Concomitant duplication of the *Cid* and *Cenp-C* genes in *Drosophila*

José R. Teixeira[1], Guilherme B. Dias[1], Marta Svartman[1], Alfredo Ruiz[2], Gustavo C. S. Kuhn[1]

[1] Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, Postal Code: 31270-901

[2] Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain, Postal Code: 08193

Corresponding author:

Prof. Dr. Gustavo C. S. Kuhn

Laboratório de Citogenômica Evolutiva

Departamento de Biologia Geral

Instituto de Ciências Biológicas

Universidade Federal de Minas Gerais

Av. Antônio Carlos 6627 – Pampulha

Postal Code: 31270-901

Belo Horizonte, MG

Brazil

Fone: +55 (31) 3409-3062

E-mail: gcskuhn@ufmg.br

Abstract

The constitutive centromeric proteins CenH3 and Cenp-C are interdependent in their role of establishing centromere identity and function. In a recent paper, Kursel and Malik (February 2017; doi: 10.1093/molbev/msx091) reported that the *Drosophila* CenH3 homologue *Cid* underwent four independent duplication events during evolution. Particularly interesting is the duplication that took place in the common ancestor of the *Drosophila* subgenus and led to the subfunctionalization and high divergence of the *Cid1* and *Cid5* paralogs. Here, we describe another independent *Cid* duplication (*Cid1* leading to *Cid6*) in the *buzzatii* cluster (*repleta* group) of the *Drosophila* subgenus. Moreover, we found that, in addition to the *Cid1*/*Cid5* duplication, *Cenp-C* was also duplicated (*Cenp-C1*, *Cenp-C2*) in the common ancestor of the *Drosophila* subgenus. Analyses of expression and tests for positive selection indicate that both *Cid5* and *Cenp-C2* are male germline-biased and evolved adaptively, indicating subfunctionalization of the *Cid* and *Cenp-C* paralogs. Our findings further highlight the strong interdependence between CenH3 and Cenp-C, paving the way to new perspectives by which centromere function and evolution can be addressed.

Keywords: CenH3, Cenp-C, gene duplication; centromere, Drosophila

Centromeres are epigenetically defined by the presence of the centromeric histone H3 variant CenH3, which creates a unique chromatin structure that is linked to outer kinetochore proteins by Cenp-C (reviewed by De Rop et al. 2012). In fact, CenH3 and Cenp-C are interdependent in their role of establishing centromere identity and function (Erhardt et al. 2008). Interestingly, this is illustrated by the fact that both *CenH3* and *Cenp-C* were lost independently in at least four lineages of insects (Drinnenberg et al. 2014). On the other hand, there is no record yet, neither in animals nor in plants, of concomitant duplications of both *CenH3* and *Cenp-C*.

In a paper published last February, Kursel and Malik (2017) reported that the *Drosophila* CenH3 homologue *Cid* underwent four independent duplication events during *Drosophila* evolution. Duplicate *Cid* genes exist in *D. eugracilis* (*Cid1*, *Cid2*) and in the *montium* subgroup (*Cid1*, *Cid3*, *Cid4*), both within the *Sophophora* subgenus, and in the entire *Drosophila* subgenus (*Cid1*, *Cid*5). Surprisingly, *Drosophila* species with a single *Cid* gene are the minority, as over one thousand *Drosophila* species encode two or more *Cid* genes (Kursel and Malik 2017).

In all analyzed species from the *Drosophila* subgenus, *Cid1* and *Cid5* are flanked by the *cbc* and *bbc* genes and the *Kr* and *CG6907* genes, respectively (Kursel and Malik 2017). However, while looking for the orthologs of *Cid1* and *Cid5* in the assembled genomes of two other species from the *Drosophila* subgenus, *D. buzzatii* and *D. seriema* (*repleta* group), we found a *Cid1* homolog, which we called *Cid6*, flanked by the *CG14341* and *IntS14* genes. Fluorescent *in situ* hybridizations on polytene chromosomes using *Cid6* probes showed one distal signal (in relation to the chromocenter) in the Muller element B of *D. buzzatii* and *D. seriema*, whereas *Cid1* probes showed one proximal signal in the Muller element C of the closely related *D. mojavensis* and in the outgroup *D. virilis* (fig. 1, upper panel). By investigating the *Cid1* locus of *D. buzzatii*, we found one 116-bp fragment of the original gene surrounded by a myriad of transposable elements (TEs; fig. 1, lower panel). We concluded that *Cid1* was degenerated by several TE insertions after the origin of *Cid6* by an inter-chromosomal duplication of *Cid1* in the lineage that gave rise to *D. buzzatii* and *D. seriema*. The *D. buzzatii* and *D. seriema* species belong to the monophyletic *D. buzzatii* cluster (Manfrin and Sene

2006) and the time of divergence between them has been estimated as ~4.6 mya (Oliveira et al. 2012). The divergence between the *D. buzzatii* and the closely related *D. mojavensis* clusters has been estimated at ~11.3 mya (Oliveira et al. 2012). Therefore, we can infer that the *Cid1* duplication happened between ~4.6 and 11.3 mya.

Why *Cid6* persevered while *Cid1* degenerated? The *D. buzzatii Cid1* locus is located in the most proximal region of the Muller element C (scaffold 115; Guillén et al. 2015). This region is very close to the pericentromeric heterochromatin, where TEs are highly abundant (Pimpinelli et al. 1995; Casals et al. 2005; Casals et al. 2006). Pericentromeric and adjacent regions are known to have low rates of crossing-over (Comeron et al. 2012; Nambiar and Smith 2016), which makes negative selection less effective in these regions (Zhang and Kishino 2004; Clément et al. 2006). Thus, it is reasonable to suggest that the presence of *Cid6* in the Muller element B alleviated the selective pressures over *Cid1* in the Muller element C, whose proximity to the pericentromeric heterochromatin fostered its degradation by several posterior TE insertions.

Given the interdependence between CenH3 and Cenp-C, we wondered if *Cenp-C* was also duplicated in species of the lineages in which *Cid* was duplicated. *D. eugracilis*, the *montium* subgroup and all the other species of the *Sophophora* subgenus have only one copy of *Cenp-C*, which is flanked by the *5-HT2B* gene. Interestingly, the species of the *Drosophila* subgenus have two copies of *Cenp-C*, which we called *Cenp-C1* and *Cenp-C2*: the former is flanked by the *5-HT2B* and *CG1427* genes, and the latter is flanked by the *CLS* and *RpL27* genes. We did not find additional *Cenp*-C duplicates in the *buzzatii* cluster. A maximum likelihood tree shows that *Cenp-C* was duplicated after the split between the *Sophophora* and *Drosophila* subgenera but before the split between *Zaprionus indianus* and the other species of the *Drosophila* subgenus (fig. 2A). Thus, we conclude that *Cenp-C2* originated from a duplication of *Cenp-C1* in the common ancestor of the *Drosophila* subgenus, at least 50 mya (Russo et al. 2013).

Both Cenp-C1 and Cenp-C2 contain all of the five major conserved Cenp-C domains (fig. 2B; Heeger et al. 2005): arginine-rich (R-rich), drosophilids Cenp-C homology (DH), nuclear localization

signal (NLS), CenH3 binding motif (also known as the Cenp-C motif), and C-terminal dimerization (Cupin). The only exception is *D. grimshawi* Cenp-C2, which lacks the Cupin domain. Interestingly, the two Cenp-C paralogs share ~65% identity at the amino acid level, with most of the divergence concentrated in inter-domain sequences. The conservation of these domains indicates that the two paralogs did not undergo neofunctionalization.

Kursel and Malik (2017) showed that *Cid5* expression is male germline-biased and proposed that *Cid1* and *Cid5* subfunctionalized and now perform nonredundant centromeric roles. In order to investigate if *Cenp-C1* and *Cenp-C2* are differentially expressed and correlated in some way with the expression of the *Cid* paralogs, we analyzed the available transcriptomes from embryos, larvae, pupae and adult females and males of *D. buzzatii* (described in Guillén et al. 2015), and from testes of *D. virilis* and *D. americana* (BioProject Accession PRJNA376405).

While *Cid6* is transcribed in all stages of development of *D. buzzatii*, *Cid5* transcription is limited to pupae and adult males and is higher than *Cid6* transcription in the latter (fig. 3A). Also, *Cid5* transcription is elevated in testes of *D. virilis* and *D. americana*, whereas *Cid1* is virtually silent (fig. 3C). Our results further support the findings of Kursel and Malik (2017) that *Cid5* display a male germline-biased expression. In this context, our finding that *Cid5* is also transcribed in pupae of *D. buzzatii* may be related to the ongoing development of the male gonads.

In contrast to the *Cid* paralogs, we found that the *Cenp-C* paralogs are always transcribed. *Cenp-C2* transcription is higher than *Cenp-C1* transcription in pupae and adult males of *D. buzzatii* (fig. 3B) and in testes of *D. virilis* (fig. 3D). On the other hand, *Cenp-C1* transcription is higher than *Cenp-C*2 transcription in embryos and adult females of *D. buzzatii*. There is no significant difference between their expression in testes of *D. americana*. Similar to what was found for the *Cid* paralogs, the differential expression between the *Cenp-C* paralogs support the hypothesis of subfunctionalization. Given their male germline-biased expression, it is likely that *Cid5* and *Cenp-C2* are interdependent in male meiosis.

Centromeres are essential for the faithful segregation of chromosomes in cell divisions, yet the centromeric DNA and both CenH3 and Cenp-C are highly variable across species (Henikoff et al. 2000; Talbert et al. 2004; Plohl et al. 2008). This paradox may be explained by the centromere drive hypothesis, which states that CenH3 and Cenp-C constantly evolve and acquire new binding preferences to rapidly evolving centromeric DNAs in an effort to suppress their selfishly spread through the population by female meiotic drive (Henikoff et al. 2001; Dawe and Henikoff 2006).

The rapid evolution of CenH3 required for the "drive suppressor" function may be disadvantageous for canonical functions (e.g. mitosis; Finseth et al. 2015; Kursel and Malik 2017). Extending this reasoning, selection may act differently in each of the *Cid* and *Cenp-C* paralogs. Thus, considering the possibility of optimization for divergent functions, we performed tests for positive selection on full-length alignments of the *Cid* and *Cenp-C* paralogs using maximum likelihood methods. As alignments with either few informative sites or too many gaps can generate insufficient data, we focused our analyses on five closely related cactophilic *Drosophila* species from the *repleta* group (*D. mojavensis*, *D. arizonae*, *D. navojoa*, *D. buzzatii* and *D. seriema*).

Consistent with the hypothesis of their interdependence, we found that both *Cid5* and *Cenp-C2* evolved adaptively (table 1). Bayes Empirical Bayes analyses identified with a posterior probability > 95% four amino acids in the N-terminal tail of Cid5 and six amino acids throughout Cenp-C2 as having evolved under positive selection. Of the six Cenp-C amino acids, one is in the DH domain, one is in the Cupin domain, and the remaining four are in inter-domain sequences.

Molecular genetic data alone cannot reveal the underlying cause of adaptive evolution. Kursel and Malik (2017) found signs of positive selection in the male germline-biased *Cid3* paralog of the *montium* subgroup and proposed that *Cid3* and *Cid5* are candidate suppressors of centromere drive given their male germline-biased expression. Our results of positive selection on both *Cid5* and *Cenp-C2* do support this hypothesis. However, there is still a need to clarify how CenH3 and Cenp-C would suppress centromere drive in male-meiosis, given that the proposed models state that suppression occurs in female meiosis (Henikoff et al. 2001; Dawe and Henikoff 2006). Additionally, male

germline-biased genes are widely known to evolve adaptively as the result of male-male or male-female competition (Ellegren and Parsch 2007; Meisel 2011). Finally, if the adaptive evolution of CenH3 and Cenp-C modulates their binding to centromeric DNAs, how come that the diverged Cid1/Cenp-C1 and Cid5/Cenp-C2 bind in different contexts (i.e. mitosis vs. meiosis) to the same set of centromeric DNAs? All things considered, we present the possibility that adaptive evolution of CenH3 and Cenp-C is somehow linked to alterations in centromeric chromatin structure.

A number of studies have shown that both CenH3 and Cenp-C not only are essential for kinetochore assembly but also coordinate the dynamics of centromeric chromatin. Although the specific function of the N-terminal tail is unknown in *Drosophila* Cid, studies in humans, fission yeast and Arabidopsis have shown that the N-terminal tail is important for recruitment and stabilization of inner kinetochore proteins, centromeric chromatin conformation and proper chromosome segregation (Bailey et al. 2013; Fachinetti et al. 2013; Folco et al. 2015; Logsdon et al. 2015; Maheshwari et al. 2015). Additionally, Cenp-C affects CenH3 nucleosome structure and dynamics (Falk et al. 2015; Falk et al. 2016), as well as meiotic Cid deposition and centromere clustering (Unhavaithaya and Orr-Weaver 2013; Kwenda et al. 2016). The specific function of the R-rich and DH domains, and the possible functions of inter-domain sequences of *Drosophila* Cenp-C are unknown, but the Cupin domain, present in all metazoans (Heeger et al. 2005), has been implicated in Cenp-C dimerization (Sugimoto et al. 1997).

Of all the described *Cid* paralogs, *Cid1* and *Cid5* are the most divergent: their N-terminal tails only share ~15% of identity, represented by the conservation of only 1-2 of the four core Cid motifs (Kursel and Malik 2017). If the N-terminal tail of Cid interacts with Cenp-C, it is possible that the duplication of *Cenp-C* allowed the higher divergence between the *Cid1* and *Cid5* paralogs. In this context, both Cid5 and Cenp-C2 could have specialized in creating a centromeric chromatin structure that is better suited for male meiosis requirements. As interfering with centromeric proteins that are specialized in meiosis would avoid disruption of essential mitotic functions (Kursel and Malik 2017),

129 functional studies on Cid5 and Cenp-C2 have the potential to elucidate the dynamics of both CenH3

130 and Cenp-C evolution.

Materials and Methods

Identification of *Cid* and *Cenp-C* orthologs and paralogs in sequenced genomes

Drosophila Cid and *Cenp-C* genes were identified by tBLASTx in sequenced genomes using the *D. melanogaster Cid1* and *Cenp-C1* as queries (FlyBase IDs FBgn0040477 and FBgn0266916, respectively). Since *Cid* is encoded by a single exon in *Drosophila*, we selected the entire open reading frame for each *Cid* gene hit, and since *Cenp-C* have multiple introns, we used the Augustus gene prediction algorithm (Stanke and Morgenstern 2005) to identify the coding sequences. For annotated genomes, we recorded the 5' and 3' flanking genes for the *Cid* and *Cenp-C* genes of each species. For genomes that were not annotated, we used the 5' and 3' nucleotide sequences flanking the *Cid* and *Cenp-C* genes as queries to the *D. melanogaster* genome using BLASTn and verified the synteny in accordance to the hits. All *Cid* and *Cenp-C* coding sequences and their database IDs can be found in Supplementary Files S1 and S2, respectively.

Fluorescent *in situ* hybridizations on polytene (FISH) chromosomes

Probes for *Cid1/Cid6* were obtained by PCR from genomic DNA of *D. buzzatii* (strain st-1), *D. seriema* (strain D73C3B), *D. mojavensis* (strain 14021-0248.25) and *D. virilis* (strain 15010-1551.51). We cloned the PCR products into the pGEM-T vector (Promega) and sequenced them to confirm identity. Recombinant plasmids were labeled with digoxigenin 11-dUTP by nick translation (Roche Applied Science). FISH on polytene chromosomes was performed as described in Dias et al. (2015). The slides were analyzed under an Axio Imager A2 epifluorescence microscope equipped with the AxioCam MRm camera (Zeiss). Images were captured with the AxioVision (Zeiss) software and edited in Adobe Photoshop.

Phylogenetic analyses

Cid and *Cenp-C* sequences were aligned at the codon level using MUSCLE (Edgar 2004) and refined manually. Using the five major conserved domains of Cenp-C (Heeger et al. 2005), we generated

maximum likelihood phylogenetic trees in MEGA6 (Tamura et al. 2013) with the GTR substitution model and 1,000 bootstrap replicates for statistical support.

Expression analyses

RNA-seq data from *D. buzzatii* (described in Guillén et al. 2015), and from *D. virilis* and *D. americana* (BioProject Accession PRJNA376405) were aligned against the *Cid* and *Cenp-C* coding sequences from each species with Bowtie2 (Langmead and Salzberg 2012), as implemented to the Galaxy server (Afgan et al. 2016). Mapped reads were normalized by the transcripts per million (TPM) method (Wagner et al. 2012), and all normalized values < 1 were set to 1 so that log 2 TPM ≥ 0.

Positive selection analyses

Cid and *Cenp-C* alignments and gene trees were used as input into the CodeML NSsites models of PAMLX version 1.3.1 (Xu and Yang 2013). To determine whether each paralog evolves under positive selection, we compared three models that do not allow dN/dS to exceed 1 (M1a, M7 and M8a) to two models that allow dN/dS > 1 (M2a and M8). Positively selected sites were classified as those with a M8 Bayes Empirical Bayes posterior probability > 95%.

Acknowledgments

We are grateful to Dr. Maura Helena Manfrin (Univesity of São Paulo) for providing us with the *D. seriema* strain used in the present work. This work was supported by a grant from "Fundação de Amparo à Pesquisa do Estado de Minas Gerais" (FAPEMIG) to G.K. (grant number APQ-01563-14).

References

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44:W3-W10.

Bailey A, Panchenko T, Sathyan K, Petkowski J, Pai P, Bai D, Russell D, Macara I, Shabanowitz J, Hunt D, et al. 2013. Posttranslational modification of CENP-A influences the conformation of centromeric chromatin. Proc Natl Acad Sci USA. 110:11827-11832.

Casals F, Cáceres M, Manfrin M, González J, Ruiz A. 2005. Molecular Characterization and Chromosomal Distribution of Galileo, Kepler and Newton, Three Foldback Transposable Elements of the Drosophila buzzatii Species Complex. Genetics. 169:2047-2059.

Casals F, González J, Ruiz A. 2006. Abundance and chromosomal distribution of six Drosophila buzzatii transposons: BuT1, BuT2, BuT3, BuT4, BuT5, and BuT6. Chromosoma. 115:403-412.

Clément Y, Tavares R, Marais G. 2006. Does lack of recombination enhance asymmetric evolution among duplicate genes? Insights from the Drosophila melanogaster genome. Gene. 385:89-95.

Comeron J, Ratnappan R, Bailin S. 2012. The Many Landscapes of Recombination in Drosophila melanogaster. PLoS Genet. 8:e1002905.

Dawe R, Henikoff S. 2006. Centromeres put epigenetics in the driver's seat. Trends Biochem Sci. 31:662-669.

De Rop V, Padeganeh A, Maddox P. 2012. CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly. Chromosoma. 121:527-538.

Dias G, Heringer P, Svartman M, Kuhn G. 2015. Helitrons shaping the genomic architecture of Drosophila: enrichment of DINE-TR1 in α- and β-heterochromatin, satellite DNA emergence, and piRNA expression. Chromosome Res. 23:597-613.

Drinnenberg I, de Young D, Henikoff S, Malik H. 2014. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. eLife. 3:e03676.

Edgar R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1797.

Erhardt S, Mellone B, Betts C, Zhang W, Karpen G, Straight A. 2008. Genome-wide analysis reveals a cell cycle–dependent mechanism controlling centromere propagation. J Cell Biol. 183:805-818.

Fachinetti D, Diego Folco H, Nechemia-Arbely Y, Valente L, Nguyen K, Wong A, Zhu Q, Holland A, Desai A, Jansen L, et al. 2013. A two-step mechanism for epigenetic specification of centromere identity and function. Nat Cell Biol. 15:1056-1066.

Falk S, Guo L, Sekulic N, Smoak E, Mani T, Logsdon G, Gupta K, Jansen L, Van Duyne G, Vinogradov S, et al. 2015. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. Science. 348:699-703.

Falk S, Lee J, Sekulic N, Sennett M, Lee T, Black B. 2016. CENP-C directs a structural transition of CENP-A nucleosomes mainly through sliding of DNA gyres. Nat Struct Mol Biol. 23:204-208.

Finseth F, Dong Y, Saunders A, Fishman L. 2015. Duplication and Adaptive Evolution of a Key Centromeric Protein in Mimulus, a Genus with Female Meiotic Drive. Mol Biol Evol. 32:2694-2706.

Folco H, Campbell C, May K, Espinoza C, Oegema K, Hardwick K, Grewal S, Desai A. 2015. The CENP-A N-Tail Confers Epigenetic Stability to Centromeres via the CENP-T Branch of the CCAN in Fission Yeast. Curr Biol. 25:348-356.

Gonzalez J, Nefedov M, Bosdet I, Casals F, Calvete O, Delprat A, Shin H, Chiu R, Mathewson C, Wye N, et al. 2005. A BAC-based physical map of the Drosophila buzzatii genome. Genome Res. 15:885-889.

Guillén Y, Rius N, Delprat A, Williford A, Muyas F, Puig M, Casillas S, Ràmia M, Egea R, Negre B, et al. 2014. Genomics of Ecological Adaptation in Cactophilic Drosophila. Genome Biol Evol. 7:349-366.

Heeger S, Leismann O, Schittenhelm R, Schraidt O, Heidmann S, Lehner C. 2005. Genetic interactions of separase regulatory subunits reveal the diverged Drosophila Cenp-C homolog. Genes Dev. 19:2041-2053.

Henikoff S, Ahmad K, Malik H. 2001. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. Science. 293:1098-1102.

Henikoff S, Ahmad K, Platero J, van Steensel B. 2017. Heterochromatic deposition of centromeric histone H3-like proteins. Proc Natl Acad Sci USA. 97:716-721.

Kohany O, Gentles A, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. 7:474.

Kuhn G, Ruiz A, Alves M, Sene F. 1996. The metaphase and polytene chromosomes of Drosophila seriema (repleta group; mulleri subgroup). Brazillian Journal of Genetics. 19:209–216.

Kursel L, Malik H. 2017. Recurrent Gene Duplication Leads to Diverse Repertoires of Centromeric Histones in Drosophila Species. Mol Biol Evol.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357-359.

Logsdon G, Barrey E, Bassett E, De Nizio J, Guo L, Panchenko T, Dawicki-McKenna J, Heun P, Black B. 2015. Both tails and the centromere targeting domain of CENP-A are required for centromere establishment. J Cell Biol. 208:521-531.

Maheshwari S, Tan E, West A, Franklin F, Comai L, Chan S. 2015. Naturally Occurring Differences in CENH3 Affect Chromosome Segregation in Zygotic Mitosis of Hybrids. PLoS Genetics. 11:e1004970.

Manfrin M, Sene F. 2006. Cactophilic Drosophila in South America: A Model for Evolutionary Studies. Genetica. 126:57-75.

Nambiar M, Smith G. 2016. Repression of harmful meiotic recombination in centromeric regions. Semin Cell Dev Biol. 54:188-197.

Oliveira D, Almeida F, O'Grady P, Armella M, DeSalle R, Etges W. 2012. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the Drosophila repleta species group. Mol Phylogenet Evol. 64:533-544.

Pimpinelli S, Berloco M, Fanti L, Dimitri P, Bonaccorsi s, Marchetti E, Caizzi R, Caggese C, Gatti M. 1995. Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. Proc Natl Acad Sci USA. 92:3804-3808.

Plohl M, Luchetti A, Meštrović N, Mantovani B. 2008. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene. 409:72-82.

Schaeffer S, Bhutkar A, McAllister B, Matsuda M, Matzkin L, O'Grady P, Rohde C, Valente V, Aguade M, Anderson W, et al. 2008. Polytene Chromosomal Maps of 11 Drosophila Species: The Order of Genomic Scaffolds Inferred From Genetic and Physical Maps. Genetics. 179:1601-1655.

Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33:W465-W467.

Sugimoto K, Kuriyama K, Shibata A, Himeno M. 2017. Characterization of internal DNA-binding and C-terminal dimerization domains of human centromere/kinetochore autoantigen CENP-C in vitro: role of DNA-binding and self-associating activities in kinetochore organization. Chromosome Res. 5:132-141.

Talbert P, Bryson T, Henikoff S. 2004. Adaptive evolution of centromere proteins in plants and animals. J Biol. 3:18.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 30:2725-2729.

Wagner G, Kin K, Lynch V. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 131:281-285.

Zhang Z, Kishino H. 2004. Genomic Background Predicts the Fate of Duplicated Genes: Evidence From the Yeast Genome. Genetics. 166:1995-1999.

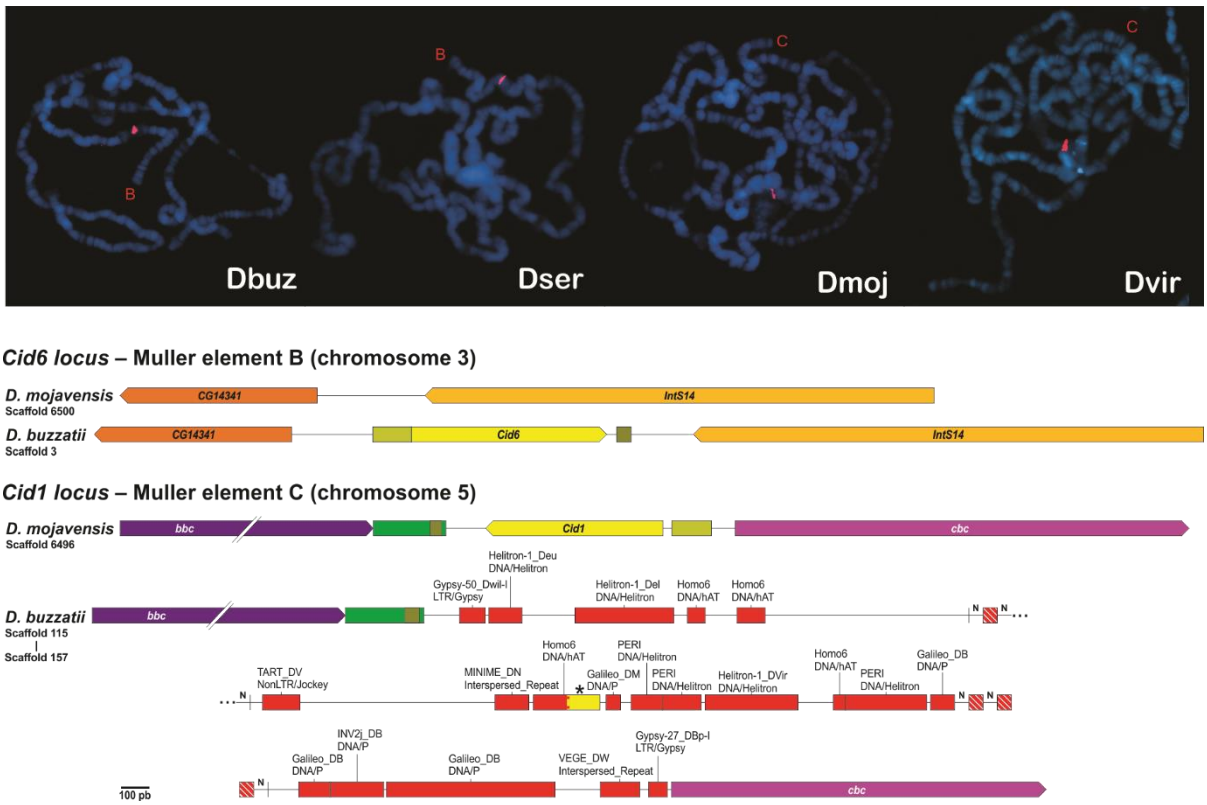

Figure 1. Inter-chromosomal duplication of *Cid1*. (Upper panel) Fluorescent in situ hybridizations on polytene chromosomes with *Cid* probes (red signal). In *D. buzzatii* (Dbuz) and *D. seriema* (Dser), the *Cid1* homolog *Cid6* is located in a distal region of the Muller element B. On the other hand, in the closely related *D. mojavensis* (Dmoj) and in the outgroup *D. virilis* (Dvir), *Cid1* is located in a proximal region of the Muller element C. Chromosome arms were identified by their morphology (Kuhn et al. 1996; González et al. 2005; Schaeffer et al. 2008) and are indicated as Muller elements. **(Lower panel)** Comparison between the *Cid1* and *Cid6* loci of *D. buzzatii* and the corresponding regions of *D. mojavensis*. The asterisk indicates the 116-bp fragment of the original gene, 'N' indicates unidentified nucleotides, and red boxes indicate transposable elements identified by the RebBase Censor tool (Kohany et al. 2006).

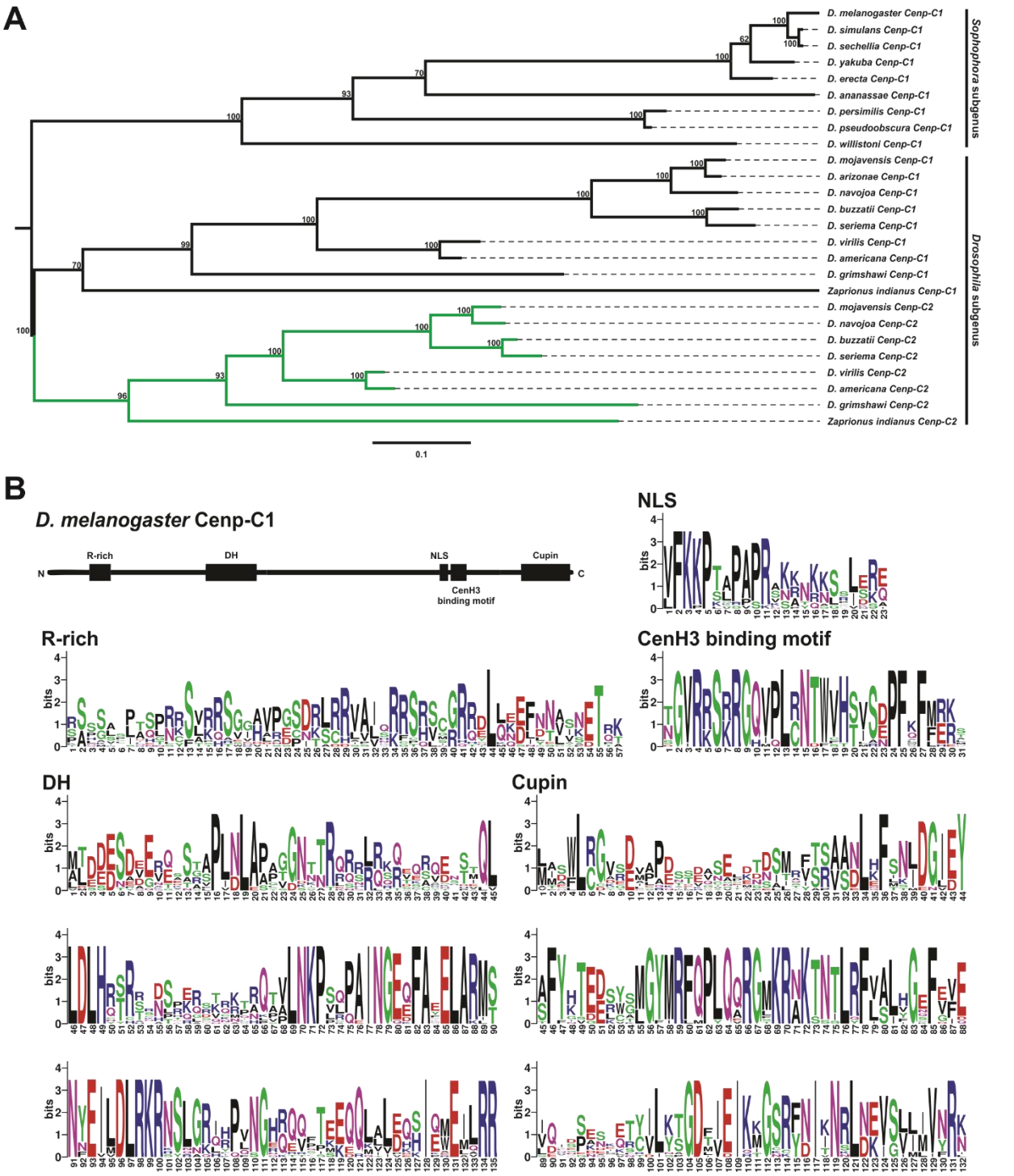

Figure 2. *Cenp-C1* was duplicated in the common ancestor of the *Drosophila* subgenus. (A) Maximum likelihood tree of the two *Cenp-C* paralogs. Bootstrap values are shown in each node. Scale bar represents number of substitutions per site. **(B)** Schematic representation of the domain structure of *D. melanogaster* Cenp-C1 and logo representation for the *Drosophila* subgenus Cenp-C1 and Cenp-C2 consensus of each domain. Domains are as follow: R-rich, arginine-rich; DH, drosophilid Cenp-C homology; NLS, nuclear localization signal; CenH3 binding motif, also known as Cenp-C motif; Cupin, a dimerization domain near the C-terminal region.

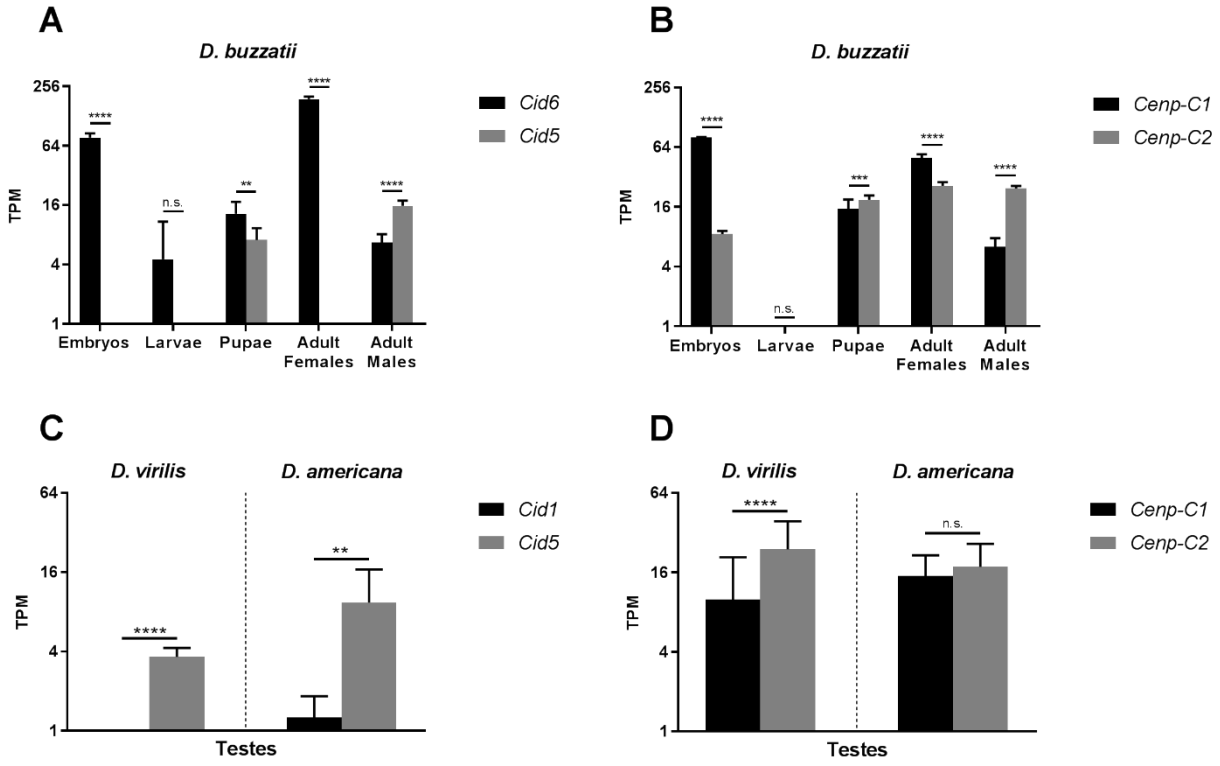

Figure 3. *Cid* and *Cenp-C* expression patterns in *D. buzzatii* (A and B) and in *D. virilis* and *D. americana* (C and D). *Cid* and *Cenp-C* transcripts are expected to be found in abundance only in mitotic active tissues. Therefore, *Cid* and *Cenp-C* transcripts are abundant in embryos and adult females likely due to oogenesis and oocyte mRNA accumulation required to sustain the high mitotic rate in early embryonic development, in adult males likely due to spermatogenesis, and in pupae likely due to morphogenesis. As most of the larval mitotic activity is limited to neuroblasts, a low abundance of *Cid* and *Cenp-C* transcripts can be expected. Data are presented as mean ± confidence interval (95%) and analyzed by one-way ANOVA (A and B) and Student's t-test (C and D): *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$; ****$P \leq 0.0001$; n.s., not significant. TPM, transcripts per million.

Table 1. Summary of tests for positive selection performed on each *Cid* and *Cenp-C* paralog.

	Alignment length (#nts)	M1a vs. M2a	M7 vs. M8	M8a vs. M8
Cid1/Cid6	609	$P = 1$	$P = 1$	$P = 0.982$
Cid5	600	$P = 0.099$	$P = 0.069$	$P = 0.025$*
Cenp-C1	3,492	$P = 0.496$	$P = 0.163$	$P = 0.210$
Cenp-C2	3,696	$P = 0.194$	$P = 0.005$*	$P = 0.068$