TESTING THE CENTROMERE-DRIVE HYPOTHESIS IN PRIMATES

L. Enrique. Gomez[1], Emily A. Beck[2], & Kirstin N. Sterner[1]

[1] University of Oregon, Department of Anthropology, [2] University of Oregon, Institute of Ecology and Evolution

BACKGROUND

Chromosomal centromeres play a critical role in the process of cell division. Centromeres act as binding sites for microtubules that pull chromosomes apart during mitosis and meiosis. Despite this conserved function, the centromeres themselves can vary in size and sequence content between species.

Rapid evolution in these regions can also drive rapid evolution in centromere-associated proteins (kinetochore). Previous work has suggested these rapid changes are likely to accumulate in one of two essential centromere components; either CENP-A or CENP-C (**Fig. 1**). Through compensatory coevolution, positive selection can subsequently cascade into other essential protein complexes resulting in hybrid incompatibility. Cascading selection from the centromere to CENP-A was previously reported in *Drosophila* by Beck et al. *Sci Rep.* 2015, demonstrating the extension of positive selection to the essential Condensin I complex (*SMC2, SMC4, NCAPD2, NCAPG, NCAPH* orthologs). Rapid evolution of both centromeric DNA and centromere-associated proteins suggest the centromere complex is undergoing an evolutionary tug-of-war between selfish centromeric DNA and kinetochore proteins struggling to maintain centromere function (the centromere drive hypothesis).

Research Objective:

To test for evidence of the centromere drive hypothesis in primates by examining genes that encode for members of two centromeric components (CENP-A and CENP-C).

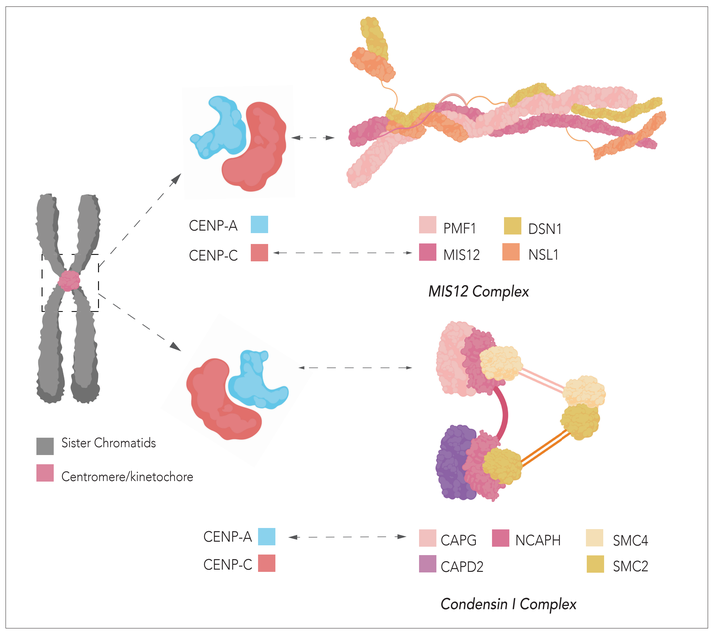
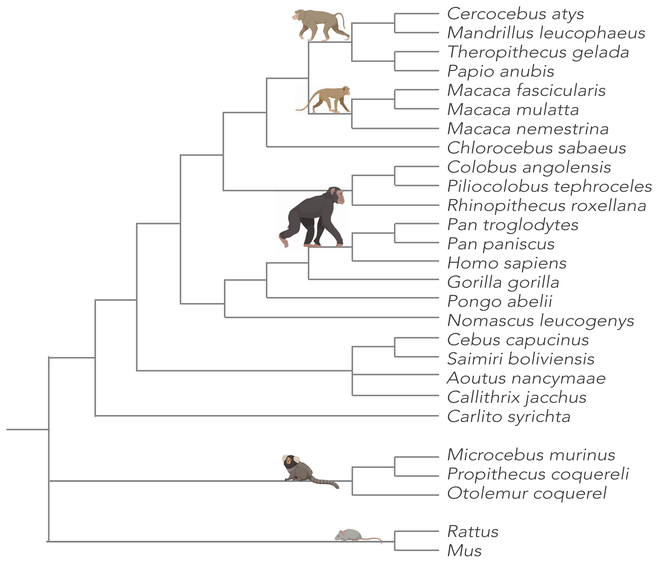


Figure 1: Representation showing the hypothesized Centromeric component (CENP-A and CENP-C) interactions. Petrovic. et. all., *Cell.* 2016. Dash line symbolizes protein interactions.

METHODS



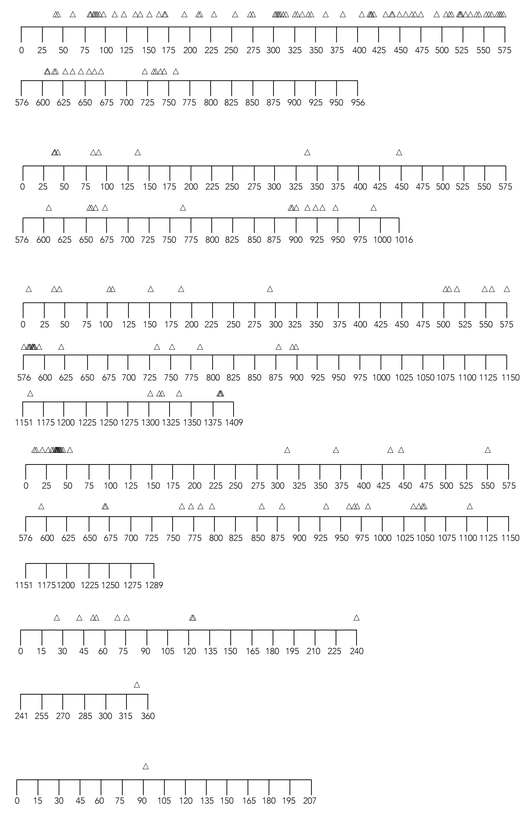
To test if kinetochore-associated proteins evolve rapidly in other animals, we examined the genome sequence of *CENP-A* and *CENP-C* and the genes of their associated protein complexes, Condensin I (*NCAPD2, NCAPG, SMC4, SMC2, NCAPH*) and Mis12 (*DSN1, MIS12, NSL1, PMF1*) respectively, across primates. Sequences were mined from publicly available genomes (Ensembl and NCBI; **Fig. 2**), aligned using Clustal-Omega and manually checked in Mesquite to ensure that protein-coding sequences conform to codon boundaries.

We tested for positive selection using the program codeml in the PAML v4.5 package. We compared the fit of our data to a nearly neutral model (M7) and a model that allows for a fraction of sites to be evolving under positive selection (M8). Model M7 allows ω to vary among codons according to a beta distribution while M8 adds an additional class of codons with ω>1. When the M8 model was a significantly better fit to our data than the M7 model, we identified individual sites under positive selection using the Bayes empirical Bayes (BEB) calculation of posterior probabilities for site classes (**Fig. 3**).

Figure 2. Taxa included in this study and species tree used in PAML analyses. Note: a few species were not available for some genes.

RESULTS

We found evidence suggestive of positive selection for *CENP-C, NCAPD2, NCAPG, SMC4, DSN1,* and *MIS12* (Figure 3). These findings support the centromere-drive hypothesis.



LOCUS	M7	M8	LLR	P
CENPC	-10135.16325	-10117.39737	35.531764	<0.0001**
NCAPD2	-14805.50788	-14798.39506	14.225642	0.0008**
NCAPG	-9867.546905	-9863.359969	8.373872	0.015*
SMC4	-11146.91968	-11134.66885	24.50167	<0.0001**
DSN1	-3705.829358	-3701.64856	8.361596	0.015*
MIS12	-2206.737261	-2202.253759	8.967004	0.011*

Figure 3: Distribution of the positively selected sites in *CENP-C, NCAPD2, NCAPG, SMC4, DSN1,* and *MIS12* (left panel) shown relative to amino acid position. Positively selected sites were identified using the BEB method and show significant results after the M8M7 test (right panel). △ denotes location of positively selected sites in protein.

DISCUSSION AND FUTURE DIRECTIONS

Our findings show evidence of positive selection on *CENP-C, NCAPD2, NCAPG, SMC4, DSN1,* and *MIS12* , suggesting a complex tug-of war between the centromere DNA and centromere-associated proteins that drives centromere function. Partial *PMF1* (a component of the MIS12 complex) also showed evidence of positive selection ($p < 0.0015$).

This work is part of a larger collaborative project that is investigating the evolution of centromere-associated proteins across Animalia (e.g., fish, wasp, mammals), with the purpose of identifying the differences in centromere function that may explain patterns of speciation.

The results seen here are consistent with positive selection, however functional tests are needed to see if substitutions identified in our analysis affect protein function and organism fitness.

We are also interested in examining variation in these genes in the context of hybridization in primates.