OBSERVATION ARTICLE

Possible identification of CENP-C in fish and the presence of the CENP-C motif in M18BP1 of vertebrates. [version 1; peer review: 1 approved with reservations]

Leos G. Kral

Department of Biology, University of West Georgia, Carrollton, GA, 30118, USA

Abstract
The centromeric protein CENP-C is a base component of the kinetochore. This protein, along with CENP-A has been shown to adaptively evolve in a number of animal and plant species. In order to determine if CENP-C also evolves in fish species, I attempted to retrieve fish CENP-C sequences from GenBank. No Teleostei CENP-C sequences were found either by name or by BLASTP searches with the vertebrate CENP-C motif sequence. A number of putative Teleostei protein sequences were identified in GenBank that have homology to the C-terminal cupin domain of vertebrate CENP-C. These proteins only have partial homology to the CENP-C motif, but evidence is presented that makes it likely that these fish proteins are orthologs of CENP-C. Interestingly, it was also discovered that the CENP-C motif sequence is also mostly present in M18BP1 proteins of fish and some other vertebrates but not in mammals. This finding may have implications for CENP-C and M18BP1 assembly in centromeric regions of different vertebrate taxa.

Keywords
CENP-C, M18BP1, centromeric proteins, teleostei, kinetochore, CENP-C motif, cupin domain protein

Open Peer Review

Approval Status

| 1 | 2 | 3 |
|---|---|---|
| view | view | view |

version 2
(revision)
20 Jan 2016

version 1
05 Aug 2015

1. Tatsuo Fukagawa, Osaka University, Osaka, Japan
2. Nikolina Sekulic, University of Oslo, Oslo, Norway
   University of Oslo, Oslo, Norway
3. Inna Lermontova, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

Any reports and responses or comments on the article can be found at the end of the article.
Corresponding author: Leos G. Kral (lkral@westga.edu)

Competing interests: No competing interests were disclosed.

Grant information: This study was supported by a Faculty Research Grant funded by the University of West Georgia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2015 Kral LG. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

How to cite this article: Kral LG. Possible identification of CENP-C in fish and the presence of the CENP-C motif in M18BP1 of vertebrates. [version 1; peer review: 1 approved with reservations] F1000Research 2015, 4:474
https://doi.org/10.12688/f1000research.6823.1

First published: 05 Aug 2015, 4:474 https://doi.org/10.12688/f1000research.6823.1
Introduction

The kinetochore is a structure that connects chromosomal centromeric DNA to microtubules during mitosis and meiosis\(^1\). The centromere is epigenetically defined by the deposition of nucleosomes that contain the histone H3 variant CENP-A\(^2\). Centromeric protein CENP-C is required for both the recruitment of new CENP-A to the centromeric region as well as the initial assembly of the kinetochore. The CENP-C protein is generally considered to be ubiquitous in all eukaryotic taxa since homologs of CENP-C have been identified in yeast\(^3\) and Drosophila\(^4\) as well as many plants and vertebrates\(^5\). While CENP-C evolves so rapidly that very little homology is observed between distantly related taxa, a conserved CENP-C motif has been identified across all lineages studied\(^6\). This conserved motif should, therefore, be of utility to identify CENP-C orthologs in other species.

CENP-A has been initially shown to evolve adaptively in Drosophila\(^7\), in members of the Bressicaceae family\(^8\) and more recently in primates\(^9\) and in percid fishes\(^10\). CENP-C has also been shown to evolve adaptively in a number of animal and plant species\(^1\) as well as in primates\(^9\). In an effort to determine if CENP-C also evolves adaptively in fish species, searches were conducted in GenBank for Teleostei proteins that had been already identified as CENP-C or for genes that had been annotated as coding for CENP-C. No such fish proteins or genes were found. BLASTP searches of just the Teleostei subset of GenBank were performed with the conserved vertebrate CENP-C motif and these too failed to find identified fish CENP-C proteins or genes. However, these searches did identify fish M18BP1, which, as will be discussed below, contains a sequence homologous to the CENP-C motif. A search of the Chondrichthyes for proteins that had been already identified as CENP-C, or for genes that had been annotated as coding for CENP-C, found one gene annotated as CENP-C in the elephant shark (Callorhinchus milii) genome. Interestingly, while this putative shark CENP-C protein contained a cupin domain at the C-terminal end homologous to the cupin domain found at the C-terminal end of other vertebrate CENP-C proteins, the conserved CENP-C motif was not found in the expected location upstream of the cupin domain. However, a region further upstream does have homology to some of the most conserved amino acids of the CENP-C motif.

BLASTP searches of the Teleostei subset of GenBank with the putative shark CENP-C protein sequence identified a number of genes that were primarily homologous to the C-terminal cupin domain. However, none of these fish sequences were annotated as CENP-C in GenBank. Upon closer analysis, as will be discussed below, these C-terminal cupin domain containing fish proteins do contain sequences that are partly homologous to the conserved CENP-C motif and, therefore, these fish genes could be CENP-C orthologs.

Methods

Standard BLASTP searches were performed on the NCBI blast server. The vertebrate CENP-C motif NVRRTKRXRLKPLEYWRGERVBY used in BLASTP searches in this study was obtained from an alignment of 25 species including amphibians, reptiles, birds and mammals (Supplementary File S1). Retrieved sequences were aligned with the MUSCLE alignment feature in Geneious (version 6.1) sequence analysis software.

Results and discussion

BLASTP searches with the vertebrate CENP-C motif identified CENP-C proteins from a variety of taxa, including plants, but did not identify any CENP-C in fish lineages. It is possible that CENP-C may be absent in fish, but the ubiquity of this protein in other lineages and the central role of this protein in centromeric function make this unlikely. A C-terminal cupin domain protein encoded by a shark gene annotated in GenBank as CENP-C was used to identify homologs in Teleostei genomes by BLASTP. The retrieved teleost fish homologs were annotated as either calponin homology domain containing protein, neurofilament heavy polypeptide-like protein, or myb-like protein. Within vertebrate CENP-C proteins the RxxRXxxxPlxYWxGERxxY sequence defines identities within the CENP-C motif located within about 100 amino acids upstream of the cupin domain (Figure 1). However, within the shark and teleost fish C-terminal cupin domain-containing protein sequences, only some of these CENP-C motif sequence identities were present (Figure 2) and, therefore, unambiguous identity of these proteins as CENP-C was not obvious.

In a recent study that examined the interaction between CENP-C conserved domains and CENP-A containing nucleosomes (or

---

**Figure 1.** Conserved domains and sequences in vertebrate CENP-C. (A) Diagram of human CENP-C. (B) Amino acids that are identical in the CENPC motif in vertebrates in which CENP-C has been identified. (C) Conserved sequence in the CENP-C central region that is homologous to part of the CENP-C motif. Amino acid locations within human CENP-C protein of conserved sequences are indicated at the beginning and end of each sequence.

---

| Sequence | Description |
|----------|-------------|
| NVRRTKRXRLKPLEYWRGERVBY | Vertebrate CENP-C motif |
| RxxRXxxxPlxYWxGERxxY | CENP-C motif homologous sequence |
| RxxRXxxx | Central region sequence |

---

**Figure 2.** Diagram of human CENP-C. (A) Diagram of human CENP-C. (B) Amino acids that are identical in the CENPC motif in vertebrates in which CENP-C has been identified. (C) Conserved sequence in the CENP-C central region that is homologous to part of the CENP-C motif. Amino acid locations within human CENP-C protein of conserved sequences are indicated at the beginning and end of each sequence.
nucleosomes containing histone H3 modified with a CENP-A C-terminal tail), Kato et al.\textsuperscript{11} identified within the conserved central region of CENP-C a R\textsubscript{x}S\textsubscript{x}xPS\textsubscript{x}WW consensus sequence (Figure 1) that is similar to the core portion of the CENP-C motif. Mutations of the arginine to alanine or the tryptophans to alanine in this sequence prevented the binding of this central region to the nucleosomes. So, functionally, the R\textsubscript{x}xxP\textsubscript{x}xWW portion of the central region sequence is important to centromeric binding of CENP-C. Furthermore, mutations of the arginine, tyrosine and tryptophan in the core CENP-C motif R\textsubscript{x}xxP\textsubscript{x}xYW also reduce the binding affinity the CENP-C to the nucleosomes\textsuperscript{11}. A mutation of arginine to alanine in this core portion of the CENP-C motif was previously shown to prevent the binding of Xenopus CENP-C to centromeres\textsuperscript{25}.

An alignment of the putative shark and teleost fish CENP-C proteins identified two conserved regions that contained the R\textsubscript{x}xxP\textsubscript{x}xWW sequences (Figure 2). The placement of these sequences corresponds roughly to the locations of the central portion and the CENP-C motif of the vertebrate CENP-C (Figure 1). Therefore, it is likely that the combination of the C-terminal cupin domain and the presence of these centromeric nucleosome binding regions in positions generally corresponding to the locations of the central region and the CENP-C motif identifies these teleost genes as possible CENP-C orthologs. It will be necessary, of course, to verify if this protein is actually found at fish centromeres. It should be noted, however, that the distance between the cupin domain and the “CENP-C motif” position is about twice as long in the putative fish CENP-C in comparison to this distance in CENP-C of other vertebrates. It is interesting that the putative shark “CENP-C motif” location lacks the tryptophans of the R\textsubscript{x}xxP\textsubscript{x}xWW sequence and that Poecilia reticulata has a replacement of the first tryptophan in the conserved central region sequence (Figure 2). However, depending on other factors acting in the assembly of the centromere in various taxa, it may be possible that just one of those conserved R\textsubscript{x}xxP\textsubscript{x}xWW sequences may be necessary for centromeric binding of the putative fish CENP-C. Interestingly, no homology to the R\textsubscript{x}xxP\textsubscript{x}xWW portion of the conserved central region is detectable in CENP-C of reptiles and birds.

Interestingly, BLASTP searches of the Teleostei subset of GenBank retrieved centromeric protein M18BP1 sequences. This protein is recruited to centromeres by CENP-C\textsuperscript{13,15} and along with centromeric proteins Mis18\textsuperscript{α} and Mis18\textsuperscript{β} functions in the recruitment of CENP-A to centromeres\textsuperscript{16}. The M18BP1 protein contains almost the entire vertebrate CENP-C motif in all vertebrates examined except in mammals (Figure 3). It appears that the

---

**Figure 2.** Conserved regions homologous to a portion of the CENP-C motif present in putative fish CENP-C. (A) Diagram of putative Stegastes partitus CENP-C. (B) Alignment of fish central region sequences that contain the conserved R\textsubscript{x}xxP\textsubscript{x}xWW sequence. (C) Alignment of the fish cupin domain proximal sequences that contain the conserved R\textsubscript{x}xxP\textsubscript{x}xWW sequence. Amino acids matching the conserved sequence identities are highlighted in red. Amino acid locations within each species’ protein are indicated at the beginning and end of each sequence.
CENP-C motif sequence is not exclusive to just CENP-C. Since both CENP-C and M18BP1 associate with centromeres and with each other, it is tempting to speculate that what has generally been regarded as a CENP-C motif sequence facilitates the interaction of both of these proteins with centromeric nucleosomes. Furthermore, since mammalian M18BP1 lacks this CENP-C motif, it is possible that mammalian M18BP1 may be more dependent on association with CENP-C to localize to the centromere than the M18BP1 of other vertebrate taxa.

CENP-C motif sequence is not exclusive to just CENP-C. Since both CENP-C and M18BP1 associate with centromeres and with each other, it is tempting to speculate that what has generally been regarded as a CENP-C motif sequence facilitates the interaction of both of these proteins with centromeric nucleosomes. Furthermore, since mammalian M18BP1 lacks this CENP-C motif, it is possible that mammalian M18BP1 may be more dependent on association with CENP-C to localize to the centromere than the M18BP1 of other vertebrate taxa.

Supplementary material

Supplementary file S1. Alignment of the C-terminal portion of vertebrate CENP-C proteins.

The vertebrate CENP-C motif containing consensus sequence utilized in BLASTP searches spans amino acids 11 to 33 and is highlighted in red.

Click here to access the data.

Competing interests

No competing interests were disclosed.

Grant information

This study was supported by a Faculty Research Grant funded by the University of West Georgia.

I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
References

1. Cheeseman IM: The kinetochore. Cold Spring Harb Perspect Biol. 2014; 6(7): a015826. PubMed Abstract | Publisher Full Text

2. Black BE, Brock MA, Bédard S, et al.: An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. Proc Natl Acad Sci U S A. 2007; 104(12): 5008–5013. PubMed Abstract | Publisher Full Text | Free Full Text

3. Meluh PB, Koshland D: Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C. Mol Biol Cell. 1995; 6(7): 793–807. PubMed Abstract | Publisher Full Text | Free Full Text

4. Heeger S, Leismann O, Schittenhelm R, et al.: Genetic interactions of separase regulatory subunits reveal the diverged Drosophila Cenp-C homolog. Genes Dev. 2005; 19(17): 2041–2053. PubMed Abstract | Publisher Full Text | Free Full Text

5. Taft M, Fryet T, Henikoff S: Adaptive evolution of centromere proteins in plants and animals. J Biol. 2004; 3(4): 18. PubMed Abstract | Publisher Full Text | Free Full Text

6. Malik HS, Henikoff S: Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics. 2001; 157(3): 1293–1298. PubMed Abstract | Free Full Text

7. Cooper JL, Henikoff S: Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol. 2004; 21(9): 1712–1718. PubMed Abstract | Publisher Full Text

8. Schuler MG, Swanson W, Thomas PJ, et al.: Adaptive evolution of foundation kinetochore proteins in primates. Mol Biol Evol. 2010; 27(7): 1585–1597. PubMed Abstract | Publisher Full Text | Free Full Text

9. Fountain DM, Kral LG: Isolation and Characterization of the Etheostoma tallapoosae (Teleostei: Percidae) CENP-A Gene. Genes (Basel). 2011; 2(4): 829–840. PubMed Abstract | Publisher Full Text | Free Full Text

10. Abbey HN, Kral LG: Adaptive Evolution of CENP-A in Percid Fishes. Genes (Basel). 2015; 6(3): 662–671. PubMed Abstract | Publisher Full Text

11. Kato H, Jiang J, Zhou BR, et al.: A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. Science. 2013; 340(6136): 1110–1113. PubMed Abstract | Publisher Full Text | Free Full Text

12. Milks K, Moree B, Straight AF: Dissection of CENP-C-directed centromere and Kinetochore assembly. Mol Biol Cell. 2009; 20(19): 4246–4255. PubMed Abstract | Publisher Full Text | Free Full Text

13. Moree B, Meyer CB, Fuller CJ, et al.: CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. J Cell Biol. 2011; 194(6): 855–871. PubMed Abstract | Publisher Full Text | Free Full Text

14. Dambacher S, Deng W, Hahn M, et al.: CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus. 2012; 3(1): 101–110. PubMed Abstract | Publisher Full Text | Free Full Text

15. Fujita Y, Hayashi T, Kiyomitsu T, et al.: Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. Dev Cel. 2007; 12(1): 17–30. PubMed Abstract | Publisher Full Text
Tatsuo Fukagawa
Graduate School of Frontier Biosciences, Osaka University, Osaka, Japan

The centromere protein CENP-C is well known as an essential component for functional kinetochore assembly. Due to importance of this molecule, CENP-C must be conserved in Fish species. The author performed BLASTP searches with the conserved CENP-C motif sequence, but any CENP-C homologues in Fish lineages were not identified with this sequence. However, as there is a putative CENP-C sequence in shark genome, BLASTP searches were carried out with C-terminal domain sequence of putative shark CENP-C. Then, the author identified CENP-C candidates from various teleost genomes. Although the author does not show that candidate proteins localize to centromeres, these candidates contain related sequences of CENP-C motif, which were a little divergent from the vertebrate consensus sequences. Interestingly, the authors found that various vertebrate M18BP1s, which are recruited to centromeres by CENP-C, contain the CENP-C motif sequence, but mammalian homologues do not contain the motif sequence.

This is an observation article and finding of the CENP-C motif in the M18BP1 sequence is interesting. However, to improve the quality of the paper, the author should revise the manuscript. My specific concerns are following.

1. As the author recognizes, it is necessary to verify whether CENP-C candidates from teleost genome really localize to Fish centromeres. As the author obtained a candidate from Zebrafish genome, such an experiment is not difficult with the Zebrafish experimental system. If the author added localization data, Figure 2 would be interesting.

2. Related to Figure 2. If the author shows sequence comparison of Cupin domain in teleost sequences, it would be helpful.

3. When the author discuss about central region of human CENP-C, it may be better to cite a recent paper by Nagpal et al. (Mol. Biol. Cell, 2015), which says that central region sequence does not exist in chicken CENP-C. Then, the author can emphasize that the central motif is really important for CENP-A binding.
4. The author described some results of the analysis in the Introduction. This is not necessary and it would be better to cut off redundant description.

**Competing Interests:** No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 12 Jan 2016

**Leos Kral,** University of West Georgia, Carrollton, USA

I have revised the manuscript to address concerns #2, #3 and #4. Unfortunately, I do not have the resources to carry out the localization experiment (concern #1). My main motivation in publishing these observations is to bring awareness of this issue to individuals who may have the resources and interest to follow up with the relevant experiments.

**Competing Interests:** No competing interests were disclosed.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com