in Centromere Complexity. *Cell*, *138*(6), 1067–1082. https://doi.org/10.1016/j.cell.2009.08.036

Mandrioli, M., & Manicardi, G. C. (2020). Holocentric chromosomes. *PLoS Genetics*, *16*(7), 1–11. https://doi.org/10.1371/journal.pgen.1008918

Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A., & Landsman, D. (2005). Histone structure and nucleosome stability. *Expert Review of Proteomics*, *2*(5), 719–729. https://doi.org/10.1586/14789450.2.5.719

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., Pellino, M., Fuchs, J., Ma, W., Kuhlmann, M., Brandt, R., Vanzela, A. L. L., Beseda, T., Šimková, H., Pedrosa-Harand, A., & Houben, A. (2015). Holocentromeres in *Rhynchospora* are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(44), 13633–13638. https://doi.org/10.1073/pnas.1512255112

Mckinley, K. L., & Cheeseman, I. M. (2016). The molecular basis for centromere identity and function. *Nature Reviews Molecular Cell Biology*, *17*(1), 16–29. https://doi.org/10.1038/nrm.2015.5

McNulty, S. M., Sullivan, L. L., & Sullivan, B. A. (2017). Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. *Developmental Cell*, *42*(3), 226-240.e6. https://doi.org/10.1016/j.devcel.2017.07.001

Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J. G., Sebra, R., Peluso, P., Eid, J., Rank, D., Garcia, J. F., Derisi, J. L., Smith, T., Tobias, C., Ross-Ibarra, J., Korf, I., & Chan, S. W. L. (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology*, *14*(1), 1–20.

Melters, D. P., Paliulis, L. V., Korf, I. F., & Chan, S. W. L. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Research*, *20*(5), 579–593. https://doi.org/10.1007/s10577-012-9292-1

Meštrović, N., Castagnone-Sereno, P., & Plohl, M. (2006a). High conservation of the differentially amplified MPA2 satellite DNA family in parthenogenetic root-knot nematodes. *Gene*, *376*(2), 260–267. https://doi.org/10.1016/j.gene.2006.04.008

Meštrović, N., Castagnone-Sereno, P., & Plohl, M. (2006b). Interplay of Selective Pressure and Stochastic Events Directs Evolution of the MEL172 Satellite DNA Library in Root-Knot Nematodes. *Molecular Biology and Evolution*, *23*(12), 2316–2325. https://doi.org/10.1093/molbev/msl119

Meštrović, N., Pavlek, M., Car, A., Castagnone-Sereno, P., Abad, P., & Plohl, M. (2013). Conserved DNA Motifs, Including the CENP-B Box-like, Are Possible Promoters of Satellite DNA Array Rearrangements in Nematodes. *PLoS ONE*, *8*(6), e67328. https://doi.org/10.1371/journal.pone.0067328

Meštrović, N., Plohl, M., & Castagnone-Sereno, P. (2009). Relevance of satellite DNA genomic distribution in phylogenetic analysis: A case study with root-knot nematodes of the genus *Meloidogyne*. *Molecular Phylogenetics and Evolution*, *50*(1), 204–208. https://doi.org/10.1016/j.ympev.2008.10.013

Meštrović, N., Plohl, M., Mravinac, B., & Ugarković, Đ. (1998). Evolution of Satellite DNAs from the genus *Palorus*-Experimental Evidence for the “Library” Hypothesis. *Molecular Biology and Evolution*, *15*(8), 1062–1068. https://doi.org/10.1093/oxfordjournals.molbev.a026005

Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G. A., Schneider, V. A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., … Phillippy, A. M. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, *585*(7823), 79–84. https://doi.org/10.1038/s41586-020-2547-7

Monen, J., Hattersley, N., Muroyama, A., Stevens, D., Oegema, K., & Desai, A. (2015). Separase Cleaves the N-Tail of the CENP-A Related Protein CPAR-1 at the Meiosis I Metaphase-Anaphase Transition in *C . elegans*. *PloS One*, *10*(4), e0125382. https://doi.org/10.1371/journal.pone.0125382

Monen, J., Maddox, P. S., Hyndman, F., Oegema, K., & Desai, A. (2005). Differential role of CENP-A

in the segregation of holocentric *C. elegans* chromosomes during meiosis and mitosis. *Nature Cell Biology*, *7*(12), 1248–1255. https://doi.org/10.1038/ncb1331

Mravinac, B., & Plohl, M. (2010). Parallelism in Evolution of Highly Repetitive DNAs in Sibling Species. *Molecular Biology and Evolution*, *27*(8), 1857–1867. https://doi.org/10.1093/molbev/msq068

Mravinac, B., Plohl, M., & Ugarković, D. (2004). Conserved patterns in the evolution of *Tribolium* satellite DNAs. *Gene*, *332*(1–2), 169–177. https://doi.org/10.1016/j.gene.2004.02.055

Muller, H., Gil, J., & Drinnenberg, I. A. (2019). The Impact of Centromeres on Spatial Genome Architecture. *Trends in Genetics*, *35*(8), 565–578. https://doi.org/10.1016/j.tig.2019.05.003

Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., Novák, P., Wanner, G., & Macas, J. (2012). Stretching the Rules: Monocentric Chromosomes with Multiple Centromere Domains. *PLoS Genetics*, *8*(6), e1002777. https://doi.org/10.1371/journal.pgen.1002777

Neumann, P., Pavlíková, Z., Koblížková, A., Fuková, I., Jedličková, V., Novák, P., & Macas, J. (2015). Centromeres Off the Hook: Massive Changes in Centromere Size and Structure Following Duplication of CenH3 Gene in *Fabeae* Species. *Molecular Biology and Evolution*, *32*(7), 1862–1879. https://doi.org/10.1093/molbev/msv070

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., … Phillippy, A. M. (2021). The complete sequence of a human genome. *BioRxiv*. https://doi.org/10.1101/2021.05.26.445798

Oliveira, L., Neumann, P., Jang, T. S., Klemme, S., Schubert, V., Koblížková, A., Houben, A., & Macas, J. (2020). Mitotic Spindle Attachment to the Holocentric Chromosomes of *Cuscuta europaea* Does Not Correlate With the Distribution of CENH3 Chromatin. *Frontiers in Plant Science*, *10*(January), 1–11. https://doi.org/10.3389/fpls.2019.01799

Panchenko, T., Sorensen, T. C., Woodcock, C. L., Kan, Z. Y., Wood, S., Resch, M. G., Luger, K., Englander, S. W., Hansen, J. C., & Black, B. E. (2011). Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(40), 16588–16593. https://doi.org/10.1073/pnas.1113621108

Pavlek, M., Gelfand, Y., Plohl, M., & Meštrović, N. (2015). Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. *DNA Research*, *22*(6), 387–401. https://doi.org/10.1093/dnares/dsv021

Peterson, C. L., & Laniel, M. A. (2004). Histones and histone modifications. *Current Biology : CB*, *14*(14), 546–551. https://doi.org/10.1016/j.cub.2004.07.007

Piras, F. M., Nergadze, S. G., Magnani, E., Bertoni, L., Attolini, C., Khoriauli, L., Raimondi, E., & Giulotto, E. (2010). Uncoupling of Satellite DNA and Centromeric Function in the Genus *Equus*. *PLoS Genetics*, *6*(2), e1000845. https://doi.org/10.1371/journal.pgen.1000845

Plohl, M., Meštrović, N., & Mravinac, B. (2014). Centromere identity from the DNA point of view.

Chromosoma, *123*(4), 313–325. https://doi.org/10.1007/s00412-014-0462-0

Pointer, M. D., Gage, M. J. G., & Spurgin, L. G. (2021). *Tribolium* beetles as a model system in evolution and ecology. *Heredity*, *126*(6), 869–883. https://doi.org/10.1038/s41437-021-00420-1

Presting, G. G. (2018). Centromeric retrotransposons and centromere function. *Current Opinion in Genetics and Development*, *49*, 79–84. https://doi.org/10.1016/j.gde.2018.03.004

Prosée, R. F., Wenda, J. M., & Steiner, F. A. (2020). Adaptations for centromere function in meiosis. *Essays in Biochemistry*, *64*(2), 193–203. https://doi.org/10.1042/ebc20190076

Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S., Denell, R., Beeman, R. W., Gibbs, R., Bucher, G., Friedrich, M., Grimmelikhuijzen, C. J. P., Klingler, M., Lorenzen, M., Roth, S., Schröder, R., Tautz, D., Zdobnov, E. M., Muzny, D., Attaway, T., Bell, S., … Grossmann, D. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, *452*(7190), 949–955. https://doi.org/10.1038/nature06784

Rošić, S., & Erhardt, S. (2016). No longer a nuisance: Long non-coding RNAs join CENP-A in epigenetic centromere regulation. *Cellular and Molecular Life Sciences*, *73*(7), 1387–1398. https://doi.org/10.1007/s00018-015-2124-7

Sansregret, L., & Swanton, C. (2017). The role of aneuploidy in cancer evolution. *Cold Spring Harbor Perspectives in Medicine*, *7*(1), 1–17. https://doi.org/10.1101/cshperspect.a028373

Schön, I., Martens, K., & Dijk, P. (2009). Lost Sex. In I. Schön, K. Martens, & P. Dijk (Eds.), *Lost Sex*. Springer Netherlands. https://doi.org/10.1007/978-90-481-2770-2

Schwenk, K., Brede, N., & Streit, B. (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1505), 2805–2811. https://doi.org/10.1098/rstb.2008.0055

Scott, K. C., & Sullivan, B. A. (2014). Neocentromeres: A place for everything and everything in its place. *Trends in Genetics*, *30*(2), 66–74. https://doi.org/10.1016/j.tig.2013.11.003

Shang, W. H., Hori, T., Toyoda, A., Kato, J., Popendorf, K., Sakakibara, Y., Fujiyama, A., & Fukagawa, T. (2010). Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Research*, *20*(9), 1219–1228. https://doi.org/10.1101/gr.106245.110

Smurova, K., & De Wulf, P. (2018). Centromere and Pericentromere Transcription: Roles and Regulation … in Sickness and in Health. *Frontiers in Genetics*, *9*(December), 1–26. https://doi.org/10.3389/fgene.2018.00674

Steiner, F. A., & Henikoff, S. (2014). Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *ELife*, *3*(e02025), 1–22. https://doi.org/10.7554/eLife.02025

Szitenberg, A., Salazar-Jaramillo, L., Blok, V. C., Laetsch, D. R., Joseph, S., Williamson, V. M., Blaxter, M. L., & Lunt, D. H. (2017). Comparative Genomics of Apomictic Root-Knot Nematodes: Hybridization, Ploidy, and Dynamic Genome Change. *Genome Biology and Evolution*, *9*(10), 2844–2861. https://doi.org/10.1093/gbe/evx201

Talbert, P. B., & Henikoff, S. (2020). What makes a centromere? *Experimental Cell Research*, *389*(2), 111895. https://doi.org/10.1016/j.yexcr.2020.111895

Ting, D. T., Lipson, D., Paul, S., Brannigan, B. W., Akhavanfard, S., Coffman, E. J., Contino, G.,

Deshpande, V., Iafrate, A. J., Letovsky, S., Rivera, M. N., Bardeesy, N., Maheswaran, S., & Haber, D. A. (2011). Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. *Science*, *331*(6017), 593–596. https://doi.org/10.1126/science.1200801

Tollervey, J. R., & Lunyak, V. V. (2012). Epigenetics: judge, jury and executioner of stem cell fate. *Epigenetics*, *7*(8), 823–840. https://doi.org/10.4161/epi.21141

Triantaphyllou, A. C. (1981). Oogenesis and the Chromosomes of the Parthenogenetic Root-knot Nematode *Meloidogyne incognita*. *Journal of Nematology*, *13*(2), 95–104.

Tyson, J. (2020). Rocky Mountain adventures in Genomic DNA sample preparation , ligation protocol optimisation / simplification and Ultra long read generation. *Protocols.Io*, 1–5. https://doi.org/dx.doi.org/10.17504/protocols.io.7euhjew

Ugarković, D., Podnar, M., & Plohl, M. (1996). Satellite DNA of the Red Flour Beetle *Tribolium castaneum* - Comparative Study of Satellites from the Genus *Tribolium*. *Molecular Biology and Evolution*, *13*(8), 1059–1066. https://doi.org/10.1093/oxfordjournals.molbev.a025668

Vlahović, I., Gluncić, M., Rosandić, M., Ugarković, D. I., & Paar, V. (2017). Regular higher order repeat structures in beetle *Tribolium castaneum* genome. *Genome Biology and Evolution*, *9*(10), 2668–2680. https://doi.org/10.1093/gbe/evw174

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, *7*. https://doi.org/10.1038/s41587-021-01108-x

Watanabe, Y. (2005). Sister chromatid cohesion along arms and at centromeres. *Trends in Genetics*, *21*(7), 405–412. https://doi.org/10.1016/j.tig.2005.05.009

Wells, J. N., & Feschotte, C. (2020). A Field Guide to Eukaryotic Transposable Elements. *Annual Review of Genetics*, *54*, 539–561. https://doi.org/10.1146/annurev-genet-040620-022145

Wesemael, W. M. L. (2007). Biology and management of the root-knot nematode *Meloidogyne chitwoodi* in field vegetable crops. Ghent University.

Wolfgruber, T. K., Sharma, A., Schneider, K. L., Albert, P. S., Koo, D. H., Shi, J., Gao, Z., Han, F., Lee, H., Xu, R., Allison, J., Birchler, J. A., Jiang, J., Dawe, R. K., & Presting, G. G. (2009). Maize centromere structure and evolution: Sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. *PLoS Genetics*, *5*(11), 13–16. https://doi.org/10.1371/journal.pgen.1000743

Wong, C. Y. Y., Ling, Y. H., Mak, J. K. H., Zhu, J., & Yuen, K. W. Y. (2020). “Lessons from the extremes: Epigenetic and genetic regulation in point monocentromere and holocentromere establishment on artificial chromosomes.” *Experimental Cell Research*, *390*(2), 111974. https://doi.org/10.1016/j.yexcr.2020.111974

Zhao, J., Bacolla, A., Wang, G., & Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences*, *67*(1), 43–62. https://doi.org/10.1007/s00018-009-0131-2

7. Summary

Centromeres are chromosomal regions essential for kinetochore assembly and attachment to the spindle, ensuring proper chromosome segregation. They are determined epigenetically by the presence of centromere specific histone H3 variant (CenH3). The protein component is associated with centromeric DNA (cenDNA) and both are required for incorporation into centromeric nucleosomes to maintain centromere function. Despite the conserved role of centromeres, associated CenH3 and DNA evolve rapidly, which is explained with centromere drive model that accounts for observed variability of centromeric region and points out female meiosis as crucial factor.

We investigated nematode centromeres in three closely related, polyploid species of the *Meloidogyne incognita* group that reproduce asexually, by mitotic parthenogenesis and in the distantly related asexual *M. hapla* which reproduce by meiotic parthenogenesis. Duplication of the CenH3 gene was found that occurred in the common ancestor of the *Meloidogyne* species. One of them (αCenH3) preserved in all extant *Meloidogyne* species whereas the other evolved rapidly under positive selection into four different variants (αβγδCenH3s). This observed evolution of CenH3 genes/proteins in *Meloidogyne* species suggests the subspecialization of CenH3s in sexual ancestral species. Immunostaining and transcription analysis performed on *M. incognita* revealed a dominant role of αCenH3 protein, whereas the other variants have lost their function in mitosis. Chromosomal localization of αCenH3 showed cluster-like centromeric organization with different levels of deposition along the chromosomes forming distinct types of centromeres. The results of native chromatin immunoprecipitation followed by sequencing (ChIP-seq) revealed αCenH3-associated DNA. This centromeric DNA is comprised of tandem repeats with divergent monomers that in spite of sequence variability share a completely conserved 19-bp box. Beside *M. incognita,* the conservation of centromeric protein and underlying DNA was also found in two closely related species. The observed pattern of preservation of the centromeric region, at the protein and DNA level, was explained by the absence of centromere drive in mitosis, thus allowing CenH3 and associated DNA to achieve an equilibrium that can persist over a long period of time.

Holocentrics have kinetochore activity distributed along the entire chromosome length while monocentrics show a single, primary constriction visible at metaphase chromosomes. In addition to the holocentric *Meloidogyne*, centromere was further

explored in the insect model organism *Tribolium castaneum*. We identified cCenH3 protein and successfully produced antibody specifically recognizing centromeric variant of histone H3, cCenH3 in *T. castaneum*. Centromere was localized on chromosomes revealing unusually elongated centromeres, making them longest insect regional centromeres, comprising approximately 40% of the chromosome length. In addition, centromere in *T. castaneum* showed metapolycentric structure with several individual cCenH3-containing domains. By applying ChIP-seq we analyzed cCenH3 chromatin that showed to be assembled upon highly abundant TCAST satellite together with several other repetitive DNAs. Also, TCAST as major centromeric constituent is not confined only to CenH3 regions but rather spreads into pericentromere. (Peri)centromere regions represent the largest part missing from genome assemblies. Similar observation is true for *Meloidogyne*, whose genome with much less repetitive content, still contains an underrepresented proportion of centromere specific sequences in the assembly.

In order to address long-range centromere organization, we opted for Nanopore sequencing that has no limitations in reads lengths. Preparation of DNA for Nanopore sequencing that is long enough offers the possibility to span over large repetitive arrays. Next, we optimized method for isolation of high molecular weight DNA in *T. castaneum* by successfully circumventing common obstacles encountered when using standardized protocols. Upon full relaxation and light sheering, DNA had optimal quality and was in sufficient quantity to perform library preparation yielding average N50 values up to 26 kb. This approach of long-read Nanopore sequencing provides a basis for further research into the organization of centromeric repeats in *Meloidogyne* and *Tribolium* species.

8. Sažetak

Centromere su kromosomske regije neophodne za sastavljanje kinetohore i vezanje na diobeno vreteno tako osiguravajući pravilnu segregaciju kromosoma. One su epigenetski određene postojanjem centromerno specifične varijante histona H3 (CenH3). Proteinska komponenta je povezana s centromernom DNA i obje su potrebne za oblikovanje centromernih nukleosoma kako bi se održala funkcija centromera. Unatoč očuvanoj ulozi centromera, pripadajući CenH3 protein i DNA brzo evoluiraju, što se objašnjava modelom centromernog usmjeravanja (eng. *centromere drive*) koji objašnjava uočenu varijabijabilnost centromerne regije i ističe mejozu kod ženki kao ključan faktor.

Istraživali smo centromere nematoda u tri blisko srodne, poliploidne vrste iz skupine *Meloidogyne incognita* koje se razmnožavaju nespolno, mitotičkom partenogenezom i u udaljenoj također nespolnoj vrsti *M. hapla* koja se razmnožava mejotskom partenogenezom. Pronađena je duplikacija CenH3 gena za koju je utvrđeno da se dogodila u zajedničkom pretku vrsta *Meloidogyne*. Jedan od pronađenih CenH3 gena/proteina (αCenH3) se pokazao očuvan u svim postojećim vrstama, dok su ostali ubrzano evoluirali pod utjecajem pozitivne selekcije u četiri različite varijante (αβγδCenH3s). Ova opažena evolucija CenH3 gena/proteina u vrstama *Meloidogyne* sugerira njegovu subspecijalizaciju u pretcima koji su se spolno razmnožavali. Imunofluorescencija i analiza transkriptoma vrste *M. incognita* pokazala je dominantnu ulogu αCenH3 proteina, dok su ostale varijante izgubile svoju funkciju u mitozi. Kromosomska lokalizacija αCenH3 proteina je pokazala organizaciju centromere u formi grupiranja uz različite međuodnose i uzorke slaganja duž kromosoma, tvoreći različite strukture centromera. Rezultati dobiveni metodom nativne kromatinske imunoprecipitacije uz sekvenciranje kratkih odsječaka (ChIP-seq) su otkrili DNA povezanu s αCenH3 proteinom. Navedena centromerna DNA se sastoji od uzastopnih ponavljanja u koja su uključeni divergentni monomeri, koji usprkos njihovoj razlici u sekvenci posjeduju potpuno konzervirani motiv od 19-pb. Osim u vrsti *M. incognita*, očuvanje centromernog proteina i pripadajuće DNA pronađeno je također i u dvije blisko srodne vrste. Primjećeni obrazac očuvanja centromerne regije, na proteinskoj i DNA razini smo objasnili odsutnošću centromernog usmjeravanja u mitozi, što dopušta CenH3 proteinu i pripadajućoj DNA da postignu ravnotežu koja može opstati tijekom dužeg vremenskog razdoblja.

Aktivnost kinetohore je u holocentričnim vrstama raspodjeljena duž kromosoma, za razliku od monocentričnih vrsta koje imaju mjesto primarne konstrikcije vidljivo na metafaznim kromosomima. Uz holocentrične predstavnike vrsta roda *Meloidogyne*, centromera je također istražena u modelnom organizmu kukca vrste *Tribolium castaneum*. Identificirali smo centromernu varijantu histona H3 kod kukca vrste *T. castaneum* i uspješno proizveli antitijelo koje specifično prepoznaje cCenH3 protein. Centromera je lokalizirana na kromosomima otkrivajući neobično izdužene centromere kod vrste *T. castaneum*, što ih čini najduljim regionalnim centromerama kukaca uz pokrivanje približno 40% duljine kromosoma. Također je primjećeno da imaju metapolicentričnu strukturu sastavljenu od nekoliko pojedinačnih domena koje sadrže cCenH3. Primjenom metode ChIP-seq smo analizirali cCenH3 kromatin za koji se pokazalo da je izgrađen od visoko zastupljene TCAST satelitne DNA zajedno s nekoliko drugih ponavljajućih DNA. Također je pronađeno da TCAST kao glavna centromerna sastavnica nije ograničena samo na CenH3 regije, već se širi u pericentromerni dio kromosoma. (Peri)centromerne regije predstavljaju najveći dio koji nedostaje u posloženim genoma. Slično opažanje vrijedi i za *Meloidogyne*, jer iako njihovi genomi sadrže daleko manje ponavljajućih sekvenci još uvijek su te sekvence podzastupljene u posloženom genomu.

Kako bismo pristupili proučavanju centomere u dugačkim potezima, odlučili smo se za Nanopore sekvenciranje koje nema ograničenja u samim duljinama očitanja. Ukoliko je DNA pripremljena za Nanopore sekvenciranje dovoljno dugačka, nudi mogućnost da se premoste veliki nizovi ponavljajućih sekvenci. Stoga smo optimizirali metodu za izolaciju visokomolekularne DNA u vrsti *T. castaneum* i uspješno nadvladali uobičajene poteškoće koje se susreću pri korištenju standardiziranih protokola za izolaciju. Između ostaloga, optimizacija je uključivala potpuno opuštanje DNA molekule i djelomičnu fragmentaciju laganim provlaćenjem krozu iglu. Tako pripremljana DNA je imala zadovoljavajuću kvalitetu i bila je u dovoljnoj količini za pripremu DNA knjižnice, dajući prosječne vrijednosti N50 do 26 kb. Navedeni pristup Nanopore sekvenciranja dugačkih očitanja daje osnovu za daljnja istraživanja organizacije centromernih ponavljanja u vrstama *Meloidogyne* i *Tribolium*.

9. Curriculum vitae and publication list

Evelin Despot Slade

Researcher Identification Number: 362280

Address: Županijska 6, Osijek, Croatia
Phone +385 (98) 190 1121
E-mail: evelin.despot.slade@irb.hr

Education

2016 –	Josip Juraj Strossmayer University of Osijek, Osijek, Croatia. University Postgraduate Interdisciplinary Doctoral Study, Molecular Biosciences
2013 – 2015	Swiss Federal Institute of Technology (ETHZ), Zurich, Switzerland. M.S. studies in Cell Biology, D-BIOL
2010 – 2013	University of Zagreb, Faculty of Science, Zagreb, Croatia. B.S. studies in Molecular Biology
2006 – 2010	Natural Science and Mathematics High School, Osijek, Croatia

Professional Experience

2017 -	Ruđer Bošković Institute, Zagreb, Croatia. Research assistant
2014 - 2015	ETH Zurich, D-HEST, Cartilage Engineering + Regeneration, Zürich, Switzerland. Master thesis: “Role of Axin2/ β-catenin in chondrocyte de/re-differentiation”
2014	ETH Zurich, D-BIOL, Institute of Molecular Health Sciences, Zürich, Switzerland, Semester project: “Physical Characterization of Mouse Haploid Embryonic Stem Cells”
2013 - 2014	ETH Zurich, D-HEST, Cartilage Engineering + Regeneration, Zürich, Switzerland, Semester project: “Lentiviral Transduction of Human Chondrocytes Using Collagen 2 Reporter”
2013	University of Zagreb, Faculty of Science, Department of Molecular biology, Zagreb, Croatia, Laboratory skill training

2012	University of Zürich, Institute of Molecular Life Sciences, Zürich, Switzerland, Research project within participating in BUSS 2012: "Natural variation in RAS signaling in *C. elegans*"

Teaching experience

2019	Dunja Kureljak, Ruđer Bošković Institute, Zagreb, Croatia, Laboratory skill training
2017	Eva Pavlinek, Ruđer Bošković Institute, Zagreb, Croatia, Laboratory skill training
2014 - 2015	ETH Zürich, Teaching Assistant in Biocompatible materials I ETH Zürich, Teaching Assistant in Principles in Tissue Engineering ETH Zürich, Teaching Assistant in Principles in Biomedical Interfaces

Publications

1. Volarić M, Veseljak D, Mravinac B, Meštrović N, **Despot-Slade E** (2021) Genes 12 (8), 1114 https://doi.org/10.3390/genes12081114 [Q2, **IF=4,1** (2020)/**4,3** (5Y), 0 citation] **corresponding author**
2. **Despot-Slade E**, Mravinac B, Širca S, Castagnone-Sereno P, Plohl M, Meštrović N (2021) Mol. Biol. Evol. 38(5):1943-1965 https://doi.org/10.1093/molbev/msaa336 **[Q1, IF=16,2** (2020)**/18,7** (5Y), 0 citation**] first author**
3. Gržan T, **Despot-Slade E**, Meštrović N, Plohl M, Mravinac B (2020) PLoS Genetics 16(10):e1009115 https://doi.org/10.1371/journal.pgen.1009115 **[Q1, IF=5.9** (2020)**/6,3** (5Y), 1 citation**] first author**
4. Cvitanic MT, Zeljko TV, Pasantes JJ, Garcia-Souto D, Grzan T, **Despot-Slade E**, Plohl M, Satovic, E (2020) Genes 11 (6), 695 https://doi.org/10.3390/genes11060695 [Q2, **IF=4,1** (2020)/**4,3** (5Y), 1 citation] **coauthor**

5. Öztürk E, **Despot-Slade E**, Pichler M, Zenobi-Wong M (2017) Exp. Cell Res. 360(2):113-124 https://doi.org/10.1016/j.yexcr.2017.08.033 [Q3, **IF=3,9** (2020)/**4,1** (5Y), 8 citations] **coauthor**
6. Öztürk E, Hobiger S, **Despot-Slade E**, Pichler M, Zenobi-Wong M (2017) Sci. Rep. 7(1):9032 https://doi.org/10.1038/s41598-017-09505-6 [Q1, **IF=4,4** (2020)/**5,1** (5Y), 10 citations] **coauthor**

Conference abstracts

1. Volarić M, **Despot-Slade E**, Meštrović N, Long-range organisation of holocentromeres. PhD Student Symposium 2021, Zagreb, Croatia. Faculty of Science, University of Zagreb, 2021. pp. 267-268 (poster)
2. **Despot-Slade E**, Mravinac B, Širca S, Castagnone-Sereno P, Plohl M, Meštrović N, Epigenomics of holocentromere. PhD Student Symposium 2021, Zagreb, Croatia. Faculty of Science, University of Zagreb, 2021. pp. 185-186 (poster)
3. Gržan T, **Despot-Slade E**, Meštrović N, Plohl M, Mravinac B, Centromere epigenomics of the three closely related *Tribolium* species. HDBMB 2019 Crossroads in Life Sciences, Zagreb, Croatia. Croatian Society for Biochemistry and Molecular Biology, 2019. pp. 86-86 (poster)
4. **Despot-Slade E**, Pavlek M, Mravinac B, Castagnone-Sereno P, Plohl M, Meštrović N, Centromere of the nematode Meloidogyne; characterization of centromere specific H3 protein (CENH3) and underlying DNA sequences. Molecular Biophysics: ABC of the puzzle of Life, Zagreb, Croatia. Ruđer Bošković Institute and Croatian Biophysical Society, 2019. pp. 82-82 (poster)
5. **Despot-Slade E**, Pavlek M, Mravinac B, Castagnone-Sereno P, Plohl M, Meštrović N, Root-knot nematodesas a model for centromere (epi)genomics and satellitome research. 4th CONGRESS OF CROATIAN GENETICISTS with international participation, Zagreb, Croatia. Croatian Genetic Society, 2018. pp. 35-35 (lecture - Meštrović N)
6. **Despot-Slade E**, Širca S, Plohl M, Meštrović N, Comparative satellitome analysis reveals the polyploid hybrid origin in holocentric nematodes. 4th CONGRESS OF CROATIAN GENETICISTS with international participation, Zagreb, Croatia. Croatian Genetic Society, 2018. pp. 43-43 (poster)

7. Born AK, **Despot-Slade E**, Zenobi-Wong M, A collagen 2 reporter for online identification of chondrogenic microenvironments. eCM XV - Cartilage & Disc: Repair and Regeneration, Davos, Switzerland, 2014. pp. 61-61 (poster)
8. Schmid T, Snoek LB; Rodriguez M, van der Bent ML, **Despot-Slade E**, Kammenga JE, Hajnal A, Natural variation in Monoamine Oxidase A modulates RAS/MAPK pathway activity during C.elegans vulval development. The 5th EMBO meeting abstract book, Amsterdam, Netherlands, 2013. pp. 49-49 (poster)

Awards and participations

2012	Recognition of outstanding student status by Faculty of Science, Zagreb, Croatia
2012	Biology Undergraduate Summer School – BUSS 2012, University of Zürich, Switzerland
2011	Participation in the 2011 Biology Night, University of Zagreb, Croatia

Workshops and educations

2021	Training for safe use and maintenance of devices and data processing CFX96 qPCR system and Maestro software package (Biorad), Labena, Zagreb, Croatia
2019	White text on black screen - A brief introduction to command-line interface on your favourite Linux/Unix OS, IT Support Department, Ruđer Bošković Institute, Zagreb, Croatia
2018	Training programme on commercialization of RDI results: collaborating and competing on the European stage: embracing EU standards and norms for technology transfer to raise an international profile, Ruđer Bošković Institute and Knowledge Network, Zagreb, Croatia
2018	Zebrafish (*Danio rerio*) as a model organism in biological, biomedical and environmental research, Laboratory for

	Molecular Ecotoxicology, Ruđer Bošković Institute, Zagreb, Croatia
2017	6th Workshop on the Application of Next Generation Sequencing to Repetitive DNA Analysis in Plants, Institute of Plant Molecular Biology, České Budějovice, Czech Republic

Languages

Croatian	native language
English	speak fluently and read/write with high proficiency (Cambridge ESOL Council of Europe – Level C1 diploma)
German	speak, read and write with medium competence