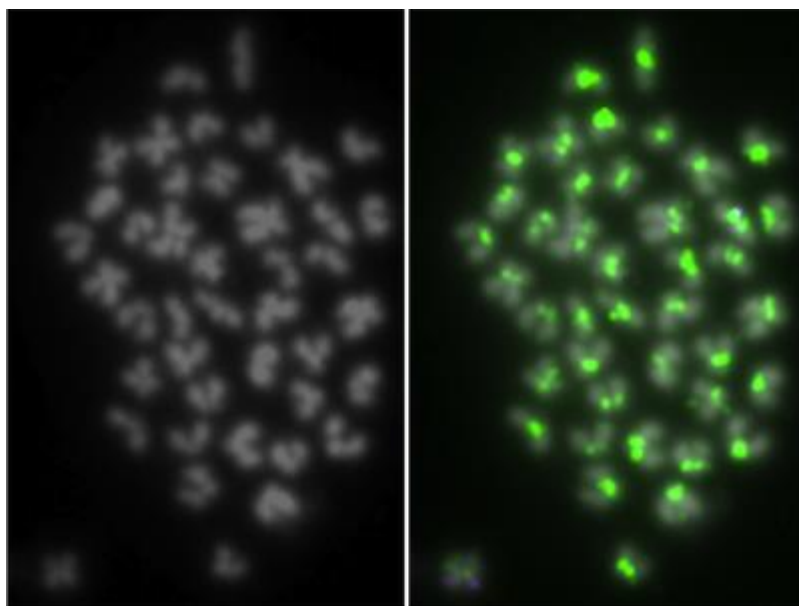


Filling in the gaps: centromere of stickleback fish identified

November 15, 2015

L Koch



Metaphase chromosome spread from threespine stickleback fish. The centromere repeat probe (green, right) hybridizes to the primary constriction on metaphase chromosomes (grey, left).

Image provided by Jennifer Cech.

Cell division is one of the most basic and important processes of living organisms. Cells must replicate and accurately segregate their chromosomes into daughter cells in order to ensure their survival. On the other hand, the regulation of cell division is often abnormal in cancer cells, which makes it an attractive target for therapeutic intervention. Accurate separation of chromosomes by the mitotic spindle depends upon their connection to spindle fibers at a highly regulated location of each chromosome called the centromere.

Despite their functional importance and the conservation of the cell division machinery across eukaryotic organisms, there is a fascinating variability in the underlying DNA sequence of the centromere as well as in many of the proteins that bind to the centromere. How can diverse organisms carry out this same process given such variability at the source of the connection between the DNA and the spindle? This is an ongoing topic of research and has been called the "centromere paradox". A recent study led by graduate student Jennifer Cech in the laboratory of Dr. Catherine Peichel (Basic Sciences) set out to identify and characterize the centromeric sequence of the threespine stickleback fish (*Gasterosteus aculeatus*), which is an important model system in evolutionary biology and genetics. Their work was recently published in *Chromosome Research*.

In addition to the variation in the sequence between organisms, another major obstacle to sequencing centromeres is the fact that they are highly repetitive and AT-rich regions, making aligning and assembling sequencing reads extremely difficult. For this reason, “we really don’t know the sequence of centromeres from many species, and centromeres have been called the dark matter of the genome” , said Dr. Peichel (Basic Sciences). To overcome this, the authors set out to identify the sequence of threespine stickleback CENP-A, a histone H3 variant that is known to be the epigenetic mark of the centromere across nearly all eukaryotes. Using the sequence of the zebrafish CENP-A, the authors identified a homologous CENP-A gene in the threespine stickleback genome. The gene encodes a 148 amino acid protein with the characteristic histone-fold domain in the C-terminus and a divergent N-terminal tail. To confirm this protein marked the centromeric histones, they raised an antibody against the N-terminus of CENP-A, which varies considerably from the canonical histone H3 protein, and showed by immunofluorescence that the antibody recognizes distinct foci that localize to the primary constriction of each chromosome in nuclei and chromosome spreads, consistent with it marking the centromeric histone H3, CENP-A.

Now the authors were ready to identify the centromeric DNA sequence of the threespine stickleback. Using their stickleback CENP-A antibody they performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). Next they used a cluster analysis pipeline developed in the laboratory of Dr. Steve Henikoff (Basic Sciences) in order to find sequences that were enriched in the immunoprecipitated DNA compared to total input DNA. A 186 base pair repeat was identified, bearing the characteristics of other known centromeric repeats such as being AT-rich and harboring a CENP-B box motif. Furthermore, at 186 nucleotides, it is a length of DNA that could wrap a single nucleosome.

In order to confirm this repeat marked the centromere of each chromosome, the authors performed FISH (fluorescence *in situ* hybridization) on the stickleback fish chromosomes using a probe against their GacCEN (*G. aculeatus* centromere) sequence. They found that the sequence localized to the primary constriction of 41 out of 42 chromosomes in nuclei and chromosome spreads. The authors hypothesized that the chromosome that was unlabeled by the GacCEN probe was the Y chromosome, given that the sequence of the Y centromere of mammals is known to be highly divergent from the centromere of the other chromosomes. Indeed, the authors were able to identify the unlabeled chromosome as the Y by also using probes that distinguish the X and Y chromosomes of the stickleback fish.

"We are now working hard to identify the Y chromosome centromere sequence," said Dr. Peichel (Basic Sciences). "Once we do this, we will be able to see how it has evolved in other stickleback species, including two species that have a fusion between the Y chromosome and another chromosome. " It will certainly be exciting to see how these fusion chromosomes have contributed to the evolution of stickleback fish species. The rapid evolution of centromeres has been hypothesized to be important for reproductive isolation of emerging species.

[Cech JN, Peichel CL](#). 2015. Identification of the centromeric repeat in the threespine stickleback fish (*Gasterosteus aculeatus*). *Chromosome Research*. 23(4):767-79. doi: 10.1007/s10577-015-9495-3.

This work was funded by National Science Foundation, National Institutes of Health and Fred Hutchinson Cancer Research Center.