Centromere identification and inactivation on a neo-Y chromosome fusion in threespine stickleback fish (*Gasterosteus aculeatus*)

Jennifer N. Cech

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Catherine Peichel, Chair

Sue Biggins

Steve Henikoff

Program Authorized to Offer Degree:

Molecular and Cellular Biology

Portions of this dissertation were adapted from the following publication:

Cech JN and Peichel CL (2015) Identification of the centromeric repeat in the threespine stickleback fish (*Gasterosteus aculeatus*). *Chromosome Research*, 23(4): 767-779

University of Washington

Abstract

Centromere identification and inactivation on a neo-Y chromosome fusion in threespine stickleback fish (*Gasterosteus aculeatus*)

Jennifer N. Cech

Dr. Catherine L. Peichel
Member, Divisions of Basic Sciences and Human Biology,
Fred Hutchinson Cancer Research Center
Affiliate Professor, Department of Biology,
University of Washington

Centromeres are the primary constriction observed on many chromosomes, and they are required for normal cell division. Having one and only one centromere per chromosome is essential for proper chromosome segregation during both mitosis and meiosis. Chromosomes containing two centromeres (dicentric) often mis-segregate during cell division, resulting in aneuploidy or chromosome breakage. Dicentric chromosomes can be stabilized by centromere inactivation, a process which re-establishes monocentric chromosomes. There are two proposed mechanisms of centromere inactivation: a solely epigenetic mechanism involving loss of the centromeric histone, also called centromere protein A (CENP-A), or a genetic mechanism involving deletion or mutation of centromeric DNA. However, little is known about this process in naturally occurring

dicentric chromosomes. For my dissertation, I characterized the mechanism of centromere inactivation on a Y chromosome-autosome fusion (referred to as a neo-Y chromosome) that has been fixed in Japan Sea threespine stickleback fish (*Gasterosteus nipponicus*).

In order to characterize the Japan Sea neo-Y chromosome, I first needed to identify the threespine stickleback centromeric DNA sequence. Centromere sequences exist as gaps in many genome assemblies due to their repetitive nature. Thus, I took an unbiased approach utilizing CENP-A chomatin immunoprecipitation followed by high-throughput sequencing to identify the centromeric repeat sequence in the closely related Pacific Ocean threespine stickleback fish (*Gasterosteus aculeatus*). A 186-bp, AT-rich repeat was validated as centromeric using both fluorescence in situ hybridization (FISH) and immunofluorescence combined with FISH (IF-FISH) on interphase nuclei and metaphase spreads. This repeat (GacCEN) hybridizes strongly to the centromere on all chromosomes, with the exception of weak hybridization to the Y chromosome.

To test whether epigenetic or genetic inactivation has occurred on the Japan Sea neo-Y chromosome, I used a combination of GacCEN FISH and CENP-A immunofluoresence on metaphase chromosome spreads. I demonstrated that there has been epigenetic inactivation of the centromere derived from the Y chromosome on the Japan Sea neo-Y chromosome. Furthermore, my data suggest that there may be genetic changes to the centromere derived from the ancestral Y chromosome, potentially contributing to its inactivation.

Together, my work provides the first validated sequence information for the threespine stickleback centromere. Additionally, the Japan Sea stickleback neo-Y is one

of the few examples of a naturally-occurring and stable dicentric chromosome involving two functionally important chromosomes that shows evidence for centromere inactivation. It is also one of the first examples showing centromere inactivation as a potential mechanism used to maintain a chromosome fusion that may play a role in the process of speciation between the Pacific Ocean and Japan Sea sticklebacks.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. Catherine Peichel for her mentorship throughout my graduate career. She instilled in me a great work ethic, taught me how to think critically, and helped me better my experimental design. I would also like to thank all the members of the Peichel lab without whom this project would never have been completed.

I thank my graduate program (MCB) for providing me with the opportunity to study at the University of Washington and the Fred Hutchinson Cancer Research Center. I thank Jeff Delrow and Ryan Basom of the Fred Hutchinson Cancer Research Center Genomics Shared Resource for help with ChIP-seq data analysis, Jaki Braggin for help with the Western blots, and the Henikoff, Malik, and Peichel labs for helpful discussions.

I would like to thank my thesis committee members: Sue Biggins, Harmit Malik, Steve Henikoff, and Christine Disteche for help and support throughout this project. I also thank Sue Biggins and Steve Henikoff for reading this thesis.

This research was supported by a National Science Foundation Graduate Research Fellowship (DGE-1256082), the National Institutes of Health Chromosome Metabolism and Cancer Training Grant (T32 CA009657), and the Fred Hutchinson Cancer Research Center.

DEDICATION

This dissertation is dedicated to my parents, Tom and Carol Cech, who have inspired in me a love of science since I was a child. Their patience and guidance has supported me in the years leading up to and during my graduate work. I am forever thankful.

Chapter 1: Background

Introduction

The ability to pass on genetic information to offspring is dependent on proper cell division. Ensuring that each daughter cell receives the correct number of chromosomes is also critical for normal cellular function and growth. Emphasizing these important functions, errors in chromosome segregation have been associated with birth defects and cancer. It is clearly important to stably maintain chromosome number during the lifetime of an organism. However, on an evolutionary timescale, chromosomes are highly dynamic, and can undergo both fission and fusion leading to changes in chromosome number. My thesis project addresses how cells maintain proper chromosome segregation amidst chromosomal changes in an evolutionary context.

Centromeres

The centromere of a chromosome was first described by Walther Flemming in the 1880's as the primary constriction observed on condensed chromosomes during both mitosis and meiosis (Flemming 1880). This primary constriction is the region of the chromosome where microtubules attach, allowing for the segregation of chromosomes during cell division. The attachment of the microtubules to the centromere occurs via a multi-protein structure called the kinetochore. While the composition of the kinetochore varies between species, from hundreds of known proteins in humans to around 40 proteins in budding yeast (Gascoigne and Cheeseman 2011; Gascoigne and Cheeseman 2012; Biggins 2013), it has a conserved function to provide a link between the microtubules and the centromeric DNA on each chromosome.

Genetic characteristics of centromeres

Chromosomes can be defined by the relative position of their centromere. Chromosomes with a centromere at the very end are called "telocentric". Those with a centromere very close, but not at the end are called "acrocentrics." Chromosome with centromeres near, but not at the middle, are called "submetacentric" and those with a centromere at the center are called "metacentric."

Chromosomes can also be defined by the type of centromere they contain. They can be divided into two main classes: holocentric or monocentric. Holocentric chromosomes contain many centromeres along the entire length of the chromosome, and are found in organisms such as nematodes (*Caenorhabditis elegans)*, butterflies (*Heliconius melpomene*), moths (*Bombyx mori*), and dragonflies (*Ladona fulva*) (Drinnenberg et al. 2014; Steiner and Henikoff 2014) Monocentric chromosomes only have a single centromeric region on each chromosome and are the most common type of centromere found in eukaryotes, including model organisms such as fission yeast (*Schizosaccharomyces pombe*), fruit fly (*Drosophila melanogaster*), and human (*Homo sapiens*) (Fukagawa and Earnshaw 2014a; McKinley and Cheeseman 2015). Monocentric chromosomes can be further divided into point and regional centromeres. Point centromeres are considered the simplest type of centromere and are found in budding yeast (*Saccharomyces cerevisiae*) (Kobayashi et al. 2015). The centromeres of budding yeast are termed "point" because a unique 125bp DNA sequence defines the single centromere on each yeast chromosome (Carbon and Clarke 1984; Clarke and Carbon 1985). By contrast, many organisms, including mammals and some plants, contain

"regional" centromeres. Regional centromeres are also found at a single location on each chromosome, but are comprised of kilobases to megabases of DNA. The sequence of these regional centromeres is often repetitive and AT rich, but the specific sequence varies dramatically between species (Henikoff et al. 2001; Alkan et al. 2011; Melters et al. 2013), and can even vary among centromeres within a species (Nagaki et al. 2004; Piras et al. 2010; Tek et al. 2010; Shang et al. 2010; Gong et al. 2012). Because of their repetitive nature, centromere sequences are difficult to assemble and often remain as gaps in many species genome assemblies. Therefore, the centromere sequence is unknown in most species.

Epigenetic characteristics of centromeres

Consistent with the highly variable sequence of the centromere, the centromeric DNA itself is neither necessary nor sufficient to define an active centromere. For example, it has been found that by artificially removing the centromeric region on a chromosome, the chromosome can avoid aneuploidy by creating a new centromere (Ishii et al. 2008; Ketel et al. 2009). These "neo" centromeres do not depend on the underlying DNA sequence, but do tend to form in gene-poor and repetitive regions.

Progress in identifying the mark of an active centromere was made in the 1980s when (Earnshaw and Rothfield 1985) found that the sera from patients with a disease called Limited scleroderma, or CREST syndrome, bound to the centromeric region on metaphase chromosomes. The proteins in this sera were further identified as centromeric proteins (CENP) A, B, and C. CENP-A is a histone variant, only found at the centromere (Palmer et al. 1987; Palmer et al. 1989), with similar characteristics to the H3 histone

(Sullivan et al. 1994), such as the C-terminal Histone Fold Domain (HFD) containing three alpha helices and two linker regions in (Sullivan et al. 1994; Vermaak et al. 2002), Unlike H3 which is nearly identical between all known organisms (Roach et al. 2012), CENP-A varies between organisms in both N-terminal tail, as well in the loop 1 region of the C-terminal HFD (Henikoff et al. 2001; Talbert 2002; Schueler et al. 2010). The CENP-A N-terminal tail is highly divergent; it varies greatly in length and sequence among CENP-A proteins between different species and can even vary between closely related species (Malik and Henikoff 2001; Henikoff et al. 2001; Malik et al. 2002; Talbert 2002; Cooper and Henikoff 2004; Ravi et al. 2010; Roach et al. 2012). Variation in the loop1 region between species is thought to be important for CENP-A function (Black et al. 2007)

The CENP-A protein is a key component of the nucleosomes found specifically at centromeres. In non-centromeric nucleosomes, approximately 170-190 bp of DNA is wrapped around 2 copies each of 4 histones (H2A, H2B, H3, and H4). In centromeric nucleosomes, H3 is replaced with CENP-A, which is sometimes called CenH3 (Figure 1-1) (Palmer et al. 1987; Palmer et al. 1989; Palmer et al. 1991; Sullivan et al. 1994). CENP-A is also an essential protein; experimental deletion of CENP-A is lethal in *S. cerevisiae*, *C. elegans*, *Candida albicans* and in chicken, mouse and human cell lines (Stoler et al. 1995; Buchwitz et al. 1999; Howman et al. 2000; Sanyal and Carbon 2002; Regnier et al. 2005; Fachinetti et al. 2013). In addition to being essential, CENP-A is also sufficient to define an active centromere. Experimentally mislocalizing CENP-A to non-centromeric DNA was sufficient to recruit other kinetochore components, and in some cases induce dicentric associated mitotic errors (Heun et al. 2006; Allshire and Karpen

2008; Olszak et al. 2011; Barnhart et al. 2011; Mendiburo et al. 2011; Sekulic and Black 2012). Additionally, it was found that neo-centromeres are made active by the presence of CENP-A (Voullaire et al. 1993; Warburton et al. 1997; Warburton 2004; Marshall et al. 2008; Shang et al. 2013; Fukagawa and Earnshaw 2014b). These critical experiments determined that CENP-A is the epigenetic mark of active centromeres.

Establishment and maintenance of centromeres

Additional histone modifications decorate the chromatin at centromeres. In an active centromere, the core centromeric chromatin ("centrochromatin") consists of CENP-A containing nucleosomes interspersed with canonical H3 containing nucleosomes that are di-methylated at the fourth lysine (H3K4me2) (Sullivan and Karpen 2004; Dunleavy et al. 2005). Interestingly, H2K4me2 is usually associated with transcriptionally active chromatin (Bernstein et al. 2002), leading researchers to propose there is a functional role for more "open" chromatin at the centromere (Sullivan and Karpen 2004), including a potential role of centromeric transcription to promote CENP-A recruitment to the centromere (Allshire and Karpen 2008; Bergmann et al. 2011). Outside of the core centromeric chromatin, the "pericentromeric" chromatin is packaged as traditional heterochromatic H3-containing nucleosomes that are di and tri-methylatated at lysine 9 (H3K9me2/3) (Sullivan and Karpen 2004), a histone modification associated with transcriptionally silent chromatin (Sullivan and Karpen 2004; Dunleavy et al. 2005). These histone modifications resemble classic repressive heterochomatin, and have been proposed to limit centromere size and function by acting as a barrier to the expansion of

the centrochromatin (Maggert and Karpen 2001; Sullivan 2002; Sullivan and Karpen 2004).

Dicentric chromosomes

While the centromere has a conserved and essential function for chromosomes, it is also essential that every chromosome has one and only one active centromere in organisms with monocentric chromosomes. A chromosome with two centromeres is called a dicentric and is unstable during cell division. Barbara McClintock first described this instability of dicentric chromosomes and termed it the breakage-fusion-bridge cycle (Figure 1-2) (McClintock 1939; McClintock 1941). When a dicentric chromosome enters mitosis or meiosis, problems arise when the microtubules attach to the centromere. A normal monocentric chromosome only has a single centromere that can bind microtubules and form a bipolar spindle, allowing normal chromosome segregation. However, a dicentric has two centromeres that can both bind microtubules. If microtubules that originated from the same spindle pole bind to both centromeres, no problems occur. However, if the microtubules from spindle poles on the opposing sides of the cell bind to the two centromeres, the chromosome will be pulled in opposite directions. This leads to merotelic kinetochore attachments where the dicentric chromosome lags in the middle during anaphase as it is being pulled to opposite poles. The physical tension on the chromosome can lead to breakage of the chromosome. When the broken ends are replicated during the next S- phase of the cell cycle, the broken ends can anneal back to each other, re-creating a dicentric. Not only does the breakage-fusion-bridge cycle start again, but the breakage and re-annealing leads to gene loss and gene

gain. This is evident in cancer cells where dicentrics are often observed, leading to cells with "increased genetic heterogeneity" (Gisselsson et al. 2000; Gisselsson et al. 2001).

Dicentric stabilization and centromere inactivation

Based on the model presented above, dicentric chromosomes should be very unstable. However, it has also been shown that if chromosomes fuse, and the two centromeres are close enough to each other, they will act as a single centromere (Koshland et al. 1987; Page and Shaffer 1998; Sullivan and Willard 1998). This is seen both in chromosome fusions that were artificially created in the lab as well as in naturally occurring Robertsonian fusions; which are fusions between two acrocentric chromosome, where a shorter intercentromeric distance allows for two active centromeres, and a long intercentromeric distance is associated with centromere inactivation (Page and Shaffer 1998; Sullivan and Willard 1998; Higgins et al. 2005; Lange et al. 2009).

There are also examples of non-Robertsonian fusions that should create dicentric chromosomes and are stable, suggesting that the chromosomes are able to alleviate the problems associated with being dicentric by inactivating one of the centromeres, thereby re-establishing monocentric chromosomes. Most of our current understanding of how dicentrics are stabilized comes from artificially created dicentric chromosomes in yeast, fruit flies, maize, and human cell lines (Sears and Câmara 1952; Agudo et al. 2000; Higgins et al. 2005; Pobiega and Marcand 2010; Zhang et al. 2010; Stimpson et al. 2010; Sato et al. 2012; Stimpson et al. 2012), and a few examples of naturally occurring dicentric chromosome in humans and wheat (Therman et al. 1986; Maraschio et al. 1990; Fisher et al. 1997; Page and Shaffer 1998; Sullivan and Willard 1998; Han et al. 2006;

Lange et al. 2009; MacKinnon and Campbell 2011; Koo et al. 2011; Stimpson et al. 2012; Liu et al. 2015) (Table 1-1).

From these studies, there are three proposed mechanisms of dicentric stabilization: chromosome re-breakage, and centromere inactivation by either: genetic deletion or epigenetic inactivation of one centromere (Figure 1-3). In chromosome re-breakage, the dicentric is simply re-broken between the two centromeres, thereby re-creating two monocentric chromosomes. The other two models both propose stabilization through two different mechanisms of centromere inactivation. In genetic inactivation, there is deletion or mutation of the DNA sequence of one centromere, preventing it from binding CENP-A and therefore microtubules. In epigenetic inactivation, the DNA of the inactivated centromere is still present, but CENP-A is not. Without CENP-A, the centromeric DNA does not recruit kinetochore components and therefore does not bind mictrotubules during cell division.

Examples of dicentric stabilization and centromere inactivation

In artificially created dicentric chromosome in fission yeast (*S. pombe*), less than 0.1% of the cells created are viable, further emphasizing the detrimental affect of having multiple centromeres. In the surviving cells, chromosome stability can be achieved by each of the three mechanisms. Epigenetic inactivation is the most common and occurs in ~72.1% of the cells, genetic deletion occurs in ~12.8% of cells, and or re-breakage occurs in the ~15.1% of cells. Additionally, this study showed that once one of the centromeres has been epigenetically inactivated, it remains inactivated through multiple cell cycles (Sato et al. 2012).

In plants, centromere inactivation has been seen in both artificial as well as natural dicentric and polycentric chromosome. Natural dicentrics are found in the supernumerary B chromosomes in maize. These chromosomes are considered dispensable, but still maintain essential centromere components. Because they are dispensable, cases of B translocations and subsequent centromere inactivation have been documented. In these studies, lack of CENP-A at one of the centromeres indicated epigenetic inactivation, while both the active and inactive centromere still contain centromeric DNA (Han et al. 2006; Liu et al. 2015). In addition to the removal of CENP-A at the inactive centromere, DNA methylation also changes at the inactivated centromere. While the active centromeres are hypomethylated at cytosine bases, the entire centromeric region of the B chromosome became hypermethylated at cytosine upon centromere inactivation (Koo et al. 2011). In artificially created tricentric wheat chromosomes (Sears and Câmara 1952), chromosome stabilization occurs via chromosome re-breakage and epigenetic inactivation (Zhang et al. 2010). In addition to CENP-A removal from the core centromere, H3K27me2/3 histone modification replaces the H3K4me2 histone modification that is usually found at active centromeres. H3K27me2/3 is involved in repressing gene transcription in plants (Zhang et al. 2007). This change in methylation and CENP-A removal inactivates the centromeric DNA and makes the chromatin more heterochromatic (Sullivan and Karpen 2004; Dunleavy et al. 2005; Zhang et al. 2010; Ribeiro et al. 2010; Bergmann et al. 2011).

Artificially created human dicentrics were made in human fibrosarcoma cells by transiently expressing a mutant protein which sequesters the telomere protein (TRF2) away from chromosome ends, ultimately disrupting telomeres and leading to

chromosome fusions (Stimpson et al. 2010). In these dicentrics, 80% involved acrocentrics with variable inter-centromeric distances. At four days after induction, functional dicentrics were still observed, but genetic deletion as well as epigenetic centromere inactivation was observed within a few weeks. Some functional dicentrics were observed even after 180 cell divisions (Stimpson et al. 2010); it would be interesting to see if this phenomena holds true in naturally occurring dicentrics.

There are examples of natural human dicentrics in both cancer cell lines as well as case studies, which often involve the Y chromosome. While it has been shown that the breakage-fusion-bridge cycle is often seen in cancer cell lines, some cancer cell lines do show evidence for centromere inactivation (MacKinnon and Campbell 2011). In contrast to studies in yeast artificial dicentric chromosomes, centromere deletion was the more common inactivation mechanism in human cancer cell lines (MacKinnon and Campbell 2011). In two examples of dicentric Y chromosomes found in clinical patients (Maraschio et al. 1990), genetic deletion of the centromere was found in one case, while epigenetic inactivation was found in the other case.

While we have learned about the mechanisms of dicentric stabilization from these artificially created dicentrics and from human disease patients (Table 1-1), there is currently nothing known about the stabilization of natural dicentric chromosomes, which are often created when chromosomes fuse in an evolutionary context.

Chromosome fusions in evolution

Chromosome fusion and fission events are common during evolution as evidenced by the extensive variation in chromosome number among species. For example, in multicellular

eukaryotes, chromosome number ranges from a diploid number of 2 in the Jack jumper ant (Crosland and Crozier 1986) to 1260 in the Addlers tongue plant (Van der Burg 2004). In eutherian mammals, the diploid chromosome number ranges from 6 to 102 in the Viscacha rat (Gallardo et al. 1999); this diversity results from many independent fusion and fission events (Ferguson-Smith and Trifonov 2007). The mammalian species with the lowest number of chromosome is the Indian muntjac with a diploid number of 6 in females or 7 in males. Interestingly, a closely related species, the Chinese muntjac has a diploid chromosome number of 46. This large difference in chromosome number does not prevent these two sister species from breeding in captivity (Shi et al. 1980; Yang et al. 1995; Wang and Lan 2000). It is speculated that a series of chromosome fusion events between the telocentric chromosomes present in the Chinese muntjac created the much larger Indian muntjac chromosomes (Wang and Lan 2000). Chromosome fusions have also occurred in the primate lineage; it has been proposed that human chromosome 2 is a fusion between chimp chromosomes 2a and 2b (Lejeune et al. 1973; IJdo et al. 1991). While debated, some argue that there is evidence for a cryptic inactivated centromere on the large arm of human chromosome 2 (Avarello et al. 1992). All of these data suggest that chromosome fusion and fission events are common and can account for much of the diversity of karyotype number we see today. Based on these findings, it is theorized that many of these chromosome fusion events must have been accompanied by centromere inactivation to stabilize the chromosome fusions. However, there are no existing data on the mechanisms involved in centromere inactivation during evolution.

Stickleback fish as a model system to study centromere inactivation

The stickleback family (Gasterosteidae) provides a unique opportunity to study the mechanisms that stabilize dicentric chromosomes during evolution. Within the stickleback family, seven species of sticklebacks have evolved in the past 35 million years (Kitano et al 2009; Ross et al 2009)(Figure 1-4). Previous work has identified two independent chromosome fusions involving the ancestral metacentric Y chromosome present in the Pacific Ocean threespine stickleback (*Gasterosteus aculeatus*). This Y chromosome has fused to the acrocentric chromosome 9 in males of the Japan Sea threespine stickleback (*Gasterosteus nipponicus*) species within the past two million years, creating a neo-Y chromosome (Kitano et al. 2009). This Y chromosome has independently fused to the acrocentric chromosome 12 in males of the blackspotted stickleback (*Gasterosteus wheatlandi*) species within the past 15 milion years, creating an independent neo-Y chromosome (Ross et al. 2009). Because the Japan Sea fusion involved the metacentric Y and an acrocentric chromosome, it should have been dicentric upon formation. Preliminary data suggested that there was no loss of a chromosome arm on the Japan Sea neo-Y chromosome fusion or re-breakage of the fusion chromosome. These data suggest that centromere inactivation has occurred on the Japan Sea neo-Y chromosome. The young evolutionary age (less than 2 million years) of the Japan Sea neo-Y fusion chromosome, the availability of the extant ancestral Pacific Ocean threespine stickleback, and genetic and genomic resources for the threespine stickleback system suggested to us that the Japan Sea neo-Y chromosome would be an excellent model system to study the mechanisms of centromere inactivation during karyotype evolution.

Goals of thesis

The goal of my thesis project was to identify the mechanisms that contribute to dicentric stabilization and centromere inactivation of the Japan Sea neo-Y fusion. Because there is no evidence for re-breakage of the neo-Y fusion, we hypothesized that either genetic deletion or epigenetic inactivation had occurred. To test for genetic deletion, I first needed to identify the centromere sequence in threespine sticklebacks (Chapter 2). With the centromere sequence in hand, I was then able to evaluate the contribution of genetic and epigenetic inactivation to stabilization of the Japan Sea neo-Y fusion (Chapter 3). Finally, I performed experiments to address the question of why the neo-Y fusion was fixed in the Japan Sea population (Chapter 4).

Table 1-1: Examples of dicentric chromosome stabilization

Organism	**Re-breakage**	**Genetic Deletion**	**Epigenetic inactivation**	**Natural or artificial**	**Reference**
Fission yeast *S. pombe*	✔	✔	✔	Artificial	Sato et al. 2012
Budding yeast *S. cerevisiae*	✔	✔		Artificial	Pobiega and Marcand 2010 Stimpson et al. 2012 (review)
Fruit fly *D. melanogaster*			✔	Artificial	Agudo et al. 2000
Maize *Z. mays*	✔		✔	Artificial	Sears and Camara 1952 Zhang et al. 2010
Maize *Z. mays*			✔	Natural B chromosomes	Han et al. 2006 Koo et al. 2011 Liu et al. 2015
Human *H. sapiens*		✔	✔	Artificial	Higgins et al 2005 Stimpson et al. 2010
Human *H. sapiens*			✔	Natural Bloom syndrome and Robertsonian fusions	Therman et al. 1989 Page and Shaffer 1998 Stimpson 2012 (review)
Human *H. sapiens*		✔	✔	Natural Y chromosome	Maraschio et al. 1990 Fisher et al. 1997 Lange et al. 2009
Human *H. sapiens*		✔	✔	Natural Cancer cell lines	MacKinnon et al. 2011
Human *H. sapiens*			✔	Natural X chromosomes	Sullivan and Willard 1998

Chapter 1: Figures

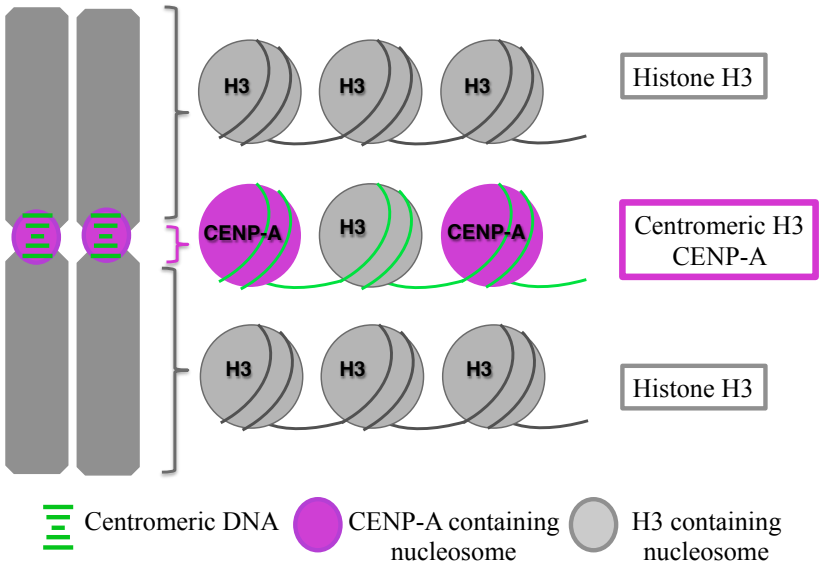


Figure 1-1. CENPA containing nucleosomes at the centromere. Active centromeres are characterized by a specific histone H3 variant (CENP-A) only found in centromeric nucleosomes. Centromeric chromatin consists of H3 nucleosomes interspersed with CENP-A containing nucleosomes. Centromeric DNA is often, but not always, comprised of repetitive DNA (green dashed lines).

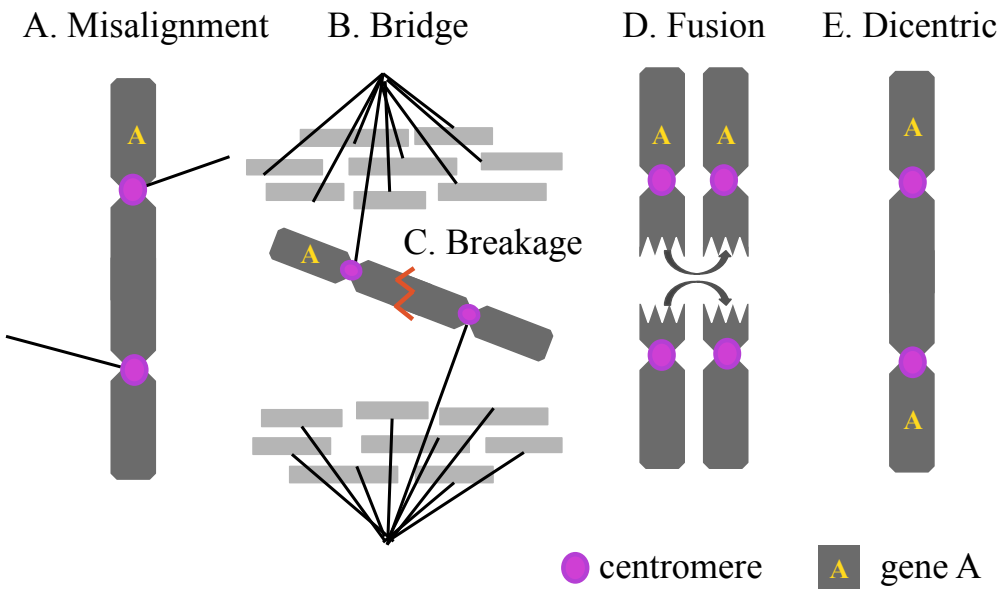


Figure 1-2. Breakage-Fusion-Bridge cycle associated with dicentric chromosomes. Dicentric chromosomes (indicated by two purple dots) are prone to errors in segregation if the two centromere are bound by microtubules emanating from opposite poles of the cell (**a**). This leads to anaphase lag or a bridge formation (**b**). The physical tension pulling on the dicentric chromosome from opposite ends can break the chromosome (**c**), resulting in fusion of the broken ends during the next round of replication (**d**). Successive rounds of the breakage-fusion-bridge cycle can lead to gene gain (as shown by gene "A") as well as gene loss (**e**).

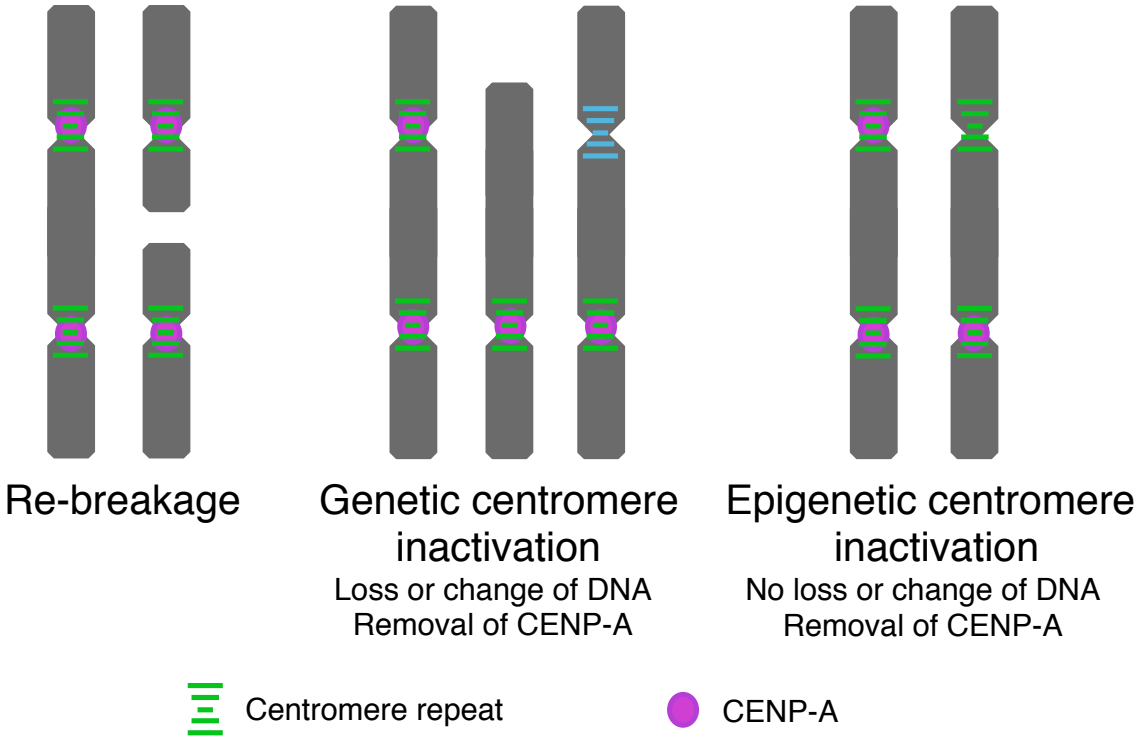


Figure 1-3. Three proposed mechanisms of dicentric chromosome stabilization. (**a**) In the re-breakage model of stabilization, the dicentric chromosome is re-broken into two monocentric chromosomes with no loss of DNA or removal of CENP-A. (**b**) There are two mechanisms of stabilization involving centromere inactivation. In genetic inactivation, the underlying DNA of one centromere is either removed or altered, resulting in removal of CENP-A. In epigenetic inactivation, there is no underlying removal or change to the centromeric DNA; inactivation occurs solely by the removal of CENP-A.

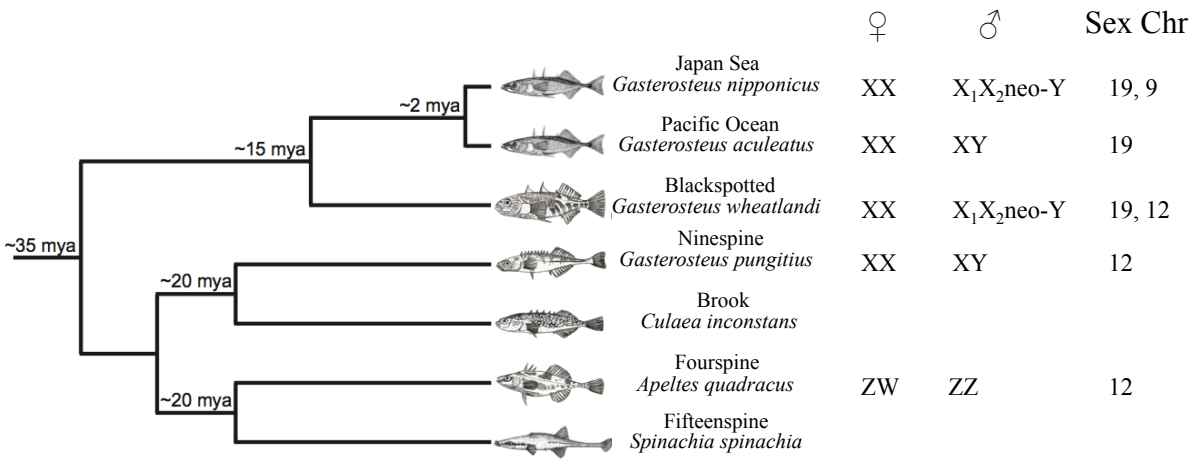


Figure 1-4. The seven species of stickleback fish. Five of the seven species have identified sex chromosomes. Japan Sea male stickleback fish have a fusion between the ancestral Y chromosome (chromosome 19) and chromosome 9 resulting in an X_1X_2neo-Y sex chromosome system. X_1 represents the ancestral X chromosome, and X_2 represents the unfused chromsome 9. The Blackspotted male stickleback fish also have a fusion between the ancestral Y chromosome (chromosome 19) and chromosome 12, resulting in an independently evolved X_1X_2neo-Y sex chromosome system.

Chapter 1 References:

Agudo M, Abad JP, Molina I, et al (2000) A dicentric chromosome of Drosophila melanogaster showing alternate centromere inactivation. Chromosoma 109:190–196

Alkan C, Cardone MF, Catacchio CR, et al (2011) Genome-wide characterization of centromeric satellites from multiple mammalian genomes. Genome Research 21:137–145

Allshire RC, Karpen GH (2008) Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet 9:923–937

Avarello R, Pedicini A, Caiulo A, et al (1992) Evidence for an ancestral alphoid domain on the long arm of human chromosome 2. Human Genetics 89:247–249

Barnhart MC, Kuich PHJL, Stellfox ME, et al (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J Cell Biol 194:229–243

Bergmann JH, Rodriguez MG, Martins NMC, et al (2011) Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. EMBO J 30:328–340

Bernstein BE, Humphrey EL, Erlich RL, et al (2002) Methylation of histone H3 Lys 4 in coding regions of active genes. Proc Natl Acad Sci USA 99:8695–8700

Biggins S (2013) The Composition, Functions, and Regulation of the Budding Yeast Kinetochore. Genetics 194:817–846

Black BE, Jansen LET, Maddox PS, et al (2007) Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Molecular Cell 25:309–322

Buchwitz BJB, Ahmad KK, Moore LLL, et al (1999) A histone-H3-like protein in C. elegans. Nature 401:547–548

Carbon J, Clarke L (1984) Structural and functional analysis of a yeast centromere (cen3). J Cell Sci Suppl 1–17

Clarke L, Carbon J (1985) The structure and function of yeast centromeres. Annu Rev Genet 29–56

Cooper JL, Henikoff S (2004) Adaptive evolution of the histone fold domain in centromeric histones. Molecular Biology and Evolution 21:1712–1718

Crosland MW, Crozier RH (1986) Myrmecia pilosula, an ant with only one pair of chromosomes. Science 231:1278–1278

Drinnenberg IA, deYoung D, Henikoff S, Malik HS (2014) Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. eLife e03676

Dunleavy EE, Pidoux AA, Allshire RR (2005) Centromeric chromatin makes its mark. Trends in Biochemical Sciences 30:4–4

Earnshaw WC, Rothfield N (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma 91:313–321

Fachinetti D, Folco HD, Nechemia-Arbely Y, et al (2013) A two-step mechanism for epigenetic specification of centromere identity and function. Nat Cell Biol 15:1056–1066

Ferguson-Smith MA, Trifonov V (2007) Mammalian karyotype evolution. Nat Rev Genet 8:950–962

Fisher AM, Al-Gazali L, Pramathan T, et al (1997) Centromeric inactivation in a dicentric human Y;21 translocation chromosome. Chromosoma 106:199–206

Flemming W (1880). Beitrage zur Kenntniss der Zelle und ihrer Lebenserscheinungen Theil II (in German). Archivfiir Mikroskopische Anatomie 18: 151-259

Fukagawa T, Earnshaw WC (2014a) The Centromere: Chromatin foundation for the kinetochore machinery. Dev Cell 30:496–508

Fukagawa T, Earnshaw WC (2014b) Neocentromeres. Current Biology 24:R946–R947

Gallardo MH, Bickham JW, Honeycutt RL, et al (1999) Discovery of tetraploidy in a mammal. Nature 401:341

Gascoigne KE, Cheeseman IM (2011) Kinetochore assembly: if you build it, they will come. Current Opinion in Cell Biology 23:102–108

Gascoigne KE, Cheeseman IM (2012) T time for point centromeres. Nat Cell Biol 14:559–561

Gisselsson D, Jonson T, Petersén A, et al (2001) Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. Proc Natl Acad Sci USA 98:12683–12688

Gisselsson D, Pettersson L, Höglund M, et al (2000) Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. Proc Natl Acad Sci USA 97:5357–5362

Gong Z, Wu Y, Koblizkova A, et al (2012) Repeatless and repeat-based centromeres in potato: implications for centromere evolution. Plant Cell 24:3559–3574

Han F, Lamb JC, Birchler JA (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. Proc Natl Acad Sci USA 103:3238–3243

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098–1102

Heun P, Erhardt S, Blower MD, et al (2006) Mislocalization of the drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Dev Cell 10:303–315

Higgins AW, Gustashaw KM, Willard HF (2005) Engineered human dicentric chromosomes show centromere plasticity. Chromosome Res 13:745–762

Howman EV, Fowler KJ, Newson AJ (2000) Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc Natl Acad Sci USA 93:1148-1153

IJdo JW, Baldini A, Ward DC, et al (1991) Origin of human chromosome 2: an ancestral telomere-telomere fusion. Proc Natl Acad Sci USA 88:9051–9055

Ishii K, Ogiyama Y, Chikashige Y, et al (2008) Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. Science 321:1088–1091

Ketel C, Wang HSW, McClellan M, et al (2009) Neocentromeres form efficiently at multiple possible loci in candida albicans. PLoS Genet 5:e1000400

Kitano J, Ross JA, Mori S, et al (2009) A role for a neo-sex chromosome in stickleback speciation. Nature 461:1079–1083

Kobayashi N, Suzuki Y, Schoenfeld LW, et al (2015) Discovery of an unconventional centromere in budding yeast redefines evolution of point centromeres. Current Biology 1–9

Koo DH, Han F, Birchler JA, Jiang J (2011) Distinct DNA methylation patterns associated with active and inactive centromeres of the maize B chromosome. Genome Research 21:908–914

Koshland D, Rutledge L, Fitzgerald-Hayes M, Hartwell LH (1987) A genetic analysis of dicentric minichromosomes in Saccharomyces cerevisiae. Cell 48:801–812

Lange J, Skaletsky H, van Daalen SKM, et al (2009) Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. Cell 138:855–869

Lejeune J, Dutrillaux B, Rethoré MO, Prieur M (1973) Comparaison de la structure fine des chromatides d'Homo sapiens et de Pan troglodytes. Chromosoma 43:423-444

Liu Y, Su H, Zhang J, et al (2015) Dynamic epigenetic states of maize centromeres. Front

Plant Sci 6:589

MacKinnon RN, Campbell LJ (2011) The role of dicentric chromosome formation and cecondary centromere deletion in the evolution of myeloid malignancy. Genetics Research International 2011:1–11

Maggert KA, Karpen GH (2001) The activation of a neocentromere in Drosophila requires proximity to an endogenous centromere. Genetics 158:1615–1628

Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics 157:1293–1298

Malik HS, Vermaak D, Henikoff S (2002) Recurrent evolution of DNA-binding motifs in the Drosophila centromeric histone. Proc Natl Acad Sci USA 99:1449–1454

Maraschio P, Zuffardi O, Caiulo A, et al (1990) Deletion of specific sequences or modification of centromeric chromatin are responsible for Y chromosome centromere inactivation. Human Genetics 85:491–494

Marshall OJ, Chueh AC, Wong LH, Choo KHA (2008) Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. The American Journal of Human Genetics 82:261–282

McClintock B (1939) The behavior in successive nuclear divisions of a chromosome broken at meiosis. Proc Natl Acad Sci USA 25:405–416

McClintock B (1941) The stability of broken ends of chromosomes in Zea Mays. Genetics 26:234–282

McKinley KL, Cheeseman IM (2015) The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol 1–14

Melters DP, Bradnam KR, Young HA, et al (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14:R10

Mendiburo MJ, Padeken J, Fulop S, et al (2011) Drosophila CENH3 is sufficient for centromere formation. Science 334:686–690

Nagaki K, Cheng Z, Ouyang S, et al (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145

Olszak AM, van Essen D, Pereira AJ, et al (2011) Heterochromatin boundaries are hotspots for de novo kinetochore formation. Nat Cell Biol 13:799–808

Page SL, Shaffer LG (1998) Chromosome stability is maintained by short intercentromeric distance in functionally dicentric human Robertsonian translocations. Chromosome Res 6:115–122

Palmer DK, O'Day K, Trong HL, et al (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci USA 88:3734–3738

Palmer DK, O'Day K, Wener MH, et al (1987) A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J Cell Biol 104:805–815

Palmer DKD, O'Day KK, Margolis RLR (1989) Biochemical analysis of CENP-A, a centromeric protein with histone-like properties. Prog Clin Biol Res 318:61–72

Piras FM, Nergadze SG, Magnani E, et al (2010) Uncoupling of satellite DNA and centromeric function in the genus equus. PLoS Genet 6:e1000845

Pobiega S, Marcand S (2010) Dicentric breakage at telomere fusions. Genes & Development 24:720–733.

Ravi M, Kwong PN, Menorca RMG, et al (2010) The rapidly evolving centromere-specific histone has stringent functional requirements in arabidopsis thaliana. Genetics 186:461–471

Regnier V, Vagnarelli P, Fukagawa T, et al (2005) CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. Molecular and Cellular Biology 25:3967–3981

Ribeiro SA, Vagnarelli P, Dong Y, et al (2010) A super-resolution map of the vertebrate kinetochore. Proc Natl Acad Sci USA 107:10484–10489

Roach KC, Ross BD, Malik HS (2012) Rapid evolution of centromeres and centromeric/kinetochore proteins. Rapidly Evolving Genes and Genetic Systems pp 83–93

Ross JA, Urton JR, Boland J, et al (2009) Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). PLoS Genet 5:e1000391

Sanyal K, Carbon J (2002) The CENP-A homolog CaCse4p in the pathogenic yeast Candida albicans is a centromere protein essential for chromosome transmission. Proc Natl Acad Sci USA 99:12969–12974

Sato H, Masuda F, Takayama Y, et al (2012) Epigenetic inactivation and subsequent heterochromatinization of a centromere stabilize dicentric chromosomes. Current Biology 22:658-667

Schueler MG, Swanson W, Thomas PJ, et al (2010) Adaptive evolution of foundation kinetochore proteins in primates. Molecular Biology and Evolution 27:1585–1597

Sears ER, Câmara A (1952) A transmissible dicentric chromosome. Genetics 37:125–135

Sekulic N, Ben E Black (2012) Molecular underpinnings of centromere identity and maintenance. Trends in Biochemical Sciences 1–10

Shang W-H, Hori T, Martins NMC, et al (2013) Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. Dev Cell 24:635–648

Shang WH, Hori T, Toyoda A, et al (2010) Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. Genome Research 20:1219–1228

Shi L, Ye Y, Duan D (1980) Comparative cytogenetic studies on the red muntjac, Chinese muntjac, and their F1 hybrids. Cytogenetics and Cell Genetics 26:22-27

Steiner FA, Henikoff S (2014) Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. eLife 3:e02025–e02025

Stimpson KM, Matheny JE, Sullivan BA (2012) Dicentric chromosomes: unique models to study centromere function and inactivation. Chromosome Res 20:595–605

Stimpson KM, Song IY, Jauch A, et al (2010) Telomere disruption results in non-random formation of de novo dicentric chromosomes involving acrocentric human chromosomes. PLoS Genet 6:e1001061

Stoler S, Keith KC, Curnick KE, Fitzgerald-Hayes M (1995) 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 & Development 9:573–586

Sullivan BA (2002) Centromere round-up at the heterochromatin corral. Trends in Biotechnology 20:89-92

Sullivan BA, Karpen GH (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat Struct Mol Biol 11:1076–1083

Sullivan BA, Willard HF (1998) Stable dicentric X chromosomes with two functional centromeres. Nat Genet 20:227–228

Sullivan KF, Hechenberger M, Masri K (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol 127:581–592

Talbert PB (2002) Centromeric localization and adaptive evolution of an arabidopsis histone H3 variant. Plant Cell 14:1053–1066

Tek AL, Kashihara K, Murata M, Nagaki K (2010) Functional centromeres in soybean include two distinct tandem repeats and a retrotransposon. Chromosome Res 18:337–347

Therman E, Trunca C, Kuhn EM, Sarto GE (1986) Dicentric chromosomes and the inactivation of the centromere. Human Genetics 72:191–195

Van der Burg WJ (2004). Ophioglossum Reticulatum. In: Grubben, GJH and Denton OA (Editors). Plant Resources of Tropical Africa 2. Vegetables. PROTA Foundation, Wageningen, Netherlands / Backuys Publishers, Leiden, Netherlands / CTA, Wageningen, Netherlands. pp 404-405

Vermaak D, Hayden HS, Henikoff S (2002) Centromere targeting element within the histone fold domain of Cid. Molecular and Cellular Biology 22:7553–7561

Voullaire LE, Slater HR, Petrovic V (1993) A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? Am J Hum. Genet 52:1153-1163

Wang W, Lan H (2000) Rapid and parallel chromosomal number reductions in muntjac deer inferred from mitochondrial DNA phylogeny. Molecular Biology and Evolution 17:1326–1333

Warburton PE (2004) Chromosomal dynamics of human neocentromere formation. Chromosome Res 12:617–626

Warburton PE, Cooke CA, Bourassa S, et al (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Current Biology 7:901–904

Yang F, Carter NP, Shi L, Ferguson-Smith MA (1995) A comparative study of karyotypes of muntjacs by chromosome painting. Chromosoma 103:642–652

Zhang W, Friebe B, Gill BS, Jiang J (2010) Centromere inactivation and epigenetic modifications of a plant chromosome with three functional centromeres. Chromosoma 119:553–563

Zhang X, Clarenz O, Cokus S, et al (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in arabidopsis. Plos Biol 5:e129

Chapter 2: Identification of the CENP-A protein and associated centromeric sequence in the threespine stickleback

This chapter is reformatted from: Cech JN and Peichel CL (2015) Identification of the centromeric repeat in the threespine stickleback fish (*Gasterosteus aculeatus*). *Chromosome Research*, 23(4): 767-779.

Introduction

Proper segregation of chromosomes is essential for the faithful and equal separation of each chromosome during both mitosis and meiosis. Mis-segregation of chromosomes can lead to aneuploidy, which is often associated with cancer, miscarriage, and birth defects such as viable trisomies (Kops et al. 2005; Morales et al. 2007; Hunt and Hassold 2010; Lister et al. 2010; Revenkova et al. 2010; Gordon et al. 2012; Ricke and van Deursen 2013). Physical separation of chromosomes occurs by the attachment of microtubule spindle fibers via the kinetochore to a primary constriction on each chromosome, called the centromere. Because of the vital role centromeres play in cell division, each chromosome must contain only a single functional centromere. Functional centromeres are defined by the presence of a centromeric histone variant known as either CenH3 or CENP-A (Palmer et al. 1991; Sullivan et al. 1994). CENP-A replaces histone H3 at functional centromeres and localizes to the primary constriction on metaphase chromosome spreads (Earnshaw and Rothfield 1985; Palmer et al. 1989; Sullivan and Schwartz 1995; Warburton et al. 1997; Amor et al. 2004).

While the presence of CENP-A at functional centromeres is almost entirely conserved among eukaryotes (Henikoff et al. 2001; Malik and Henikoff 2009; Drinnenberg et al. 2014), the genetic sequence at centromeres varies dramatically between species (Henikoff et al. 2001; Alkan et al. 2011; Melters et al. 2013) and even

among chromosomes of the same species or closely related species (Shang et al. 2010; Tek et al. 2010; Piras et al. 2010; Gong et al. 2012). Despite variation at the sequence level, most centromere sequences have conserved characteristics, such as being highly repetitive and AT-rich (Melters et al. 2013). In addition, many mammalian centromere repeats contain a conserved CENP-B box, a 17 bp DNA motif that recruits the CENP-B protein, thought to aid in de novo centromere formation (Masumoto et al. 1989; Ohzeki et al. 2002; Alkan et al. 2011). Previous methods used to identify centromere sequences include isolation of satellite DNA, restriction digest approaches, and BAC (bacterial artificial chromosome) cloning (Maio 1971; Manuelidis 1978; Haaf et al. 1993; Garrido-Ramos et al. 1994; Crollius et al. 2000; Edwards and Murray 2005; Tek et al. 2010; Shang et al. 2010). However, these approaches can be limited by the restriction enzymes used, by biases against cloning repetitive DNA, and assumptions that the most repetitive sequences in the genome will be the centromere. Chromatin immunoprecipitation (ChIP) with a CENP-A antibody followed by cloning of the ChIP DNA has recently been used to both confirm putative centromere sequences (Zhong 2002; Nagaki et al. 2003; Nagaki et al. 2004; Nagaki and Murata 2005; Houben et al. 2007), and to identify novel repeats (Lee et al. 2005; Edwards and Murray 2005; Nagaki et al. 2008; Tek et al. 2010; Shang et al. 2010; Alkan et al. 2011). However, CENP-A ChIP cloning is limited both by the number of clones that can be sequenced and biases against cloning repetitive DNA. Chromatin immunoprecipitation with a CENP-A antibody followed by high throughput sequencing (ChIP-seq) is the most unbiased approach to identify functional centromeric associated DNA (Henikoff et al. 2015), but it has not yet been extensively used (Gong et al. 2012; Henikoff et al. 2015).

Here we sought to identify the centromere sequence in threespine stickleback fish (*Gasterosteus aculeatus*), an emerging model organism used to study the genetic, genomic, and molecular basis for evolution (Kingsley and Peichel 2007). The threespine stickleback has an assembled genome (Jones et al. 2012), yet the centromere sequence is still unknown. Gaps in the genome assembly correspond with the cytological constriction on most chromosomes (Urton et al. 2011), suggesting that the stickleback centromere is comprised of repetitive sequences (Henikoff 2002; Rudd and Willard 2004). To identify the threespine stickleback centromere sequence, we first identified the complete threespine stickleback CENP-A coding sequence, and then performed ChIP-seq using a threespine stickleback specific CENP-A antibody. We confirmed that this 186 bp, AT-rich sequence is centromeric using both fluorescence in situ hybridization (FISH) and CENP-A immunofluorescence coupled with FISH (IF-FISH).

Materials and Methods

Fish use and care

Three populations of *G. aculeatus* were used in this study: wild caught fish from Lake Union (Seattle, Washington, USA), laboratory-reared Pacific Ocean fish, and laboratory-reared Japan Sea fish. Both of these laboratory populations were derived from wild-caught fish collected in Akkeshi on Hokkaido Island, Japan (Kitano and Mori 2007; Kitano et al. 2009). Fish were housed in 29-gallon aquarium tanks in summer lighting conditions (16 h light, 8 h dark) at approximately 16°C in 0.35% saltwater (3.5 g/L Instant Ocean salt (Spectrum Brands, USA); 0.4 ml/l sodium bicarbonate). Fish were fed with live brine shrimp nauplii and frozen *Mysis* shrimp twice daily. All institutional and

national guidelines for the care and use of laboratory animals were followed, and all procedures were approved by the Fred Hutchinson Cancer Research Center Institutional Animal Care and Use Committee (protocol 1575). Fish were caught from Lake Union with permission of the Washington Department of Fish and Wildlife (Scientific Collection permits 12-057, 13-039, 14-065, and 15-033).

Determining the Cenpa coding sequence

To find the putative *G. aculeatus Cenpa* gene, the *Danio rerio* CENP-A protein sequence was used to search for a homologous sequence in the threespine stickleback genome assembly (Ensembl BROAD S1; Feb 2006) using blastp. The putative threespine stickleback CENP-A protein (ENSGACP00000024534) is a product of the gene (ENSGACG00000018561.1). However, this predicted gene was missing a stop codon. To determine the full *Cenpa* coding sequence, RNA was extracted using Trizol (Life Technologies, USA) from a Japan Sea male stickleback kidney. An Invitrogen 3' RACE (rapid amplification of cDNA ends) kit (Life Technologies, USA) was used to make cDNA. Internal primer JC06 (5' GAGAGGTGTGCCAGAGCTTCTC 3') was used in conjunction with the 3' RACE kit primer AUAP to amplify and sequence the entire 3' end of the *Cenpa* gene. This coding sequence was used to design primers (5' ATGCGTCACAATTCATCTACC 3' and 5' TTACAGGTTGTCCACCCC 3') to amplify the entire cDNA. To determine whether the Japan Sea and Pacific Ocean *Cenpa* cDNA sequences are the same, Pacific Ocean male RNA was extracted from liver tissue using Trizol, cDNA was synthesized using the SuperScript III First-Strand Synthesis

System kit (Life Technologies, USA), and the *Cenpa* cDNA was Sanger sequenced following polymerase chain reaction (PCR) amplification with the primers above.

Antibody design

A *G. aculeatus* CENP-A specific antibody was designed to amino acids 1-22 (MRHNSSTSRRKGKTPQHRPPLA) of the N-terminal end of the threespine stickleback CENP-A protein (Figure 2-1b). The rabbit IgG affinity purified polyclonal antibody was produced by Covance Research Products (USA).

Western blot

50 mg of Pacific Ocean female kidney and liver tissue was dissected in 400 µl RIPA (150 mM sodium chloride, 1% NP-40, 0.5 % sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 50 µM Tris pH 8.0, 0.3 mM phenylmethanesulfonylfluoride (PMSF)). Using a pellet pestle, tissue was homogenized in 400 µl RIPA, 0.3 mM PMSF on ice. The pestle was rinsed with an additional 600 µl RIPA with PMSF, then the sample was rotated 2 h at 4^{o}C. Cells were spun for 20 min at 12,000 rpm and 4^{o}C. The pellet was resuspended in 30 µl of 2% SDS, and 6 µl of 6X Laemmli buffer was added. The samples were boiled at 95^{o}C for 5 min and then run on a 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel in running buffer (25 mM Tris, 0.192 M glycine, 0.2% SDS). Proteins were transferred from the gel to a polyvinylidene difluoride membrane for 1 h at 14 volts in transfer buffer (25 mM Tris, 192 mM glycine, 10% methanol). The membrane was blocked in I-BLOCK (Applied Biosystems, USA) for 1 h. The membrane was incubated with the primary antibody (rabbit, anti-stickleback

CENP-A) at 1:1000 in I-BLOCK plus 5% bovine serum albumin (BSA) on a shaking nutator overnight. The membrane was washed 3x5 min with phosphate-buffered saline Tween-20 (PBST), then incubated with alkaline phosphatase conjugated goat, anti-rabbit secondary antibody (Applied Biosystems, USA) at 1:10,000 in I-BLOCK for 1 h. The membrane was washed 2x5 min with PBST, then 2x5 min in assay buffer (20 mM Tris pH 9.8, 1 mM magnesium chloride). A few drops of CDP-star Chemiluminescent substrate (Sigma-Aldrich, USA) was added to membrane. Film was exposed for 5 sec, then developed.

Immunoflourescence on metaphase spreads and interphase nuclei

Immunofluorescence was conducted using a protocol adapted from (Blower et al. 2002). For metaphase spreads, Lake Union male spleen tissue was homogenized in 1 ml of 0.56% potassium chloride (KCl) with a glass dounce. For interphase nuclei, 15 Pacific Ocean embryos at 48 days post fertilization (dpf) were homogenized in 10 ml of 0.56% KCl with a glass dounce. Cells were incubated for 45 min on ice to swell, and 350 μl (metaphase spreads) or 250 μl (interphase nuclei) of the cell suspension was spun for 10 min at 1950 rpm onto Fisherfinest Premium Superfrost microscope slides through a Fisher single cytology funnel (Fisher Scientific, USA) using a Cytospin 3 (Shandon, USA). Slides were immediately placed in ice-cold methanol for 20 min followed by a 10 min room temperature incubation in 4% formaldehyde in phosphate-buffered saline (PBS). Slides were washed 10 min in PBST (PBS, 0.05% Tween-20) and then incubated with block solution (PBST, 1% BSA, 0.02% sodium azide) for 20 min at room temperature. Primary antibody (rabbit, anti-stickleback CENP-A) was added at 1:300 in

block solution for 4 h at room temperature. Slides were washed 3x5 min in PBST at room temperature. A secondary Alexa Fluor 488 chicken, anti-rabbit IgG antibody (Life Technologies, USA) was added at 1:500 in block solution and incubated overnight at 4°C. Slides were washed 3x5 min in PBST, then counterstained with vectashield 4',6'-Diamidino-2-phenlyindole (DAPI) stain (Vector Laboratories, USA) for 10 min before imaging as described in the *Microscopy* section.

CENP-A chromatin immunoprecipitation (ChIP)

Chromatin preparation from tissue: The ChIP protocol was performed using the SimpleChIP Plus Enzymatic Chromatin IP kit (Magnetic Beads) (Cell Signaling Technology, USA). Approximately 0.1 g of kidney and liver tissue was dissected from two Pacific Ocean females for two independent samples. The tissue was placed into 2 ml of 1X protease inhibitor cocktail (PIC) buffer. The tissue was dounced 3 times with a glass dounce, and the dounce was rinsed with an additional 2 ml of 1X PIC buffer. 4 ml of the tissue suspension was incubated with 180 µl of 37% formaldehyde for 20 min at room temperature on a nutator. From this point, the Simple ChIP protocol was followed with the following modifications: cells were incubated for 30 min at 37°C with 1.6 µl of micrococcal nuclease (MNase) to create mono and dinucleosomes. MNase treatment was stopped with addition of 40 µl 0.5 M ethylenediaminetetraacetic acid (EDTA). Cells were spun at 13,000 rpm for 1 min at 4°C and resuspended in 400 µl 1X ChIP buffer. Nuclear membranes were disrupted with 3x30 sec sonication using a probe sonicator on high, with a 30 sec incubation on wet ice between pulses. Lysates were clarified by spinning at

10,000 rpm for 10 min at 4°C. 30 μl of the supernatant was taken for chromatin analysis and quantification, and the rest was stored at -80°C until immunopreciptation (IP).

Immunoprecipitation: 2 x 10 μg of chromatin was diluted in 500 μl 1X ChIP buffer. For each sample, 10 μl was taken and stored at -20°C for use as the input samples. For each 10 μg IP, chromatin was incubated with 4.64 μg of rabbit, anti-stickleback CENP-A antibody in 1X ChIP buffer overnight at 4°C with rotation. 30 μl of magnetic beads were added to each 10 μg IP. Following a 2 h incubation at 4°C with rotation, experimental methods were carried out following the Simple ChiP kit instructions. Input and IP DNA was purified using a MinElute PCR purification kit (Qiagen, USA). Two IP and two input samples were each eluted in 11 μl EB buffer. Paired end 150 bp sequencing was performed on the Illumina HiSeq 2500 (Illumina, USA) at the Fred Hutchinson Cancer Research Center Genomics Shared Resource. Fastq files containing the raw sequencing data have been deposited to the NCBI Sequence Read Archive (study accession SRP063504).

ChIP analysis: CENP-A ChIP data was analyzed using the pipeline described by (Henikoff et al. 2015). Paired reads were merged using SeqPrep (https://github.com/jstjohn/SeqPrep). We then discarded any CENP-A ChIP reads that aligned to the BROADS1 assembly of the *G. aculeatus* genome (Jones et al. 2012), using bwa v0.7.12 (Li and Durbin 2009). The remaining merged pairs (9,182,498 from Pacific Ocean female 1 (POF1) and 10,208,973 from Pacific Ocean female 2 (POF2)) were run through CD-HIT-EST (Li and Godzik 2006, Fu et al. 2012) to create a set of the most

abundant IP clusters, each cluster sharing at least 90% sequence identity. The longest read for each cluster was used a reference sequence for that cluster. The clusters were ranked by the number of reads, and the top 500 cluster reference sequences were identified. For each cluster reference sequence, we first counted the number of reads in both the total IP and input samples (all mapped and unmapped reads), and then normalized these counts by calculating the reads per million (RPM) for both the IP and input samples. We calculated the fold enrichment of each cluster reference sequence in the IP relative to input by dividing the IP RPM by the input RPM. The top 500 most abundant cluster references were then ranked by order of fold enrichment. Sequences with a fold enrichment of less than one were discarded. The total average fold enrichment for all of the top enriched clusters combined was calculated by dividing the total IP RPM by the total input RPM.

The cluster reference sequences determined from above were aligned to the most abundant cluster sequence (Figure 2-2). The Geneious (Biomatters, New Zealand) sequence global alignment program (parameter: cost matrix 65% similarity, gap open penalty = 12, gap extension penalty = 3) was used for alignments. From this alignment, a consensus sequence repeating unit was determined using the >50% identity "strict" parameter in Geneious. A consensus repeating unit was determined for each independent IP sample. The two consensus repeats were then aligned to create a single consensus centromeric repeat sequence (CEN) (Figure 2-3).

To find the percentage of reads from each ChIP experiment that contained sequences with homology to the CEN, we used BWA-MEM (Li 2013) to align all reads from either the input or the IP to the consensus repeat sequence. Only reads with a score

>30 were counted where the score is dependent on: B INT ([4]), O INT [,INT] ([6,6]), E INT [,INT] ('{O}+{E}*k'[1,1], and L INT [,INT] ([5,5]).

To find assembled scaffolds that contain sequences homologous to the centromeric repeat sequence, we conducted a BLAST search for the CEN sequence in the threespine stickleback genome assembly (Ensembl BROAD S1; Feb 2006). We identified sequences with homology to the CEN sequence in scaffolds from regions of the genome that were not assigned to chromosome assemblies and in scaffolds at the edges of gaps in chromosome assemblies that correspond to the putative position of centromeres as determined using the p:q chromosome arm length ratios (Urton et al. 2011). Up to 5 kb of sequence data from each region was extracted, and the consensus centromeric repeat was aligned to these sequences in Geneious (Biomatters, New Zealand) to identify tandem repeats and to determine percent identity between the consensus centromere repeat and the repeats present in the genome assembly.

Fluorescence in situ hybridization (FISH) on metaphase spreads and interphase nuclei

CEN FISH probe: PCR primers JC103 (5'GGTGCTAGATTTAGGAAAACA 3') and JC106 (5'GTGCATTCATGACTTTTAAGG 3') were used to amplify the threespine stickleback centromeric repeat from genomic DNA. The PCR product appeared as a ladder. The entire ladder was extracted using the QIAquick gel extraction kit (Qiagen, USA). The purified PCR product was cloned into the PCR2.1 vector (Life Technologies, USA). The clone used as a template to make the FISH probe contained 1.5 copies of the centromeric repeat. This clone was amplified using primers JC103 and JC106 with PCR fluorescein labeling mix (Roche Applied Science, Switzerland) to make a fluorescein-12-

dUTP labeled probe of 288 bp. The probe was lyophilized at 55°C for 3 hours, then resuspended in 10 μl hybridization buffer (50% formamide, 2X saline-sodium citrate (SSC), 10% dextran sulfate, 0.2 mM Ethylenediaminetetraacetic acid (EDTA), 2 mM Tris pH 8.0). The probe was denatured at 72°C for 5 min and then stored at 37°C until use.

Sex chromosome FISH probe: The *G. aculeatus* BAC clone CHORI-213 101E08 was used as a FISH probe to distinguish the X and Y chromosomes. This BAC clone hybridizes to the end of the p arm of the Ychromosome and to the middle of the p arm of the X chromosome (Ross and Peichel 2008). The BAC FISH probe was labeled with Alexa-568 following the protocol in (Urton et al. 2011).

Metaphase spreads and interphase nuclei: Wild-caught fish from Lake Union were injected with 10 μl of 1% colchicine in PBS for 16 h to arrest mitotic cells. Spleen and liver tissue was dissected and homogenized in 2 ml of 0.56% KCl using a glass dounce. Metaphase spreads and interphase nuclei were prepared following the protocol described in (Ross and Peichel 2008).

Hybridization: Slides were washed in 2X SSC for 5 min, then in denature solution (2X SSC, 70% formamide) for 2 min in a 72°C water bath. Slides were dehydrated in 70% cold ethanol for 2 min, followed by a room temperature dehydration series for 2 min each in 85%, 90%, and 100% ethanol. After the slides were air dried for 5 min, 10 μl of either the CEN probe, the sex chromosome probe, or both was added. Then, the slides were

covered with a coverslip, sealed with rubber cement, and hybridized at 37°C overnight in a humid chamber. Slides were washed in a 42°C water bath 3x5 min in wash solution 1 (2X SSC, 50% formamide), followed by 3x5 min in 2X SSC. Slides were air dried 5 min, then counterstained and mounted with vectashield DAPI stain. Images were taken as described in the *Microscopy* section.

Immunofluorescence-FISH

For IF-FISH, 15 Lake Union embryos at 48 dpf were dounced in 10 ml 0.56% KCl and 0.1% Tween-20 using a glass dounce. Metaphase and interphase nuclei slides were prepared following the protocol in the *Immunofluorescence* section, with the following modifications: primary rabbit, anti-stickleback CENP-A antibody was diluted 1:100 in block, and 100 µl was added to the slide, which was covered with a glass coverslip and incubated in a humid chamber overnight at 4°C. Slides were washed 3x5 min with PBST. The secondary goat, anti-rabbit 568 antibody (Invitrogen, USA) was diluted 1:600 in block, and 100 µl was added to the slide and incubated 4 h at room temperature in the dark. Slides were washed 3x5 min in PBST, then post fixed with 4% formaldehyde in PBS for 10 min at room temperature. 10 µl of PCR labeled CEN probe was lyophilized at 55°C for 4 h and resuspended in 10 µl of hybridization buffer (50% formamide, 2X SSC, 10% dextran sulfate, 20% TE pH 8). The probe was denatured at 72°C for 5 min in a water bath and then put at 37°C until use. Hybridization with the CEN probe was performed following the protocol in the *Hybridization* section.

Microscopy

Slides were imaged using a Nikon Eclipse 80i microscope (Nikon, Japan) with an automated filter turret (Chroma filters 31000v2 (DAPI), 41001 (FITC), and 41004 (Texas Red); Chroma, USA). Images were taken using the 100X objective, using the Photometrics CoolSNAP ES2 camera (Photometrics, USA). NIS Elements imaging software (BR 3.00, SP7, Hotfix8, Build 548, Nikon, Japan) was used to pseudo-color the images. For CENP-A immunofluorescence images, the Alexa Fluor-488 labeled antibody and DAPI were pseudo-colored purple and grey, respectively. For FISH images, the Alexa Fluor-568 labeled BAC probe and DAPI were pseudo-colored purple and grey, respectively; the fluorescein labeled CEN probe is green. For IF-FISH images, the Alexa Flour-568 labeled CENP-A antibody, and DAPI were pseudo-colored purple and grey, respectively.

Results

Identification of the threespine stickleback CENP-A coding sequence

Using 3' RACE, PCR, and cDNA sequencing, we identified the entire coding sequence of the threespine stickleback CENP-A gene, which is 447 bp and encodes a protein of 148 amino acids (Figure 2-1a). The amino acid sequences between Pacific Ocean and Japan Sea populations differ by a single amino acid at position 36 (GenBank accession numbers KT321854 (Pacific Ocean), KT321855 (Japan Sea)). The Pacific Ocean sequence is used as the reference here. Like other known CenH3-like proteins, the threespine stickleback CENP-A protein has an N-terminal tail that is highly diverged from the threespine stickleback H3 protein (Figure 2-1b). The N-terminal tail is also

highly diverged among CENP-A proteins from different fish species, while the C-terminal histone fold domain is more conserved (Fountain and Kral 2011).

Threespine stickleback CENP-A is centromere specific

To determine whether the threespine stickleback CENP-A protein is a centromeric histone variant, we designed an antibody that would recognize the N-terminal end of the CENP-A, but not the H3, protein (Figure 2-1b). A Western blot on protein lysate from threespine stickleback tissue shows that the CENP-A antibody is specific to a prominent band of around 17 kD, consistent with its predicted atomic weight of 16.7 kD (Figure 2-1c). The antibody hybridizes to punctate spots on interphase nuclei (Figure 2-4a) and to the primary constriction of all chromosomes on metaphase spreads (Figure 2-4b, c). Taken together, these data suggest that this is a bona fide centromeric histone.

CENP-A associated sequences in threespine stickleback

To identify the centromere sequence, we performed CENP-A ChIP-seq in two independent female fish from the Pacific Ocean population. Using the cluster analysis pipeline described in (Henikoff et al. 2015), we identified the 500 most abundant sequences in each IP sample and then determined that the majority of these sequences (320 in POF1 and 299 in POF2) were enriched in the IP sequences relative to input sequences (Table 2-1). After aligning these enriched sequences (Figure 2-2), we found a single consensus repeating unit for each individual IP sample. These two consensus repeats are 99.5% identical and were aligned to create a single consensus repeat (Figure 2-3). The consensus repeat is 186 bp with an AT-content of 62.9% (Figure 2-5a), and was

17.04 or 17.68-fold enriched in the IP relative to input in two independent ChIP-seq experiments (Table 2-1). This repeat made up 4.43 and 4.29% of total reads from each input sample, highlighting its overall abundance in the genome. Like many other centromeric repeats, this consensus repeat is AT-rich (Melters et al. 2013), contains a putative centromere protein B (CENP-B) box (Figure 2-5b) (Masumoto et al. 2004; Edwards and Murray 2005; Henikoff et al. 2015), and has a sequence length consistent with wrapping around a single nucleosome (Willard 1991; Shelby et al. 1997; Henikoff et al. 2001).

Nearly all of the sequences (316/320 from POF1; 289/299 from POF2) that were enriched in the CENP-A IP relative to the input aligned to the consensus repeat (Table 2-1). The few enriched sequences that did not align to the consensus repeat (4 from POF1; 10 from POF2) did not align to each other and were low in abundance, with less than 150 total reads in the IP and 84 total reads in the input. It is possible that these low abundance sequences are bound by CENP-A and are interspersed among long stretches of the consensus tandem repeat. However, our data suggests that the majority of sequences bound to CENP-A in the threespine stickleback are variants of the consensus repeat that we identified (Table 2-1).

The CENP-A associated sequence is centromere specific

To determine whether the sequence is the bona fide centromere, we generated a fluorescently labeled probe of the consensus repeat sequence identified by ChIP-seq and found that it hybridizes to distinct loci in interphase nuclei (Figure 2-6a), as well as to the constriction on 41 of 42 metaphase chromosomes in male threespine sticklebacks (Figure

2-6b, c). IF-FISH also shows that the centromere repeat co-localizes with CENP-A in both interphase nuclei (Figure 2-7a), as well as on metaphase spreads (Figure 2-7b, c). All of these results suggest that the 186 bp repeat, now referred to as GacCEN (GenBank accession number KT321856), is the centromeric repeat in threespine stickleback fish.

The GacCEN probe shows weak hybridization to the Y chromosome

In mammals, the Y chromosome centromere can differ from the centromere on the other chromosomes. For example, the house mouse Y centromere is comprised of a novel satellite repeat (Pertile et al. 2009), and the human Y centromere is a divergent alpha satellite repeat (Wolfe et al. 1985; Miga et al. 2014). Thus, we hypothesized that the single chromosome that did not hybridize to the GacCEN probe in males was the Y (Figure 2-8b). Indeed, when we performed FISH with the GacCEN probe and a BAC probe that distinguishes the X and Y chromosomes (Ross and Peichel 2008), we found that the Y chromosome centromere shows very weak hybridization to this centromeric repeat, while the X chromosome shows strong hybridization (Figure 2-8).

GacCEN repeat shows a decrease in percent identity outside of the core centromere

We searched for sequences with homology to the consensus centromeric repeat in the current stickleback genome assembly (Jones et al. 2012; Ensembl BROADS1) to see if we could uncover any large scaffolds containing the GacCEN repeat. We found multiple copies of a similar repeat on four scaffolds that did not map to chromosome assemblies and at the edges of ten different scaffolds from nine assembled chromosomes, corresponding to the locations of centromeric constrictions in metaphase spreads (Urton

et al. 2011). As predicted (Smith 1976; Schueler et al. 2001; Henikoff 2002; Shepelev et al. 2009; Henikoff et al. 2015), the repeats found of the edge of the assembled contigs show a decrease in percent identity to the GacCEN as they move away from the core centromere (Figure 2-9). None of the GacCEN containing contigs appear to have any higher order repeat (HOR) structure. However, until longer centromere-containing reads are obtained, it will be difficult to determine whether stickleback centromeres in fact have a HOR structure.

Discussion

By identifying the complete coding sequence of the threespine stickleback CENP-A protein and designing a species-specific antibody, we were able to perform an unbiased CENP-A ChIP-seq experiment to identify the threespine stickleback centromere repeat. We validated this repeat as centromeric by performing FISH, as well as IF-FISH with the CENP-A antibody on metaphase spreads. This repeat shares similar characteristics to other centromere repeats in its size, AT content, and putative CENP-B box (Alkan et al. 2011, Melters et al. 2013). It should be noted that the putative *G. aculeatus* CENP-B box shows weak conservation to the evolutionarily conserved human CENP-B domains (Alkan et al. 2011), yet shows the same number of conserved bases as the putative dog, platypus, and *Xenopus* CENP-B boxes (Figure 2-5b). The lack of a fully conserved CENP-B box was consistent across all individual cluster repeats, suggesting there are no small subsets of active CENP-B box containing repeats. Interestingly, we also find no evidence for a CENP-B gene in the *G. aculeatus* genome assembly. Although it is not

clear whether the CENP-B box in the GacCEN is functional, it is interesting that the repeat shares similar sequence characteristics with other known centromeric repeats.

A previous study (Melters et al. 2013) used a bioinformatic pipeline to identify candidate centromere sequences in 282 species, including two putative centromere sequences of 313 bp and 186 bp in threespine stickleback (Table S4 in Melters et al. 2013). While they could not identify the source of the discordance between the two potential repeats, they assumed that the 313 bp sequence was the centromere because of its higher abundance in the genome. However, we tested this repeat using FISH and found that it does not hybridize to the centromere on metaphase chromosomes (data not shown). In fact, the 186 bp repeat that they identified is 93.0% similar to the GacCEN repeat identified in our study (D. Melters personal communication, July 7, 2015). These results underscore the importance of FISH and/or CENP-A ChIP validation to discern between multiple candidate sequences identified by other methods.

We also previously attempted to identify the *G. aculeatus* centromere sequence using methods that do not rely on a species-specific CENP-A antibody. We used the RepeatNet program (Alkan et al. 2011), which identifies the most abundant sequences in a sequenced genome. In addition, we used a restriction digest approach to identify abundant repeats in the genome. While both computational and restriction digest methods are easy to use and cheaper than ChIP-seq, they generated a large list of potential repeats. To narrow down our list of putative sequences, we assumed, as is common, that the centromeric repeat would not map to the assembled regions of the genome, and that it would be one of the most abundant repeats in the unassembled genome sequences (Alkan et al. 2011; Melters et al.2013). Based on these criteria, we made a list of candidates to

test by FISH, but none hybridized to the centromere. However, after performing the ChIP-seq experiment, we realized that the GacCEN repeat was actually identified using both the computational and the restriction digest methods. Because the edges of many genome assembly scaffold edges contained a few, more diverged, copies of this repeat, our analysis showed that the putative sequences mapped to distinct chromosomes. Furthermore, the GacCEN repeat was not as abundant as other repeats we identified. For these two reasons, the GacCEN repeat sequence was not among the first to be tested. Our experience highlights that other approaches can be used to identify centromeric repeats in organisms without a known CENP-A antibody or an available epitope-tagged CENP-A protein. However, CENP-A ChIP-seq is still the most unbiased way to identify or validate a centromeric repeat.

Our work also demonstrates that it is possible in some cases to identify remnants of centromere repeats on the edges of genome assembly gaps. We found some repeats on the edges of assembled scaffolds that were slightly more divergent than the core centromere sequences identified by ChIP-seq (Figure 2-9). This is consistent with what has been seen in human CENP-A ChIP-seq experiments in which the sequence homology among α-satellite repeats decreases with distance from the core functional centromere (Henikoff et al. 2015). Unequal crossing over at nearly identical repeats is thought to lead to the homogenization of the core centromere, with mutation leading to divergence and unique differences in repeats outside of the core centromere (Smith 1976; Schueler et al. 2001; Henikoff 2002; Shepelev et al. 2009; Henikoff et al. 2015). These differences allow for the assembly of the centromeric repeats. However, Henikoff et al. 2015 showed that more diverged repeats may lose their centromere function, implying that the core,

unassembled centromere repeats make up functional centromeres. Our initial filtering step to remove all CENP-A ChIP reads that mapped to the assembled genome actually discarded the repeats we later found on the edges of genome assembly gaps. While these similar repeat may still have centromeric function, by discarding any repeats that mapped to the genome, we ensured that our methods enriched for the core, functional centromere sequences.

Finally, while we have determined that 41 out of the 42 threespine stickleback chromosomes show hybridization with the GacCEN probe in males, the Y centromere appears to show weak hybridization. This could be due to either a different or divergent Y specific centromere repeat (Pertile et al. 2009; Miga et al. 2014), or to a decrease in the number of repeats on the Y centromere array. Future work will use a similar unbiased ChIP-seq approach in both males and females to identify the Y-chromosome centromere repeat.

Table 2-1. Summary of CENP-A ChIP seq data from two independent Pacific Ocean females.

	Pacific Ocean female 1	**Pacific Ocean female 2**
Number enriched clusters in 500 most abundant	320	299
Number enriched clusters with GacCEN repeat	316	289
Average fold enrichment of IP/input	17.04	17.68
Range of fold enrichment across clusters	1.25 - 87.60	1.01 - 134.00
Total normalized IP RPM for enriched GacCEN containing clusters	307,600.96	308,096.65
Total normalized input RPM for enriched GacCEN containing clusters	18,046.51	17,426.30
Total IP reads	28,309,375	34,297,458
Total IP reads that map to GacCEN (% total)	20,245,271 (71.51%)	23,695,808 (69.09%)
Total input reads	26,102,993	30,137,027
Total input reads that map to GacGEN (% total)	1,155,758 (4.43%)	1,291,918 (4.29%)

Chapter 2 Figures:

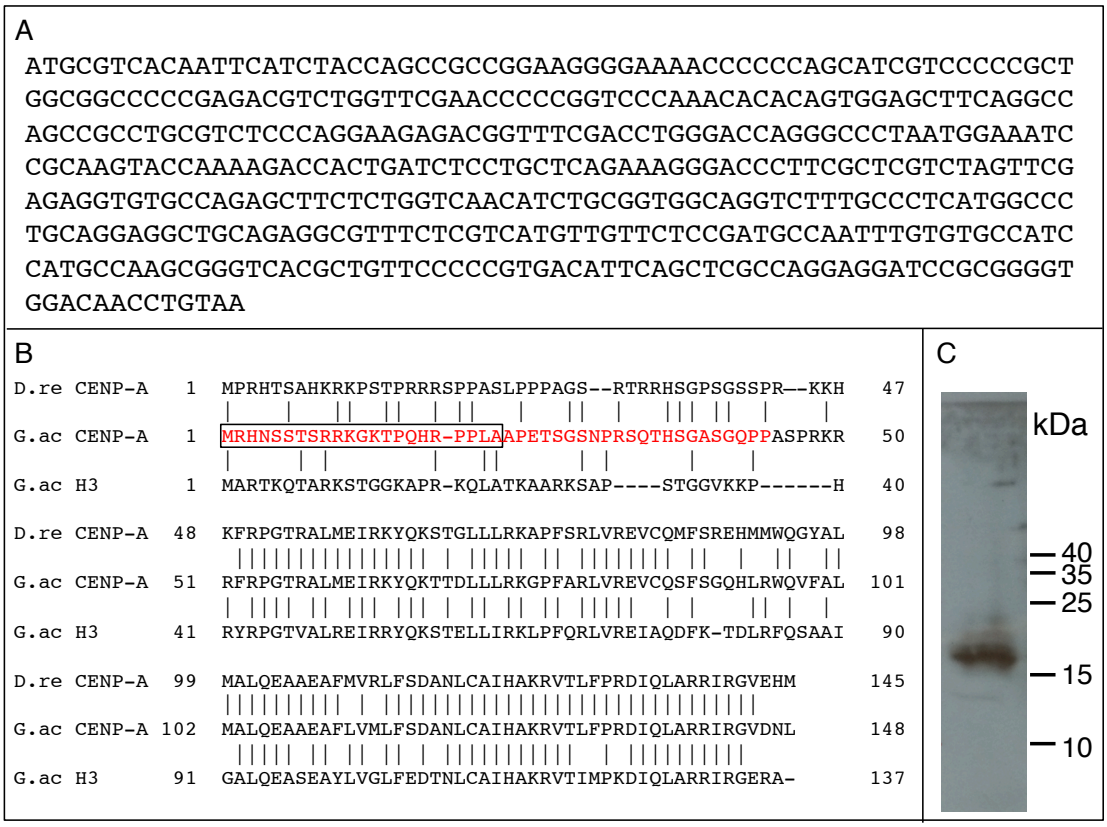


Figure 2-1. Threespine stickleback Pacific Ocean CENP-A cDNA, protein, and antibody. **a** Full-length cDNA sequence for the threespine stickleback CENP-A protein. This sequence has been deposited in GenBank (accession number KT321854). **b** Alignment between the *Danio rerio* CENP-A protein (D.re CENP-A), the threespine stickleback CENP-A protein (G.ac CENP-A), and the threespine stickleback H3 protein (G.ac H3) reveals extensive divergence in the N-terminal tail (*red*). The threespine stickleback H3 protein sequence is a product of the gene ENSGACG00000005779 (Ensembl BROAD S1; Feb 2006). The antibody was raised against the peptide sequence highlighted with a box. **c** Western blot of the CENP-A antibody on protein lysates from threespine stickleback liver and kidney.

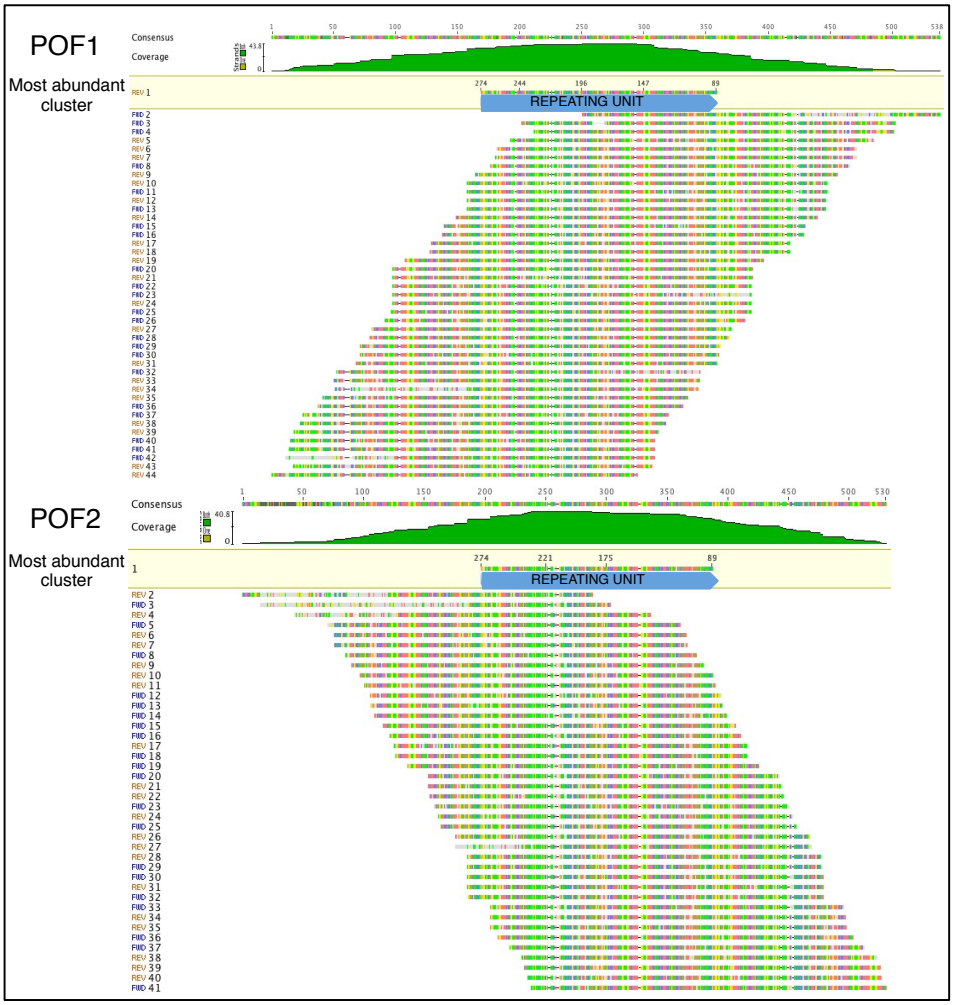


Figure 2-2. Alignment of the enriched repeats found in the most abundant clusters from both ChIP-seq samples. All repeats that were enriched in the IP relative to input from the 500 most abundant clusters in each sample were aligned in Geneious (Biomatters, New Zealand), but only the 44 repeats (Pacific Ocean female 1 (POF1)) and 41 repeats (Pacific Ocean female 2 (POF2)) that were enriched in the IP relative to input from the 50 most abundant clusters are shown here for simplicity. The blue box indicates the 186 bp repeating unit found in all enriched clusters.

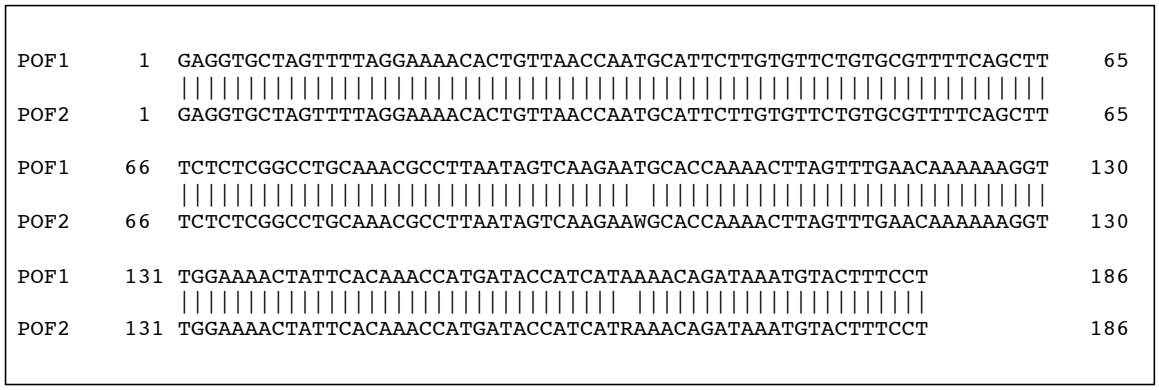


Figure 2-3. Alignment of the consensus repeating unit from each independent ChIP-seq sample. The consensus repeat sequences from the Pacific Ocean female 1 (POF1) and Pacific Ocean female 2 (POF2) were aligned to create the final consensus centromere repeat sequence shown in Figure 2-5. Nucleotide ambiguities: R = A or G; W = A or T.

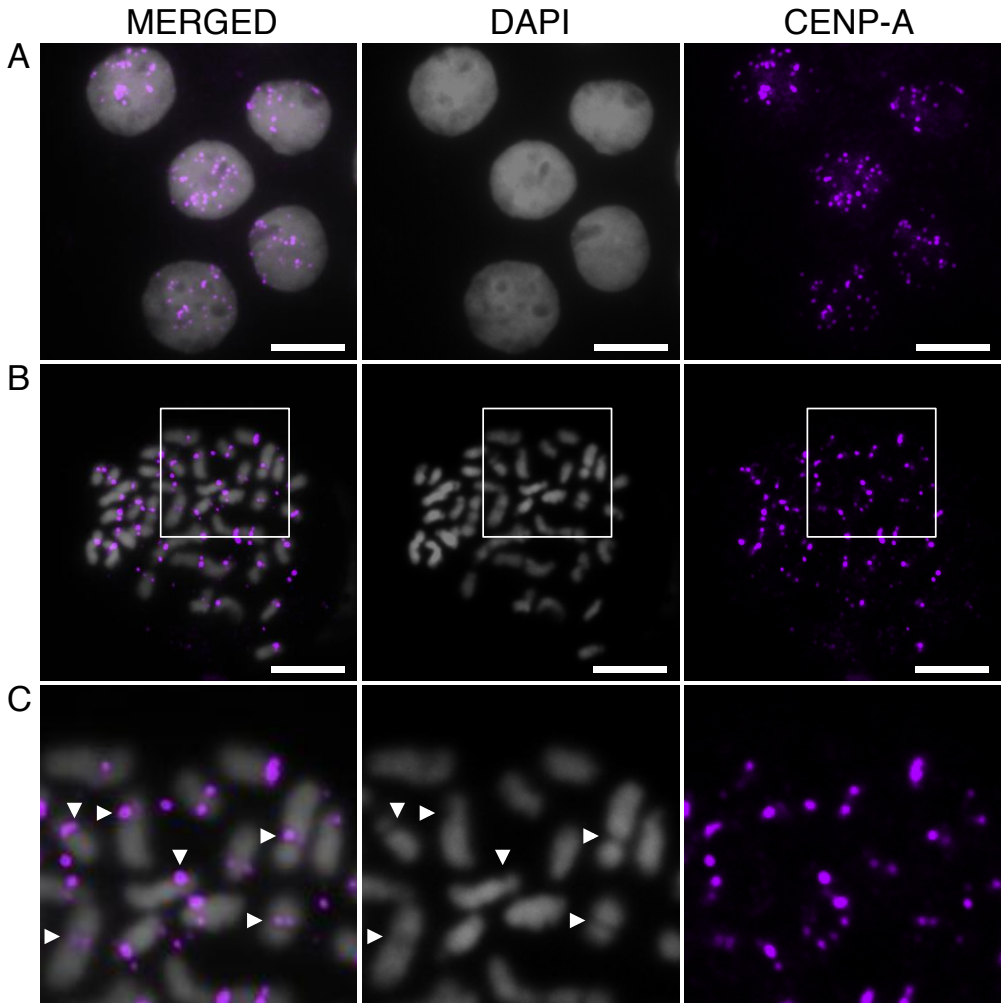


Figure 2-4. The threespine stickleback CENP-A antibody localizes to the centromere. Immunofluorescence reveals that the CENP-A antibody (*purple*) hybridizes to **a** distinct puncta in interohase nuclei from Pacific Ocean embryos, **b** a single region on each chromosome in a metaphase spread from a Lake Union male, and **c** the primary constriction (*arrowheads*) on each chromosome in a higher magnification view of the boxed region in panel (**b**). *Scale bar, 5μm*

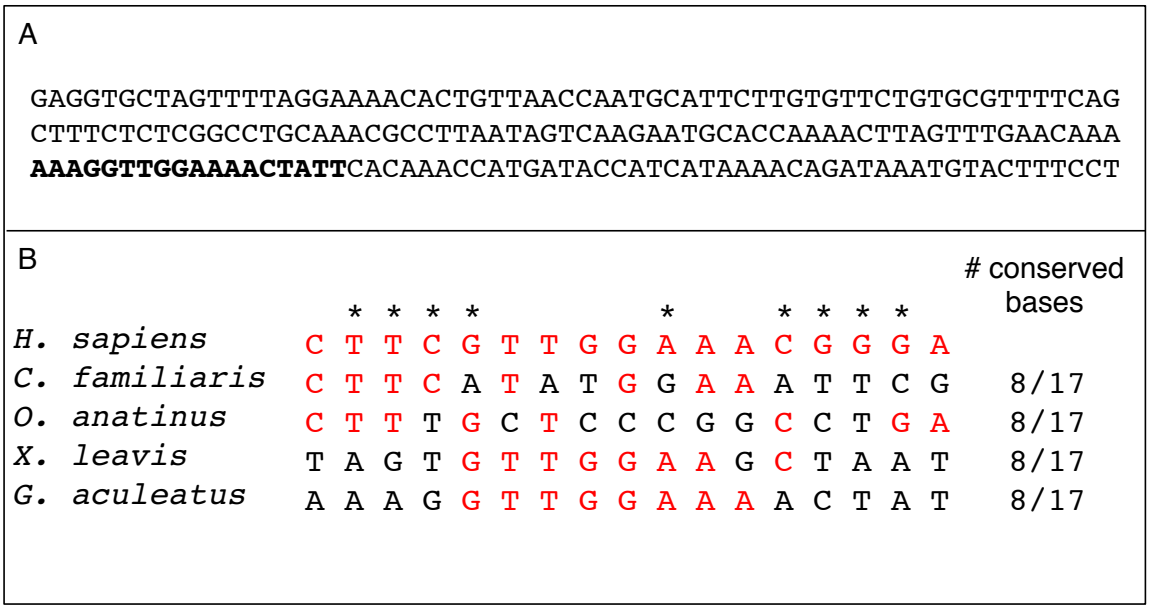


Figure 2-5. Threespine stickleback centromere repeat sequence. **a** The consensus GacCEN repeat is 186 bp and has an AT content of 62.9%. The putative CENP-B box is highlighted in *bold.* **b** The putative CENP-B box in the GacCEN shows sequence similarity to the CENP-B box in humans (*Homo sapiens),* dog (*Canis familiaris),* platypus (*Ornithorhynchus anatinus),* and the African clawed frog (*Xenopus laevis)*. Identical nucleotides to the human CENP-B box are in *red*, and *asterisks* denote the evolutionary conserved domains in humans (Ohzeki et al. 2002; Alkan et al. 2011). This sequence has been deposited in GenBank (accession number KT321856).

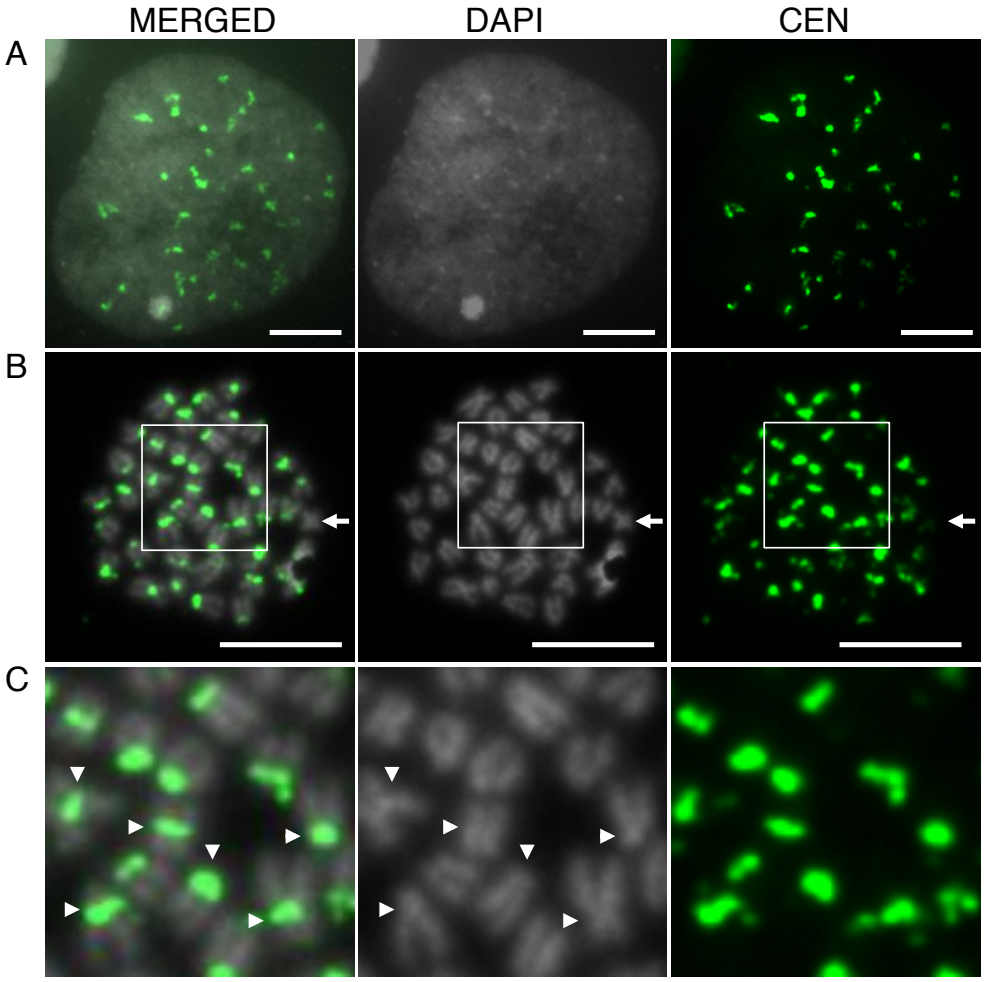


Figure 2-6. The threespine stickleback CEN repeat hybridizes to the centromere. FISH reveals that the CEN probe (*green*) hybridizes to **a** distinct puntca in interphase nuclei from Lake Union embryos, **b** a single region on each chromosome except one (*arrow*) in a metaphase spread from a Lake Union male, and **c** the primary constriction (*arrowheads*) on each chromosome in a higher magnification view of the boxed region in panel (**b**). *Scale bar, 5 μm*

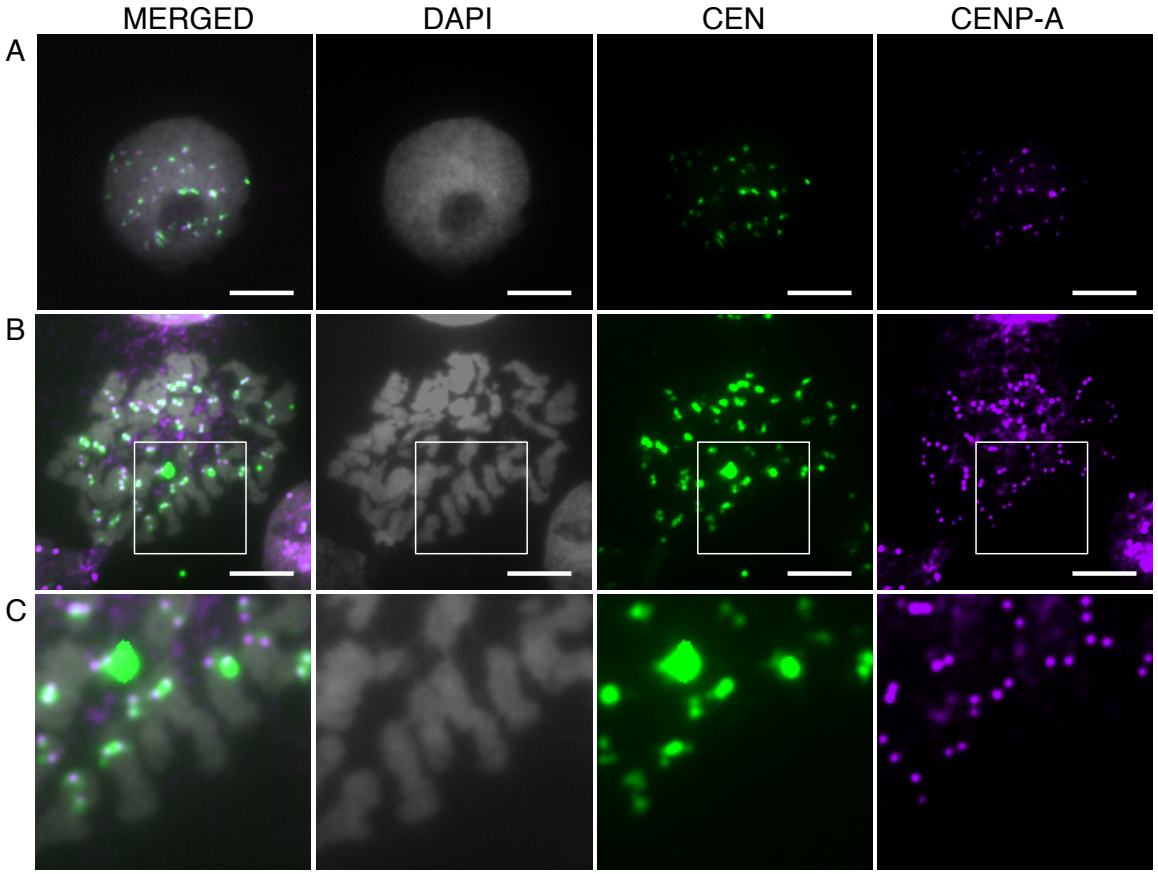


Figure 2-7. The threespine stickleback centromere repeat colocalized with CENP-A. IF-FISH with CENP-A antibody (*purple)* shows colocalization with the CEN probe (*green) in* **a** interphase nuclei and **b,c** metaphase chromosomes from Lake Union embryos. Panel (**c**) shows a higher magnification view of the boxed region in panel (**b**). *Scale bar, 5 μm*

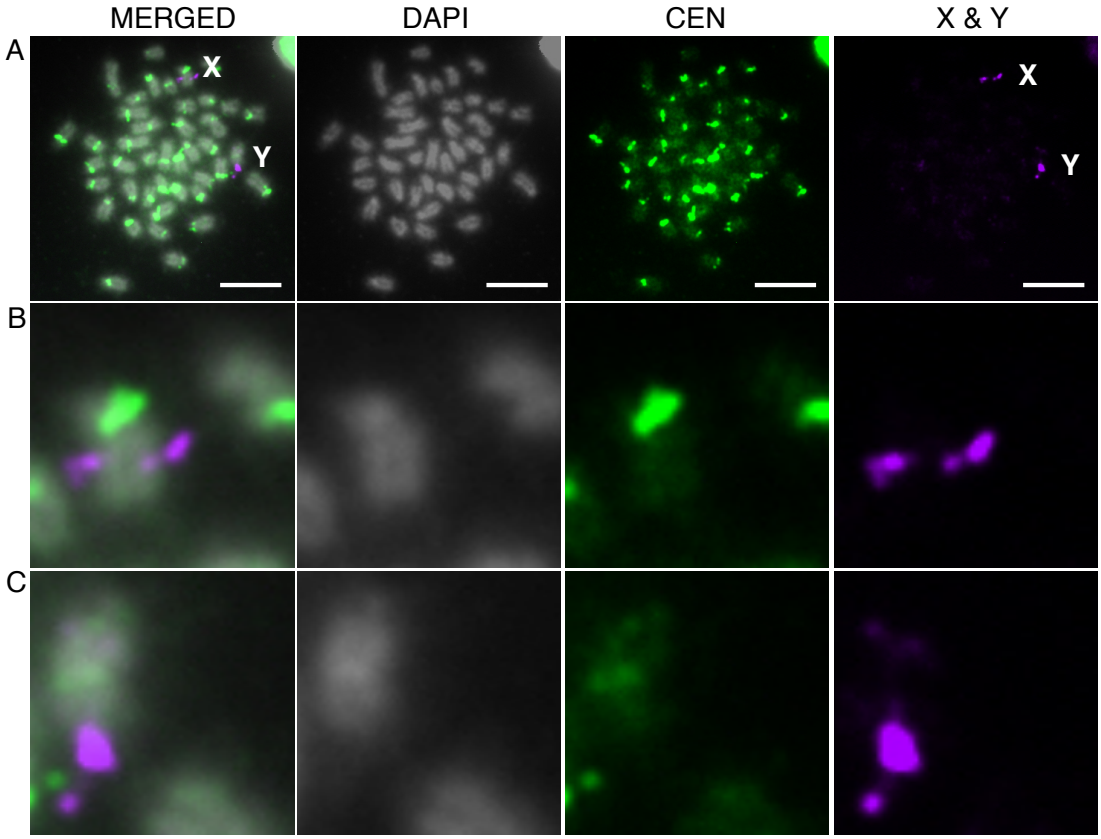


Figure 2-8. The threespine stickleback centromere repeat shows weak hybridization to the Y chromosome centromere. **a** FISH reveals that the CEN probe (*green)* shows weak hybridization to the Y chromosome but strong hybridization to the X chromosome on a metaphase spread from a Pacific Ocean male. The X and Y chromosome can be identified and distinguished by hybridization to a FISH probe from BAC 101E08 (*purple)* (Ross and Peichel 2008). Higher magnification views of hybridization to the X (**b**) and Y (**c**) chromosomes are shown. *Scale bar, 5 μm*

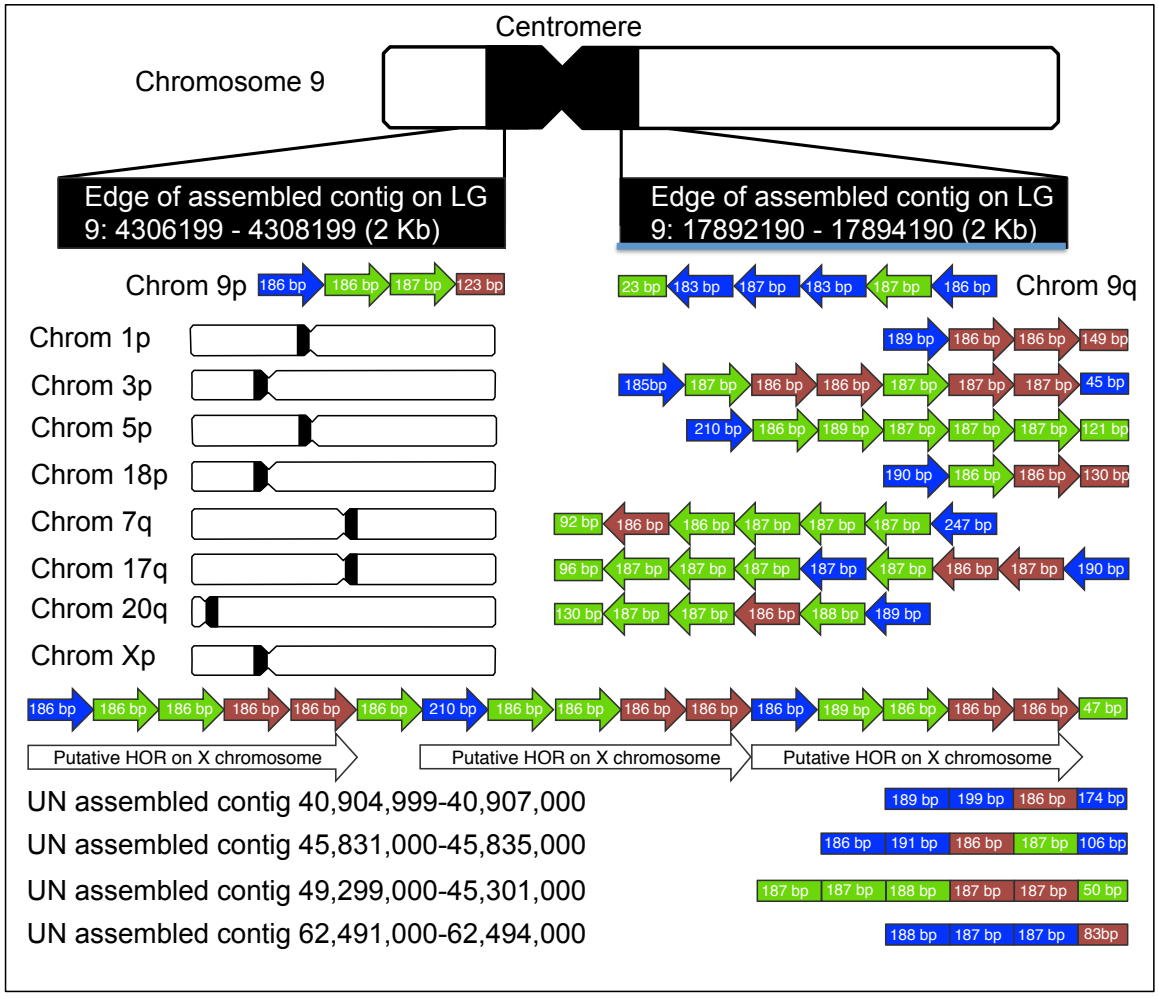


Figure 2-9. Scaffolds containing centromere repeats. Assembled scaffolds containing centromere repeats were found on both edges of the gap in the chromosome 9 assembly corresponding to the position of the centromere, and on a single edge of the gaps corresponding to the centromeres of chromosomes 1, 3, 5, 7, 17, 18, 20, and the X chromosome (Urton et al. 2011). Additional GacCEN like repeats were found on four scaffolds from unassembled regions of the genome. Each arrow shows the size of the repeat variant, and the percent identity to GacCEN is represented by color: blue <90 %; green 90-95 %; red >95 %. The direction of the arrow points towards the center of the genome assembly gap. The dark shaded regions in each chromosome drawing depicts the region where the GacCEN containing scaffold is found; relative position of the centromeres is based on Urton et al. 2011.

Chapter 2 References:

Alkan C, Cardone MF, Catacchio CR, et al. (2011) Genome-wide characterization of centromeric satellites from multiple mammalian genomes. Genome Res 21:137–145

Amor DJ, Bentley K, Ryan J, et al. (2004) Human centromere repositioning "in progress". Proc Natl Acad Sci USA 101:6542–6547

Blower MD, Sullivan BA, Karpen GH (2002) Conserved organization of centromeric chromatin in flies and humans. Dev Cell 2:319–330

Crollius HR, Jaillon O, Dasilva C, et al. (2000) Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. Genome Res 10:939–949

Drinnenberg IA, deYoung D, Henikoff S, Malik HS (2014) Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. eLife 3: e03676

Earnshaw WC, Rothfield N (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma 91:313–321

Edwards NS, Murray AW (2005) Identification of *Xenopus* CENP-A and an associated centromeric DNA repeat. Mol Biol Cell 16:1800–1810

Fountain DM, Kral LG (2011) Isolation and characterization of the *Etheostoma tallapoosae* (Teleostei: Percidae) CENP-A gene. Genes 2:829–840

Fu L, Niu B, Zhu Z, et al. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152

Garrido-Ramos MA, Jamilena M, Lozano R, et al. (1994) Cloning and characterization of a fish centromeric satellite DNA. Cytogenet Cell Genet 65:233–237

Gong Z, Wu Y, Koblizkova A, et al. (2012) Repeatless and repeat-based centromeres in potato: implications for centromere evolution. Plant Cell 24:3559–3574

Gordon DJ, Resio B, Pellman D (2012) Causes and consequences of aneuploidy in cancer. Nat Rev Genet 13:189–203

Haaf T, Schmid M, Steinlein C, et al. (1993) Organization and molecular cytogenetics of a satellite DNA family from *Hoplias malabaricus* (Pisces, Erythrinidae). Chromosome Res 1:77–86

Henikoff JG, Thakur J, Kasinathan S, Henikoff S (2015) A unique chromatin complex occupies young α -satellite arrays of human centromeres. Science Advances 1:e1400234

Henikoff S (2002) Near the edge of a chromosome's "black hole". Trends Genet 18:165–167

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098–1102

Houben A, Schroeder-Reiter E, Nagaki K, et al. (2007) CENH3 interacts with the centromeric retrotransposon *cereba* and GC-rich satellites and locates to centromeric substructures in barley. Chromosoma 116:275–283

Hunt P, Hassold T (2010) Female meiosis: coming unglued with age. Curr Biol 20:R699–R702

Jones FC, Grabherr MG, Chan YF, et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55-61

Kingsley DM, Peichel CL (2007) The molecular genetics of evolutionary change in sticklebacks. In: Östlund-Nilsson S, Mayer I, Huntingford FA, eds. Biology of the Three-Spined Stickleback. Boca Raton: CRC press, pp 41–81

Kitano J, Mori S, Peichel CL (2007) Phenotypic divergence and reproductive isolation between sympatric forms of Japanese threespine sticklebacks. Biol J Linn Soc 91:671–685

Kitano J, Ross JA, Mori S, et al. (2009) A role for a neo-sex chromosome in stickleback speciation. Nature 461:1079–1083

Kops GJPL, Weaver BAA, Cleveland DW (2005) On the road to cancer: aneuploidy and the mitotic checkpoint. Nat Rev Cancer 5:773–785

Lee H-R, Zhang W, Langdon T, et al. (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. Proc Natl Acad Sci USA 102:11793–11798

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659

Lister LM, Kouznetsova A, Hyslop LA, et al. (2010) Age-related meiotic segregation errors in mammalian oocytes are preceded by depletion of cohesin and Sgo2. Curr Biol 20:1511–1521

Maio JJ (1971) DNA strand reassociation and polyribonucleotide binding in the African

green monkey, *Cercopithecus aethiops*. J Mol Biol 56:579–595

Malik HS, Henikoff S (2009) Major evolutionary transitions in centromere complexity. Cell 138:1067–1082

Manuelidis L (1978) Chromosomal localization of complex and simple repeated human DNAs. Chromosoma 66:23–32

Masumoto H, Masukata H, Muro Y, et al. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J Cell Biol 109:1963–1973

Masumoto H, Nakano M, Ohzeki J-I (2004) The role of CENP-B and alpha-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres. Chromosome Res 12:543–556

Melters DP, Bradnam KR, Young HA, et al. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14:R10

Miga KH, Newton Y, Jain M, et al. (2014) Centromere reference models for human chromosomes X and Y satellite arrays. Genome Res 24:697–707

Morales C, Sánchez A, Bruguera J, et al. (2007) Cytogenetic study of spontaneous abortions using semi-direct analysis of chorionic villi samples detects the broadest spectrum of chromosome abnormalities. Am J Med Genet 146A:66–70

Nagaki K, Cheng Z, Ouyang S, et al. (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145

Nagaki K, Kashihara K, Murata M (2008) A centromeric DNA sequence colocalized with a centromere-specific histone H3 in tobacco. Chromosoma 118:249–257

Nagaki K, Murata M (2005) Characterization of CENH3 and centromere-associated DNA sequences in sugarcane. Chromosome Res 13:195–203

Nagaki K, Talbert PB, Zhong CX, et al. (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics 163:1221–1225

Ohzeki J-I, Nakano M, Okada T, Masumoto H (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J Cell Biol 159:765–775

Palmer DKD, O'Day KK, Margolis RLR (1989) Biochemical analysis of CENP-A, a centromeric protein with histone-like properties. Prog Clin Biol Res 318:61–72

Palmer DK, O'Day K, Trong HL, et al. (1991) Purification of the centromere-specific

protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci USA 88:3734–3738

Pertile MD, Graham AN, Choo KHA, Kalitsis P (2009) Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability. Genome Res 19:2202–2213

Piras FM, Nergadze SG, Magnani E, et al. (2010) Uncoupling of satellite DNA and centromeric function in the genus *Equus*. PLoS Genet 6:e1000845

Revenkova E, Herrmann K, Adelfalk C, Jessberger R (2010) Oocyte cohesin expression restricted to predictyate stages provides full fertility and prevents aneuploidy. Curr Biol 20:1529–1533

Ricke RM, van Deursen JM (2013) Aneuploidy in health, disease, and aging. J Cell Biol 201:11–21

Ross JA, Peichel CL (2008) Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. Genetics 179:2173–2182

Rudd MK, Willard HF (2004) Analysis of the centromeric regions of the human genome assembly. Trends Genet 20:529–533

Schueler MG, Higgins AW, Rudd MK, et al. (2001) Genomic and genetic definition of a functional human centromere. Science 294:109-115

Shang WH, Hori T, Toyoda A, et al. (2010) Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. Genome Res 20:1219–1228

Shelby RD, Vafa O, Sullivan KF (1997) Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. J Cell Biol 136:501–513

Shepelev VA, Alexandrov A, Yurov YB, et al. (2009) The evolutionary origin of man can be traced in the layers of defunct ancestral alpha satellites flanking the active centromeres of human chromosomes. PLoS Genet 5: e1000641

Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. Science 191:528–535

Sullivan BA, Schwartz S (1995) Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. Hum Mol Genet 4:2189–2197

Sullivan KF, Hechenberger M, Masri K (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol 127:581–592

Tek AL, Kashihara K, Murata M, Nagaki K (2010) Functional centromeres in soybean include two distinct tandem repeats and a retrotransposon. Chromosome Res 18:337–347

Urton JR, McCann SR, Peichel CL (2011) Karyotype differentiation between two stickleback species (Gasterosteidae). Cytogenet Genome Res 135:150–159

Warburton PE, Cooke CA, Bourassa S, et al. (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr Biol 7:901–904

Willard HF (1991) Evolution of alpha satellite. Curr Opin Genet Dev 1:509-514

Wolfe J, Darling SM, Erickson RP, et al. (1985) Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. J Mol Biol 182:477–485

Zhong CX (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

Chapter 3: Centromere inactivation on a neo-Y fusion chromosome in stickleback fish

Introduction

The centromere of a chromosome was first described by Walther Flemming in the 1880's as the primary constriction observed on condensed chromosomes during both mitosis and meiosis (Flemming 1880). This primary constriction is the region of the chromosome where microtubules attach, allowing for the segregation of chromosomes during cell division. Many organisms, including mammals and some plants, contain "regional" centromeres. Regional centromeres are found at a single location on each chromosome, and are comprised of kilobases to megabases of DNA. The sequence of these regional centromeres is often repetitive and AT rich, but the specific sequence varies dramatically between species (Henikoff et al. 2001; Alkan et al. 2011; Melters et al. 2013), and can even vary among centromeres within a species (Nagaki et al. 2004; Piras et al. 2010; Tek et al. 2010; Shang et al. 2010; Gong et al. 2012). Regional centromeres have a common epigenetic characteristic; the presence of a histone variant, centromere protein A (CENP-A) (Palmer et al. 1987; Palmer et al. 1989; Palmer et al. 1991; Sullivan et al. 1994). CENP-A replaces histone H3 in centromeric nucleosomes, and its presence is considered the hallmark of an active centromere (Voullaire et al. 1993; Warburton et al. 1997; Warburton 2004; Heun et al. 2006; Foltz et al. 2006; Marshall et al. 2008; Allshire and Karpen 2008; Silva and Jansen 2009; Olszak et al. 2011; Barnhart et al. 2011; Mendiburo et al. 2011; Sekulic and Black 2012; Shang et al. 2013; Fukagawa and Earnshaw 2014).

While the centromere has conserved and essential functions for chromosomes, it is also essential that every chromosome has one and only one active centromere in organisms with monocentric chromosomes. A normal monocentric chromosome only has

a single centromere that can bind microtubules and form a bipolar spindle, allowing normal chromosome segregation. However, a dicentric chromosome has two centromeres that can both bind microtubules, leading to instability during cell division. Barbara McClintock first described this instability and termed it the breakage-fusion-bridge cycle (McClintock 1939; McClintock 1941). If the microtubules from spindle poles on the opposing sides of the cell bind to the two centromeres, the chromosome will be pulled in opposite directions. This leads to "anaphase lag" where the dicentric chromosome lags in the middle during anaphase as it is being pulled to opposite poles. The physical tension on the chromosome can lead to breakage of the chromosome. When the broken ends are replicated during the next S- phase of the cell cycle, the broken ends can anneal back to each other, re-creating a dicentric. Not only does the breakage-fusion-bridge cycle start again, but the breakage and re-annealing leads to gene loss and gene gain. This is evident in cancer cells where dicentrics are often observed, leading to cells with "increased genetic heterogeneity" (Gisselsson et al. 2000; Gisselsson et al. 2001).

Although dicentric chromosomes should be very unstable, it has also been shown that if two centromeres on a fusion chromosome are close enough to each other, they will act as a single centromere (Koshland et al. 1987; Page and Shaffer 1998; Sullivan and Willard 1998). This is seen both in chromosome fusions that were artificially created in the lab as well as in naturally occurring Robertsonian fusions, where a shorter intercentromeric distance allows for two active centromeres, and a long intercentromeric distance is associated with centromere inactivation (Page and Shaffer 1998; Sullivan and Willard 1998; Higgins et al. 2005; Lange et al. 2009).

There are also examples of non-Robertsonian fusions that should create dicentric chromosomes and are stable, suggesting that the chromosomes are able to alleviate the problems associated with being dicentric by inactivating one of the centromeres, thereby re-establishing monocentric chromosomes. Most of our current understanding of how dicentrics are stabilized comes from artificially created dicentric chromosomes in yeast, fruit flies, maize, and human cell lines (Sears and Câmara 1952; Agudo et al. 2000; Higgins et al. 2005; Pobiega and Marcand 2010; Zhang et al. 2010; Stimpson et al. 2010; Sato et al. 2012; Stimpson et al. 2012), and a few examples of naturally occurring dicentric chromosome in humans and wheat (Therman et al. 1989; Maraschio et al. 1990; Fisher et al. 1997; Page and Shaffer 1998; Sullivan and Willard 1998; Higgins et al. 2005; Han et al. 2006; MacKinnon and Campbell 2011; Koo et al. 2011; Stimpson et al. 2012; Flemming 1880; Liu et al. 2015).

From these studies, there are three proposed mechanisms of dicentric stabilization: chromosome re-breakage, and centromere inactivation by either genetic or epigenetic inactivation of one centromere. In chromosome re-breakage, the dicentric is simply re-broken between the two centromeres, thereby re-creating two monocentric chromosomes. The other two models both propose stabilization through two different mechanisms of centromere inactivation. In genetic inactivation, deleterious mutation of the DNA sequence of one centromere prevents it from binding CENP-A and therefore microtubules. In epigenetic inactivation, the DNA of the inactivated centromere is still present, but CENP-A is not. Without CENP-A, the centromeric DNA does not recruit kinetochore components and therefore does not bind mictrotubules during cell division.

While we have learned about the possible mechanisms of dicentric stabilization

from artificially created dicentrics and from human disease patients, there is currently almost nothing known about the stabilization of natural dicentric chromosomes, which are often created when chromosomes fuse in an evolutionary context. Chromosome fusion and fission events are common during evolution as evidenced by the extensive variation in chromosome number among species. For example, in multicellular eukaryotes, chromosome number ranges from a diploid number of 2 in the Jack jumper ant (Crosland and Crozier 1986) to 1260 in the Addlers tongue plant (Van der Burg *2004*). In eutherian mammals, the diploid chromosome number ranges from 6 to 102 in the Viscacha rat (Gallardo et al. 1999); this diversity results from many independent fusion and fission events (Ferguson-Smith and Trifonov 2007). The mammalian species with the lowest number of chromosome is the Indian muntjac with a diploid number of 6 in females or 7 in males. Interestingly, a closely related species, the Chinese muntjac has a diploid chromosome number of 46. This large difference in chromosome number does not prevent these two sister species from breeding in captivity (Shi et al. 1980; Yang et al. 1995; Wang and Lan 2000). It is speculated that a series of chromosome fusion events between the telocentric chromosomes present in the Chinese muntjac created the much larger Indian muntjac chromosomes (Wang and Lan 2000). Chromosome fusions have also occurred in the primate lineage; human chromosome 2 is a fusion between chimp chromosomes 2a and 2b (Lejeune et al. 1973; IJdo et al. 1991; Avarello et al. 1992). All of these data suggests that chromosome fusion and fission events are common and can account for much of the diversity of karyotype number we see today. Based on these findings, it is theorized that many of these chromosome fusion events must have been accompanied by centromere inactivation to stabilize the chromosome fusions. However,

almost nothing is known about the mechanisms involved in centromere inactivation during evolution.

The stickleback family (Gasterosteidae) provides a unique opportunity to study the mechanisms that stabilize natural dicentric chromosomes during evolution because there have been several chromosome fusion events that have occurred within in the past 35 million years (Ross et al. 2009; Kitano et al. 2009). Previous work has identified a chromosome fusion involving the ancestral metacentric Y chromosome present in the Pacific Ocean threespine stickleback (*Gasterosteus aculeatus*). This Y chromosome has fused to the acrocentric chromosome 9 in males of the Japan Sea threespine stickleback (*Gasterosteus nipponicus*) species within the past two million years, creating a neo-Y chromosome (Kitano et al. 2009) (Figure 3-1). Because this fusion involved the metacentric Y and an acrocentric chromosome, the neo-Y should have been dicentric upon formation. Preliminary data suggested that there was no loss of a chromosome arm on the Japan Sea neo-Y chromosome fusion or re-breakage of the fusion chromosome. These data suggest that centromere inactivation has occurred on the Japan Sea neo-Y chromosome.

In this study, we aimed to first determine if centromere inactivation has occurred on the neo-Y, and to then test whether inactivation is due to a genetic or epigenetic mechanism. Both the genetic and epigenetic mechanisms of centromere inactivation ultimately result in only one functional centromere, as defined by the presence of CENP-A staining. If centromere inactivation has occurred, we expect to see only one region on the neo-Y with CENP-A staining. If no centromere inactivation has occurred, we expect to see two regions of CENP-A staining on the neo-Y. Using CENP-A antibody staining,

we determined that the Japan Sea neo-Y has only one active centromere, as defined by presence of CENP-A. In an extension of our previous study (Cech and Peichel 2015), we then performed chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) to determine the centromere sequence in the Japan Sea species. We then used fluorescence in situ hybridization (FISH) to determine that there have likely been genetic changes to the ancestral Y centromere on the Japan Sea neo-Y, leading to inactivation of this centromere.

Materials and Methods

Fish use and care

Two species of lab reared threespine stickleback fish were used in this study: Pacific Ocean (*Gasterosteus aculeatus*) and Japan Sea (*Gasterosteus nipponicus*). Lab-reared fish were derived from wild caught fish collected in Akkeshi on Hokkaido Island, Japan (Kitano et al. 2007, 2009). Fish were kept in in 3.5% saltwater (3.5 g/l Instant Ocean salt (Spectrum Brands, USA); 0.4 ml/l sodium bicarbonate) in 29-gal aquarium tanks, at 16 °C and 16 hr light / 8 hr dark lighting conditions. Fish were fed twice daily with live *Artemia* nauplii and once daily with frozen *Mysis* shrimp. All institutional and national guidelines for the care and use of laboratory animals were followed, and all procedures were approved by the Fred Hutchinson Cancer Research Center Institutional Animal Care and Use Committee (protocol 1575).

CENP-A immmunoprecipitation followed by high throughput sequencing (ChIP-seq)

Chromatin immunopreciptation with the stickleback specific CENP-A antibody was performed using the SimpleChIP Plus Enzymatic Chromatin IP Kit (Magnetic Beads; Cell Signaling Technology, USA). 0.1 g of liver and tissue was taken from each of two Pacific Ocean males, two Japan Sea females, and two Japan Sea male sticklebacks. The chromatin immunoprecipitation was performed at the same time and following the same protocol as for the two Pacific Ocean female samples described in Cech and Peichel (2015). To identify the Japan Sea consensus centromere sequence, ChIP analysis was performed using the same methods as in Cech and Peichel (2015).

To search for the Pacific Ocean Y centromere sequence, the top 500 enriched reads for both Pacific Ocean males and Japan Sea males and females were determined as in Cech and Peichel (2015). Using CD-EST-HIT2d (Li and Dogzik 2006; Fu et al. 2012) Pacific Ocean male enriched reads were compared to Pacific Ocean female enriched reads to find Pacific Ocean male unique reads that were less than 80% similar to female samples (Cech and Peichel 2015). Pacific Ocean male enriched reads were also compared to Japan Sea male enriched reads to find Pacific Ocean male unique reads not found in Japan Sea male samples. These two data sets were then aligned to the threespine stickleback centromeric repeat sequence (GacCEN) to identify reads that differed from the consensus GacCEN (Cech and Peichel 2015).

Fluorescence in situ hybridization (FISH) on metaphase spreads and interphase nuclei

Lab reared Pacific Ocean and Japan Sea male and female lab fish were injected with 10 µl of 1 % colchicine in PBS for 14-16 hours to arrest metaphase cells. Metaphase and interphase cells were obtained following the same protocol described in Ross and Peichel

(2008), with the exception that metaphase slides were dried at 50 °C, not 37 °C. The GacCEN probe was made as described in Cech and Peichel (2015). Bacterial artificial chromosome (BAC) probes listed in Table 1 were prepared following the protocol in Urton et al. (2011) and labeled with either Alexa-488 or Alexa-568.

Metaphase slides were first washed for 5 min with phosphate buffered saline Tween-20, fixed for 10 min in 4% paraformaldehyde in phosphate buffered saline, then washed again in phosphate buffered saline Tween-20 for 5 min. Subsequent FISH hybridization with just GacCEN or just BAC probes was performed as described in Cech and Peichel (2015). For metaphase slides hybridized with combined GacCEN and BAC probes, 10 µl of the GacCEN probe was first lyophilized at 50 °C for four hours. The dried GacCEN probe was then resuspended with either 10 µl of a single BAC probe in hybridization buffer, or 20 µl of equal parts of two BAC probes in hybridization buffer.

Telomere staining

Telomere staining was performed according to manufacturers instructions using the DAKO telomere PNA FISH kit/FITC no. K5325 (Agilent Technologies, USA).

Immunofluorescence-FISH (IF-FISH)

For IF-FISH, 15 Pacific Ocean or Japan Sea embryos at 48 hours days post fertilization were dounced in 10 ml 0.56 % potassium chloride using a glass dounce. The remainder of the IF-FISH protocol follows Cech and Peichel (2015).

Microscopy

Images were taken using the Nikon Eclipse 80i microscope (Nikon, Japan) with an automated filter turret (Chroma filters 31000v2 (DAPI), 41001 (FITC), and 41004 (Texas Red); Chroma, USA) using the 100x objective. Following Cech and Peichel (2015), images were pseudo-colored using NIS Elements imaging software (BR 3.00, SP7, Hotfix8, Build 548, Nikon, Japan).

Results

No loss of chromosome arms on the Japan Sea neo-Y fusion chromosome

To determine if the neo-Y fusion was a complete fusion between both chromosomes, or if large portions of each or one of the respective p-arms was lost, we performed two types of FISH experiments. First, we conducted hybridization with BAC probes homologous to the ends of chromosome 9 (44L12) and the Y chromosome (101E8) that are near the fusion site, as well as the end of the Y chromosome (188J19) opposite to the fusion site (Table 3-1). Hybridization of these probes on Japan Sea male metaphase chromosomes not only shows only three chromosomes (X_1, X_2, neo-Y) with hybridization, but also demonstrates that loss of an entire chromosome arm has not occurred on the neo-Y fusion chromosome (Figure 3-2). Hybridization of these same BAC probes on Pacific Ocean male metaphase chromosomes further demonstrates that there have not been gross rearrangements between the Japan Sea and Pacific Ocean sticklebacks on either chromosome 9 or the Y chromosome (Figure 3-3).

Second, to test if there was truly no loss of any genetic material, we also used a FISH probe to the 6bp telomere sequence (TTAGGG) found in all vertebrates (Meyne et al. 1989). Telomere staining on the neo-Y chromosome shows no internal telomere

signal, indicating that at least some loss of genetic material occurred on the neo-Y fusion (Figure 3-4). This is not surprising, as telomeric signals are normally found at the ends of chromosomes (Meyne et al. 1989). Interestingly, telomere loss can induce chromosome fusions (Gisselsson et al. 2001; Maser and DePinho 2002; Bailey 2006; Pobiega and Marcand 2010; Murnane 2010; Stimpson et al. 2010). However, we do not know whether loss of telomeres is a cause or consequence of the neo-Y fusion. Taken together, these data do suggest that the neo-Y chromosome resulted from a fusion between two nearly complete monocentric chromosomes, creating a dicentric neo-Y chromosome. The Japan Sea fish are fertile, and we observe no evidence for the chromosome segregation associated aneuploidy (Chapter 4) often seen with dicentric chromosomes. Furthermore, there is only evidence of only one constriction on the neo-Y in metaphase chromosome spreads (Figure 3-4). Thus, we hypothesized that the neo-Y chromosome has been stabilized by one of the three mechanisms described above. There is no evidence of re-breakage of the Japan Sea neo-Y, suggesting that the stabilization occurred by centromere inactivation via either genetic or epigenetic inactivation.

Evidence for centromere inactivation on the Japan Sea neo-Y chromosome

Using CENP-A antibody staining in combination with FISH, we determined that only one region of CENP-A staining is found on the neo-Y in multiple independent metaphase spreads from Japan Sea males (Figure 3-5). The CENP-A staining was flanked by BAC probes present at the end of the Y near the fusion breakpoint and at the distal end of chromosome 9, indicating that the active centromere is on chromosome 9 (Figure 3-5). The CENP-A staining on the neo-Y is in a similar location to the CENP-A staining found

on the unfused chromosome 9, further suggesting that the active centromere on the neo-Y is retained from the ancestral acrocentric chromosome 9 (Figure 3-5f).

To confirm that the ancestral Y chromosome centromere was active before the fusion that created the neo-Y occurred, we performed CENP-A staining in combination with FISH using a BAC probe (91G03; Table 3-1) that specifically labels the Y chromosome on Pacific Ocean male chromosomes (Figure 3-6). We found normal CENP-A staining on the ancestral Y centromere (Figure 3-7). This is evidence that the ancestral Y chromosome had an active centromere, and that the Y chromosome centromere was inactivated after the fusion to chromosome 9 in Japan Sea males.

Evidence for genetic inactivation on the Japan Sea neo-Y chromosome

Because we found evidence for only a single active centromere on the Japan Sea neo-Y chromosome, we wanted to determine whether this was due to genetic inactivation (i.e. deletion or alteration) of the ancestral Y chromosome centromere. We had previously identified the centromere sequence (GacCEN) in the Pacific Ocean threespine stickleback (Cech and Peichel 2015). Although this centromere sequence appears to be present on all autosomes, including chromosome 9 and the X chromosome, we observed very weak hybridization of the centromere sequence to the Y chromosome in the ancestral Pacific Ocean population (Cech and Peichel 2015). It was unclear whether the centromere on the Y chromosome has a completely different sequence, a highly divergent sequence, or a reduced number of repeats. Here, we further explored these possibilities in order to analyze the fate of the ancestral Y chromosome centromere on the Japan Sea neo-Y. We first performed additional FISH experiments on Pacific Ocean males with the centromere

probe and confirmed that centromere hybridization on the ancestral Y is weak and variable. Over 12 independent FISH experiments, we counted 28 metaphase spreads with positive GacCEN staining, and 27 with negative GacCEN staining. This variability is demonstrated by both positive and negative staining on metaphase spreads from the same male (Figure 3-8). These data suggest that the Y centromere is likely not a completely different sequence, but instead might represent a divergent repeat sequence and/or a reduced number of the same repeat sequence.

To further determine whether there is a unique Y centromere sequence, we utilized CENP-A ChIP sequencing data obtained from Pacific Ocean male and female samples. Using analyses pipeline that we previously developed to identify the GacCEN in Pacific Ocean female samples (Cech and Peichel 2015), we first identified the 500 most abundant sequence clusters in two Pacific Ocean males. Of these 500 most abundant sequence clusters, 328 and 309 were enriched in the CENP-A immunoprecipitation (IP) sequences relative to input sequences in the two independent Pacific Ocean male samples. We then compared the top enriched clusters from Pacific Ocean males to those identified previously in Pacific Ocean females (Cech and Peichel 2015) to find CENP-A associated sequences unique to Pacific Ocean males. Of these 140 male unique sequence clusters, 131 had homology ranging from 76.9% to 93.5% similarity to GacCEN. The remaining nine sequence clusters all had a very low number of overall reads in both the input and IP sample, comprising only 0.003% of the 11,582,843 total IP reads from the top 637 enriched clusters. Seven of these reads are mostly CA dinucleotide repeats. Taken together with the FISH data, these analyses suggest that there is not a unique centromere sequence on the ancestral Pacific Ocean Y chromosome. However, until we

are able to completely sequence the Pacific Ocean Y chromosome, including the centromere, we will be unable to verify this hypothesis.

To determine whether the centromere sequence in the Japan Sea species is the same as in the Pacific Ocean species, we also performed CENP-A ChIP-seq in Japan Sea males and females at the same time we performed the Pacific Ocean CENP-A ChIP-seq (Cech and Peichel 2015). We had previously demonstrated that the Pacific Ocean and Japan Sea CENP-A genes only differ by a single amino acid (Cech and Peichel 2015), which is not in the region of the protein used for the antibody design (Figure 3-9). Using previously described methods (Cech and Peichel 2015), we identified a consensus CENP-A associated putative centromere sequence in both Japan Sea males and females. The Japan Sea putative centromere sequence is 98.7% similar to the Pacific Ocean GacCEN, and is also 186 bp and 61.2 % AT rich (Figure 3-10). Similar to the Pacific Ocean population, the Japan Sea GacCEN probe shows hybridization to the constriction on metaphase spreads (Figure 3-11) as well as co-localization with CENP-A (Figure 3-12).

We performed FISH using the GacCEN probe in combination with BAC probes on Japan Sea metaphase spreads. There is strong hybridization of the GacCEN probe to the ancestral chromosome 9 centromere, providing evidence that this centromere has not been deleted on the neo-Y (Figure 3-13). Although we also observed strong GacCEN staining on both X chromosomes in Japan Sea females (Figures 3-14), there was no hybridization of the GacCEN probe to the ancestral Y centromere on the neo-Y (Figure 3-13). This lack of hybridization to the Y chromosome was consistently observed in seven different metaphase spreads from three different experiments. Because the GacCEN probe staining is weak on the ancestral Y centromere, we cannot determine

whether the lack of staining to the Y centromere on the neo-Y is due to deletion of the centromeric DNA on the neo-Y or divergence of the centromeric DNA on the neo-Y. However, these data do suggest that there has been a genetic alteration to the ancestral Y chromosome centromere on the Japan Sea neo-Y.

To further assess whether there has been a deletion of the ancestral Y centromere on the neo-Y, we examined the hybridization pattern of BAC probes flanking the ancestral Y centromere on the neo-Y. Previous work had identified two BAC probes (171H24 and 180J08; Table 3-1) flanking the centromeric constriction on both the X and Y chromosomes in the Pacific Ocean species (Ross and Peichel 2008). Note that the order of these probes is different between the X and Y chromosomes due to the presence of several inversions on the Y chromosome relative to the X (Ross and Peichel 2008). There is no loss of hybridization with either probe on the Japan Sea neo-Y (Figure 3-15), indicating that these large regions flanking the centromere are still present. However, when compared to hybridization on the Japan Sea X, these probes do appear to be closer together on the neo-Y (Figure 3-15). Additionally, hybridization with these two probes and the GacCEN probe shows that there is also a slight distance between these two probes on the ancestral Pacific Ocean Y chromosome, along with weak internal centromere staining (Figure 3-16). By contrast, there is no internal GacCEN signal between these two BACs on the Japan Sea neo-Y (Figure 3-17) although there is clear separation between these BACs and internal GacCEN signal on both X chromosomes in Japan Sea females (Figure 3-18). Taken together, these data suggest, but do not definitely prove, that there has been a deletion of the ancestral Y chromosome centromere on the Japan Sea neo-Y. Taken together, our data support a model in which there has been a

genetic inactivation of the ancestral Y centromere on the Japan Sea neo-Y, leading to a single functional centromere on this chromosome fusion.

Discussion

In order to determine if the ancestral Y was truly deleted or has just diverged, further experiments will be needed to sequence the ancestral Y centromere. As in most organisms besides humans, a few other primates and mice (Wolfe et al. 1985; Pertile et al. 2009), there is no Y chromosome sequence assembly in stickleback. This work is currently in progress and will hopefully include the sequence of the Y centromere. Once the Y chromosome is assembled, we can further test for genetic deletion by fine mapping the centromere boundaries using new BAC probes.

While we cannot rule out if the ancestral Y centromere region has been deleted or has changed in some other way on the neo-Y, our experiments suggest there have been genetic changes, and likely some deletion on the neo-Y ancestral Y centromere. Our evidence for a partial deletion or loss of centromeric DNA is consistent with genetic deletions observed in inactivated human centromeres (Maraschio et al. 1990; Fisher et al. 1997; Stimpson et al. 2010). These results suggest that epigenetic and genetic changes may both play a role in centromere inactivation.

Overall, we have shown that the Japan Sea neo-Y chromosome fusion has been stabilized by centromere inactivation. The ancestral Y centromere has been inactivated as evidenced by loss of CENP-A staining, and the active neo-Y centromere comes from chromosome 9 (Figure 3-19). Future work should further characterize this inactivation by looking for additional histone and DNA methylation changes. Centromere inactivation in

maize shows the replacement of H3K4me2 with H3K27me2/3 (Zhang et al. 2010), as well as hypermethylation of DNA at inactivated centromeres (Koo et al. 2011). It would be interesting to characterize the inactive and active states of stickleback centromeres at the histone and DNA methylation level.

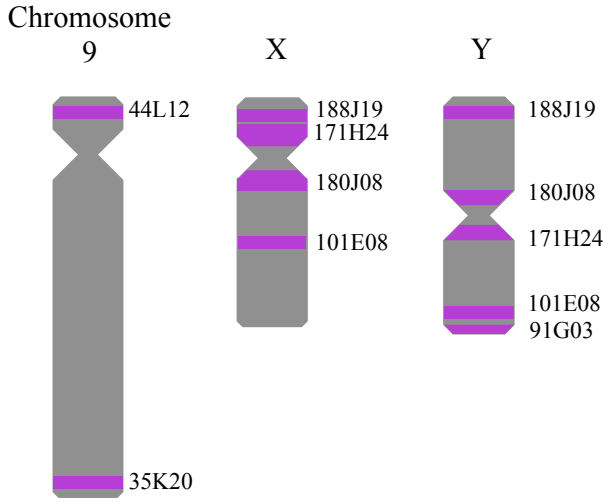
Interestingly, the ancestral Y chromosome has independently fused to the acrocentric chromosome 12 in males of the blackspotted stickleback (*Gasterosteus wheatlandi*) species within the past 15 million years, creating an independent neo-Y chromosome (Ross et al. 2009). From an evolutionary perspective, it is intriguing that of the seven known species of stickleback, two have undergone fixation of a Y-autosome fusion. Future work will test whether the Y chromosome centromere is also inactivated on the blackspotted neo-Y, leaving the ancestral chromosome12 centromere functional.

The Japan Sea neo-Y is one of the first examples in which centromere inactivation has been demonstrated as mechanism to stabilize a chromosome fusion in an evolutionary context. Because dicentrics are so detrimental, we hypothesize that there was likely a reason the fusion led to fixation. Further experiments (Chapter 4) will test for why this fusion was selected, and how it potentially led to the speciation of the Japan Sea sticklebacks from the Pacific Ocean ancestral population (Kitano et al. 2009). Another interesting evolutionary question would be to assess the reason for the fusion in the blackspotted stickleback. Because both the blackspotted and the Japan Sea species have a fusion between an autosome and the Y chromosome, comparing the genes found on chromosome 9 in Japan Sea and chromosome 12 in blackspotted sticklebacks would allow us to test for similar genes or gene families that might be play a role in the fusion.

Table 3-1: List of BACs used in this study

Chr	BAC	Library	Genomic position (bp)	Cytogenetic position*
9	44L12	CHORI-213	497,589 – 729,621	Distal end of short arm
9	35K20	CHORI-215	18,760,858 – 18,914,793	Distal end of long arm
X	188J19	CHORI-213	2,384,200 – 2,563,816	Distal end of short arm
Y	188J19	CHORI-213	unknown	Distal end of short arm
X	180J08	CHORI-213	3,377,138 – unknown	Proximal end of short arm
Y	180J08	CHORI-213	unknown	Proximal end of long arm
X	171H24	CHORI-213	5,465,262 – 5,643,381	Proximal end of long arm
Y	171H24	CHORI-213	unknown	Proximal end of short arm
X	101E08	CHORI-213	11,446,792 – 11,651,387	Middle of long arm
Y	101E08	CHORI-213	unknown	Distal end of long arm
Y	91G03	CHORI-215	unknown	Distal end of long arm

*Position is relative to the centromere; proximal is closer to the centromere and distal is further from the centromere. The genomic positions of the BACs on chromosome 9 and the X chromosome were determined by BLASTing the publicly available BAC end sequences (Kingsley and Peichel 2007) to the updated threespine stickleback genome assembly (Glazer et al. 2015). The genomic positions of BACs on the Y chromosome are unknown because a complete genome assembly of the Y chromosome is not available.



Chapter 3 Figures:

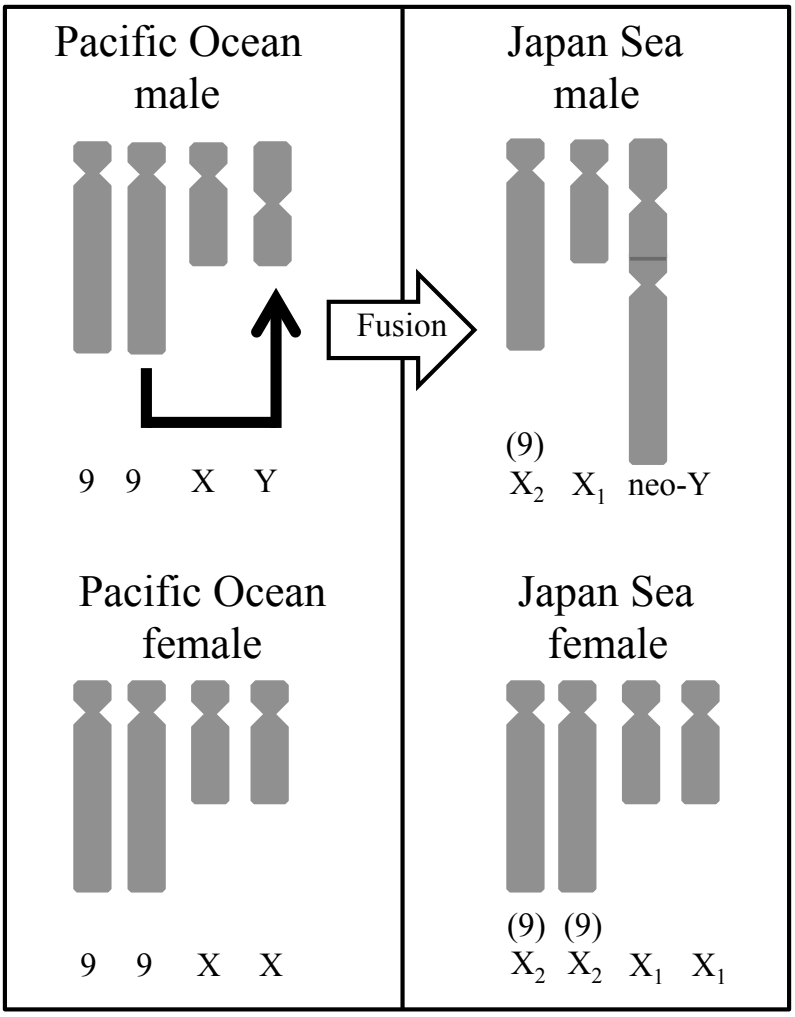


Figure 3-1. Pacific Ocean (ancestral) and Japan Sea (derived) chromosomes used in this study. A fusion between the acrocentric chromosome 9 and the metacentric Y chromosome from the ancestral Pacific Ocean species gave rise to the neo-Y chromosome in the Japan Sea species around two million years ago. The Japan Sea sticklebacks still retain the ancestral submetacentric X chromosome (now termed X_1), and the unfused acrocentric chromosome 9 (now termed X_2).

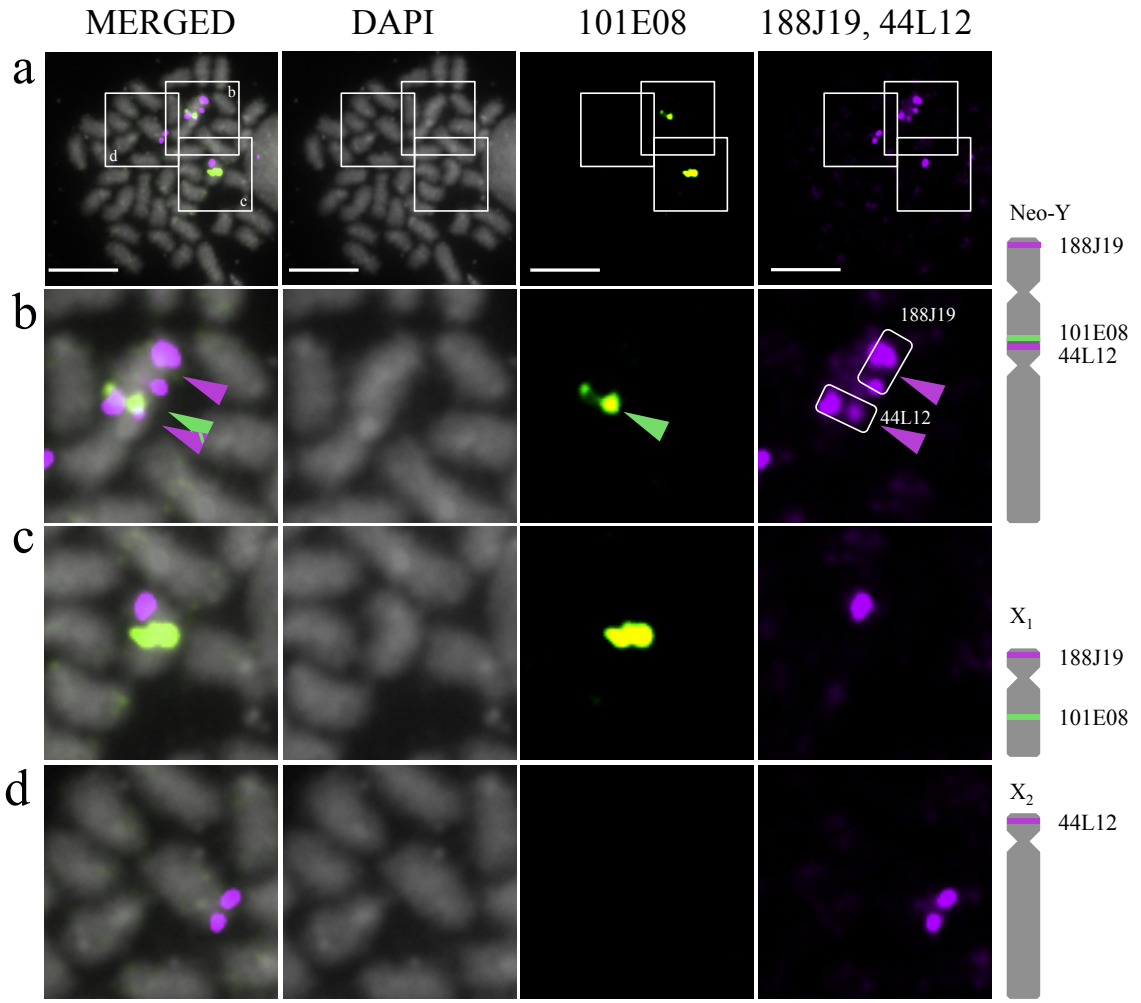


Figure 3-2. The Japan Sea neo-Y is a fusion between two complete chromosomes. FISH with the ancestral X and Y chromosome BACs 188J19 (*purple*) and101E08 (*green*) and the chromosome 9 BAC 44L12 (*purple*) on a Japan Sea male metaphase spread shows (**a**) three chromosomes with BAC hybridization each highlighted by square boxes. Higher magnification of the boxed regions in panel (**a**) shows the neo-Y fusion (**b**), the X_1 chromosome, which is derived from the ancestral X chromosome (**c**) and the X_2 chromosome, which is derived from the ancestral chromosome 9 (**d**). *Scale bar*, 5 μm

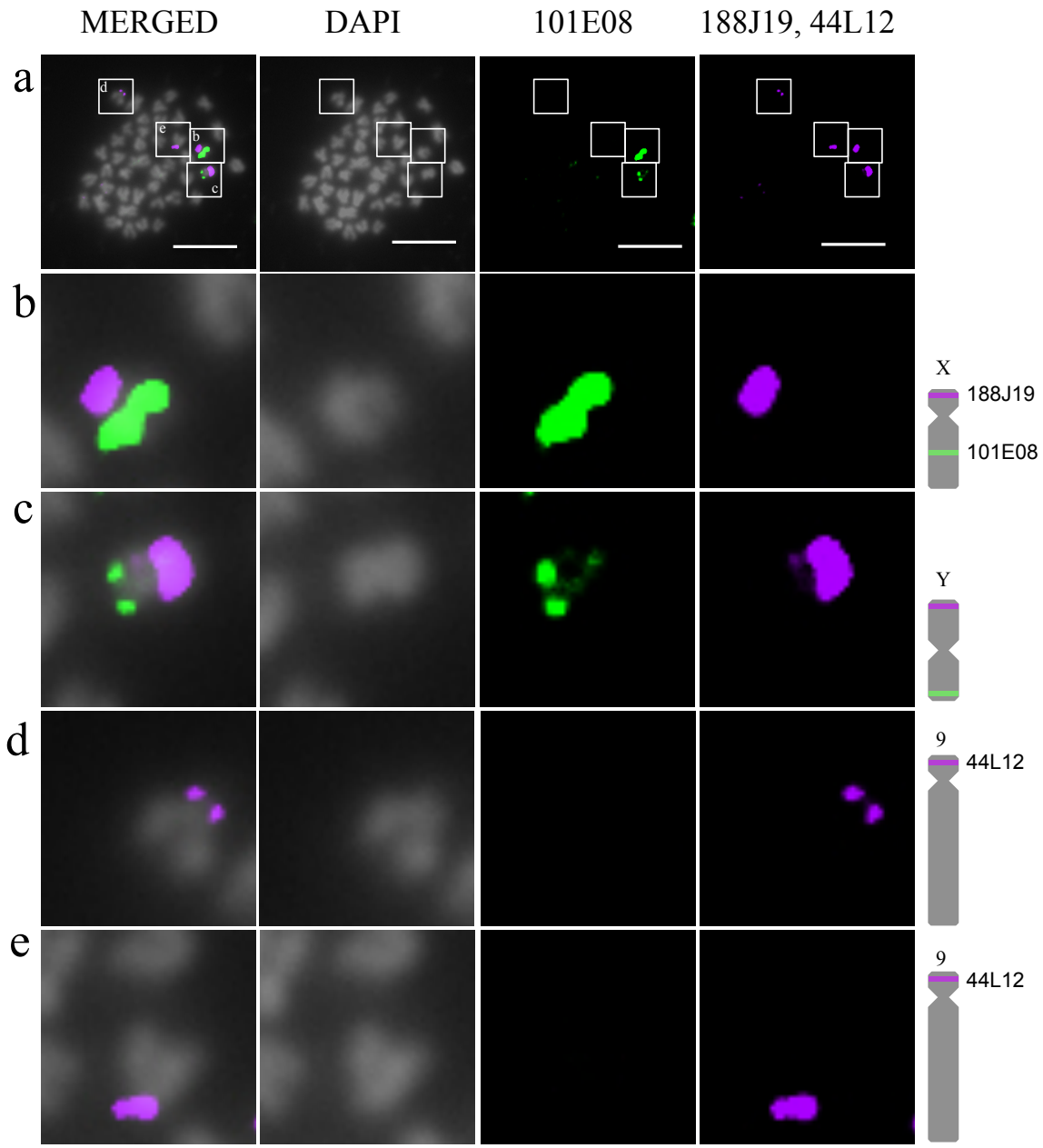


Figure 3-3. The ancestral state of chromosome 9, and the X and Y in Pacific Ocean stickleback fish. FISH with the ancestral X and Y chromosome BACs 188J19 (*purple*) and101E08 (*green*) and the chromosome 9 BAC 44L12 (*purple*) on a Pacific Ocean male metaphase spread shows (**a**) the two unfused chromosome 9s, the X chromosome, and the unfused Y chromosome each highlighted with a square box. Higher magnification of the boxed regions in panel (**a**) shows the X chromosome with two regions of BAC hybridization(**b**), the unfused Y chromosome with two regions of BAC hybridization (**c**), and the two unfused chromosome 9s (**d**, **e**). *Scale bar*, 5 µm

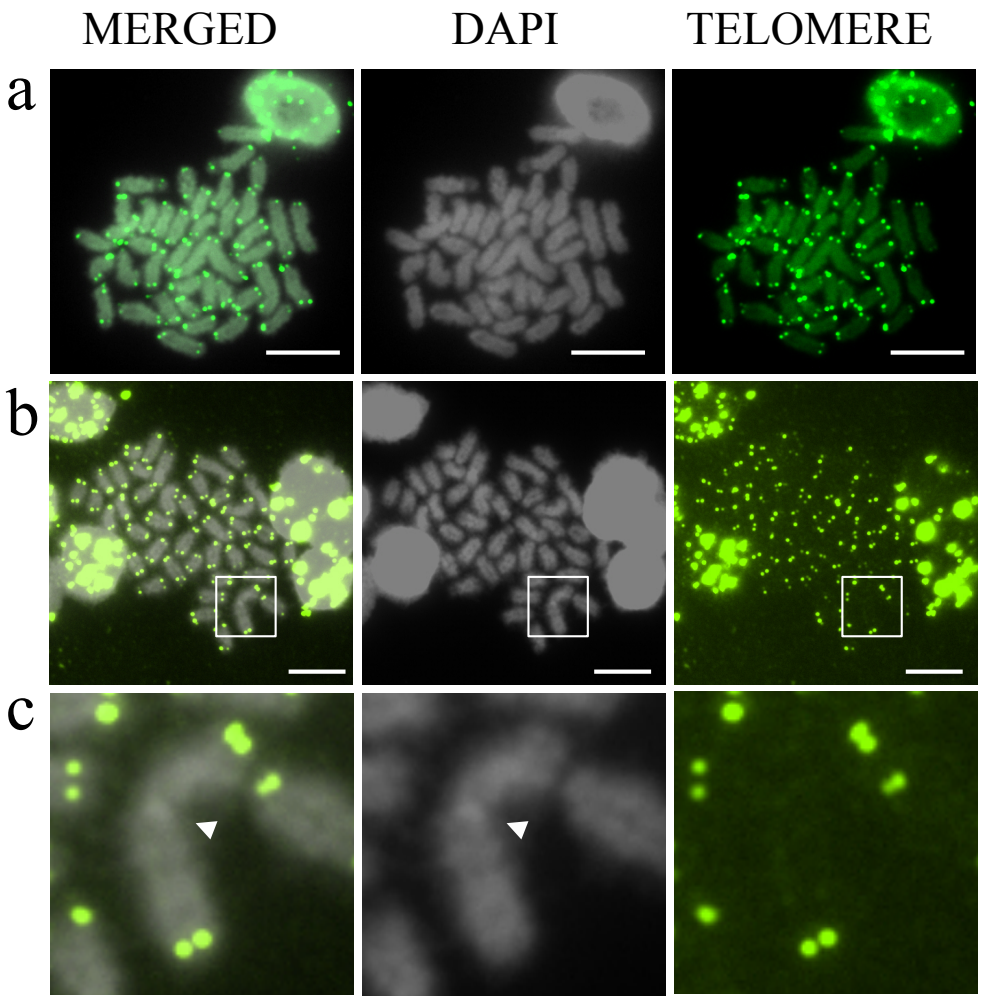


Figure 3-4. Telomere staining in Japan Sea male and female metaphase spreads. Telomere staining is seen on the ends of chromosomes in (**a**) Japan Sea female metaphase chromosomes and (**b**) Japan sea male metaphase chromosomes. The neo-Y is the largest chromosome and is highlighted by a box in panel (**b**). Panel (**c**) shows a higher magnification view of the neo-Y with no internal telomere signal. The primary centromeric constriction on the neo-Y is indicated by the white arrowhead. *Scale bar*, 5 μm

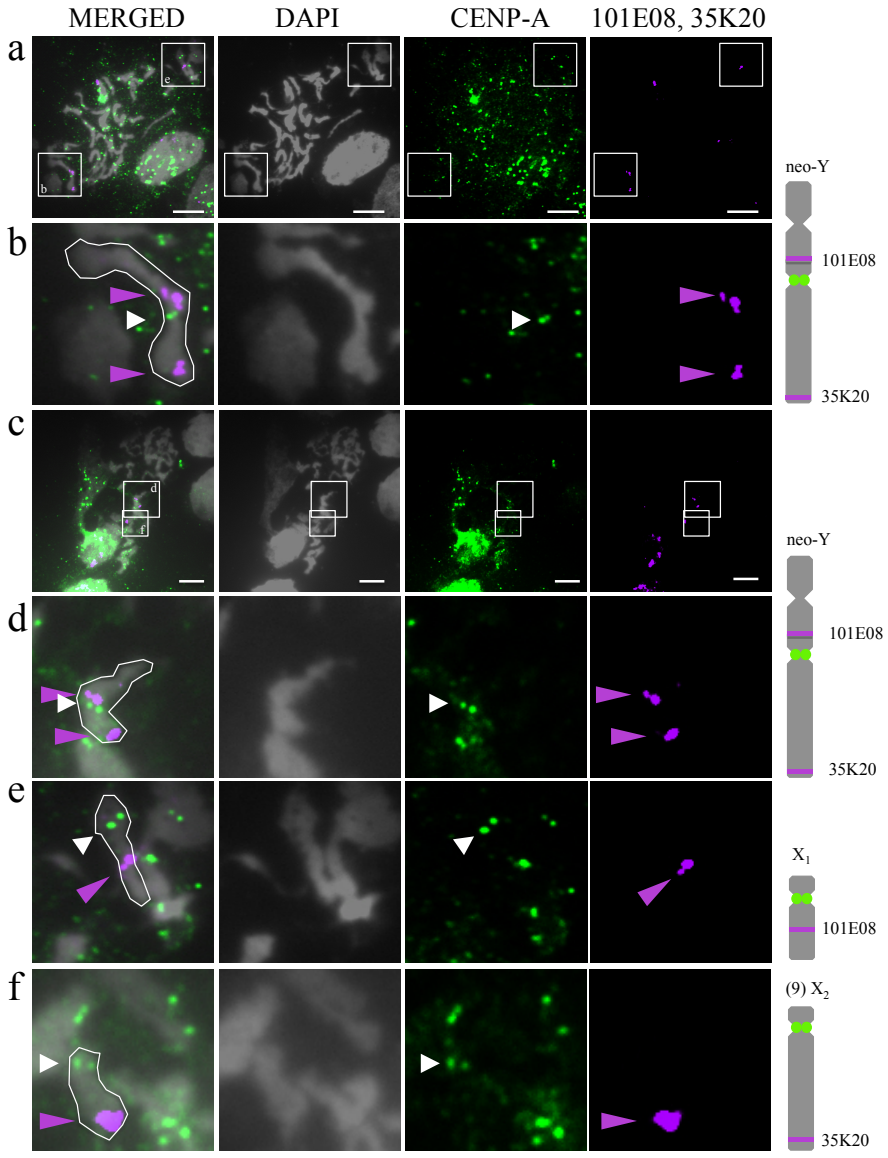


Figure 3-5. The CENP-A antibody only localizes to one region on the Japan sea neo-Y chromosome. Panels (**a**) and (**c**) show two independent metaphase spreads from Japan Sea embryos stained with the CENPA antibody (*green*) as well as the X and Y chromosome BAC 101E08 (*purple*) and the chromosome 9 BAC 35K20 (*purple*). Panels (**b**) and (**d**) are higher magnifications of the boxed regions in panels (**a**) and (**c**)**,** highlighting the neo-Y chromosome with the two regions of BAC staining (*purple arrowheads*) flanking two distinct CENP-A puncta (*white arrowhead*). The CENP-A staining is located on the chromosome 9-derived part of the neo-Y chromosome. Panel (**e**) is a higher magnification of the boxed region in (**a**), highlighting the ancestral X_1 chromosome; 101E08 hybridizes to the middle of the long arm of X_1 (*purple arrowhead*), and there is one region of two CENP-A puncta (*white arrowhead*). Panel (**f**) is a higher magnification of the boxed region in (**c**), highlighting the unfused chromosome 9 (X_2); 35K20 hybridizes to the end of the long arm of X_2 (*purple arrowhead*), and there is one region of two CENP-A puncta (*white arrowhead*). *Scale bar*, 5 μm

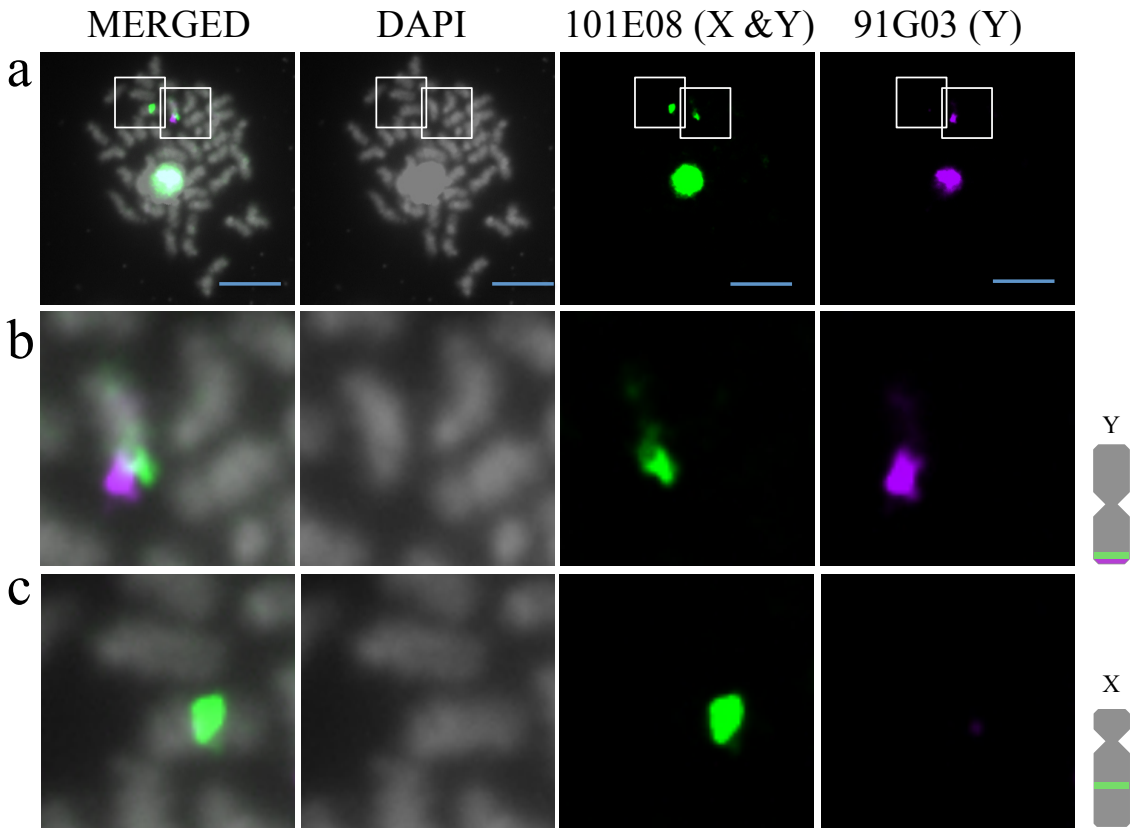


Figure 3-6. The BAC clone 91G03 is a Y specific BAC. (**a**) FISH was performed on a Pacific Ocean male metaphase spread with BACs 101E08 (*green*), and 91G03 (*purple*). 101E08 hybridizes to the X and Y chromosome, while 91G03 only hybridizes to the Y. Panel (**b**) is a magnification of the Y chromosome from (**a**) showing hybridization of the known sex chromosome BAC 101E08 and BAC 91G03 to the very end of the Y chromosome. Panel (**c**) is a magnification of the X chromosome from (**a**), with hybridization of BAC 101E08 to the middle of the long arm, and no hybridization of BAC 91G03. *Scale bar*, 5 μm

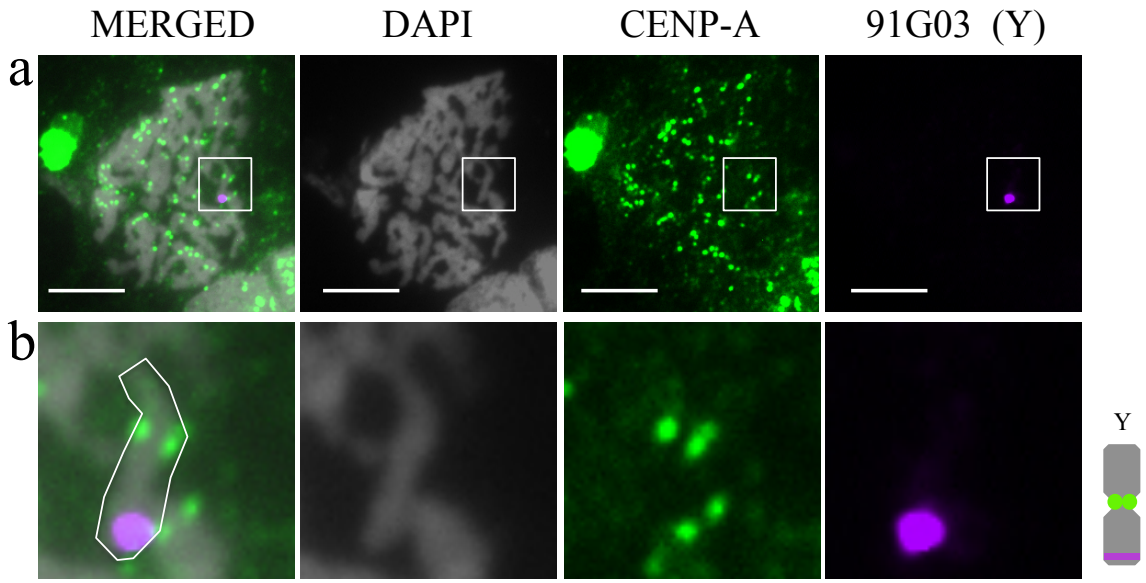


Figure 3-7. CENP-A antibody staining on the ancestral Pacific Ocean Y chromosome. (**a**) A metaphase spread from Pacific Ocean embryos was stained with the CENPA antibody (*green*) and the Y chromosome specific BAC 91G03 (*purple*). Panel (**b**) is a magnification of the boxed region in panel (**a**), highlighting the Y chromosome with 91G03 staining on the end of the long arm, and two distinct CENP-A puncta hybridizing to the centromere of each sister chromatid on the Y chromosome. *Scale bar*, 5 µm

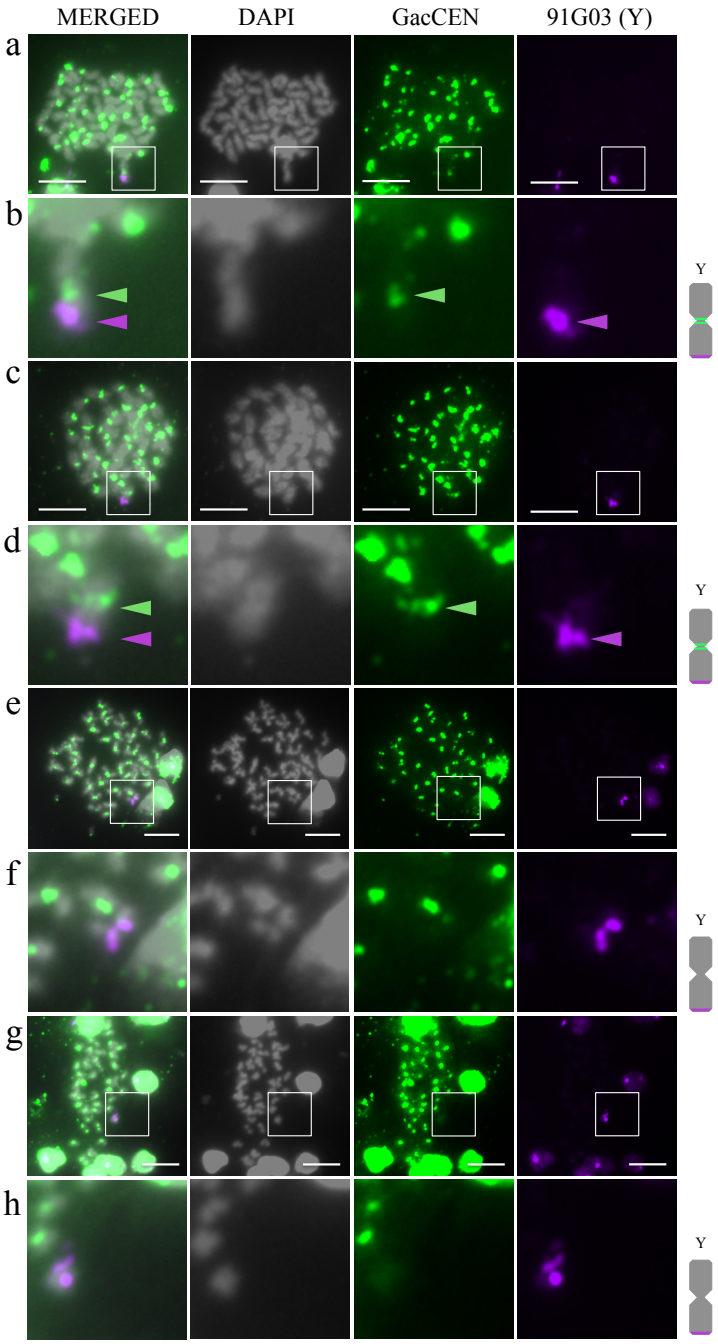


Figure 3-8. GacCEN staining is weak and variable on the ancestral Pacific Ocean male Y chromosome centromere. FISH with the GacGEN probe (*green*) and Y chromosome BAC 91G03 (*purple*) on four different metaphase spreads from the same Pacific Ocean male. Two metaphase spreads show GacCEN staining on the Y chromosome (**a-d**) while two metaphase spreads lack GacCEN staining on the Y chromosome (**e-h**). Panels (**b**, **d**) show higher magnification of the Y chromosome with the weak GacCEN staining (*green arrowhead*), and panels (**f**, **h**) show higher magnification of the Y chromosome with no GacCEN staining. *Scale bar*, 5 μm

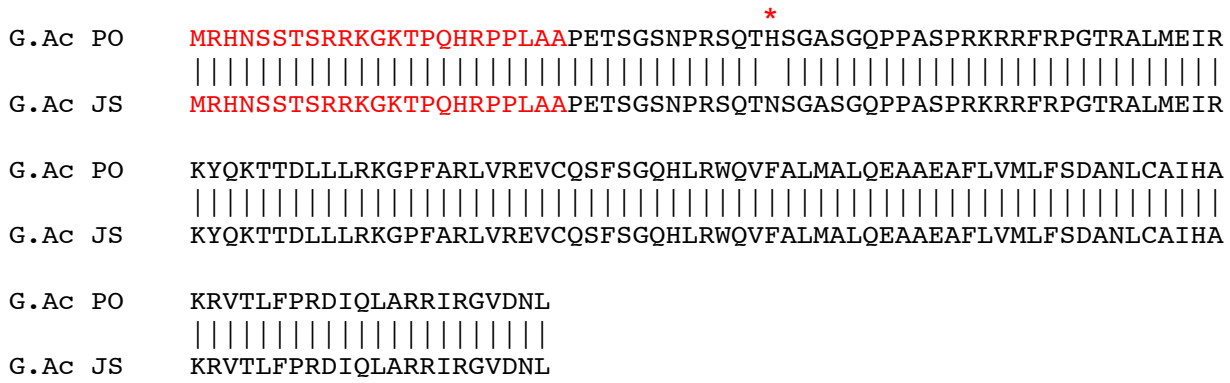


Figure 3-9. Comparison of the CENP-A protein amino acid sequence between the Pacific Ocean (PO) and Japan Sea (JS) sticklebacks. There is only one amino acid difference (red asterisk) between the two proteins, which is not in the amino acid sequence targeted by the CENP-A antibody (red letters).

Figure 3-10. Comparison of the CENP-A associated centromeric sequence between the Pacific Ocean (PO) and Japan Sea (JS) stickleback species. The Japan Sea consensus sequence is 98.9% similar to the GacCEN sequence previously identified in the Pacific Ocean species (Cech and Peichel 2015). The red asterisk denotes the only three nucleotide differences between these two consensus sequences. Nucleotide ambiguities: Y = C or T; R = A or G.

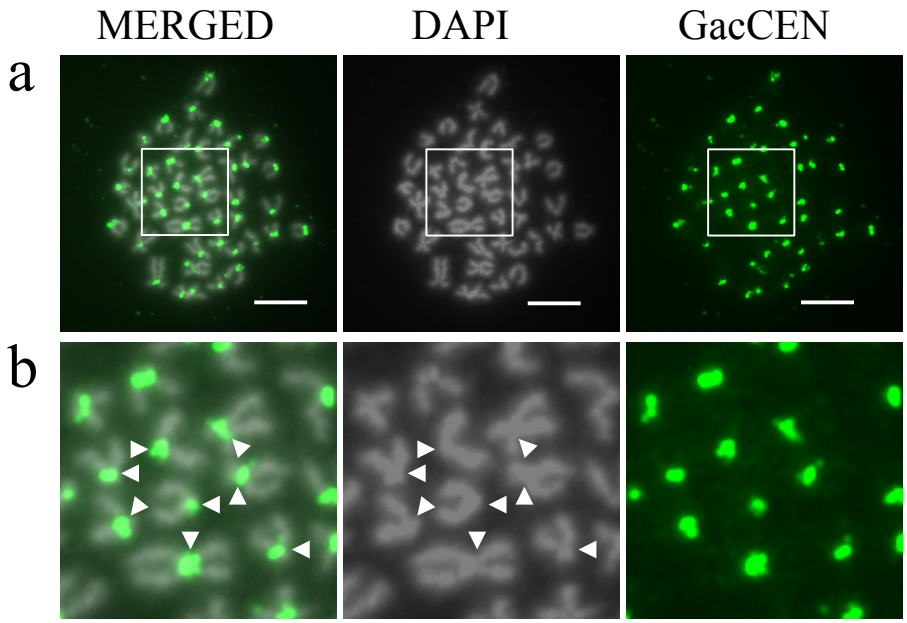


Figure 3-11. The GacCEN probe hybridizes to the centromere on Japan Sea chromosomes. (**a**) The GacCEN probe hybridizes to a single region on each chromosome in a metaphase spread from a Japan Sea male. Panel (**b**) shows a magnification of the boxed region in (**a**), highlighting the hybridization of the GacCEN probe to the primary constriction (*white arrowheads*) on each chromosome. *Scale bar*, 5 μm

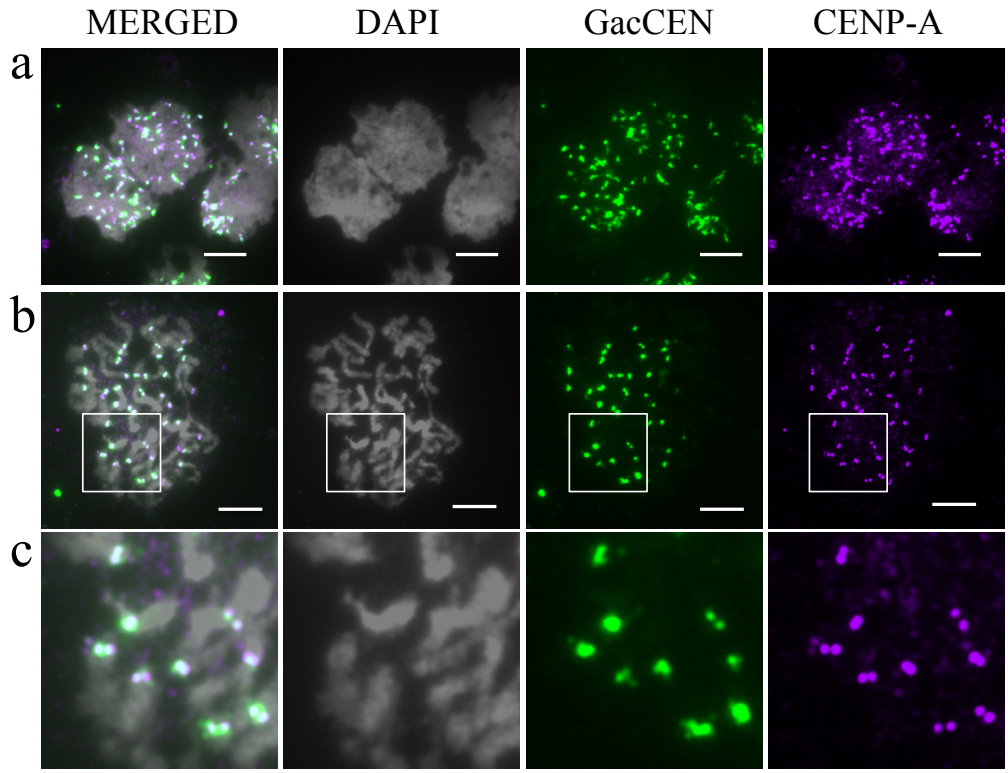


Figure 3-12. The GacCEN probe colocalizes with CENP-A on Japan Sea chromosomes. The GacCEN probe (*green*) colocalizes with the CENP-A antibody (*purple*) at distinct puncta in interphase nuclei (**a**) as well as to a single region on each chromosome in a metaphase spread from a Japan Sea embryo (**b**). Panel (**c**) shows a magnification of the metaphase spread shown in (**b**). *Scale bar*, 5 μm

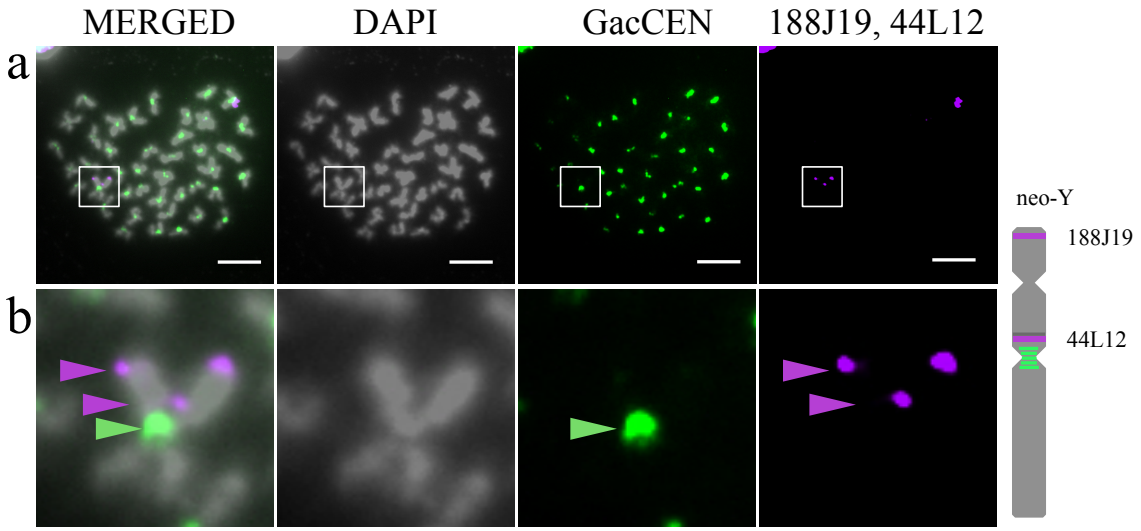


Figure 3-13. GacCEN hybridizes to the centromere of chromosome 9 on the Japan Sea neo-Y chromosome. FISH with the GacCEN probe (*green*), Y chromosome BAC probe 188J19 (*purple*), and a chromosome 9 BAC probe 44L12 (*purple*) on a metaphase spread from a Japan Sea male is shown in panel (**a**). Panel (**b**) is a magnification of the neo-Y chromosome from the boxed region in panel (**a**). The GacGEN probe only localizes to the region of the ancestral chromosome 9 centromere on the neo-Y. *Scale bar*, 5 μm

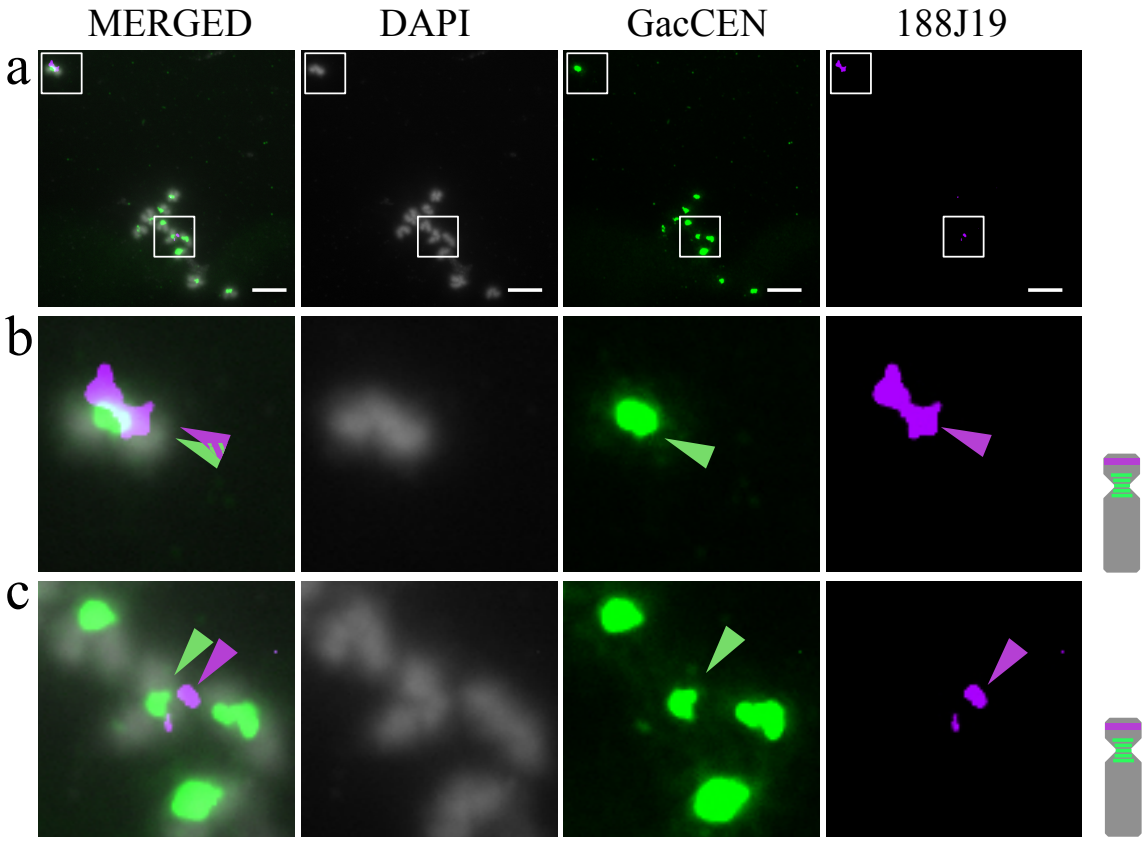


Figure 3-14. Both submetacentric X_1 chromosomes in the Japan Sea female show strong GacCEN hybridization. FISH with an X chromosome BAC 188J19 (*purple*), and GacCEN (*green*) on a Japan Sea female metaphase spread in shown in panel (**a**). Panels (**b**) and (**c**) are magnifications of the boxed regions in panel (**a**), showing the two ancestral X_1 chromosomes, with strong GacCEN staining (green arrowhead) consistent with the submetacentric centromere positioning. *Scale bar*, 5 μm

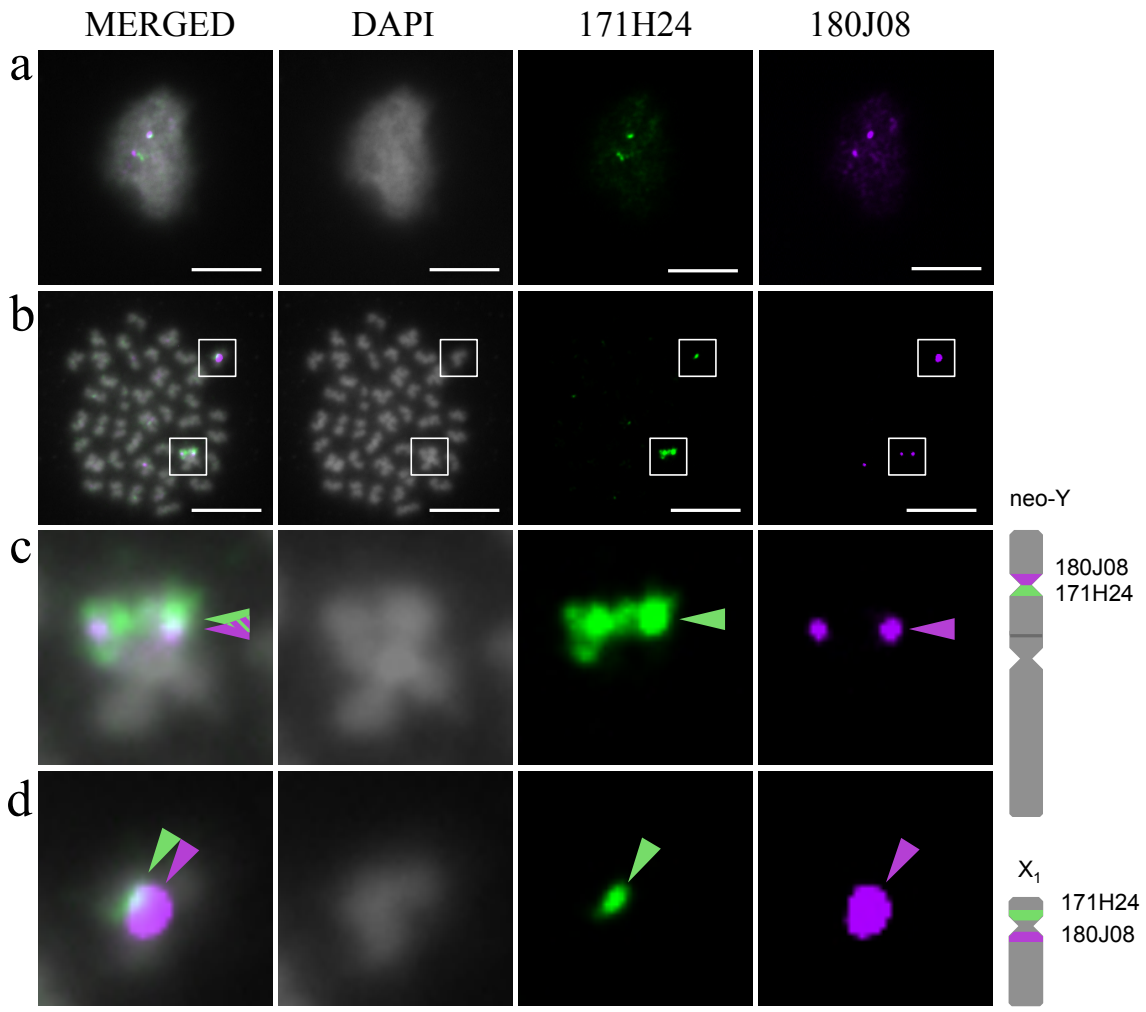


Figure 3-15. Mapping the regions flanking the ancestral Y centromere on the neo-Y fusion. FISH with the X and Y BACs 180J08 (*purple*) and 171H24 (*green*) on Japan Sea male interphase nuclei (**a**) and metaphase spreads (**b**). Both probes are present on the neo-Y (**c**) and the ancestral X_1 (**d**), although the probes appear to be closer together on the neo-Y (**c**) as compared to the ancestral X_1 (**d**)**.** *Scale bar*, 5 µm

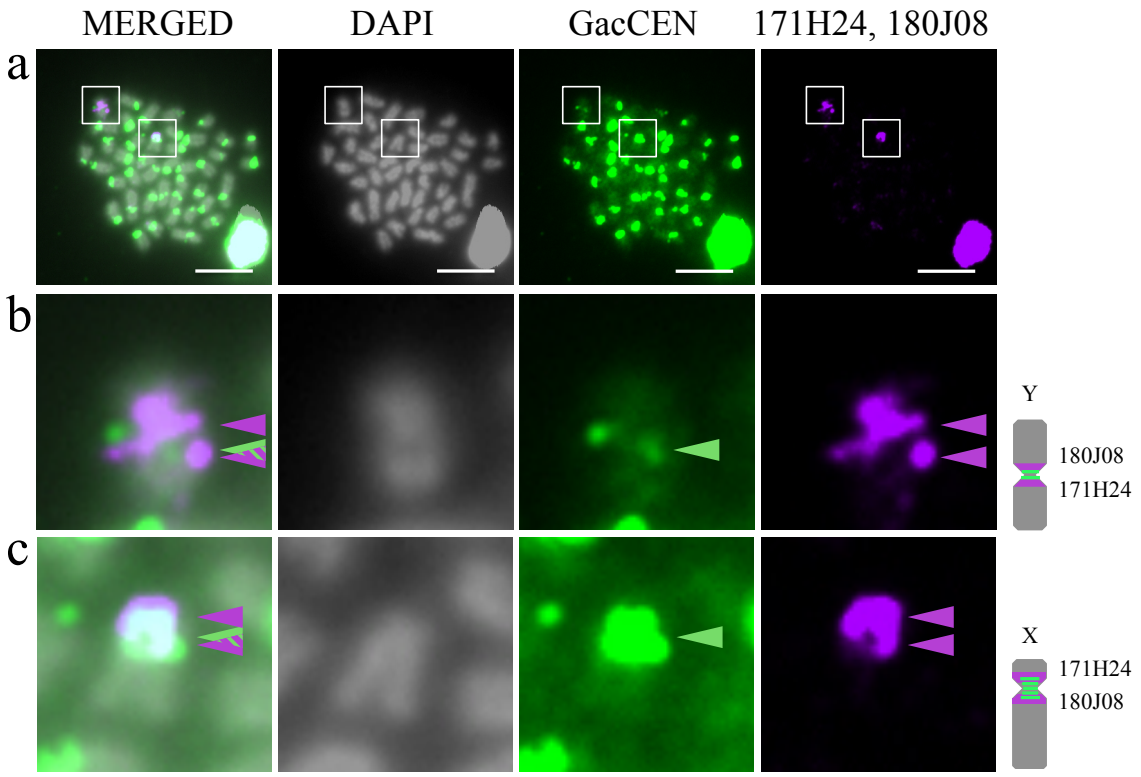


Figure 3-16. GacCEN and centromere flanking regions on the Pacific Ocean X and Y chromosomes. FISH with the X and Y BACs 180J08 and 171H24 (*purple*) and GacCEN (*green*) on a metaphase spread from a Pacific Ocean male is shown in panel (**a**). Panels (**b**) and (**c**) are magnifications of the boxed regions in panel (**a**) showing two distinct regions of staining (*purple arrowheads*) flanking weak GacCEN staining (*green arrowhead*) on the Y (**b**), and two distinct regions of staining (*purple arrowheads*) flanking strong GacCEN staining (*green arrowhead*) on the X (**c**). *Scale bar*, 5 µm

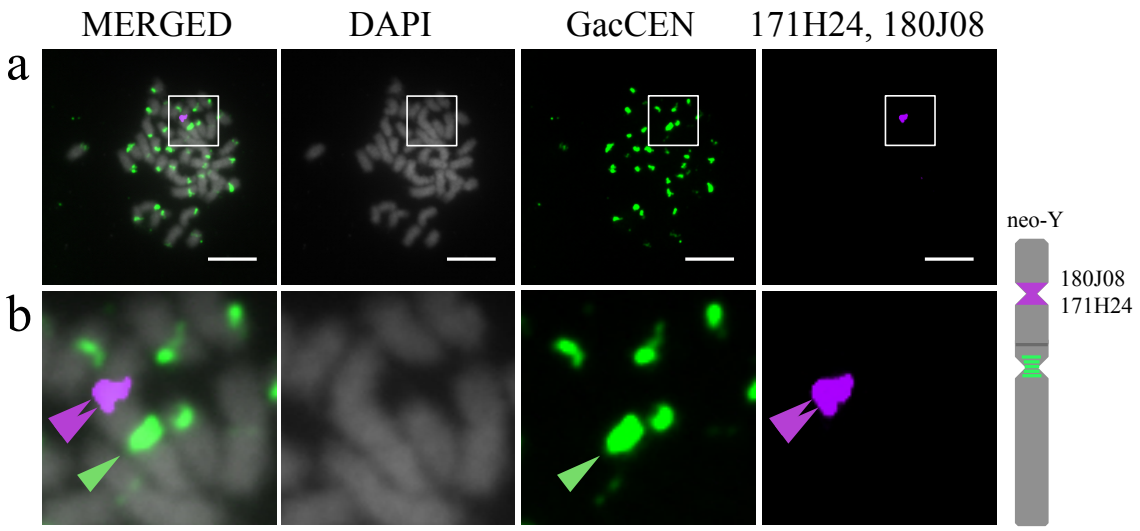


Figure 3-17. GacCEN and centromere flanking regions on the Japan Sea neo-Y chromosome. FISH with the Y BACs 180J08 and 171H24 (*purple*) and GacCEN (*green*) on a metaphase spread from a Japan Sea male is shown in panel (**a**). Panel (**b**) is a magnification of the boxed regions in panel (**a**) showing the two centromere flanking probes (*purple arrowheads*) are very close together on the neo-Y with no GacCEN staining in between. The strong GacCEN staining (*green arrowhead*) is from the chromosome 9 centromere on the neo-Y. *Scale bar*, 5 μm

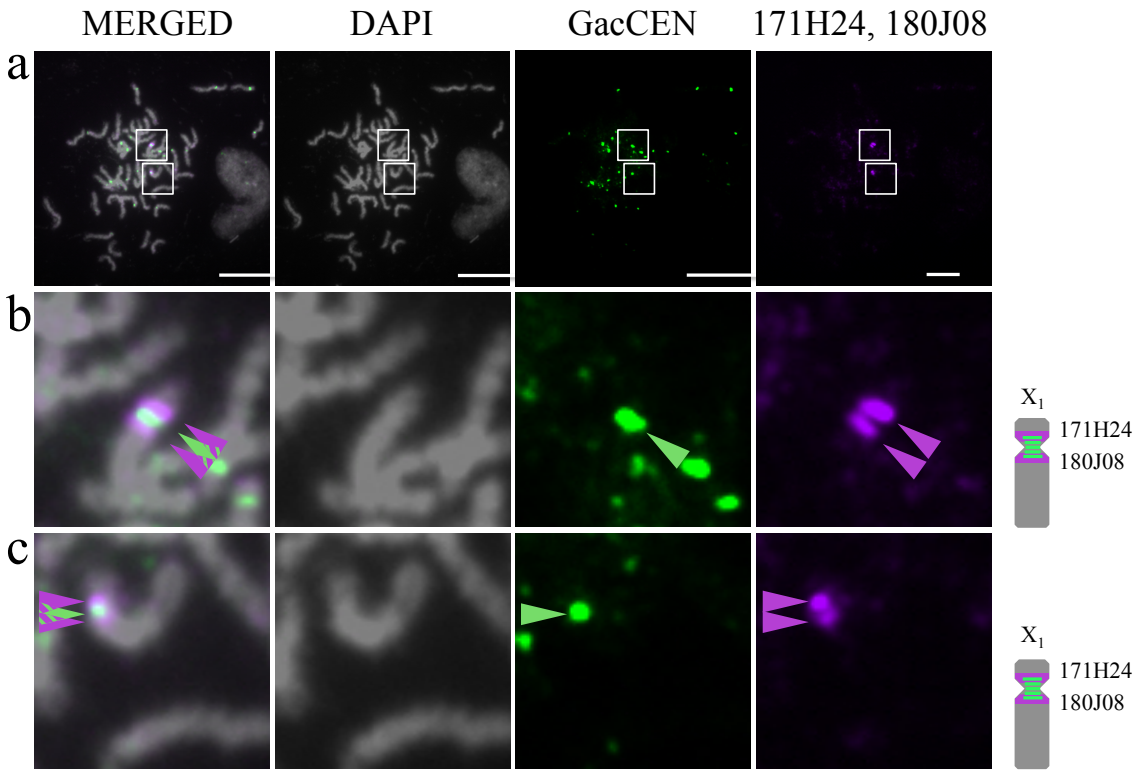


Figure 3-18. GacCEN and centromere flanking regions on Japan Sea female X_1 chromosomes. FISH with the X and Y BACs 180J08 and 171H24 (*purple*) and GacCEN (*green*) on a metaphase spread from a Japan Sea female is shown in panel (**a**). Panels (**b**) and (**c**) are magnifications of the boxed regions in (**a**) showing two distinct regions of BAC hybridization (*purple arrowheads*) flanking strong GacCEN staining (*green arrowhead*) on both X_1 chromosomes. *Scale bar*, 5 μm

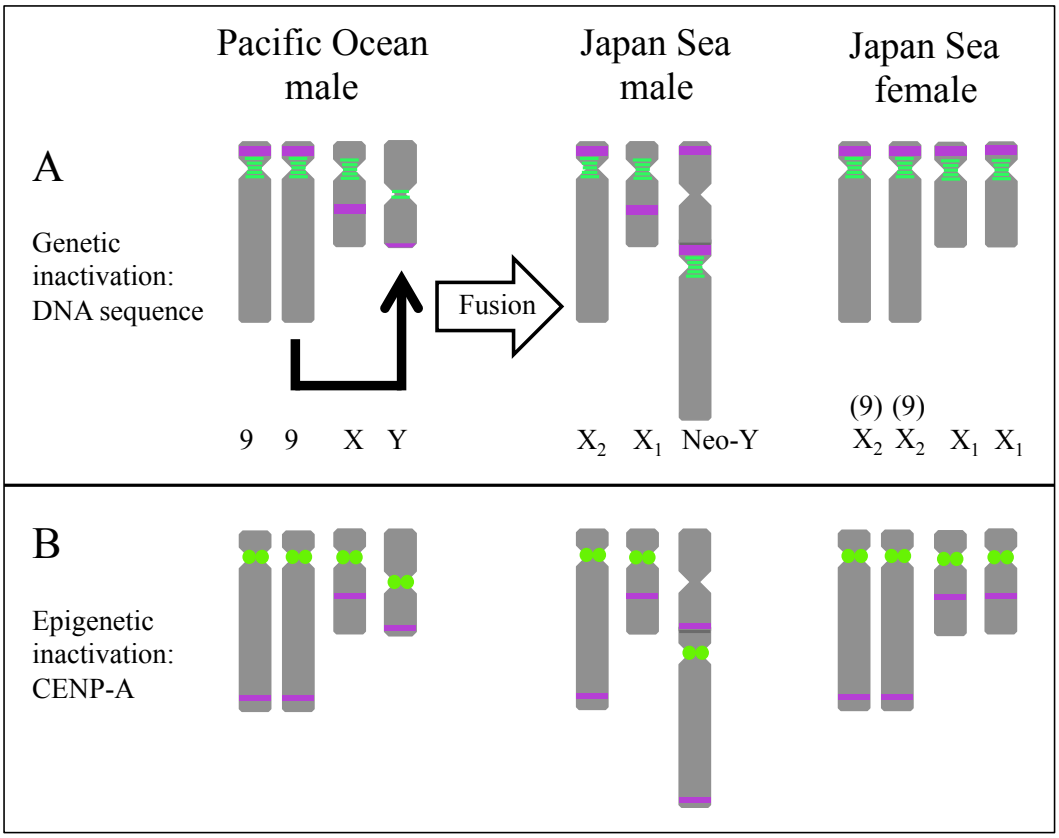


Figure 3-19. Summary of GacCEN and CENP-A staining. Strong GacCEN (*green dashed lines*) staining is present on chromosome 9 and the X chromosome in Pacific Ocean males. We find evidence for weak and variable GacCEN staining on the Pacific Ocean Y centromere. The Japan Sea sticklebacks also show strong GacCEN staining on the unfused chromosome 9 (X_2) and the ancestral X_1, but only one region of strong GacCEN staining at the chromosome 9 centromere in the neo-Y fusion. Similarly, we see CENP-A staining (*green dots*) on the ancestral chromosome 9, and both the X and Y chromosomes in Pacific Ocean males. The Japan Sea sticklebacks also show CENP-A staining on the unfused chromosome 9 (X_2) and the ancestral X_1, but only one region of strong CENP-A staining at the chromosome 9 centromere in the neo-Y fusion. Together this is evidence for centromere inactivation of the ancestral Y centromere on the neo-Y chromosome. Purple lines indicate BAC probes used for FISH.

Chapter 3 References:

Agudo M, Abad JP, Molina I, et al (2000) A dicentric chromosome of Drosophila melanogaster showing alternate centromere inactivation. Chromosoma 109:190–196

Alkan C, Cardone MF, Catacchio CR, et al (2011) Genome-wide characterization of centromeric satellites from multiple mammalian genomes. Genome Research 21:137–145

Allshire RC, Karpen GH (2008) Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet 9:923–937

Avarello R, Pedicini A, Caiulo A, et al (1992) Evidence for an ancestral alphoid domain on the long arm of human chromosome 2. Human Genetics 89:247–249

Bailey SM (2006) Telomeres, chromosome instability and cancer. Nucleic Acids Res 34:2408–2417

Barnhart MC, Kuich PHJL, Stellfox ME, et al (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J Cell Biol 194:229–243

Cech JN, Peichel CL (2015) Identification of the centromeric repeat in the threespine stickleback fish (Gasterosteus aculeatus). Chromosome Res 23:767-779

Crosland MW, Crozier RH (1986) Myrmecia pilosula, an ant with only one pair of chromosomes. Science 231:1278–1278

Ferguson-Smith MA, Trifonov V (2007) Mammalian karyotype evolution. Nat Rev Genet 8:950–962

Fisher AM, Al-Gazali L, Pramathan T, et al (1997) Centromeric inactivation in a dicentric human Y;21 translocation chromosome. Chromosoma 106:199–206

Flemming W (1880). Beitrage zur Kenntniss der Zelle und ihrer Lebenserscheinungen Theil II (in German). Archivfiir Mikroskopische Anatomie 18: 151-259.

Foltz DR, Jansen LET, Black BE, et al (2006) The human CENP-A centromeric nucleosome-associated complex. Nat Cell Biol 8:458–469

Fu L, Niu B, Zhu Z et al (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28: 3150–3152

Fukagawa T, Earnshaw WC (2014) Neocentromeres. Current Biology 24:R946–R947

Gallardo MH, Bickham JW, Honeycutt RL, et al (1999) Discovery of tetraploidy in a mammal. Nature 401:341

Gisselsson D, Jonson T, Petersén A, et al (2001) Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. Proc Natl Acad Sci USA 98:12683–12688

Gisselsson D, Pettersson L, Höglund M, et al (2000) Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. Proc Natl Acad Sci USA 97:5357–5362

Glazer AM, Killingbeck EE, Mitros T, et al (2015) Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. G3 (Bethesda) 5:1463–1472

Gong Z, Wu Y, Koblizkova A, et al (2012) Repeatless and repeat-based centromeres in potato: implications for centromere evolution. Plant Cell 24:3559–3574

Han F, Lamb JC, Birchler JA (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. Proc Natl Acad Sci USA 103:3238–3243

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098–1102

Heun P, Erhardt S, Blower MD, et al (2006) Mislocalization of the drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Dev Cell 10:303–315

Higgins AW, Gustashaw KM, Willard HF (2005) Engineered human dicentric chromosomes show centromere plasticity. Chromosome Res 13:745–762

IJdo JW, Baldini A, Ward DC, et al (1991) Origin of human chromosome 2: an ancestral telomere-telomere fusion. Proc Natl Acad Sci USA 88:9051–9055

Kingsley DM, Peichel CL (2007) The molecular genetics of evolutionary change in sticklebacks. In: Östlund-Nilsson S, Mayer I, Huntingford FA (eds) Biology of the Three-Spined Stickleback. CRC press, Boca Raton, pp 41–81

Kitano J, Ross JA, Mori S, et al (2009) A role for a neo-sex chromosome in stickleback speciation. Nature 461:1079–1083

Koo DH, Han F, Birchler JA, Jiang J (2011) Distinct DNA methylation patterns associated with active and inactive centromeres of the maize B chromosome. Genome Research 21:908–914

Koshland D, Rutledge L, Fitzgerald-Hayes M, Hartwell LH (1987) A genetic analysis of dicentric minichromosomes in Saccharomyces cerevisiae. Cell 48:801–812

Lange J, Skaletsky H, van Daalen SKM, et al (2009) Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. Cell 138:855–869

Lejeune J, Dutrillaux B, Rethoré MO, Prieur M (1973) Comparaison de la structure fine des chromatides d'Homo sapiens et de Pan troglodytes. Chromosoma 43:423-444.

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659

Liu Y, Su H, Zhang J, et al (2015) Dynamic epigenetic states of maize centromeres. Front Plant Sci 6:589

MacKinnon RN, Campbell LJ (2011) The role of dicentric chromosome formation and secondary centromere deletion in the evolution of myeloid malignancy. Genetics Research International 2011:1–11

Maraschio P, Zuffardi O, Caiulo A, et al (1990) Deletion of specific sequences or modification of centromeric chromatin are responsible for Y chromosome centromere inactivation. Human Genetics 85:491–494

Marshall OJ, Chueh AC, Wong LH, Choo KHA (2008) Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. The American Journal of Human Genetics 82:261–282

Maser RS, DePinho RA (2002) Connecting chromosomes, crisis, and cancer. Science 297:565–569

McClintock B (1939) The behavior in successive nuclear divisions of a chromosome broken at meiosis. Proc Natl Acad Sci USA 25:405–416

McClintock B (1941) The stability of broken ends of chromosomes in Zea Mays. Genetics 26:234–282

Melters DP, Bradnam KR, Young HA, et al (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14:R10

Mendiburo MJ, Padeken J, Fulop S, et al (2011) Drosophila CENH3 Is sufficient for centromere formation. Science 334:686–690

Meyne J, Ratliff RL, Moyzis RK (1989) Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci USA 86:7049–7053

Murnane JP (2010) Telomere loss as a mechanism for chromosome instability in human cancer. Cancer Research 70:4255–4259

Nagaki K, Cheng Z, Ouyang S, et al (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145

Olszak AM, van Essen D, Pereira AJ, et al (2011) Heterochromatin boundaries are hotspots for de novo kinetochore formation. Nat Cell Biol 13:799–808

Page SL, Shaffer LG (1998) Chromosome stability is maintained by short intercentromeric distance in functionally dicentric human Robertsonian translocations. Chromosome Res 6:115–122

Palmer DK, O'Day K, Trong HL, et al (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci USA 88:3734–3738

Palmer DK, O'Day K, Wener MH, et al (1987) A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J Cell Biol 104:805–815

Palmer DKD, O'Day KK, Margolis RLR (1989) Biochemical analysis of CENP-A, a centromeric protein with histone-like properties. Prog Clin Biol Res 318:61–72

Pertile MD, Graham AN, Choo KHA, Kalitsis P (2009) Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability. Genome Research 19:2202–2213

Piras FM, Nergadze SG, Magnani E, et al (2010) Uncoupling of satellite DNA and centromeric function in the genus equus. PLoS Genet 6:e1000845

Pobiega S, Marcand S (2010) Dicentric breakage at telomere fusions. Genes & Development 24:720–733

Ross JA, Peichel CL (2008) Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. Genetics 179:2173–2182

Ross JA, Urton JR, Boland J, et al (2009) Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). PLoS Genet 5:e1000391

Sato H, Masuda F, Takayama Y, et al (2012) Epigenetic inactivation and subsequent heterochromatinization of a centromere stabilize dicentric chromosomes. Current Biology 22:658-667

Sears ER, Câmara A (1952) A transmissible dicentric chromosome. Genetics 37:125–135

Sekulic N, Black BE (2012) Molecular underpinnings of centromere identity and

maintenance. Trends in Biochemical Sciences 1–10

Shang W-H, Hori T, Martins NMC, et al (2013) Chromosome engineering allows the efficient isolation of certebrate neocentromeres. Dev Cell 24:635–648

Shang WH, Hori T, Toyoda A, et al (2010) Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. Genome Research 20:1219–1228

Shi L, Ye Y, Duan D (1980) Comparative cytogenetic studies on the red muntjac, Chinese muntjac, and their F1 hybrids. Cytogenet Cell Genet 26:22-27

Silva M, Jansen L (2009) At the right place at the right time: novel CENP-A binding proteins shed light on centromere assembly. Chromosoma 118:567–574

Stimpson KM, Matheny JE, Sullivan BA (2012) Dicentric chromosomes: unique models to study centromere function and inactivation. Chromosome Res 20:595–605. doi: 10.1007/s10577-012-9302-3

Stimpson KM, Song IY, Jauch A, et al (2010) Telomere disruption results in non-random formation of de novo dicentric chromosomes involving acrocentric human chromosomes. PLoS Genet 6:e1001061

Sullivan BA, Willard HF (1998) Stable dicentric X chromosomes with two functional centromeres. Nat Genet 20:227–228

Sullivan KF, Hechenberger M, Masri K (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol 127:581–592

Tek AL, Kashihara K, Murata M, Nagaki K (2010) Functional centromeres in soybean include two distinct tandem repeats and a retrotransposon. Chromosome Res 18:337–347

Therman E, Susman B, Denniston C (1989) The nonrandom participation of human acrocentric chromosomes in Robertsonian translocations. Ann Hum Genet 53:49–65

Urton JR, McCann SR, Peichel CL (2011) Karyotype differentiation between two stickleback species (Gasterosteidae). Cytogenet Genome Res 135:150–159

Van der Burg WJ (2004). Ophioglossum Reticulatum. In: Grubben, GJH and Denton OA (Editors). Plant Resources of Tropical Africa 2. Vegetables. PROTA Foundation, Wageningen, Netherlands / Backuys Publishers, Leiden, Netherlands / CTA, Wageningen, Netherlands. pp 404-405

Voullaire LE, Slater HR, Petrovic V (1993) A functional marker centromere with no

detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? American Journal of Human Genetics 52:1153-1163

Wang W, Lan H (2000) Rapid and parallel chromosomal number reductions in muntjac deer inferred from mitochondrial DNA phylogeny. Molecular Biology and Evolution 17:1326–1333

Warburton PE (2004) Chromosomal dynamics of human neocentromere formation. Chromosome Res 12:617–626

Warburton PE, Cooke CA, Bourassa S, et al (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Current Biology 7:901–904

Wolfe J, Darling SM, Erickson RP, et al (1985) Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. J Mol Biol 182:477–485

Yang F, Carter NP, Shi L, Ferguson-Smith MA (1995) A comparative study of karyotypes of muntjacs by chromosome painting. Chromosoma 103:642–652

Zhang W, Friebe B, Gill BS, Jiang J (2010) Centromere inactivation and epigenetic modifications of a plant chromosome with three functional centromeres. Chromosoma 119:553–563

Chapter 4: Why did the neo-Y chromosome fusion evolve in the Japan Sea stickleback?

Introduction

The Japan Sea neo-Y fusion described in Chapter 3 is not only interesting from the perspective of understanding the mechanisms that underlie centromere inactivation, but also raises interesting questions about the evolutionary forces that lead to chromosomal fusions. In experiments where dicentric chromosome fusions are artificially created, less than 0.5% of the cells survive, emphasizing the detrimental affect of dicentric chromosome fusions (Sato et al. 2012). Yet the neo-Y fusion has been fixed in the Japan Sea stickleback species. We have already shown that this fusion was stabilized by centromere inactivation. Thus, one major question remains unanswered: why did the neo-Y chromosome fusion occur and became fixed in the population? We have four main hypotheses as to why the neo-Y may have become fixed in the Japan Sea population: 1) centromere hitchhiking, 2) meiotic drive, 3) sexually antagonistic selection, and 4) genetic drift.

Centromere hitchhiking

The first hypothesis is "centromere hitchhiking." In this model, we hypothesized that if the ancestral Y chromosome had a weak centromere that did not function optimally, the Y chromosome would not always properly segregate leading to aneuploidy. To alleviate having a suboptimal centromere, the Y chromosome could "hitch" on to another chromosome in order to maintain proper chromosome segregation. Currently, the measure of centromere strength is determined by CENP-A staining intensity (Chmátal et

al. 2014). It has been shown on both mouse and human Y chromosome centromeres that reduced CENP-A staining may be indicative of a weakened centromere (Nath et al. 1995; Irvine et al. 2004; Graham and Kalitsis 2014). Further supporting our hypothesis, in a mouse cell line with a partial Y centromere deletion, there are reduced CENPA levels and increased Y chromosome loss (Graham and Kalitsis 2014). Nath et al. 1995 similarly shows in samples of human male cells ranging from newborn to 79 years old, there was a significant increase in Y chromosome aneuploidy in older men, likely due to weakened kinetochore on aging Y chromosome centromeres. To test if the cause of the neo-Y fusion was a weak centromere, we examined levels of CENP-A staining on the ancestral Pacific Ocean Y chromosome. We also tested for evidence of aneuploidy, or loss of the Y chromosome, by examining sex ratios of crosses between Pacific Ocean females and males.

Meiotic drive

The second hypothesis is "meiotic drive." Meiotic drive (Unckless et al. 2015; Lindholm et al. 2016) is a phenomenon where chromosomes defy the Mendelian rules of segregation and are not transmitted in equal ratios to progeny. Many examples of meiotic drive are found in female meiosis because only one of the four potential chromosomes becomes an egg, while the other three become polar bodies (Pardo-Manuel de Villena and Sapienza 2001; Ubeda et al. 2014). Female meiotic drive predicts that one of a pair of homologous chromosomes or chromatids may preferentially segregate to the cell destined to become the egg, while the other of the pair becomes an evolutionary dead end in the polar bodies. However, the Japan Sea fusion involves the Y chromosome, which is never

subjected to female meiotic drive. Male meiotic drive also occurs, and is most easily recognized as sex ratio distortion when it involves the sex chromosomes (Lindholm et al. 2016). Because all four products of male meiosis become active sperm, male meiotic drive results from a difference in the ability of the sperm containing different chromosomes to fertilize eggs. If male meiotic drive was driving fixation of the Japan Sea neo-Y fusion, we hypothesized that in the ancestral Pacific Ocean population, unfused X and unfused Y containing sperm are equally able to fertilize eggs, resulting in an optimal sex ratio of 50:50 male to female offspring. In the Japan Sea population, we hypothesized that sperm with the neo-Y outcompeted sperm with the X to fertilize eggs, leading to a skewed sex ratio with more than 50% neo-Y males in the initial population. This skewed sex ratio may have further led to the evolution of a suppressor of male meiotic drive on the Japan Sea X chromosomes. Such a suppressor should not be present on the Pacific Ocean X chromosome. Thus, to test if there is any evidence for meiotic drive on the neo-Y fusion, we designed an experiment to test the ability of neo-Y containing sperm vs ancestral X containing sperm to fertilize eggs (recapitulating the fertilization ability of sperm at the time of the fusion).

Sexually antagonistic selection

The third hypothesis is "sexually antagonistic selection." This theory has been proposed as a mechanism to drive sex chromosome-autosome fusions (Charlesworth and Charlesworth 1980). In a system with already established heterogametic sexes, an autosome – Y chromosome fusion might evolve if the autosome contained a gene that has alleles with differential fitness effects in the two sexes. More specifically, if an autosome

contained a gene in which an allele is male beneficial but female detrimental, selection would favor the fusion of the autosome containing the male beneficial allele to the Y chromosome, where reduced recombination would lead to linkage with the sex determination gene. In this way, only males would carry the allele that is beneficial and females would not carry the allele that is detrimental. It is possible that sexually antagonistic selection was a driving force in the evolution of the Japan Sea neo-Y fusion. Interestingly, two male mating traits (dorsal spine length and dorsal pricking behavior) that are important to reproductive isolation between the Pacific Ocean females and Japan Sea males map to the Japan Sea neo-X chromosome (Kitano et al. 2009). However, this study was unable to assess whether these traits map to the neo-Y chromosome, as would be predicted by evolutionary theory (Charlesworth and Charlesworth 1980). Current work in the Peichel Lab is aimed at testing this hypothesis by determining whether genes on the Japan Sea neo-Y show genomic signatures of sexually antagonistic selection (Kirkpatrick and Guerrero 2014).

Genetic drift

The fourth hypothesis is that genetic drift due to a bottleneck allowed for the fixation of the fusion. The Japan Sea and Pacific Ocean species are thought to have diverged two million years ago when the Japan Sea and the Pacific Ocean were geographically isolated (Kitano and Mori 2007). After the last glacial period 10-70,000 years ago, coastal lines changed such that these two species do currently co-occur in coastal waters around Japan (Higuchi et al. 2014). However, this period of geographical isolation may have allowed for genetic and chromosomal changes to occur leading to their current reproductive

isolation when they co-occur (Kitano and Mori 2007; Kitano et al. 2009). Although not tested here, it is possible that there was not direct selection for the fixation of the Japan Sea neo-Y fusion and that it fixed via genetic drift because the Japan Sea went through a bottleneck when it was isolated from the Pacific Ocean. Interestingly, a recent study demonstrated that Y-autosome fusions are more common that X-autosome fusions, and that XY systems (male heterogametic) are more likely to have sex chromosome-autosome fusions than ZW systems (female heterogametic) in both fish and squamate reptiles (Pennell et al. 2015). Population genetic modeling studies suggest that this pattern is not driven by meiotic drive or sexually antagonistic selection; rather the prevalence of Y-autosome fusions is best explained if the fusions occur more frequently in males, are slightly deleterious, and fix by genetic drift (Pennell et al. 2015). Thus, fixation by genetic drift remains a plausible hypothesis for the Japan Sea neo-Y chromosome.

Materials and Methods

Crosses

To test for deviations in sex ratios, three separate crosses were made. We intercrossed Pacific Ocean males and females as well as Japan Sea males and females to test for the baseline sex ratios in the Pacific Ocean and Japan Sea species. The Pacific Ocean crosses also allowed us to test whether there is evidence for loss of the Y chromosome during meiosis due to a weak centromere. To test for whether meiotic drive of the initial neo-Y fusion might have occurred, we first crossed a Pacific Ocean female to a Japan Sea male to create F1 hybrid males with both the ancestral X and the neo-Y. By crossing these F1

males back to a Pacific Ocean female, we tested the ability of the F1 male sperm containing either the ancestral X or the neo-Y to fertilize eggs containing the ancestral X (Figure 4-1).

For each cross, male testes were dissected and stored in 1 ml of Ginzberg solution (Hart and Messina 1972). Eggs from a single female were then fertilized with around 100 μl of sperm from a male of the appropriate population. Eggs were allowed to fertilize for 10 min, then put in fish water in a petri dish. At 24 hours, any non-fertilized eggs were removed. At 48 hours, any embryos that had died were also removed, and the remaining embryos were fixed in 100% ethanol until genotyping.

DNA extraction from embryos

A single embryo from each cross was individually placed in a PCR plate well containing 60 μl of 50mM sodium hydroxide (NaOH). A plate seal was placed on the PCR plate and the embryos were digested at 95 ^{o}C for 30 minutes. The embryos were dissociated by manual pipetting with individual, sterile, single use pipette tips. A new plate seal was added to the plate, and the embryos were digested again at 95 ^{o}C for an additional 10 min. The plate seal was removed and 8 μl 1M Tris-HCl pH 7.4 was added to each well to neutralize the solution. 1 μl of this DNA extraction was used for sex genotyping by PCR.

Sex genotyping

Sex genotyping was performed using two sets of PCR primers. For the Pacific Ocean crosses, primers flanking the sex-linked IDH gene were used (Peichel et al. 2004). These primers amplify a single product from the X chromosome of around 300 bp in females,

but two products of 300 bp (X chromosome) and 270 bp (Y chromosome) in males. The sequence of the IDH primers are: F3 5'- GGGACGAGCAAGATTTATT and R4 5'- TTATCGTTAGCCAGGAGATGG. PCR parameters are described below. PCR products were run on a 2% TBE gel and then imaged. For both the backcross and Japan Sea crosses, we used PCR primers in LRRC61, a gene next to IDH. These primers amplify a single product of ~500 bp in females, and two products (~500 bp and ~180 bp) in males. The sequence of the LRRC61 primers are: F1 5'- TGGGGAGTGTATGAGTTTAGAGAA and R1 5'- CTCAAAGGCTGAAGGCAGTG. PCR products were run on a 1.5% TBE gel and then imaged. PCR parameters were as follows for the different primers:

IDH PCR:

(95 ^{o}C 2:00, 56 ^{o}C 1:00, 72 ^{o}C 2:00), (5X: {94 ^{o}C 1:00, 56 ^{o}C 1:00, 72 ^{o}C 2:00}), (30X: {90 ^{o}C 1:00, 56 ^{o}C 1:00, 72 ^{o}C 2:00}), (72 ^{o}C 5:00)

LRRC61 PCR:

(94 ^{o}C 1:00), (5X: {94 ^{o}C 0:30s, 65 ^{o}C 0:30s, 72 ^{o}C 3:00}), (94 ^{o}C 0:30s, 64 ^{o}C 0:30s, 72 ^{o}C 3:00), (94 ^{o}C 0:30s, 63 ^{o}C 0:30s, 72 ^{o}C 3:00), (94 ^{o}C 0:30s, 62 ^{o}C 0:30s, 72 ^{o}C 3:00), (94 ^{o}C 0:30s, 61 ^{o}C 0:30s, 72 ^{o}C 3:00), (25x: {94 ^{o}C 0:30s, 60 ^{o}C 0:30s, 72 ^{o}C 3:00}), (72 ^{o}C 10:00)

Results

Centromere hitchhiking

If the ancestral Y centromere has a weakened centromere function, we would expect CENP-A staining to be decreased, or lacking on the Y chromosome. We tested this

hypothesis by performing CENP-A staining in combination with BAC probes to distinguish the Y chromosome in Pacific Ocean males (see Chapter 3 for Methods). Immunofluorescence CENP-A staining on Pacific Ocean male metaphase spreads shows no evidence for decreased staining on the Y chromosome compared to the other chromosomes on the metaphase spread (Figure 3-7). While we did not directly quantify the levels of staining, these data suggest that the ancestral Y chromosome does not likely have a weak or non-functioning centromere. If the Y centromere was weak or suboptimal, we might expect to see mis-segregation or aneuploidy of the Y chromosome, yet we see equal sex ratios in 2517 offspring of Pacific Ocean crosses (Table 4-1). Thus, there is no evidence of aneuploidy associated with a suboptimal centromere on the ancestral Y chromosome, further supporting the conclusion that the ancestral Y chromosome centromere is active.

Meiotic drive

Because having equal sex ratios is so important, there is the possibility that the current Japan Sea X chromosome has evolved a suppressor allele to combat meiotic drive on the neo-Y, re-establishing a 50:50 sex ratio in the current Japan Sea population (Ubeda et al. 2014; Unckless et al. 2015). Therefore, we backcrossed a Pacific Ocean x Japan Sea F1 male to a Pacific Ocean female (Figure 4-1) and genotyped the sex of 2978 offspring. This allowed us to test the ability of the F1 sperm containing either the Japan Sea neo-Y or the Pacific Ocean ancestral X (with no potential suppressor), to fertilize Pacific Ocean ancestral X containing eggs. If the neo-Y shows drive against the ancestral X, the neo-Y containing sperm will out compete the ancestral X containing sperm and there will be a

greater proportion of male offspring. As a control, we also tested the sex ratio of 2517 Pacific Ocean embryos and 1038 Japan Sea embryos. Our results show that there are no deviations from equal sex ratios in any of the crosses (Table 4-1). These results suggest that meiotic drive was not likely the reason for fixation of the Japan Sea neo-Y fusion.

Discussion

While we have not yet tested all of the hypotheses, we predict that either sexually antagonistic selection or genetic drift led fixation of the neo-Y. Because we know dicentric chromosomes are inherently unstable, it is likely that some driving force (like sexually antagonistic selection) was strong enough to overcome the negative effects of the dicentric fusion, allowing the neo-Y to persist until centromere inactivation took place on the neo-Y, ultimately allowing for its fixation. However, it is also possible that this neo-Y was fixed in the Japan Sea population due to a bottleneck before the last glacial period, consistent with recent theory on the evolution of sex chromosome fusions (Pennell et al. 2015). Current work on the Peichel Lab is focused on testing whether there are genes with sexually antagonistic effects on the neo-Y chromosome, which will shed further light on the evolution of the Japan Sea neo-Y chromosome fusion.

Table 4-1. Sex genotyping to test for meiotic drive

Cross	Total embryos genotyped	# male embryos (percent of total)	# female embryos (percent of total)	Chi-sq P -value
Backcross	2878	1416 (49.2 %)	1462 (50.80%)	0.39
Pacific Ocean	2517	1261 (50.10%)	1256 (49.90%)	0.92
Japan Sea	1038	521 (50.19%)	517 (49.81%)	0.90

Chapter 4 Figure:

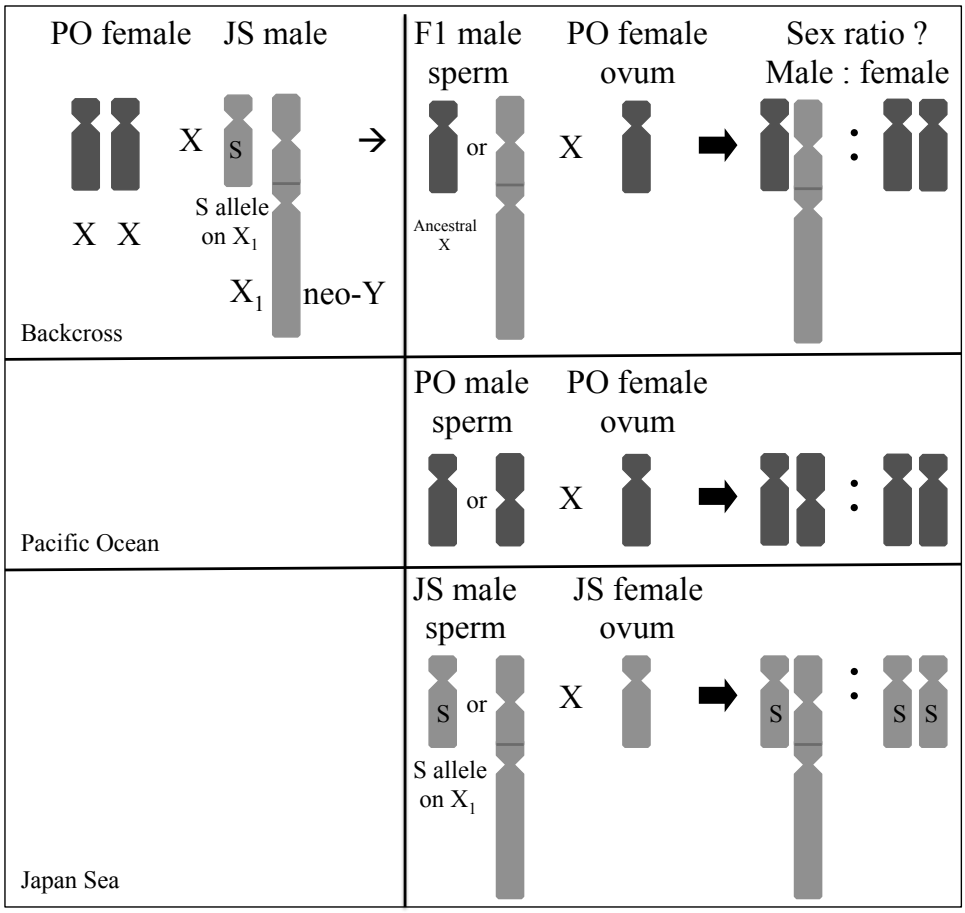


Figure 4-1. Crosses to test for meiotic drive and aneuploidy. For the backcross, a Pacific Ocean female was crossed to a Japan Sea male creating F1 hybrids. A male F1 hybrid would have sperm containing either the ancestral X or the neo-Y. F1 males were backcrossed to Pacific Ocean females to test if the neo-Y containing sperm fertilized more eggs than the ancestral X containing sperm. We also crossed Pacific Ocean males to Pacific Ocean females, as well as Japan Sea males to Japan Sea females to assess the current sex ratios for these species. The "S" on the Japan Sea X_1 chromosome designates the presence of a possible suppressor of meiotic drive.

Chapter 4 References:

Charlesworth D, Charlesworth B (1980) Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genet Res (Camb) 35:205–214

Chmátal L, Gabriel SI, Mitsainas GP, et al (2014) Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. Current Biology 1–6

Graham AN, Kalitsis P (2014) Chromosome Y centromere array deletion leads to impaired centromere function. PLoS ONE 9:e86875–e86875

Hart NH, Messina M (1972) Artificial insemination of ripe eggs in the zebra fish, Brachydanio rerio. Copeia 1972:302-305

Higuchi M, Sakai H, Goto A (2014) A new threespine stickleback, Gasterosteus nipponicus sp. nov. (Teleostei: Gasterosteidae), from the Japan Sea region. Ichthyol Res 61:341–351

Irvine DV, Amor DJ, Perry J, et al (2004) Chromosome size and origin as determinants of the level of CENP-A incorporation into human centromeres. Chromosome Res 12:805–815

Kirkpatrick M, Guerrero RF (2014) Signatures of sex-antagonistic selection on recombining sex chromosomes. Genetics 197:531–541

Kitano J, Mori S (2007) Phenotypic divergence and reproductive isolation between sympatric forms of Japanese threespine sticklebacks. Biol. J. Linn. Soc 91:671-685

Kitano J, Ross JA, Mori S, et al (2009) A role for a neo-sex chromosome in stickleback speciation. Nature 461:1079–1083

Lindholm AK, Dyer KA, Firman RC et al 2016. The ecology and evolutionary dynamics of meiotic drive. Trends in Ecology and Evolution 31:315-326

Nath J, Tucker JD, Hando JC (1995) Y chromosome aneuploidy, micronuclei, kinetochores and aging in men. Chromosoma 103:725–731

Pardo-Manuel de Villena F, Sapienza C (2001) Female meiosis drives karyotypic evolution in mammals. Genetics 159:1179–1189

Peichel CL, Ross JA, Matson CK, et al (2004) The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. Current Biology 14:1416–1424

Pennell MW, Kirkpatrick M, Otto SP, et al (2015) Y fuse? sex chromosome fusions in fishes and reptiles. PLoS Genet 11:e1005237

Sato H, Masuda F, Takayama Y, et al (2012) Epigenetic inactivation and subsequent heterochromatinization of a centromere stabilize dicentric chromosomes. Current Biology 22:658-667

Ubeda F, Patten MM, Wild G (2014) On the origin of sex chromosomes from meiotic drive. Proceedings of the Royal Society B: Biological Sciences 282:20141932–20141932

Unckless RL, Larracuente AM, Clark AG (2015) Sex-ratio meiotic drive and Y-linked resistance in drosophila affinis. Genetics 199:831-840

Chapter 5: Perspectives and future work

Further characterization of centromere inactivation on the neo-Y

We have shown that the Japan Sea neo-Y was a fusion between two complete chromosomes leading to an initially dicentric chromosome. We have further shown that there is only one active centromere on the neo-Y, and that the mechanism of dicentric stabilization involved centromere inactivation. Using a combination of CENP-A Immunoflourescence, and GacCEN and BAC probe FISH staining, we show the ancestral Y centromere on the neo-Y is no longer active, as it lacks CENP-A staining. Furthermore, our data suggest that there have been additional genetic changes to the inactivated centromere on the neo-Y. While lack of CENP-A is characteristic of inactive centromeres, future work should further characterize this inactivation by looking for additional histone and DNA methylation changes on the inactivated neo-Y centromere. Centromere inactivation in maize involves the replacement of H3K4me2 with H3K27me2/3 (Zhang et al. 2010), as well as hypermethylation of DNA at inactivated centromeres (Koo et al. 2011). These changes alter the centromeric chromatin to a more repressive heterochromatic state. It would be interesting to characterize the inactive and active states of stickleback centromeres at the histone and DNA methylation level.

Centromere inactivation on the blackspotted Stickleback neo-Y

Interestingly, the ancestral Y chromosome has independently fused to the acrocentric chromosome 12 in males of the blackspotted stickleback (*Gasterosteus wheatlandi*) species within the past 15 million years, creating an independent neo-Y chromosome fusion (Ross et al. 2009). From an evolutionary perspective, it is intriguing that two of

the seven known species of stickleback have undergone fixation of a Y-autosome fusion. Future work should test to see if the Y chromosome centromere is also inactivated on the blackspotted neo-Y, leaving the ancestral chromosome 12 centromere functional. Initial hybridization of the GacCEN sequence on metaphase spreads from blackspotted sticklebacks show little signal, suggesting that the blackspotted stickleback may have a different centromere sequence than the threespine stickleback. If the CENP-A protein sequence is the same in the blackspotted and threespine stickleback, we can use our CENP-A antibody in a ChiP-seq experiment to identify the blackspotted stickleback centromere sequence. If the CENP-A protein is different in the two species, a new antibody would need to be developed first. It would also be interesting to assess the reason for the fusion in this species.

Y chromosome centromeres

Our current evidence supports some level of genetic changes occurring on the Y centromere in both the ancestral Pacific Ocean Y chromosome and the Japan Sea neo-Y chromosomes. However, in order to determine if the ancestral Y centromere was truly deleted on the neo-Y or has just diverged, further experiments will be needed to sequence the ancestral Y centromere in the Pacific Ocean population. Sequencing Y centromeres has revealed that both the mouse and the human Y chromosome centromere sequences have diverged relative to centromeres from the autosomes and the X chromosomes (Wolfe et al. 1985; Pertile et al. 2009; Miga et al. 2014). The mouse Y min satellite is ~77% similar to the centromere sequence found on other chromosomes, and the human Y chromosome centromere sequence is about 70% similar to the X chromosome centromere

sequence. Higher mutation rates in males and lack of recombination between X and Y chromosomes (Pennell et al. 2015) might cause the increased divergence of the Y chromosome centromeres. Thus, it is possible that the stickleback Y centromere sequence is simply a divergent form of the GacCEN, or contains fewer repeats, leading to weak and variable GacCEN hybridization using our current FISH methods. While our preliminary ChiP-seq data supports the hypothesis that the ancestral Y is likely a divergent repeat, and not a unique sequence, we can only verify this by sequencing the Y chromosome centromere.

Y chromosome fusions

Another evolutionary question that remains is why we often see certain chromosomes (like the Y) involved in fusions and why certain centromeres are preferentially inactivated. Is there something about particular centromeres that make them more likely to inactivate upon fusion? A study of centromere inactivation among Robertsonian fusions found a hierarchy of which centromere was inactivated (Sullivan et al. 1994). In 83% (29/35) of Robertonsian translocations involving chromosome 14, the chromosome 14 centromere remained active. Similarly, in all of the Robertsonian translocations involving chromosome 15, the chromosome 15 centromere was always inactivated. One hypothesis is that the centromeres differ in strength and size and that the stronger centromere remains active over the weaker centromere upon fusion. Recent work in plants shows that the smaller of the two centromeres is always inactivated on dicentric chromosomes (Liu et al. 2015).

These data suggest that the Y chromosome centromere in sticklebacks in smaller or weaker. In humans, the Y chromosome centromere was on average the smallest (680kb), while the chromosome 11 centromere was on average the largest (3000kb) centromere array (Wevrick and Willard 1989). More recent work compared centromere array size between human X and Y chromosomes and also found the Y chromosome centromeres are on average smaller than the X chromosome centromeres (~1Mb vs ~3.5Mb respectively) (Miga et al. 2014). CENP-A domain size is proportional to centromere array size with Y centromeres having less CENP-A than X chromosomes (Sullivan et al. 2011). Consistent with this, in human cells lines, the Y chromosome centromere has the least amount of CENPA (Irvine et al. 2004). If having more CENP-A imparts a "stronger" centromere, and a larger centromere array has more CENP-A, are Y centromeres inherently weaker when compared to other centromeres because they are smaller? Consistent with this idea, a mouse line with high rates of Y chromosome instability was due to Y centromere deletion and subsequent reduction in CENP-A levels. (Graham and Kalitsis 2014). Using this rationale, I hypothesize that while a smaller centromere on a Y chromosome can function independently and contain an active centromere, when a Y chromosome fuses to a chromosome with a stronger centromere, the Y chromosome centromere becomes "weaker" in comparison and may be more easily inactivated. Interestingly, there are significantly more Y-autosome fusions in fish than X-autosome fusions in fish (Pennell et al. 2015). If the Y centromeres are smaller than the X centromeres in fish, a Y-autosome fusion might be less detrimental, as the Y centromeres might more easily be inactivated. In this model, X-autosome fusions might result more often in the BFB cycle than Y-autosome fusions in fish species. Future work

should compare the size of the ancestral Y centromere to the ancestral chromosome 9 centromere to see if this prediction holds true in the neo-Y fusion as well. If the ancestral Y centromere is smaller than the chromosome 12 centromere in the blackspotted stickleback, we would then hypothesize that the ancestral Y centromere would be inactivated on the blackspotted neo-Y as well. It would also be interesting to compare the centromeres of other fish species with Y-autosome fusions to see if the Y centromere is always inactivated.

The Japan Sea stickleback neo-Y is one of the few examples of a natural, and stable, dicentric chromosome involving two functionally important chromosomes that shows evidence for centromere inactivation. It is also one of the first examples showing centromere inactivation as a potential mechanism used to maintain a chromosome fusion that may play a role in the formation of new species. My thesis work paves way for future work that will aim to further characterize the Y chromosome centromere, and better understand its role in stickleback chromosome fusions and speciation.

Chapter 5 References:

Graham AN, Kalitsis P (2014) Chromosome Y centromere array deletion leads to impaired centromere function. PLoS ONE 9:e86875–e86875

Irvine DV, Amor DJ, Perry J, et al (2004) Chromosome size and origin as determinants of the level of CENP-A incorporation into human centromeres. Chromosome Res 12:805–815

Koo DH, Han F, Birchler JA, Jiang J (2011) Distinct DNA methylation patterns associated with active and inactive centromeres of the maize B chromosome. Genome Research 21:908–914

Liu Y, Su H, Pang J, et al (2015) Sequential de novo centromere formation and inactivation on a chromosomal fragment in maize. Proc Natl Acad Sci USA 201418248

Miga KH, Newton Y, Jain M, et al (2014) Centromere reference models for human chromosomes X and Y satellite arrays. Genome Research 24:697–707

Pennell MW, Kirkpatrick M, Otto SP, et al (2015) Y fuse? sex chromosome fusions in fishes and reptiles. PLoS Genet 11:e1005237

Pertile MD, Graham AN, Choo KHA, Kalitsis P (2009) Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability. Genome Res 19:2202–2213

Ross JA, Urton JR, Boland J, et al (2009) Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). PLoS Genet 5:e1000391

Sullivan BA, Wolff DJ, Schwartz S (1994) Analysis of centromeric activity in Robertsonian translocations: implications for a functional acrocentric hierarchy. Chromosoma 103:459–467

Sullivan LL, Boivin CD, Mravinac B, et al (2011) Genomic size of CENP-A domain is proportional to total alpha satellite array size at human centromeres and expands in cancer cells. Chromosome Res 19:457–470

Wevrick R, Willard HF (1989) Long-range organization of tandem arrays of alpha satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. Proc Natl Acad Sci USA 86:9394–9398

Wolfe J, Darling SM, Erickson RP, et al (1985) Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. J Mol Biol 182:477–485

Zhang W, Friebe B, Gill BS, Jiang J (2010) Centromere inactivation and epigenetic modifications of a plant chromosome with three functional centromeres. Chromosoma 119:553–563