EXPLORING THE DYNAMICS OF PAIRING AND DOUBLE STRAND BREAK REPAIR ON THE SEX CHROMOSOMES OF THREESPINE STICKLEBACKS

by

SHIVANGI NATH

(Under the Direction of Michael A. White)

ABSTRACT

Sex chromosomes degenerate and evolve independently in the absence of recombination. Homologous chromosomes pair during meiosis and exchange genetic information in the form of crossovers and non-crossovers. However, this poses a challenge to sex chromosome in terms of pairing as well as double stranded break repair as they are typically only homologous at the small, shared region called pseudoautosomal region (PAR). Threespine sticklebacks (*Gasterosteus aculeatus)* have a young sex chromosomes system which is approximately 14 million years old and retains much of homology between sex chromosomes. While ancient sex chromosomes only pair at PAR and rarely form breaks outside PAR, we sought to address how young sex chromosomes navigate meiosis. In this dissertation the patterns of pairing, double stranded break formation on sex chromosomes and the mechanisms of repair of such breaks is explored. The sex chromosomes despite inversions and divergence were shown to pair fully and undergo synaptic adjustment in late prophase of meiosis I, revealing that the degeneration and divergence of Y chromosome is not sufficient to abrogate pairing in this system. The sex chromosomes also were shown to form double stranded breaks at the same rate and

spatial pattern as autosomes. This indicates that DSBs are not suppressed in our system which is in contrast with ancient sex chromosomes on mice where DSB suppression on sex chromosomes is well documented. DSBs on sex chromosome systems were often found to be repaired through gene conversion between sex chromosomes in more recently formed strata or the regions where recombination was halted. Together these results show that meiotic nature of young sex chromosomes is similar to autosomes and ongoing genetic exchange between the sex chromosomes.

INDEX WORDS: Sex Chromosomes; Meiosis; Gene conversion; Synapsis; Genome Assembly; Threespine stickleback fish

EXPLORING THE DYNAMICS OF PAIRING AND DOUBLE STRAND BREAK REPAIR ON THE SEX CHROMOSOMES OF THREESPINE STICKLEBACKS

by

SHIVANGI NATH

B.E., Birla Institute of Technology and Sciences, India, 2015

M.S., Birla Institute of Technology and Sciences, India, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

EXPLORING THE DYNAMICS OF PAIRING AND DOUBLE STRAND BREAK REPAIR ON THE SEX CHROMOSOMES OF THREESPINE STICKLEBACKS

by

SHIVANGI NATH

Major Professor:	Michael White
Committee:	Mary Goll
	Kelly Dawe
	Kelly Dyer
	James Leebens-Mack

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2021

ACKNOWLEDGEMENTS

To my advisor Mike White for all his guidance, feedback and assurances when things got rough at times. I cannot thank him enough for always looking out for me and lending a keen ear to my concerns, academic or otherwise.

To my committee members Kelly Dawe, Mary Goll, Jim Leebans-Mack and Kelly Dyer for their expertise and guidance.

To Rich Meagher for teaching me microscopy and always taking time to talk to me about troubleshooting.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

DNA double strand break formation and chromosome synapsis during meiosis

Meiosis is a specialized form of cell division that results in the formation of haploid gametes. Meiosis can be divided broadly into two stages: Meiosis-I and Meiosis-II. During meiosis-I, the homologous chromosomes undergo recombination and segregate into two daughter cells, while during meiosis-II, the sister chromatids separate from each other into two additional daughter cells, forming a total of four haploid gametes. Meiosis-I and meiosis-II typically consists of four different phases: prophase, metaphase, anaphase and telophase. Prophase of meiosis-I can be further divided into leptotene, zygotene, pachytene and diplotene. Prophase is initiated when double stranded breaks (DSBs) form throughout the chromatids in leptotene. The single-stranded DNA at each break facilitates the homology search, ultimately leading to co-alignment of the homologous chromosomes. The co-aligned homologous chromosomes then form a synaptonemal complex between them in zygotene and the DSBs finish repairing through homologous-based repair, either as a cross over or non-crossover (i.e., gene conversion) event (ZICKLER and KLECKNER 2015; ZICKLER 2020).

The process of DSB formation and repair is a highly conserved process throughout the tree of life. The DSB formation in meiosis is catalyzed by a topoisomerase-like protein Spo11(KEENEY *et al.* 1997). The DNA around DSBs is processed by exonucleases Mre11 and/or Exo1 to expose the single stranded DNA

(ssDNA) (GARCIA *et al.* 2011). The ssDNA is then loaded with RPA, Rad51 and/or DMC1 (PLUG *et al.* 1997; MOENS *et al.* 2002; EDLINGER and SCHLÖGELHOFER 2011) which initiate strand invasion and homologous repair (HINCH *et al.* 2020) (Fig. 1.1). Concomitantly, chromosomes undergo chromatin condensation and formation of axial, lateral and central elements which help with homologous repair and proper segregation (BOLCUN-FILAS and HANDEL 2018). Axial (*SMC3, STAG3, Rec8*) and lateral elements (*SCP3, SCP2*) are deposited on all chromosomes and aid in chromosome condensation. Central elements hold the homologs undergoing synapsis. One notable central element is SCP1 that dimerizes between the synaptonemal complex of the homologous chromosomes and holds them together (MEUWISSEN *et al.* 1997; PAGE and HAWLEY 2004; DE VRIES *et al.* 2005) (Fig. 1.2).

Double stranded breaks that form during meiosis can be repaired through cross-over or non-crossover (gene conversion) mechanisms via multiple pathways. One or both re-sectioned DNA strands can invade the intact homologous chromosome and use the homologous template to synthesize new DNA strand. While single strand displacement and annealing (also called SDSA) results in gene conversion (non-crossover), the intermediates with second end capture can be resolved as crossover or non-crossover products (JASIN and ROTHSTEIN 2013; CECCALDI *et al.* 2016; SULLIVAN and BERNSTEIN 2018). Crossing over is important for proper segregation of homologs in some systems but typically only one DSB per chromosome is repaired as a CO and the vast majority of breaks are repaired as gene conversions (BAUDAT and DE MASSY 2007; MANCERA *et al.* 2008; CECCALDI *et al.* 2016). While crossing over involves exchange of long stretches of

chromatids, gene conversions result in non-reciprocal repair of the lesion using the other homolog (MANSAI *et al.* 2011; PADHUKASAHASRAM and RANNALA 2013).

Heteromorphic sex chromosomes often share limited sequence homology for synapsis and DSB repair

Sex chromosomes evolve from autosomes when one of the chromosomes acquire a sex determining gene and stop recombining. Sex chromosomes can vary widely in the extent of their heteromorphy. In the absence of recombination, the deleterious mutations and large scale indels are not effectively purged from the non-recombining sex chromosomes and accumulate irreversibly on Y/W (Bachtrog 2013; Rodrigues *et al.* 2017). Hence sex chromosomes evolve independently beyond PAR and are no longer homologous in the absence of recombination. Sex chromosome systems of mammals (~180 million years old) and birds (~140 million years old) are ancient and show extreme divergence from each other (CORTEZ *et al.* 2014). The human Y only contains around 3% of the genes that were present on ancestral autosomes, illustrating the extent of degeneration (BELLOTT *et al.* 2014).

Since pairing of chromosomes in meiosis is dependent on homology, sex chromosomes differ from autosomes in their meiotic behavior (CHANDLEY *et al.* 1984; KAUPPI *et al.* 2011). Sex chromosomes usually share a small region of homology called the pseudoautosomal region (PAR) where synapsis and crossing over must occur. In mammals, sex chromosomes only pair in the PAR and their pairing is delayed (RASMUSSEN and HOLM 1978; KAUPPI *et al.* 2011; DUMONT *et al.* 2018). However, pairing is not limited to the PAR on all sex chromosomes. The ancient heterogametic ZW sex chromosome system of birds exhibits pairing that extends beyond the PAR, resulting

in some non-homologous pairing (GUIOLI *et al.* 2012; PIGOZZI 2016). Furthermore, in marsupials sex chromosome pairing has been completely lost and instead protein structures like "dense plate" and "protein bridges" may aid in proper segregation (CARNERO *et al.* 1991; JIMÉNEZ *et al.* 1991; DE LA FUENTE *et al.* 2007). Voles have lost sex chromosomes pairing independently in many species, suggesting that while in some systems PAR pairing is obligate, others have evolved mechanisms to render PAR synapsis dispensable (BORODIN *et al.* 2012; DUMONT *et al.* 2018). While we have extensive knowledge of sex chromosome pairing from mammalian systems and a handful of non-mammalian systems, pairing has not been explored in evolutionary young sex chromosome systems till date, where sequence homology between the X and Y or Z and W is much higher. Studying pairing in young sex chromosome system will shed light on how sex chromosomes pair when much of the homology is still intact. If young sex chromosomes also pair only in the PAR, it will inform our view on how quickly an impediment to non-homologous pairing may evolve.

Double strand break formation and repair on heteromorphic sex chromosomes

Double stranded breaks forming on sex chromosomes also show atypical patterns in the PAR as well as in the non-PAR region. In house mice, the density of DSBs in the PAR is extremely high to ensure the obligate crossover event, whereas the DSB density on the non-PAR regions is extremely low (KAUPPI *et al.* 2011; LANGE *et al.* 2016). This suggests there can be mechanisms to enrich or suppress DSB's among regions of the sex chromosomes. One of factors determining the number of DSBs forming on chromosomes is the chromatin organization during meiosis. Chromatin is organized as loops around the axis formed by various cohesins. Depending on the overall length of the synaptonemal

axis, the DNA can be packaged in tighter or broader loops around the axis. The DSBs are formed on the loops, in proximity of the region of chromatin tethered to axial elements. Thus, smaller loop size will allow for more loops in a given chromosome segment and will provide more potential DSB sites which can then be tethered to axis and proceed repair. Similarly, a longer loop will suppress the number of potential DSBs (LAM and KEENEY 2014; ZICKLER and KLECKNER 2015; RUIZ-HERRERA *et al.* 2017). Interestingly, loop size has been implicated in regulating DSB rate on sex chromosomes in house mice. Within the house mouse PAR, the synaptonemal axis is lengthened in males, resulting in much smaller loops and a higher DSB rate relative to outside of the PAR. This configuration is thought to be one mechanism to ensure a crossover event. In the non-PAR region of Y chromosomes, the DNA is packaged in much longer loops, which may be why there are very low rates of DSB formation along the remainder of the Y chromosome (KAUPPI *et al.* 2011; LANGE *et al.* 2016).

Double strand breaks that do form on sex chromosomes can be repaired utilizing three separate templates. First, Y chromosomes often have palindromic repeats, which can function in intra-chromosomal repair. Second, sex chromosomes can be repaired using the sister chromatid (i.e., X-X repair or Y-Y repair). This kind of repair is proposed to be suppressed in early stages of prophase and is allowed in the later stages (Pachytene) (LANGE *et al.* 2016; TORAASON *et al.* 2020). Third, sex chromosomes can find tracts of local homology with the other gametolog and repair (i.e. X-Y repair) (TROMBETTA and CRUCIANI 2017) (Fig. 1.3).

Inter-gametolog gene conversion outside the PAR has been identified between the human X and Y chromosomes. Among these repair events, thirteen gene conversion

hotspots have been identified from population samples (ROSSER *et al.* 2009; TROMBETTA *et al.* 2010; TROMBETTA *et al.* 2014; TROMBETTA *et al.* 2016; TROMBETTA *et al.* 2017). Interestingly, most of the hotspots overlap with regions of known translocations involved in sex reversal phenotype. It is hypothesized that some of the repair events in those hotspots result in translocations and rest get resolved as GCs (CHANDLEY *et al.* 1984). While nine of them are still active in humans, the rest are ancient events (reviewed in (TROMBETTA *et al.* 2017). Some of these hotspots are also shared with chimpanzees which shows their remarkable longevity (TROMBETTA *et al.* 2014; TROMBETTA *et al.* 2016; TROMBETTA *et al.* 2017). One such hotspot is *HSA* which has been re-sequenced in multiple human populations to show gene conversions in both X->Y and Y->X directions (ROSSER *et al.* 2009). Inter-gametolog genetic conversion has also been implicated in the *VCY/VCX* gene pair (TROMBETTA *et al.* 2010). The *VCY/VCX* genes share 95% homology which could be due to strong purifying selection or gene conversion. Notably, the inter-chromosomal gene conversion rate on sex chromosomes is at least a magnitude higher than that of mutation rate, which indicates that gene conversion events have a major role in shaping (ROSSER *et al.* 2009; TROMBETTA *et al.* 2010; TROMBETTA *et al.* 2016; TROMBETTA and CRUCIANI 2017; TROMBETTA *et al.* 2017).

Intrachromosomal gene conversion has been identified within the human and house mouse Y chromosomes. Sex chromosomes can have multiple copies of same genes (*BPY2, CDY, DAZ, PRY, RBMY1A1, TSPY* on human Y and *SLY, SSTY1, SSTY2, SRSY* in mice) and duplicated segments (P1-P8 in humans) that contain multiple genes. The DSBs can be repaired using such templates on the same chromosome and homogenize the genes resulting in duplicated genes with extremely low sequence

divergence (SKALETSKY *et al.* 2003; BOSCH *et al.* 2004; SOH *et al.* 2014; TROMBETTA *et al.* 2016; TROMBETTA and CRUCIANI 2017; LUCOTTE *et al.* 2018).

Although gene conversion on the sex chromosomes has been widely studied in humans and mice, there is a growing body of evidence that this process occurs in other species as well. Recently, it has been shown in young sex chromosomes of *Poecilia reticulata and Poecilia wingei* how recombination is suppressed over almost the entire sex chromosomes beyond PAR, but the suppression is not complete. Instead of clustering by gametologs, some genes cluster by species showing that gene conversion or occasional recombination maybe homogenizing some genes (DAROLTI *et al.* 2020). On the other hand, some genes in old and degenerated sex chromosomes in *A. platyrhynchos* (mallards) have also been shown to harbor inter-gametolog gene conversions. This evidence bolsters the idea that recombination suppression is not often complete, and the leaky recombination can play a major role in shaping the dynamics of Y chromosome degradation (WRIGHT *et al.* 2014). Inter-chromosomal and intra-chromosomal gene conversions can play a major role in increasing the longevity of single copy genes on sex chromosomes or maintain multiple functional copies of dosage sensitive genes (MARAIS *et al.* 2010; BELLOTT *et al.* 2014). Additional detailed characterization of DSB initiation and repair using young sex chromosomes is necessary to understand whether ongoing gene conversion between the X and Y or within the Y can shape early sequence evolution.

Using threespine stickleback fish as a model species to understand double strand break initiation and repair on young sex chromosomes

Threespine stickleback fish (*Gasterosteus aculeatus*) are an evolutionary and ecologically important model organism (BELL and FOSTER 1994; BELL 1995; SCHLUTER *et al.* 2010; JONES *et al.* 2012). Multiple populations of threespine stickleback fish exist in freshwater as well as marine ecosystems throughout the world, making it a suitable system to study speciation and, parallel and convergent evolution (CRESKO *et al.* 2004; HART *et al.* 2018; FANG *et al.* 2020). Threespine sticklebacks have a XY sex chromosome system which emerged almost the same time as its split from ninespine sticklebacks around 15 million years ago (BELL *et al.* 2009; ROSS *et al.* 2009; VARADHARAJAN *et al.* 2019; SARDELL *et al.* 2020). This XY system is relatively young and Y chromosome shows limited degeneration compared to the old sex chromosome systems like birds and mammals (~120 and 180 million years old, respectively) (BELLOTT *et al.* 2014; CORTEZ *et al.* 2014; WHITE *et al.* 2015; PEICHEL *et al.* 2020). Sex chromosomes in threespine sticklebacks share a small pseudoautosomal region (2.75 Mb) on one end. The sex chromosomes harbor three major inversions corresponding to three strata where recombination was suppressed in stepwise fashion (ROSS and PEICHEL 2008; PEICHEL *et al.* 2020). Whole genome as well as sex chromosomes in this system have been assembled which provides a great opportunity to use genomic tools to study evolution and molecular biology of young sex chromosomes (PEICHEL *et al.* 2020). Threespine sticklebacks are also ideal model for studying meiosis as meiocyte collection is relatively simple and abundant. Males undergo seasonal meiosis which enables

collection of large number of meiocytes in desired stages of meiosis when they are sexually mature (CUÑADO *et al.* 2002; ISHIKAWA and KITANO 2020).

Research Aims

The broad goal of my dissertation is to explore how young sex chromosomes navigate meiosis. In this dissertation, I use the threespine stickleback to study pairing and patterns of double stranded break formation and repair on young sex chromosomes and also improve the existing assembly by filling in the gaps using PacBio Long reads. I investigate the broad hypothesis that young sex chromosomes have higher rates of DSB formation and repair via gene conversion between the X and Y than what is observed on more ancient sex chromosomes.

In chapter two of my dissertation, I use a molecular cytogenetic approach to explore the overall rate of DSB formation in threespine stickleback spermatocytes on the autosomes and sex chromosomes. I also explore the conformations of sex chromosome pairing during prophase of meiosis I. If the sex chromosomes are not pairing beyond PAR, we would expect the DSBs to be mainly repaired through sister chromatids and intra-chromosomal repeats.

In chapter three of my dissertation, I used long-read sequencing technologies to improve the threespine stickleback reference genome assembly. This highly refined assembly was important for generating the most accurate sequence alignments with the long reads used in the fourth chapter.

In chapter four of my dissertation, I explore how double stranded breaks were repaired on sex chromosomes by sequencing a pool of mature sperm. I used high-fidelity circular consensus PacBio sequencing of long haploid sperm DNA molecules in order to

localize individual crossover and gene conversion events throughout the autosomes and sex chromosomes.

In chapter five, I discuss the future studies that can be conducted to address the mechanisms to further study pairing of sex chromosomes at a finer scale and to study gene conversion on sex chromosomes that can be discovered using alternative sequencing approaches.

Figures

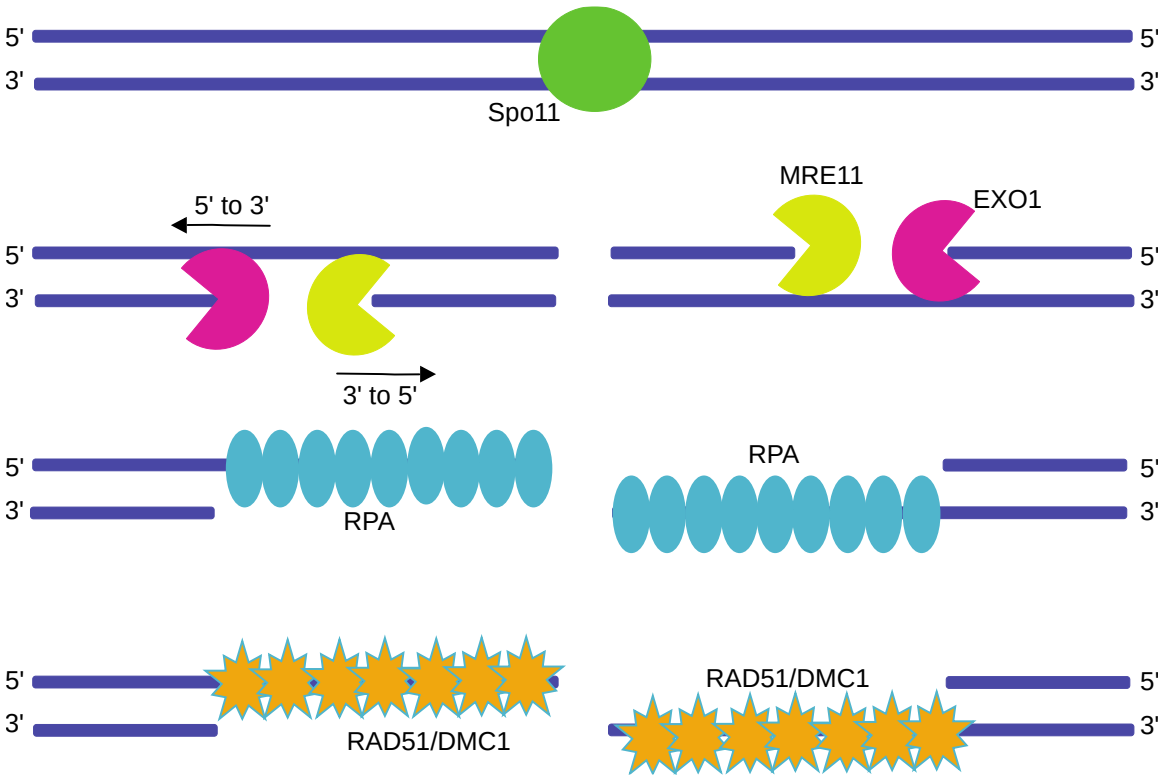


Figure1.1 Processing of double stranded breaks during meiosis. The Double stranded breaks are formed by *Spo11* which are then processes by exonucleases (*Exo1* and *Mre11*) to expose the single stranded DNA (ssDNA) at the lesion. *RPA* stabilizes the ssDNA which is then replaced by Rad51 and DMC1 that initiate strand invasion and repair.

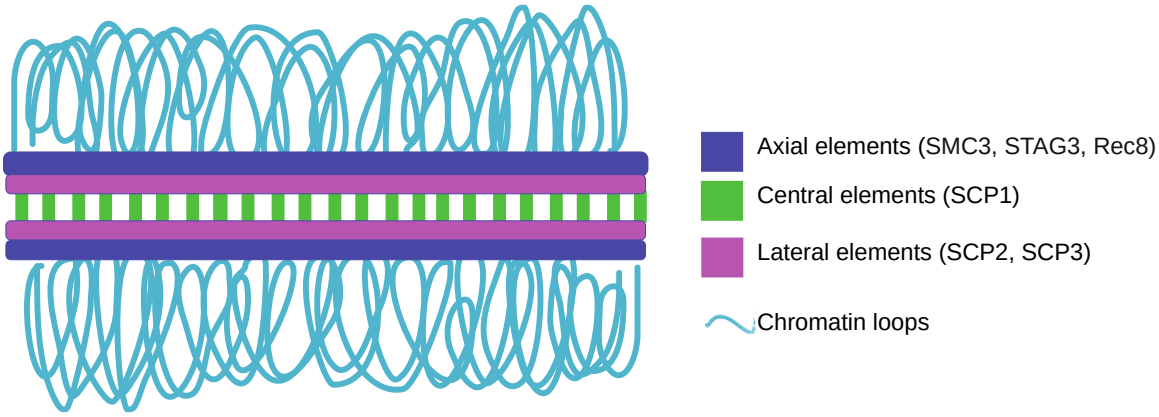


Figure1.2 The arrangement of chromatin loops around synaptonemal complex during meiosis. The axial elements form the axis between sister chromatids, lateral elements deposit on all homologous chromosomes and finally central elements dimerize between the homologous chromosomes to complete synapsis.

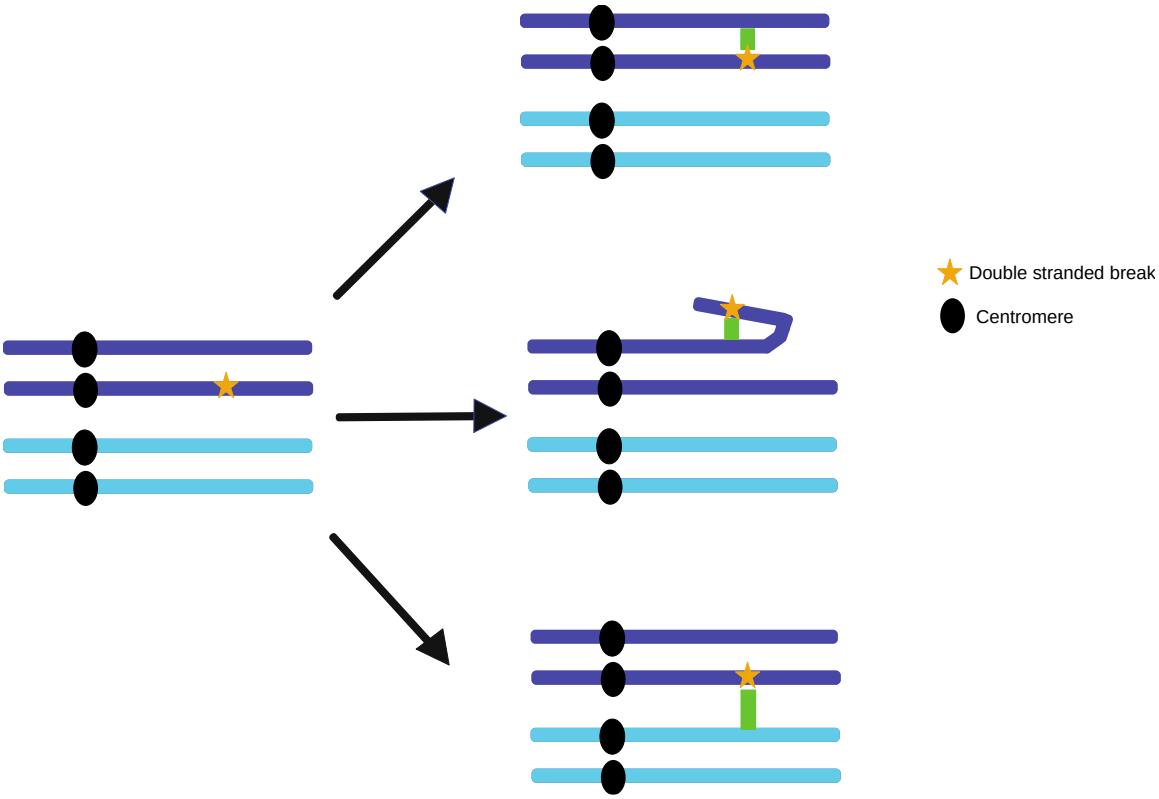


Figure1.3 The three possible fates of Double stranded breaks on sex chromosomes. The DSBs on sex chromosomes can be either repaired by using sister chromatid as a template (top), or the duplicated sequences within the chromosome (middle) and, finally by gene conversion or non-reciprocal exchange with the other gametolog in regions of high sequence identity (bottom).

References

Bachtrog, D., 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet 14**:** 113-124.

Baudat, F., and B. de Massy, 2007 Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Res 15**:** 565-577.

Bell, M., and S. A. Foster, 1994 The evolutionary biology of the threespine sticklebacks, pp. Oxford University Press.

Bell, M. A., 1995 Sticklebacks: a model for behavior evolution. Trends Ecol Evol 10**:** 101-103.

Bell, M. A., J. D. Stewart and P. J. Park, 2009 The World's Oldest Fossil Threespine Stickleback Fish. Copeia 2009**:** 256-265, 210.

Bellott, D. W., J. F. Hughes, H. Skaletsky, L. G. Brown, T. Pyntikova *et al.*, 2014 Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508**:** 494-499.

Bolcun-Filas, E., and M. A. Handel, 2018 Meiosis: the chromosomal foundation of reproduction. Biol Reprod 99**:** 112-126.

Borodin, P. M., E. A. Basheva, A. A. Torgasheva, O. A. Dashkevich, F. N. Golenishchev *et al.*, 2012 Multiple independent evolutionary losses of XY pairing at meiosis in the grey voles. Chromosome Res 20**:** 259-268.

Bosch, E., M. E. Hurles, A. Navarro and M. A. Jobling, 2004 Dynamics of a human interparalog gene conversion hotspot. Genome Res 14**:** 835-844.

Carnero, A., R. Jiménez, M. Burgos, A. Sánchez and R. Díaz de la Guardia, 1991 Achiasmatic sex chromosomes in Pitymys duodecimcostatus: mechanisms of association and segregation. Cytogenet Cell Genet 56**:** 78-81.

Ceccaldi, R., B. Rondinelli and A. D. D'Andrea, 2016 Repair Pathway Choices and Consequences at the Double-Strand Break. Trends Cell Biol 26**:** 52-64.

Chandley, A. C., P. Goetz, T. B. Hargreave, A. M. Joseph and R. M. Speed, 1984 On the nature and extent of XY pairing at meiotic prophase in man. Cytogenet Cell Genet 38**:** 241-247.

Cortez, D., R. Marin, D. Toledo-Flores, L. Froidevaux, A. Liechti *et al.*, 2014 Origins and functional evolution of Y chromosomes across mammals. Nature 508**:** 488-493.

Cresko, W. A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al.*, 2004 Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proc Natl Acad Sci U S A 101**:** 6050-6055.

Cuñado, N., J. Barrios, E. S. Miguel, R. Amaro, C. Fernández *et al.*, 2002 Synaptonemal complex analysis in oocytes and spermatocytes of threespine stickleback Gasterosteus aculeatus (Teleostei, Gasterosteidae). Genetica 114**:** 53-56.

Darolti, I., A. E. Wright and J. E. Mank, 2020 Guppy Y Chromosome Integrity Maintained by Incomplete Recombination Suppression. Genome Biol Evol 12**:** 965-977.

de la Fuente, R., M. T. Parra, A. Viera, A. Calvente, R. Gómez *et al.*, 2007 Meiotic pairing and segregation of achiasmate sex chromosomes in eutherian mammals: the role of SYCP3 protein. PLoS Genet 3**:** e198.

de Vries, F. A., E. de Boer, M. van den Bosch, W. M. Baarends, M. Ooms *et al.*, 2005 Mouse Sycp1 functions in synaptonemal complex assembly, meiotic recombination, and XY body formation. Genes Dev 19**:** 1376-1389.

Dumont, B. L., C. L. Williams, B. L. Ng, V. Horncastle, C. L. Chambers *et al.*, 2018 Relationship Between Sequence Homology, Genome Architecture, and Meiotic Behavior of the Sex Chromosomes in North American Voles. Genetics 210**:** 83-97.

Edlinger, B., and P. Schlögelhofer, 2011 Have a break: determinants of meiotic DNA double strand break (DSB) formation and processing in plants. J Exp Bot 62**:** 1545-1563.

Fang, B., P. Kemppainen, P. Momigliano, X. Feng and J. Merilä, 2020 On the causes of geographically heterogeneous parallel evolution in sticklebacks. Nat Ecol Evol 4**:** 1105-1115.

Garcia, V., S. E. Phelps, S. Gray and M. J. Neale, 2011 Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. Nature 479**:** 241-244.

Guioli, S., R. Lovell-Badge and J. M. Turner, 2012 Error-prone ZW pairing and no evidence for meiotic sex chromosome inactivation in the chicken germ line. PLoS Genet 8**:** e1002560.

Hart, J. C., N. A. Ellis, M. B. Eisen and C. T. Miller, 2018 Convergent evolution of gene expression in two high-toothed stickleback populations. PLoS Genet 14**:** e1007443.

Hinch, A. G., P. W. Becker, T. Li, D. Moralli, G. Zhang *et al.*, 2020 The Configuration of RPA, RAD51, and DMC1 Binding in Meiosis Reveals the Nature of Critical Recombination Intermediates. Mol Cell 79**:** 689-701.e610.

Ishikawa, A., and J. Kitano, 2020 Diversity in reproductive seasonality in the three-spined stickleback,. J Exp Biol 223.

Jasin, M., and R. Rothstein, 2013 Repair of strand breaks by homologous recombination. Cold Spring Harb Perspect Biol 5**:** a012740.

Jiménez, R., A. Carnero, M. Burgos, A. Sánchez and R. Díaz de la Guardia, 1991 Achiasmatic giant sex chromosomes in the vole Microtus cabrerae (Rodentia, Microtidae). Cytogenet Cell Genet 57**:** 56-58.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli *et al.*, 2012 The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484**:** 55-61.

Kauppi, L., M. Barchi, F. Baudat, P. J. Romanienko, S. Keeney *et al.*, 2011 Distinct properties of the XY pseudoautosomal region crucial for male meiosis. Science 331**:** 916-920.

Keeney, S., C. N. Giroux and N. Kleckner, 1997 Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell 88**:** 375-384.

Lam, I., and S. Keeney, 2014 Mechanism and regulation of meiotic recombination initiation. Cold Spring Harb Perspect Biol 7**:** a016634.

Lange, J., S. Yamada, S. E. Tischfield, J. Pan, S. Kim *et al.*, 2016 The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair. Cell 167**:** 695-708 e616.

Lucotte, E. A., L. Skov, J. M. Jensen, M. C. Macià, K. Munch *et al.*, 2018 Dynamic Copy Number Evolution of X- and Y-Linked Ampliconic Genes in Human Populations. Genetics 209**:** 907-920.

Mancera, E., R. Bourgon, A. Brozzi, W. Huber and L. M. Steinmetz, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature 454**:** 479-485.

Mansai, S. P., T. Kado and H. Innan, 2011 The Rate and Tract Length of Gene Conversion between Duplicated Genes. Genes (Basel) 2**:** 313-331.

Marais, G. A., P. R. Campos and I. Gordo, 2010 Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. Genome Biol Evol 2**:** 347-357.

Meuwissen, R. L., I. Meerts, J. M. Hoovers, N. J. Leschot and C. Heyting, 1997 Human synaptonemal complex protein 1 (SCP1): isolation and characterization of the cDNA and chromosomal localization of the gene. Genomics 39**:** 377-384.

Moens, P. B., N. K. Kolas, M. Tarsounas, E. Marcon, P. E. Cohen *et al.*, 2002 The time course and chromosomal localization of recombination-related proteins at meiosis in the mouse are compatible with models that can resolve the early DNA-DNA interactions without reciprocal recombination. J Cell Sci 115**:** 1611-1622.

Padhukasahasram, B., and B. Rannala, 2013 Meiotic gene-conversion rate and tract length variation in the human genome. Eur J Hum Genet.

Page, S. L., and R. S. Hawley, 2004 The genetics and molecular biology of the synaptonemal complex. Annu Rev Cell Dev Biol 20**:** 525-558.

Peichel, C. L., S. R. McCann, J. A. Ross, A. F. S. Naftaly, J. R. Urton *et al.*, 2020 Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol 21**:** 177.

Pigozzi, M. I., 2016 The Chromosomes of Birds during Meiosis. Cytogenet Genome Res 150**:** 128-138.

Plug, A. W., A. H. Peters, Y. Xu, K. S. Keegan, M. F. Hoekstra *et al.*, 1997 ATM and RPA in meiotic chromosome synapsis and recombination. Nat Genet 17**:** 457-461.

Rasmussen, S. W., and P. B. Holm, 1978 Human meiosis II. Chromosome pairing and recombination nodules in human spermatocytes. Carlsberg Research Communications 43**:** 275-327.

Rodrigues, N., T. Studer, C. Dufresnes and N. Perrin, 2017 Sex-Chromosome Recombination in Common Frogs Brings Water to the Fountain-of-Youth. Mol Biol Evol 35**:** 942-948.

Ross, J. A., and C. L. Peichel, 2008 Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. Genetics 179**:** 2173-2182.

Ross, J. A., J. R. Urton, J. Boland, M. D. Shapiro and C. L. Peichel, 2009 Turnover of sex chromosomes in the stickleback fishes (gasterosteidae). PLoS Genet 5**:** e1000391.

Rosser, Z. H., P. Balaresque and M. A. Jobling, 2009 Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. Am J Hum Genet 85**:** 130-134.

Ruiz-Herrera, A., M. Vozdova, J. Fernández, H. Sebestova, L. Capilla *et al.*, 2017 Recombination correlates with synaptonemal complex length and chromatin loop size in bovids-insights into mammalian meiotic chromosomal organization. Chromosoma 126**:** 615-631.

Schluter, D., K. B. Marchinko, R. D. Barrett and S. M. Rogers, 2010 Natural selection and the genetics of adaptation in threespine stickleback. Philos Trans R Soc Lond B Biol Sci 365**:** 2479-2486.

Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier *et al.*, 2003 The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423**:** 825-837.

Soh, Y. Q., J. Alfoldi, T. Pyntikova, L. G. Brown, T. Graves *et al.*, 2014 Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell 159**:** 800-813.

Sullivan, M. R., and K. A. Bernstein, 2018 RAD-ical New Insights into RAD51 Regulation. Genes (Basel) 9.

Toraason, E., C. Clark, A. Horacek, M. L. Glover, A. Salagean *et al.*, 2020 Sister chromatid repair maintains genomic integrity during meiosis in Caenorhabditis elegans. bioRxiv**:** 2020.2007.2022.216143.

Trombetta, B., and F. Cruciani, 2017 Y chromosome palindromes and gene conversion. Hum Genet 136**:** 605-619.

Trombetta, B., F. Cruciani, P. A. Underhill, D. Sellitto and R. Scozzari, 2010 Footprints of X-to-Y gene conversion in recent human evolution. Mol Biol Evol 27**:** 714-725.

Trombetta, B., E. D'Atanasio and F. Cruciani, 2017 Patterns of Inter-Chromosomal Gene Conversion on the Male-Specific Region of the Human Y Chromosome. Front Genet 8**:** 54.

Trombetta, B., G. Fantini, E. D'Atanasio, D. Sellitto and F. Cruciani, 2016 Evidence of extensive non-allelic gene conversion among LTR elements in the human genome. Sci Rep 6**:** 28710.

Trombetta, B., D. Sellitto, R. Scozzari and F. Cruciani, 2014 Inter- and intraspecies phylogenetic analyses reveal extensive X-Y gene conversion in the evolution of gametologous sequences of human sex chromosomes. Mol Biol Evol 31**:** 2108-2123.

Varadharajan, S., P. Rastas, A. Löytynoja, M. Matschiner, F. C. F. Calboli *et al.*, 2019 A High-Quality Assembly of the Nine-Spined Stickleback (Pungitius pungitius) Genome. Genome Biol Evol 11**:** 3291-3308.

White, M. A., J. Kitano and C. L. Peichel, 2015 Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. Mol Biol Evol 32**:** 1981-1995.

Wright, A. E., P. W. Harrison, S. H. Montgomery, M. A. Pointer and J. E. Mank, 2014 Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. Evolution 68**:** 3281-3295.

Zickler, D., 2020 Diter von Wettstein and The Meiotic Program of Pairing and Recombination. Methods Mol Biol 2124**:** 19-35.

Zickler, D., and N. Kleckner, 2015 Recombination, Pairing, and Synapsis of Homologs during Meiosis. Cold Spring Harb Perspect Biol 7.

CHAPTER 2

NON-HOMOLOGOUS PAIRING AND DOUBLE STRANDED BREAK FORMATION IN THREESPINE STICKLEBACKS

Introduction

Meiosis is a specialized form of cellular division process that results in the formation of haploid gametes. In order to produce functional gametes, the chromosomes have to segregate properly such that each gamete receives only one copy of each homologous chromosome pair. If chromosomes fail to pair and segregate correctly, severe meiotic errors can result (reviewed in (HASSOLD and HUNT 2001). In some species, pairing is initiated by the formation of double stranded breaks (DSB) that are then repaired using intact sequence from the homologous chromosome as a template (GIROUX *et al.* 1989; BAUDAT *et al.* 2000; ROMANIENKO and CAMERINI-OTERO 2000; GRELON *et al.* 2001). Heteromorphic sex chromosomes pairs (X/Y or Z/W sex chromosomes that do not share extensive sequence homology) present unique challenges in terms of pairing and DSB repair during meiosis (PAGE *et al.* 2005; DE LA FUENTE *et al.* 2007; KAUPPI *et al.* 2011; BORODIN *et al.* 2012; DE LA FUENTE *et al.* 2012; GUIOLI *et al.* 2012; DUMONT *et al.* 2018).

Our understanding of double strand break initiation and repair and pairing of heteromorphic sex chromosomes has largely been informed by detailed studies of ancient Y chromosomes in mammals and birds (PAGE *et al.* 2005; DE LA FUENTE *et al.* 2007; KAUPPI *et al.* 2011; BORODIN *et al.* 2012; DE LA FUENTE *et al.* 2012; GUIOLI *et al.* 2012;

DUMONT *et al.* 2018). Heteromorphic sex chromosomes typically have a pseudoautosomal region that is shared between the sex chromosomes that undergoes an obligate crossover each meiosis. During meiosis, the sex chromosomes in most mammals exhibit delayed pairing and they only synapse within the PAR, while the sex chromosome specific regions (non-PAR) remain largely unpaired (RASMUSSEN and HOLM 1978; CHANDLEY *et al.* 1984; KAUPPI *et al.* 2011; FEDERICI *et al.* 2015; DUMONT *et al.* 2018). In house mice, double strand break formation on sex chromosomes has been studied in great detail. Double stranded breaks on non-PAR sex chromosomes in male mice are highly suppressed and are repaired late in the meiosis when rest of the autosomal DSBs have been repaired (KAUPPI *et al.* 2011; LANGE *et al.* 2016). Since the non-PAR regions of sex chromosomes do not physically par in mice, the DSBs formed in this region are either repaired intrachromosomally using repeated sequence as a template or using sister chromatids. DSB patterns in organisms besides house mice have not been studied to date.

Sex chromosome pairing configurations are not always obligate and limited to the PAR. It can be highly variable within and across species. Sex chromosomes in marsupials have lost the PAR and hence, do not physically synapse. The components of the synaptonemal complex (SYCP3) aid in proper segregation by forming a "dense plate" structure (ROCHE *et al.* ; SOLARI and BIANCHI 1975; SHARP 1982; SELUJA *et al.* 1987; PAGE *et al.* 2005). While sex chromosomes in voles retain a PAR, they have independently lost the PAR pairing in many species (BORODIN *et al.* 2012; DUMONT *et al.* 2018). In one of the species, the sex chromosomes show PAR synapsis and complete asynapsis in equal frequency (DUMONT *et al.* 2018). Additionally, human sex chromosomes show variable levels of synapsis ranging between 10%-70% of Y axis in

continuum with PAR (RASMUSSEN and HOLM 1978; CHANDLEY *et al.* 1984). This non homologous synapsis could be responsible for inter-chromosomal genetic exchange in non-PAR regions of sex chromosomes as reported relatively recently through genomic studies at the population level (ROSSER *et al.* 2009; TROMBETTA *et al.* 2010; TROMBETTA *et al.* 2017). Furthermore, in birds, the ZW sex chromosomes pair at PAR and also in a non-homologous fashion throughout their axes in majority (88%) of meiocytes (GUIOLI *et al.* 2012; PIGOZZI 2016). Hence, sex chromosome pairing is variable across systems ranging from asynapsis in marsupials to non-homologous synapsis in birds and different mechanisms ensure faithful segregation of sex chromosomes during meiosis.

Notably, all these organisms that inform our knowledge of sex chromosome pairing have old, degenerated sex chromosomes. It is imperative to explore pairing in young sex chromosomes to study the evolution of modifications in pairing once sex chromosomes arise and start degenerating from their ancestral autosomal counterparts. Young sex chromosomes are characterized by higher sequence homology between sex chromosomes as compared to old sex chromosomes (BACHTROG 2013). We hypothesize that young sex chromosomes with intact PARs will synapse at this region during meiosis. Non-par pairing will depend on the structural arrangements, like inversions, which can impede pairing during meiosis. In guppies, for instance, inversions between sex chromosomes have not been observed and sex chromosome do not show extensive degeneration (DAROLTI *et al.* 2020; ALMEIDA *et al.* 2021). We would expect complete synapsis in sex chromosomes of this system. Alternately, due to the lack of extensive degeneration young sex chromosomes may synapse in non-PAR regions. Leaky recombination between sex chromosomes has also been observed in guppies that may

have an impact on Y chromosome degeneration (DAROLTI *et al.* 2020). This indirectly suggests that perhaps DSB rates are not as suppressed on young sex chromosomes of guppies as compared to old sex chromosome systems and higher sequence homology allows for leaky genetic exchange between sex chromosomes. However, there is no direct evidence of lack of DSB suppression in young sex chromosome systems due to lack of studies in this field. Detailed study of pairing and DSB formation and repair will yield insights into our understanding of evolution of these processes on young sex chromosomes.

Threespine sticklebacks (*Gasterosteus aculeatus*) have a young sex chromosome system which is around 14 million years old (PEICHEL *et al.* 2020; SARDELL *et al.* 2020). X and Y chromosomes in this system have three strata, as shown by the pattern of divergence of shared genes between the sex chromosomes (d_S : stratum1=0.155, stratum2=0.052, stratum3=0.033) (PEICHEL *et al.* 2020). The three strata correspond to three major inversions that distinguish the sex chromosomes as revealed by cytogenetics as well as a high-quality genome assembly (including Y) (ROSS *et al.* 2009; PEICHEL *et al.* 2020). In contrast, the synonymous divergence between shared genes on human sex chromosomes is much higher (d_S: stratum1= 2.58, stratum2/3=0.51, stratum4/5=0.15) (BELLOTT *et al.* 2014). Heteromorphic sex chromosomes can be distinguished during meiosis due to their atypical synapsis (RASMUSSEN and HOLM 1978; CHANDLEY *et al.* 1984; KAUPPI *et al.* 2011). However, it was reported that sex chromosome pair is indistinguishable during meiosis in this system (CUÑADO *et al.* 2002). Threespine sticklebacks are an idea system to study patterns of pairing and DSB formation on young

sex chromosomes that have accumulated inversions but relatively younger in terms of sequence degeneration.

We developed cytogenetics tools to study pairing and double strand break formation on the threespine stickleback sex chromosomes. We adopted surface spreading techniques form zebrafish and used immunofluorescence to delineate chromosomes and DSB foci. Furthermore, we used fluorescence in-situ hybridization probes to distinguish and study sex chromosomes and their pairing patterns.

Materials and Methods

Ethics statement

All procedures using threespine stickleback fish were approved by the University of Georgia Animal Care and Use Committee (protocol A2018 10-003-A8).

Preparation of chromosome spreads

Chromosome spreads were prepared using a modified protocol from Blokhina et al 2019 (BLOKHINA *et al.* 2019). We targeted fish that were approximately five to eight months after hatch (standard length 5.4 cm), where testes were actively undergoing meiosis (NAFTALY *et al.* 2020). Whole testes were dissected and macerated using a Dounce homogenizer in a volume of 200 ul PBS buffer. The cell suspension was centrifuged for five min at 200 xg and the pellet was resuspended in 200 ul of 100 mM sucrose. The cell suspension was incubated for five min at room temperature. 20 ul of the suspension was pipetted across a clear slide and then 100 ul of fixative (1% paraformaldehyde with 0.15% Triton-X 100) was added and left overnight in a humid chamber. The slides were washed the next day in 1X PBS three times for 15, 10, and 5 min and then stored at –20ºC.

Immunofluorescence of SMC3 and RAD51

We adapted an immunofluorescence protocol from Dumont et al 2011 (DUMONT and PAYSEUR 2011). The spreads were permeabilized by pipetting 1.5 ml of permeabilization solution (1X PBS, 1 mM EDTA, and 1% TritonX-100) on the slide and incubated for 20 min at RT (DAWE *et al.* 2018). The slides were then blocked with 1.5ml of 1X ADB using same procedure as described above. After blocking, the slides were incubated with diluted primary antibodies (1:100 anti-SMC3, ab9263; 1:40 anti-RAD51, MA5-14419) in 40 ul of 1X PBS per slide. The slides were sealed with a coverslip and rubber cement and then incubated at 4º C overnight. After the primary incubation coverslip was removed, and the slides were blocked with 1X antibody dilution buffer for 20 min. Secondary antibodies (Goat anti-rabbit was ab150077/ab150080/ab150079 and Goat anti-mouse ab150113) were applied at the same dilution as primary antibodies, sealed with a coverslip and rubber cement, and incubated for two hours at 37º C. After the secondary incubation, the slides were washed in 1X PBS three times (15 min., 5 min., and 5 min.). The slides were sealed with a coverslip and Vectashield Antifade Mounting Media (Vector Labs).

Synthesizing probes for fluorescence *in situ* hybridization

BACs were extracted from 200 ml bacterial cultures using the NucleoBond Xtra Midi kit (Takara Bio). The probes were made using a Vysis Nick Translation Kit (Abbott molecular) following manufacturer's instructions in a batch of 50ul reactions. Two µg of input DNA was used for the reaction instead of the one µg suggested as one µg did not yield any hybridization. The reagents were mixed and incubated in a PCR block for 16 hours at 15º C followed by inactivation for 15 minutes at 70º C. FISH protocol was

adapted from Ross et al. The probes generated from nick translation reaction were precipitated with salmon sperm DNA (Invitrogen 15632-011), ethanol and Sodium Acetate. For each 10ul probe, 1 ul of salmon sperm DNA (15632011, ThermoFisher Scientific), 30.25 ul of 100% ethanol and 1.1 ul of 3M sodium acetate was used and the probes were precipitated for at least 15 at room temperature in dark. The probes were then spun down for 30 minutes at 4^{o} C at 12000 rpm. The supernatant was discarded, and the pellet was left to dry at RT for 15 minutes. The pellet was then reconstituted in 2 µl of TE (pH. 8) and 8 µl of hybridization buffer (5 mL 100% formamide, 1 mL 20x SSC pH 7.0, 2 mL 50% dextran sulfate BP1585-100).

Fluorescence *in situ* hybridization on chromosome spreads

After immunofluorescence, FISH probes were hybridized to the same slides. The slides were incubated with 2X SSC at 75^{o} C for five minutes and then treated with denaturing solution for two minutes. The slides were then passed through an ethanol series (70%, 85% and 100%) for two minutes each. The slides were dried in a slanted position for 10 minutes. Then, 10 µl of the precipitated and reconstituted probe was put on the slide and cover slipped with a small square coverslip and sealed with rubber cement. The slides were incubated overnight at 37^{o} C and washed next day 3X in 45°C wash solution 1 (2x SSC/50% formamide, pH 7) and then 3X in 45°C wash solution 2 (2X SSC) for five minutes each. After drying, the slides were mounted with Vectashield Antifade Mounting Media (Vector Laboratories). All slides were imaged using Leica DM6000 B upright microscope at 63x magnification with DAPI, TRITC, Alexa633 and FITC filter sets. Images were captured using a Hamamatsu ORCA-ER digital camera.

Sex chromosome pairing

The spreads stained with SMC3 and *idh* FISH were imaged and spreads in which all chromosomes were paired were used in further analysis. In order to study the relative enrichment of each conformation in different prophase stages and their correlation with DSB repair, the spreads stained with SMC3, Rad51 and *idh* FISH probes were used. The spreads were staged into different meiotic stages by the following criteria (ACQUAVIVA *et al.* 2020). Leptotene spreads were defined by small SMC3 stretches forming along chromosomes. Zygotene was defined by long SMC3 stretches covering entire chromosomes and undergoing synapsis. The number of chromosomes in this stage was between 42 and 22, indicating not all sister chromatids had paired with their homolog (threespine stickleback fish have a haploid chromosome count of 42). If all the chromosomes were paired, they were characterized as pachytene stage. Sex chromosomes from pachytene and zygotene spreads from four different males were categorized into the three conformations: one, intermediate and end. Presence and absence of DSBs was also recorded for each spread in the three conformations. These spreads were divided into three categories according to the position of *idh* probes. If the spreads only had one probe, it was named "one", if the spreads had two probes with one signal on the tip of the sex chromosome pair and the other in the middle of the chromosome pair, it was named "end", and if both signals were interstitially located then it was named "intermediate".

The length of *X-idh* and *Y-idh* loops was measured using ImageJ (https://imagej.nih.gov/ij/). The freehand tool was used to draw a line connecting the two ends of each *idh* probe in the end and intermediate conformations. We could not measure the "one" configuration because the *X-idh* and *Y-idh* probes co-localize. The measure

utility was clicked to measure length in pixels and pixels were converted to microns by multiplying by a scaling factor of 0.13.

For the DSB distribution analysis the length of each chromosomes was measured using ImageJ and then divided into five equal parts based on the total length of the chromosome. The number of DSBs were then recorded in each interval. For the chromosomes that had no DSBs, zero was recorded for each interval. The average value was plotted for each interval along with standard errors. A total of 420 autosomes and 47 sex chromosomes from spreads in zygotene/pachytene were used for this analysis.

Double strand break counting and normalization

The chromosome spreads were staged as described in the "*Sex chromosome pairing"* section**.** For each spread, the total number of double strand breaks per spread and the number of DSBs that occurred on the sex chromosome were recorded. The counts were repeated three times and the average was recoded. The user counting the spreads was blinded to the stage of the spread. The density of DSBs per Mb was calculating by dividing the genome-wide counts by 900 Mb (assuming genome size to be roughly 450Mb) while the number of DSBs on sex chromosomes was divided by 37 Mb (the total size of the X chromosome and Y chromosome combined).

Results

The threespine stickleback X and Y chromosomes pair non-homologously during meiosis

The Y chromosome of the threespine stickleback has undergone at least three major inversions since it diverged from the common autosome ancestor with the X chromosome (ROSS *et al.* 2009; PEICHEL *et al.* 2020). The Y chromosome has also

undergone varied degrees of sequence degeneration across much of its length, resulting in a shorter estimated chromosome size (X chromosome: 20.6Mb, Y chromosome :17.8 Mb) (PEICHEL *et al.* 2020). Combined, chromosome-wide synteny between the X and Y chromosomes outside of the pseudo-autosomal region has been lost, suggesting the X and Y chromosomes may not fully pair during meiosis, similar to the ancient sex chromosomes of mammals (KAUPPI *et al.* 2011). To observe how the X and Y chromosomes pair, we used a well-characterized cytogenetic marker *(Idh)* to differentiate the chromosomes throughout prophase I (*Idh* is located in the middle of the X chromosome at 11.5 Mb and at the end of the Y chromosome at 17.8 Mb distal to PAR) (ROSS *et al.* 2009; PEICHEL *et al.* 2020). Using this marker as a probe, we find the X and Y chromosomes synapse fully during late zygotene and pachytene, when synapsis is also complete on the autosomes (Fig.2.1). We found the X-linked *Idh (X-Idh)* is in the middle of the sex chromosome pair while *Idh* on the Y chromosome *(Y-Idh)* is located further distal from the PAR on the synapsed chromosome pair (total number of spreads observed=51; full synapsis observed=50). Given the large-scale structural re-arrangements between the sex chromosomes, our results highlight that synapsis is proceeding in a non-homologous fashion and is not delayed relative to autosomes.

DNA double strand breaks form at a similar frequency on the autosomes and sex chromosomes

To quantify the total number of DSBs throughout the genome, we counted RAD51 foci, a marker for breaks undergoing repair (SHINOHARA *et al.* 1992; VAN DEN BOSCH *et al.* 2002). We observed that RAD51 dependent DSBs occurred at a density of 0.041 DSBs/Mb in leptotene. As prophase proceeds, double strand breaks are repaired

and RAD51 foci decrease (BARLOW *et al.* 1997; KOUZNETSOVA *et al.* 2011). Consistent with this, we observed significantly different DSB counts at each stage, decreasing as prophase proceeded (leptotene, n=34, mean: 37.06; zygotene, n=53, mean: 23.74; pachytene, n=34, mean: 5.35; leptotene vs. zygotene $P < 0.001$; zygotene vs pachytene $P < 0.001$; leptotene vs pachytene $P < 0.001$) (Fig. 2.2).

On the ancient Y chromosome of mammals, DSBs are suppressed relative to the autosomes (KAUPPI *et al.* 2011; LANGE *et al.* 2016). We therefore tested whether the younger X and Y chromosomes of threespine stickleback fish exhibit any suppression of DSBs or whether they exhibit counts that more closely resemble autosomes. We found that the density of DSBs sex chromosomes is not significantly different from that of autosomes in spreads in zygotene and pachytene stages (n=47; P=0.89; Wilcoxon test) (Fig 2.3A-C). We also explored the distribution of DSBs across sex chromosome pair and autosomes in zygotene and pachytene stages (Fig 2.3D). We found that most DSBs are concentrated at the ends of chromosomes in both autosomes and sex chromosomes, we did not see a significant difference between the sex chromosomes and autosomes in terms of enrichment of DSBs in these intervals. Interval five corresponded to the PAR, intervals one and two corresponded to stratum one (the oldest region), interval three to stratum two and interval four to stratum three (the two younger regions) relative to the X chromosome (PEICHEL *et al.* 2020).

The X chromosome undergoes synaptic adjustment during male meiosis

Among the fully paired X and Y chromosomes, we observed variation in the relative positions of the two *Idh* probes within the synapsed pair. The placement of these markers could be categorized into three groups. In the first group, we observed the *X-Idh*

and *Y-Idh* probes were located very close together and were indistinguishable at this level of resolution ("One probe"; n=7/50). In the second group, we observed the *X-Idh* and *Y-Idh* probes were located further apart, but the Y-*Idh* probe was not at the end of the sex chromosome axis ("Intermediate"; n=20/50, Fig. 2.1B). In the third group, the *X-Idh* was located in the middle of the SC axis and the *Y-Idh* was located at the tip of the sex chromosome axis ("End"; n=23/50, Fig. 2.1C). While the second group is the expected conformation given the sizes of sex chromosomes and positions of the *Idh* loci, we hypothesize the existence of three conformations is due to synaptic adjustment.

To identify what stage in meiosis each configuration corresponds to, we examined chromosome spreads across four males. Two of these males had spreads mostly in zygotene and pachytenes, while the rest two had spreads mostly in pachytene. We examined these two groups independently and found that the end-to-end conformations and intermediate conformations occurred more often in pachytene whereas the single focus conformations were found to be enriched during the earlier zygotene stage of prophase. For further analysis, we combined the four males and performed a pairwise Fishers exact test to test if the three conformations show differential enrichment between zygotene and pachytene stages (pachytene n=37, E=22, I=12, O=3; zygotene n=13, E=1, I=8, O=4) (Supplemental Fig 2.1, Fig 2.4). The pairwise Fisher's test showed that "End" conformation is enriched in pachytene (P=0.001) while "Intermediate" (P=0.100) and "One probe" (P=0.069) conformations do not show significant enrichment for either stages.

Synaptic adjustment occurs between chromosomes when homologs differ in length, either due to large insertions or deletions (SOLARI 1992; ZICKLER and KLECKNER

1999). This always occurs through shortening of the longer chromosomes (ASHLEY 1990; GREENBAUM *et al.* 1990; ZICKLER and KLECKNER 1999). We explored whether synaptic adjustment was occurring in the threespine stickleback fish through shortening of the X chromosome. The length of the DNA loops that extend from the chromosome axes are a useful indicator of the overall level of compaction of DNA (KLECKNER 2006). Given the same physical length of DNA, shorter axes require longer loops to extend from the axis, whereas longer axes result in smaller extending loops (KAUPPI *et al.* 2011). We therefore measured the width of DNA loops at the *X-Idh* and *Y-Idh* as a proxy for loop size in the intermediate (n=10) and end (n=18) conformations. If the X or Y chromosome are changing in length over the course of prophase I, we should see a corresponding change in DNA loop size on the respective chromosome. We found that while the *Y-Idh* loop size is the same in both conformations (intermediate conformation: 0.72µm; end conformation: 0.87µm; $P = 0.55$, Wilcoxon 2-sample test), the *X-Idh* loop increases in length (intermediate conformation: 1.13µm; end conformation: 2.03µm; $P = 0.02$, Wilcoxon 2-sample test) in the end conformation (Fig. 2.5). This suggests that the synaptic adjustment is largely due to shortening of the X chromosome axis as prophase proceeds.

Discussion

In this paper we characterized pairing and DSB formation during meiosis in male threespine stickleback meiocytes to understand how young X and Y chromosomes pair and initiate double strand break repair. The XY sex chromosome system in Threespine sticklebacks is ~14 million years old and the SCs have accumulated three major inversions which distinguish the sex chromosomes. The Y chromosomes has degenerated

over evolutionary time and lost ~3 Mb sequence, however the sex chromosomes still share a lot of homology (WHITE *et al.* 2015; PEICHEL *et al.* 2020).

Non-homologous pairing of SCs is frequently observed in nature including humans, which show variation in the extent of Y axes involved in pairing with X (CHANDLEY *et al.* 1984). We showed that SCs in this young sex chromosome system synapse fully and without delay compared to their autosomal counterparts. This is in agreement with a previous study done in threespine stickleback system to elucidate sex specific differences in meiosis (CUÑADO *et al.* 2002). The homologs including sex chromosomes in this system are not paired prior to entering meiosis as in drosophila so the absence of any delay is not due to somatic pairing of SCs (STACK and BROWN 1969). While we observed three different conformations of synapsed axes, we do not see any mouse-like conformations where only PAR is involved in synapsis. This is similar to non-homologous pairing observed in ZW SCs in chickens where SCs pair fully in 88% of the oocytes (GUIOLI *et al.* 2012). We also showed that the synaptic adjustment during later stages of prophase is driven by shortening of X chromosome relative to Y to minimize the un-synapsed region. It remains to be addressed if the non-homologous interaction between SCs in this system is driven due to homology searching by DSB repair pathway. Knocking out *Spo11* will perhaps address this issue (ROMANIENKO and CAMERINI-OTERO 2000). In the absence of DSBs formed by *Spo11* the sex chromosomes may fail to pair fully, then it will reveal that the pairing of sex chromosomes in threespine sticklebacks is perhaps aided by homology search in order to repair DSBs. Inversions offer a simple and abrupt solution to the problem of recombination events that may unlink a sexually antagonistic gene from its desired chromosome. However, inversions

can still allow for gene conversions and double crossovers (KORUNES and NOOR 2019). Some sex chromosome systems undergo gradual recombination suppression in the absence of inversions. *Poecilia reticulata* and *P. wingei* have a small stratum 1 (~3Mb) and a larger stratum 2 (~17 Mb). The recombination is suppressed on almost the entire length of the sex chromosomes; however, the recombination is not complete in the second stratum and occasional recombination events may hinder the degeneration of sex chromosomes (DAROLTI *et al.* 2020).

Stratum one on X chromosomes remains unsynapsed till Pachytene when the synaptic adjustment brings the X and Y to same length. Hence DSBs forming on stratum one must get repaired by using template from sister chromatids.

Sex chromosomes can vary in their pairing behavior even in some mammals, as observed in closely related species of voles (BORODIN *et al.* 2012; DUMONT *et al.* 2018). While sex chromosomes in most Voles pair at PAR, this pairing has been lost independently in multiple species of voles. Closely related species of Threespine sticklebacks also have young sex chromosomes all ranging between 1-20 million years old: a neo-sex chromosome system in *G. wheatlandi* and *G. nipponicus*, XY in *P. pungitius* and ZW in *A. quadracus* (ROSS *et al.* 2009; DIXON *et al.* 2019; SARDELL *et al.* 2020). The *G. wheatlandi* and Threespine sticklebacks share the same ancestral XY sex chromosome system (using linkage group 19) with the exception that Y chromosomes in G. wheatlandi have fused with a different chromosome (linkage group 12) giving rise to a neo-Y. The LG19 region of Y chromosome in *G. wheatlandi* has undergone a more extensive degeneration as compared to the Threespine sticklebacks although they have a common ancestor. It would be interesting to explore if sex chromosomes in these

species, specifically in more degenerated *G. wheatlandi*, also show similar patterns as Threespine sticklebacks and define the ancestral pairing state in this clade.

We also assessed the frequency and patterns of DSB formation on autosomes and sex chromosomes in this system. We observe the highest number of breaks in leptotene and a reduction through zygotene and pachytene. Since we Since we only counted RAD51 dependent DSBs in our experiments, we could be potentially missing the DMC1 dependent breaks and underestimating the total number of DSBs. While in mouse, RAD51 and DMC1 colocalize strongly (89%) on all DSB foci, in *Arabidopsis thaliana* RAD51 and DMC1 only localize on around 40% of the foci in zygotene (TARSOUNAS *et al.* 1999; KURZBAUER *et al.* 2012). We also observed spreads with highest number of DSBs to be in leptotene which decrease as the prophase progresses. This pattern is distinct from mammals where most DSBs are observed in early zygotene stage. We interpret that this pattern could be due to quick repair of DSBs or displacement of RAD51 as they are forming and, hence, we are perhaps not capturing all the DSBs formed (BLOKHINA *et al.* 2019). Despite the caveats, we observe the similar rates of DSB formation in sex chromosomes and autosomes. This shows that perhaps the DSB suppression is a function of sequence degeneration and haplotypes lacking DSBs have not been under strong selection. However, due to complete synapsis, we could not distinguish the number of DSBs forming on X chromosomes vs Y chromosomes and can be better assessed using ChIP-Seq against Rad51 or ssDNA pulldown with Spo11 (LANGE *et al.* 2016; HINCH *et al.* 2020). Since asynapsis and unrepaired DSBs during meiosis triggers silencing of sex chromosomes in some organisms, we expect Threespine stickleback SCs to lack sex chromosome-wide MSCI. It would be interesting to study the

sex chromosome gene expression and regulation during meiosis in Threespine sticklebacks. Interestingly, ZW chromosomes in chickens lack sex chromosome-wide MSCI as well (GUIOLI *et al.* 2012).

Figures

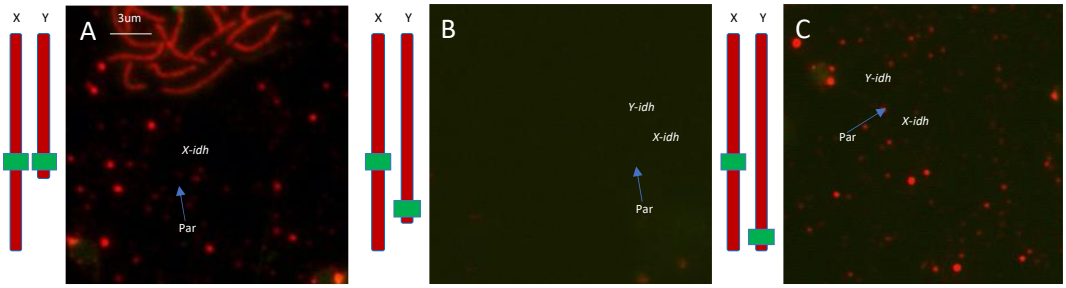


Figure2.1**.** (A) shows the "One probe" conformation of sex chromosome pair during meiosis. (B) shows the "Intermediate" state while (C) shows the "End" state. SMC3 (red) marks the chromosomal axes while FISH with *idh* probe (green) distinguishes the sex chromosomes from autosomes. Signals corresponding to X chromosome is marked X-idh while signal corresponding to Y chromosome is marked Y-idh.

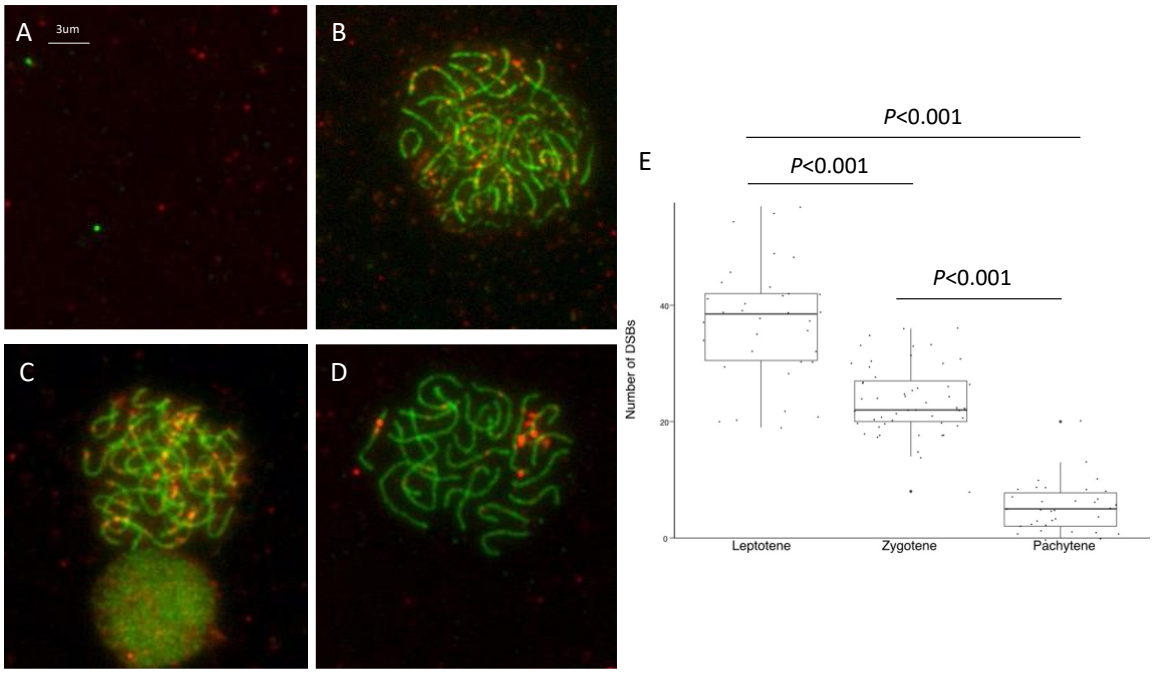


Figure2.2. (A) leptotene (B) Early-zygotene (C) Late-zygotene and (D) pachytene spreads stained with SMC3(green) and Rad51(red). (E) Number of DSBs on chromosome spreads across leptotene, zygotene and pachytene stages of meiosis assayed using Rad51.

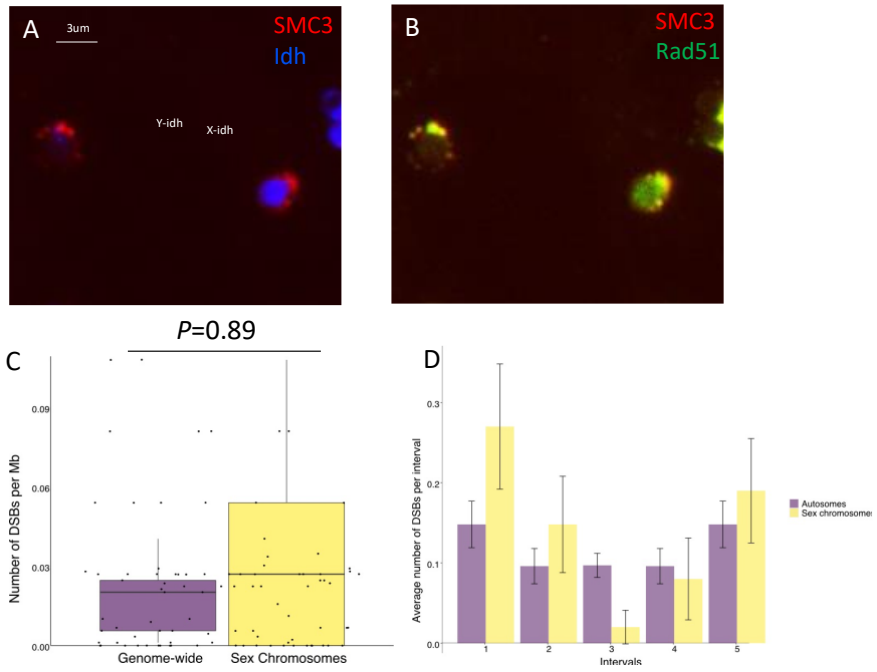


Figure2.3. Late zygotene spread stained with SMC3, Rad51 and probed with *idh*. (A) A spread with SMC3 and idh probe and the same image stained for Rad51 in (B). (C) Density of DSBs in Sex chromosomes vs Genome-wide in late prophase (late zygotene and pachytene). (D) Distribution of DSBs across autosomes and sex chromosome pair during late prophase (late zygotene and pachytene)

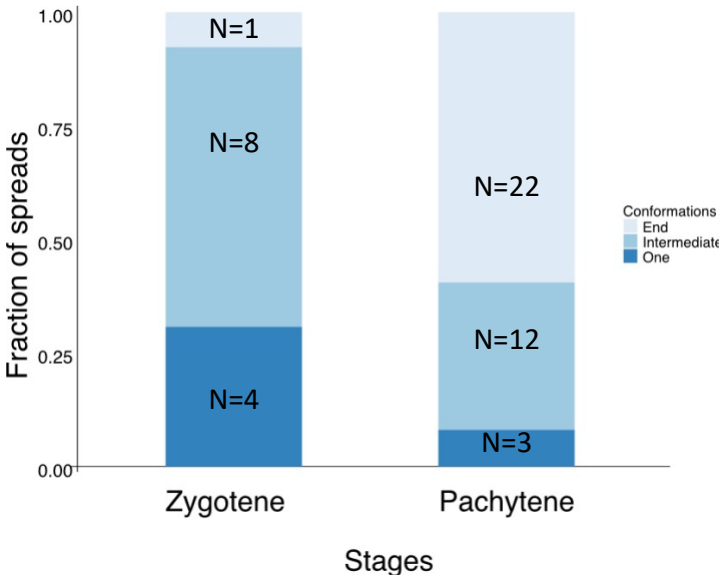


Figure2.4. Relative enrichment of “One”, “End” and “Intermediate” conformations in zygotene and pachytene spreads.

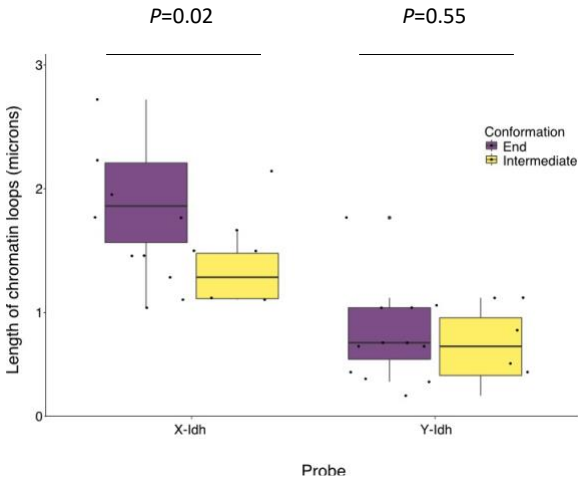


Figure2.5. Loop width of *X-idh* and *Y-idh* in “End” and “Intermediate” conformations.

References

Acquaviva, L., M. Boekhout, M. E. Karasu, K. Brick, F. Pratto *et al.*, 2020 Ensuring meiotic DNA break formation in the mouse pseudoautosomal region. Nature 582**:** 426-431.

Almeida, P., B. A. Sandkam, J. Morris, I. Darolti, F. Breden *et al.*, 2021 Divergence and Remarkable Diversity of the Y Chromosome in Guppies. Mol Biol Evol 38**:** 619-633.

Ashley, T., 1990 Axial shortening during pachynema unrelated to nonhomologous synapsis. Cytogenet Cell Genet 53**:** 185-190.

Bachtrog, D., 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet 14**:** 113-124.

Barlow, A. L., F. E. Benson, S. C. West and M. A. Hultén, 1997 Distribution of the Rad51 recombinase in human and mouse spermatocytes. EMBO J 16**:** 5207-5215.

Baudat, F., K. Manova, J. P. Yuen, M. Jasin and S. Keeney, 2000 Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. Mol Cell 6**:** 989-998.

Bellott, D. W., J. F. Hughes, H. Skaletsky, L. G. Brown, T. Pyntikova *et al.*, 2014 Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508**:** 494-499.

Blokhina, Y. P., A. D. Nguyen, B. W. Draper and S. M. Burgess, 2019 The telomere bouquet is a hub where meiotic double-strand breaks, synapsis, and stable homolog juxtaposition are coordinated in the zebrafish, Danio rerio. PLoS Genet 15**:** e1007730.

Borodin, P. M., E. A. Basheva, A. A. Torgasheva, O. A. Dashkevich, F. N. Golenishchev *et al.*, 2012 Multiple independent evolutionary losses of XY pairing at meiosis in the grey voles. Chromosome Res 20**:** 259-268.

Chandley, A. C., P. Goetz, T. B. Hargreave, A. M. Joseph and R. M. Speed, 1984 On the nature and extent of XY pairing at meiotic prophase in man. Cytogenet Cell Genet 38**:** 241-247.

Cuñado, N., J. Barrios, E. S. Miguel, R. Amaro, C. Fernández *et al.*, 2002 Synaptonemal complex analysis in oocytes and spermatocytes of threespine stickleback Gasterosteus aculeatus (Teleostei, Gasterosteidae). Genetica 114**:** 53-56.

Darolti, I., A. E. Wright and J. E. Mank, 2020 Guppy Y Chromosome Integrity Maintained by Incomplete Recombination Suppression. Genome Biol Evol 12**:** 965-977.

Dawe, R. K., E. G. Lowry, J. I. Gent, M. C. Stitzer, K. W. Swentowsky *et al.*, 2018 A Kinesin-14 Motor Activates Neocentromeres to Promote Meiotic Drive in Maize. Cell 173**:** 839-850.e818.

de la Fuente, R., M. T. Parra, A. Viera, A. Calvente, R. Gómez *et al.*, 2007 Meiotic pairing and segregation of achiasmate sex chromosomes in eutherian mammals: the role of SYCP3 protein. PLoS Genet 3**:** e198.

de la Fuente, R., A. Sánchez, J. A. Marchal, A. Viera, M. T. Parra *et al.*, 2012 A synaptonemal complex-derived mechanism for meiotic segregation precedes the evolutionary loss of homology between sex chromosomes in arvicolid mammals. Chromosoma 121**:** 433-446.

Dixon, G., J. Kitano and M. Kirkpatrick, 2019 The Origin of a New Sex Chromosome by Introgression between Two Stickleback Fishes. Mol Biol Evol 36**:** 28-38.

Dumont, B. L., and B. A. Payseur, 2011 Genetic analysis of genome-scale recombination rate evolution in house mice. PLoS Genet 7**:** e1002116.

Dumont, B. L., C. L. Williams, B. L. Ng, V. Horncastle, C. L. Chambers *et al.*, 2018 Relationship Between Sequence Homology, Genome Architecture, and Meiotic Behavior of the Sex Chromosomes in North American Voles. Genetics 210**:** 83-97.

Federici, F., E. Mulugeta, S. Schoenmakers, E. Wassenaar, J. W. Hoogerbrugge *et al.*, 2015 Incomplete meiotic sex chromosome inactivation in the domestic dog. BMC Genomics 16**:** 291.

Giroux, C. N., M. E. Dresser and H. F. Tiano, 1989 Genetic control of chromosome synapsis in yeast meiosis. Genome 31**:** 88-94.

Greenbaum, I. F., D. W. Hale, P. D. Sudman and E. Nevo, 1990 Synaptonemal complex analysis of mole rats (Spalax ehrenbergi): unusual polymorphisms of chromosome 1. Genome 33**:** 898-902.

Grelon, M., D. Vezon, G. Gendrot and G. Pelletier, 2001 AtSPO11-1 is necessary for efficient meiotic recombination in plants. EMBO J 20**:** 589-600.

Guioli, S., R. Lovell-Badge and J. M. Turner, 2012 Error-prone ZW pairing and no evidence for meiotic sex chromosome inactivation in the chicken germ line. PLoS Genet 8**:** e1002560.

Hassold, T., and P. Hunt, 2001 To err (meiotically) is human: the genesis of human aneuploidy. Nat Rev Genet 2**:** 280-291.

Hinch, A. G., P. W. Becker, T. Li, D. Moralli, G. Zhang *et al.*, 2020 The Configuration of RPA, RAD51, and DMC1 Binding in Meiosis Reveals the Nature of Critical Recombination Intermediates. Mol Cell 79**:** 689-701.e610.

Kauppi, L., M. Barchi, F. Baudat, P. J. Romanienko, S. Keeney *et al.*, 2011 Distinct properties of the XY pseudoautosomal region crucial for male meiosis. Science 331**:** 916-920.

Kleckner, N., 2006 Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex. Chromosoma 115**:** 175-194.

Korunes, K. L., and M. A. F. Noor, 2019 Pervasive gene conversion in chromosomal inversion heterozygotes. Mol Ecol 28**:** 1302-1315.

Kouznetsova, A., R. Benavente, A. Pastink and C. Höög, 2011 Meiosis in mice without a synaptonemal complex. PLoS One 6**:** e28255.

Kurzbauer, M. T., C. Uanschou, D. Chen and P. Schlögelhofer, 2012 The recombinases DMC1 and RAD51 are functionally and spatially separated during meiosis in Arabidopsis. Plant Cell 24**:** 2058-2070.

Lange, J., S. Yamada, S. E. Tischfield, J. Pan, S. Kim *et al.*, 2016 The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair. Cell 167**:** 695-708 e616.

Naftaly, A. S., S. Pau and M. A. White, 2020 Long-read RNA sequencing reveals widespread sex-specific alternative splicing in threespine stickleback fish. bioRxiv**:** 2020.2011.2012.380428.

Page, J., S. Berríos, M. T. Parra, A. Viera, J. A. Suja *et al.*, 2005 The program of sex chromosome pairing in meiosis is highly conserved across marsupial species: implications for sex chromosome evolution. Genetics 170**:** 793-799.

Peichel, C. L., S. R. McCann, J. A. Ross, A. F. S. Naftaly, J. R. Urton *et al.*, 2020 Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol 21**:** 177.

Pigozzi, M. I., 2016 The Chromosomes of Birds during Meiosis. Cytogenet Genome Res 150**:** 128-138.

Rasmussen, S. W., and P. B. Holm, 1978 Human meiosis II. Chromosome pairing and recombination nodules in human spermatocytes. Carlsberg Research Communications 43**:** 275-327.

Roche, L., G. Seluja and R. Wettstein, The meiotic behaviour of the XY pair in Lutreolina crassicaudata (Marsupialia: Didelphoidea).

Romanienko, P. J., and R. D. Camerini-Otero, 2000 The mouse Spo11 gene is required for meiotic chromosome synapsis. Mol Cell 6**:** 975-987.

Ross, J. A., J. R. Urton, J. Boland, M. D. Shapiro and C. L. Peichel, 2009 Turnover of sex chromosomes in the stickleback fishes (gasterosteidae). PLoS Genet 5**:** e1000391.

Rosser, Z. H., P. Balaresque and M. A. Jobling, 2009 Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. Am J Hum Genet 85**:** 130-134.

Sardell, J. M., M. P. Josephson, A. C. Dalziel, C. L. Peichel and M. Kirkpatrick, 2020 Contrasting tempos of sex chromosome degeneration in sticklebacks. bioRxiv**:** 2020.2009.2017.300236.

Seluja, G. A., L. Roche and A. J. Solari, 1987 Male meiotic prophase in Didelphis albiventris: A comparative cytological and electron microscopical study. Journal of Heredity 78**:** 218-222.

Sharp, P., 1982 Sex chromosome pairing during male meiosis in marsupials. Chromosoma 86**:** 27-47.

Shinohara, A., H. Ogawa and T. Ogawa, 1992 Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. Cell 69**:** 457-470.

Solari, A. J., 1992 Equalization of Z and W axes in chicken and quail oocytes. Cytogenet Cell Genet 59**:** 52-56.

Solari, A. J., and N. O. Bianchi, 1975 The synaptic behaviour of the X and Y chromosomes in the marsupial Monodelphis dimidiata. Chromosoma 52**:** 11-25.

Stack, S. M., and W. V. Brown, 1969 Somatic pairing, reduction and recombination: an evolutionary hypothesis of meiosis. Nature 222**:** 1275-1276.

Tarsounas, M., T. Morita, R. E. Pearlman and P. B. Moens, 1999 RAD51 and DMC1 form mixed complexes associated with mouse meiotic chromosome cores and synaptonemal complexes. J Cell Biol 147**:** 207-220.

Trombetta, B., F. Cruciani, P. A. Underhill, D. Sellitto and R. Scozzari, 2010 Footprints of X-to-Y gene conversion in recent human evolution. Mol Biol Evol 27**:** 714-725.

Trombetta, B., E. D'Atanasio and F. Cruciani, 2017 Patterns of Inter-Chromosomal Gene Conversion on the Male-Specific Region of the Human Y Chromosome. Front Genet 8**:** 54.

van den Bosch, M., P. H. Lohman and A. Pastink, 2002 DNA double-strand break repair by homologous recombination. Biol Chem 383**:** 873-892.

White, M. A., J. Kitano and C. L. Peichel, 2015 Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. Mol Biol Evol 32**:** 1981-1995.

Zickler, D., and N. Kleckner, 1999 Meiotic chromosomes: integrating structure and function. Annu Rev Genet 33**:** 603-754.

CHAPTER 3

IMPROVED CONTIGUITY OF THE THREESPINE STICKLEBACK GENOME USING LONG-READ SEQUENCING[1]

[1] Shivangi Nath, Daniel E Shaw, Michael A White, Improved contiguity of the threespine stickleback genome using long-read sequencing, G3 Genes|Genomes|Genetics, Volume 11, Issue 2, February 2021, jkab007, https://doi.org/10.1093/g3journal/jkab007. Reprinted here with permission from the publisher.

Abstract

While the cost and time for assembling a genome has drastically decreased, it still remains a challenge to assemble a highly contiguous genome. These challenges are rapidly being overcome by the integration of long-read sequencing technologies. Here, we use long-read sequencing to improve the contiguity of the threespine stickleback fish (*Gasterosteus aculeatus*) genome, a prominent genetic model species. Using Pacific Biosciences sequencing, we assembled a highly contiguous genome of a freshwater fish from Paxton Lake. Using contigs from this genome, we were able to fill over 76% of the gaps in the existing reference genome assembly, improving contiguity over five-fold. Our gap filling approach was highly accurate, validated by 10X Genomics long-distance linked-reads. In addition to closing a majority of gaps, we were able to assemble segments of telomeres and centromeres throughout the genome. This highlights the power of using long sequencing reads to assemble highly repetitive and difficult to assemble regions of genomes. This latest genome build has been released through a newly designed community genome browser that aims to consolidate the growing number of genomics datasets available for the threespine stickleback fish.

Introduction

Reference genome assemblies have been invaluable in the discovery of genes, the annotation of regulatory regions, and for providing a scaffold for understanding genetic variation within a species. With the advent of new sequencing technologies and the reduction of cost, there has been a rapid increase in the total number of reference genomes available across taxa. Although it has become much simpler to produce a draft reference assembly, the completion of a high-quality, contiguous assembly remains a

great challenge. There are many regions within individual genomes that are unassembled. These regions are enriched for highly repetitive sequence that cannot be assembled using sequencing technologies that produce short fragments (GNERRE *et al.* 2011; NAGARAJAN and POP 2013). Even the most highly refined genomes, like the human genome still have many gaps, which often are composed of long segmental duplications (SCHNEIDER *et al.* 2017).

Long-read sequencing technologies (Oxford Nanopore and Pacific Biosciences) have shown promise in spanning highly repetitive regions of genomes, bridging previously intractable gaps in assemblies to improve overall contiguity. Within the human genome, many highly repetitive regions have been resolved, such as pericentromeres (VOLLGER *et al.* 2020), complete centromeres (JAIN *et al.* 2018b), telomeres (JAIN *et al.* 2018a) and the entire major histocompatibility complex (JAIN *et al.* 2018a). *De novo* assemblies of highly repetitive Y chromosomes have also become feasible using long-read sequencing (CHANG and LARRACUENTE 2019; PEICHEL *et al.* 2020). Overall, long-read sequencing has enabled chromosome-scale assemblies in multiple species, including many teleost fish (CONTE *et al.* 2019; ZHOU *et al.* 2019; HE *et al.* 2020; HERAS *et al.* 2020; LIU *et al.* 2020; MIGA *et al.* 2020; PROST *et al.* 2020). It is clear that hybrid assembly approaches incorporating long-read sequencing have greatly improved contiguity of genomes.

Here we use long-read sequencing to generate a *de novo* Paxton Lake male genome assembly and improve the most recent version of the threespine stickleback reference assembly. The threespine stickleback fish has been an important model system to understand evolution, ecology, physiology, and toxicology (WOOTTON 1976; BELL and

FOSTER 1994). The identification of the genetic mechanisms underlying many adaptative traits was facilitated by the release of a high-quality reference assembly (JONES *et al.* 2012). This reference assembly was constructed from a single female fish from Bear Paw Lake (Alaska, USA) using paired-end Sanger sequencing of multiple genomic libraries. Contigs were scaffolded to genetic linkage maps, which resulted in 21 chromosome-level scaffolds (400.4 Mb), with 60.7 Mb of unplaced scaffolds. The assembly has undergone several revisions, using high-density genetic linkage maps from multiple populations (ROESTI *et al.* 2013; GLAZER *et al.* 2015), and a Hi-C proximity-guided assembly from a male from Paxton Lake (PEICHEL *et al.* 2017). Despite multiple revisions, the latest version of the assembly (v. 4) still contains 13,538 gaps and 20.6 Mb of unplaced scaffolds (PEICHEL *et al.* 2017). The gaps between contigs in the chromosome scaffolds likely represent repetitive regions or GC-rich regions, which have been shown to be recalcitrant to traditional assembly methods (BENJAMINI and SPEED 2012; ROSS *et al.* 2013).

We first generated a *de novo* assembly of a Paxton Lake benthic male threespine stickleback fish. Paxton Lake, in particular, has been a focal population of threespine stickleback fish to understand the genomic basis of sympatric speciation (MCPHAIL 1992; HATFIELD and SCHLUTER 1999; ARNEGARD *et al.* 2014). Chromosome-level scaffolds of the X and Y from a Paxton lake benthic male were recently assembled using a combination of PacBio sequencing and chromatin conformation capture sequencing (Hi-C) (PEICHEL *et al.* 2020). We used the remaining assembled autosomal contigs, combined with Hi-C sequencing and optical mapping to produce contiguous chromosome-level autosome scaffolds. We show this assembly is highly colinear with the reference Bear

Paw Lake reference genome (v. 4). To improve the contiguity of the existing v. 4 assembly, we used the Paxton Lake contigs to fill gaps in the assembly. We were able to close 76.7% of the gaps, incorporating 13.5% of the previously unplaced scaffolds. Closed gaps were highly accurate, verified through long-distance linked-read information. In addition, we were able to extend sequence of many of chromosomes into telomeres. This new v. 5 assembly represents a noteworthy improvement, allowing researchers to interrogate many previously inaccessible repetitive regions, and highlights the power of long-read sequencing to substantially improve genome contiguity.

Methods

Ethics statement

All procedures using threespine stickleback fish were approved by the University of Georgia Animal Care and Use Committee (protocol A2018 10-003-Y2-A5).

Paxton Lake benthic male *de novo* assembly

A male Paxton Lake benthic threespine stickleback fish (Texada Island, British Columbia) was previously sequenced using PacBio to approximately 75x coverage (NCBI BioProject database accession PRJNA591630; (PEICHEL *et al.* 2020) and assembled into contigs using Canu (KOREN *et al.* 2017). The Canu contigs were previously polished using Arrow (PEICHEL *et al.* 2020). This assembly had a total of 3593 contigs (N50: 683 kb) from across the genome. X- and Y-linked reads were previously separated from this set of contigs (PEICHEL *et al.* 2020), leaving a total of 3134 contigs from across the remainder of the genome. Contigs were assembled into scaffolds using Hi-C proximity guided scaffolding, derived from a different male from the Paxton Lake benthic population (NCBI SRA database: PRJNA336561;(PEICHEL *et al.* 2017). Hi-C

reads were aligned to the autosome contigs using Juicer (v. 1.5.6) (DURAND *et al.* 2016) . Autosomes were scaffolded using 3D-DNA (v. 180114 with --editor repeat coverage 11 (DURAND *et al.* 2016; DUDCHENKO *et al.* 2017; PEICHEL *et al.* 2020). Accuracy of the scaffolding was verified using BioNano optical maps (Table S1). Previously produced optical maps from a different male from the Paxton Lake benthic population (PEICHEL *et al.* 2017) were aligned to Paxton Lake *de novo* assembly using HybridScaffold.pl within the BionanoSolve software package (v. 3.4). We removed contigs from the Hi-C scaffolds that were not also supported by the optical contigs. A contig was not supported if less than 50% of its length did not overlap with an optical contig. Alignments between the Paxton Lake assembly and the optical contigs were visualized using MapOptics (Burgin et al. 2018). Unsupported contigs were removed using a custom Perl script. Collinearity between the Paxton Lake assembly and the v. 4 reference assembly was assessed using nucmer in the MUMmer software package (KURTZ *et al.* 2004). Nucmer was run with default parameters and --mum. Alignments were filtered using delta-filter, retaining alignments with an alignment identity greater than 98% and alignment lengths greater than 4 kb.

Closing gaps in the reference assembly

Version four of the threespine stickleback reference assembly contains 1263 unplaced contigs (chr. Un) that were narrowed to chromosomes but were not placed into specific gaps (there was a total of 3378 chr. Un contigs: 1263 contigs were narrowed broadly to chromosomes and 2115 could not be localized to any chromosome) (PEICHEL *et al.* 2017). We used the 1263 chr. Un contigs that were previously narrowed to chromosomes in combination with the Paxton Lake Canu contigs to independently fill the

remaining gaps in the reference assembly. To create the v. 5 assembly, we closed gaps in the v. 4 reference assembly using LR_Gapcloser with the parameter -a 1 (XU *et al.* 2019). We increased the allowed deviation between gap length and the inserted sequence length to provide additional flexibility for gap size that was not inferred accurately in the v. 4 reference assembly. LR_Gapcloser fills existing gaps in the reference assembly by identifying contigs which span a gap completely or partially from either end. Three Paxton Lake Canu contigs caused a reduction in total chromosome size after placement into gaps. Alignment of these contigs to the v. 4 reference assembly shows a small region of homology not linear with the rest of the contig which caused LR_Gapcloser to erroneously ligate the two ends of the gaps (Fig. S3.1). We omitted these three contigs from further analysis. We used BLAT (v. 3.5; (KENT 2002) to identify which of the 1263 previously narrowed chr. Un contigs from the reference assembly were placed within a gap. We filtered for stringent alignments by only retaining matches where at least 90% of the query length aligned to the assembly and the total aligned region had 2% or less mismatches.

Many chr. Un contigs that were not placed in the v. 4 reference assembly may be represented in the v. 5 assembly if they were contained completely within a Paxton Lake Canu contig (PEICHEL *et al.* 2020). To test this, we used BLAT to align the 3378 chr. Un contigs to the new v. 5 assembly. We filtered for stringent alignments by only retaining matches where at least 90% of the query length aligned to the assembly and the total aligned region had 2% or less mismatches. Chr. Un contigs that did not align to the assembly were retained as unassembled and concatenated into a single fasta sequence,

with each contig separated by 100 N's (total length: 19.88 Mb with N's; 19.59 Mb without N's). Our assembly pipeline is summarized in Fig. S3.2

Validation of the closed gaps in the v. 5 reference assembly

We verified that gaps were closed correctly in the reference assembly using two approaches. First, we validated that the contigs LR_Gapcloser used to close gaps in the reference assembly were collinear with the Paxton Lake *de novo* assembly. Sequence from the closed gaps from the v. 5 reference assembly were aligned to the Paxton Lake *de novo* assembly using the nucmer utility in MUMMER (v. 4) (KURTZ *et al.* 2004). We also aligned the longer v. 5 contigs, split at gaps that were not closed, to the v. 4 reference assembly. Alignments were stringently filtered for an overall sequence identity greater than 98% and for a minimum length aligned of 4kb. Second, we used long-distance linked-read sequencing from a female fish from a different freshwater population (Lake Washington, Washington, USA). Segments supported by two independently derived freshwater populations (Paxton Lake and Lake Washington) would suggest gaps closed in the reference assembly (Bear Paw Lake) represent the ancestral state, likely shared among all populations of threespine stickleback fish.

For the linked-read sequencing, we extracted high molecular weight DNA from blood using alkaline lysis. Blood was collected from euthanized fish into 0.85x SSC buffer. The cells were collected by centrifuging for two minutes at 2000 xg. Pelleted cells were resuspended in five ml of 0.85x SSC and 27 µl of 20 µg/ml Proteinase K solution. To lyse the cells, five ml of 2x SDS buffer (80mM EDTA, 100mM Tris pH 8.0, and 1% SDS) was added to the suspension and the solution was incubated at 55°C for two minutes. After incubation, 10 ml of buffered phenol/chloroform/isoamyl-alcohol was

added to the suspension. The suspension was incubated at room temperature under slow rotation for 30 minutes. The suspension was centrifuged for one minute at 2000 xg at 4°C to separate phases. The aqueous phase was extracted, 10 ml of chloroform was added, and the suspension was rotated for one hour. The chloroform extraction step was repeated twice. After all extractions, the aqueous phase was separated and mixed with ice cold 100% ethanol and one ml of 3M sodium-acetate (pH 5.5). The tube was gently inverted until a spool of DNA was observed. The DNA spool was transferred to a two ml tube filled with 70% cold ethanol and pelleted at 500 xg for two minutes. The DNA was allowed to dry at room temperature and resuspended in nuclease free water. Wide bore pipette tips were used for the whole procedure to minimize shearing. The integrity and size of the high molecular weight DNA was verified using a high sensitivity large fragment analysis on a fragment analyzer (Advanced Analytical Technologies, CA, USA). Genomic DNA was size selected to exclude fragments below 50 kb. Linked-read library preparation and sequencing (one Illumina NextSeq 2x150 bp high-output flow cell) was conducted by the Georgia Genomics and Bioinformatics Core (University of Georgia, GA, USA). Longranger (v. 2.2.2) was used to trim barcodes from the raw sequences and align the trimmed sequences to the new v. 5 assembly in wgs mode with default parameters (https://github.com/10XGenomics/longranger). The overall alignment rate of linked-reads to the assembly was 84.4%, resulting in a genome-wide mean read depth of 26.1X.

Assessing the completeness of the v. 5 reference assembly

We assessed the completeness of the v. 5 reference assembly by identifying universal single copy orthologs (BUSCO) throughout the new assembly, compared to the

previous v. 4 assembly (PEICHEL *et al.* 2017). BUSCO (v. 3.0.2) was run using default parameters, comparing against the Actinopterygii lineage dataset (4584 total single copy orthologs; OrthoDB v. 9) (SIMÃO *et al.* 2015). Actinopterygii was used because threespine stickleback fish are teleosts, which is the largest infraclass of Actinopterygii.

Identification of telomeric sequences

PacBio long reads with highly repetitive regions are often not assembled into contigs. We identified the telomeric reads by searching for the ancestral metazoan telomeric motif 'TTAGGG' or 'CCCTAA' (MOYZIS *et al.* 1988; MEYNE *et al.* 1989; TRAUT *et al.* 2007) in the raw PacBio reads. We searched for the motif and their respective counts in each read using the awk command-line utility. Reads were considered for further analyses if they had more than 50 occurrences of the motif. These reads were aligned to the v. 5 reference assembly using minimap2 (v. 2.17) (LI 2018) with default parameters to map to PacBio genomic reads (-ax map-pb). Only the primary alignments were retained. Telomeric reads were assigned to a specific chromosome if greater than 10 kb of unique sequence overlapped with one end of a chromosome. Positive telomeric alignments were merged with the v. 5 reference assembly. Repetitive sequence content within telomeres were visualized using the dotplot function in Geneious Prime (v2019 1.1) (https://www.geneious.com).

Identification of centromeric sequences

BLAST+ (blastn; v. 2.7.1) (CAMACHO *et al.* 2009) was used to identify the 186 bp threespine stickleback CENP-A monomer repeat (CECH and PEICHEL 2015) in the PacBio Canu assembled contigs. Contigs containing CENP-A repeats were mapped to the new v. 5 repeat masked assembly (see Genome annotation and repeat masking) using minimap2

(LI 2018) with default parameters to map to PacBio genomic reads (-ax map-pb). Contigs were only retained if greater than 10 kb of sequence mapped uniquely to one chromosome side. The number of CENP-A repeats per chromosome were counted using blastn. Dotplots were generated using Geneious Prime (v2019 1.1) (https://www.geneious.com).

Genome annotation and repeat masking

Genome features were lifted over from the previous reference assembly (v. 4) using a hybrid approach. Genome features were first lifted over to the new assembly using the software package flo (PRACANA *et al.* 2017). Most of the features were lifted over successfully (98.1%). We used BLAT to lift over the remaining fraction. The sequence for the features not lifted over with flo were extracted from the version four assembly using samtools faidx. These sequences were then aligned to the new assembly using BLAT. For each feature, the longest alignment was chosen.

Many of the closed gaps were not represented in the previous reference assembly (v. 4) and were therefore unannotated. We annotated these regions using the MAKER (v. 3.01.02) genome annotation pipeline (CANTAREL *et al.* 2008; HOLT and YANDELL 2011). These annotations combined evidence from multiple RNA-seq transcriptomes, all predicted Ensembl protein sequences (release 95), and *ab initio* gene predictions from SNAP (v. 2006-07-28) (KORF 2004) and Augustus (v. 3.3.2) (STANKE *et al.* 2006). MAKER was run over three rounds using the RNA-seq transcriptomes and methods previously described (PEICHEL *et al.* 2020).

Repeats were annotated across the genome using a combination of RepeatModeler (v. 1.0.11) and RepeatMasker (v. 4.0.5) (http://www.repeatmasker.org). Repeats were

first modeled using the default parameters of RepeatModeler. Repeats were then annotated and masked using RepeatMasker with default parameters and the custom RepeatModeler database.

We tested for enrichment of repeats and genes in closed gaps throughout the genome by comparing to randomly drawn 10 Mb segments (we placed 9.9 Mb of sequence within gaps; see Results). We also tested for enrichment of repeats and transposable elements in the remainder of the unplaced chr. Un contigs by comparing to randomly drawn 20 Mb segments throughout the assembled genome (19.59 Mb of chr. Un contigs remained unplaced; see Results). We generated a null distributions by randomly drawing 10,000 segments throughout the genome using bedtools (v. 2.29.2) shuffle (QUINLAN and HALL 2010). We then used bedtools intersect to count the number of repeats (with option -c for both 10Mb and 15Mb segments) as well as the number of bases that overlapped genes (with option -wao for 10 Mb segments) within each random segment.

Data availability

Long-distance linked-read sequencing, the v. 5 reference assembly, and the Paxton Lake *de novo* assembly, are available on the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA639125. The v. 5 reference assembly is also available for download and browsing from the threespine stickleback genome browser (https://stickleback.genetics.uga.edu). All supplemental material has been uploaded to figshare.

Results and Discussion

The Paxton Lake genome was assembled into chromosome-level scaffolds

A total of 3134 contigs from across the autosomes were assembled into 20 chromosome-level scaffolds. The initial Hi-C proximity guided scaffolded assembly resulted in a total autosome length that was considerably larger than the v. 4 reference assembly (v. 4 reference assembly: 416.97 Mb; Paxton Lake assembly: 473.16 Mb), suggesting there were contigs that were erroneously scaffolded into each chromosome. To explore this, we incorporated long-distance optical mapping contigs (N50: 1.35 Mb) from a different Paxton Lake male fish to refine the assembly (PEICHEL *et al.* 2017). Consistent with some contigs being erroneously scaffolded, the average percent coverage of the Paxton lake assembly by the aligned optical maps across autosomes was only 87.0% (Fig. 3.1, Fig. S3.3). We improved the assembly by removing individual contigs within chromosome scaffolds that were not supported well by the optical alignments (see methods). After removing contigs, the average percent coverage by aligned optical maps across chromosomes improved (95.2%). In addition, the total length of autosomes of the Paxton Lake *de novo* assembly more closely matched the v. 4 reference assembly (v. 4 reference assembly: 416.97 Mb; Paxton Lake assembly: 427.45 Mb; Table 1. With the addition of the previously assembled Paxton Lake X chromosome sequence (chr. XIX; (PEICHEL *et al.* 2020), the total genome length was 448.50 Mb (Table 1).

Genome wide, the Paxton Lake assembly was highly collinear with the v. 4 reference assembly (Fig. S3.4). However, unlike the v. 4 reference assembly, the Paxton Lake assembly was more contiguous. The Paxton Lake assembly had longer contigs

(N50: 1.25Mb) and only 1484 gaps across the autosomes, whereas the v. 4 reference assembly had a total of 12,960 autosomal gaps between shorter contigs (N50: 91.7kb). Across the genome, we detected 16 small inversions between the Paxton Lake assembly and the v. 4 reference assembly (Fig. S3.4; Table S3.2). We compared these breakpoints with the aligned optical map contigs to identify whether these were true inversions within the Paxton Lake population. An inversion would be supported if it was embedded within an optical mapping contig that was completely collinear with the assembly. All inversion breakpoints either fell at the edge of an optical contig or were not located within an optical contig, suggesting these may reflect assembly errors in the Paxton Lake assembly. Additional refinement will be necessary to determine if these small inversions reflect assembly errors or true structural variants within the Paxton Lake population.

A majority of gaps were closed across the threespine stickleback reference assembly

Since the *de novo* Paxton Lake assembly was more contiguous than the v. 4 reference assembly, we used the Paxton Lake Canu contigs in conjunction with the 1263 v. 4 chr. Un contigs that had been narrowed to chromosomes to attempt to close the 13,538 gaps in the v. 4 reference assembly. Using LR_Gapcloser we closed 10,394 of the gaps (76.8%), leaving only 3,144 gaps in the v. 5 assembly (XU *et al.* 2019) (File S1, File S2,Fig. S3.5). In addition to the fully closed gaps, 146 gaps were partially closed. A total of 9,928,283 bases were added to gaps in the assembly. This resulted in an overall greater contiguity of the genome, with a 5.57-fold greater N50 contig length within scaffolds compared to the previous reference assembly (v. 5 N50: 510.8 kb; v. 4 N50: 91.7 kb) (Table 2).

Genome contiguity and annotation completeness is often assessed by BUSCO (benchmarking universal single copy orthologs) statistics (WATERHOUSE *et al.* 2018). We determined if the additional sequence in the v. 5 reference assembly contained coding sequence that improved overall BUSCO metrics. Of the 3640 genes within the database, we found a total of 3521 BUSCO genes in the assembly (96.7%). This represented an increase of 99 genes compared to the previous assembly. In addition, the total number of fragmented BUSCO genes decreased to 14, compared to 108 in the v. 4 reference assembly (Table S3).

Of the 3378 chr. Un contigs from the v. 4 reference assembly, we determined how many were represented in the closed gaps of the new v. 5 reference assembly. Of the 3378 contigs, 457 contigs were placed within gaps (13.5%). The previous assembly used a Hi-C-based proximity-guided assembly method that was able to narrow some of the chr. Un contigs (1263) to chromosomes, but was not able to place these contigs into specific gaps (PEICHEL *et al.* 2017). We used this information to verify whether our contig placement was corroborated by the Hi-C sequencing. Of the 1263 previously narrowed chr. Un contigs, we placed 90 of into gaps. A majority of these contigs (80.0%) fell within the same chromosome they were assigned to previously by the Hi-C proximity-guided method. This high concordance further confirms the reliability of our methodology and closure of the gaps.

Across all closed gaps, we added 9.93 Mb of sequence to the genome. 1.13 Mb of this newly added sequence was from chr. Un contigs previously sequenced, but not placed in chromosomes. The remaining 8.80 Mb represented new regions from the long-read sequencing. Many of the gaps in the genome likely represent highly repetitive

regions that are challenging to assemble. We compared the repetitive sequence content between the 9.93 Mb of newly added sequence and the remainder of the genome. Indeed, we found newly closed gaps are enriched for repeat sequences (simple and interspersed repeats; 10,000 permutations; $P < 0.001$; Fig. S3.6). Overall, 17.4% of newly added bases contained repetitive DNA compared to 13.5% in the remainder of the genome. Across all newly added gap sequence, we found an overlap with a total of 1602 protein coding genes in v. 5. 1280 of such the genes that were fragmented in v. 4 are now contiguous in v. 5 (File S3). The newly placed regions overall exhibit a slightly lower density of coding sequence compared to the remainder of the genome (Fig S3.7; 10,000 permutations; $P < 0.083$). Only 7.9% of the closed gap bases were contained within coding regions. Across the remainder of the genome, 28.3% of bases in the v. 5 reference assembly were contained within coding regions. Combined, our results suggest the highly repetitive nature of the sequence contained within these gaps may have prevented assembly of these regions.

Although we closed a majority of gaps in the assembly, we were still unable to determine where 2921 of the chr. Un contigs belonged in the assembly (total length: 19.59 Mb). One possibility why we were unable to place these contigs is that they are even more difficult to assemble due to higher repetitive sequence content. Consistent with this, the unplaced contigs were highly enriched for Gypsy retrotransposons compared to the placed chr. Un contigs ($P < 0.001$; Fig. S3.6). 9.7% of the bases in unplaced contigs overlapped with Gypsy elements compared to 1.3% of the bases across the remainder of the genome. It is also possible that these contigs represent segments of the genome outside of gaps that are mis-assembled. Our method only focused on closing

gaps between contigs. Additional work will be necessary to determine whether these contigs integrate elsewhere in the genome. Assembly of these contigs may be facilitated by using additional de novo genome assemblies from other populations of threespine stickleback fish (BERNER *et al.* 2019).

Gap closing was validated by long-distance linked reads and collinear alignments with the Paxton Lake assembly

We aligned all gap sequences that were closed in the v. 5 assembly by LR_Gapcloser back to the *de novo* assembled Paxton Lake assembly to see if they were independently placed in the same chromosomal position by the two approaches. Of the 10,394 gaps closed in the v. 5 assembly, we were able to align 8552 (82.3%) back to the Paxton Lake assembly (Fig 3.2). The missing 1842 gap contigs that were placed in the v. 5 reference assembly by LR_Gapcloser were not assembled In Paxton Lake using the *de novo* assembly pipeline. Of the 8552 aligned gaps only 78 (0.01%) aligned to different chromosomes in the two assemblies. The remaining contigs exhibited highly collinear placements in the two assemblies (Fig. 3.2), supporting accurate gap closing in the v. 5 reference genome.

We used long-distance linked-reads to also validate placement of the new sequence within gaps. Linked-read molecules that support closure of a gap would exhibit aligned short-reads throughout the closed gap, whereas linked-read molecules that do not support closure of a gap would have aligned short-reads outside of the gap, but a lack of alignment within the gap (Fig. 3.3). Similar to the alignment between the Paxton Lake and v. 5 reference assembly, the gap closures were highly supported by the linked-read alignments. We only observed 36 gaps (0.3%) that were not supported by linked-reads

(i.e., a lack of short-read alignments over the newly added sequence). The remainder of the 10,394 gaps in this analysis that were closed (99.7%) were supported by the long-distance linked-read dataset (Fig. 3.3). We did not remove the small percentage of gaps that were not supported by the linked-read molecules or with alignment to the Paxton Lake assembly. It is possible this small number of closed gaps reflected true structural variation between the different populations. We therefore included them in the final assembly.

Telomere repeats and centromere repeats were identified within PacBio long reads

The telomeres of threespine stickleback fish contain a tandemly repeated G-rich hexanucleotide repeat that is conserved across metazoans (MOYZIS *et al.* 1988; MEYNE *et al.* 1989; TRAUT *et al.* 2007; OCALEWICZ 2013). Although DNA probes targeting these repeats clearly hybridize at the ends of all chromosomes in threespine stickleback fish, the underlying sequence of these regions is missing from the reference assembly. We therefore searched for the ancestral metazoan telomeric motif 'TTAGGG' or 'CCCTAA' in the raw PacBio reads to identify putative telomere caps (OCALEWICZ *et al.* 2011). We identified 3525 PacBio reads that contained telomere motifs. Seven of these reads contained enough unique sequence to align to the end of individual chromosomes (chromosomes IV, VII, VIII, X, XIV, XV, XVII). These reads showed an abundance of the ancestral metazoan telomeric motif at one end with little to no higher order structure (Fig. 3.4; Fig. S3.8). The telomeric motif was repeated 114-492 times throughout the sequence on different chromosomes.

We also searched for centromere repeats within the PacBio assembled contigs. We identified the core 186 bp CENP-A repeat (CECH and PEICHEL 2015) within 91

contigs (the length of repetitive DNA among contigs ranges from 12.61 – 125.17 kb). 48 of these contigs contained enough unique sequence to align to all 21 chromosomes (Fig. 3.5; File S4; File S5). 11 chromosomes had centromere contigs that map to both sides of the gap, 9 chromosomes had a centromere contig only on one side of the centromere, and one contig contained a full centromere sequence, spanning the entire gap (chromosome IX). Interestingly, on many of the chromosomes, the repeat length was long enough to discern clear higher order repeat structure (Fig. 3.5; Fig. S3.9). Our results are similar to the variability in higher order repeat among the autosomes and X chromosome of humans (WILLARD 1985; WILLARD *et al.* 1986; ALEXANDROV *et al.* 1993; SHEPELEV *et al.* 2015; HARTLEY and O'NEILL 2019). We detected multiple contigs mapping to either side of the centromere gap for all chromosomes (File S4), indicating the male fish used for sequencing is likely heterozygous for centromeric arrays. This is consistent with high polymorphism of centromere arrays observed within other species (WILLARD *et al.* 1986; DEVILEE *et al.* 1988; WEVRICK and WILLARD 1989; MAHTANI and WILLARD 1990; GREIG *et al.* 1991).

Y chromosomes in mammals have also been documented to have highly variable centromeric repeats that are divergent from their counterparts across the remainder of the genome (WOLFE *et al.* 1985; PERTILE *et al.* 2009; MIGA *et al.* 2014). Assembly of segments of the threespine stickleback Y chromosome centromere (PEICHEL *et al.* 2020) revealed an alpha satellite monomer repeat that was divergent from the consensus monomeric repeat identified from the remainder of the genome (CECH and PEICHEL 2015). With the assembly of larger tracks of centromeric sequence from the autosomes and the X chromosome, we now show the Y chromosome centromere is also divergent

from the remainder of the genome at the level of higher order repeats (PEICHEL *et al.* 2020), matching other rapidly evolving Y chromosomes. Although our assembly has uncovered a large fraction of the centromeric sequence for each chromosome, we were unable to assemble complete centromere sequences outside of the 46.5 kb centromere of chromosome IX. It therefore remains unknown how centromere length varies throughout the threespine stickleback genome. Complete characterization of the centromeric repetitive arrays will be aided by future sequencing of ultra-long reads (JAIN *et al.* 2018b; MIGA *et al.* 2020).

Conclusions

By using long-read sequencing we were able to substantially improve the overall contiguity of the threespine stickleback reference genome assembly, increasing the N50 length of contigs over five-fold. Our assembly also highlights the power of using long-read sequencing technologies to assemble previously inaccessible regions of the genome, like centromeres and telomeres. We have released this assembly through a new threespine stickleback fish community genome browser (https://stickleback.genetics.uga.edu). The v. 5 reference assembly and the Paxton Lake *de novo* assembly will be useful additions to the rapidly expanding functional genomics toolkit available in threespine stickleback fish.

Acknowledgements

This research was funded by the National Science Foundation IOS 1645170 (M.A.W.), the National Science Foundation MCB 1943283 (MAW), the Office of the Vice President of Research at the University of Georgia (M.A.W.), and the University of Georgia Research Foundation (D.E.S.). We thank the Georgia Genomics and

Bioinformatics Core at the University of Georgia for help with the long-distance linked-read sequencing. We also thank Brigitte Hofmeister and the Franklin College Office of Information Technology at the University of Georgia for help building the threespine stickleback genome browser.

Figures

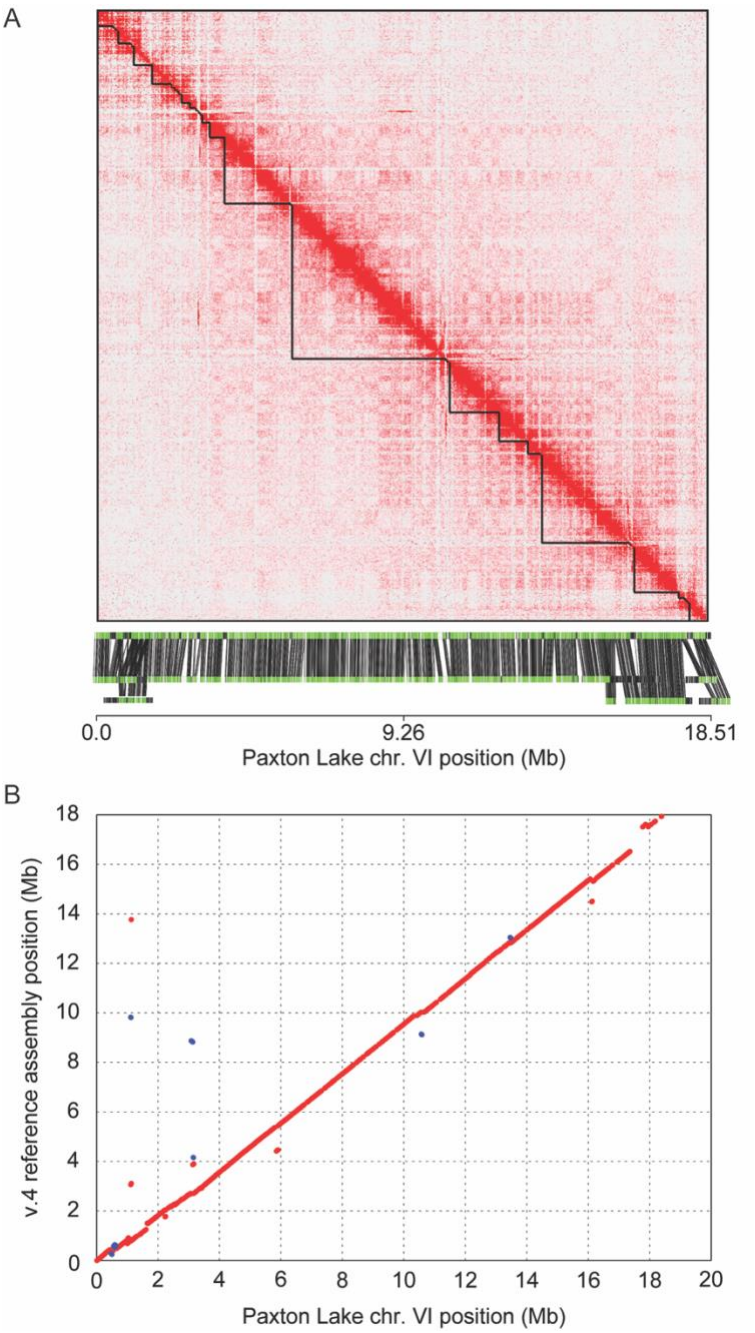


Figure 3.1. The Paxton Lake de novo assembly is collinear with the v. 4 reference assembly. (A) Hi-C chromosome conformation capture sequencing generated a single chromosome VI scaffold. The contact map revealed an enrichment of interactions between contigs that are in close proximity, visualized along the diagonal. Contig boundaries within the scaffold are denoted by black triangles along the diagonal. The corresponding Paxton Lake optical map contigs are concordant with the Hi-C scaffolding. The reference sequence is shown on the top and the optical contigs are shown on the bottom. 97.2% of the chromosome is covered by optical map contigs. (B) Nucleotide alignments between chromosome VI of the two assemblies reveal a syntenic ordering. Blue dots represent small regions of the chromosomes that are in an inverted region relative to the remainder of the alignment (red). The remaining Hi-C scaffold maps, dotplots and optical alignments are shown in supplemental Fig. S3.10, S3.4 and S3.3.

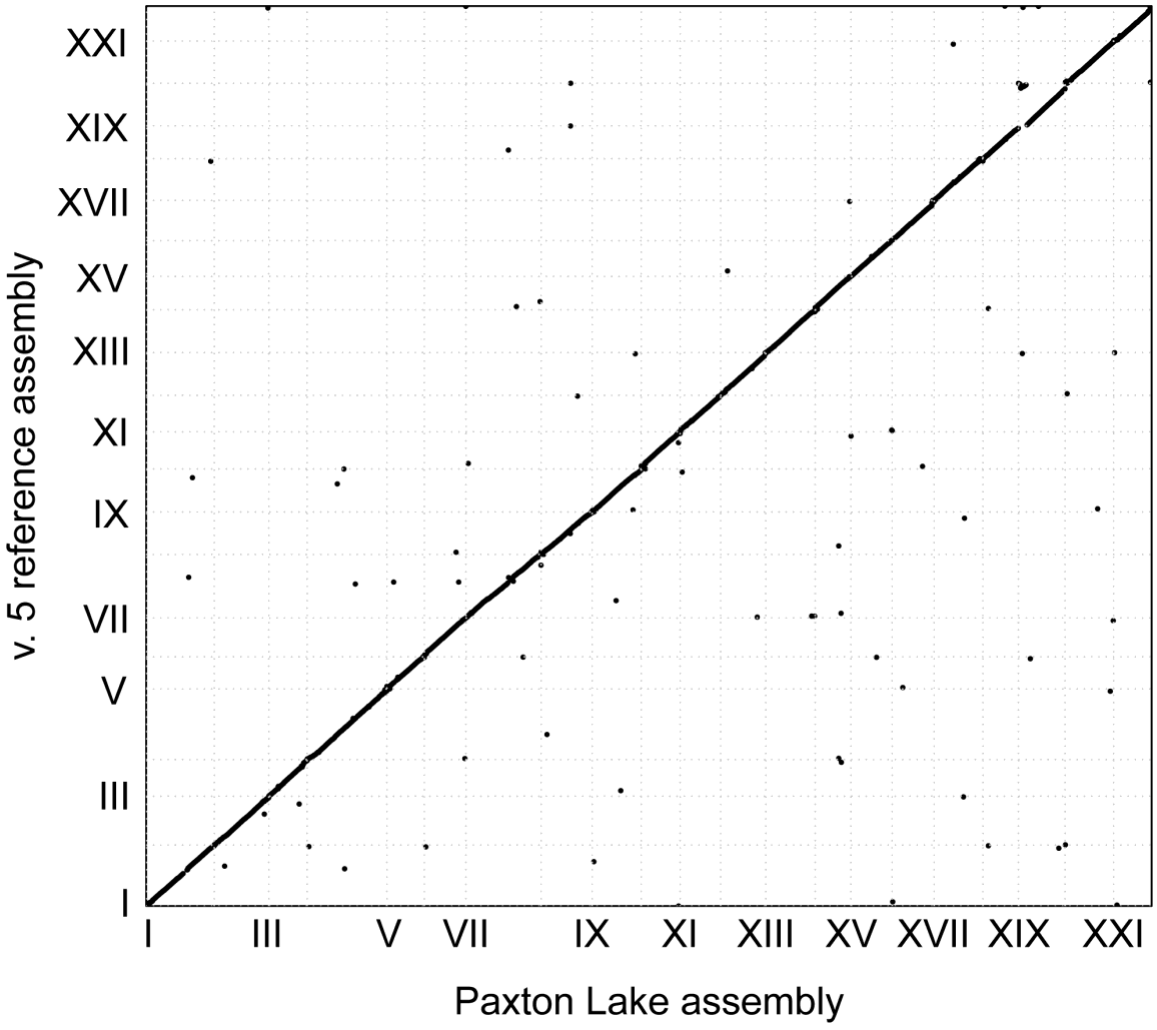


Figure 3.2. Gaps closed in the v. 5 reference assembly are collinear the Paxton Lake assembly. Gaps were closed in the v. 5 reference assembly using the unassembled Paxton Lake contigs. LR_Gapcloser placed sequences in the v. 5 reference assembly in a collinear order with the Paxton Lake de novo assembly. Only 0.01% of the gaps aligned to different chromosomes in the two assemblies, either reflecting true structural variation between the reference assembly and Paxton Lake or assembly error in either of the populations.

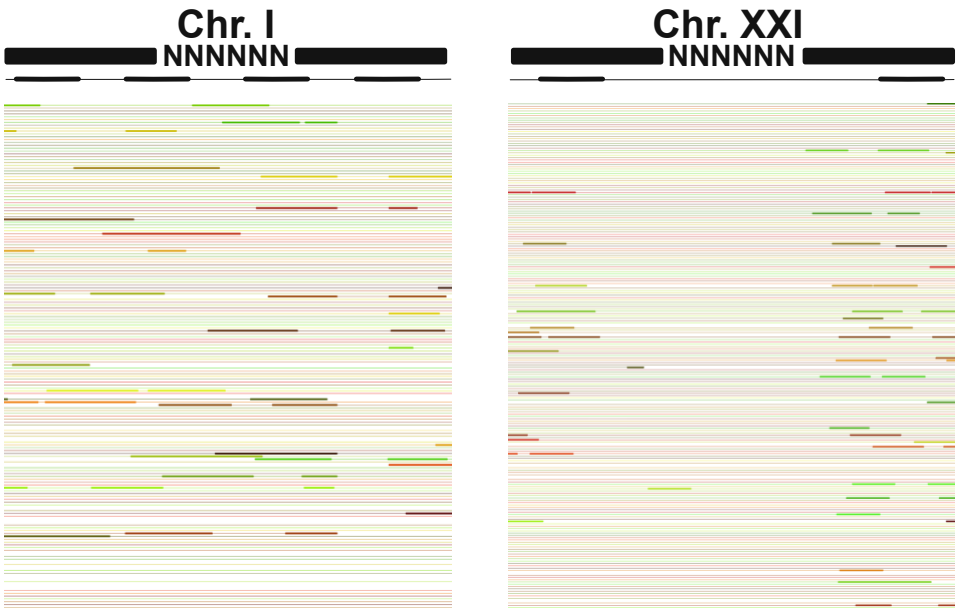


Figure 3.3. 10X Genomics linked-reads validate most of the closed gaps. 99.7% of closed gaps exhibit linked-read alignments throughout the gap region, indicating a correctly closed gap (e.g. Chr. I:192,954-193,294 bp with flanking region). 0.03% of gaps were not validated by the linked-read sequencing. In these regions, alignments of the linked-reads only occur outside of the gap (e.g. Chr. XXI: 9,436,991-9,437,593 bp with flanking region). A representative schematic outlining how the linked-reads should align is shown in black. The actual aligned linked-reads are shown by bolded color lines. Thin lines indicate gaps between the linked-reads. Average read depth of linked-reads across the genome was 26.1X. A subset of reads aligning is shown here for simplicity.

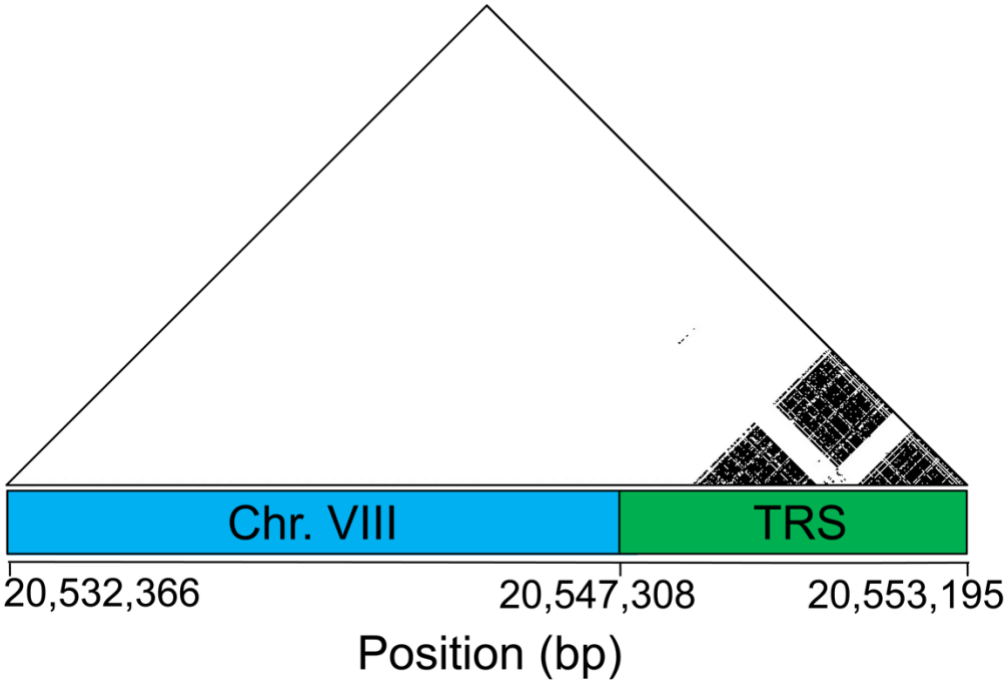


Figure 3.4. Telomeres exhibit a high density of the conserved metazoan telomere motif. Dots represent 100% sequence identity between matching windows of 15 bp. The blue box represents the end of chr. VIII where the long read aligns uniquely to positions 20,532,366-20,547,308. The green box denotes a ~10 kb segment rich with telomeric repeat sequence (TRS). The remaining telomeres are shown in Fig S3.8.

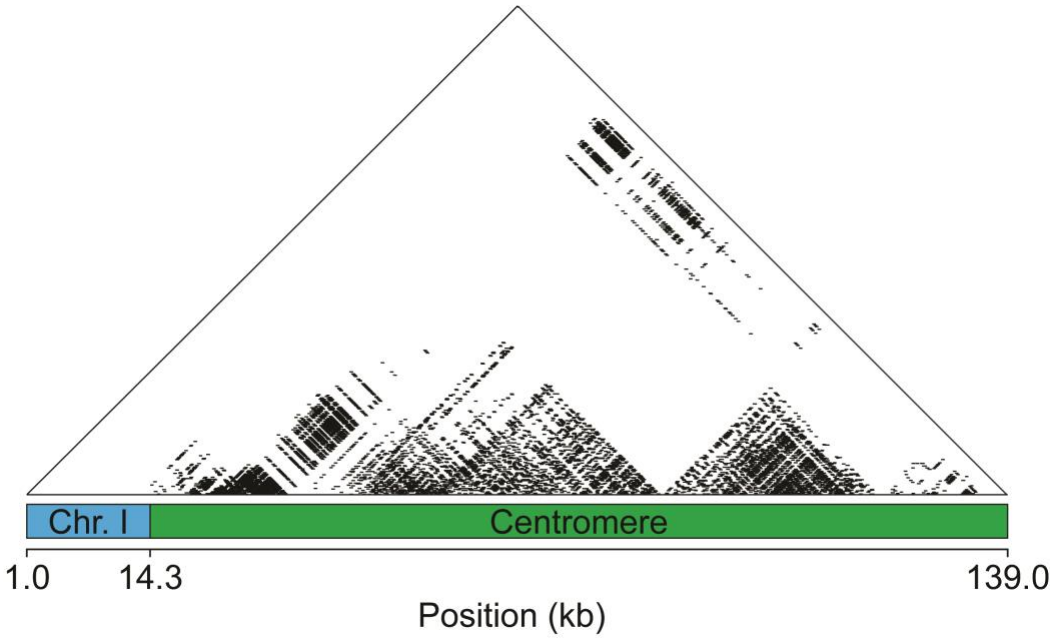


Figure 3.5. Centromeres display higher order repeat structure. On chr. I, the centromere contig contains 673 186 bp monomer repeats. Sequence identity between repeats is depicted by black dots matching windows of 300 bp with 100% sequence identity. The blue region denotes the side of chr. I (20,330,007-20,344,665 bp) with unique sequence that aligns to the periphery of the centromere. The green region is the newly aligned centromere contig. The other side of the centromere did not align to the other arm of chromosome I. The remaining centromeres are shown in Fig. S3.9.

TABLE 3.1. Chromosomal lengths of the Paxton Lake Assembly and the v. 4 reference assembly.

Chr.	v. 4 reference assembly	Paxton Lake assembly
I	29714595	30291332
II	23752435	24322974
III	17815537	17075251
IV	34244925	35558100
V	15579443	16703751
VI	18862055	18511259
VII	30864241	33510017
VIII	20606801	22749572
IX	20880404	21858600
X	18035923	17341068
XI	17683359	18065382
XII	20811783	20111693
XIII	20800062	22032464
XIV	16179395	15892220
XV	17375354	18397630
XVI	19558478	18726499
XVII	20254007	21751973
XVIII	15989023	15795861
XX	20484364	21645813
XXI	17480265	17105228
Autosome total	**416972449**	**427446687**

XIX	20618466	20783580
Genome total	**437590915**	**448230267**

TABLE3. 2. Improved contiguity of the threespine stickleback genome.

	v. 5 assembly	v.4 assembly
Assembly size (without Ns and chrUn)	448.67 B	441.86MB
Number of Gaps	3144	13538
L50	233	1291
L90	983	5378
N50	510.82kb	91.68kb
N90	94.65kb	18.17kb

References

Alexandrov, I.A., L.I. Medvedev, T.D. Mashkova, L.L. Kisselev, L.Y. Romanova *et al.*, 1993 Definition of a new alpha satellite suprachromosomal family characterized by monomeric organization. *Nucleic Acids Res* 21 (9):2209-2215.

Arnegard, M.E., M.D. McGee, B. Matthews, K.B. Marchinko, G.L. Conte *et al.*, 2014 Genetics of ecological divergence during speciation. *Nature* 511 (7509):307-311.

Bell, M., and S.A. Foster, 1994 The evolutionary biology of the threespine sticklebacks. Oxford University Press.

Benjamini, Y., and T.P. Speed, 2012 Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res* 40 (10):e72.

Berner, D., M. Roesti, S. Bilobram, S.K. Chan, H. Kirk *et al.*, 2019 Sequencing, Assembly, and Annotation of Four Threespine Stickleback Genomes Based on Microfluidic Partitioned DNA Libraries. *Genes (Basel)* 10 (6).

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

Cantarel, B.L., I. Korf, S.M. Robb, G. Parra, E. Ross *et al.*, 2008 MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 18 (1):188-196.

Cech, J.N., and C.L. Peichel, 2015 Identification of the centromeric repeat in the threespine stickleback fish (Gasterosteus aculeatus). *Chromosome Res* 23 (4):767-779.

Chang, C.H., and A.M. Larracuente, 2019 Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the. *Genetics* 211 (1):333-348.

Conte, M.A., R. Joshi, E.C. Moore, S.P. Nandamuri, W.J. Gammerdinger *et al.*, 2019 Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *Gigascience* 8 (4).

Devilee, P., T. Kievits, J.S. Waye, P.L. Pearson, and H.F. Willard, 1988 Chromosome-specific alpha satellite DNA: isolation and mapping of a polymorphic alphoid repeat from human chromosome 10. *Genomics* 3 (1):1-7.

Dudchenko, O., S.S. Batra, A.D. Omer, S.K. Nyquist, M. Hoeger *et al.*, 2017 De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356 (6333):92-95.

Durand, N.C., M.S. Shamim, I. Machol, S.S. Rao, M.H. Huntley *et al.*, 2016 Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* 3 (1):95-98.

Glazer, A.M., E.E. Killingbeck, T. Mitros, D.S. Rokhsar, and C.T. Miller, 2015 Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in Sticklebacks with Genotyping-by-Sequencing. *G3* 5 (7):1463-1472.

Gnerre, S., I. Maccallum, D. Przybylski, F.J. Ribeiro, J.N. Burton *et al.*, 2011 High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* 108 (4):1513-1518.

Greig, G.M., S. Parikh, J. George, V.E. Powers, and H.F. Willard, 1991 Molecular cytogenetics of alpha satellite DNA from chromosome 12: fluorescence in situ hybridization and description of DNA and array length polymorphisms. *Cytogenet Cell Genet* 56 (3-4):144-148.

Hartley, G., and R.J. O'Neill, 2019 Centromere Repeats: Hidden Gems of the Genome. *Genes (Basel)* 10 (3).

Hatfield, T., and D. Schluter, 1999 Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution* 53 (3):866-873.

He, S., L. Li, L.Y. Lv, W.J. Cai, Y.Q. Dou *et al.*, 2020 Mandarin fish (Sinipercidae) genomes provide insights into innate predatory feeding. *Commun Biol* 3 (1):361.

Heras, J., M. Chakraborty, J.J. Emerson, and D.P. German, 2020 Genomic and biochemical evidence of dietary adaptation in a marine herbivorous fish. *Proc Biol Sci* 287 (1921):20192327.

Holt, C., and M. Yandell, 2011 MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.

Jain, M., S. Koren, K.H. Miga, J. Quick, A.C. Rand *et al.*, 2018a Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 36 (4):338-345.

Jain, M., H.E. Olsen, D.J. Turner, D. Stoddart, K.V. Bulazel *et al.*, 2018b Linear assembly of a human centromere on the Y chromosome. *Nat Biotechnol* 36 (4):321-323.

Jones, F.C., M.G. Grabherr, Y.F. Chan, P. Russell, E. Mauceli *et al.*, 2012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484 (7392):55-61.

Kent, W.J., 2002 BLAT--the BLAST-like alignment tool. *Genome Res* 12 (4):656-664.

Koren, S., B.P. Walenz, K. Berlin, J.R. Miller, N.H. Bergman *et al.*, 2017 Canu: scalable and accurate long-read assembly via adaptive. *Genome Res* 27 (5):722-736.

Korf, I., 2004 Gene finding in novel genomes. *BMC Bioinformatics* 5:59.

Kurtz, S., A. Phillippy, A.L. Delcher, M. Smoot, M. Shumway *et al.*, 2004 Versatile and open software for comparing large genomes. *Genome Biol* 5 (2):R12.

Li, H., 2018 Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18):3094-3100.

Liu, J., A.S. Seetharam, K. Chougule, S. Ou, K.W. Swentowsky *et al.*, 2020 Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol* 21 (1):121.

Mahtani, M.M., and H.F. Willard, 1990 Pulsed-field gel analysis of alpha-satellite DNA at the human X chromosome centromere: high-frequency polymorphisms and array size estimate. *Genomics* 7 (4):607-613.

McPhail, J.D., 1992 Ecology and evolution of sympatric sticklebacks (Gasterosteus): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia. *Canadian Journal of Zoology* 70 (2):361-369.

Meyne, J., R.L. Ratliff, and R.K. Moyzis, 1989 Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc Natl Acad Sci U S A* 86 (18):7049-7053.

Miga, K.H., S. Koren, A. Rhie, M.R. Vollger, A. Gershman *et al.*, 2020 Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 585 (7823):79-84.

Miga, K.H., Y. Newton, M. Jain, N. Altemose, H.F. Willard *et al.*, 2014 Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Res* 24 (4):697-707.

Moyzis, R.K., J.M. Buckingham, L.S. Cram, M. Dani, L.L. Deaven *et al.*, 1988 A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A* 85 (18):6622-6626.

Nagarajan, N., and M. Pop, 2013 Sequence assembly demystified. *Nat Rev Genet* 14 (3):157-167.

Ocalewicz, K., 2013 Telomeres in fishes. *Cytogenet Genome Res* 141 (2-3):114-125.

Ocalewicz, K., P. Woznicki, G. Furgala-Selezniow, and M. Jankun, 2011 Chromosomal location of Ag/CMA 3 -NORs, 5S rDNA and telomeric repeats in two stickleback species. *Italian Journal of Zoology* 78 (1):12-19.

Peichel, C.L., S.R. McCann, J.A. Ross, A.F.S. Naftaly, J.R. Urton *et al.*, 2020 Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol* 21 (1):177.

Peichel, C.L., S.T. Sullivan, I. Liachko, and M.A. White, 2017 Improvement of the Threespine Stickleback Genome Using a Hi-C-Based Proximity-Guided Assembly. *J Hered* 108 (6):693-700.

Pertile, M.D., A.N. Graham, K.H. Choo, and P. Kalitsis, 2009 Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability. *Genome Res* 19 (12):2202-2213.

Pracana, R., A. Priyam, I. Levantis, R.A. Nichols, and Y. Wurm, 2017 The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. *Mol Ecol* 26 (11):2864-2879.

Prost, S., M. Petersen, M. Grethlein, S.J. Hahn, N. Kuschik-Maczollek *et al.*, 2020 Improving the Chromosome-Level Genome Assembly of the Siamese Fighting Fish. *G3* 10 (7):2179-2183.

Quinlan, A.R., and I.M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6):841-842.

Roesti, M., D. Moser, and D. Berner, 2013 Recombination in the threespine stickleback genome--patterns and consequences. *Mol Ecol* 22 (11):3014-3027.

Ross, M.G., C. Russ, M. Costello, A. Hollinger, N.J. Lennon *et al.*, 2013 Characterizing and measuring bias in sequence data. *Genome Biol* 14 (5):R51.

Schneider, V.A., T. Graves-Lindsay, K. Howe, N. Bouk, H.C. Chen *et al.*, 2017 Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res* 27 (5):849-864.

Shepelev, V.A., L.I. Uralsky, A.A. Alexandrov, Y.B. Yurov, E.I. Rogaev *et al.*, 2015 Annotation of suprachromosomal families reveals uncommon types of alpha satellite organization in pericentromeric regions of hg38 human genome assembly. *Genom Data* 5:139-146.

Simão, F.A., R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, and E.M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19):3210-3212.

Stanke, M., A. Tzvetkova, and B. Morgenstern, 2006 AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol* 7 Suppl 1:S11.11-18.

Traut, W., M. Szczepanowski, M. Vítková, C. Opitz, F. Marec *et al.*, 2007 The telomere repeat motif of basal Metazoa. *Chromosome Res* 15 (3):371-382.

Vollger, M.R., G.A. Logsdon, P.A. Audano, A. Sulovari, D. Porubsky *et al.*, 2020 Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads. *Ann Hum Genet* 84 (2):125-140.

Waterhouse, R.M., M. Seppey, F.A. Simão, M. Manni, P. Ioannidis *et al.*, 2018 BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* 35 (3):543-548.

Wevrick, R., and H.F. Willard, 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 U S A* 86 (23):9394-9398.

Willard, H.F., 1985 Chromosome-specific organization of human alpha satellite DNA. *Am J Hum Genet* 37 (3):524-532.

Willard, H.F., J.S. Waye, M.H. Skolnick, C.E. Schwartz, V.E. Powers *et al.*, 1986 Detection of restriction fragment length polymorphisms at the centromeres of human chromosomes by using chromosome-specific alpha satellite DNA probes: implications for development of centromere-based genetic linkage maps. *Proc Natl Acad Sci U S A* 83 (15):5611-5615.

Wolfe, J., S.M. Darling, R.P. Erickson, I.W. Craig, V.J. Buckle *et al.*, 1985 Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. *J Mol Biol* 182 (4):477-485.

Wootton, R., 1976 The Biology of Sticklebacks. Academic Press.

Xu, G.C., T.J. Xu, R. Zhu, Y. Zhang, S.Q. Li *et al.*, 2019 LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* 8 (1).

Zhou, Y., S. Xiao, G. Lin, D. Chen, W. Cen *et al.*, 2019 Chromosome genome assembly and annotation of the yellowbelly pufferfish with PacBio and Hi-C sequencing data. *Sci Data* 6 (1):267.

CHAPTER 4

GENE CONVERSION BETWEEN SEX CHROMOSOMES IN THREESPINE STICKLEBACKS

Introduction

Sex chromosomes evolve from autosomes when one chromosome acquires a sex determining gene. The subsequent accumulation of sexually antagonistic genes on sex chromosomes favors recombination suppression between them (BACHTROG 2013; BACHTROG et al. 2014; ALISON E WRIGHT 2016). The recombinant haplotypes maybe selected against in the population due to reduced fitness. Sex chromosomes diverge from each other when recombination is suppressed in a stepwise fashion, often by means of inversions. Inversions offer a major structural impediment to homologous repair during meiosis and suppresses recombination in large blocks (CHARLESWORTH et al. 2005). Following recombination suppression, sex chromosomes typically share a small region of homology called a pseudoautosomal region where all crossing over occurs in males every meiosis (CHANDLEY et al. 1984; PIGOZZI 2016). While sex chromosomes have long been considered to be non-recombinant, this view has been challenged recently. In primates, extensive population resequencing studies have shown continue exchange can still occur between the sex chromosomes in the form of non-crossover repair (i.e. gene conversion). (ROSSER et al. 2009; TROMBETTA et al. 2010; TROMBETTA et al. 2014; TROMBETTA et al. 2016; TROMBETTA et al. 2017). Additional work is necessary in other species to understand whether this is a universal phenomenon.

The double stranded breaks formed during meiosis can be repaired by using homologous chromosome or sister chromatids as a template. While only one DSB per chromosome is resolved as a crossover, majority of DSBs are repaired as gene conversions. Crossing over involves reciprocal exchange of long tracts of DNA at DSBs while gene conversions are typically smaller non-reciprocal repair of a break using homologous chromosome or sister chromatid as a template (CHEN et al. 2007; PADHUKASAHASRAM and RANNALA 2013). The DSBs on sex chromosomes can be repaired using duplicated sequences within chromosomes, sister chromatids and even gametologs when the sequences are similar. The inter chromosomal gene conversion tracts can have single or multiple donor sequences. While the presence of multiple "donors" increases nucleotide and haplotype diversity in the hotspots by introducing new SNPs, the hotspots with single donors have a homogenizing effect and will be eventually undetectable (TROMBETTA et al. 2017). While this observation is at odds with the model that genetic exchange on sex chromosomes will be deleterious by introducing X-linked variants on Y and vice versa, for the single copy genes that are expressed in both sexes in multiple tissues and have housekeeping functions this may be a mechanism to purge out deleterious mutations on Y chromosomes (BELLOTT et al. 2014; CORTEZ et al. 2014). Inter-chromosomal gene conversion in ZW chromosomes of birds has also been implicated in homogenizing the genes suggesting that recombination suppression is not complete (WRIGHT et al. 2014).

Gene conversion has been most widely studied on the human sex chromosomes (TROMBETTA et al. 2010; TROMBETTA et al. 2014; TROMBETTA et al. 2016; TROMBETTA and CRUCIANI 2017; TROMBETTA et al. 2017). The sex chromosomes originated in the

common ancestor of all mammals at least 180 million years old and is composed of at least five strata on which recombination was suppressed in a stepwise fashion (SKALETSKY et al. 2003). The youngest stratum which is around 30 million years old shares a 95% homology between the two sex chromosomes (SKALETSKY et al. 2003; ROSS et al. 2005). Human sex chromosomes also share two pseudo-autosomal regions at the ends where recombination can occur during meiosis (CHANDLEY et al. 1984). In the absence of chromosome wide recombination (apart from PAR) between sex chromosomes, the X and Y chromosomes have diverged over time and only share a handful of genes (SKALETSKY et al. 2003; BELLOTT et al. 2014). The genes retained on Y chromosomes despite the extensive degeneration have been shown to be dosage sensitive and crucial for fertility (BELLOTT et al. 2014). The human Y chromosome also has ampliconic regions including palindromes (P1-P8) and harbor the duplicated genes. The palindromic arms share >99.9% homology, potentially due to extensive gene conversions between them that effectively homogenize the genes in the two arms (SKALETSKY et al. 2003; TROMBETTA and CRUCIANI 2017). Purifying selection and intra gene conversion between multi-copy genes has been proposed as main mechanism for maintaining these genes on the Y chromosomes in the face of extensive degeneration (SKALETSKY et al. 2003; BELLOTT et al. 2014). However, thirteen hotspots of inter-chromosomal gene conversion have been identified on human sex chromosomes in the youngest strata (stratum4 and stratum5) where genetic exchange can take place in small tracts. Nine of these tracts represent gene conversion hotspots active in human lineage and some shared with chimpanzees as well (reviewed in (TROMBETTA et al. 2017). Since homologous recombination hotspots are minimally shared between humans and chimpanzees, shared

gene conversion hotspots on sex chromosomes shows that non-allelic recombination hotspots have a longer lifespan. The hotspots are ununiformly distributed on the Y chromosomes with nine out of thirteen all observed conversion tracts limited to just 3 genes in stratum 5 and 4 (TROMBETTA et al. 2016).

Given the evidence of gene conversion in old sex chromosomes of humans and birds, we sought to address the meiotic rate of gene conversion in young sex chromosomes and how this genetic exchange may affect the evolution of sex chromosomes that share extensive homology. Since young sex chromosomes are not extensively degenerated, the microhomology in certain regions can allow for such aberrant gene conversion events by using the gametolog as a template.

We used the threespine stickleback fish (Gasterosteus aculeatus) as it has a young XY sex chromosome system which is almost 20 million years old (ROSS et al. 2009; PEICHEL et al. 2020). The sex chromosomes share a PAR (2.5Mb) and have three strata corresponding to three major inversions which disrupt the synteny. Despite Y chromosome degeneration, the sex chromosomes still share extensive homology and a total of 605 genes outside PAR (PEICHEL et al. 2020). The Y chromosome in threespine sticklebacks have 42.5% of the X-linked ancestral genes outside of PAR, while Y chromosome in humans only retains 3% of ancestral X-linked genes (BELLOTT et al. 2014; PEICHEL et al. 2020). The sex chromosomes in our system also have 3 inversions that distinguish it from X chromosome which potentially hinder genetic exchange. In order to study the rates of gene conversion on sex chromosomes, we sequenced the DNA from blood and mature sperm using long reads technology and analyzed the reads to extract the chimeric reads that have evidence of a crossover or noncrossover event.

Methods

High molecular weight DNA extraction from blood

High molecular weight DNA was extracted from sperm and blood of a single adult male using alkaline lysis. Blood was collected from euthanized fish into 0.85X SSC buffer. The cells were collected by centrifuging for two minutes at 2000 xg. Pelleted cells were resuspended in five ml of 0.85X SSC and 27 µl of 20 µg/ml Proteinase K solution. To lyse the cells, five ml of 2X SDS buffer (80mM EDTA, 100mM Tris pH 8.0, and 1% SDS) was added to the suspension and the solution was incubated at 55°C for two minutes. After incubation, 10 ml of buffered phenol/chloroform/isoamyl-alcohol was added to the suspension. The suspension was incubated at room temperature under slow rotation for 30 minutes. The suspension was centrifuged for one minute at 2000 xg at 4°C to separate phases. The aqueous phase was extracted, 10 ml of chloroform was added, and the suspension was rotated for one hour. The chloroform extraction step was repeated twice. After all extractions, the aqueous phase was separated and mixed with ice cold 100% ethanol and one ml of 3M sodium-acetate (pH 5.5). The tube was gently inverted until a spool of DNA was observed. The DNA spool was transferred to a two ml tube filled with 70% cold ethanol and pelleted at 500 xg for two minutes. The DNA was allowed to dry at room temperature and resuspended in nuclease free water. Wide bore pipette tips were used for the whole procedure to minimize shearing. The integrity and size of the high molecular weight DNA was verified using a high sensitivity large fragment analysis on a fragment analyzer (Advanced Analytical Technologies, CA, USA).

High molecular weight DNA extraction from sperm

The procedure was adapted from Jain et al 2018 (JAIN et al. 2018). Testes were dissected from the same adult male and gently macerated in PBS buffer with the Dounce homogenizer to release sperm. Only the suspension was used for DNA extraction and the tissue bits were discarded. The suspension was spun down at 4500 xg for 10 minutes. The pellet was resuspended by pipette mixing in 100 µl sterile PBS
And then 10 ml of TLB (100 mM NaCl, 10 mM Tris-Cl, pH 8.0, 25 mM EDTA, pH 8.0, 0.5% (w/v) SDS, 20 µg/ml RNase A) was added to suspension and vortexed at full speed for 5 seconds. The suspension was incubated at 37°C for 1 hour and then Proteinase K (20 mg/mL) was added to a final concentration of 100 ug/ml. It was then mixed by slowly rotating end-over-end 10 times and incubating at 50°C for 1 hour on a slow shaker plate. The viscous lysate was then split into two 15 mil tubes and 5 ml of TE-saturated phenol was added to each tube. The mix was then again placed on a slow shaker for 10 minutes. The mixture was then centrifuged at 4500 rpm for 10 minutes and aqueous phase was extracted using a wide-bore pipette tip. A second extraction was done using 5 ml of saturated phenol-chloroform-isoamyl alcohol. The aqueous phases were combined from two tubes into a new 50 ml Falcon tube and precipitated using the same procedure as described in “High molecular weight DNA extraction from blood” section.

Library preparation and sequencing

The DNA fragment size range and was determined to be >40kb in length using a fragment analyzer. The DNA was sheared to appropriate size range and size selected using Blue Pippin (final average molecule size Sperm: 18822 bp; Blood:10335bp). The libraries were made using SMRTbell Express Template Prep Kit 2.0 for HiFi CCS

sequencing and sequenced on the Sequel II platform at Georgia Genomics and Bioinformatics Center at University of Georgia.

Short read illumina sequencing of DNA and Blood

DNA from the same samples was also sequenced using an Illumina NovaSeq 6000 (paired-end 150 bp). Library preparations and sequencing was conducted by Genewiz. 151 million reads were obtained from sperm and 137 million reads from blood. The reads were aligned using bowtie2 (v 2.4.1) and 91.1% of reads from sperm aligned while 78.8% reads mapped from blood. Only the alignments above mapping quality of 21 were retained which resulted in retention of 115 million reads from sperm and 82 million from blood. The duplicate reads were removed using MarkDuplicates utility of picard (v. 2.21.6). The subsequent reads were then realigned around indels and genotypes were called using GATK (3.8.1). Final read depth of sperm was 41.6X and blood was 31.7X.

Genome Alignment of CCS reads

The CCS reads were generated using CCS software (v. 3.4.1) with –noPolish option and minimum subread pass of 8. The reads were aligned to Threespine stickleback v. 5 genome assembly (without chr. Y) using pbmm2 (v. 1.0.0_conda) using --CCS preset. The total number of reads aligned was 614,489 (25.1X) in sperm and 970,557 (21.8X) in blood. Only the primary alignments and alignments with more than 2kb of length were considered for further analysis. Any secondary alignments were removed. If a read aligns to multiple places in a genome unambiguously then one of the alignments is arbitrarily chosen as a primary alignment by the alignment software while rest of the alignments of the same reads are referred to as secondary alignments. Read accuracy was calculated using the formula number of matching bases /(matches +mismatches

+deletions + insertions). 90% of the reads were at least 98.8% accurate in both sperm and blood. This is the standard measure of base accuracy of PacBio Hifi reads (WENGER et al. 2019).

Phasing and finding gene conversion and crossover tracts

The two parental haplotypes in the F1 fish were phased using whatshap (v 1.1) (MARTIN et al. 2016) with the blood DNA. Briefly, the blood long reads aligned to the genome and variant call format (VCF) file of the genotyped calls using blood illumina reads were used with the "phase" utility of whatshap program. This generated phased VCF files of illumina reads. This information was then used to phase long reads from the sperm sample and each read was classified as Haplotype1 (H1) or Haplotype2 (H2) depending on their phase. The reads that could not be assigned to any phase were tagged as "none". The consensus bases were determined for each haplotype at each position by counting the number of reads supporting each allele at that position and assigning the allele with highest number of reads for each haplotype using custom scripts. As expected, at most positions the bases were same in each haplotype. The positions at which the bases differed were termed as SNPs and reflect the differences in the two parental genomes. Then each read was examined by comparing bases at each position with that of the consensus sequence constructed in the previous step to determine what haplotype bases at each position belonged to. Most reads have all bases belonging to just one haplotype (614,489 reads) and 2,410 reads were a chimera(Fig. 1). The chimeric reads obtained from our analysis were further divided into the two categories of gene conversion tracts and crossover tracts. The crossover tracts were defined as reads that exhibited a breakpoint somewhere within the read that split the read into two parental haplotypes .

The gene conversion tracts were defined as reads that contained a segment of DNA from one of the parental haplotypes flanked on either side with the other parental haplotype.

Results

We examined the chromosomal distribution of crossover and gene conversion tracts in autosomes as well as the sex chromosomes. In total, 16 gene conversion tracts were found on sex chromosomes (Y=17.8 Mb, X=20.6 Mb) and 12/17 were in PAR, 2/17 in stratum 3, 2/17 in stratum 2 and 0/17 in stratum 3 (Fig 2). The minimum tract length and maximum tract lengths were observed to be 70.0 bp and 807.3 bp respectively. The 5 crossover events observed were all in PAR and each tract had at least 5 SNPs from each haplotype. For the autosome XI (17.6 Mb), the 62 gene conversion events 12 and crossover events were distributed throughout the chromosome length (Fig 3). The chr. XI was chosen as this the diploid size of this chromosome is very close to the size of sex chromosome pair (~38Mb of sex chromosomes).

Crossovers are only expected with the pseudoautosomal region in male meiosis. One reason we may not observe events outside of the PAR is mapping bias against aligning reads to the X chromosome. We tested whether the skewed distribution of gene conversion and crossover reads on PAR was caused by this by examining the read depth in the PAR, and the three strata and found that the read depths in the different regions of X chromosome are similar and comparable to autosomes (read depth in PAR: 16.0X, stratum2:15.4X, stratum3: 14.6X, stratum1: 11.0X, chrXI:16.5). The reduced read depth on stratum1 is perhaps due to non-alignment of Y-linked reads from the regions on Y chromosomes that may have undergone extensive degeneration and even deletion which

will reduce the overall diploid read depth. However, if the regions are extremely degenerated, we will not expect inter-chromosomal gene conversions in such regions.

Since gene conversion is a non-reciprocal form of homologous recombination, we were interested in the direction of conversion in the tracts. While the reads aligning in PAR cannot be differentiated according to their origin of chromosome (X or Y) in this sample, we can differentiate the non-PAR reads and determine if they belonged to X or Y chromosome based on their percent identity of alignment. We determined that 2/4 non-PAR GC tracts were in X to Y direction and 2/4 were in X->Y (summarized in Table. 1). Due to small sample size of these tracts we can't speculate on the relative frequency of direction in which gene conversion occurs. In humans the X to Y direction of gene conversion is more well documented compared to Y to X. The recombination on X chromosomes in females fragments the conversion tracts that may have been transposed to X from Y making it difficult to detect them. While in males, the tracts transposed from X to Y remain stable and can be detected. It is not clear if either direction is favored more over the other (ROSSER et al. 2009).

All the gene conversion tracts that we found on the sex chromosomes were either in non-genic regions or in introns (4/16 in introns and 12/16 in intergenic regions). This is in contrast to observations in human gene conversion hotspots which tend to be enriched in exons (TROMBETTA et al. 2014; TROMBETTA et al. 2017).

I have shown here that gene conversion may occur on sex chromosomes in non-PAR regions. Since I observed 2 crossover events in ~15 individuals in stratum2 and stratum3 each, I calculated the rate of conversion per generation per base pair. The upper bound and lower bound of this rate will depend on the average minimum and maximum

tract length (stratum2: min=9 bp, max=172 bp; stratum3: min=89 bp, max=1680.5 bp). Number of bases converted was divided by read depth in the particular stratum and the total number of bases in observation which will be a number of bases in that stratum in both sex chromosomes. This results in the rate of $1.2x10^{-7}$ -$2.4x10^{-6}$ bases in stratum2 that will be converted per individual and $1.1x10^{-6}$-$2.04x10^{-5}$ bases in stratum3 that will be converted per individual. This rate should be compared with mutation rate in Threespine sticklebacks to be placed in a meaningful context of how these events will affect sequence evolution. The gene conversion rate in the PAR was computed to be $1.1x10^{-5}$-$1.1x10^{-4}$ converted bases per individual.

Discussion

The crossovers we found in the autosomes are concentrated in the end of the chromosome which is an expected pattern in males and lends some credibility to our pipeline that we created to find crossover and gene conversion tracts. However, we still find tracts in the non-telomeric regions of the chromosomes in high numbers. In mice, 90% of the DSBs are formed are inferred to be repaired through homologous gene conversion across all autosomes (Li *et al.* 2019). Assuming the same conversion rate and 37 DSBs per meiosis as inferred from Chapter 2, we expect around 33 gene conversion events, however we observe 97 events per meiosis from our pipeline. One of the reasons for such patterns could be that reads aligning from paralogous regions could give us a false signal of conversion or crossover and a more stringent filter is needed. This can be exacerbated if such regions are highly repetitive. Such repetitive regions should be excluded in next steps of the pipeline.

We did not find any crossovers beyond PAR in our dataset which is expected as such events may be deleterious. However, aberrant recombination occurs at a low rate in sex chromosomes of humans which sometimes leads to sex reversal and by sequencing at a higher depth we can recover such events (FERGUSON-SMITH 1965; SCHIEBEL *et al.* 1997).

Figures

Figure1. Schematic for detecting gene conversion tracts (A) The consensus sequence for each haplotype was determined by counting the allele in maximum number of reads aligning to each position. While nucleotides in most positions were same in both haplotypes, some position differed and were termed as SNPs. The sequencing errors/mutations are shown in red (B) Most reads matched the consensus sequences perfectly (read 11 and read 13) but some showed a conversion patter (read 12) or a crossover pattern (read 14)

Figure 2. Distribution of Gene conversion and Crossover tracts on X chromosome in PAR, stratum1, stratum2 and stratum3.

Figure 3. Distribution of Gene conversion and Crossover tracts on chr. XI in 2 Mb windows.

Table 1. Summary of gene conversion tracts involving sex chromosomes.

References

Alison E Wright, R. D., Fabian Zimmer, Judith E. Mark 2016 How to make a sex chromosome. Nature Communications.

Bachtrog, D., 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet 14**:** 113-124.

Bachtrog, D., J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto *et al.*, 2014 Sex determination: why so many ways of doing it? PLoS Biol 12**:** e1001899.

Bellott, D. W., J. F. Hughes, H. Skaletsky, L. G. Brown, T. Pyntikova *et al.*, 2014 Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508**:** 494-499.

Chandley, A. C., P. Goetz, T. B. Hargreave, A. M. Joseph and R. M. Speed, 1984 On the nature and extent of XY pairing at meiotic prophase in man. Cytogenet Cell Genet 38**:** 241-247.

Charlesworth, D., B. Charlesworth and G. Marais, 2005 Steps in the evolution of heteromorphic sex chromosomes. Heredity (Edinb) 95**:** 118-128.

Chen, J. M., D. N. Cooper, N. Chuzhanova, C. Ferec and G. P. Patrinos, 2007 Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8**:** 762-775.

Cortez, D., R. Marin, D. Toledo-Flores, L. Froidevaux, A. Liechti *et al.*, 2014 Origins and functional evolution of Y chromosomes across mammals. Nature 508**:** 488-493.

FERGUSON-SMITH, M. A., 1965 KARYOTYPE-PHENOTYPE CORRELATIONS IN GONADAL DYSGENESIS AND THEIR BEARING ON THE PATHOGENESIS OF MALFORMATIONS. J Med Genet 2**:** 142-155.

Jain, M., S. Koren, K. H. Miga, J. Quick, A. C. Rand *et al.*, 2018 Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36**:** 338-345.

Li, R., E. Bitoun, N. Altemose, R. W. Davies, B. Davies *et al.*, 2019 A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination. Nat Commun 10**:** 3900.

Martin, M., M. Patterson, S. Garg, S. O Fischer, N. Pisanti *et al.*, 2016 WhatsHap: fast and accurate read-based phasing. bioRxiv**:** 085050.

Padhukasahasram, B., and B. Rannala, 2013 Meiotic gene-conversion rate and tract length variation in the human genome. Eur J Hum Genet.

Peichel, C. L., S. R. McCann, J. A. Ross, A. F. S. Naftaly, J. R. Urton *et al.*, 2020 Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol 21**:** 177.

Pigozzi, M. I., 2016 The Chromosomes of Birds during Meiosis. Cytogenet Genome Res 150**:** 128-138.

Ross, J. A., J. R. Urton, J. Boland, M. D. Shapiro and C. L. Peichel, 2009 Turnover of sex chromosomes in the stickleback fishes (gasterosteidae). PLoS Genet 5**:** e1000391.

Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. McLay *et al.*, 2005 The DNA sequence of the human X chromosome. Nature 434**:** 325-337.

Rosser, Z. H., P. Balaresque and M. A. Jobling, 2009 Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. Am J Hum Genet 85**:** 130-134.

Schiebel, K., M. Winkelmann, A. Mertz, X. Xu, D. C. Page *et al.*, 1997 Abnormal XY interchange between a novel isolated protein kinase gene, PRKY, and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y-)XY females. Hum Mol Genet 6**:** 1985-1989.

Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier *et al.*, 2003 The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423**:** 825-837.

Trombetta, B., and F. Cruciani, 2017 Y chromosome palindromes and gene conversion. Hum Genet 136**:** 605-619.

Trombetta, B., F. Cruciani, P. A. Underhill, D. Sellitto and R. Scozzari, 2010 Footprints of X-to-Y gene conversion in recent human evolution. Mol Biol Evol 27**:** 714-725.

Trombetta, B., E. D'Atanasio and F. Cruciani, 2017 Patterns of Inter-Chromosomal Gene Conversion on the Male-Specific Region of the Human Y Chromosome. Front Genet 8**:** 54.

Trombetta, B., G. Fantini, E. D'Atanasio, D. Sellitto and F. Cruciani, 2016 Evidence of extensive non-allelic gene conversion among LTR elements in the human genome. Sci Rep 6**:** 28710.

Trombetta, B., D. Sellitto, R. Scozzari and F. Cruciani, 2014 Inter- and intraspecies phylogenetic analyses reveal extensive X-Y gene conversion in the evolution of gametologous sequences of human sex chromosomes. Mol Biol Evol 31**:** 2108-2123.

Wenger, A. M., P. Peluso, W. J. Rowell, P. C. Chang, R. J. Hall *et al.*, 2019 Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol 37**:** 1155-1162.

Wright, A. E., P. W. Harrison, S. H. Montgomery, M. A. Pointer and J. E. Mank, 2014 Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. Evolution 68**:** 3281-3295.

Figures

A

Position \ Read	read 1	read 2	read 3	read 4	read 5	Consensus H1
1	A	A	A	A	A	A
2	G	G	C	G	G	G
3	T	T	T	T	A	T
4	A	A	A	G	A	A
5	A	A	A	A	A	A
6	C	C	C	C	C	C
7	G	T	T	T	T	T
8	A	A	C	A	A	A
9	C	C	C	C	C	C
10	G	T	T	T	T	T
11	A	A	C	A	A	A

Position \ Read	read 6	read 7	read 8	read 9	read 10	Consensus H2
1	A	A	A	A	A	A
2	G	G	C	G	G	G
3	C	C	C	C	A	C
4	A	A	A	G	A	A
5	A	A	A	A	A	A
6	G	G	G	G	G	G
7	G	T	T	T	T	T
8	A	A	C	A	A	A
9	G	G	G	G	G	G
10	G	T	T	T	T	T
11	A	A	C	A	A	A

B

Consensus H1	--T--C--C--
Consensus H2	--C--G--G--
Read 11	--C--G--G--
Read 12	--T--G--C--
Read 13	--T--C--C--
Read 14	--T--G--G--

Figure4.1. Schematic for detecting gene conversion tracts (A) The consensus sequence for each haplotype was determined by counting the allele in maximum number of reads aligning to each position. While nucleotides in most positions were same in both haplotypes, some position differed and were termed as SNPs. The sequencing errors/mutations are shown in red (B) Most reads matched the consensus sequences perfectly (read 11 and read 13) but some showed a conversion patter (read 12) or a crossover pattern (read 14)

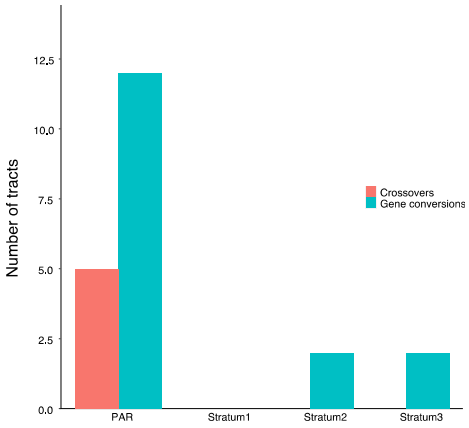


Figure 4.2. Distribution of Gene conversion and Crossover tracts on X chromosome in PAR, stratum1, stratum2 and stratum3.

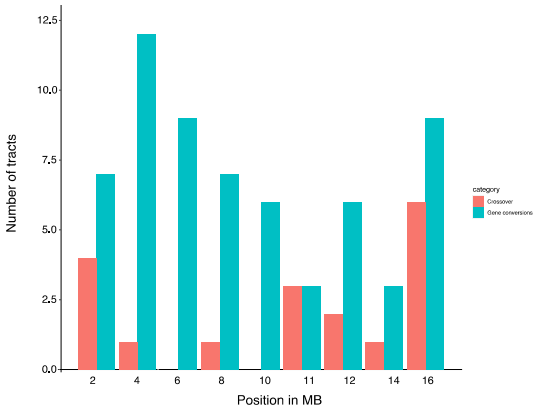


Figure 4.3. Distribution of Gene conversion and Crossover tracts on chr. XI in 2 Mb windows.

Table 4.1. Summary of gene conversion tracts involving sex chromosomes.

XIX positions	Region	Feature overlap	Min tract length (bp)	Max tract length (bp)	
260325	PAR	intron	32	211	
496644	PAR	intergenic	13	199	
505804	PAR	intergenic	43	2,819	

633038	PAR	intron	8	549	
1356865	PAR	intergenic	293	497	
1720751	PAR	intergenic	55	124	
1915161	PAR	intergenic	36	3223	
1984130	PAR	intergenic	187	418	
2188041	PAR	intergenic	97	234	
2286850	PAR	intron	62	217	
2343284	PAR	intergenic	7	91	
2370746	PAR	intergenic	92	630	
4197853	Stratum2	intergenic	12	46	X->Y
5653640	Stratum2	intron	6	298	Y->X
10569694	Stratum2	intergenic	10	871	X->Y
11188193	Stratum3	intergenic	168	2490	Y->X
Average Tract length			70.06	807.31	

CHAPTER 5

FUTURE DIRECTIONS

Studying Synaptic adjustment at a finer scale

X chromosome in Threespine stickleback system is almost 3Mb longer than Y which poses the issue of unsynapsed extra sequence on X chromosomes during meiosis (PEICHEL *et al.* 2020). We showed evidence for synaptic adjustment on sex chromosomes in late prophase in my first chapter. Briefly, the X chromosome shortens its axis to minimize the unsynapsed region on the distal end of the PAR by increasing its chromatin loop size. However, we did not show how the loop size changes at a local level on X chromosomes. It is possible that the loop size along the entire lengths of the X chromosomes is not being remodeled and a local change in the loop size around stratum1 is sufficient to fully adjust the chromosomal axes. The stratum2 and stratum3 of X and Y chromosomes are largely syntenic and adjacent to each other in both sex chromosomes in Threespine sticklebacks. Stratum 1 is close to PAR in Y chromosomes while stratum 1 is distal to PAR in X chromosomes. With the help of FISH probes for the PAR, stratum2 and stratum3 on X chromosomes, we will be able to address if whole X chromosome is remodeled while undergoing synaptic adjustment. I predict that while PAR will maintain its chromatin conformation intact through synaptic adjustment, the regions closest to stratum1 will show the greatest change in loops sizes. In my third chapter, we found four inter gametolog conversion tracts that are in stratum2 and stratum3. If some DSBs on sex chromosomes indeed repair through intergametolog gene conversion, widespread

adjustment in stratum2 and 3 will be detrimental to such repair foci. However, if the adjustment is shown to be largely local and confined to stratum3, then most regions of stratum2 and 3 will not be affected by such modifications.

Modifications in synaptonemal complex during synaptic adjustment

While synaptic adjustment is well documented in other systems, the dynamics of synaptonemal complex during adjustment is not explored. The synaptonemal complex is a structure composing multiple classes of proteins localizing between sister chromatids and homologous chromosomes, that aid in the condensation of chromosomes during meiosis, homology search and repair (BOLCUN-FILAS and HANDEL 2018). Of particular interest is SCP1 that dimerizes between the homologous chromosomes like a "zipper" and specifically marks the regions that are synapsed (MEUWISSEN *et al.* 1997). During synaptic adjustment, the SCP1 association may undergo disassociation to aid is chromosome movement. Alternatively, if temporary disassociation is not observed then it will show that synaptonemal complex is perhaps intact during adjustment and does not aid directly in the process.

Using stickleback clade to study evolution of meiotic processes

The mammalian sex chromosome system arose ~200 million years ago when *SRY* evolved as a sex determining gene. The Eutherians and Marsupials diverged ~180 million years ago and share the two oldest strata. Since the split the Eutherians and Marsupials have evolved independently in terms of gene retention, gene trafficking and even evolved distinct pairing mechanisms during meiosis (BELLOTT *et al.* 2014; CORTEZ *et al.* 2014). While sex chromosomes in eutherians mostly pair at the PAR, marsupial sex chromosomes do not synapse (SOLARI and BIANCHI 1975; PAGE *et al.* 2005; KAUPPI *et al.*

2011; ACQUAVIVA *et al.* 2020). The sex chromosomes in marsupials are held in close proximity by proteinaceous structures sometimes referred to as “dense plate”. This loss of pairing is due to loss of PAR. Some members of Gerbillinae and Arvicolinae families are a notable exception to the canonical pairing conformation in eutherians (DE LA FUENTE *et al.* 2007; BORODIN *et al.* 2012; DUMONT *et al.* 2018). The sex chromosomes in these systems don’t pair and may or may not be accompanied by loss of PAR. The young sex chromosomes of threespine sticklebacks and related species also offer a great system to study variation of pairing in sex chromosomes (ROSS *et al.* 2009). The ninespine sticklebacks have a XY sex chromosome system involving linkage groups (LG) 12, the fourspine sticklebacks have a ZW sex chromosome system using yet unknown LGs and finally blackspotted has a XXY sex chromosome system with LG 12 and 19 as their sex chromosomes. The Blackspotted and Threespine sticklebacks diverged from each other ~15 million years ago while their ancestor diverged from the clade with ninespine and fourspine around 35 million years ago (ROSS *et al.* 2009; DIXON *et al.* 2019; SARDELL *et al.* 2020). More recently, the Y chromosomes in Japan sea population of threespine sticklebacks was discovered to be fused with LG 9 and is now characterized as a separate species. As evident, the sex chromosome systems have been turned over multiple times in this clade and can be useful in studying how meiosis specific processes like sex chromosome pairing, Double stranded break formation and repair have been conserved or diverged across this clade. The sex chromosomes in these systems also use multiple different LGs as their sex chromosomes and will be very useful in exploring if specific chromosomes can be associated with different pairing conformations or meiotic behaviors.

The ZW sex chromosome system in birds also show an extensive non-homologous pairing despite an extremely degenerated W chromosome (PIGOZZI 2016). The sex chromosomes in chicken synapse ~88% of the times and due to the absence of any un-synapsed regions, MSCI (Meiotic Sex Chromosome Inactivation) is not observed (GUIOLI *et al.* 2012; PIGOZZI 2016). Mammals show MSCI due to PAR limited pairing of sex chromosomes which harbor un-synapsed non-PAR regions (TURNER *et al.* 2006; DE VRIES *et al.* 2012; FEDERICI *et al.* 2015). Since Threespine sticklebacks synapse fully and may only have small un-synapsed regions before chromosomes fully adjust, we expect MSCI markers like phosphorylated H2AX to only accumulate transiently. However, this needs to be tested in Threespine sticklebacks.

Patterns of DSB formation on sex chromosomes

We found in this dissertation that sex chromosome pair in Zygotene and Pachytene has similar number of DSBs as autosomes. This observation could reflect young sex chromosomes do not show repression of DSBs as seen in old sex chromosomes. However, we did not show if the sex chromosome pairing in threespine sticklebacks is dependent on DSB formation. In systems like mouse and yeast the chromosomes do not synapse in the absence of DSBs as synapsis is mediated by homology search in order to repair the DSBs (ROMANIENKO and CAMERINI-OTERO 2000). On the other hand, in systems like fly and worms *Spo11* is dispensable (DERNBURG *et al.* 1998; MACQUEEN *et al.* 2002). Pairing in latter is independent of DSBs. Dependence of DSBs on pairing can be addressed by knocking out *Spo11*. If pairing in non-PAR regions of sex chromosomes of threespine sticklebacks is shown to be dependent on DSBs then,

we can interpret that the DSBs in non-PAR regions are involved in homology search and may get repaired using the other gametolog.

One of the major caveats of our approach using cytology was that we could not distinguish the two chromosomes from each other in early stages of prophase before they paired. Hence, we could not report the number of DSBs forming on X and Y chromosomes separately. Hence, it is possible that most DSBs were forming on just one of the gametologs while they were suppressed on the other. In order to distinguish the rate of DSBs forming on each sex chromosomes, chromatin immuno-precipitation with Rad51 or Spo11 can be used (LANGE *et al.* 2016; HINCH *et al.* 2020).

Further approaches to study gene conversion on sex chromosomes

Gene conversion on sex chromosomes in humans and other animals has historically been studied with population data. Typically, Y chromosomes have a very low haplotype diversity compared to autosomes and X chromosomes (WILSON SAYRES *et al.* 2014). If a SNP arises in Y chromosomes in a population, sequencing data can be used to determine if the SNPs and/or polymorphic sites observed in Y chromosomes of certain individuals are shared with X chromosomes (ROSSER *et al.* 2009; TROMBETTA *et al.* 2010; TROMBETTA *et al.* 2014; TROMBETTA *et al.* 2016; TROMBETTA and CRUCIANI 2017; TROMBETTA *et al.* 2017). If more than one SNP is close proximity is shared between X and Y chromosomes, then it strongly suggests a gene conversion event. This population-based approach can also be used in Threespine sticklebacks to detect gene conversion between sex chromosomes. Threespine sticklebacks also have multiple freshwater and marine populations across northern hemisphere and SNP patterns can be compared

between populations to reveal population specific patterns of distribution of ongoing gene conversion events (FANG *et al.* 2018).

When constructing a gene tree across multiple species of mammals or birds with genes that are shared between sex chromosomes, the genes cluster by chromosomes instead of species of origin (WRIGHT *et al.* 2014). However, this pattern can be disrupted if gene conversion event has been active in one of the species which then leads to genes clustering together by species. Similarly, when comparing multiple populations in Threespine sticklebacks, a phylogenetic approach can be used to explore ancestral gene conversion events between sex chromosomes between shared genes.

Gene conversion between gametologs can have a protective effect on the Y linked genes which can escape degeneration by purging deleterious mutations using this mechanism. The rate of gene conversion inferred from frameworks mentioned in previous paragraphs can be used to model the effect of such events on the evolution of sex chromosomes. Specifically, it can be used to model the effect of gene conversion on degeneration of Y-linked alleles. Previously, the effect of intra-chromosomal gene conversions has been modeled and predicted how duplicates on sex chromosomes have a longer lifespan as they can use the duplicated genes for gene conversion (CONNALLON and CLARK 2010; MARAIS *et al.* 2010).

References

Acquaviva, L., M. Boekhout, M. E. Karasu, K. Brick, F. Pratto *et al.*, 2020 Ensuring meiotic DNA break formation in the mouse pseudoautosomal region. Nature 582**:** 426-431.

Alexandrov, I. A., L. I. Medvedev, T. D. Mashkova, L. L. Kisselev, L. Y. Romanova *et al.*, 1993 Definition of a new alpha satellite suprachromosomal family characterized by monomeric organization. Nucleic Acids Res 21**:** 2209-2215.

Alison E Wright, R. D., Fabian Zimmer, Judith E. Mark 2016 How to make a sex chromosome. Nature Communications.

Almeida, P., B. A. Sandkam, J. Morris, I. Darolti, F. Breden *et al.*, 2021 Divergence and Remarkable Diversity of the Y Chromosome in Guppies. Mol Biol Evol 38**:** 619-633.

Arnegard, M. E., M. D. McGee, B. Matthews, K. B. Marchinko, G. L. Conte *et al.*, 2014 Genetics of ecological divergence during speciation. Nature 511**:** 307-311.

Ashley, T., 1990 Axial shortening during pachynema unrelated to nonhomologous synapsis. Cytogenet Cell Genet 53**:** 185-190.

Bachtrog, D., 2013 Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet 14**:** 113-124.

Bachtrog, D., J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto *et al.*, 2014 Sex determination: why so many ways of doing it? PLoS Biol 12**:** e1001899.

Barlow, A. L., F. E. Benson, S. C. West and M. A. Hultén, 1997 Distribution of the Rad51 recombinase in human and mouse spermatocytes. EMBO J 16**:** 5207-5215.

Baudat, F., and B. de Massy, 2007 Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. Chromosome Res 15**:** 565-577.

Baudat, F., K. Manova, J. P. Yuen, M. Jasin and S. Keeney, 2000 Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. Mol Cell 6**:** 989-998.

Bell, M., and S. A. Foster, 1994 The evolutionary biology of the threespine sticklebacks, pp. Oxford University Press.

Bell, M. A., 1995 Sticklebacks: a model for behavior evolution. Trends Ecol Evol 10**:** 101-103.

Bell, M. A., J. D. Stewart and P. J. Park, 2009 The World's Oldest Fossil Threespine Stickleback Fish. Copeia 2009**:** 256-265, 210.

Bellott, D. W., J. F. Hughes, H. Skaletsky, L. G. Brown, T. Pyntikova *et al.*, 2014 Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature 508**:** 494-499.

Benjamini, Y., and T. P. Speed, 2012 Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res 40**:** e72.

Berner, D., M. Roesti, S. Bilobram, S. K. Chan, H. Kirk *et al.*, 2019 Sequencing, Assembly, and Annotation of Four Threespine Stickleback Genomes Based on Microfluidic Partitioned DNA Libraries. Genes (Basel) 10.

Blokhina, Y. P., A. D. Nguyen, B. W. Draper and S. M. Burgess, 2019 The telomere bouquet is a hub where meiotic double-strand breaks, synapsis, and stable homolog juxtaposition are coordinated in the zebrafish, Danio rerio. PLoS Genet 15**:** e1007730.

Bolcun-Filas, E., and M. A. Handel, 2018 Meiosis: the chromosomal foundation of reproduction. Biol Reprod 99**:** 112-126.

Borodin, P. M., E. A. Basheva, A. A. Torgasheva, O. A. Dashkevich, F. N. Golenishchev *et al.*, 2012 Multiple independent evolutionary losses of XY pairing at meiosis in the grey voles. Chromosome Res 20**:** 259-268.

Bosch, E., M. E. Hurles, A. Navarro and M. A. Jobling, 2004 Dynamics of a human interparalog gene conversion hotspot. Genome Res 14**:** 835-844.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10**:** 421.

Cantarel, B. L., I. Korf, S. M. Robb, G. Parra, E. Ross *et al.*, 2008 MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18**:** 188-196.

Carnero, A., R. Jiménez, M. Burgos, A. Sánchez and R. Díaz de la Guardia, 1991 Achiasmatic sex chromosomes in Pitymys duodecimcostatus: mechanisms of association and segregation. Cytogenet Cell Genet 56**:** 78-81.

Ceccaldi, R., B. Rondinelli and A. D. D'Andrea, 2016 Repair Pathway Choices and Consequences at the Double-Strand Break. Trends Cell Biol 26**:** 52-64.

Cech, J. N., and C. L. Peichel, 2015 Identification of the centromeric repeat in the threespine stickleback fish (Gasterosteus aculeatus). Chromosome Res 23**:** 767-779.

Chandley, A. C., P. Goetz, T. B. Hargreave, A. M. Joseph and R. M. Speed, 1984 On the nature and extent of XY pairing at meiotic prophase in man. Cytogenet Cell Genet 38**:** 241-247.

Chang, C. H., and A. M. Larracuente, 2019 Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the. Genetics 211**:** 333-348.

Charlesworth, D., B. Charlesworth and G. Marais, 2005 Steps in the evolution of heteromorphic sex chromosomes. Heredity (Edinb) 95**:** 118-128.

Chen, J. M., D. N. Cooper, N. Chuzhanova, C. Ferec and G. P. Patrinos, 2007 Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8**:** 762-775.

Connallon, T., and A. G. Clark, 2010 Gene duplication, gene conversion and the evolution of the Y chromosome. Genetics 186**:** 277-286.

Conte, M. A., R. Joshi, E. C. Moore, S. P. Nandamuri, W. J. Gammerdinger *et al.*, 2019 Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. Gigascience 8.

Cortez, D., R. Marin, D. Toledo-Flores, L. Froidevaux, A. Liechti *et al.*, 2014 Origins and functional evolution of Y chromosomes across mammals. Nature 508**:** 488-493.

Cresko, W. A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al.*, 2004 Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proc Natl Acad Sci U S A 101**:** 6050-6055.

Cuñado, N., J. Barrios, E. S. Miguel, R. Amaro, C. Fernández *et al.*, 2002 Synaptonemal complex analysis in oocytes and spermatocytes of threespine stickleback Gasterosteus aculeatus (Teleostei, Gasterosteidae). Genetica 114**:** 53-56.

Darolti, I., A. E. Wright and J. E. Mank, 2020 Guppy Y Chromosome Integrity Maintained by Incomplete Recombination Suppression. Genome Biol Evol 12**:** 965-977.

Dawe, R. K., E. G. Lowry, J. I. Gent, M. C. Stitzer, K. W. Swentowsky *et al.*, 2018 A Kinesin-14 Motor Activates Neocentromeres to Promote Meiotic Drive in Maize. Cell 173**:** 839-850.e818.

de la Fuente, R., M. T. Parra, A. Viera, A. Calvente, R. Gómez *et al.*, 2007 Meiotic pairing and segregation of achiasmate sex chromosomes in eutherian mammals: the role of SYCP3 protein. PLoS Genet 3**:** e198.

de la Fuente, R., A. Sánchez, J. A. Marchal, A. Viera, M. T. Parra *et al.*, 2012 A synaptonemal complex-derived mechanism for meiotic segregation precedes the evolutionary loss of homology between sex chromosomes in arvicolid mammals. Chromosoma 121**:** 433-446.

de Vries, F. A., E. de Boer, M. van den Bosch, W. M. Baarends, M. Ooms *et al.*, 2005 Mouse Sycp1 functions in synaptonemal complex assembly, meiotic recombination, and XY body formation. Genes Dev 19**:** 1376-1389.

de Vries, M., S. Vosters, G. Merkx, K. D'Hauwers, D. G. Wansink *et al.*, 2012 Human male meiotic sex chromosome inactivation. PLoS One 7**:** e31485.

Dernburg, A. F., K. McDonald, G. Moulder, R. Barstead, M. Dresser *et al.*, 1998 Meiotic recombination in C. elegans initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. Cell 94**:** 387-398.

Devilee, P., T. Kievits, J. S. Waye, P. L. Pearson and H. F. Willard, 1988 Chromosome-specific alpha satellite DNA: isolation and mapping of a polymorphic alphoid repeat from human chromosome 10. Genomics 3**:** 1-7.

Dixon, G., J. Kitano and M. Kirkpatrick, 2019 The Origin of a New Sex Chromosome by Introgression between Two Stickleback Fishes. Mol Biol Evol 36**:** 28-38.

Dudchenko, O., S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger *et al.*, 2017 De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356**:** 92-95.

Dumont, B. L., and B. A. Payseur, 2011 Genetic analysis of genome-scale recombination rate evolution in house mice. PLoS Genet 7**:** e1002116.

Dumont, B. L., C. L. Williams, B. L. Ng, V. Horncastle, C. L. Chambers *et al.*, 2018 Relationship Between Sequence Homology, Genome Architecture, and Meiotic Behavior of the Sex Chromosomes in North American Voles. Genetics 210**:** 83-97.

Durand, N. C., M. S. Shamim, I. Machol, S. S. Rao, M. H. Huntley *et al.*, 2016 Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 3**:** 95-98.

Edlinger, B., and P. Schlögelhofer, 2011 Have a break: determinants of meiotic DNA double strand break (DSB) formation and processing in plants. J Exp Bot 62**:** 1545-1563.

Fang, B., P. Kemppainen, P. Momigliano, X. Feng and J. Merilä, 2020 On the causes of geographically heterogeneous parallel evolution in sticklebacks. Nat Ecol Evol 4**:** 1105-1115.

Fang, B., J. Merilä, F. Ribeiro, C. M. Alexandre and P. Momigliano, 2018 Worldwide phylogeny of three-spined sticklebacks. Mol Phylogenet Evol 127**:** 613-625.

Federici, F., E. Mulugeta, S. Schoenmakers, E. Wassenaar, J. W. Hoogerbrugge *et al.*, 2015 Incomplete meiotic sex chromosome inactivation in the domestic dog. BMC Genomics 16**:** 291.

Garcia, V., S. E. Phelps, S. Gray and M. J. Neale, 2011 Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. Nature 479**:** 241-244.

Giroux, C. N., M. E. Dresser and H. F. Tiano, 1989 Genetic control of chromosome synapsis in yeast meiosis. Genome 31**:** 88-94.

Glazer, A. M., E. E. Killingbeck, T. Mitros, D. S. Rokhsar and C. T. Miller, 2015 Genome Assembly Improvement and Mapping Convergently Evolved Skeletal Traits in Sticklebacks with Genotyping-by-Sequencing. G3 5**:** 1463-1472.

Gnerre, S., I. Maccallum, D. Przybylski, F. J. Ribeiro, J. N. Burton *et al.*, 2011 High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A 108**:** 1513-1518.

Greenbaum, I. F., D. W. Hale, P. D. Sudman and E. Nevo, 1990 Synaptonemal complex analysis of mole rats (Spalax ehrenbergi): unusual polymorphisms of chromosome 1. Genome 33**:** 898-902.

Greig, G. M., S. Parikh, J. George, V. E. Powers and H. F. Willard, 1991 Molecular cytogenetics of alpha satellite DNA from chromosome 12: fluorescence in situ hybridization and description of DNA and array length polymorphisms. Cytogenet Cell Genet 56**:** 144-148.

Grelon, M., D. Vezon, G. Gendrot and G. Pelletier, 2001 AtSPO11-1 is necessary for efficient meiotic recombination in plants. EMBO J 20**:** 589-600.

Guioli, S., R. Lovell-Badge and J. M. Turner, 2012 Error-prone ZW pairing and no evidence for meiotic sex chromosome inactivation in the chicken germ line. PLoS Genet 8**:** e1002560.

Hart, J. C., N. A. Ellis, M. B. Eisen and C. T. Miller, 2018 Convergent evolution of gene expression in two high-toothed stickleback populations. PLoS Genet 14**:** e1007443.

Hartley, G., and R. J. O'Neill, 2019 Centromere Repeats: Hidden Gems of the Genome. Genes (Basel) 10.

Hassold, T., and P. Hunt, 2001 To err (meiotically) is human: the genesis of human aneuploidy. Nat Rev Genet 2**:** 280-291.

Hatfield, T., and D. Schluter, 1999 Ecological speciation in sticklebacks: environment-dependent hybrid fitness. Evolution 53**:** 866-873.

He, S., L. Li, L. Y. Lv, W. J. Cai, Y. Q. Dou *et al.*, 2020 Mandarin fish (Sinipercidae) genomes provide insights into innate predatory feeding. Commun Biol 3**:** 361.

Heras, J., M. Chakraborty, J. J. Emerson and D. P. German, 2020 Genomic and biochemical evidence of dietary adaptation in a marine herbivorous fish. Proc Biol Sci 287**:** 20192327.

Hinch, A. G., P. W. Becker, T. Li, D. Moralli, G. Zhang *et al.*, 2020 The Configuration of RPA, RAD51, and DMC1 Binding in Meiosis Reveals the Nature of Critical Recombination Intermediates. Mol Cell 79**:** 689-701.e610.

Holt, C., and M. Yandell, 2011 MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12**:** 491.

Ishikawa, A., and J. Kitano, 2020 Diversity in reproductive seasonality in the three-spined stickleback,. J Exp Biol 223.

Jain, M., S. Koren, K. H. Miga, J. Quick, A. C. Rand *et al.*, 2018a Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36**:** 338-345.

Jain, M., H. E. Olsen, D. J. Turner, D. Stoddart, K. V. Bulazel *et al.*, 2018b Linear assembly of a human centromere on the Y chromosome. Nat Biotechnol 36**:** 321-323.

Jasin, M., and R. Rothstein, 2013 Repair of strand breaks by homologous recombination. Cold Spring Harb Perspect Biol 5**:** a012740.

Jiménez, R., A. Carnero, M. Burgos, A. Sánchez and R. Díaz de la Guardia, 1991 Achiasmatic giant sex chromosomes in the vole Microtus cabrerae (Rodentia, Microtidae). Cytogenet Cell Genet 57**:** 56-58.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli *et al.*, 2012 The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484**:** 55-61.

Kauppi, L., M. Barchi, F. Baudat, P. J. Romanienko, S. Keeney *et al.*, 2011 Distinct properties of the XY pseudoautosomal region crucial for male meiosis. Science 331**:** 916-920.

Keeney, S., C. N. Giroux and N. Kleckner, 1997 Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell 88**:** 375-384.

Kent, W. J., 2002 BLAT--the BLAST-like alignment tool. Genome Res 12**:** 656-664.

Kleckner, N., 2006 Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex. Chromosoma 115**:** 175-194.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman *et al.*, 2017 Canu: scalable and accurate long-read assembly via adaptive. Genome Res 27**:** 722-736.

Korf, I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5**:** 59.

Korunes, K. L., and M. A. F. Noor, 2019 Pervasive gene conversion in chromosomal inversion heterozygotes. Mol Ecol 28**:** 1302-1315.

Kouznetsova, A., R. Benavente, A. Pastink and C. Höög, 2011 Meiosis in mice without a synaptonemal complex. PLoS One 6**:** e28255.

Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway *et al.*, 2004 Versatile and open software for comparing large genomes. Genome Biol 5**:** R12.

Kurzbauer, M. T., C. Uanschou, D. Chen and P. Schlögelhofer, 2012 The recombinases DMC1 and RAD51 are functionally and spatially separated during meiosis in Arabidopsis. Plant Cell 24**:** 2058-2070.

Lam, I., and S. Keeney, 2014 Mechanism and regulation of meiotic recombination initiation. Cold Spring Harb Perspect Biol 7**:** a016634.

Lange, J., S. Yamada, S. E. Tischfield, J. Pan, S. Kim *et al.*, 2016 The Landscape of Mouse Meiotic Double-Strand Break Formation, Processing, and Repair. Cell 167**:** 695-708 e616.

Li, H., 2018 Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34**:** 3094-3100.

Liu, J., A. S. Seetharam, K. Chougule, S. Ou, K. W. Swentowsky *et al.*, 2020 Gapless assembly of maize chromosomes using long-read technologies. Genome Biol 21**:** 121.

Lucotte, E. A., L. Skov, J. M. Jensen, M. C. Macià, K. Munch *et al.*, 2018 Dynamic Copy Number Evolution of X- and Y-Linked Ampliconic Genes in Human Populations. Genetics 209**:** 907-920.

MacQueen, A. J., M. P. Colaiácovo, K. McDonald and A. M. Villeneuve, 2002 Synapsis-dependent and -independent mechanisms stabilize homolog pairing during meiotic prophase in C. elegans. Genes Dev 16**:** 2428-2442.

Mahtani, M. M., and H. F. Willard, 1990 Pulsed-field gel analysis of alpha-satellite DNA at the human X chromosome centromere: high-frequency polymorphisms and array size estimate. Genomics 7**:** 607-613.

Mancera, E., R. Bourgon, A. Brozzi, W. Huber and L. M. Steinmetz, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature 454**:** 479-485.

Mansai, S. P., T. Kado and H. Innan, 2011 The Rate and Tract Length of Gene Conversion between Duplicated Genes. Genes (Basel) 2**:** 313-331.

Marais, G. A., P. R. Campos and I. Gordo, 2010 Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. Genome Biol Evol 2**:** 347-357.

Martin, M., M. Patterson, S. Garg, S. O Fischer, N. Pisanti *et al.*, 2016 WhatsHap: fast and accurate read-based phasing. bioRxiv**:** 085050.

McPhail, J. D., 1992 Ecology and evolution of sympatric sticklebacks (Gasterosteus): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia. Canadian Journal of Zoology 70**:** 361-369.

Meuwissen, R. L., I. Meerts, J. M. Hoovers, N. J. Leschot and C. Heyting, 1997 Human synaptonemal complex protein 1 (SCP1): isolation and characterization of the cDNA and chromosomal localization of the gene. Genomics 39**:** 377-384.

Meyne, J., R. L. Ratliff and R. K. Moyzis, 1989 Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci U S A 86**:** 7049-7053.

Miga, K. H., S. Koren, A. Rhie, M. R. Vollger, A. Gershman *et al.*, 2020 Telomere-to-telomere assembly of a complete human X chromosome. Nature 585**:** 79-84.

Miga, K. H., Y. Newton, M. Jain, N. Altemose, H. F. Willard *et al.*, 2014 Centromere reference models for human chromosomes X and Y satellite arrays. Genome Res 24**:** 697-707.

Moens, P. B., N. K. Kolas, M. Tarsounas, E. Marcon, P. E. Cohen *et al.*, 2002 The time course and chromosomal localization of recombination-related proteins at meiosis in the mouse are compatible with models that can resolve the early DNA-DNA interactions without reciprocal recombination. J Cell Sci 115**:** 1611-1622.

Moyzis, R. K., J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven *et al.*, 1988 A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci U S A 85**:** 6622-6626.

Naftaly, A. S., S. Pau and M. A. White, 2020 Long-read RNA sequencing reveals widespread sex-specific alternative splicing in threespine stickleback fish. bioRxiv**:** 2020.2011.2012.380428.

Nagarajan, N., and M. Pop, 2013 Sequence assembly demystified. Nat Rev Genet 14**:** 157-167.

Ocalewicz, K., 2013 Telomeres in fishes. Cytogenet Genome Res 141**:** 114-125.

Ocalewicz, K., P. Woznicki, G. Furgala-Selezniow and M. Jankun, 2011 Chromosomal location of Ag/CMA 3 -NORs, 5S rDNA and telomeric repeats in two stickleback species. Italian Journal of Zoology 78**:** 12-19.

Padhukasahasram, B., and B. Rannala, 2013 Meiotic gene-conversion rate and tract length variation in the human genome. Eur J Hum Genet.

Page, J., S. Berríos, M. T. Parra, A. Viera, J. A. Suja *et al.*, 2005 The program of sex chromosome pairing in meiosis is highly conserved across marsupial species: implications for sex chromosome evolution. Genetics 170**:** 793-799.

Page, S. L., and R. S. Hawley, 2004 The genetics and molecular biology of the synaptonemal complex. Annu Rev Cell Dev Biol 20**:** 525-558.

Peichel, C. L., S. R. McCann, J. A. Ross, A. F. S. Naftaly, J. R. Urton *et al.*, 2020 Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol 21**:** 177.

Peichel, C. L., S. T. Sullivan, I. Liachko and M. A. White, 2017 Improvement of the Threespine Stickleback Genome Using a Hi-C-Based Proximity-Guided Assembly. J Hered 108**:** 693-700.

Pertile, M. D., A. N. Graham, K. H. Choo and P. Kalitsis, 2009 Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability. Genome Res 19**:** 2202-2213.

Pigozzi, M. I., 2016 The Chromosomes of Birds during Meiosis. Cytogenet Genome Res 150**:** 128-138.

Plug, A. W., A. H. Peters, Y. Xu, K. S. Keegan, M. F. Hoekstra *et al.*, 1997 ATM and RPA in meiotic chromosome synapsis and recombination. Nat Genet 17**:** 457-461.

Pracana, R., A. Priyam, I. Levantis, R. A. Nichols and Y. Wurm, 2017 The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. Mol Ecol 26**:** 2864-2879.

Prost, S., M. Petersen, M. Grethlein, S. J. Hahn, N. Kuschik-Maczollek *et al.*, 2020 Improving the Chromosome-Level Genome Assembly of the Siamese Fighting Fish. G3 10**:** 2179-2183.

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26**:** 841-842.

Rasmussen, S. W., and P. B. Holm, 1978 Human meiosis II. Chromosome pairing and recombination nodules in human spermatocytes. Carlsberg Research Communications 43**:** 275-327.

Roche, L., G. Seluja and R. Wettstein, The meiotic behaviour of the XY pair in Lutreolina crassicaudata (Marsupialia: Didelphoidea).

Rodrigues, N., T. Studer, C. Dufresnes and N. Perrin, 2017 Sex-Chromosome Recombination in Common Frogs Brings Water to the Fountain-of-Youth. Mol Biol Evol 35**:** 942-948.

Roesti, M., D. Moser and D. Berner, 2013 Recombination in the threespine stickleback genome--patterns and consequences. Mol Ecol 22**:** 3014-3027.

Romanienko, P. J., and R. D. Camerini-Otero, 2000 The mouse Spo11 gene is required for meiotic chromosome synapsis. Mol Cell 6**:** 975-987.

Ross, J. A., and C. L. Peichel, 2008 Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. Genetics 179**:** 2173-2182.

Ross, J. A., J. R. Urton, J. Boland, M. D. Shapiro and C. L. Peichel, 2009 Turnover of sex chromosomes in the stickleback fishes (gasterosteidae). PLoS Genet 5**:** e1000391.

Ross, M. G., C. Russ, M. Costello, A. Hollinger, N. J. Lennon *et al.*, 2013 Characterizing and measuring bias in sequence data. Genome Biol 14**:** R51.

Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. McLay *et al.*, 2005 The DNA sequence of the human X chromosome. Nature 434**:** 325-337.

Rosser, Z. H., P. Balaresque and M. A. Jobling, 2009 Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. Am J Hum Genet 85**:** 130-134.

Ruiz-Herrera, A., M. Vozdova, J. Fernández, H. Sebestova, L. Capilla *et al.*, 2017 Recombination correlates with synaptonemal complex length and chromatin loop size in bovids-insights into mammalian meiotic chromosomal organization. Chromosoma 126**:** 615-631.

Sardell, J. M., M. P. Josephson, A. C. Dalziel, C. L. Peichel and M. Kirkpatrick, 2020 Contrasting tempos of sex chromosome degeneration in sticklebacks. bioRxiv**:** 2020.2009.2017.300236.

Schluter, D., K. B. Marchinko, R. D. Barrett and S. M. Rogers, 2010 Natural selection and the genetics of adaptation in threespine stickleback. Philos Trans R Soc Lond B Biol Sci 365**:** 2479-2486.

Schneider, V. A., T. Graves-Lindsay, K. Howe, N. Bouk, H. C. Chen *et al.*, 2017 Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res 27**:** 849-864.

Seluja, G. A., L. Roche and A. J. Solari, 1987 Male meiotic prophase in Didelphis albiventris: A comparative cytological and electron microscopical study. Journal of Heredity 78**:** 218-222.

Sharp, P., 1982 Sex chromosome pairing during male meiosis in marsupials. Chromosoma 86**:** 27-47.

Shepelev, V. A., L. I. Uralsky, A. A. Alexandrov, Y. B. Yurov, E. I. Rogaev *et al.*, 2015 Annotation of suprachromosomal families reveals uncommon types of alpha satellite organization in pericentromeric regions of hg38 human genome assembly. Genom Data 5**:** 139-146.

Shinohara, A., H. Ogawa and T. Ogawa, 1992 Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. Cell 69**:** 457-470.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31**:** 3210-3212.

Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier *et al.*, 2003 The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423**:** 825-837.

Soh, Y. Q., J. Alfoldi, T. Pyntikova, L. G. Brown, T. Graves *et al.*, 2014 Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell 159**:** 800-813.

Solari, A. J., 1992 Equalization of Z and W axes in chicken and quail oocytes. Cytogenet Cell Genet 59**:** 52-56.

Solari, A. J., and N. O. Bianchi, 1975 The synaptic behaviour of the X and Y chromosomes in the marsupial Monodelphis dimidiata. Chromosoma 52**:** 11-25.

Stack, S. M., and W. V. Brown, 1969 Somatic pairing, reduction and recombination: an evolutionary hypothesis of meiosis. Nature 222**:** 1275-1276.

Stanke, M., A. Tzvetkova and B. Morgenstern, 2006 AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. Genome Biol 7 Suppl 1**:** S11.11-18.

Sullivan, M. R., and K. A. Bernstein, 2018 RAD-ical New Insights into RAD51 Regulation. Genes (Basel) 9.

Tarsounas, M., T. Morita, R. E. Pearlman and P. B. Moens, 1999 RAD51 and DMC1 form mixed complexes associated with mouse meiotic chromosome cores and synaptonemal complexes. J Cell Biol 147**:** 207-220.

Toraason, E., C. Clark, A. Horacek, M. L. Glover, A. Salagean *et al.*, 2020 Sister chromatid repair maintains genomic integrity during meiosis in Caenorhabditis elegans. bioRxiv**:** 2020.2007.2022.216143.

Traut, W., M. Szczepanowski, M. Vítková, C. Opitz, F. Marec *et al.*, 2007 The telomere repeat motif of basal Metazoa. Chromosome Res 15**:** 371-382.

Trombetta, B., and F. Cruciani, 2017 Y chromosome palindromes and gene conversion. Hum Genet 136**:** 605-619.

Trombetta, B., F. Cruciani, P. A. Underhill, D. Sellitto and R. Scozzari, 2010 Footprints of X-to-Y gene conversion in recent human evolution. Mol Biol Evol 27**:** 714-725.

Trombetta, B., E. D'Atanasio and F. Cruciani, 2017 Patterns of Inter-Chromosomal Gene Conversion on the Male-Specific Region of the Human Y Chromosome. Front Genet 8**:** 54.

Trombetta, B., G. Fantini, E. D'Atanasio, D. Sellitto and F. Cruciani, 2016 Evidence of extensive non-allelic gene conversion among LTR elements in the human genome. Sci Rep 6**:** 28710.

Trombetta, B., D. Sellitto, R. Scozzari and F. Cruciani, 2014 Inter- and intraspecies phylogenetic analyses reveal extensive X-Y gene conversion in the evolution of gametologous sequences of human sex chromosomes. Mol Biol Evol 31**:** 2108-2123.

Turner, J. M., S. K. Mahadevaiah, P. J. Ellis, M. J. Mitchell and P. S. Burgoyne, 2006 Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids. Dev Cell 10**:** 521-529.

van den Bosch, M., P. H. Lohman and A. Pastink, 2002 DNA double-strand break repair by homologous recombination. Biol Chem 383**:** 873-892.

Varadharajan, S., P. Rastas, A. Löytynoja, M. Matschiner, F. C. F. Calboli *et al.*, 2019 A High-Quality Assembly of the Nine-Spined Stickleback (Pungitius pungitius) Genome. Genome Biol Evol 11**:** 3291-3308.

Vollger, M. R., G. A. Logsdon, P. A. Audano, A. Sulovari, D. Porubsky *et al.*, 2020 Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads. Ann Hum Genet 84**:** 125-140.

Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis *et al.*, 2018 BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol Biol Evol 35**:** 543-548.

Wenger, A. M., P. Peluso, W. J. Rowell, P. C. Chang, R. J. Hall *et al.*, 2019 Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol 37**:** 1155-1162.

Wevrick, R., and H. F. Willard, 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 U S A 86**:** 9394-9398.

White, M. A., J. Kitano and C. L. Peichel, 2015 Purifying Selection Maintains Dosage-Sensitive Genes during Degeneration of the Threespine Stickleback Y Chromosome. Mol Biol Evol 32**:** 1981-1995.

Willard, H. F., 1985 Chromosome-specific organization of human alpha satellite DNA. Am J Hum Genet 37**:** 524-532.

Willard, H. F., J. S. Waye, M. H. Skolnick, C. E. Schwartz, V. E. Powers *et al.*, 1986 Detection of restriction fragment length polymorphisms at the centromeres of human chromosomes by using chromosome-specific alpha satellite DNA probes: implications for development of centromere-based genetic linkage maps. Proc Natl Acad Sci U S A 83**:** 5611-5615.

Wilson Sayres, M. A., K. E. Lohmueller and R. Nielsen, 2014 Natural selection reduced diversity on human y chromosomes. PLoS Genet 10**:** e1004064.

Wolfe, J., S. M. Darling, R. P. Erickson, I. W. Craig, V. J. Buckle *et al.*, 1985 Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. J Mol Biol 182**:** 477-485.

Wootton, R., 1976 The Biology of Sticklebacks, pp. Academic Press.

Wright, A. E., P. W. Harrison, S. H. Montgomery, M. A. Pointer and J. E. Mank, 2014 Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. Evolution 68**:** 3281-3295.

Xu, G. C., T. J. Xu, R. Zhu, Y. Zhang, S. Q. Li *et al.*, 2019 LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. Gigascience 8.

Zhou, Y., S. Xiao, G. Lin, D. Chen, W. Cen *et al.*, 2019 Chromosome genome assembly and annotation of the yellowbelly pufferfish with PacBio and Hi-C sequencing data. Sci Data 6**:** 267.

Zickler, D., 2020 Diter von Wettstein and The Meiotic Program of Pairing and Recombination. Methods Mol Biol 2124**:** 19-35.

Zickler, D., and N. Kleckner, 1999 Meiotic chromosomes: integrating structure and function. Annu Rev Genet 33**:** 603-754.

Zickler, D., and N. Kleckner, 2015 Recombination, Pairing, and Synapsis of Homologs during Meiosis. Cold Spring Harb Perspect Biol 7.

APPENDIX

SUPPLEMENTAL FIGURES

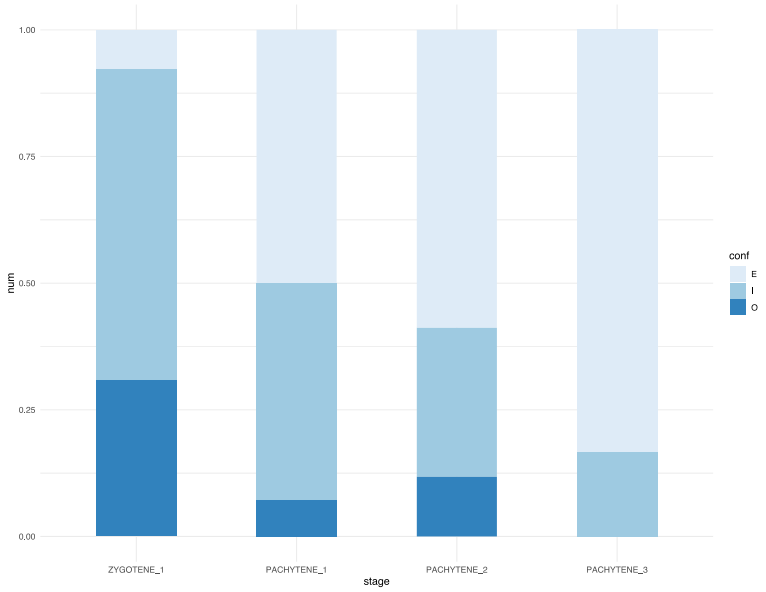


Figure S2.1.

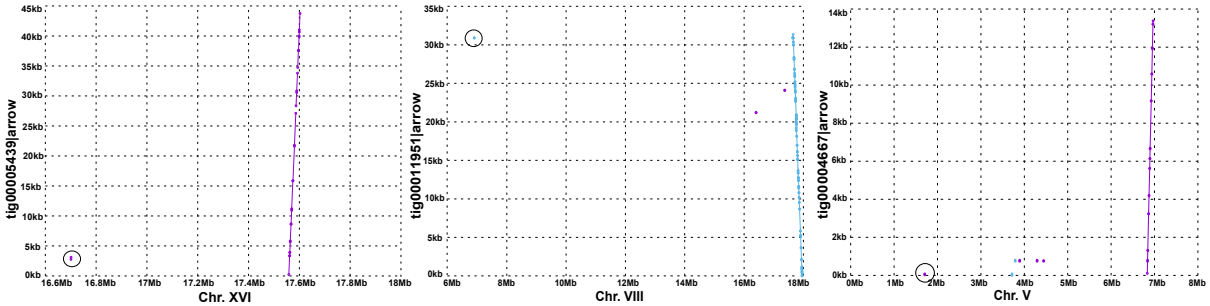


Figure S3.1. Three Paxton Lake contigs contained regions that did not align in a linear fashion. These contigs caused erroneous gap closing of the v. 5 reference assembly using LR_Gapcloser. The small regions aligning incorrectly are shown with a circle. This region was used by LR_Gapcloser to erroneously ligate the two regions of the contig, causing a decrease in chromosome scaffold length. These three contigs were removed from the assembly.

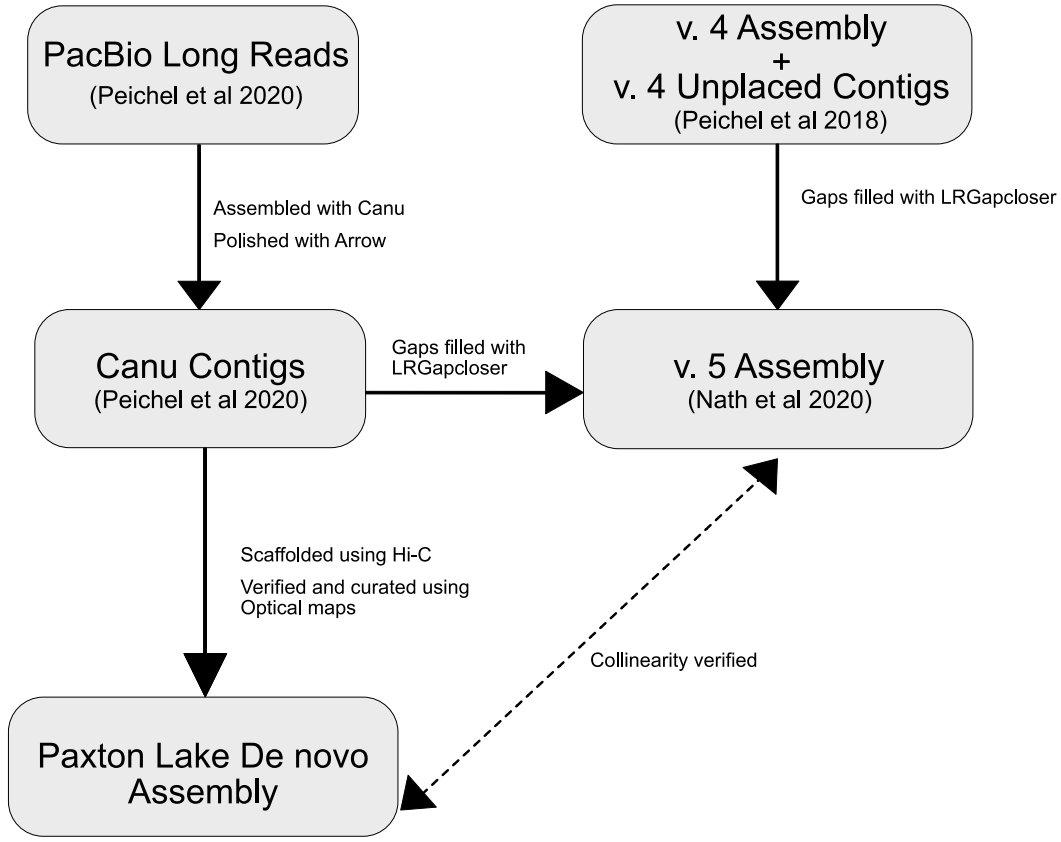


Figure S3.2. The new threespine stickleback genome (v. 5) incorporated contigs from a Paxton Lake male fish and unplaced contigs from the v. 4 genome.

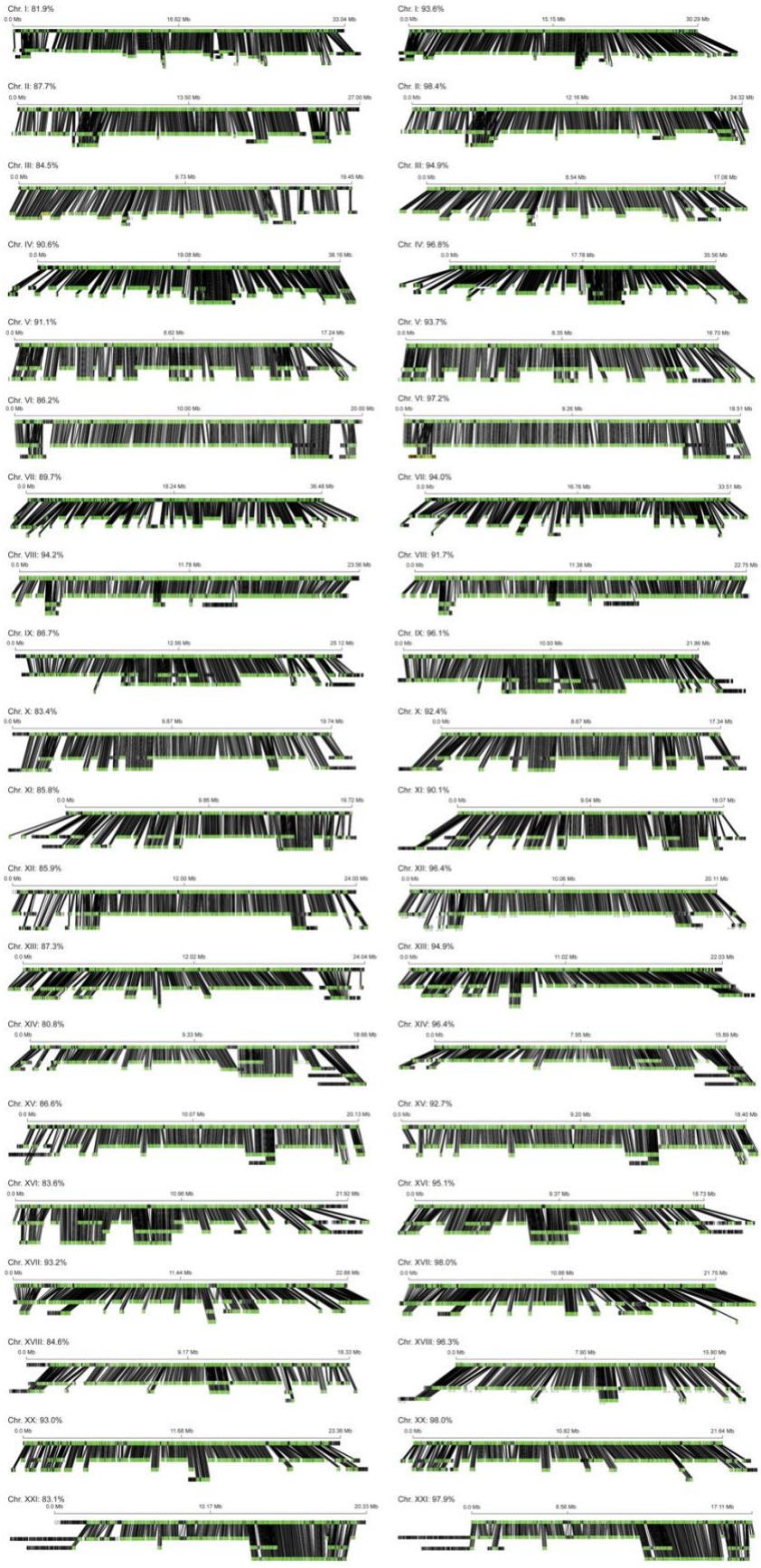
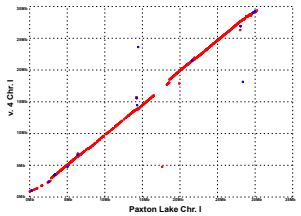
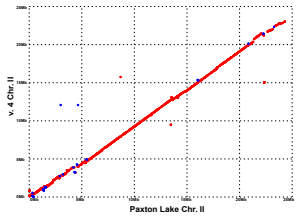
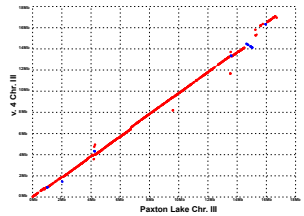
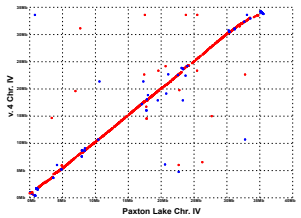

Chr. I: 81.9%
Chr. I: 93.6%
Chr. II: 87.7%
Chr. II: 98.4%
Chr. III: 84.5%
Chr. III: 94.9%
Chr. IV: 90.6%
Chr. IV: 96.8%
Chr. V: 91.1%
Chr. V: 93.7%
Chr. VI: 86.2%
Chr. VI: 97.2%
Chr. VII: 89.7%
Chr. VII: 94.0%
Chr. VIII: 94.2%
Chr. VIII: 91.7%
Chr. IX: 86.7%
Chr. IX: 96.1%
Chr. X: 83.4%
Chr. X: 92.4%
Chr. XI: 85.8%
Chr. XI: 90.1%
Chr. XII: 85.9%
Chr. XII: 96.4%
Chr. XIII: 87.3%
Chr. XIII: 94.9%
Chr. XIV: 80.8%
Chr. XIV: 96.4%
Chr. XV: 86.6%
Chr. XV: 92.7%
Chr. XVI: 83.6%
Chr. XVI: 95.1%
Chr. XVII: 93.2%
Chr. XVII: 98.0%
Chr. XVIII: 84.6%
Chr. XVIII: 96.3%
Chr. XX: 93.0%
Chr. XX: 98.0%
Chr. XXI: 83.1%
Chr. XXI: 97.9%

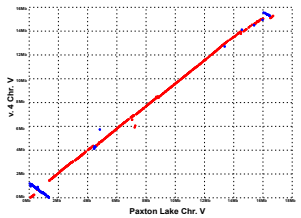
Figure S3.3. BioNano optical maps align with high coverage to the Paxton Lake *de novo* genome assembly. Optical contigs were aligned to the initial assembly (left). Contigs were removed from the initial assembly if less than 50% of its length did not overlap with an optical contig, creating the refined assembly (right). For each chromosome, the reference sequence is shown on the top and the optical contigs are shown on the bottom. Percent coverage of each chromosome by optical contigs is shown.

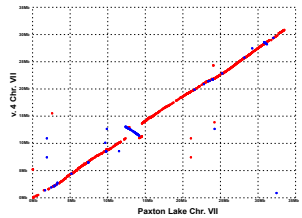

v. 4 Chr. I
Paxton Lake Chr. I

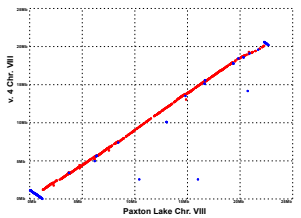

v. 4 Chr. II
Paxton Lake Chr. II

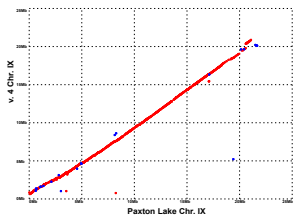

v. 4 Chr. III
Paxton Lake Chr. III

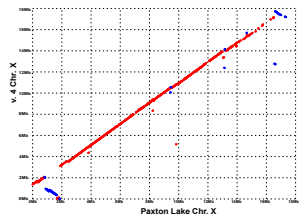

v. 4 Chr. IV
Paxton Lake Chr. IV

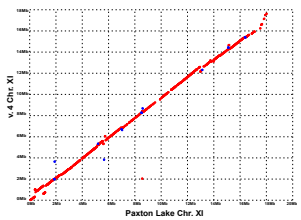

v. 4 Chr. V
Paxton Lake Chr. V

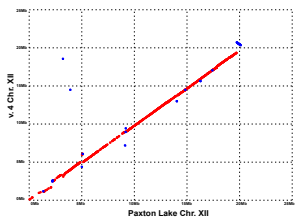

v. 4 Chr. VII
Paxton Lake Chr. VII

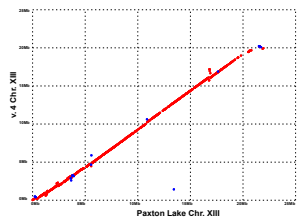

v. 4 Chr. VIII
Paxton Lake Chr. VIII

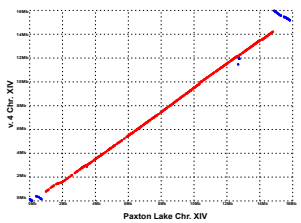

v. 4 Chr. IX
Paxton Lake Chr. IX

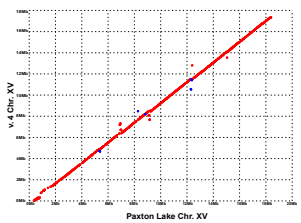

v. 4 Chr. X
Paxton Lake Chr. X

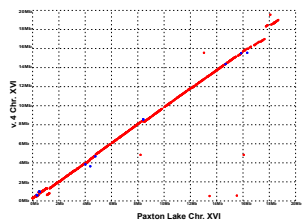

v. 4 Chr. XI
Paxton Lake Chr. XI

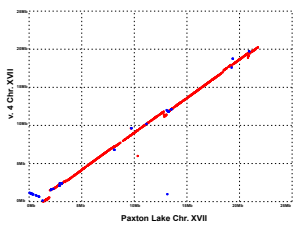

v. 4 Chr. XII
Paxton Lake Chr. XII

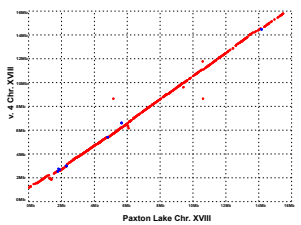

v. 4 Chr. XIII
Paxton Lake Chr. XIII

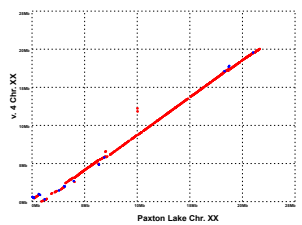

v. 4 Chr. XIV
Paxton Lake Chr. XIV


v. 4 Chr. XV
Paxton Lake Chr. XV


v. 4 Chr. XVI
Paxton Lake Chr. XVI


v. 4 Chr. XVII
Paxton Lake Chr. XVII


v. 4 Chr. XVIII
Paxton Lake Chr. XVIII


v. 4 Chr. XX
Paxton Lake Chr. XX

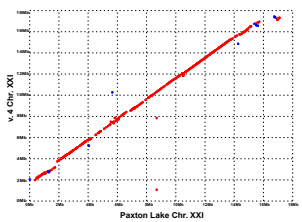

v. 4 Chr. XXI
Paxton Lake Chr. XXI

Figure S3.4. The v. 4 reference assembly and the Paxton Lake *de nevo* assembly are highly collinear. Both genomes were aligned using nucmer. Alignments are only shown if they exceed 98% sequence identity and had an alignment length of at least 4000 bp.

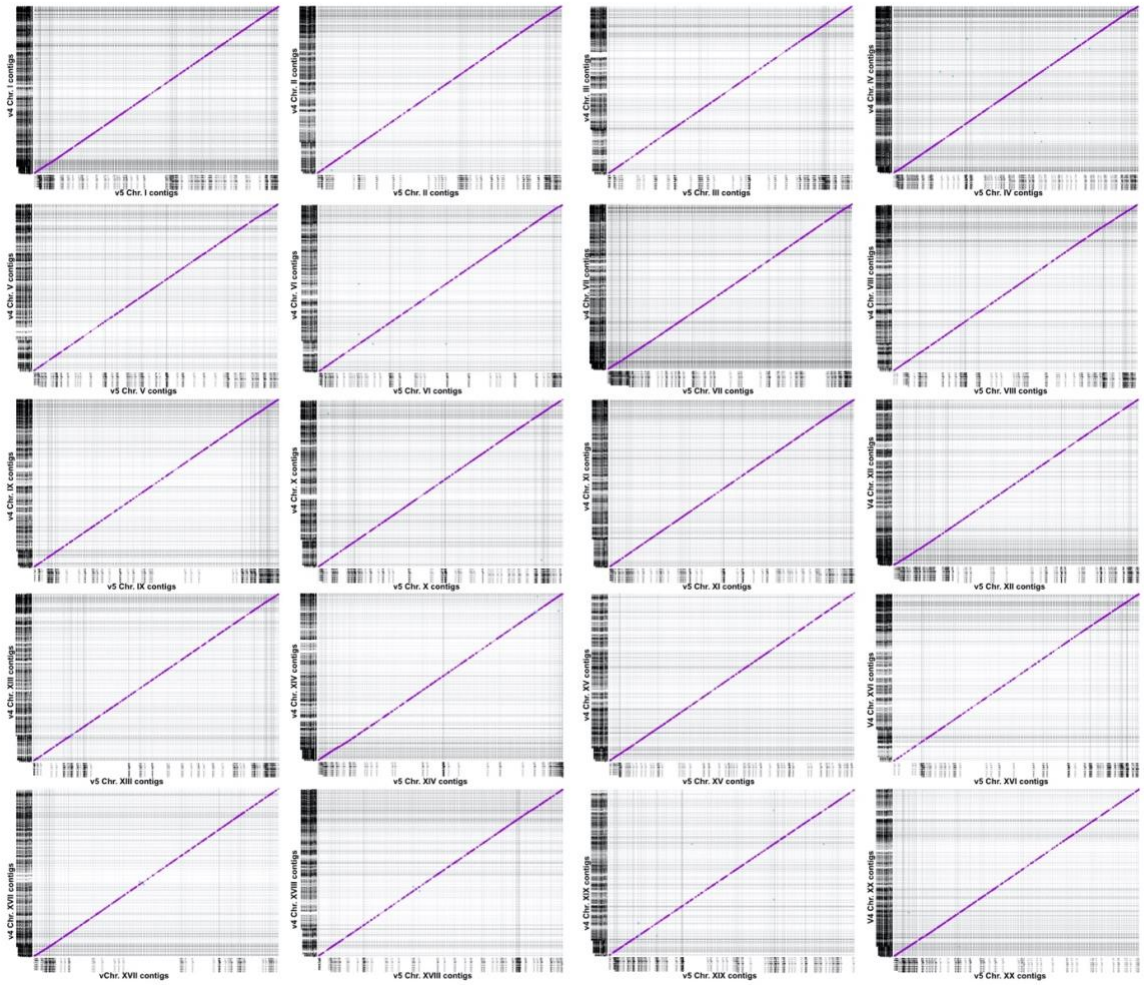


Figure S3.5. The v. 4 reference assembly and the v. 5 reference assembly are collinear. There are fewer contigs in the v. 5 assembly, revealing a more contiguous assembly across all chromosomes. Chromosomes in the v. 4 assembly were broken at N's into individual contigs and aligned to the contigs in the v. 5 reference assembly using nucmer.

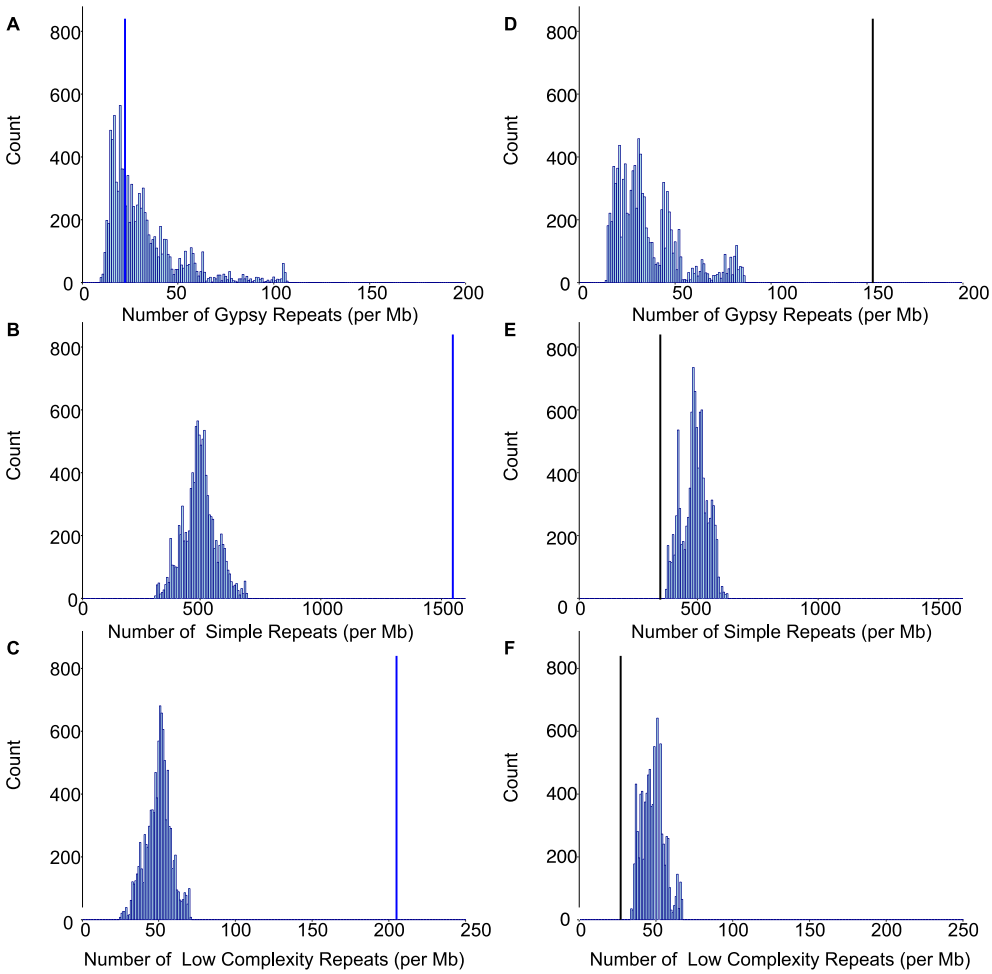


Figure S3.6. Unplaced chr. Un contigs are enriched for LTR-gypsy retrotransposons, but not low complexity repeats or simple repeats. The total density of LTR-gypsy retrotransposons (A, D), Simple repeats (B, E), and low complexity repeats (C, F) were quantified in closed gaps (blue line) and the remaining contigs that could not be placed (chr. Un; black line). 10,000 10 Mb windows were randomly drawn across the genome to generate a null background distribution for the filled gaps and 10,000 15 Mb windows were randomly drawn across the genome to generate a null background distribution for the chr. Un contigs.

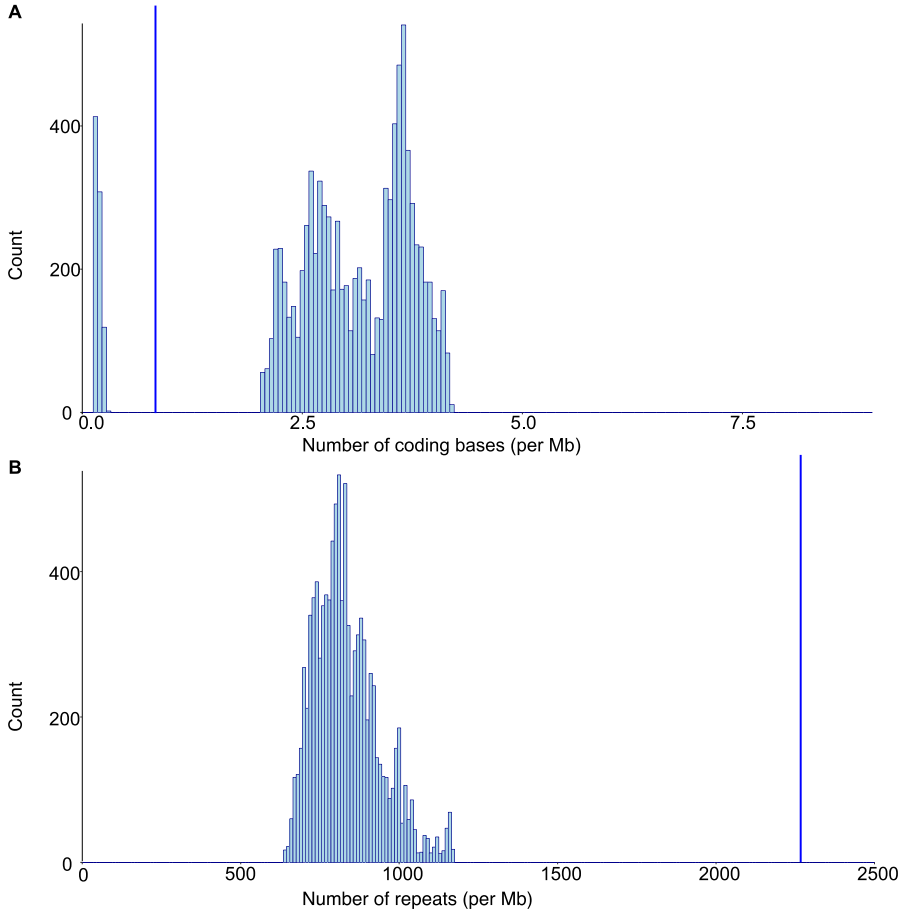


Figure S3.7. Newly closed gap sequences are enriched for repeats and under-enriched for coding regions. 10,000 10 Mb windows were randomly drawn across the genome to generate a null background distribution. The observed density of coding sequence (A) or repeats (B) within gap sequence is indicated by a blue line.

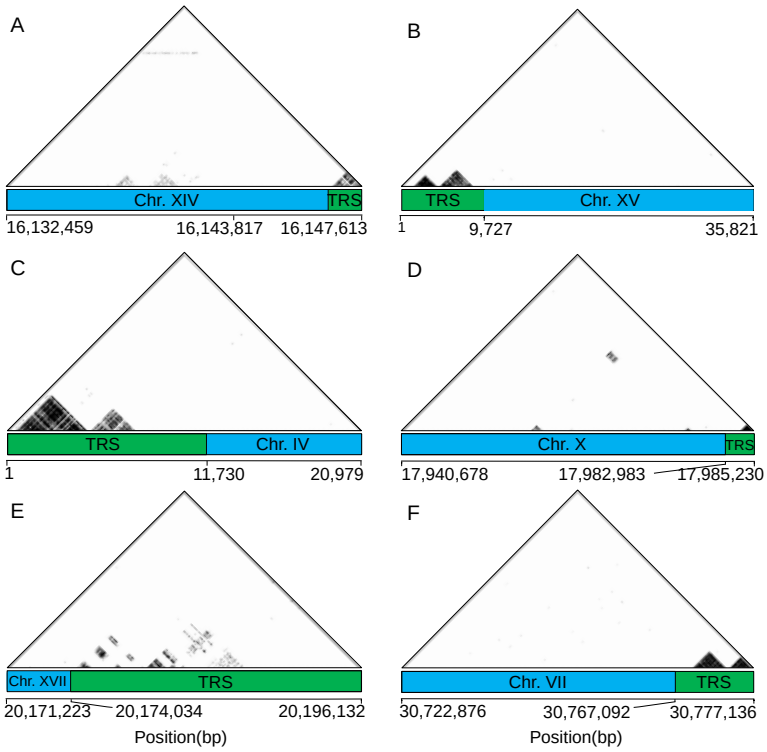


Figure S3.8. Telomeres exhibit a high density of the conserved metazoan telomere motif. Dots represent 100% sequence identity between matching windows of 15 bp. The blue box represents the end of a chromosome where the long read aligned uniquely. The green box denotes the segment rich with telomeric repeat sequence (TRS).

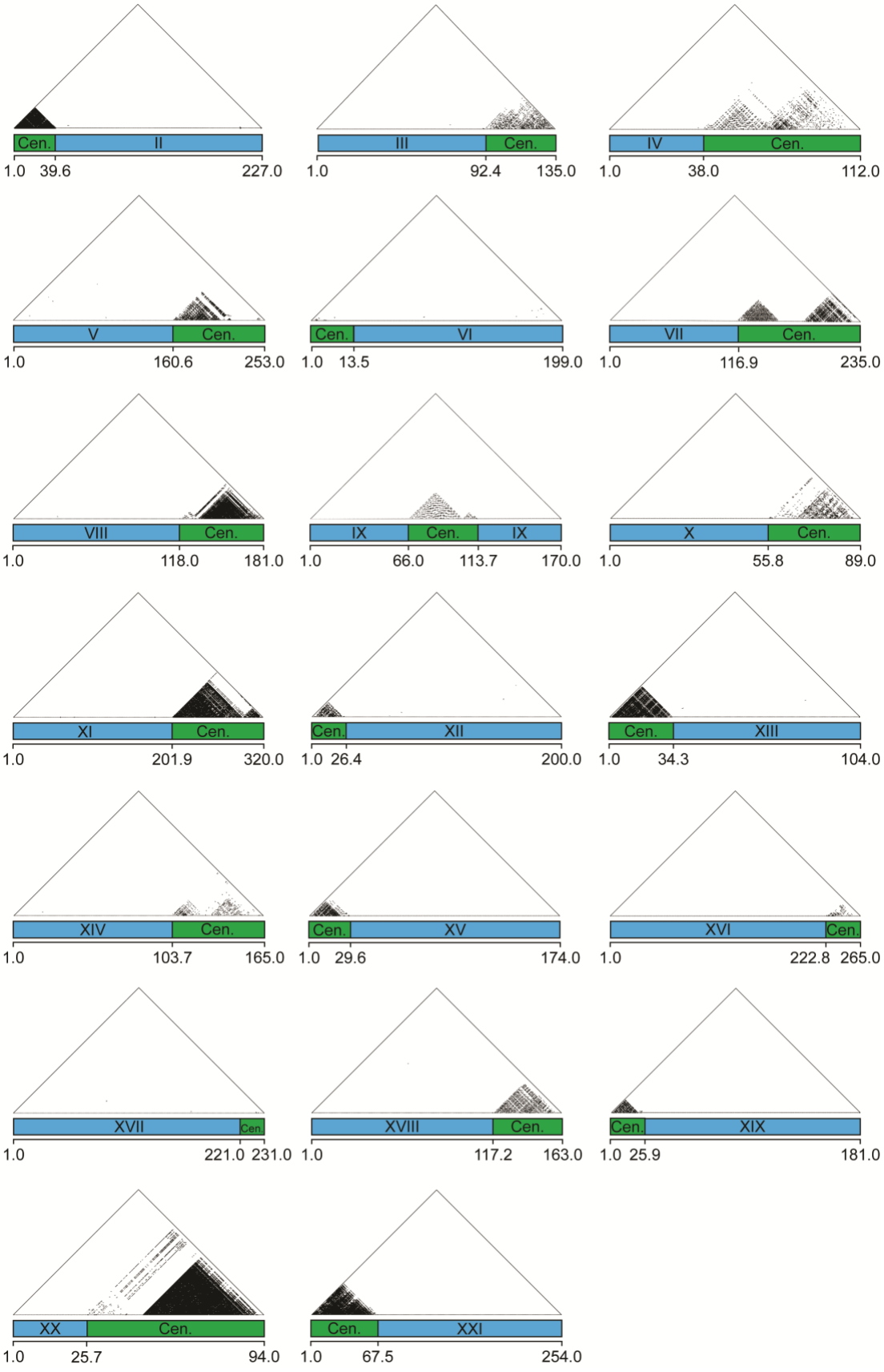

Cen. II
1.0 39.6 227.0
III Cen.
1.0 92.4 135.0
IV Cen.
1.0 38.0 112.0
V Cen.
1.0 160.6 253.0
Cen. VI
1.0 13.5 199.0
VII Cen.
1.0 116.9 235.0
VIII Cen.
1.0 118.0 181.0
IX Cen. IX
1.0 66.0 113.7 170.0
X Cen.
1.0 55.8 89.0
XI Cen.
1.0 201.9 320.0
Cen. XII
1.0 26.4 200.0
Cen. XIII
1.0 34.3 104.0
XIV Cen.
1.0 103.7 165.0
Cen. XV
1.0 29.6 174.0
XVI Cen.
1.0 222.8 265.0
XVII Cen.
1.0 221.0 231.0
XVIII Cen.
1.0 117.2 163.0
Cen. XIX
1.0 25.9 181.0
XX Cen.
1.0 25.7 94.0
Cen. XXI
1.0 67.5 254.0

Figure S3.9. Centromeres display higher order repeat structure. Sequence identity between repeats is depicted by black dots matching windows of 180 bp with 100% sequence identity. The blue region denotes the end of a chromosome with unique sequence. The green region is the newly aligned centromere contig and positions are shown in kb.

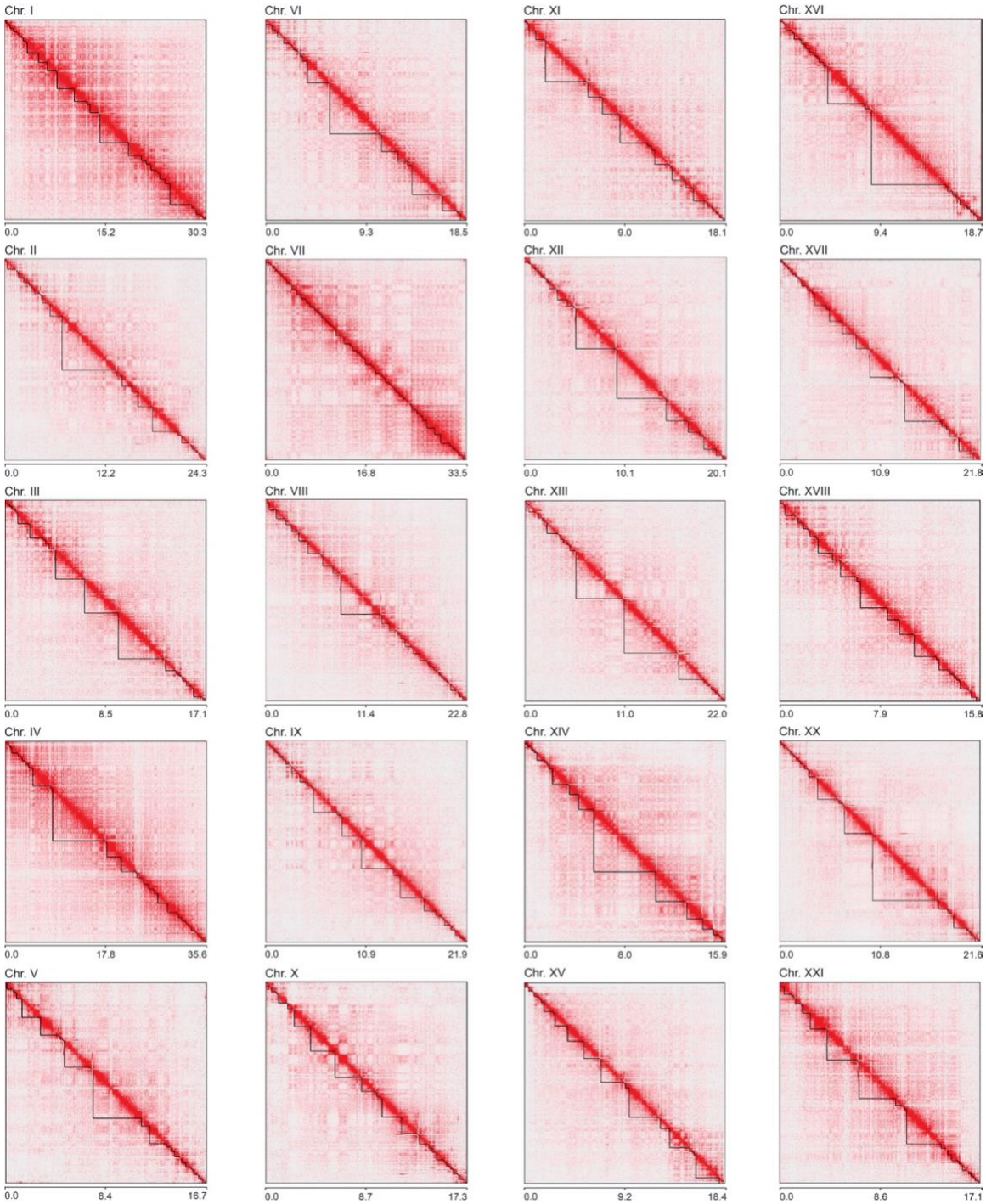


Figure S3.10. Hi-C chromosome conformation capture sequencing and proximity-guided scaffolding generated 20 main autosomal scaffolds. Contact maps for each autosome are shown. There is an enrichment of interactions between contigs that are in close proximity,

visualized along the diagonal. Contig boundaries within the scaffolds are denoted by black triangles along the diagonal.