2017

CHARACTERIZATION OF A LARGE VERTEBRATE GENOME AND HOMOMORPHIC SEX CHROMOSOMES IN THE AXOLOTL, *AMBYSTOMA MEXICANUM*

Melissa Keinath
University of Kentucky, mckein2@g.uky.edu
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.508

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Melissa Keinath, Student

Dr. Jeramiah J. Smith, Major Professor

Dr. David F. Westneat, Director of Graduate Studies

CHARACTERIZATION OF A LARGE VERTEBRATE GENOME AND HOMOMORPHIC SEX CHROMOSOMES IN THE AXOLOTL, *AMBYSTOMA MEXICANUM*

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By
Melissa Keinath

Lexington, Kentucky

Director: Dr. Jeramiah J. Smith, Associate Professor of Biology
and Dr. S. Randal Voss, Full Professor of Biology

Lexington, Kentucky

2017

ABSTRACT OF DISSERTATION

CHARACTERIZATION OF A LARGE VERTEBRATE GENOME AND HOMOMORPHIC SEX CHROMOSOMES IN THE AXOLOTL, *AMBYSTOMA MEXICANUM*

Changes in the structure, content and morphology of chromosomes accumulate over evolutionary time and contribute to cell, developmental and organismal biology. The axolotl (*Ambystoma mexicanum*) is an important model for studying these changes because: 1) it provides important phylogenetic perspective for reconstructing the evolution of vertebrate genomes and amphibian karyotypes, 2) its genome has evolved to a large size (~10X larger than human) but has maintained gene orders, and 3) it possesses potentially young sex chromosomes that have not undergone extensive differentiation in the structure that is typical of many other vertebrate sex chromosomes (e.g. mammalian XY chromosomes and avian ZW chromosomes). Early chromosomal studies were performed through cytogenetics, but more recent methods involving next generation sequencing and comparative genomics can reveal new information. Due to the large size and inherent complexity of the axolotl genome, multiple approaches are needed to cultivate the genomic and molecular resources essential for expanding its utility in modern scientific inquiries.

This dissertation describes our efforts to improve the genomic and molecular resources for the axolotl and other salamanders, with the aim of better understanding the events that have driven the evolution of vertebrate (and amphibian) chromosomes. First, I review our current state of knowledge with

respect to genome and karyotype evolution in the amphibians, present a case for studying sex chromosome evolution in the axolotl, and discuss solutions for performing analyses of large vertebrate genomes. In the second chapter, I present a study that resulted in the optimization of methods for the capture and sequencing of individual chromosomes and demonstrate the utility of the approach in improving the existing *Ambystoma* linkage map and generating targeted assemblies of individual chromosomes. In the third chapter, I present a published work that focuses on using this approach to characterize the two smallest chromosomes and provides an initial characterization of the huge axolotl genome. In the fourth chapter, I present another study that details the development of a dense linkage map for a newt, *Notophthalmus viridescens*, and its use in comparative analyses, including the discovery of a specific chromosomal fusion event in *Ambystoma* at the site of a major effect quantitative trait locus for metamorphic timing. I then describe the characterization of the relatively undifferentiated axolotl sex chromosomes, identification of a tiny sex-specific (W-linked) region, and a strong candidate for the axolotl sex-determining gene. Finally, I provide a brief discussion that recapitulates the main findings of each study, their utility in current studies, and future research directions.

The research in this dissertation has enriched this important model with genomic and molecular resources that enhance its use in modern scientific research. The information provided from evolutionary studies in axolotl chromosomes shed critical light on vertebrate genome and chromosome evolution, specifically among amphibians, an underrepresented vertebrate clade in genomics, and in homomorphic sex chromosomes, which have been largely unstudied in amphibians.

KEYWORDS: genomics, large vertebrate genome, sex chromosome evolution, *Ambystoma mexicanum*, amphibian

Melissa Keinath

November 7, 2017

CHARACTERIZATION OF A LARGE VERTEBRATE GENOME AND HOMOMORPHIC SEX CHROMOSOMES IN THE AXOLOTL, *AMBYSTOMA MEXICANUM*

By

Melissa Keinath

Jeramiah J. Smith, Ph.D.
Co-Director of Dissertation

S. Randal Voss, Ph.D.
Co-Director of Dissertation

David F. Westneat, Ph.D.
Director of Graduate Studies

July 7, 2017

To my parents, who have supported me in every possible way.

To my sister and brother for being my best friends.

To my paternal grandparents, who are forward-thinking and wise.

In memory of my maternal grandparents, who I think about every single day.

ACKNOWLEDGEMENTS

This dissertation would never have reached its final form had it not been for the support, assistance, and input from many humans, axolotls, and newts. I must thank my primary advisor, Jeramiah Smith, for his endless commitment to making me a better scientist and the immense amount of time he spent with me in all areas significant for scientific development (e.g. discussing ideas, formulating hypotheses, experimenting, analyzing data, writing, drinking beer/bourbon, etc.) It was impossible to fill my brain up with the amount of knowledge maybe only his brain can hold, but he certainly tried. Next, I need to thank my co-advisor, Randal Voss, for being an incredible educator and scientist, who provided critical insight in all of my projects, funded most of my work, and improved my writing. I must also thank the rest of my committee –Vincent Cassone and Mike Fried –for finding time to attend multiple meetings and talks for me, giving me reasons to think outside my very specific area of interest and guiding my development professionally and personally. I could listen to stories from Mike all day, and I'd live in Vinnie's kitchen if I could.

I want to specifically thank those members of the Smith and Voss labs that have been crucial to my graduate years in one way or many. Particularly the brilliant and talented postdocs, Nataliya and Vlad, without whom, I would certainly still be in graduate school. In the order in which you came into my life, thanks to the rest of the labs: Tyler, Carlena (and her smiliest baby, Zoë), Sarah, Joseph, Nour, Alex, Sri, Kevin, Will, Stephanie, Kalen, Zach, Patrick, Aum, Ryan, Varun, Cody and Courtney. You all know your important roles in my graduate years, and I won't soon forget your contributions. There is not a chance I can name all of the PIs, postdocs, lab technicians, graduate and undergraduate students that have in some way impacted my graduate years, but I should list at least several of the most significant of those relationships: Charlie, Shishir, Scott, Paul, Rose, Ashley, Megan, Wen, Lakshmi, Chelsea, Jacqueline, Cagney, Tom, Gillian, Jen, Devin, Jiffin, Taylor, Danielle, Robin, Shreyas, Kim, Sruthi, Sandeep, Jim, Brittany, Daip, and Swagata. Thanks to Lina and Chris for making my

course and conference experience far better than average. I can't thank Jacquie and Bev enough for their endless support and effort that we don't always deserve. Thanks to Seth for a million and one computer/technology favors. Thanks to Chris and Laura Muzinic for letting me leave the colony to pursue a Ph.D., providing animals/embryos when I needed them, and a special thanks to Laura for helping so much near the end of my graduate years on a side project I care about and swear will finish eventually.

Thanks to those people who collaborated with us on all of the projects I was fortunate enough to play some role, especially Dan and Oleg. I fondly remember Takis, and his willingness to provide funding for my work with newts. I've been honored to be the recipient of various awards and grants that have enabled me to attend courses and conferences, so I'd like to thank the following groups for providing those opportunities for me: Okinawa Institute of Science and Technology, University of Kentucky Department of Biology, the Gertrude F. Ribble fund, University of Kentucky Graduate School, University of Kentucky Woman's Club, Cold Spring Harbor Laboratories, Museum of Natural History, and American Genetic Association. I also want to thank Rubicon Genomics and Hudson Alpha Institute for Biotechnology Genomic Services Laboratory for their helpful insight toward various genomic approaches.

Most of the work in this dissertation was funded by grants from the National Institutes of Health (NIH) (R24OD010435 and EY10540) and Department of Defense (DOD) (W911NF1110475). All of the axolotls and embryos were provided by the Ambystoma Genetic Stock Center, which is currently funded by the NIH (P40OD019794) and previously by the National Science Foundation (NSF) (DBI-0951484).

Finally, thanks to my closest friends and the rest of my family, who are not in academia! I appreciate Laurie, Mallori, Julia (AKA Fuchs) and her sweet kleiner Fuchs, Chloe. In addition to the family I dedicated this dissertation to, I want to thank my Aunt Carol, Uncle Richard, Daniel, Uncle Steve, Kennetha, Stephen, Whitney, Braden, Ansley, Mallory, Dre, Asher, and Tante Carol for their continued support and love throughout the years. An meine deutsche Familie,

vielen Dank, dass sie mich besuchen lassen. Ich vermisse dich, Ingo, Hildegard, Marvin, Gerrit, Karl und Gretel.

I’m sure I’ve forgotten one or more humans, but how many of them read the acknowledgements anyway?

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

GENOME, KARYOTYPE AND SEX CHROMOSOME EVOLUTION IN AMPHIBIANS: CHALLENGES AND OPPORTUNITIES FOR THE FUTURE

Genome and karyotype evolution, an amphibian perspective

More than 7,000 of the estimated 66,000 vertebrate species are amphibians (The World Conservation Union 2014). New taxa of frogs, toads, salamanders and caecilians are reported every year, suggesting far more await discovery (Köhler et al. 2005; Koepfli et al. 2015). Compared to other vertebrates, the number of described species from the class Amphibia have risen immensely, with as many as 1,500 in the last 12 years (AmphibiaWeb 2017). This group of ectothermic tetrapods is comprised of three distinct clades: Anura (frogs and toads), Gymnophiona (caecilians) and Caudata (salamanders). Members of these clades have rich biodiversity with nearly 6,000 anurans, almost 600 salamanders and just under 200 caecilians (AmphibiaWeb 2017). Appearance alone provides a glimpse at the impressive diversity that exists among these orders, the worm-like bodies of caecilians, the long hind legs of the tailless frogs, and the lizard-like bodies of the salamanders. With the exception of marine environments, amphibian habitats are found all over the world and vary immensely from species to species. In addition to variability in anatomy, habitat and life history traits, amphibians have remarkably diverse genome size and karyotype.

The genome sizes of amphibians are extremely variable, ranging from ~930Mb in the ornate burrowing frog to ~117Gb in a salamander, the Neuse River waterdog (AmphibiaWeb 2017). The complexities of these genomes differ as well. Besides a slightly larger genome than that of the Western clawed frog (*Xenopus tropicalis*), the Tibetan Plateau frog (*Nanorana parkeri*) shows a different distribution and frequency of transposable elements (Sun et al. 2015). In addition, a potentially amphibian-specific region was identified that contained just over 200 genes (Sun et al. 2015). Compared to the genomes of frogs and caecilians, all salamander genomes are greatly enlarged due to an expansion of long terminal repeat retrotransposons at the base of the salamander clade (Sun et al. 2012). Salamander genomes contain a relatively high transposable element

(TE) content, including several elements that seem to have undergone recent (or continuing) proliferation (Sun et al. 2012). While salamander genomes display relatively high levels of repetition compared to mammals and birds, the repetitive portion of the genome has continued to evolve over the 200 million years since the expansions that occurred in the basal salamander lineage, and much of it is now effectively single copy (Keinath et al. 2015). Furthermore, a slower rate of DNA loss has been shown in some salamanders compared to other vertebrates, adding to the preservation of a large genome size in salamanders (Sun and Mueller 2014).

The genomes of amphibians are organized into variable numbers of chromosomes, making up the karyotype. Cytogenetic and molecular studies reveal amphibian chromosome numbers differ from fewer 9 pairs in a frog, *Dendrobates truncates* and 10 pairs in *Xenopus* to over 50 pairs with variation within and among closely related species (Green and Sessions 1991). All 4 types of chromosomes, metacentric, submetacentric, acrocentric and telocentric are present as well as both macro- and micro-chromosomes in these karyotypes (Sessions 2008). The variation in chromosome number and type can lead to speciation, macroevolution or divergent adaptation and affects recombination and segregation, among other processes (Nachman and Searle 1995; Guerrero and Kirkpatrick 2014; Pennell et al. 2015). The main avenues by which chromosome number changes during evolution are through chromosomal fusion and fission events. A fusion occurs when two acrocentric chromosomes come together through reciprocal translocation, and the result is a reduction in the number of chromosomes in the karyotype (White 1973). Conversely, the splitting of a metacentric chromosome into 2 chromosomes through fission increases the number of chromosomes in the karyotype (Schubert et al. 1995).

Comparative studies have revealed that these chromosomes are made up of blocks of conserved genes, or syntenic regions that are rearranged throughout evolutionary time (Nadeau and Taylor 1984; Pevzner and Tesler 2003). Some regions show highly conserved gene orders, and in eukaryotes, these genes are often related in terms of transcriptional control, may be functionally related, or

may have short intergenic regions (Cohen et al. 2000; Lercher et al. 2002; Hurst et al. 2004; Davila Lopez et al. 2010). Other regions may show significant rearrangement as a result of recombinational events. Comparative gene mapping in a salamander, the Mexican axolotl, (*Ambystoma mexicanum*) with the chicken genome revealed extensive conservation of gene orders (Smith and Voss 2006). Despite having only 14 haploid chromosomes and a genome more than 30 times the size of the chicken genome, orthologs in axolotl shows segmental homology with the microchromosomes of the chicken (Smith and Voss 2006). The axolotl shows high levels of conservation with other vertebrates, which provides insight for inferring the content and structure of ancestral chromosomes (Voss et al. 2011). Comparative genomics can also reveal these signatures (Pevzner and Tesler 2003; Krzywinski et al. 2009; Fishman et al. 2014). Orthologous genes in frog genomes have shown high conservation with those syntenic regions of other vertebrates, including human (Hellsten et al. 2010; Blitz 2012; Uno et al. 2013; Sun et al. 2015; Palomar et al. 2017).

Conserved synteny or breaks in conserved synteny can identify fission and fusion events, which help reconstruct ancestral karyotypes and resolve phylogenetic relationships (Maguire et al. 2014; Smith and Keinath 2015). While the evolutionary forces that enable chromosomal fission and fusion events remain largely unknown, drift, selection for recombination rate, meiotic drive and kinetochore reproduction may be drivers for these fission events (Kolnicki 2000; 2000; Pardo-Manuel de Villena and Sapienza 2001; Guerrero and Kirkpatrick 2014). As some of the most rapidly evolving parts of the genome, sex chromosomes have been used to help understand the forces that allow for the establishment of chromosomal fusions (Pennell et al. 2015).

Salamanders are important models for evolutionary studies on genomes and sex chromosomes

With the exception of the lungfish, salamanders have the largest genomes of all vertebrates, which reflect how their genomes evolved over the last 400 MY, but why they have remained large has been debated. Before advances in

sequencing technologies, genome size was linked to phenotypic features of salamanders, such as 1) rate of embryonic development (Gregory 2002) 2) complexity of brain morphology, specifically morphology of visual centers of the brain (Roth et al. 1994) 3) changes in their erythrocytes, which impacts morphological details of the nervous and visual systems in attenuated salamanders (Villolobos et al. 1988; Mueller et al. 2008) 4) rate or manifestation of metamorphosis (Gregory 2002) 5) increased nucleus and cell size (Cavalier-Smith 1978; Sessions and Larson 1987) 6) slower rates of cell division and differentiation (Sun and Mueller 2014) and 7) limitations on the rate of regeneration (Sessions and Larson 1987). Many studies of large genomes have been done in angiosperms, with genomes rivaling some salamander genomes, but far fewer studies have analyzed large vertebrate genomes in much detail. Variation in genome size within plants with similarly large genomes has been correlated with life history, geography, and ecology, but no definitive conclusions (Grime and Mowforth 1982).

It has been shown in many multicellular eukaryotes that a small population size can cause weak purifying selection (Koonin 2009). If natural selection purges deleterious mutations less effectively in organisms with smaller effective population sizes, genomes can potentially accumulate DNA, grow larger and eventually become more complex (Lynch and Conery 2003; Lynch 2007; Koonin 2009). This suggests salamanders may have historically had relatively small effective population sizes, but studies on population size in salamanders and estimates for extant salamander species contradict that idea (Frankham 2007; Organ and Shedlock 2009; Sun et al. 2012). Still the forces that cause genome expansion are poorly understood, and large vertebrate genome studies are needed to elucidate the details. With more than 600 diverse salamander genomes, the clade of salamanders seems to be an obvious choice for divulging causes and effects of genome expansion and better complete the evolutionary history of vertebrate genomes.

Without the inclusion of salamander genomes, the study of vertebrate genome evolution not only lacks an entirely unique genomic landscape of the

salamander but also an entire branch of tetrapod animals. Amphibians are the sister species to amniotes and offer crucial perspective on the transition from water to land. Karyotypic studies have highlighted the exceptional diversity of chromosomes within many salamander genomes; genomic studies are necessary to elucidate other key features, such as evolution and functional bases of genetic sex determination.

Chromosomal sex determination has arisen independently many times throughout the tree of life (Bull 1983; Bachtrog 2006; Cortez et al. 2014). In most species, the chromosomes that determine sex are the most rapidly evolving in the genome (Bachtrog et al. 2014; Beukeboom and Perrin 2014). Genomic studies on sex chromosome have been performed on all major branches of the vertebrate lineage with a major deficit among amphibians (Waters et al. 2007; Kikuchi and Hamaguchi 2013; Zhou et al. 2014; Chalopin et al. 2015). While genomic studies of several frog sex chromosomes exist (reviewed in (Malcom et al. 2014)), the genomic information for sex chromosomes in salamanders is missing completely. Like most attributes about salamanders, sex chromosomes differ incredibly among species.

Genetic sex determination has been accepted for all known salamander species that show 50/50 male/female ratio, however, few species have been investigated for sex determination from a single mendelian factor (2014). Of the known sex chromosomes of salamanders, both male heterogamety (XY) and female heterogamety (ZW) exist. Some of these sex chromosomes display homomorphy, or morphologically indistinguishable chromosomes, while others exhibit differentiation with variability in the size difference between the pair (Sessions 2008; 2014). Phylogenetic analyses indicate female heterogamety as the likely ancestral state for salamanders and the rest of amphibians, but male heterogamety has evolved independently multiple times and can vary even among closely related species (Hillis and Green 1990; Green and Sessions 1991; Sessions 2008). The large size of salamander chromosomes lend themselves to insightful cytogenetic techniques, such as lampbrush chromosomes from female oocytes as well as pairing arrangements in male

meiosis, which can elucidate fine structure differences between homomorphic sex chromosomes (Green and Sessions 1991). Characterizing homomorphic sex chromosomes may deliver valuable insight into the early stages of heteromorphic sex chromosome evolution or provide support for the fountain-of-youth or high-turnover hypotheses, surrounding sex chromosomes that will remain homomorphic throughout evolutionary time.

As described below, homomorphic chromosomes present particularly interesting targets for study because they can potentially shed light on early evolutionary processes that drive extreme differentiation in sex chromosomes. The axolotl meets all of the conditions necessary for studying homomorphic sex chromosomes, having sex chromosomes with no visible differeniation and many biological replicates for future comparative studies. Incorporating genomic sex chromosome studies from the axolotl will not only enrich sex chromsome evolutionary theory but also set the foundation for similar studies in large vertebrate genomes and provide new amphibian resources for comparative tools.

Evolution of sex chromosomes

Sex chromosomes originate from a normal pair of autosomes when a mutation arises that becomes a critical sex-determining factor. Sexually antagonistic alleles, which are beneficial for one sex and not for the other, favor the suppression of recombination between the new sex chromosomes. Over time, the non-recombining portion of the sex chromosomes (Y or W) will degrade and shrink, making the sex chromosomes appear morphologically different, forming heteromorphic sex chromosomes. (Bull 1983; Rice 1984)

Whereas many sex chromosomes become highly differentiated over time, several hypotheses suggest that some sex chromosomes are homomorphic throughout evolutionary time and are not on a road to heteromorphy (reviewed in (Stock et al. 2013; Bachtrog et al. 2014)). An early hypothesis proposed was that lack of significant sexually antagonistic selection may not send homomorphic sex chromosomes down the path to become heteromorphic (Rice 1987). In some cases, such as that of the emu, sex-biased gene expression of sex-linked genes

may minimize the deleterious effects of sexually antagonistic alleles and preserve a less differentiated appearance (Vicoso et al. 2013b). The fountain-of-youth hypothesis proposes a mechanism where by rare sex-reversed animals allow for recombination between the X and Y or Z and W chromosomes (Perrin 2009). Evidence for this hypothesis is supported by several frog species that can experience sex reversal early in development by way of temperature even when the sex chromosomes are characteristic of the opposite sex (Stock et al. 2011b). Finally, a high-turnover hypothesis proposes that lack of decay between chromosomes is a direct result of regular replacement of sex chromosomes by autosomes that develop sex-determining mutations (Schartl 2004; Stock et al. 2013). Studies on turnovers and transitions in many fishes and some amphibians have provided evidence for this hypothesis (Miura 2007; Tanaka et al. 2007; Stock et al. 2011a; Kitano and Peichel 2012; Yoshida et al. 2014). Regardless of whether homomorphic sex chromosomes are young or old, they provide important perspective on sex chromosome evolutionary theory.

Ambystoma sex chromosomes

The axolotl belongs to a group of closely related tiger salamander species. This tiger salamander complex comprises 8 major clades with 76 nemed species in the US and Mexico (Shaffer and McKnight 1996). These salamanders show incredible diversity of life history traits, including metamorphic and paedomorphic (i.e. non-metamorphic) life histories (Collins et al. 1980; Shaffer 1984b; Voss and Shaffer 1996; Voss et al. 2012). Gene trees for the complex suggest a recent history of rapid bursts of speciation, and geological evidence points the Sierran uplift and subsequent drying of North American deserts ~5 Myathat first isolated *Ambystoma californiense* and other species, marking the oldest speciation event in the complex (Axelrod 1980; Unruh 1991; Shaffer and McKnight 1996; O'Neill et al. 2013). Single locus gene tree studies and genealogical tests of species boundaries have provided evidence that despite the divergent life history adaptations and morphologies among these species, there is surprisingly low variation within the mitochondrial D-loop sequence, and the mtDNA haplotypes

are minimally diverged from one another (Martin et al. 1992; Rand et al. 1994; Shaffer and McKnight 1996; McKnight and Shaffer 1997). Even with some variation, we expect the sex locus to track the mtDNA in these species due to maternal inheritance, adding to the benefits associated with using this complex to study early sex chromosome evolution comparatively, which means the salamander complex provides a host of biological replicates.

Many experiments revealed female heterogamety in the axolotl (Humphrey 1948; Humphrey and Frankhauser 1957; Armstrong 1984). By grafting migrating primordial germ cells from one embryo to another, Humphrey surgically sex-reversed axolotls who retained their original DNA but developed physically into the opposite sex (Humphrey 1945). Sex-reversed females animals (ZW) were crossed with normal females (ZW) in order to determine the genetic sex-determining mechanism (Humphrey 1945). If the axolotl had male heterogamety, then the sex-reversed female (XX) crossed to a normal female (XX) would produce all females, but with female heterogamety, the offspring would produce 75% female and 25% male. Later mapping studies used this result to measure the distance of genes from centromeres (Lindsley et al. 1956). These techniques were used to imprecisely place the sex locus distal to the centromere of a chromosome (i.e. completely separated by recombination), near terminal end of a chromosomal arm (Armstrong 1984). Comparative cytogenetic studies analyzing banding pattern differences between *A. mexicanum* (the axolotl) and *A. tigrinum* (the tiger salamander) revealed no definitive differences between most chromosomes, however, a small terminal deletion was identified a chromosome that may correspond to the presumptive sex chromosome for *A. mexicanum* that is reported as part of this thesis [53-55].

Linkage analyses together with genetic association studies performed by Smith, *et al*. in 2009 identified a marker (*E24C3*) that was associated with segregation of the sex phenotype in the axolotl genome (Smith and Voss 2009). Using a backcross design to mate female *A. mexicanum*/*A. tigrinum* hybrids with male *A. mexicanum*, genetic screens were performed for sex-associated regions. The marker localized the sex locus to the tip of *Ambystoma* LG5 (Smith and Voss

2009), which was later reassigned to LG9 (Chapter 5). In addition, they found no evidence for different recombination frequencies between the sexes suggestive of recent evolution of sex chromosomes, however, these studies did not sample markers in close proximity to the sex locus. (Smith and Voss 2009)

Coverage analyses (presented in Chapter 5) have now identified a region of W-specific sequence in the axolotl. By pooling DNA sequence from 22 females and from 26 males and aligning the reads to a draft *Ambystoma* genome, read depth of coverage could be assessed across each scaffold. Sex-specific candidate regions had nearly zero coverage in males and about half normal coverage in females, as those W-specific sequences are only present on one sex chromosome. Each candidate region was PCR validated, and a total of 36 out of 154 were found to be sex-specific. Future analyses will be aimed at lengthening the sequence known for the *ATRW* exon, assessing expression levels during gonadogenesis through RT-PCR in developing females and verifying sex-specificity in other tiger salamander species.

Challenges and solutions

In order to improve evolutionary studies of vertebrate genomes and sex chromosomes, it is imperative that genomic resources for salamanders are improved. The genomics revolution has led to the release of hundreds of published genomes for vertebrates, including many mammals, fish, birds and reptiles, however, only 3 amphibian genome assemblies exist today, and they are all frogs, two of which are closely related: *Xenopus tropicalis(Hellsten et al. 2010)*, *Xenopus laevis* (Session et al. 2016)and *Nanorana parkeri(Sun et al. 2015)*. No representatives from the salamander clade have been assembled due to the significant challenge their large and complex genomes present, with the exception of the assembly resulting in part from this thesis, which was released in July 2017 (http://ambystoma.org/). Accordingly, until now genomic sex chromosome studies have not been performed.

Advances in sequencing technology have been vital for the growth in genomic studies, providing solutions for speed, reliability and affordability of

sequencing challenging genomes. What began in the mid 1970's with Sanger sequencing of tiny virus genomes (Sanger and Coulson 1975; Sanger et al. 1978) matured to the point of sequencing the human genome by 2001 (Venter et al. 2001), setting the foundation for all future genomic studies. As sequencing technologies changed and output increased exponentially, analytical tools required for assembly improved concordantly. A shift to Illumina short read chemistries (Bentley et al. 2008) from the longer Sanger sequences changed the assembly process from overlap consensus methods (Sanger et al. 1978) to center around de Bruijn graph assemblers, which are still widely used today (Pevzner et al. 2001). Improved scaffolding can be achieved through the use of newer technologies offered on a variety of platforms, including, paired-end reads (reads from two ends of the same DNA molecule with some distance between them), optical mapping (long-range restriction mapping), physical mapping (relating genomic positions with physical distances), and proximity ligation (crosslinking sequences that are physically proximal) (Phillippy 2017). The newest methods by Pacific Biosciences use parallelized single molecule DNA sequencing to produce longer reads (Eid et al. 2009), and Oxford Nanopore, which identifies nucleotides via electrical conductivity as it passes through a biological pore (Lu et al. 2016).

Although long reads, like those from Pacific Biosciences, can help provide more contiguous assembly, they are expensive and carry a higher error rate (Levy and Myers 2016). Shorter reads come with a lower error rate (Liu et al. 2012) and are more affordable, but the associated downstream assemblies often contain more gaps, are biased due to GC content and miss structural variation and repeats (Baker 2012; Bradnam et al. 2013). Current methods for assembly of complex genomes include combinations of short and long read data. In addition to the limitations of sequencing, analytical methods carry a computational burden, cost time, and in some cases, a large memory overhead that grows with more genomic data (Goodwin et al. 2016). These challenges are present for all vertebrate genome assemblies, however, due to the size and complexity of the

salamander genome, the issues become much more severe, often requiring huge allocation of memory, processors and time (Keinath et al. 2015).

Next generation sequencing initiatives now exist to help fill in the amphibian (with special attention to the salamander) gap among published vertebrate genome assemblies. The Genome 10K is project with a goal of sequencing the genomes of 10,000 vertebrates, with at least one individual from each vertebrate (Koepfli et al. 2015). In addition, the Amphibian Survival Alliance and Amphibian Specialist Group has a formed an initiative to provide assemblies from every family of amphibians (Amphibian Survival Alliance and Amphibian Specialist Group 2014).

Substantial progress has been made with respect to the development of molecular resources and tools for urodeles, particularly the newt, *Notophthalmus viridescens*, and the Mexican axolotl, *Ambystoma mexicanum*. There are presently proteomes (Rao et al. 2009; Bruckskotten et al. 2012; Abdullayev et al. 2013), transcriptomes (Putta et al. 2004; Abdullayev et al. 2013; Looso et al. 2013; Bryant et al. 2017), linkage maps (Smith et al. 2005a; Keinath et al. 2017), and many, many gene expression studies for the species. In the axolotl, a major effect QTL was discovered for metamorphic timing (Page et al. 2013) multiple bacterial artificial chromosome (BAC) libraries have been generated (Habermann et al. 2004; Putta et al. 2004), knockouts are possible (Fei et al. 2014) and gene functions can be manipulated through transgenics (Sobkow et al. 2006; Khattak et al. 2009; Khattak et al. 2013). Recently our group has made a fragmentary axolotl draft genome assembly publicly available (http://ambystoma.org/), however, no assembly is yet published (Smith et al. 2005b; Keinath et al. 2015).

Conclusion

Multiple approaches are necessary to cultivate the resources necessary for studying the axolotl. The following chapters will highlight some of our efforts to improve the genomic and molecular resources for salamanders, including a development of methods for laser-capture chromosome sequencing, individual chromosome assembly, a dense linkage map for the newt, and a draft genome

assembly for the axolotl. Using the resources we developed, we performed comparative genomic studies between newt, axolotl and *Xenopus* and characterized several axolotl chromosomes, including the homomorphic sex chromosomes, marking the first genomic study to identify sex-specific sequences in the axolotl.

CHAPTER TWO

DEVELOPMENT OF METHODS FOR LASER CAPTURE CHROMOSOME SEQUENCING IN AXOLOTL

Abstract

The Mexican axolotl (*Ambystoma mexicanum*) is a member of the amphibian lineage, which diverged from all other tetrapods approximately 300 million years ago. Like other Urodele amphibians, the axolotl possesses a massive genome (~32Gb). Despite its size, the genome consists of only 14 pairs of chromosomes, and its gene orders are highly conserved with reptilian and mammalian genomes. As such, the axolotl has served as an important model organism for studying evolution of vertebrate genomes, particularly with respect to the changes in gene order, chromosome structure and genome size.

Due to current limitations on assembling large genomes with many repeats and lengthy introns, an alternative approach was developed in order to better resolve the structure and content of the genome; sequencing one chromosome at a time (Smith et al. 2009). This approach involves the isolation of single dyads from axolotl chromosome spreads via laser capture microdissection (LCM). Captured dyads are then amplified and used to prepare sequencing libraries (Illumina HiSeq). The initial experiments largely targeted a single, morphologically distinct chromosome (chromosome 3). The resulting reads from this pilot study were aligned to a set of genes that were previously placed on the *Ambystoma* linkage map, and the resulting alignments revealed a strong enrichment for genes on linkage group 3 (LG3), covering more than 80% (75/92) of the markers on that LG. Sequence data from other individual chromosomes (dyads) either mapped to entire linkage groups or provided evidence for the splitting or merging of parts of linkage groups, both validating and improving the linkage map

These findings indicate that our LCM based approach yields robustly targeted sequencing, which builds on our existing linkage map and provides scaffolding data for our genome assembly. The results from this pilot study demonstrate that laser-capture sequencing is a useful approach that, in

combination with whole genome sequencing, should aid in the development of a contiguous genome assembly for *A. mexicanum*.

Introduction

The generation of a genome assembly is critical for expanding the utility of an organism to modern scientific inquiries. Improvements in sequencing technology and analytical tools have made it possible to produce genome assemblies and perform large-scale genomic studies on non-model organisms (Ellegren 2014; da Fonseca et al. 2016). Despite the extraordinary progress made in these technologies, building assemblies for large genomes remains challenging (Sun and Mueller 2014; Keinath et al. 2015; Geng et al. 2017). The Mexican Axolotl, *Ambystoma mexicanum*, is a salamander with a gigantic genome (~10X the size of the human genome), and it is also a species highly studied for development and their robust ability to regenerate complex structures, including limbs, tail, spinal cord, lens and parts of major organ systems (Carlson 1970; Voss et al. 2009; Ferris et al. 2010; Voss et al. 2013a). Regenerative studies illuminate important factors and details pertaining to the regrowth of a variety of tissues and have significant implications for regenerative medicine including, tissue engineering, regenerative cell therapy, wound healing and perhaps ultimately, regrowth of whole limbs/organs for humans (Putta et al. 2004; Brockes and Kumar 2005; Kragl and Tanaka 2009b; Kragl and Tanaka 2009a; Godwin et al. 2013; McCusker and Gardiner 2014). Researchers in regeneration hope to find a genomic basis for this, and the availability of a genome assembly should aid these studies by facilitating functional genomic approaches, such as genome editing.

The size of the axolotl genome ~32Gb, but it is not remarkable among salamanders (Licht and Lowcock 1991). It is thought that a repeat expansion 200 MY ago left all salamander genomes greatly expanded compared to other extant tetrapods, with sizes ranging from 14Gb to ~120Gb (Gregory et al. 2007; Smith et al. 2009; Zhang and Wake 2009). While salamander genomes display relatively high levels of repetition compared to mammals and birds, the repetitive

portion from the original expansion has had 200 MY to evolve, and much of it is effectively single copy (Chapter 3) (Keinath et al. 2015). Despite the enormous size of the salamander genome, the 14 pairs of chromosomes (2N=28) (Fankhauser and Humphrey 1942) show high conservation with other vertebrate lineages, retaining large syntenic blocks from the ancestral vertebrate karytoype (Voss et al. 2011). The repetitive portion of the genome accounts for approximately one third of the genome size and is broadly distributed throughout the genome (Morescalchi and Serra 1974; Keinath et al. 2015). Axolotl genes are predicted to be about 5 times as large as human genes with long introns, ~10 times the size of orthologous vertebrate introns (Smith et al. 2009).

Our initial attempts to sequence and assemble whole genome shotgun data recurrently failed because of insufficient memory despite the availability of 1 terabyte of RAM for assembly calculations (Keinath et al. 2015). In order to reduce the computational burden, we took advantage of the fact that the genome is naturally packaged into 14 smaller (mammal genome-sized) compartments: chromosomes. I therefore strived to develop a method to assemble chromosomes one by one. Chromosome 3 was chosen for the majority of samples sequenced in the laser capture chromosome study, as it is the easiest to identify in a spread of chromosomes. It has greater arm asymmetry with a ratio of about 5:3 and a constriction subterminally in the short (q) arm, denoting the location of the single nucleolar organizer region (NOR) (Callan 1966). Polymorphisms in the NOR were previously reported to be associated with inheritance of the *white* mutation and recent studies have demonstrated that a mutation in the *endothelin 3* (*edn3*), likely underlies the white phenotype (Woodcock et al. 2017).

This article details my efforts to develop methods to sequence and assemble individual chromosomes, including 3 different sampling strategies. These suggest that the best strategy is the third method using 1.0mm PEN membrane slides and PicoPlex DNAseq kit for amplification. In addition, I discuss the impacts that these chromosomal libraries have had toward the development of molecular and genomic resources for the axolotl.

Results & Discussion

Initial attempts to amplify and sequence chromosome 3

Initial attempts to sequence material from whole genome amplified (WGA) reactions yielded no sequence information that could be attributed to axolotl transcripts or genomic DNA sequence. In an attempt to identify the underlying cause of these sequencing failures, we examined individual images from Illumina sequencing runs. During each extension cycle, an image is generated as a fluorescently labeled nucleotide is being incorporated to the cluster being synthesized. Signal intensity measurements identify the nucleotide, and specific parameters related to nucleotide diversity (at least one G or A and at least one C or T) are required in the first 7 cycles for the generation of reads. Figure 2.1A shows the fluorescence detected on an Illumina flow cell during the first attempt of sequencing. In the flow cell image, a single cluster will appear as a bright spot. The scant fluorescence seen in this image depicts the incorporation of C-bases into of DNA molecules from phiX, a control sequence spiked into the reaction (Bentley et al. 2008). Whole-lane fluorescence profiles revealed heavy GT bias, which inhibited cluster identification during the initial sequencing cycles (Figure 2.1B). Reads with heavy GT bias in the first several bases were not recognized as valid extension products by the base-calling algorithm, and were therefore, not output as sequence. We presumed that this reflected failure to read “salamander” reads due to the fact that the amplification process resulted in the inclusion of a low-complexity (proprietary) random primer sequence that emitted amplified bases similar to those expected for poly-inosine.

In order to read through the presumptive WGA leader sequence, dark cycles (template synthesis with no imaging) were employed to allow polymerization for 36 cycles before cluster identification and sequence acquisition. The fluorescence image shows the presence of more clusters after the 36 dark cycles (Figure 2.2). Several caveats exist for the dark cycle solution. First, the base quality of reads generally drops toward the end of a read. Second,

the dark cycling affects all samples in a single run, and unless an entire flow cell is being utilized by a particular project, the reduced quality is less desirable.

The number of reads from these samples ranged from 7.7M to nearly 20M. To determine enrichment for markers on a specific *Ambystoma* LG, the proportions of reads aligning to each linkage group were calculated (Table 2.1). Both libraries that showed specificity to linkage groups when mapped were generated from dyads excised from thin (0.17 mm) polyethylene naphthalate (PEN) membrane slides, and not from normal (1.0 mm) glass microslides. Reads from these libraries were also mapped to human and bacterial genomes to assess potential contamination, and the percent of concordantly mapped reads to bacterial genomes were less than 2% for each sample but ranged from 21.58% to 57.91% when mapped to human (Table 2.1).

Given their initial quality metrics, samples A1 and A2 were sequenced further and the resulting reads were found to align to additional genes that have been previously mapped to LG3 and LG7, respectively. The chromosomal library for sample A1, which contained 6 chromosome (chr) 3 dyads, yielded 1737 reads that aligned to 25 LG3 markers. Further sequencing of this sample yielded 1792 additional reads that aligned to 30 total LG3 markers, covering 32.6 % of the known genes on LG3 (Table 2.1). In the library from sample A2, derived from 1 small dyad, 132 reads aligned to 15 LG7 markers. Combined with additional sequencing, a total of 272 reads aligned to 19 LG7 contigs, covering a 54.3% of the known genes on LG7 (Table 2.1).

In-line adapter ligation and comparison of PEN and PET membrane slides

As libraries generated by dyads excised from thin PEN membrane slides resulted in more specific mapping to transcripts from the *Ambystoma* linkage map, relative to normal microslides, only membrane slides were used in subsequent microdissections. Due to the fragility of the thin (0.17 mm) membrane slides, I elected to use 1.0 mm membrane slides and tested the performance of polyethylene terephthalate (PET) relative to PEN slides. In order to reduce the

contamination that was observed in the first strategy, all membrane slides were UV-treated prior to chromosome spreading and stored in a sealed desiccator.

Although dark cycling allowed for cluster identification and generation of sequence data from amplified material, the approach presented practical difficulties, as the entire flowcell must be subjected to the dark cycling. To get around this issue, the library preparation steps for these captured dyads included ligation of random hexamers to both ends of each amplicons, which provided nucleotide diversity over the first 10 cycles, enabling cluster seeding by the imaging algorithm. Due to this addition of sequence, the resulting reads were trimmed to eliminate all leader sequences, reducing the usable read lengths for the chromosomal libraries to ~60bp.

The number of reads acquired using this approach ranged from ~18.5M to ~24.2M per library (Table 2.2). The percent of reads aligning to bacteria was under 0.6% but for 1 sample with nearly 7%. The percent of human reads were high with all chromosomal libraries showing greater than 28% alignment to human. Of the 5 samples sequenced, 2 out of 5 libraries showed specificity to a single linkage group. In the chromosomal library for sample B1, containing 5 chr 3 dyads excised from a PET membrane slide, 86 reads aligned to 24 LG3 contigs (26%). In the chromosomal library for sample B2, which contained a single chr 3 dyad excised from a PEN membrane slide, 1843 reads aligned to 66 LG3 contigs, or 71% of LG3 contigs. Overall, this single chr 3 dyad excised from a normal PEN slide from this strategy yielded the best results compared to PET slides, which generally yielded smaller numbers of mapped reads (Table 2.2). Coupled with the results from dyads excised from the thin PEN membrane slides in the first strategy, normal PEN slides seem to be the best option generating amplified libraries.

Test amplifications with the PicoPlex WGA DNAseq kit

Discussion with Rubicon Genomics of previous sequencing strategies to eliminate issues preventing cluster identification from amplified libraries (dark cycles and random hexamer ligation) led to changes in the Pico Plex Whole

Genome Amplification (WGA) kit. An alpha test kit was used in this approach to eliminate the need for dark cycles or extra ligation step. Because my previous analyses had indicated that UV-treated normal PEN membrane slides yielded the best material for library preparation, these were used for all subsequent experiments.

Libraries were generated for 10 samples and sequenced on a single lane of Illumina HiSeq1500. The number of reads generated under this strategy ranged from ~15.4M to ~22.9M per sample (N = 10) (Table 2.3). The percent of reads from each axolotl chromosome library that aligned to human varied from just over 5% to <20%, with two additional libraries exceeding 50% (61.4 and 86.8%). The two libraries with excessively high numbers of human reads were generated from a single dyad that was apparently overwhelmed by contamination from human DNA via dust or mishandling of equipment or amplification kit. All samples had negligible bacterial contamination.

In axolotl, identification of individual chromosomes is challenging due to the fact that several chromosomes are similar in size and that the degree of compaction often varies within and between cells (Callan 1966) (Figure 2.3). It was therefore important to test if the generation of libraries from single dyads was a viable approach. In order to better assess the effect pooling dyads (vs individual) has on the resulting libraries, several collections were made with multiple dyads while others contained only a single dyad (Table 2.3). The libraries for samples C1 and C2 each contained ten chr 3 dyads, and yielded a total of ~15.4M and ~16.2M reads, reads for sample C1 included 1,373 reads that aligned to 74 LG3 contigs (80.4% coverage) and reads from sample C2 included 1,253 reads that aligned to 68 LG3 contigs (73.9% coverage). Another library (sample C3) was generated from five chr 3 dyads and yielded ~17.5M reads, of which 984 reads aligned to 62 LG3 contigs (67.4% coverage). To test the single dyad approach, another library (sample C4) was generated from a single chr 3 dyad and yielded ~21.9M reads with 114 of those reads aligning to 29 LG3 contigs (31.5% coverage). While the number of reads increased for this sample, there was also significant human contamination. Because the library

contained only a single dyad, I reasoned that even a small quantity of human contamination could potentially overwhelm the single dyad sample. The library for sample C8, containing 1 large dyad that did not contain the NOR. This library yielded 424 reads that aligned to 9 LG2 contigs (8.2 % coverage). The final chromosomal library for sample C9, containing a large dyad, had 84 reads with alignments to 11 contigs on the bottom half of LG4 (19% coverage) and 139 reads with alignments to 6 LG13 contigs (46.2% coverage). The human control showed 96.8% of the reads aligned to the human genome, providing evidence that the WGA kit and downstream sequencing are effective particularly when a well-developed genome is available for read mapping.

In order to better assess the extent to which our amplification approach sampled specific chromosomes, additional sequencing was performed on several samples (the two samples containing 10 chr 3 dyads and the sample containing the single dyad that aligned to LG4 and 13). The samples chosen for LG3 contained the highest proportion of coverage to the LG. Sample C9, a library generated by a single large dyad, was also selected in order to better assess the linkage for one segment of LG4 with LG13. These samples were each sequenced on a full lane of Illumina HiSeq1500. The additional sequencing of samples C1, C2 and C9 more than doubled the amount of sequence for each sample, and the proportion of contigs hit from each target LG increased correspondingly. For the first sample, the percentage of LG3 markers that were covered by sequence data rose from 80.4% to 81.5%, while the second sample coverage remained 79.3% with an increase of reads aligning to these contigs in both libraries (all 3 strategies for LG3 included in Figure 2.4). The original libraries and additional sequencing provided more data for LG3 and represent those samples with the most coverage for the LG. The proportion of reads aligning to LG4 and 13 contigs remained the same in sample C9, but the number of reads aligning increased from to 84 reads to LG4 and 139 reads to on LG13 to 92 and 162, respectively. Closer examination of the markers detected by sample C9 (a large dyad), showed sequences mapped to the genes on the lower half of LG4 (after a >30cM gap) and all of LG13 (Figure 2.4). This suggests that a

portion of LG4 (from 276.1 cM to 414.1 cM) may be linked to LG13. This idea is further reinforced by the fact that these two segments are ancestrally linked, sharing synteny with chicken chr 3 and *Xenopus* chr 9 (Voss et al. 2011).

In comparison to previous approaches, amplification using the PicoPlex WGA DNAseq kit and chromosomes dissected from 1.0mm PEN slides appeared to yield the best sequence data in terms of reducing sample contamination (Figure 2.6) and increasing both the numbers and proportions of on-target reads (Figure 2.7).

Additional chromosome sequencing projects using the PicoPlex WGA DNAseq kit

To further improve our understanding of the content of salamander chromosomes we leveraged our optimized approach to generate sequence for an additional 5 individual dyads and barcoded 24 samples that were sequenced as a pool on an Illumina HiSeq 2000 platform (Table 2.4). One small dyad sequenced mapped almost exclusively to the genes in the upper half of LG4 (Figure 2.8), above a large gap that separates the upper half of the chromosome from the lower half that had been previously identified as being linked to LG13 (Figure 2.5). Taken together, these libraries provide evidence that LG4 should be divided into 2 LGs, and that the lower half of LG4 should be linked to LG13. Six other libraries targeted smaller dyads. Two of these libraries aligned to genes on LG14 and the other 4 aligned to the genes on both LG15 and 17, revealing that the 2 LGs can be combined to make one LG for AM13. More detailed analyses of these libraries has been previously published (Keinath et al. 2015). Four other amplified dyads yielded alignments to 31.25% of the genes on LG5 (Figure 2.9), 41% of the genes on LG6 (Figure 2.10), 51% of the genes on LG10 (Figure 2.11) and 39% of the genes on LG9 (Chapter 5).

A whole additional lane of Illumina HiSeq 2000 sequencing was performed for two samples and improved our resolution of two chromosomes that have been of particular interest with respect to studies of axolotl biology (Table 2.4).

The first of these represented the chromosome corresponding to LG2 (sample D4). LG2 contains the QTL (*met*), which contributes significantly to the metamorphic timing and the expression of metamorphic vs paedomorphic life histories in hybrid crosses (Voss and Smith 2005). The resulting library yielded 16963 reads that aligned to 117 LG2 contigs, which is 80% of the genes on LG2 (Figure 2.12). Analyses of these date have been previously published (Keinath et al. 2017) and are presented in Chapter 4. An additional whole lane of sequencing was also generated for a library that aligned to LG9 (sample D2). This library yielded 68844 reads that aligned to 40 LG9 contigs, covering nearly 70% of the genes on LG9. LG9 contains the sex locus for the species and the results of additional sequencing will be discussed in Chapter 5 of this dissertation.

In every chromosomal library, we identified reads that aligned to a common list of gene markers that had been placed on separate linkage groups. The fact that these markers were identified in multiple independent libraries is considered evidence that these markers (or portions of their sequence) may not be chromosome-specific but instead are repeated throughout the genome. Our individual dyad collection also provided resolution for the linkage map and established genes known to each of the collected chromosomes, enabling the identification of specific chromosomes via FISH (unpublished, Timoshevskiy).

Comparing data from pooled vs. individual chromosomes

In our initial experiments, amplified libraries were generated from either single dyads or pools of 5-10 chromosomes. In some cases, more alignments to genes on LG3 were found in libraries generated from pooled samples; however, there are some caveats to a pooled approach. In order to capture more chromosomes, the adhesive cap used for collection of microdissected chromosomes is left open longer, exposing it to potential contamination. Additionally, the pooled sample approach is only feasible on chromosomes that can be easily identified in every spread. Unfortunately most of the mitotic chromosomes are not easily identifiable in *Ambystoma*, and they often vary in size from spread to spread. Although chr 3 is easily identifiable, Figure 2.3 reveals the stark differences in appearance this

chromosome shows in multiple spreads. While there is a smaller chance for contamination to occur in a single dyad sample, our studies suggest that a single dust particle may hold more DNA than an individual dyad, overwhelming the sample with another source (Figure 2.9).

K-mer based analyses and assembly of chromosomal libraries

To correct for possible errors that may have occurred during amplification of the captured material, reads were error corrected using whole genome shotgun data as done previously (Chapter 3). K-mer based analyses for the merged LG3 libraries show a lack of Gaussian distribution, as expected in random shotgun sequencing (Figure 2.13). Instead the distribution shows a high peak at a multiplicity of 1, where most sequencing errors will fall, and a steady drop off across all other multiplicities. One explanation for this distribution is bias in the amplification of the libraries. Similar amplification biases have been shown in another study using multiple displacement amplification (MDA) in samples compared to non-amplified sequence, resulting in a comparable k-mer distribution (Chen et al. 2014). Because amplification bias causes uneven or variable amplification in different sequences, some sequences amplify significantly more than others, and some sequence may not amplify at all. I speculate that this variation in amplification could account for the distribution in the k-mer plot, as it appears that sequences were not randomly sampled.

Attempts to assemble the chromosomal libraries from LG3 yielded fragmentary assemblies, even when all data from LG3 were merged (Keinath et al. 2015). I speculate that the fragmentary nature of these assemblies derives (partially) from amplification biases that are inherent to these amplified data. With respect to assembly algorithms, underamplified regions may not be samples sufficiently as to permit their incorporation into the assembly, whereas over-amplified regions are likely to be interpreted as repeats and not properly integrated into the assembly. Additionally, nucleotide distribution plots show GT bias in all samples, so base calling bias may contribute errors to the chromosomal libraries. Just as GC bias has been shown to reduce the accuracy

and completeness of a genome assembly, any base calling bias may show a similar effect (Chen et al. 2013).

Conclusions

Altogether the results of this study provide significant resolution to the linkage map and assign specific markers to individual chromosomes for use in future cytogenetic studies. The project optimized the approach for individual chromosomes and led to the overall improvement in the compatibility of the Rubicon PicoPlex whole genome amplification kit with the HighSeq platform. Additionally, these studies laid the foundation for future laser capture amplification projects, such as those performed on individual sperm cells of use in mapping the haplotype of a species.

Methods

General methods are described below then divided by strategy along with associated changes made to the original methods in each strategy specifically.

Preparation of cells for metaphase spreads

For each chromosome preparation, one hundred eggs were collected from a pair of wildtype axolotls crossed in the *Ambystoma* Genetic Stock Center and incubated at 18°C until stage 17 of development (Schreckenberg and Jacobson 1975) (the late neurula stage). Embryos were placed in agarose-coated, disposable plastic petri dishes filled with 10% Holtfretter's solution (Armstrong et al. 1989) and dechorionated gently using fine tip forceps. After dechorionation, the embryos were carefully transferred via plastic transfer pipette with a cut tip (to create a larger opening) into agarose-coated 24 well plates containing ~2ml 0.1% colchicine in 10% Holtfretter's solution to arrest the cells in metaphase. Up to two embryos were placed in each well. Plates containing embryos were incubated at 18°C for 48 hours, removing any dead embryos after the first 24 hours. After 48 hours, the embryos were removed using a plastic transfer pipette with a cut tip and washed with fresh 10% Holtfretter's solution then placed in a Dounce

homogenizer containing 15 ml of 0.075M KCl. Using 5 passes with a loose pestle, the embryos were disaggregated, then cells were allowed to swell for 45 minutes at room temperature. Being careful to avoid large clumps of cell debris at the bottom of the homogenizer, 12 ml of the mixture is removed (6ml at a time) via a 50 ml pipettor and added to two 15 ml tubes. The cells were fixed using 3:2 methanol:glacial acetic acid, reduced to ~2ml each, and stored in a -20°C incubator.

Metaphase chromosome spreading

Fixed cells were spread on a variety of slides throughout the various capture and sequencing strategies. For the first strategy, normal (1.0mm) microslides without membranes and thin (0.17 mm) polyethylene naphthalate (PEN) membrane slides were utilized. In the second strategy, normal (1.0mm) PEN and polyethylene terephthalate (PET) membrane slides were used. After the first strategy, all membrane slides were UV-treated for 30 minutes prior to spreading. The final strategy utilized only normal UV-treated PEN slides. In the following steps, the slides were handled in the same way for chromosome spreading.

Slides were inverted one at a time over a steam bath, made using a 35ml plastic disposable petri dish full of distilled water set on a 60°C hotplate, for 7 seconds. Immediately after steaming, a 20-200µl micropipettor is used to drop 100µl across the middle of the slide, lengthwise, in a sweeping motion. Each slide was immediately placed in a steam chamber at ~35°C for 1 minute, then set on the hot plate for 5 minutes. After slides are dry, chromosomes were stained via immersion in freshly made Giemsa stain for 2 minutes, rinsed in 95% ethanol, rinsed in distilled water, then allowed to dry in a desiccator until used.

Laser capture microdissection

Chromosomes were dissected individually using a Zeiss PALM Laser Microbeam Microscope at 40X magnification. Microdissected chromosomes were catapulted into Zeiss adhesive cap tubes using the laser energy from the scope. Following capture, the tubes were closed, labeled and held in plastic bags until transported

back to the laboratory for amplification. The time a tube is allowed to sit ranges from 10 minutes to several hours (during capture of other dyads) with no obvious effect on the resulting library.

Amplification of chromosomal libraries

10µl of a chromatin digestion buffer (Keinath et al. 2015) was pipetted into the adhesive caps. They were kept inverted and incubated overnight at 55°C. After incubation, the samples were centrifuged briefly in a tabletop centrifuge and placed in the thermal cycler to inactivate the Proteinase K at 75°C for 10 minutes and 95°C for 4 minutes.

Following this step, chromosomes were either stored at -20°C or carried through full amplification via Rubicon Whole Genome Amplification (WGA) kit. In the first 2 strategies, the original amplification kit was employed, but in the third strategy and all subsequent sequencing attempts, a newly developed PicoPlex WGA kit (now called PicoPlex DNA-seq kit) was used. The standard manufacturer protocol was used with the exception of the cell extraction step, as chromatin digestion buffer was used prior to the second step.

Using an Agilent 2100 Bioanalyzer and accompanying DNA 12000 kit, the samples were assessed for approximate concentration and size distribution. Only those samples with at least 9ng/µl were considered for sequencing. Those that fell below this threshold were further amplified using the suggested additional cycles from the kit and re-run on the Bioanalyzer. In every instance of chromosome capture, a piece of glass or membrane (depending on slide type used) was captured and processed on the Bioanalyzer as a negative control. The chosen samples were stored at -80°C until they were sent on dry ice to Hudson Alpha Institute for Biotechnology in Huntsville, Al. Sequencing techniques were different for each strategy and described below.

Comparison of thin (0.17mm) membrane slides vs. normal (1.0mm) microslides

For the first strategy, 6 samples were collected. The first two samples contained 20 chromosome (chr) 3 dyads excised from normal (1.0mm) glass microslides. The third sample contained 6 chr 3 dyads excised from thin membrane slides. The last three samples contained individual dyads from chromosomes other than chr 3, the first was taken from a thin membrane slide, and the others taken from regular slides. The samples were sent for sequencing 100bp paired-end reads on an Illumina HiSeq 1000. After initial attempts to sequence yielded no salamander data (see results), sequencing was preceded by 36 dark cycles in order to initiate cluster identification after polymerase had extended through proprietary random primer sequences.

Resulting reads were trimmed single ended using Trimmomatic (Bolger et al. 2014) to remove leader sequences and trimmed using the sliding window option to trim once the average quality within the window of 40 nucleotides fell below a threshold of Q30. Reads shorter than 40 nucleotides were removed from the data. Each library was then aligned to the human reference genome and bacterial genomes using Bowtie 2 with the paired ended mapping option to identify exact matches. Concordantly mapped reads for bacteria were removed. Human reads were removed when more than 20% of reads mapped to human concordantly. Reads were then aligned to model transcripts from the *Ambystoma* linkage map (Voss et al. 2011) using the Burrows Wheeler Aligner with the single-end mapping option and BWA-MEM algorithm (Li and Durbin 2009). Proportions of reads mapping to individual linkage groups (LG) were assessed and those showing a higher proportion of reads on an individual chromosome were compared to proportion to all other LGs to determine precision to a LG. Further sequencing using the same methods (including 36 dark cycles) was performed on the sample containing 6 chr 3 dyads.

In-line adapter ligation and comparison of PEN and PET slides

5 samples were collected for the second sequencing strategy. Because the thin (PEN 0.17) membrane slides worked best in the previous attempt for sequencing but were broke easily, thicker (normal, 1.0mm) membrane slides of both varieties (PEN and PET) were used in this set of experiments. Two samples contained individual dyads that were collected from PET membrane slides. One sample contained a chromosome 3 dyad collected from a PEN membrane slide, another contained a chromosome 3 dyad collected from a PET slide, and the final sample contained 5 chromosome 3 dyads taken from PET membrane slides.

Prior to library preparation, random hexamers were ligated to the WGA fragments to promote cluster identification. By incorporating a diverse nucleotide sequence in the 5' region of all fragments, diversity requirements by the image processing algorithms were met and allowed far more sequence to be produced. One caveat associated with this approach is that much of the data that are generated by the sequencer originates from barcodes, primer and adapter sequence, rather than the target chromosome DNA. Downstream data processing (as described in the previous section) involved the trimming of reads to remove hexamer and random primer sequences, and resulted in short, 64 bp reads from the original 100bp read.

Amplification with the PicoPlex WGA DNAseq kit

9 samples were collected. Two samples contained 10 chr 3 dyads, 2 samples contained 5 chr 3 dyads, 2 samples contained 1 chr 3 dyads, and the last 3 samples contained individual dyads of varying sizes. All chromosomes were excised from a UV-treated normal (1.0) PEN membrane slide. Amplification was accomplished through the implementation a modified version of the PicoPlex WGA kit that was kindly provided by Rubicon Genomics for alpha testing. The kit included a set of 12 barcodes and illumine-specifiic priming sequences, so library preparation is no longer outsourced as done in the first two strategies. In a final version of the PicoPlex kit (DNAseq), a set of 48 i5 and i7 barcodes are included. A control human DNA sample was amplified, barcoded and sequenced alongside

the chromosome samples. The new barcoding allowed for 48 samples to be sequenced on a single lane, increasing throughput but decreasing the depth of coverage for each sample. In this way, samples with hits to individual linkage groups could be selected and sequenced more deeply. The amplified samples were sequenced on an illumina HiSeq 2000 for paired-end 100bp reads. All sequences were trimmed as done previously, but reads on average were much longer than in previous strategies, averaging 85bp.

Reads were mapped to bacterial, human and *Ambystoma* transcripts and assessed as described previously. Both samples containing 10 chr 3 dyads and one sample containing a large individual dyad were further sequenced.

Table 2.1. Initial attempts to sequence individual and pooled WGA-amplified chromosomes

Sample ID	Tube contents	Number of Reads	Microslide	Reads mapped (proportion)	% Reads Human	% Reads Bacteria
A1	6 chr 3 dyads	19959046	PEN 0.17	1737 reads on 25 LG3 contigs (0.27)	49.71	0.51
A2	1 small dyad	7768580	PEN 0.17	132 reads on 15 LG7 contigs (0.426)	21.58	0.77
A3	20 chr 3 dyads	10049905	Normal	Not specific	23.3	1.9
A4	20 chr 3 dyads	11261099	Normal	Not specific	57.91	0.26
A5	1 small dyad	9328152	Normal	Not specific	40.07	1.87
A6	1 small dyad	11768042	Normal	Not specific	26.7	0.42
Additional sequencing						
A1	6 chr 3 dyads	7357412 combined: 27316458	PEN 0.17	Combined: 1792 reads on 30 LG3 contigs (0.326)	50.66	0.57%
A2	1 small dyad	4550396 combined: 12318976	PEN 0.17	Combined: 272 reads on 19 LG7 contigs (0.543)	5.4	0.92

Table 2.2. Sequencing of individual and pooled WGA-amplified chromosome following in-line adapter ligation

Sample ID	Tube contents	Reads	Microslide	Reads mapped	% Reads Human	% Reads Bacteria
B1	5 chr 3 dyads	21823936	PET 1.0	86 reads on 24 LG3 contigs (0.26)	34.28	0.12
B2	1 chr 3 dyad	22094221	PEN 1.0	1843 reads on 66 LG3 contigs (0.717)	28.83	0.28
B3	1 small dyad	18549440	PET 1.0	Not specific	35.36	6.96
B4	1 small dyad	20663477	PET 1.0	Not specific	75.98	0.57
B5	1 chr 3 dyad	24288480	PET 1.0	Not specific	61.01	0.22

Table 2.3. Sequencing of individual and pooled WGA-amplified chromosomes using the PicoPlex WGA DNAseq kit

Sample ID	Tube contents	Reads	Microslide	Reads mapped	% Reads Human	% Reads Bacteria
C1	10 chr 3 dyads	15409195	PEN 1.0	1373 reads on 74 LG3 contigs (0.804)	5.08	>0.05
C2	10 chr 3 dyads	16178340	PEN 1.0	1253 reads on 68 LG3 contigs (0.739)	6.37	>0.05
C3	5 chr 3 dyads	17527597	PEN 1.0	984 reads on 62 LG3 contigs (0.674)	18.96	>0.05
C4	1 chr 3 dyad	21909822	PEN 1.0	114 reads on 29 LG3 contigs (0.315)	86.83	>0.05
C5	5 chr 3 dyads	18704479	PEN 1.0	Not specific	7.46	>0.05
C6	1 chr 3 dyad	16341084	PEN 1.0	Not specific	61.44	>0.05
C7	1 medium dyad	22883652	PEN 1.0	Not specific	16.19	>0.05
C8	1 large dyad	21013292	PEN 1.0	424 reads on 9 LG2 contigs (0.082)	17.87	>0.05
C9	1 large dyad	19145257	PEN 1.0	84 reads on 11 LG4 contigs (0.19); 139 reads on 6 LG13 contigs (0.462)	9.37	>0.05
C10	Human control	16331215	PEN 1.0		96.8	>0.05
Additional sequencing						
C1	10 chr 3 dyads	18564847 combined: 33974042	PEN 1.0	Combined: 1692 reads on 75 LG3 contigs (0.815)	4.98	>0.05
C2	10 chr 3 dyads	58144138 combined: 74322478	PEN 1.0	Combined: 2157 reads on 73 LG3 contigs (0.793)	6.53	>0.05
C9	1 large dyad	22146212 combined: 41291469	PEN 1.0	Combined: 92 reads on 11 LG4 contigs (0.19); 162 reads on 6 LG13 contigs (0.462)	9.2	>0.05

Table 2.4. Summary of additional chromosomal libraries that were sequenced using my optimized approach

Sample	Contents	Mapped reads
D1	1 small dyad	116 reads on 18 LG4 contigs (0.25)
D2	1 medium dyad	995 reads on 23 LG9 contigs (0.40)
D3	1 medium dyad	446 reads on 21 LG10 contigs (0.51)
D4	1 large dyad	531 reads on 71 LG2 contigs
D5	1 medium dyad	121 reads on 15 LG5 contigs
Whole lane Sequencing		
D2	1 medium dyad	16963 reads on 117 LG2 contigs (0.80)
D4	1 large dyad	68844 reads on 40 LG9 contigs (0.70)

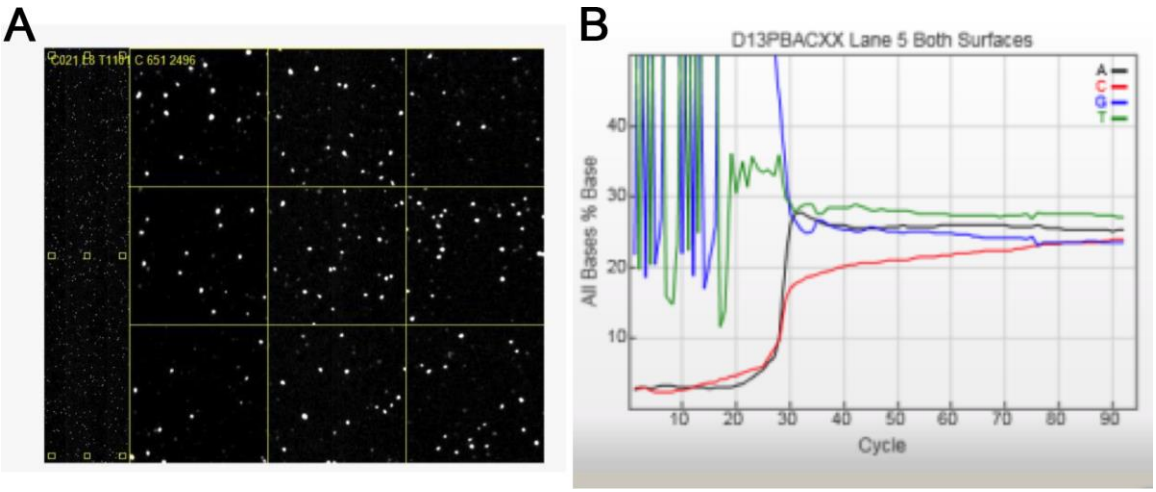

Figure 2.1. Flow cell fluorescence and nucleotide diversity for initial attempts to sequence chromosome 3.

A) Image from flow cell fluorescence from the C-channel at cycle 8 shows poor cluster detection. The left side is an image taken of the entire lane on which this DNA is being sequenced. On the right is a grid of 9 enlarged squares from the full lane showing bright spots, marking the locations of clusters that are incorporating a C nucleotide. The low-density results from a paucity of C bases being generated from amplified libraries during the first several extension cycles. The signals seen here correspond to PhiX DNA spiked into the lane as an internal control. **B)** Nucleotide distribution of the first sequencing strategy shows that among the first 30 cycles, there is extremely low nucleotide diversity with heavy GT bias shown in blue and green.

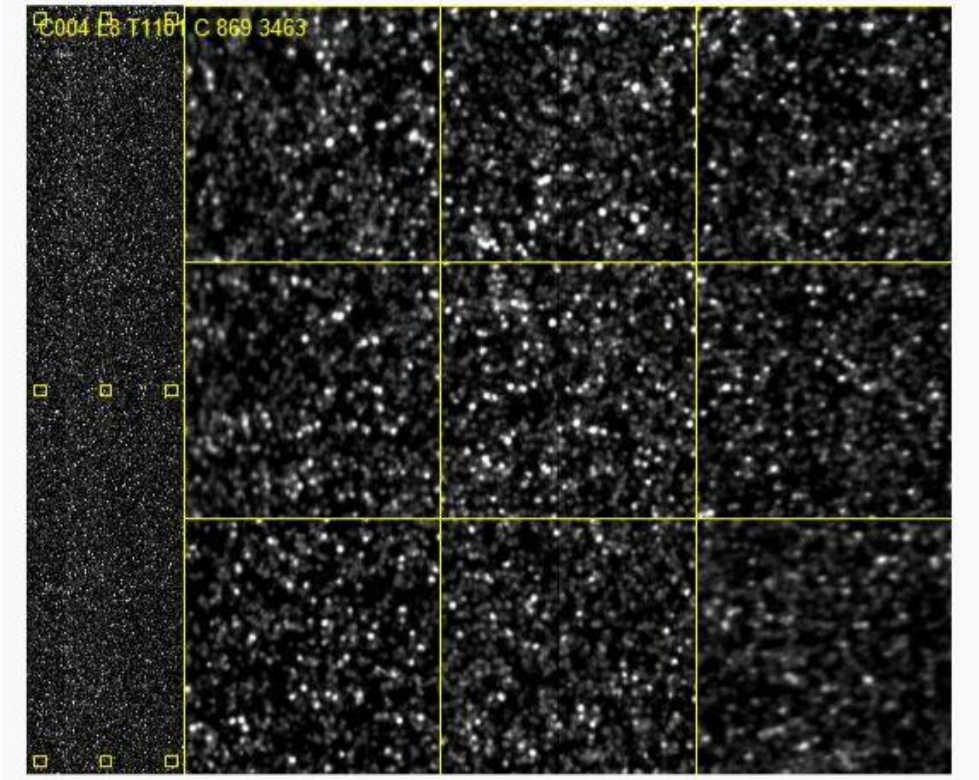

Figure 2.2. Flow cell fluorescence for initial attempts to sequence chromosome 3 after addition of 36 dark cycles.

On the left is an image of the full lane of the flow cell on which the chromosome 3 is being sequenced. This image was taken from the C-channel at cycle 4 after 36 dark cycles. The image shows that compared to the flow cell image from Figure 2.1A, far more clusters are incorporating C bases, and thus, sequences from amplified libraries are also represented.

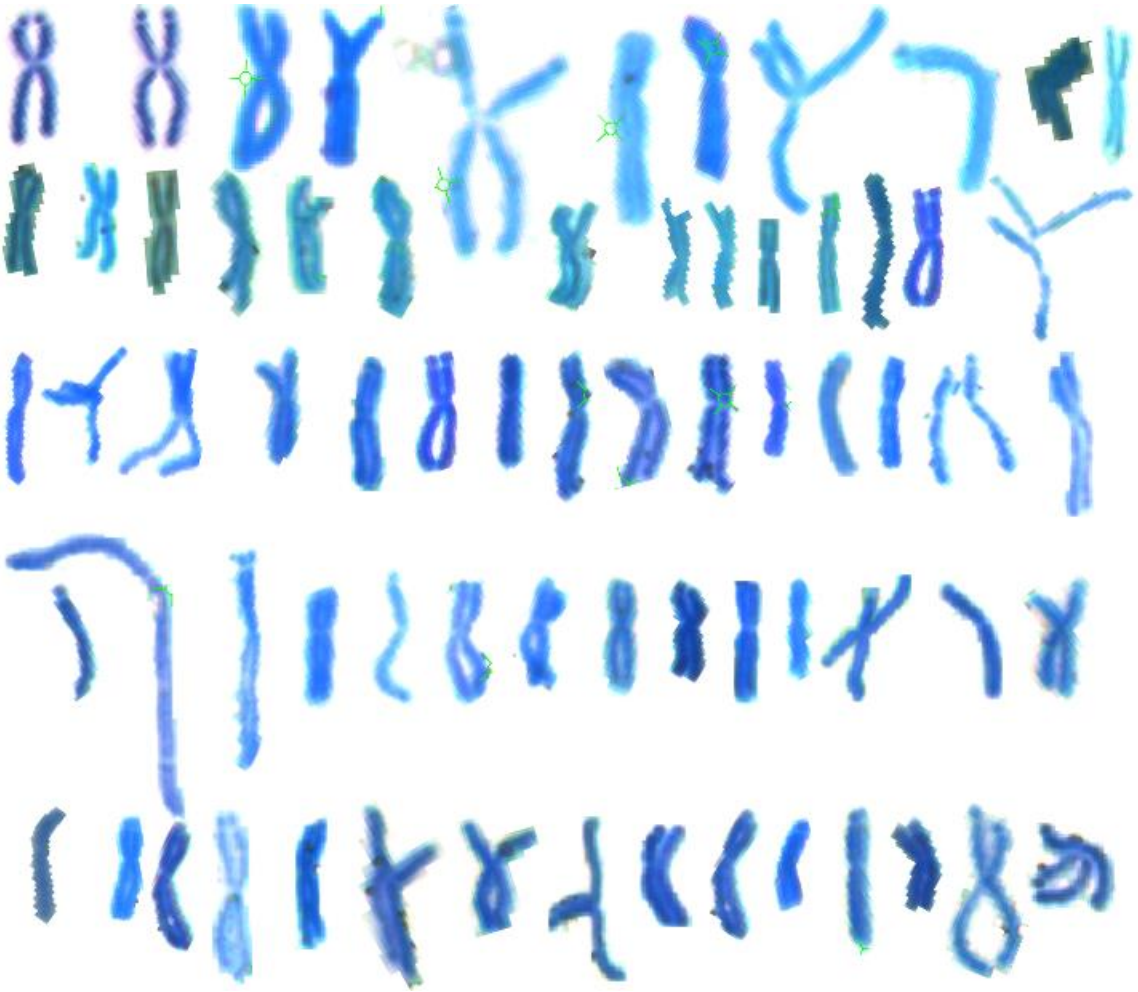

Figure 2.3. Images of individual chr 3 (LG3) dyads from a variety of mitotic spreads.

The figure depicts 69 images of chromosome 3 dyads taken from multiple metaphase Giemsa-stained axolotl chromosome spreads at 40X magnification. While chromosome 3 is easily identified within a spread of chromosomes, the contrast in sizes and shapes show the difficulty of identifying other chromosomes without a defining feature, such as the nucleolar organizer region in this chromosome, across multiple spreads.

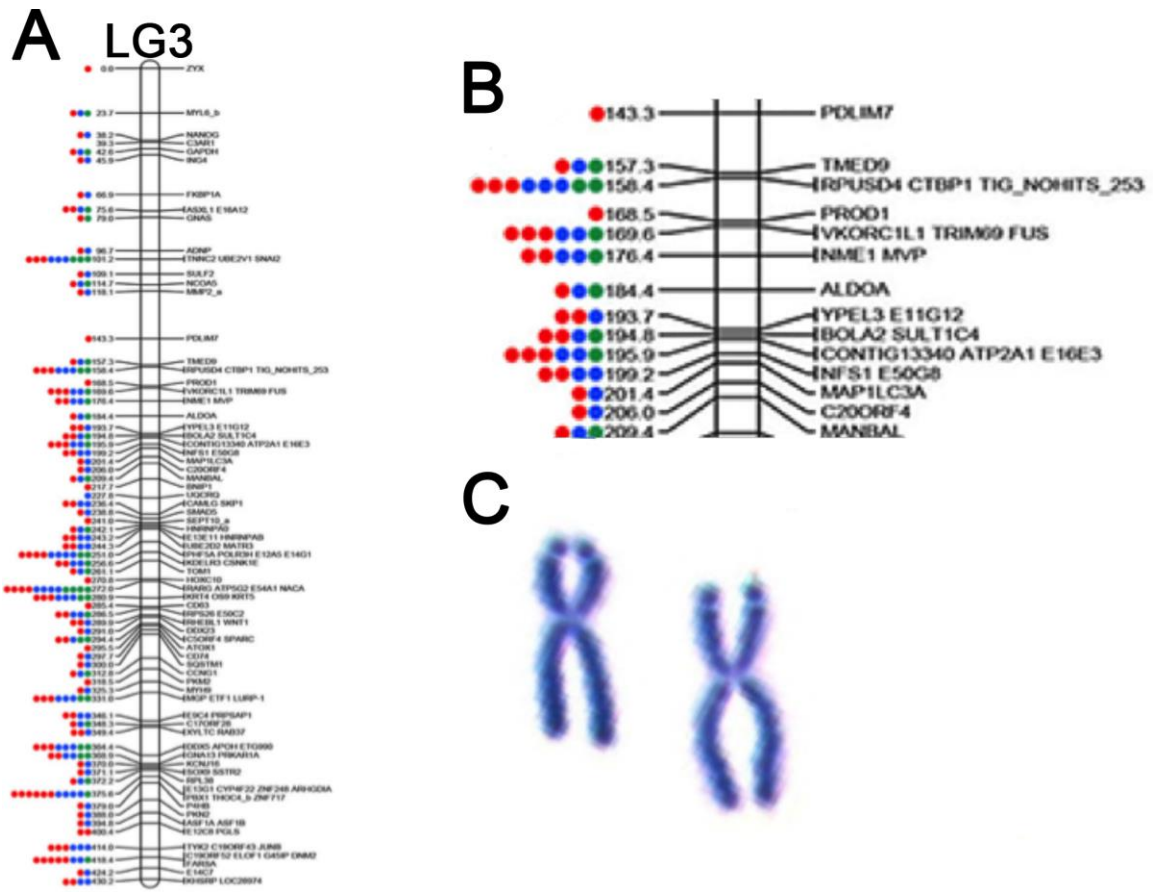

Figure 2.4. Distribution of markers to LG3 from 3 sequencing strategies. **A)** The distribution of markers sampled across linkage group 3, within three series of experiments. Green denotes mapped reads from a sequencing strategy using 36 dark cycles, blue denotes mapped reads from a strategy using in-line adapter ligation, and red denotes mapped reads from a strategy using the PicoPlex DNAseq kit. **B)** An enlarged image depicting markers hit between 143.3cM and 206.0cM. **C)** Two chromosome 3 dyads stained with Giemsa that were captured and imaged in the first sequencing strategy. The constriction subterminally in the short arm is the site of a nucleolar organizer region.

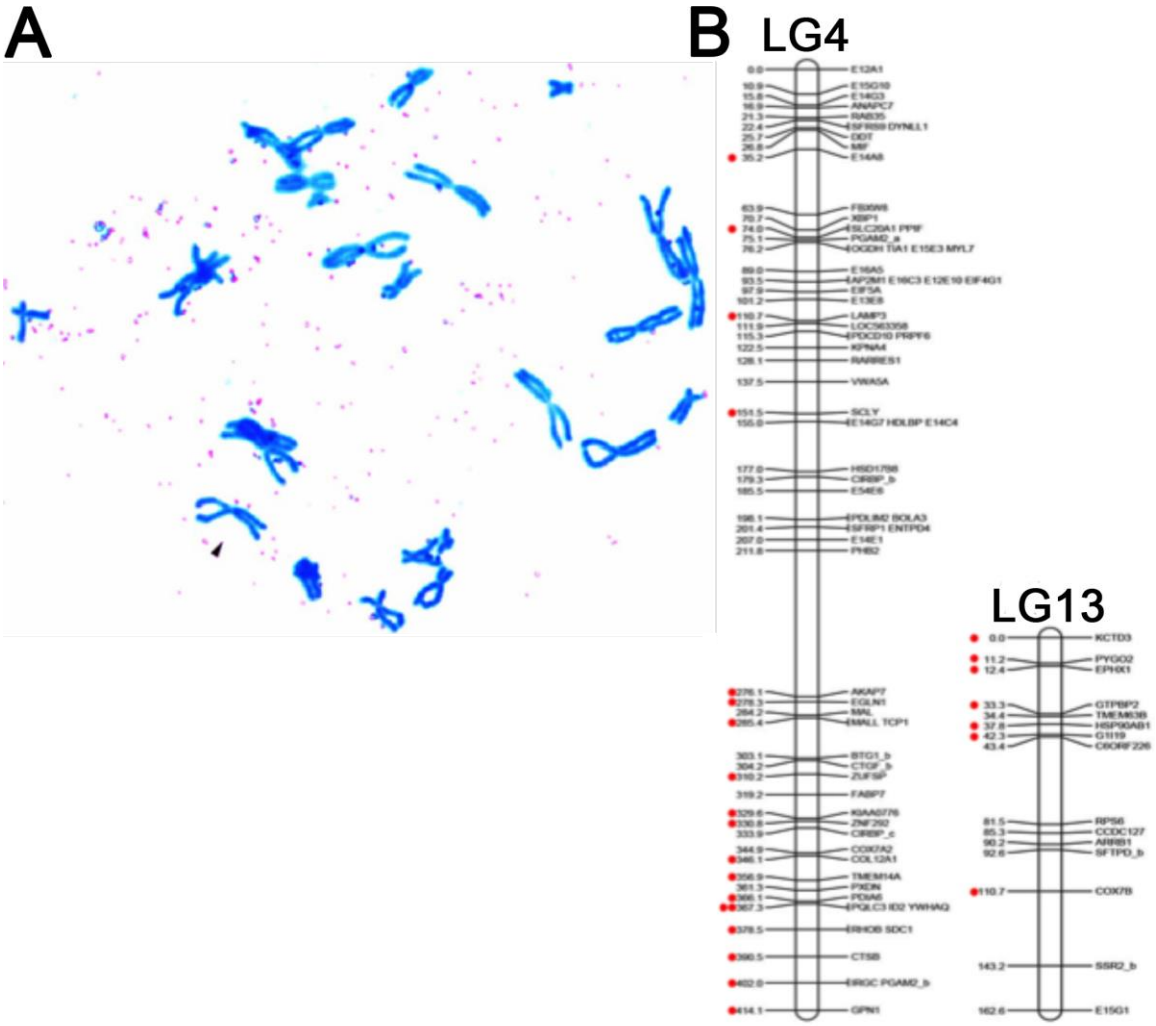

Figure 2.5. A single dyad corresponding to LG4 and 13.

Read mapping was used to assess the specificity of laser capture, amplified libraries of an individual dyad. **A)** A metaphase spread of Giemsa-stained axolotl chromosomes on a membrane slide. The arrow denotes the dyad that was captured and sequenced. **B)** The distribution of markers sampled from the individual dyad library on linkage groups 4 and 13 (labeled LG4 and LG13, respectively). Dots represent markers with mapped reads with nearly perfect matches from the individual dyad library from A.

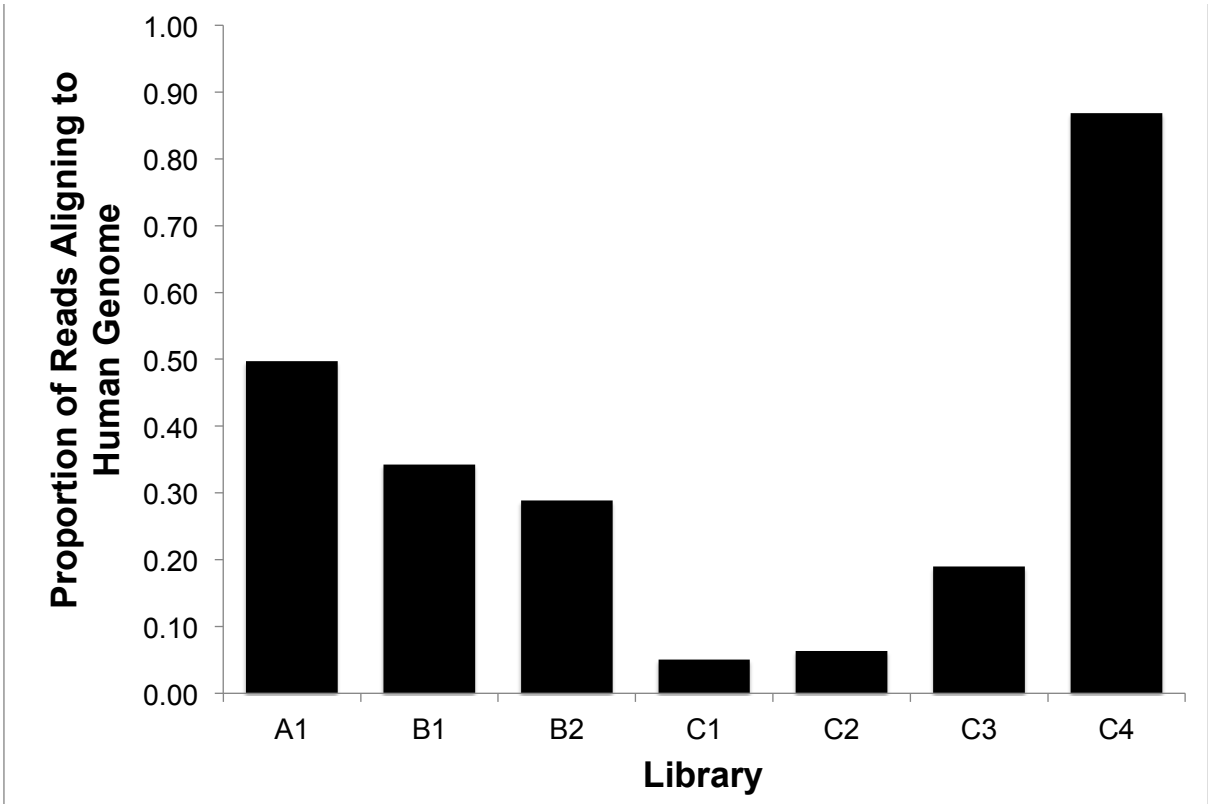

Figure 2.6. Proportion of reads aligning to human genome from LG3 libraries.

The proportion of reads aligning to the human genome, from each of the 7 LG3 chromosomal libraries. A1 represents the chromosomal library from the initial sequencing attempts of LG3 containing 6 chromosome 3 dyads. B1 and B2 represent the two chromosome (chr) 3 libraries that used in-line adapter ligation. B1 contains a single chr 3 dyad, and B2 contains 5 chr 3 dyads. C1-C4 represents the four chr 3 libraries that were generated using the new PicoPlex DNAseq kit. C1 and C2 contain ten chr 3 dyads each, C3 contains five chr 3 dyads, and C4 contains a single chr 3 dyad. The proportion of reads mapping to human decrease from initial sequencing strategies to the final strategy, except for the final sample that was generated from a single dyad and could be easily overwhelmed by little human contamination.

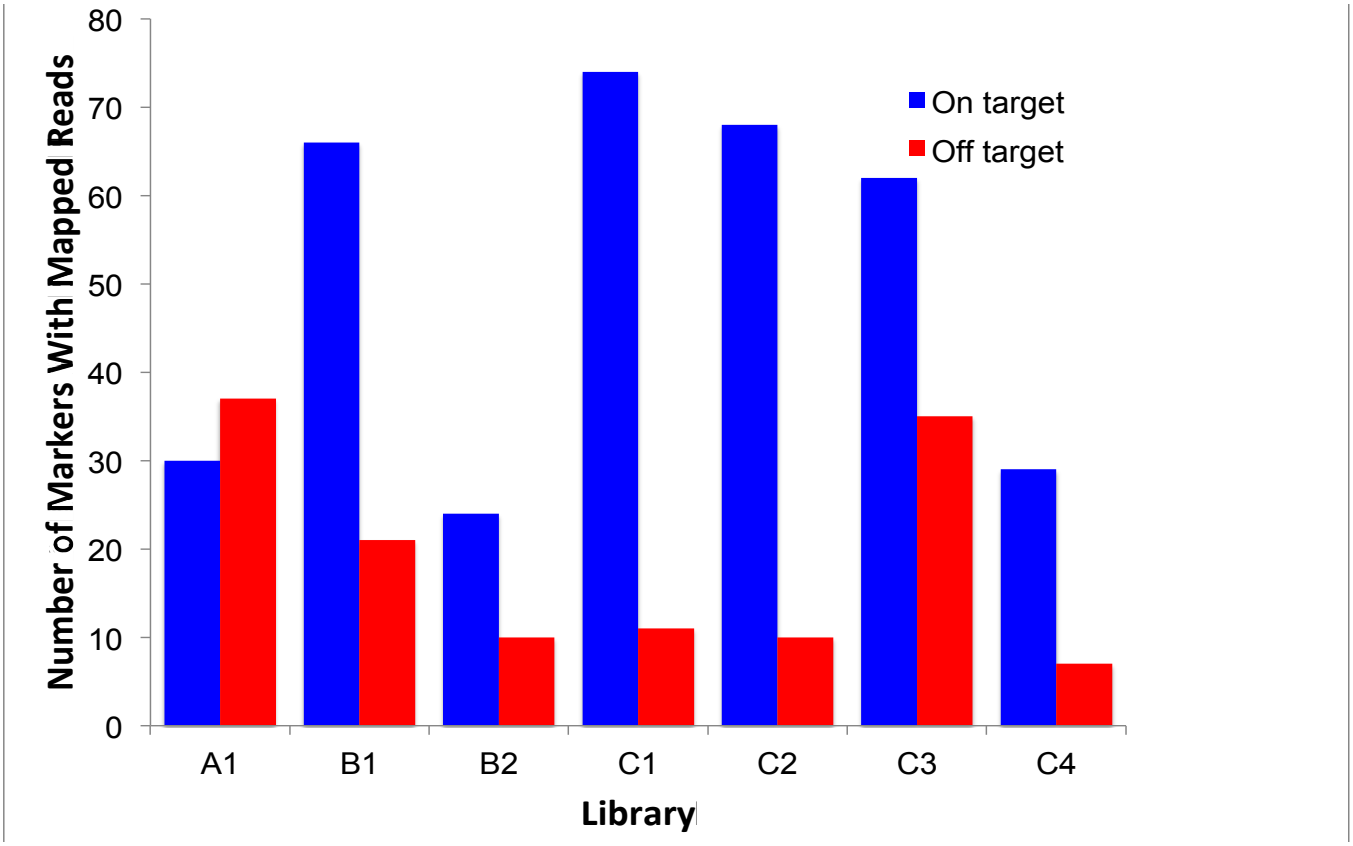

Figure 2.7. Proportion of markers on vs. off LG3.

The proportion of markers that were located on vs. off linkage group 3 and aligned to reads from each of 7 libraries that targeted the NOR chromosome. A1 represents the chromosomal library from the initial sequencing attempts of LG3, a sample containing 6 chr 3 dyads. B1 and B2 represent the two chromosome (chr) 3 libraries that used in-line adapter ligation. B1 contains a single chr 3 dyad, and B2 contains 5 chr 3 dyads. C1-C4 represents the 4 chr 3 libraries from the final sequencing attempt using the new PicoPlex DNAseq kit. C1 and C2 contain 10 chr 3 dyads each, C3 contains 5 chr 3 dyads, and C4 contains a single chr 3 dyad. The proportion of markers on linkage group 3 improved from the initial sequencing strategy to the final sequencing strategy with pooled amplification of 10 chr 3 dyads yielding the best results.

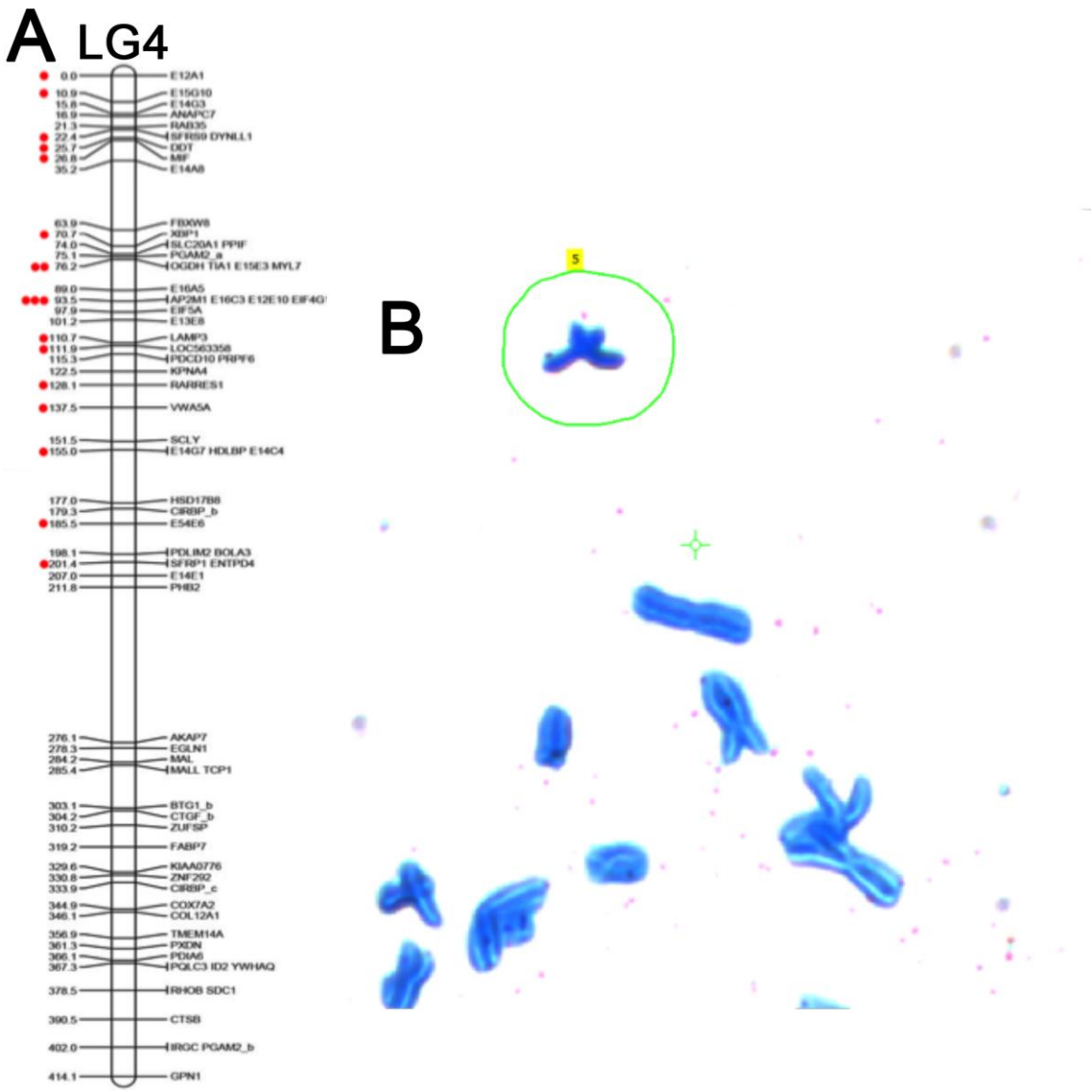

Figure 2.8. An Individual dyad that aligned to LG4.

A) The distribution of markers sampled from the individual dyad library on LG4. Dots represent markers with mapped reads with nearly perfect matches from the individual dyad library from B. Notably, all of the mapped reads are found on the top half of LG4, prior to the >30cM gap. **B)** A metaphase spread of Giemsa-stained axolotl chromosomes on a membrane slide. The green circle is the precise laser cut site for the excision of the dyad that was captured and sequenced.

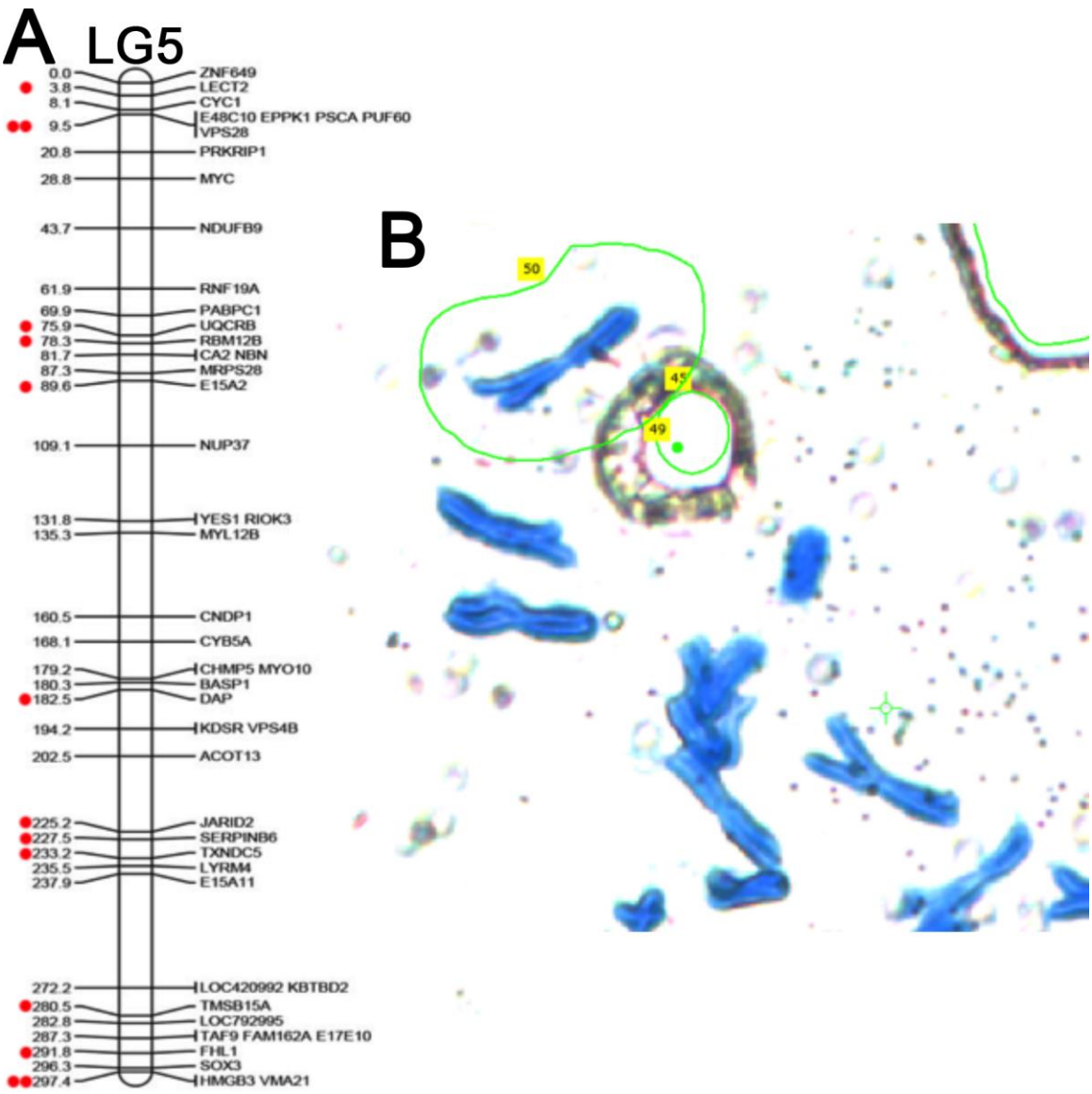

Figure 2.9. An individual dyad that aligned to LG5.

A) The distribution of markers sampled from the individual dyad library on linkage group 5. Dots represent markers with mapped reads with nearly perfect matches from the individual dyad library from B. **B)** A metaphase spread of Giemsa-stained axolotl chromosomes on a membrane slide. The green circle is the precise laser cut site for the excision of the dyad that was captured and sequenced.

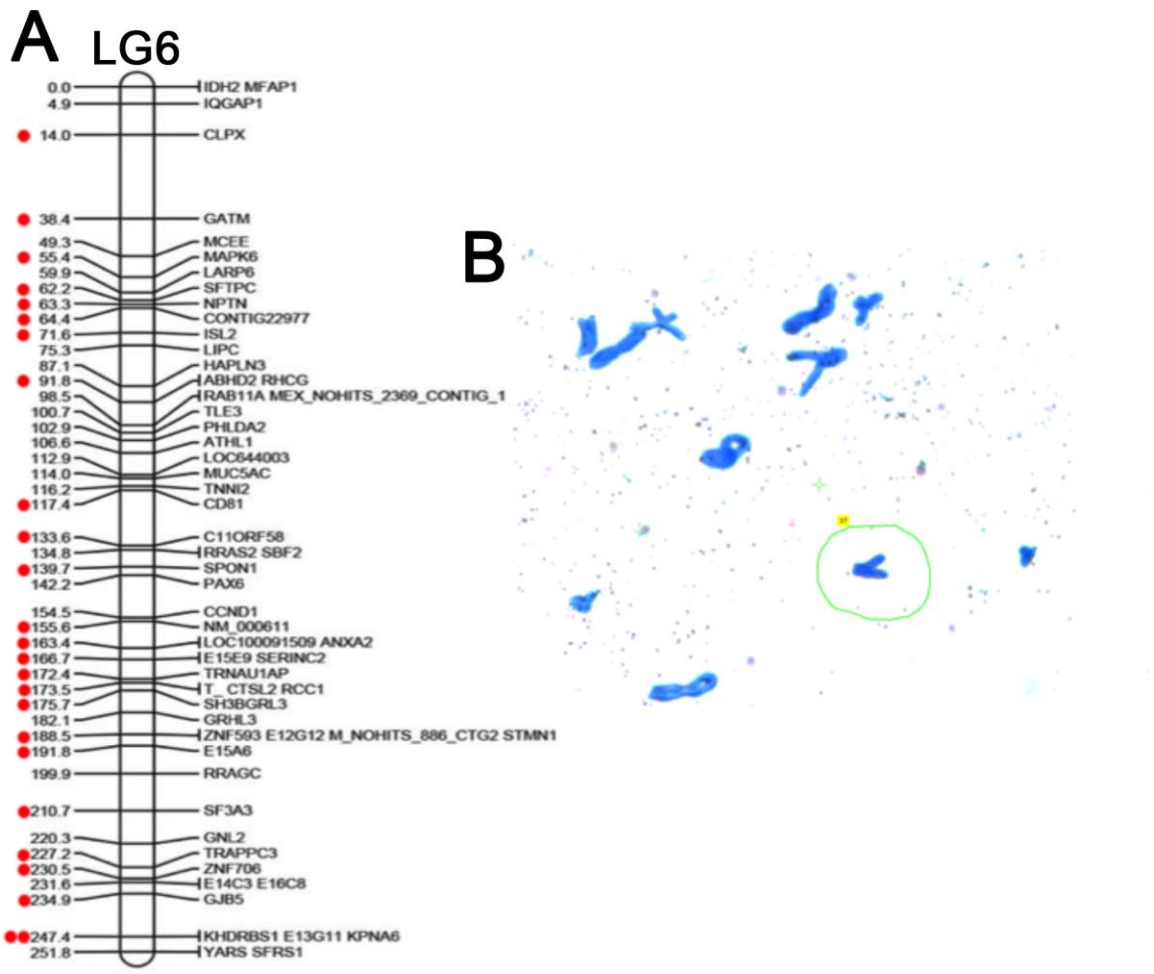

Figure 2.10. An individual dyad that aligned to LG6.

A) The distribution of markers sampled from the individual dyad library on linkage group 6. Dots represent markers with mapped reads with nearly perfect matches from the individual dyad library from B. **B)** A metaphase spread of Giemsa-stained axolotl chromosomes on a membrane slide. The green circle is the precise laser cut site for the excision of the dyad that was captured and sequenced.

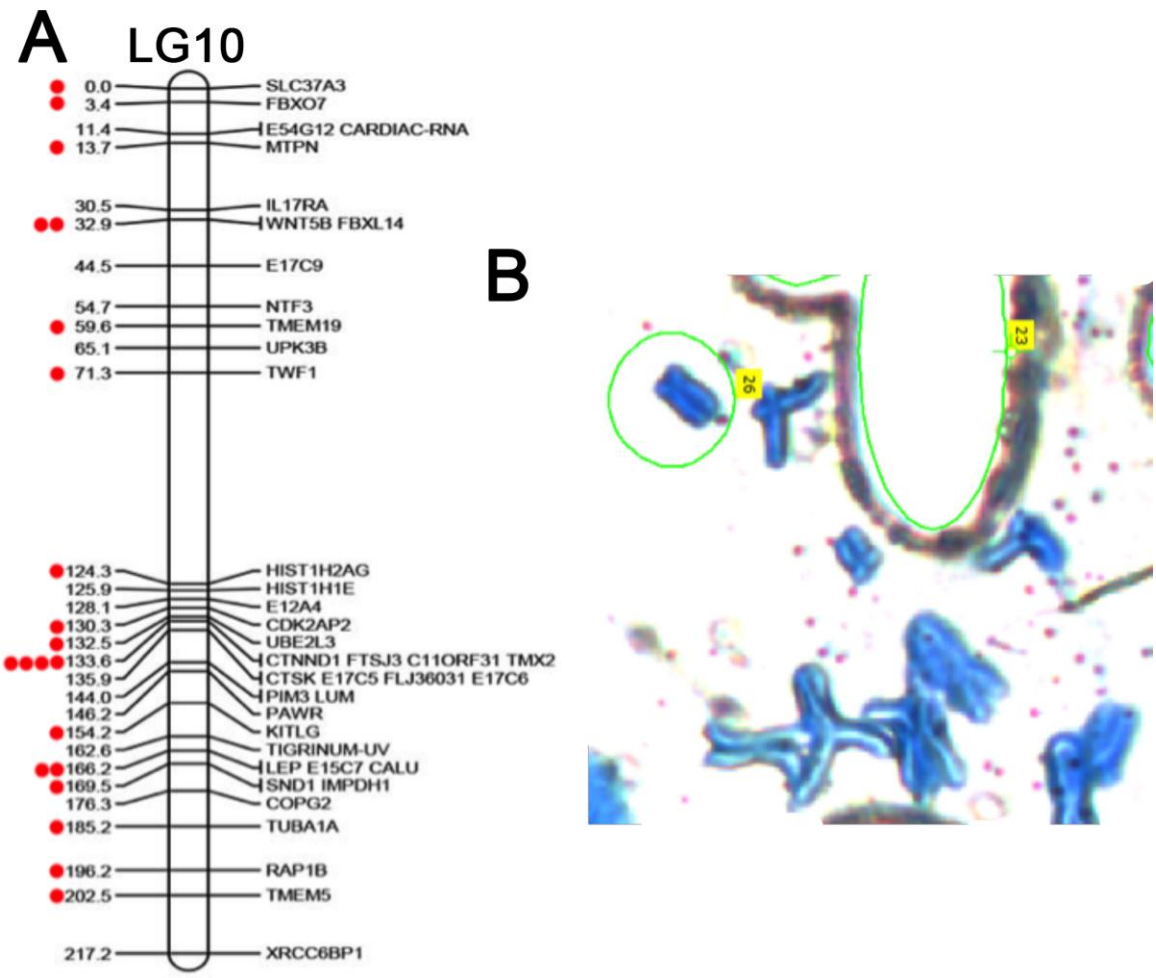

Figure 2.11. An individual dyad that aligned to LG10.

A) The distribution of markers sampled from the individual dyad library on linkage group 10. Dots represent markers with mapped reads with nearly perfect matches from the individual dyad library from B. **B)** A metaphase spread of Giemsa-stained axolotl chromosomes on a membrane slide. The green circle is the precise laser cut site for the excision of the dyad that was captured and sequenced.

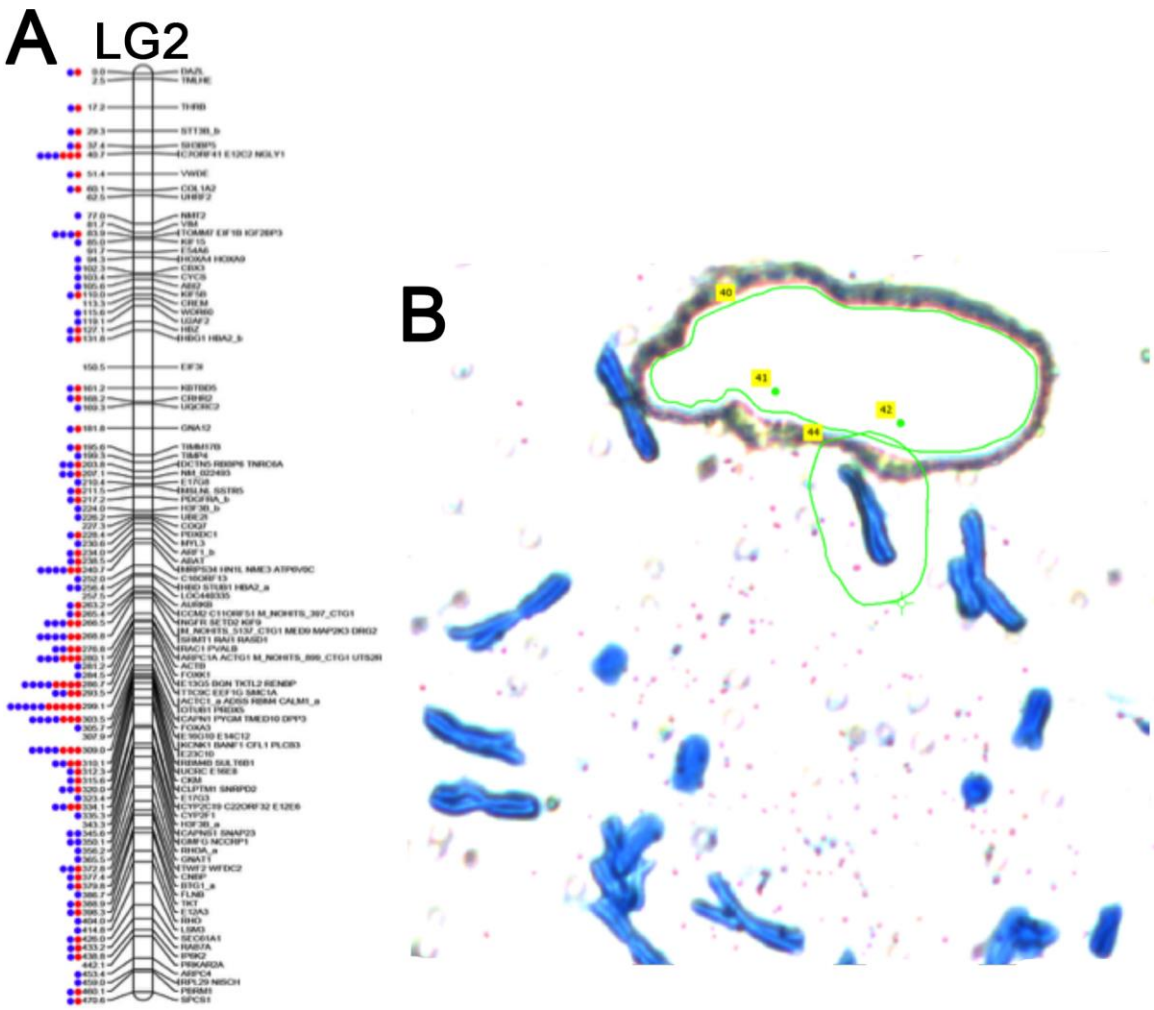

Figure 2.12. An individual dyad that aligned to LG2, and was targeted for further sequencing.

A) The distribution of markers sampled from the individual dyad library on linkage group 2. Dots represent markers with mapped reads with nearly perfect matches from the individual dyad library from B. The red denotes alignments from the first sequencing of this dyad, and blue denotes the alignments from a full lane of sequencing performed on the same chromosomal library. **B)** A metaphase spread of Giemsa-stained axolotl chromosomes on a membrane slide. The green circle is the precise laser cut site for the excision of the dyad that was captured and sequenced.

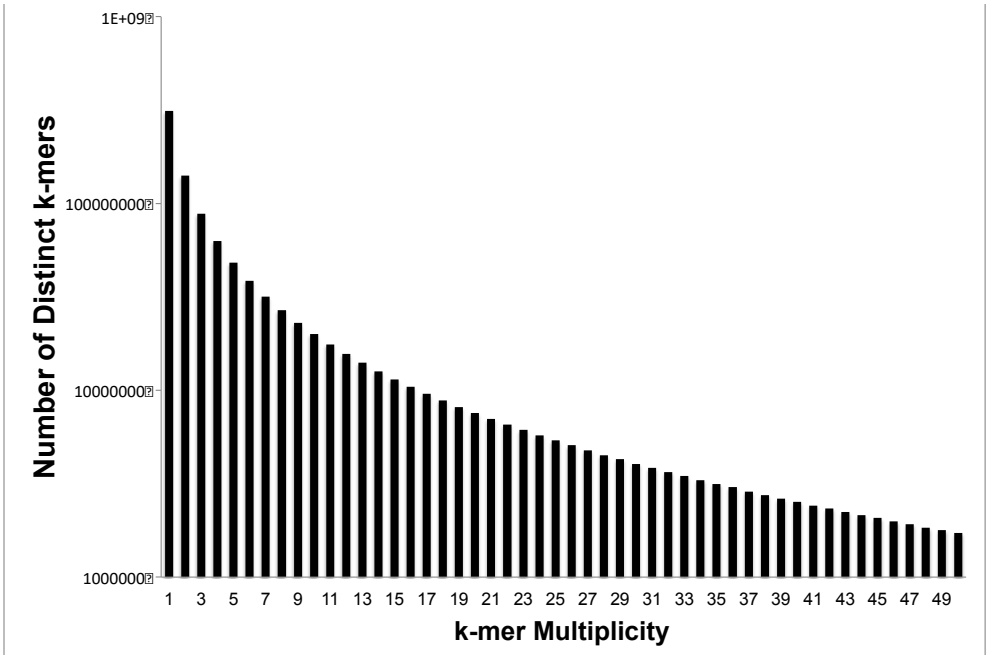

Figure 2.13. Distribution of 23-mer frequencies among the quality filtered sequence data generated from seven chr 3 libraries that used all three sequencing strategies.

The observed distribution of 23-mer frequencies from LG3 libraries (A1, B1, B2, and C1-C4) that shows no hump distribution but instead a steady decline across k-mer multiplicities. This may be a direct result of amplification bias associated with the amplification kit, as it is expected that the single-copy fraction of the chromosome would form a curve with the average depth of coverage presenting at the peak.

CHAPTER THREE

INITIAL CHARACTERIZATION OF THE LARGE GENOME OF THE SALAMANDER *AMBYSTOMA MEXICANUM* USING SHOTGUN AND LASER CAPTURE CHROMOSOME SEQUENCING

Reproduced from: Keinath MC, Timoshevskiy VA, Timoshevskaya NY, Tsonis PA, Voss SR & Smith JJ (2015) Initial characterization of the large genome of the salamander *Ambystoma mexicanum* using shotgun and laser capture chromosome sequencing. *Scientific Reports*. DOI: 10.1038/srep16413

Abstract

Vertebrates exhibit substantial diversity in genome size, and some of the largest genomes exist in species that uniquely inform diverse areas of basic and biomedical research. For example, the salamander *Ambystoma mexicanum* (the Mexican axolotl) is a model organism for studies of regeneration, development and genome evolution, yet its genome is ~10X larger than the human genome. As part of a hierarchical approach toward improving genome resources for the species, we generated 600 Gb of shotgun sequence data and developed methods for sequencing individual laser-captured chromosomes. Based on these data, we estimate that the *A. mexicanum* genome is ~32 Gb. Notably, as much as 19 Gb of the *A. mexicanum* genome can potentially be considered single copy, which presumably reflects the evolutionary diversification of mobile elements that accumulated during an ancient episode of genome expansion. Chromosome-targeted sequencing permitted the development of assemblies within the constraints of modern computational platforms, allowed us to place 2062 genes on the two smallest *A. mexicanum* chromosomes and resolves key events in the history of vertebrate genome evolution. Our analyses show that the capture and sequencing of individual chromosomes is likely to provide valuable information for the systematic sequencing, assembly and scaffolding of large genomes.

Introduction

Vertebrate genomes encompass a broad range of genome sizes ranging from 340 Mb to ~130 Gb (Gregory et al. 2007). Notably, some of the largest and most complex genomes exist in species that are of critical importance to biomedicine and evolution. Tackling the challenge of assembling large vertebrate genomes will likely require new, hierarchical approaches for sequencing, assembly and scaffolding. Considerable progress has been made in recent years to enhance genomic and molecular resources for a species with a genome size that falls toward the upper end of the vertebrate range, the primary salamander model *Ambystoma mexicanum* (Mexican axolotl). For example, large-scale transcript sequencing efforts have enabled transcriptome, proteome, and quantitative trait locus analyses of vertebrate characteristics for which axolotls are particularly informative, including tissue regeneration, thyroid hormone signalling, paedomorphosis, and karyotype evolution (Voss and Smith 2005; Smith and Voss 2006; Page et al. 2008; Monaghan et al. 2009; Voss et al. 2011; Page et al. 2013; Voss et al. 2013a; Voss et al. 2015). Methods to create transgenic axolotls and manipulate gene functions are developing rapidly and are permitting in-depth functional analyses of known candidate genes (Monaghan and Maden 2012; Khattak et al. 2013; Flowers et al. 2014; Yasue et al. 2014). Even with these advances, the axolotl lacks a fundamental resource to facilitate the use of modern, sequence-based methods of inquiry - a complete genome assembly.

The axolotl genome consists of 14 chromosome pairs (2N = 28) (Fankhauser and Humphrey 1942) and estimates of its physical size range from 21 - 48 gigabases (Edstrom and Kawiak 1961; Capriglione et al. 1987; Licht and Lowcock 1991). Early DNA reannealing studies suggested that repetitive elements constitute at least 70% of the *A. tigrinum* genome, a close relative of the axolotl; similarly large repetitive fractions are predicted for other salamander genomes (Britten and Kohne 1968; Straus 1971; Morescalchi et al. 1974; Rosbash et al. 1974; Baldari and Amaldi 1976). Indeed, large genome size is a common feature of all extant salamanders, suggesting a shared period of genome expansion prior to the basal salamander divergence during the early

Jurassic, ~180 million years ago (Zhang and Wake 2009). Recent studies have documented expansions of intergenic and intronic regions in the axolotl genome, and increases in the lengths of these regions are associated with increases in the prevalence of potential functional elements and younger repetitive sequences (Smith et al. 2009). Given our existing state of knowledge, it seems reasonable to conclude that axolotl genome structure was shaped by ancient expansions of mobile repetitive elements, and this created fertile landscapes for mobile element dynamics and DNA sequence evolution on more contemporary timescales.

Large genome size and repetitive DNA content are often cited as challenges for genome assembly, and the impact of these factors is dependent on the genomic distribution of repetitive and single copy sequences. If repetitive sequences are broadly interspersed throughout the genome, each copy (depending on length and identity with other copies) may be associated with a distinct break in assembly contiguity, potentially ranging from dozens to millions of breaks per repeat family. However, clustered repeats may be associated with a substantially smaller number of breaks. For example, centromeric repeats often represent a substantial fraction of a genome's repetitive landscape but are localized to discreet genomic segments. In total, centromeric repeats are expected to break sequence contiguity at relatively few positions within a genome assembly. If the majority of repetitive sequences in salamander genomes are organized similarly, the development of a contiguous genome assembly may become a more tractable problem, with repetitive regions being localized to relatively few large assembly gaps. Thus far, little is known about large-scale distribution of repetitive elements in salamander genomes. A few large DNA fragments (~150 kb) for genic regions of the axolotl and newt genomes have been sequenced and assembled, indicating that the structure and distribution of repeats is compatible with genome assembly of sub-megabase intervals (Smith et al. 2009; Voss et al. 2013b). However, no salamander genome has been sampled sufficiently to establish the overall size and sequence composition of repetitive (and single copy) regions, and thus the inherent complexity of the genome assembly problem.

Here, we describe initial sequencing and analysis of the *A. mexicanum* genome. Over 6 billion shotgun sequence reads were generated to achieve 19x coverage of the genome. From these sequence data, we estimate the total size of the genome to be ~32 Gb, with the repetitive fraction representing approximately 40% of the total sequence length. Attempts to directly assemble these shotgun sequence data reveal the computational complexity of assembling the largest vertebrate genomes: assemblies fail due to memory limitations (beyond 1 terabyte of RAM). Given these computational constraints, we developed and implemented chromosome capture and sequencing approaches that permitted the assembly to be divided into tractable subsets, while also providing intrinsic scaffolding information. Sequence data from the two smallest axolotl chromosomes were used to: 1) validate and extend previous linkage mapping studies and 2) develop strategies for assembly and anchoring of individually sequenced chromosomes. Altogether these studies point to a multipronged approach that will permit the development of high quality genome assemblies for *A. mexicanum* and other salamander models.

Results

A total of 16 lanes of Illumina shotgun sequence data (2 x 100 bp) were generated for a single female *A. mexicanum*, obtained from the standard laboratory strain maintained by the *Ambystoma* Genetic Stock Center (Animal # 13003.1), yielding > 0.6 Tb of raw sequence data. Analyses of k-mer frequencies were performed to estimate sequence coverage, genome size and repeat content. Filtered and quality trimmed sequences consisted of nearly 24 billion k-mers (at k = 31), with multiplicities ranging from one to 16 million. The k-mer sampling distribution contained a distinct peak at a coverage of ~13 (Figure 3.1). Accounting for k-mer sampling across reads and ignoring presumptively erroneous k-mers, we estimated that this shotgun sequence dataset averaged ~19x sequence coverage across the genome. Correspondingly, we estimate the length of the *A. mexicanum* genome to be slightly over 32 Gb (32,148,237,452 bp).

This shotgun sequence dataset was also used to generate an estimate of the relative size of the repetitive fraction of the axolotl genome, specifically as it relates to the assembly of sequences generated from short read chemistries. Based on k-mer sampling, we found that 60.1% of the genome (~19 Gb: modal coverage ± 3 s.d.) can be considered effectively single copy, with the remaining repetitive fraction (13 Gb) occurring at copy numbers ranging from two to just over one million. In general the distribution of repetitive k-mers indicates that the diversity of repetitive sequences scales inversely with copy number, consistent with interpretations of previous DNA reannealing studies (Straus 1971; Morescalchi and Serra 1974) (Figure 3.1C).

To provide additional perspective on sequence coverage and repeat content, we also aligned our shotgun sequence dataset to a collection of 24 large genomic intervals that were assembled by BAC sequencing (covering 2.6 Mb, or just under 0.01% of the genome). These analyses corroborate our k-mer based coverage estimates and provide a similar perspective on the content of repetitive elements. In total ~40% of the assembled BAC sequence can be characterized as ~single copy (with depth of coverage ranging between 1 and 40X, Figure 3.2). The remaining repetitive fraction shows a pattern consistent with k-mer based analyses but does not capture repeats at copy numbers in excess of 200,000. Extrapolating these estimates to the entire genome yields size estimates of ~13 Gb for the single copy fraction and ~19 Gb for the repetitive fraction. Presumably differences between alignment-based and k-mer based estimates reflect both increased sensitivity of alignment-based methods (which permit inexact matches) and the fraction of the genome considered.

While shotgun sequence data permit an assessment of the content of high-identity repetitive elements, they provide less information regarding the genome-wide distribution of high-identity repetitive elements or the abundance and distribution of transposable elements that have accumulated mutations over evolutionary time. To further assess the genome-wide distribution of high-identity repetitive elements, we performed in situ hybridization using the repetitive fraction of the genome (the most rapidly reannealing 40%, approximately

corresponding to copy number >1 in Figure 3.1C). Hybridization of sequences revealed that repetitive DNAs are both strongly localised to the centromeres and also interspersed across chromosomal arms at varying densities (Figure 3.3). These analyses show that a relatively large fraction of repetitive DNA exists as interspersed elements. As discussed above, these interspersed repeats are expected to have the largest impact on assembly.

In practice, the repeat content/distribution and size of the axolotl genome has presented major challenges to genome assembly. Attempts to directly assemble our shotgun sequence dataset failed due to memory limitations associated with traversing the de Bruijn graph structure during initial phases of contig extraction, despite the availability of one terabyte of RAM for these operations [Supplementary text from: (Keinath et al. 2015)]. In an attempt to circumvent these constraints, we turned our attention to developing a targeted sequencing approach that could rapidly generate data from smaller and discrete partitions of the genome while providing Gb-scale scaffolding/anchoring information. Using laser capture microscopy, individual mitotic dyads corresponding to the two smallest chromosomes of the axolotl karyotype were isolated for DNA amplification and sequencing. Notably, the axolotl karyotype is characterized by a graded series of chromosomal morphologies, which hampers the definitive identification of individual chromosomes. As such, we attempted to sequence individually amplified dyads in order to prevent cross contamination. In total, we generated and sequenced 12 barcoded libraries that are each derived from an individual dyad. Resulting reads were aligned to 918 transcribed sequences that were anchored to the *Ambystoma* genetic map. Initial analyses identified six libraries that were enriched for markers on specific *Ambystoma* linkage groups (Figure 3.4A) and these were prioritized for further sequencing. Four of these libraries yielded nearly complete coverage of two separate LGs (LG 15 and 17; Figure 3.4B). Replicated sampling of LG15 and LG17 markers among four independent libraries strongly suggests that these linkage groups comprise a single chromosome (AM13). Two additional libraries yielded almost complete coverage of a single linkage group (LG14; Figure 3.4C), confirming that

this small linkage group corresponds to a single chromosome (AM14). Overall, the fact that these libraries were heavily enriched in reads that map to the smallest *Ambystoma* linkage groups provided strong evidence that the sequencing approach was accurate and precise.

To assess the utility of laser capture libraries in generating chromosomal assemblies, we performed several analyses using combined short-insert data from each of the two target chromosomes. In an attempt to correct for sequence errors associated with the preparation of amplified libraries, we also performed parallel assemblies that leveraged whole genome shotgun data to perform error correction and read filtering prior to assembly (Greenfield et al. 2014). The relative decrease in the size and complexity of chromosome-targeted datasets (relative to the whole genome shotgun dataset) dramatically decreased the computational resources required for assembly. This allowed us to perform several rounds of assembly and optimize parameters for constructing and traversing de Bruijn graph structures. Each round of assembly required fewer than 6 hours to complete on a server with 512 GB RAM. The four optimized assemblies each exceeded 100 Mb in total length, with N50 scaffold lengths approaching 1 kb (Table 3.1). Error correction of amplified datasets decreased the number of singleton contigs, increased assembly N50 lengths and improved local scaffolding of both chromosomes. Alignment of chromosome-specific assemblies to our whole genome shotgun dataset also provided further corroboration of k-mer and BAC-based estimates of sequence coverage (Figure 3.5).

To complement and evaluate out targeted assemblies, we aligned all reference *A. mexicanum* transcripts (Smith et al. 2005b) to our draft chromosome assemblies. In doing so, we were able to place a total of 1141 reference genes on AM13 and 921 on AM14 [Supplementary Table 1 from (Keinath et al. 2015)]. To independently assess the validity of these annotations, *A. mexicanum* genes were aligned to the chicken genome. Previous analyses have revealed strong conservation of synteny between chicken and *A. mexicanum*, and we therefore expected that similar patterns of synteny should be apparent among the larger

set of genes that were annotated to AM13 and 14 (Smith and Voss 2006; Voss et al. 2011). These alignments confirmed that AM13 homologs were heavily enriched on chicken chromosomes GG26 and GG27 (Figure 3.6A). Similarly, AM14 homologs were heavily enriched on GG5. Closer examination of the distribution of AM14 homologs across GG5 revealed that AM14 homologs are distributed across two discreet regions, suggesting that GG5 was shaped by an ancestral fusion event and a subsequent pericentric inversion (Figure 3.6B), with the remainder of GG5 being orthologous to *Ambystoma* LG6 (Voss et al. 2011). Alternately, patterns of conserved synteny between AM14 and disjunct regions of GG5 might be explained by an ancient subtelomeric duplication (paralogs of spectrin beta chain occur in both subregions of the chicken genome) or possibly errors in genome assembly. Altogether, these analyses indicate that our chromosome-specific assemblies provide an accurate, though fragmentary, representation of *A. mexicanum* coding regions and their associated flanking sequences.

Our chromosome-targeted assemblies also provided an opportunity to gain further insight into the abundance and distribution of divergent repetitive element copies that were active in the past but have subsequently accumulated mutations. These more divergent copies are more amenable to assembly and shed some light on the past activity of transposable elements and are expected to be underrepresented in the analyses described above. Repeat content was assessed using RepeatModeler/RepeatMasker, which classified 22% of assembled sequences as corresponding to identifiable repeat classes (Smit and Hubley 2015; Smit et al. 2015). In total, 7% of the chromosomal assemblies could be assigned to known classes of repeats, although only 1.7% was assignable prior to de novo classification of salamander repetitive elements [Figure 3.7, 3.8, Supplementary Table 2 from (Keinath et al. 2015). Repeat counts for the separate chromosomal assemblies were remarkably similar and identified gypsy and LINE 1/2 elements as major contributors to the divergent repetitive fraction of the *A. mexicanum* genome (Figure 3.7). Repetitive elements identified by this approach were typically divergent from their consensus

sequence, with the typical element being ~20% divergent (Figure 3.8). In total, 21.4% of the assemblable fraction could be attributed to a known or *de novo* identified repeat class.

Discussion

Analyses of ~19x coverage shotgun sequence data from *A. mexicanum* provides an independent estimate of genome size that falls within the range of previous fluorometric estimates and further reveals that repetitive sequences are highly diverse in terms of sequence and copy number. In contrast to previous studies, we find that the majority of the genome consists of unique sequence (Britten and Kohne 1968; Morescalchi et al. 1974; Rosbash et al. 1974; Baldari and Amaldi 1976), at least with respect to assembly-relevant (short read) fragments. As might be anticipated, large genome size and complex repetitive environment present major challenges toward the development of a contiguous genome assembly for *A. mexicanum*. These initial studies indicate that chromosome-targeted sequencing presents an efficient strategy for simultaneously reducing assembly complexity and generating broad-scale scaffolding/anchoring information.

Our k-mer based analyses indicate as much as 62% of the genome (19.5 Gb) may be effectively single copy and that an additional 10 - 12% is potentially single-copy with respect to single chromosomes (i.e. at copy number less than ~20; Figure 3.1C). Perhaps not surprisingly, alignment-based analyses yield smaller estimates for the single-copy fraction, yet at this scale ~12 Gb can be considered single copy (Figure 3.2B). However, it is important to recognize that the designation “repetitive” is only relevant to its operational definition and that the designations “single-copy”, “low-copy” and “repetitive” are perhaps more relevant to the computational task of assembling genomes than they are predictive of the functionality of the underlying DNA segments (Treangen and Salzberg 2012).

Regardless of the method used to identify repetitive sequences, it appears that high-identity repeats only partly account for the dramatic difference in

genome size between salamanders and other tetrapod groups. The large size of the single/low-copy fraction is seemingly consistent with phylogenetic evidence suggesting that the axolotl's large genome size traces its origins to an ancient expansion event. Several repetitive elements are identifiable in the assemblable fractions of AM13 and AM14, and on average these repetitive elements were 20% divergent from their consensus sequence. As has been observed for other salamander species, gypsy and LINE 1/2 comprise major fractions of identifiable repeat classes, suggesting that these elements have undergone active transposition in the relatively recent evolutionary history of several salamander lineages (Sun et al. 2012; Sun and Mueller 2014), though notably LINE elements tended to be slightly more divergent than other repetitive element classes. With respect to large (and even human-sized) vertebrate genomes, it is likely that the vast majority of genomic DNA is derived from repetitive sequences (de Koning et al. 2011). The amplification of one or more repetitive elements almost certainly contributed significantly to the expansion of the ancestral salamander genome, although it may not be surprising that many of these sequences would have been heavily altered by mutations occurring over the last 180 - 200 million years. It seems plausible that this large volume of DNA may have provided raw material for the evolution of new functions in salamander genomes. Indeed, previous BAC sequencing studies have shown that salamander genes contain exceptionally long introns (which are transcribed into RNA) and that these introns contain a greater number and diversity of potentially functional secondary structures than their human counterparts (Smith et al. 2009).

The preliminary assemblies presented here yielded hundreds of Mb of sequence data from two chromosomes that are scaffolded at an ~300 bp scale and anchored to individual chromosomes. We anticipate that laser-capture sequencing approaches will provide important information as computational resources and sequencing strategies continue to evolve. Chromosome-targeted sequencing approaches provide two major benefits with respect to the assembly of large genomes, namely large-scale anchoring and partitioning of the assembly into computationally tractable subsets. We anticipate that the generation of

chromosome-targeted sequence data will become increasingly useful as a tool for subdividing genome assemblies, especially as it becomes possible to incorporate long read chemistries and associated algorithms into modern assembly pipelines (Roberts et al. 2013; Laszlo et al. 2014; Lee et al. 2014).

The results presented here also demonstrate the general feasibility of amplifying and deeply sequencing material from individual chromosomes that have been imaged and physically captured from the surface of a slide. Analyses of our shotgun datasets and draft assemblies illustrate the sensitivity and specificity of the approach and shed new light on the structure and gene content of the *A. mexicanum* genome. For example, these analyses confirm that *Ambystoma* LG14 corresponds to a single chromosome and improved the meiotic map by establishing that markers from LGs 15 and 17 should be coalesced into a single group, as previously proposed (Voss et al. 2011). We anticipate that chromosome-specific libraries will continue to yield critical information for the hierarchical processes of scaffolding and assembling of the remainder of the *A. mexicanum* genome. We also anticipate that current assemblies will be immediately useful for studies that leverage *A. mexicanum* for basic and biomedical research, including the identification of proximate promoters and intron/exon boundaries, the development of molecular probes and the design of targeted mutagenesis constructs.

Based on previous cytogenetic observations, we anticipate that the smallest salamander chromosomes should be slightly larger than the largest human chromosomes, likely exceeding 250 Mb (Callan 1966; Macgregor 1978), indicating that our laser capture/amplification approach will be applicable to a diversity of organisms, including those with genomes that are substantially smaller than that of *A. mexicanum*. Previous studies sequenced material from small numbers of pooled chromosomes that were captured from slides (Seifertova et al. 2013) or within microfluidic devices (Voskoboynik et al. 2013), which also provide invaluable scaffolding/anchoring information. The techniques employed here expand on these previous studies by circumventing the need to generate pooled samples, while generating deep sequence coverage of target

molecules. Because our protocol uses commercially available reagents and reactions are performed at microliter scale, the approach should be feasible for any lab that has access to a laser capture microscope and standard laboratory equipment. The general approach outlined here can be readily adapted to a diversity of biological questions, including genomic characterization of microscopically identifiable cells (e.g. cancer or germ cells) or the development of chromosome-scale scaffolds for organisms that are not amenable to meiotic mapping or laboratory culture.

Materials and methods

Ethics

All methods related to animal use were performed in accordance with AAALAC guidelines and regulations, under supervision of Division of Laboratory Animal Resources. Tissue collection was performed in accordance with protocol number 01087L2006, which was approved by the University of Kentucky Office of Research Integrity and Institutional Animal Care and Use Committee.

Generation and analysis of shotgun sequence data

Library preparation and shotgun sequencing (Illumina HiSeq 2000) were outsourced to Hudson Alpha Institute for Biotechnology (Huntsville, Al). Resulting sequences were filtered to remove sequencing adapters using Trimmomatic (Lohse et al. 2012) and common contaminants (e.g. phiX) using Bowtie 2 (Langmead et al. 2009). Initial k-mer analyses were performed using several values of k, as implemented by jellyfish (Marcais and Kingsford 2011), and the final k-mer distribution (used to estimate genomic parameters and perform error correction) was calculated using Blue (Greenfield et al. 2014). K-mer based estimates of sequence coverage, genome size and the size of the single copy fraction were generated using the method of Li et al (Li et al. 2010). K-mer based estimates of genome size assume a symmetrical sampling distribution for single copy regions and account for k-mer undersampling at the ends of short reads.

Alignment-based analyses were performed by mapping individual WGS reads to assembled BAC sequences (GenBank accession numbers: 194293375 - 194293390, 325260854, 325260856, 325260858, 325260859, 325260861, 325260863, 325260865 and 325260867) using bwa mem (v.0.7.10) with default parameters (Li 2012). Alignment files were filtered to remove unaligned reads using samtools (v.1.2), (Li and Durbin 2009) and both quality filtering and calculation of read depths were performed using sambamba (v0.5.4) (Tarasov et al. 2015).

Preparation of chromosomes

One hundred eggs from a wildtype axolotl cross were obtained from the *Ambystoma* Genetic Stock Center (AGSC) and maintained at 18°C incubator until they reached neurula stage (stage 17) (Schreckenberg and Jacobson 1975). Embryos were manually dechorionated with fine tip dissecting forceps and treated with 0.1% colchicine in 10% Holtfretter's solution (Armstrong et al. 1989) for 48 hours at 18°C to promote the accumulation of metaphases. Colchicine-treated embryos were washed with fresh 10% Holtfretter's solution then disaggregated using a Dounce homogenizer with loose pestle in a 0.075 M KCl solution (about 5 passes). After 45 minutes the swollen cells were fixed with 3:2 methanol:glacial acetic acid and stored in a -20°C incubator. Fixed cells were spread on UV-treated membrane slides (0.17 mm PET Zeiss 415190-9071-000 for library "A" and 1.0 mm PEN Zeiss 415190-9041-000 for all other libraries) by pipetting 100 □l of fixed cell suspension directly onto the surface of the slide. Slides were pretreated by inversion over a steam bath for seven seconds immediately prior to cell spreading. Following spreading, slides were immediately placed in a steam chamber at approximately 35°C for one minute, then dried by placing on a 60°C hot plate for five minutes. The chromosomes were stained by immersion in a modified Giemsa stain (Sigma-Aldrich GS500-500ML: 0.4% Giemsa, 0.7g/L KH_2PO_4, 1.0g/L Na_2HPO_4) for 2 minutes, rinsed in 95% ethanol, distilled water, and then air dried.

Laser capture microdissection (LCM)

Single chromosomes were microdissected using a Zeiss PALM Laser Microbeam Microscope at 40X magnification. Microdissected chromosomes were pressure catapulted into clear adhesive cap tubes (Zeiss 415190-9191-000) and immediately processed through amplification.

Preparation of amplified DNA

Dyads were released into solution by incubating overnight at 55°C with 10µl of a chromatin digestion buffer (1mM EDTA, 20mM TRIS pH 8.0, 0.2 mg/ml Proteinase K, 0.001% Triton X, in nuclease free water) added to the cap of the tube. Digested chromatin samples were briefly centrifuged and heat-treated for 10 minutes at 75°C and 4 minutes at 95°C to inactivate Proteinase K.

Chromosomal DNA was amplified using a Rubicon PicoPlex Whole Genome Amplification (WGA) kit (R30050) following standard manufacturer protocol, but substituting chromatin digestion buffer (above) for the cell extraction mix. The concentration and size distribution of amplified fragments were assayed using an Agilent 2100 Bioanalyzer (Agilent DNA 12000 Kit 5067-1508), and samples with less than 9 ng/µl were further amplified by performing two additional annealing/extension cycles. As an internal negative control, a piece of empty membrane was processed with each set of chromosome samples and run on the Bioanalyzer. Following initial quality control, 12 libraries were selected for outsourced sequencing on an Illumina HiSeq 2000 (Hudson Alpha Institute for Biotechnology, Huntsville, Al).

Sequence analysis of amplified DNA

Leader sequences and common contaminants (e.g. phiX) were removed using Trimmomatic (Lohse et al. 2012) and the resulting reads were aligned to model transcripts of the *Ambystoma* linkage map (Voss et al. 2011) or human reference genome, to identify nearly exact matches. To assess sequence coverage of the linkage map, chromosomal reads were mapped to model transcripts using the Burrows Wheeler Aligner (single-end mapping via the BWA-MEM algorithm) (Li

and Durbin 2009). Reads were also aligned to the human genome using Bowtie 2 (paired-end mapping) (Langmead et al. 2009), in order to assess the degree to which off-target sequences might contribute to chromosomal fragment libraries. Concordantly mapping reads were considered potential contaminants.

Assemblies were generated usingSOAPdenovo2 (Luo et al. 2012). Assemblies of whole genome shotgun data were attempted using “pregraph” and “sparse pregraph” methods for constructing de Bruijn graphs, although neither approach yielded an assembly. Several iterations of chromosome specific assembly were performed, with the best assemblies employing error corrected data (Greenfield et al. 2014) and broader coverage cutoffs to account for amplification bias during library construction (i.e. -c 0.05 -C 20) (Luo et al. 2012). Alignments between *A. mexicanum* assemblies and reference transcripts were performed using megablast (Zhang et al. 2000) and alignments between *A. mexicanum* reference transcripts and the chicken genome were performed using blast (tblastx) (Altschul et al. 1997). Alignment-based estimates of genome coverage and repeat content were performed as described above.

Preparation and labelling of C$_O$T DNA

DNA was isolated using standard phenol/chloroform extraction(Sambrook and Russell 2001). DNA was adjusted to a concentration of 1,000 ng/µl in 1.2x SSC (1 ml), then sheared and denatured by heating to 120 °C for 2 min in a prewarmed aluminium block. Following denaturation, reannealing was performed by immediately placing the tube at 60 °C for 15 min and then immediately on ice. Remaining single stranded DNA was removed by adding S1 nuclease (Thermo Scientific # EN0321) to a final concentration of 100 U per 1 mg of DNA in 1X buffer, followed by incubation at 42 °C for 1 hr. C$_O$T DNA was then purified via isopropanol precipitation (Sambrook and Russell 2001), reconstituted in TE buffer and labelled via degenerate oligonucleotide PCR. Briefly, 0.5 □g template DNA was amplified using the primer CCGACTCGAGNNNNNNATGTGG, GoTaq® DNA Polymerase and buffer, 200 □M d[A,C,G]TP, 100 □M dTTP and 100 □M Cy3-dUTP (Enzo) at a 25 □l reaction volume. Thermal cycling conditions

were: 6 minute initial denaturation at 96°C, 30 cycles of 94°C for 1 minutes, 56°C for 1.5 minutes and 72° for 2 minutes, followed by a final elongation 72°C for 8 minutes.

Classification of divergent repetitive elements

To characterize that divergent fraction of repetitive elements within the salamander genome (those not represented by high-count k-mers) we performed de novo searches for repetitive elements using RepeatModeler/RepeatMasker (Smit and Hubley 2015; Smit et al. 2015). Repetitive elements were identified de novo using RepeatModeler and combined data from the AM13 and AM14 assemblies. Final repeat annotations were made using a combined dataset of elements from RepeatModeler and all known vertebrate repetitive elements contained in the RepBase 20.02 libraries. Estimates of sequence divergence were generated using RepeatMasker and a database consisting solely of repeats that were identified in the AM13 and AM14 assemblies.

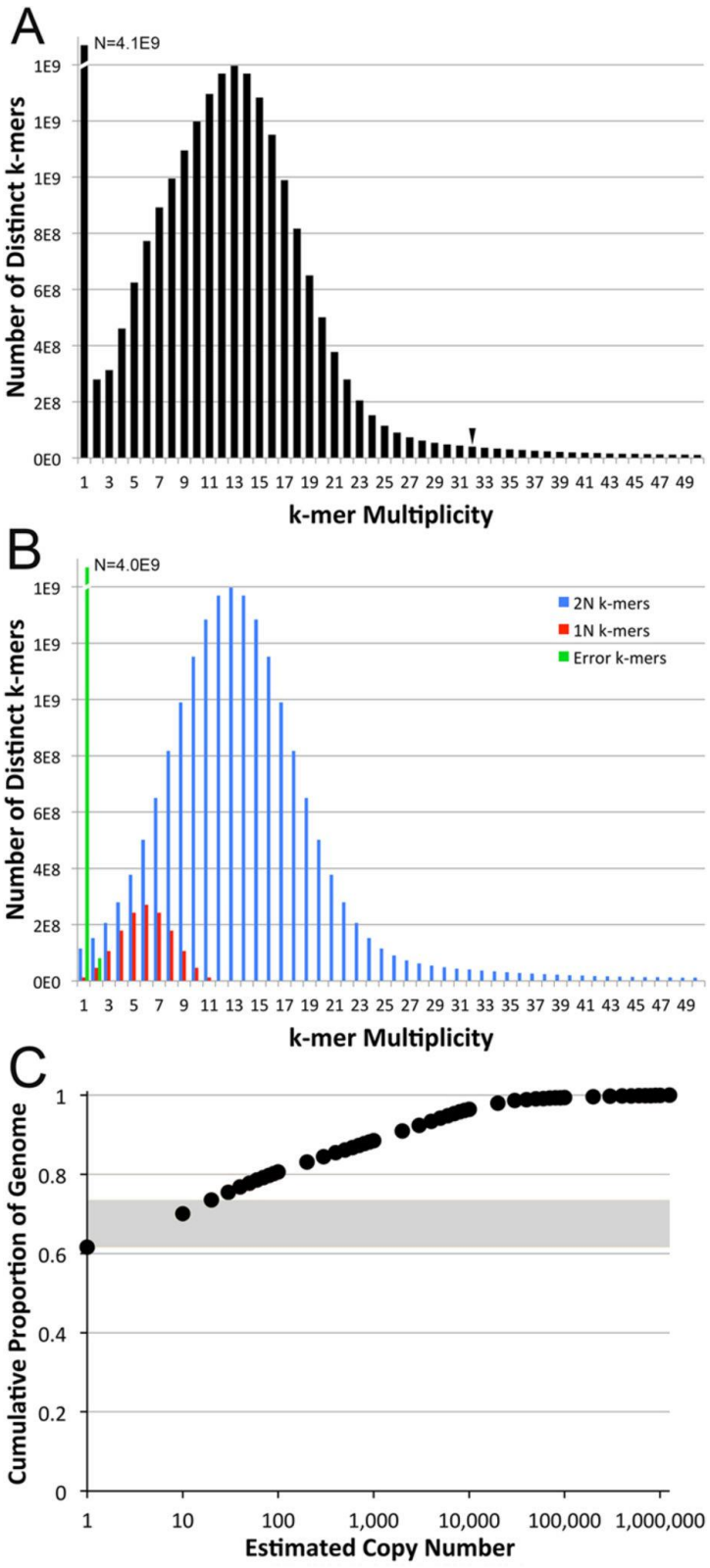
A
N=4.1E9
Number of Distinct k-mers
1E9
1E9
1E9
8E8
6E8
4E8
2E8
0E0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
k-mer Multiplicity
B
N=4.0E9
2N k-mers
1N k-mers
Error k-mers
Number of Distinct k-mers
1E9
1E9
1E9
8E8
6E8
4E8
2E8
0E0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
k-mer Multiplicity
C
Cumulative Proportion of Genome
1
0.8
0.6
0.4
0.2
0
1
10
100
1,000
10,000
100,000
1,000,000
Estimated Copy Number

Figure 3.1. Distribution of 31-mer frequencies among >0.6 terabases of quality filtered sequence data generated from a single female *A. mexicanum*.

A) The observed distribution is humped with a peak at k-mer multiplicities of 13 and 14 (estimated mean of 13.50), presumably corresponding to k-mers that were sampled from the single-copy fraction of the genome. The k-mer multiplicity corresponding to 3 standard deviations above the mean of the single copy distribution (33.67) is marked by an arrow. **B)** Decomposition of the observed distribution assuming symmetrical single-copy (diploid: 2N) and allelic (1N) k-mer distributions. The sum of all bins at a given multiplicity in panel B is equal to the observed multiplicity presented in Panel A. **C)** Low-copy k-mers account for the majority of *Ambystoma* shotgun sequence data and k-mers present at increasing copy number represent decreasing fractions of the shotgun dataset, suggesting that the diversity of repetitive sequences scales inversely with copy number. The region of the plot highlighted in grey represents copy number ranges that could plausibly exist at a copy number of ~1 per chromosome. The X-axis is plotted on a log scale to aid in visualization of patterns at lower estimated copy numbers.

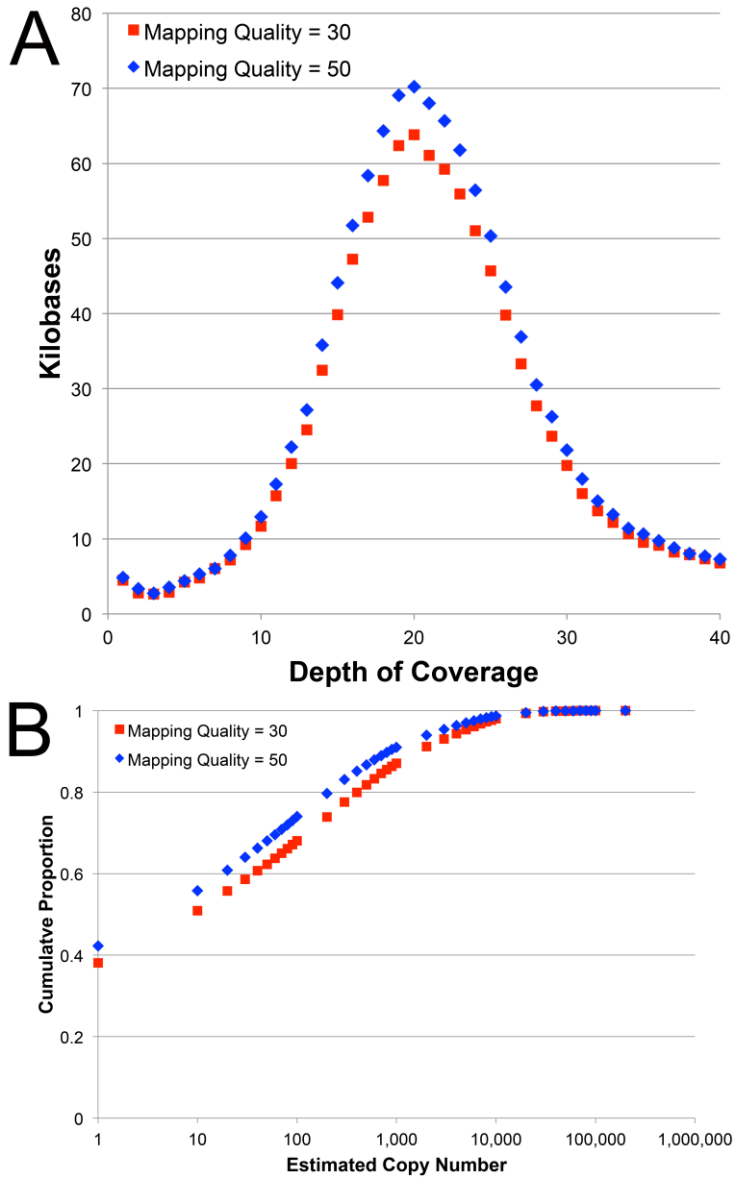

Figure 3.2. Estimation of sequence coverage and repeat content by alignment to assembled BAC clones.

A) The observed distributions are humped with peak depths of coverage between 19 and 20, consistent with estimates from analysis of k-mer frequencies. **B)** Low-coverage bases account for ~40% of *Ambystoma* BAC sequence data and bases present at increasing copy numbers represent decreasing fractions of the BAC sequences, further suggesting that the diversity of repetitive sequences scales inversely with copy number.

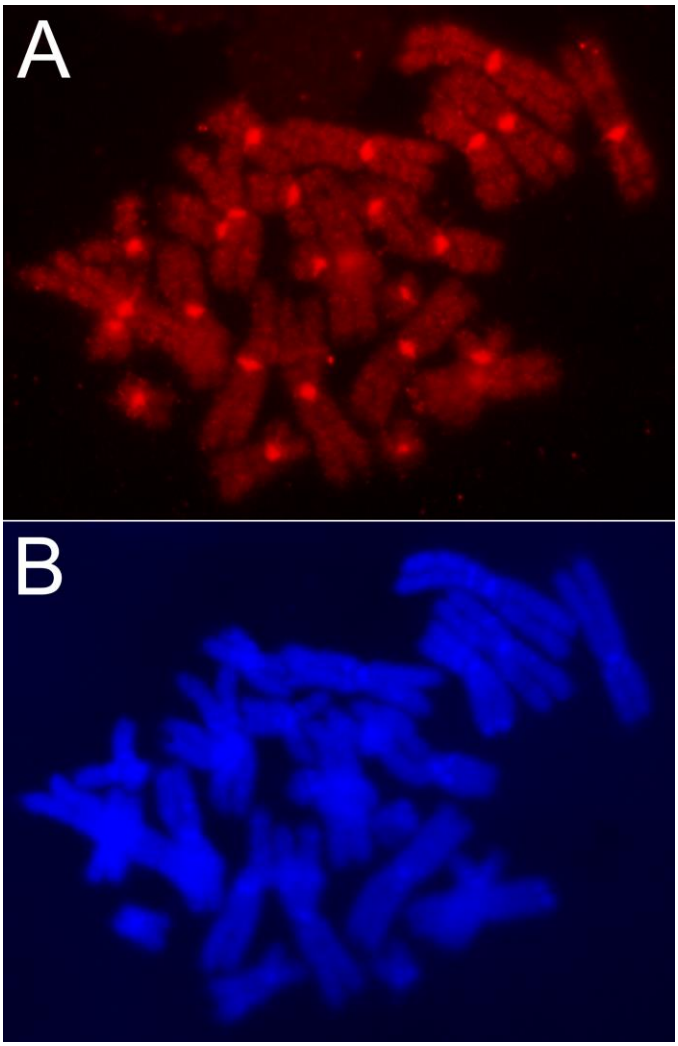

Figure 3.3. Distribution of repetitive elements in the axolotl genome. Chromosomes were hybridized with Cy3-dUTP labelled C_OT DNA (Panel **A**) and stained with DAPI (Panel **B**). This fraction of C_OT DNA contains the rapidly annealing (repetitive) portion of the genome and comprises ~45% of input DNA. Hybridization patterns show that repetitive DNA is heavily clustered at the centromeres and broadly distributed across all chromosomal arms.

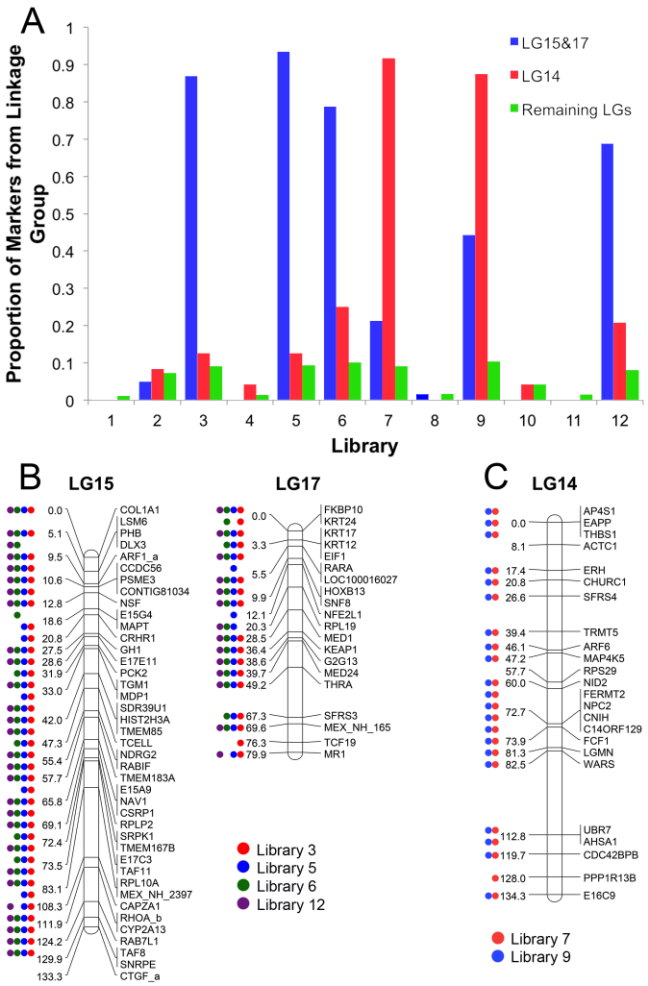

Figure 3.4. Mapping of reads generated by laser capture sequencing. Read mapping was used to assess the sensitivity and specificity of laser capture and amplification libraries. **A)** The proportion of *Ambystoma* markers with nearly identical reads recovered from chromosome-targeted sequencing. Markers from target vs. off target linkage groups are presented separately. **B)** The distribution of markers sampled from chromosome 13 (LGs 15 and 17) via targeted sequencing. Dots represent markers with mapped reads from each experimental series. Red, blue, green and purple dots denote markers that were sampled by reads (near perfect matches) from libraries 3, 5, 6 and 12, respectively. **C)** The distribution of markers sampled from chromosome 14 (LG 14) via targeted sequencing. Red and blue dots denote markers that were sampled by reads from libraries 7 and 9, respectively.

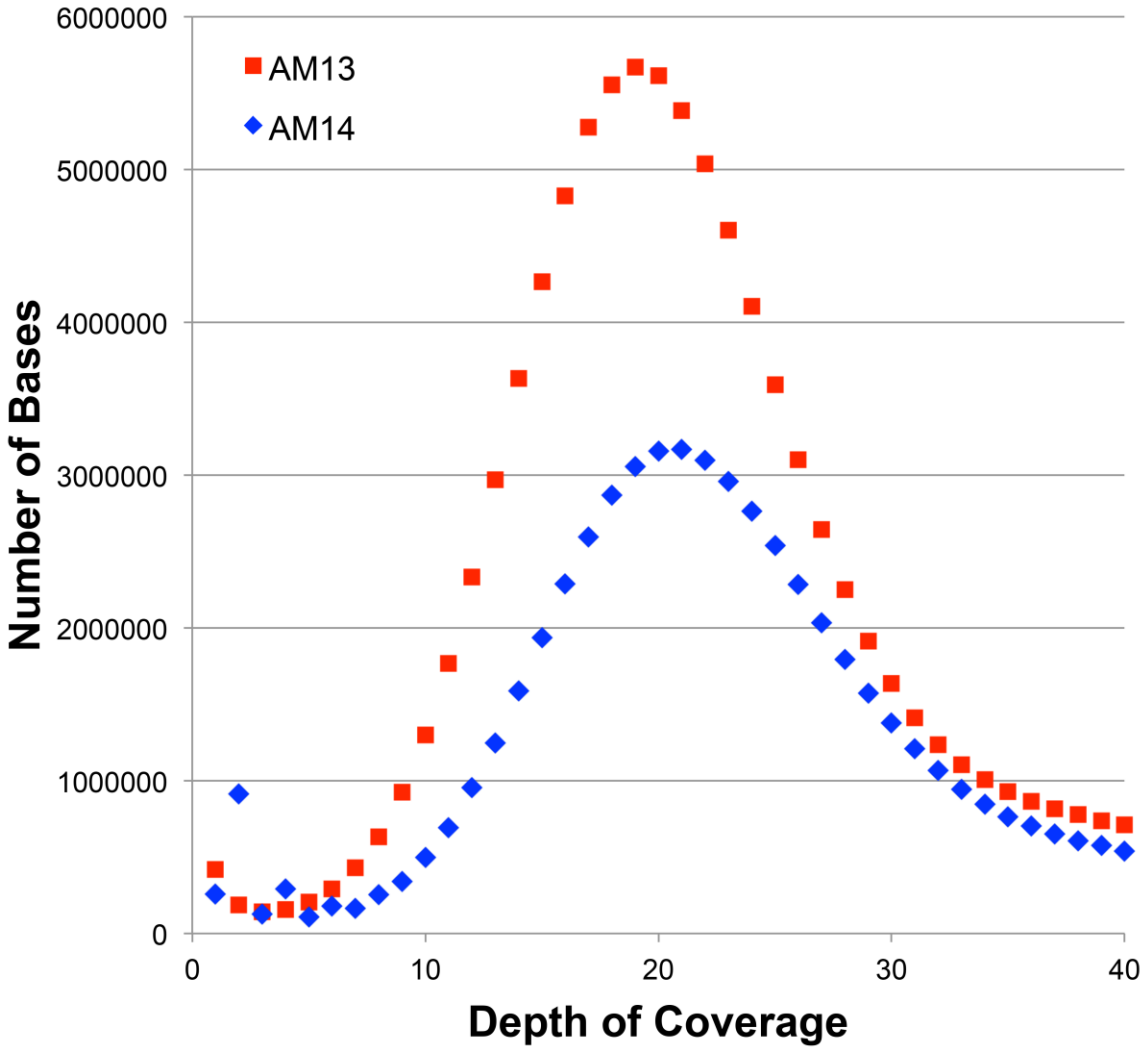

Figure 3.5. Estimation of coverage by alignment to assembled contigs from AM13 and AM14.

The observed distributions are humped with peak depths of coverage between 19 and 20, consistent with estimates from alignments to BAC clones and analysis of k-mer frequencies. MQ30 = data are filtered to include only alignments with a map quality >= 30, MQ50 = data are filtered to include only alignments with a map quality >= 50.

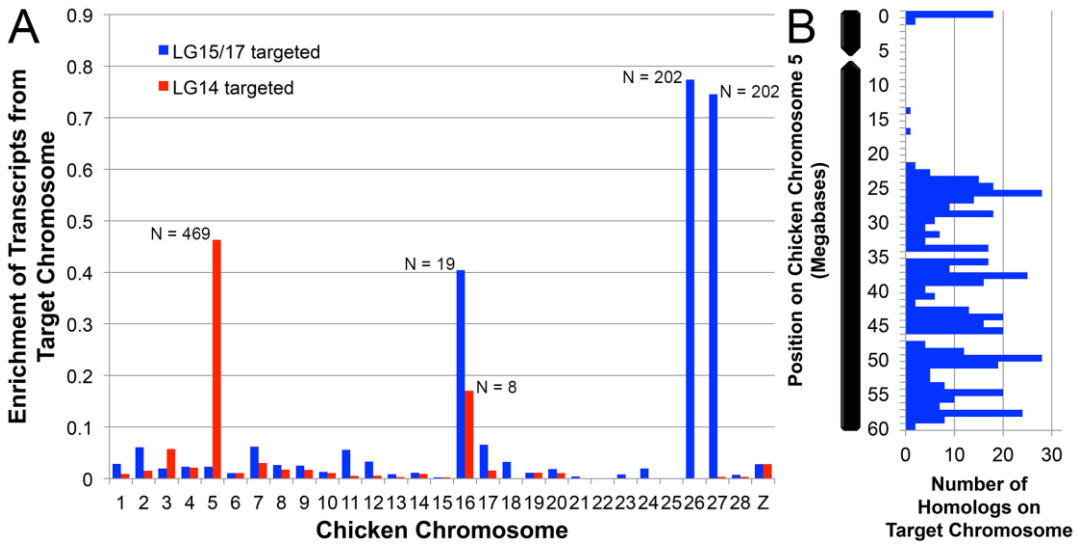

Figure 3.6. Conserved synteny between assembled *A. mexicanum* chromosomes and the chicken genome.

A) Tests for enrichment of AM13 (LG15/17 targeted) and AM14 (LG14 targeted) presumptive gene orthologs across all assembled chicken chromosomes. "Enrichment" is defined as the observed number of orthologs divided by the total number of genes that have been annotated to the chromosome (Cunningham et al. 2015). **B)** The distribution of AM14 orthologs along chicken chromosome 5 reveals a discontinuous distribution consistent with the interpretation that chicken chromosome 5 was shaped by an ancestral fusion event, and a subsequent pericentric inversion.

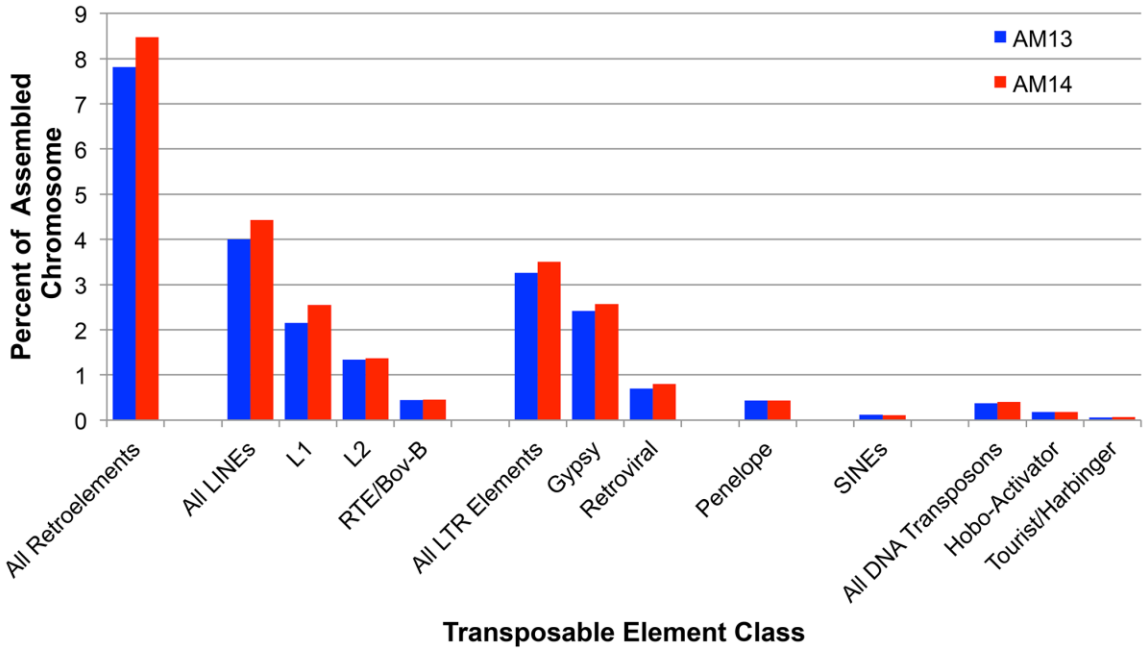

Figure 3.7. Summary of major repetitive element classes identified within assembled chromosomes.

Percentages are shown separately for the two chromosomal assemblies. LINEs (Long Interspersed Nuclear Elements), LTRs (Long Terminal Repeat), Penelope and SINEs (Short Interspersed Nuclear Elements) are retroelement subclasses. Hobo-Activator and Tourist/Harbinger elements are DNA transposon subclasses. L1, L2 and RTE/Bov-B elements are LINE subclasses. Gypsy and Retroviral elements are LTR subclasses.

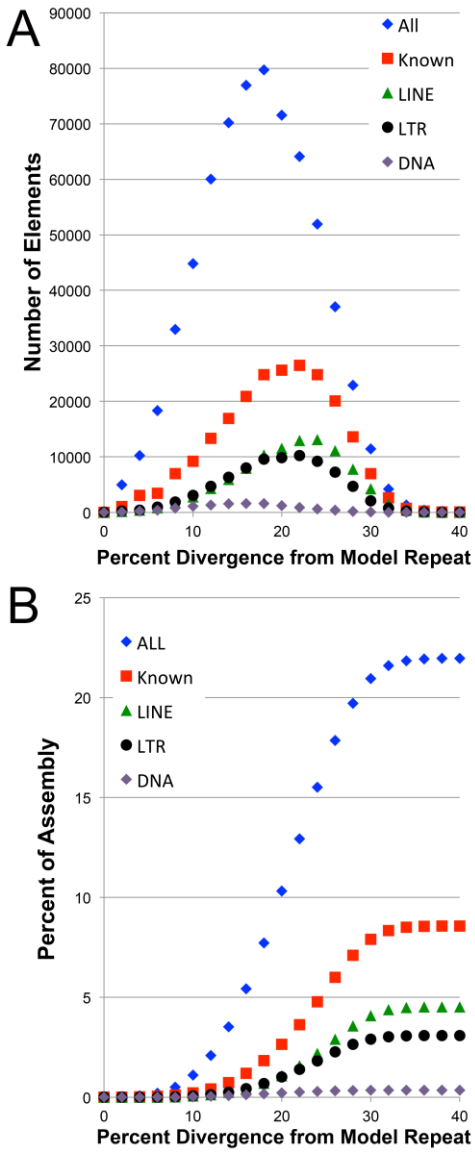

Figure 3.8. Diversity and abundance of repetitive elements in assembled scaffolds from AM13 and AM14.

A) Divergence between identified repeats and their RepeatMasker consensus sequence, using only information from *A. mexicanum* (model repeat). **B)** The cumulative contribution (by length) of these same repeat classes. In both panels, patterns are shown for several classes. Known elements are comprised of LINEs, LTRs, DNA elements and other classes that are present at lower abundances (see Figure 3.7). The class "All" consists of both known and unknown repeat classes.

Accession codes

Sequence data are deposited at the NCBI short read archives (http://www.ncbi.nlm.nih.gov/sra) under study number PRJNA269757. Assemblies are deposited at the NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank/) under accession number JXRH00000000.

Competing financial interests

The authors declare no competing interests.

Acknowledgements

We thank Rubicon Genomics and Hudson Alpha Institute for Biotechnology Genome Services Laboratory for their advice in the development of approaches to generating sequence from amplified material. This work was funded by grants from the National Institutes of Health (NIH) (R24OD010435 and EY10540) and Department of Defence (DOD) (W911NF1110475). The axolotl for this study was provided by the Ambystoma Genetic Stock Center, which is currently funded by the NIH (P40OD019794) and previously by the National Science Foundation (NSF) (DBI-0951484). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of NIH, DOD or NSF.

Contributions

J.S. and S.R.V conceived of the study. J.S., M.K., S.R.V and P.T. wrote the manuscript. M.K. performed chromosome amplification and sequencing experiments. V.T. performed in situ hybridizations. N.T., J.S. and M.K. performed computational analyses.

CHAPTER FOUR

A LINKAGE MAP FOR THE NEWT *NOTOPHTHALMUS VIRIDESCENS*: INSIGHTS IN VERTEBRATE GENOME AND CHROMOSOME EVOLUTION

Reproduced from: Keinath MC, Voss SR, Tsonis PA & Smith JJ (2017) A Linkage Map for the Newt *Notophthalmus viridescens*: Insights in Vertebrate Genome and Chromosome Evolution. *Developmental Biology*. DOI: 10.1016/j.ydbio.2016.05.027

Abstract

Genetic linkage maps are fundamental resources that enable diverse genetic and genomic approaches, including quantitative trait locus (QTL) analyses and comparative studies of genome evolution. It is straightforward to build linkage maps for species that are amenable to laboratory culture and genetic crossing designs, and that have relatively small genomes and few chromosomes. It is more difficult to generate linkage maps for species that do not meet these criteria. Here, we introduce a method to rapidly build linkage maps for salamanders, which are known for their enormous genome sizes. As proof of principle, we developed a linkage map with thousands of molecular markers (N=2349) for the Eastern newt (*Notophthalmus viridescens*). The map contains 12 linkage groups (152.3-934.7cM), only one more than the number of chromosome pairs. Importantly, this map was generated using RNA isolated from a single wild caught female and her 28 offspring. We used the map to reveal chromosome-scale conservation of synteny among *N. viridescens, A. mexicanum* (Urodela), and chicken (Amniota), and to identify large conserved segments between *N. viridescens* and *Xenopus tropicalis* (Anura). We also show that *met1,* a major effect QTL that regulates the expression of alternate metamorphic and paedomorphic modes of development in *Ambystoma,* associates with a chromosomal fusion that is not found in the *N. viridescens* map. Our results shed new light on the ancestral amphibian karyotype and reveal specific fusion and translocation events that shaped the genomes of three amphibian model taxa. The ability to rapidly build linkage maps for large salamander genomes will enable genetic and genomic analyses within this

important vertebrate group, and more generally, empower comparative studies of vertebrate biology and evolution.

Introduction

From the dawn of science to present, amphibians have greatly enriched our understanding of biology. In part, this reflects the amazing diversity of phenotypes that are expressed within and among amphibian species. For example, just among salamanders three different life cycles are observed, including direct development, indirect development/metamorphosis, and paedomorphosis, which is not observed in any other vertebrate group. Early studies of these different life cycles yielded theories for how phenotypes arise and evolve as a result of changes in developmental timing –heterochrony (Gould 1977). More recently in the example of salamander paedomorphosis, heterochrony was shown to depend upon thyroid-hormone responsive alleles that segregate among quantitative trait loci (QTL) (Voss and Shaffer 1997; Voss and Smith 2005; Page et al. 2013). Almost certainly this model system will be expanded upon in the future to enrich understanding of evolution, as will countless other amphibian models that are best suited for developmental studies.

In addition to the identification of developmental timing QTL in the example above the *Ambystoma* linkage map has permitted the identification of genomic regions contributing to sex determination and tail regeneration (Smith and Voss 2009; Voss et al. 2013a). Linkage maps are valuable resources that not only facilitate the identification of QTL and mapped-based cloning, they are also useful for reconstructing the evolution of chromosomal rearrangements and chromosome number. Comparative mapping studies have leveraged the *Ambystoma* linkage map to reveal insights about vertebrate genome evolution (Voss et al. 2001; Smith and Voss 2006; Voss et al. 2011). These studies found extensive conservation of microchromosomes and chromosomal segments from the tetrapod ancestor and independent decreases in chromosome number among lineages leading to *Ambystoma, Xenopus* and mammals. Similarly, comparative FISH mapping within the genus *Xenopus* identified a relatively small

number of rearrangement events that have occurred over the last ~60 MY of evolution in the *Xenopus* lineage, including one inversion in *X. tropicalis*, inversions on two paralogous chromosomes in *X. laevis* and one fusion (involving orthologs of *X. tropicalis* chromosomes 9 and 10) that predated a whole genome duplication in the *X. laevis* lineage (Uno et al. 2013). While these studies have shed light on the deep evolutionary history of vertebrate genomes and shallower patterns within the *Xenopus* lineage, they only represent a small number of lineages with relatively derived karyotypes. As such, data from additional species are needed to reconstruct chromosome evolution in amphibian and basal vertebrate lineages.

Amphibian genomes exhibit substantial variation in DNA content, with salamander lineages presenting exceptionally large genomes relative to other extant tetrapods. Estimates of salamander genome sizes range from 10 - 120 gigabases (Smith et al. 2009; Gregory 2015), with the smallest salamander genome exceeding the size of the largest anuran genome. Because all salamanders have large to extremely large genomes, it seems likely that genome size increased in the basal lineage that gave rise to all extant salamanders, between ~300 and 180 MYA (Zhang and Wake 2009; Hedges et al. 2015; Keinath et al. 2015). This increase in size is thought to reflect an ancient expansion of repetitive DNA sequences (Keinath et al. 2015). This expansion appears to have affected salamander genome structure in a global sense because introns, intergenic regions and linkage map size are dramatically expanded in the axolotl (Smith et al. 2005a; Smith et al. 2009; Voss et al. 2011). Interestingly, genome expansion does not appear to influence the rate of chromosome evolution. Relatively few interchromosomal rearrangements are predicted for the axolotl genome during evolution, in stark contrast to what is predicted for mammalian chromosomes and especially rodents (Smith and Voss 2006).

While analysis of a few amphibian genomes has shed critical light on the biology and evolutionary history of vertebrate genomes, it is important to recognize that our current understanding of amphibian (and vertebrate) genome

evolution is based on information from a relatively small number of species. This is due, in part, to the inherent challenges of generating controlled crosses for amphibians because many species are not amenable to the establishment of laboratory stocks. Also, in the case of salamanders, the construction of gene-anchored linkage maps is made more difficult by the enormous size of the genome. Large salamander genomes contain approximately the same number of genes as other vertebrates (Smith et al. 2009; Gregory 2015). To move toward a more comprehensive understanding of vertebrate karyotype evolution, we sought to develop an approach that permits the rapid construction of robust gene-anchored maps and circumvents the need to maintain living laboratory stocks.

Here we report the first high-density gene-anchored linkage map for the Eastern newt (*Notophthalamus viridescens:* the second for any salamander) and use this map to refine our understanding of the tempo and mode of karyotype evolution in the amphibian lineage. The ancestral lineages that gave rise to newt and *Ambystoma mexicanum* (axolotl) diverged ~150 million years ago (MYA) and ~150 MY after the frog/salamander divergence (Figure 4.1). The construction of this map leveraged RNA sequencing data to identify a large number of SNPs segregating within genic regions, which permit the phasing of polymorphisms and the construction of a map from the offspring of a single outbred female. Analysis of conserved syntenies between newt and other vertebrates revealed a small number of fusions that occurred prior to the divergence of ancestral amphibian lineages that gave rise to frogs and salamanders, in the axolotl/newt ancestor and still others that occurred after the divergence of salamander and newt. Notably one fusion that occurred in the *Ambystoma* lineage co-localized with a major effect QTL that regulates developmental timing the expression of alternate, metamorphic vs. paedomorphic, life histories (*met1*: (Voss and Shaffer 1997; Voss and Smith 2005; Page et al. 2013).

Materials and Methods:

Newt collection, embryo sampling, RNA extraction and sequencing

Male and female newts were collected from a pond near the head of Gray's Arch trail in Kentucky's Red River Gorge under Kentucky Department of Fish and Wildlife educational collection permit number SC1311325. Females were kept in 5 gallon buckets containing water and live plants from the same pond. Eggs were collected each morning and kept in crystallization dishes with 20% Holtfreter's solution at 20°C. One of these females produced 29 eggs, all of which developed normally until the date they were sampled at approximately 14 days of development (prior to hatching). Embryos were dechorionated using fine tip forceps and immediately placed in 1.7 ml tubes with RNA*later*® (ThermoFisher Scientific AM7020). RNA was extracted from several tissues from this female (blood, liver, spleen, heart, eyes, brain and skin from the dorsal surface of the head) and from individual whole embryos, using standard trizol extraction.

The RNA extracted was assessed for quality on a bioanalyzer (Agilent 2100 Bioanalyzer; Agilent RNA 6000 Nano Kit). Samples were sent to Hudson Alpha Institute for Biotechnology (Huntsville, AL) where libraries were prepared and barcoded with 36 unique tags and sequenced (100bp paired end reads) on a single lane using Illumina HiSeq 2000. One embryo sample yielded less than 500ng of RNA and was underrepresented in the resulting sequence dataset. Sequence data are deposited at the NCBI short read archives (http://www.ncbi.nlm.nih.gov/sra) under accession number SRP067290, BioProject # PRJNA305738.

Genotyping and linkage analysis

To identify and genotype polymorphisms segregating within this family, we mapped RNAseq reads to the published newt transcriptome using bwa-0.7.5a (Li and Durbin 2009), performed de-duplication using picard-tools-1.97 (http://broadinstitute.github.io/picard) and genotype calling using the *HaplotypeCaller* pipeline in GATK-2.7 (McKenna et al. 2010; Looso et al. 2013).

Genotypes were post-filtered to include only those markers with presumptive maternal or paternal polymorphisms segregating among the 28 embryos sequenced for this study, that had variant quality scores greater equal to or greater than 100 and that were represented by a minimum of 4 reads in all individuals. For each transcript, the polymorphism with the highest variant quality score was used in subsequent linkage analyses.

Although maternity was known for all individuals, the matings that produced these offspring presumably occurred before the maternal individual was placed in isolation. To test for multiple paternities, we calculated the Jaccard distance for multilocus paternal genotypes between each pair of the 28 offspring using the “Similarity of Individuals” summary generated by JoinMap (Van Ooijen 2011). These analyses revealed two distinct clusters, one cluster consisting of 18 individuals and a second consisting of 10 individuals, indicative of mixed parentage [paternity; Supplementary Table1 from (Keinath et al. 2017)]. These relationships were further verified by calculation of relatedness statistics using the *relatedness2* function of VCFtools (Danecek et al. 2011) [Supplementary Table 1 from (Keinath et al. 2017)]. As such, only the group of 18 presumably full sibs was used to estimate male meiotic recombination rates (the paternal map). All siblings were used to estimate recombination rates during female meiosis (the maternal map). Given the small number of meioses, the paternal map was only used for cross-validation of the maternal map.

Linkage analysis was performed via maximum likelihood mapping using JoinMap software package and default parameters, except that the number of optimization rounds was increased to ten to improve ordering of markers (Stam 1993; Van Ooijen 2011). This approach yielded 3,142 markers (transcripts) that segregated one or more maternal polymorphisms, of which 2,349 could be confidently placed on the maternal map and 6,546 markers that segregated candidate paternal polymorphisms. Linkage groups (LGs) containing at least 50 markers linked at a minimum LOD score (log of odds) of 3.0 were considered in downstream analyses. Linkage groups were manually curated to break linkages at >30 cM, except in one case where markers within a large syntenic block were

within 35 cM and synteny was supported by sequence data from individual laser captured chromosomes. Patterns of linkage were further evaluated using the software suite Lep-MAP2 using default settings and increasing the number of optimization rounds to ten in order to best match parameters used in JoinMap (Rastas et al. 2016). These analyses largely recapitulate patterns of linkage generated by JoinMap (including suspect joins that were broken by manual curation), except that bi-allelic and tetra-allelic markers from the same linkage group (and component transcripts) were frequently assigned to distinct Lep-MAP2 linkage groups [Supplementary Tables 2 and 3 from (Keinath et al. 2017)].

A similar F1 outbred mapping strategy has been successfully used to generate dense meiotic maps for several other species using the RAD-seq approach (rather than RNAseq data) to identify polymorphisms from genomic DNA (Amores et al. 2011; Palaiokostas et al. 2013; You et al. 2013; Kai et al. 2014; Smith and Keinath 2015). Subsampling experiments have demonstrated that this approach permits the construction of accurate meiotic maps using as few as 20 offspring (Amores et al. 2011).

Laser capture chromosome sequencing

To more accurately assess the LGs and inform potential breaks in the larger maternal LGs, we performed low-coverage shotgun sequencing of a small number of laser captured chromosomes. Chromosome preparations, spreads, staining, capture and amplification were generated with the same methods described for axolotl (Keinath et al. 2015) with a few modifications: newts were dechorionated shortly after late neurula stage of development, embryos were dechorionated in 40% Holtfreter's solution to account for the higher internal pressure of chorionic fluid, and chromosome spreading was performed in a high-humidity chamber held at 60°C (Keinath et al. 2015). We collected 24 individual chromosomes using laser capture microdissection as previously done in salamander (Keinath et al. 2015). Library preparation was performed using a Rubicon whole genome amplification (WGA) PicoPLEX™ DNA-seq (R300381), a bioanalyzer was used to check for presence of DNA, and resulting amplicons

were sequenced on a HiSeq 2000 platform at Hudson Alpha Institute for Biotechnology. Sequences were demultiplexed and trimmed to remove leader sequences that were incorporated during amplification and common contaminants (e.g. PhiX) using Trimmomatic 30.2 (Bolger et al. 2014). These remaining high quality reads were aligned to the human genome using a paired-end mapping mode in Bowtie 2 (Hormozdiari et al. 2011) to detect potential human contaminants. Reads that aligned to human were removed from those libraries in which more than 10% of total reads mapped to human. The trimmed sequence libraries consisted of 210 million read pairs, averaging 8.8 million read pairs per library (minimum = 0.1 million, maximum = 15.4 million). Sequence data are deposited at the NCBI short read archives (http://www.ncbi.nlm.nih.gov/sra) under accession number SRP072907, BioProject # PRJNA317478. In order to evaluate these chromosomal libraries, reads were aligned to the newt transcriptome using single-end mapping mode via BWA-MEM (v.0.7.10) (Li and Durbin 2009).

Comparative mapping/conserved synteny

The complete set of transcripts that were assigned to newt LGs was used to query the set of transcripts that have been placed on the *Ambystoma* meiotic map, and masked genome assemblies for chicken and *X. tropicalis*, using tblastx (Altschul et al. 1990). The best matching sequence was considered the most likely ortholog (broad sense) for each newt transcript provided: 1) alignment was ≥ 40 amino acids, 2) alignment bitscore was ≥100 and 3) the aligning sequences had at least 40% amino acid identity. The percent identity cutoff was chosen by examining the distribution of percent identity statistics among all best blast hits between newt and chicken [Supplementary Figure 1 from (Keinath et al. 2017)]. Counts of orthologs on each pairwise combination of linkage groups or chromosomes were compared to expected values derived from randomly distributing orthologies across linkage groups and chromosomes with the same number of mapped loci. In order to detect statistically significant regions of conserved synteny (and control for potential ortholog miscalls), the distribution of

orthologs was assessed using chi-square tests with Yates' correction for continuity and Bonferroni corrections for multiple testing as previously described (Smith and Keinath 2015).

Results and Discussion

One lane of Illumina RNAseq data (2 x 100 bp) was generated for 35 barcoded and pooled samples; 28 newt embryos and 7 adult tissues. In total, this sequencing run yielded 43.7 Gb of raw sequence data. These sequences were used to identify segregating polymorphisms by alignment to 120,922 previously characterized newt transcripts (Looso et al. 2013). A total of 6,460 candidate segregating maternal polymorphisms were identified on 3,142 transcripts. A total of 2,349 transcripts could be confidently placed on a set of 8 LGs with at least 50 markers linked at a minimum LOD score (log of odds) of 3.0 (Looso et al. 2013). After initial map construction, newt linkage groups were examined to identify gaps exceeding 30 cM. In cases where conserved syntentic blocks or low-pass laser-capture chromosome sequencing data did not support linkage, linkage analyses were performed separately for the segments flanking these gaps. For one large linkage group, laser-capture chromosome sequencing data indicated that both ends of the linkage group were portions of the same chromosome, exclusive of an ~700 cM internal segment. Reanalysis yielded two well-supported linkage groups [Figure 4.2, LG1 and LG5, Supplementary Figure 2 from (Keinath et al. 2017)].

The curated linkage map resolves 12 linkage groups; a number one greater than the haploid number of chromosomes in the newt karyotype and spans a total of 6,161.9 cM, with the largest group spanning 934.7 cM and the smallest spanning 152.3 cM [Figure 4.2, Supplementary Table 4 from (Keinath et al. 2017)]. Although our sibship sampling strategy (and the reproductive biology of *N. viridescens*) permitted sampling of a modest number of meioses, our marker sampling strategy yielded dense sampling across chromosomes: the average number of recombinations between adjacent markers is less than 1

(0.73). This high marker density facilitates accurate map reconstruction using robust maximum likelihood algorithms, which leverage genotypic information from several adjacent markers to identify linkage groups and order markers, effectively minimizing the impact of genotyping errors (Stam 1993; Van Ooijen 2011). Cross validation of maternal and paternal genetic maps, patterns of conserved synteny and chromosome library alignments provided additional support for the grouping and ordering of markers that were incorporated into the final map (see below). In total, our low-coverage laser capture sequence data support six linkage groups as corresponding to discrete chromosomes (LGs 1, 2, 8, 9, 10 11), whereas other chromosomes could not be directly resolved at current sampling depths [Supplementary Figure 3 from (Keinath et al. 2017)]. The relatively large size of the newt linkage map and individual LGs is consistent with previous linkage analyses in another salamander species (*Ambystoma*) and microscopic observations of chiasmata in salamander oocytes (Callan 1966; Smith et al. 2005a; Voss et al. 2011). As such, the large recombinational size of the newt linkage map lends support to the idea that the ancestral salamander genome expansion resulted in a proportional increase in rates of meiotic recombination (Smith et al. 2005a; Voss et al. 2011).

Comparison of newt, *Ambystoma*, *Xenopus* and chicken genomes revealed extensive conservation of chromosomal segments across all taxa [Figure 4.2, Supplementary Tables 5-9 from (Keinath et al. 2017)]. Examining homolog sampling depth across chicken chromosomes reveals uniform patterns of coverage across most 10 Mb intervals (newt homologs were mapped for ~10% of chicken loci sampled per interval), with the exception of the distal region of chicken chromosome 2, which was sampled at ~1/2 the average frequency [Supplementary Figure 4 from (Keinath et al. 2017)]. As previously observed, many conserved segments correspond to large portions of chicken macrochromosomes or entire microchromosomes. Previous analyses showed that these segments and microchromosomes were derived from individual chromosomes that trace their ancestry at least to the common ancestor of all bony vertebrates (Voss et al. 2011; Venkatesh et al. 2014; Braasch et al. 2016).

By examining the distribution of these conserved segments in newt, *Ambystoma* and *Xenopus*, it is possible to reconstruct several evolutionary events that define the karyotypes of these three model amphibian taxa (Table 4.1).

In general, fusions in salamander and frog (*Xenopus*) lineages appear to be largely independent, with the exception of one ancestrally conserved linkage (GG4/Z). Notably, segments of GG4 and GGZ are syntenic on two gar chromosomes and on several scaffolds in coelacanth, indicating that this linkage may have existed in the common ancestor of all bony vertebrates as indicated by earlier comparative genomic studies (Voss et al. 2001; Amemiya et al. 2013; Braasch et al. 2016). Alternatively, this pair of chromosomes might have experienced recurrent fusions in several basal vertebrate taxa. The patterns of conserved synteny in amphibians indicate that a relatively small number of fusions (perhaps none) occurred within the ancestral amphibian lineage, an interpretation that is consistent with the observation of chicken-like karyotypes within basal frog and salamander lineages (Sessions 2008) and suggests that the relatively compact karyotypes of *Xenopus* (1N = 10) and newt (1N = 11) are largely the product of convergent evolution.

These findings appear to be consistent with cytogenetic studies that showed amphibian genomes to vary considerably in structure and content, including chromosome number, size and morphology (Morescalchi et al. 1974; Duellman and Trueb 1986; Green and Sessions 1991; Vinogradov 1998). Caecilian karyotypes vary from asymmetric-bimodal karyotypes with higher numbers of chromosomes to symmetric-unimodal karyotypes with fewer chromosomes (Nussbaum and Wilkinson 1989; Sessions 2008). Among amphibians, several basal lineages possess karyotypes characterized by the presence of both micro- and macrochromosomes (e.g. Cryptobranchid salamanders and the tailed frog *Ascaphus truei*), whereas other taxa possess karyotypes consisting of a small number of macrochromosomes and no microchromosomes (e.g. *Xenopus tropicalis*: 1N = 10) (Wickbom 1950; Duellman and Trueb 1986; Sun and Mueller 2014). Similar to *X. tropicalis*, the Mexican axolotl (*Ambystoma mexicanum*) and Eastern newt (*Notophthalmus viridescens*)

possess karyotypes consisting of relatively small numbers of macrochromosomes and no microchromosomes (1N = 14 and 11, respectively) (Humphrey 1975; Hutchison and Pardue 1975; Voss et al. 2011). These patterns suggest the ancestral amphibian karyotype consisted of a large number of chromosomes with two distinct morphologies, both macro- and microchromosomes, with more compact karyotypes representing derived states (Morescalchi et al. 1973; Morescalchi et al. 1974; Sessions 2008).

Our results suggest that relatively more fusions occurred after the divergence of the ancestral lineages that gave rise to frogs and salamanders, than occurred in basal amphibian lineages. The comparative mapping data from newt allows us to better resolve the timing of several fusion events that occurred within the salamander lineage. The majority of fusions in newt (13/19) and *Ambystoma* (13/16) appear to have occurred before the divergence of their ancestral lineages, approximately 150-160 MYA (Table 4.1). The larger number of fusions in the newt lineage appears to be sufficient to explain the difference in chromosome number between newt (1N = 11) and *Ambystoma* (1N = 14). Among the three derived fusions detected in *Ambystoma,* one (corresponding to orthologous segments of newt LGs 6 and 8) is particularly notable with respect to the co-localization of it's fusion boundary with *met*1, a major QTL that has a strong influence on both metamorphic timing and expression of metamorphic vs. paedomorphic life histories (Figure 4.4) (Voss et al. 2012). The evolutionary perspective provided by the newt meiotic map dramatically improves our understanding of the timing of this event and places an upper limit on the age of this fusion at ~150 MYA. Notably, the genus *Ambystoma* is characterized by a highly uniform karyotype (1N = 14), suggesting a likely lower bound at the base of the *Ambystoma* clade (~50 MYA) (Morescalchi et al. 1974; Licht and Lowcock 1991).

Resolving the precise timing of fusions in the amphibian lineage (including, the fusion that overlaps *met1*) and other rearrangements will require the generation of chromosome scale linkage data for several additional amphibian taxa, including several lineages that are not amenable to laboratory

culture and possess genome sizes that exceed the salamander species represented here. Our mapping strategy represents one approach that should permit the rapid generation of linkage maps for any taxon wherein fresh tissues can be collected from a small group of siblings and at least one parent.

Summary

We have generated the first high-density gene-anchored linkage map for *Notophthalmus viridescens* by generating a maternal meiotic map consisting of 2,349 expressed markers, using offspring derived from a single wild-caught female. The number of LGs equals one more than the haploid chromosome number in *N. viridescens* (Hutchison and Pardue 1975). Comparative maps with newt to other amphibians (*Xenopus* and *Ambystoma*) and chicken reveal strong conservation of chromosome-scale synteny across evolutionary time. These syntenic blocks allow us to better resolve the evolutionary history of vertebrate genomes and chromosomes. Moreover, the current study serves as proof of principle for one approach that can be used to rapidly generate chromosome-scale and gene-anchored linkage maps for taxa that have been previously considered intractable.

Acknowledgements

This work was funded by grants from the National Institutes of Health (NIH) (R24OD010435 and EY10540) and Department of Defense (DOD) (W911NF1110475). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of NIH or DOD.

Table 4.1. Fusions detected in amphibian lineages.

Numbers provided in the columns “Newt” “Axolotl” and “Xenopus” correspond to the derived chromosome or linkage group that contains the segments listed in the column “Fused Chicken Chromosomes”. In several cases multiple fusions are observed to occur within a single lineage such that the ordering cannot be resolved with the available taxa. We assume that these are derived from two (e.g. 5 / 10 / 23) or three (e.g. 1 / 19 / 23 / 26) independent fusion events that occurred within a single ancestral or derived lineage. Groups of chromosomes enclosed by parentheses represent presumptive ancestral fusions that experienced additional fusion events in individual (newt, axolotl, or *Xenopus*) lineages.

Fused Chicken Chromosomes	Newt	Ambystoma	Xenopus
4 / Z	5	8	1
5 / 10 / 23	2	6	-
11 / 25 / 28	3	1	-
13 / 18 / 20	3	3	-
17 / 21	4	16	-
26 / 27	4	15	-
2 / 14	6	2	-
7 / 19 / 24	7	9	-
9 / 15 / 22	10	4	-
1 / 8	1	-	-
3 / (5/10/23)	2	-	-
(11/25/28) / (13/18/20)	3	-	-
6 / (17/21) / (26/27)	4	-	-
5 / 12	8	-	-
8 / (11/25/28)	-	1	-
12 / (2/14)	-	2	-
3 / (9/15/22)	-	4	-
15 / 28 / (4/Z)	-	-	1
1 / 19 / 23 / 26	-	-	2
1 / 10 / 13 / 22	-	-	3
5 / 8 / 11 / 12	-	-	4
3 / 9	-	-	5
2 / 20	-	-	6
6 / 21 / 24	-	-	7
4 / 5 / 17 / 25	-	-	8
7 / 14	-	-	9
18 / 27	-	-	10

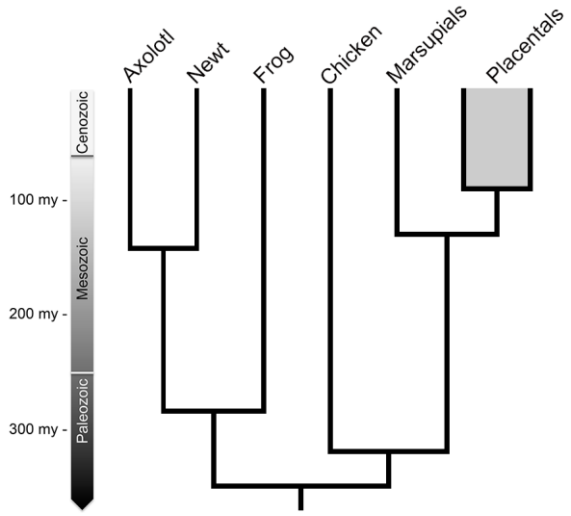

Figure 4.1. An abridged vertebrate phylogeny showing estimated divergence times between species included in this study (newt, axolotl, *Xenopus* and chicken).

The ancestral lineages that gave rise to axolotl and newt diverged in the mid-Mesozoic and the ancestral lineages that gave rise to salamanders and frogs diverged in the late-Paleozoic. Estimated divergence dates are from (Hedges et al. 2015). The eutherian/placental divergence is included to provide perspective on the newt/axolotl divergence. Several extant taxa are not shown, including caecilians and non-avian reptiles.

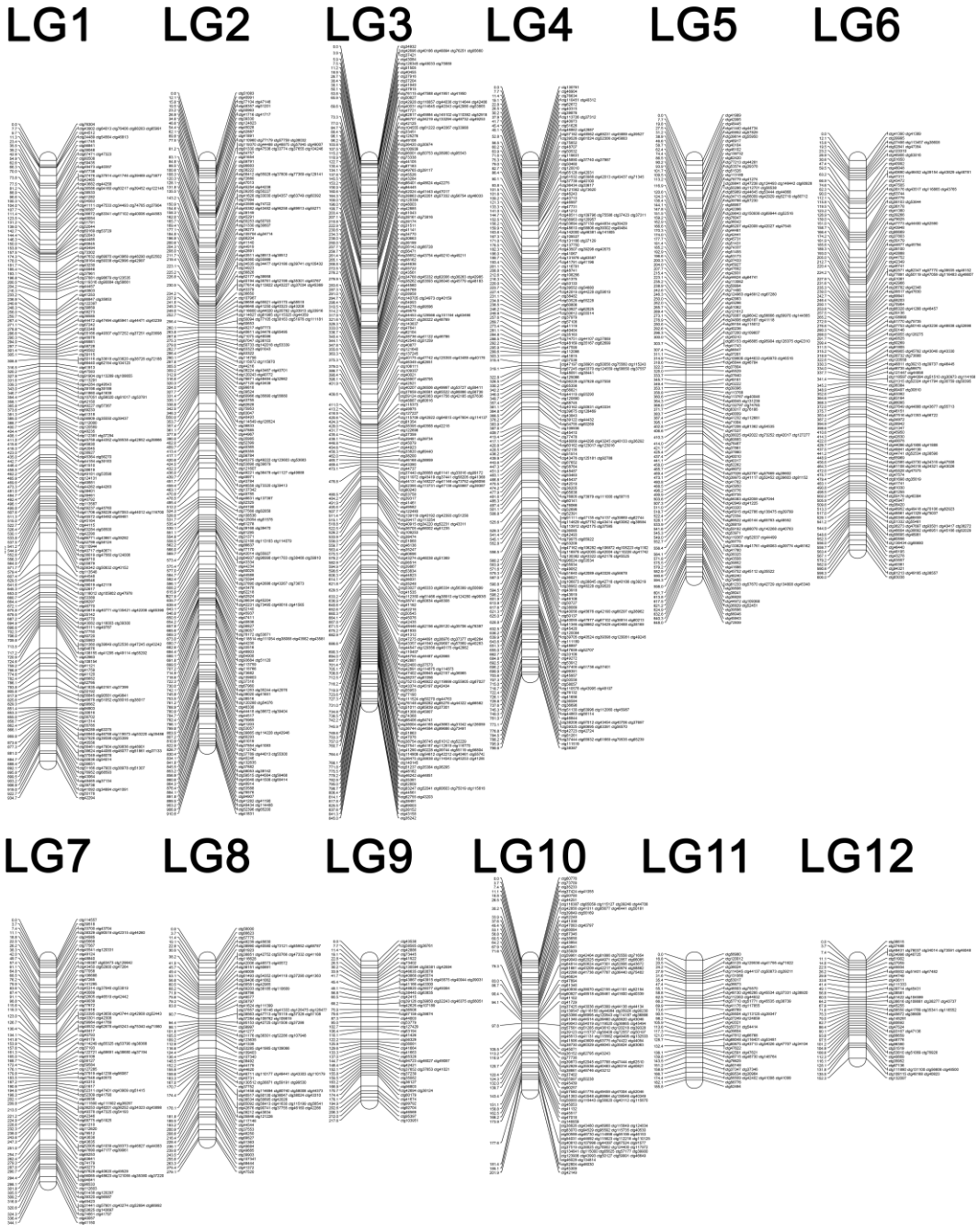

Figure 4.2. The newt meiotic map.

The newt meiotic map consists of 12 linkage groups that range in size form 152.3 to 934.7 centiMorgans (cM) and cover a combined distance of 6,161.9 cM. Marker names refer to previously assembled transcripts (Looso et al. 2013) with “Contig” abbreviated as “ctg”. Distances are shown in cM. Linkage groups are ordered with respect to estimated length, in cM.

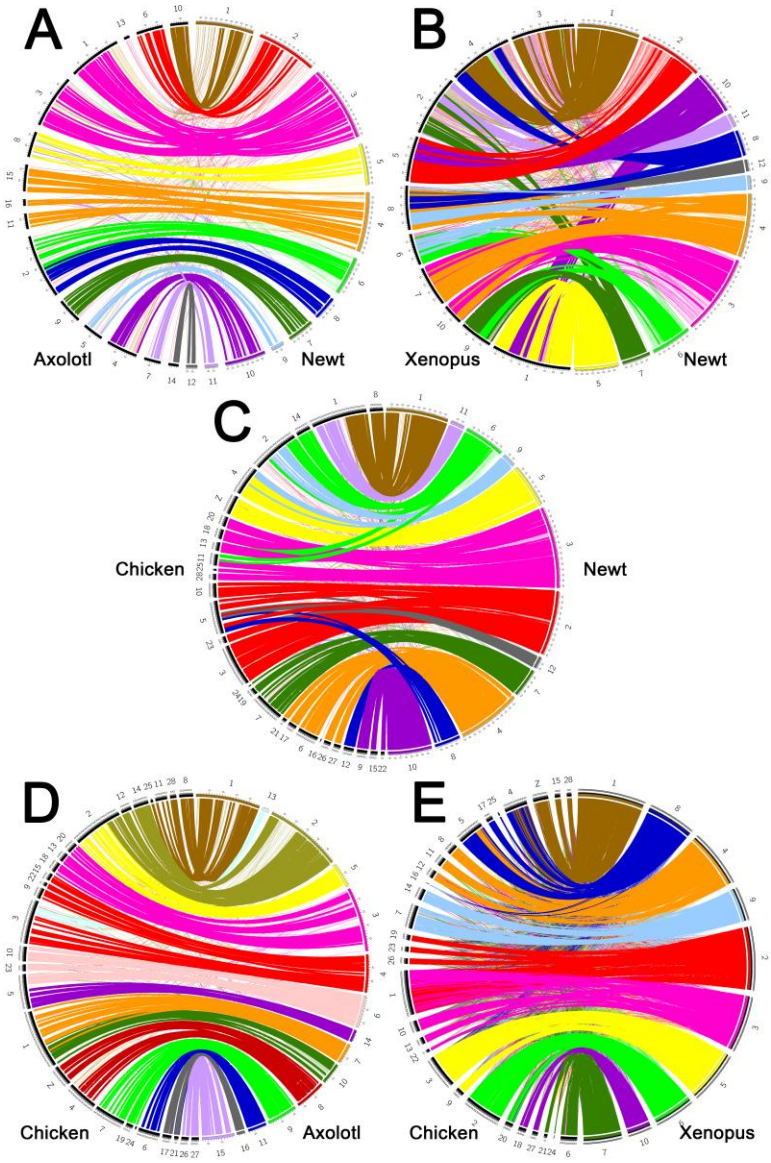

Figure 4.3. Salamander comparative maps reveal fusions fissions and translocations that define the karyotypes of three model amphibian taxa. Comparative maps showing the location of presumptive orthologs in A) axolotl and newt, B) *Xenopus* and newt, C) chicken and newt, D) axolotl and chicken, and E) *Xenopus* and chicken. Lines connecting orthologs are colored according to their location in the newt (A-C), axolotl (D) or *Xenopus* (E) genome. Bold lines connect presumptive orthologs that exist with statistically significant conserved syntenic regions.

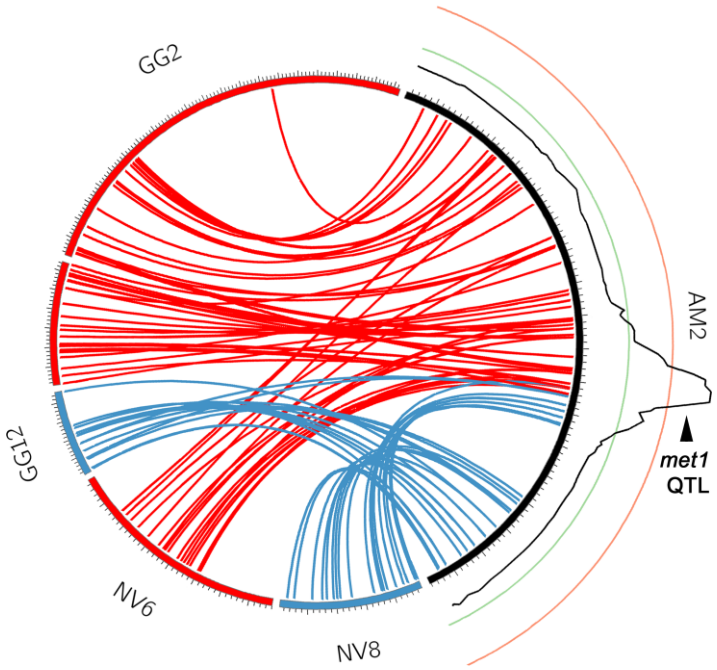

Figure 4.4. Comparative mapping of the *met1* containing linkage group (AM2).

Comparative maps showing the location of presumptive *Ambystoma* orthologs on newt and chicken chromosomes. Lines connecting orthologous genes that are predicted to lie on newt linkage group 6 (NV6) are labeled in red and lines connecting orthologous genes that are predicted to lie on newt linkage group 8 (NV8) are labeled in blue. The log of odds profile for association between individual genotype and expression of metamorphic vs. paedomorphic life history is depicted as a black line adjacent to AM2 (Page et al. 2013). Green and red lines correspond to log of odds ratios of 3 and 10, respectively.

CHAPTER FIVE

MINISCULE DIFFERENCES BETWEEN SEX CHROMOSOMES IN A GIANT VERTEBRATE (SALAMANDER) GENOME

Abstract

In the model salamander (*Ambystoma mexicanum*) sex is known to be determined by a single Mendelian factor, however, the sex chromosomes of this species do not exhibit morphological differentiation that is typical of many vertebrate taxa that possess a single sex-determining gene. Differentiated sex chromosomes are though to evolve rapidly in the context of a Mendelian sex-determining gene and, therefore, undifferentiated chromosomes provide an exceptional opportunity to witness early events in sex chromosome evolution. Whole chromosome sequencing, whole genome resequencing (48 individuals from a backcross of axolotl and tiger salamander) and *in situ* hybridization were used to identify a homomorphic chromosome that carries that *A. mexicanum* sex determining factor and identify sequences that comprise a relatively-small (300 kb) region that is present only on the W chromosome. This region represents ~1/100,000th of the ~32 Gb genome and contains a duplicated copy of the autosomal *ATRX* homolog, named *ATRW*. *ATRW* is one of the few functional (non-repetitive) genes in the chromosomal segment and maps to the tip of chromosome 9 near the marker *E24C3*, which was previously found to be linked to the sex-determining locus.

Introduction

In many species, sex is determined by inheritance of highly differentiated (heteromorphic) sex chromosomes, which have evolved independently many times throughout the tree of life (Bull 1983; Bachtrog 2006; Cortez et al. 2014). Often these chromosomes differ dramatically in morphology and gene content (Rice 1984; Charlesworth et al. 2005; Beukeboom and Perrin 2014). In mammals, males have a large, gene rich X-chromosome and a degraded, gene poor Y-chromosome while females have two X chromosomes. In birds and many other eukaryotes, females are the heterogametic sex with a large Z and smaller

W chromosome while males are homozygous, carrying two Z chromosomes. Differentiated sex chromosomes are thought to arise through a conventional process that begins when a sex-determining gene arises on a pair of homologous autosomes (Charlesworth et al. 2005; Beukeboom and Perrin 2014). The acquisitions of sexually antagonistic alleles, alleles that benefit one sex and are detrimental to the other, favor the suppression of recombination (Charlesworth 1996; Connallon and Clark 2010). Recombination suppression can lead to the accumulation of additional sexually antagonistic mutations and repetitive elements, and over time this results in the loss of nonessential parts of the Y or W chromosome, resulting in the formation of heteromorphic sex chromosomes (Charlesworth and Charlesworth 2000).

Unlike the majority of mammals and birds with stable sex-determining systems and heteromorphic sex chromosomes, amphibians have undergone evolutionary transitions between XY and ZW-type mechanisms several times, and many exhibit sex chromosomes that are morphologically indistinguishable, or homomorphic, like those of the axolotl (Hillis and Green 1990; Schmid et al. 1991; Ogata et al. 2003; Ezaz et al. 2006; 2014). Homomorphic sex chromosomes are not altogether rare among animals with examples in fish (Kamiya et al. 2012), birds (Vicoso et al. 2013b), reptiles (Vicoso et al. 2013a) and amphibians (Stock et al. 2011b). Among most amphibians that have been investigated, homomorphy is prevalent (Green and Sessions 1991; Schmid and Steinlein 2001; Stock et al. 2011b). It has been suggested that a majority of salamanders have homomorphic sex chromosomes (Green and Sessions 1991; 2014; Sessions et al. 2016), however, evidence for genetic sex determination in most species is largely based on observation of 1:1 sex ratios from clutches without thorough demonstration of Mendelian inheritance.

Early developmental experiments revealed a ZW type sex-determining mechanism for *A. mexicanum* (Humphrey 1948; Humphrey and Frankhauser 1957; Armstrong 1984). The first experiment to test for female heterogamety converted female germ cells into sperm using wildtype and white (a mutant caused by a recessive mutation in *EDN3*) animals, by grafting primordial germ

cells from a male wildtype embryo to a female white (homozygous recessive) animal. The sex-reversed male was then crossed with a normal white female (Humphrey 1945). It was expected that if the female was homozygous for sex (XX), the offspring would all be white and female. If the female were heterozygous for sex (ZW), then the white offspring would be approximately 25% male. The pair produced all white offspring with 3:1 ratio of females to males, indicating the male was a sex-reversed female with ZW chromosomes (Humphrey 1945; Humphrey 1948). Mapping studies using the frequency of equatorial separation and the map distance between the centromere and a linked marker concluded that the the sex gene ws located toward the end of the chromosomal arm (Lindsley et al. 1956), and later extimated the sex locus to be 59.1cM distal to the centromere (essentially freely recombining) of an undefined chromosome (Armstrong 1984).

Karyotyping studies of the axolotl later indicated that the smallest chromosomes were heteromorphic in *Ambystoma* species, suggesting that the smallest pair of chromosomes carried the Mendelian sex determining factor in *A. mexicanum* (Cuny and Malacinski 1985) and in the *A. jeffersonianum* species complex (Sessions 1982). However, more recent linkage mapping studies indicated that sex was determined by a locus on one of the larger linkage groups (Cuny and Malacinski 1985; Voss et al. 2013a), and chromosome sequencing studies have demonstrated that the smallest chromosomes do not carry the sex determining region (Smith and Voss 2009; Keinath et al. 2015). Notably, extensive cytogenetic studies performed by Callan, including the use of cold treatments to add constrictions to chromosomes and lampbrush chromosome techniques, revealed no features that could be associated with differentiated sex chromosomes (Callan 1966). The sex chromosomes not only appear identical to one another, but Callan found that mitotic chromosomes 9, 10 and 11 were essentially indistinguishable from one another (Callan 1966). Banding patterns between the closely related tiger salamander and axolotl revealed that axolotls retained all bands found in the tiger salamander karyotype (the reverse is not true), but for one fewer band in the axolotl chromosome 9 compared to tiger

salamander (Cuny and Malacinski 1985), the presumptive sex chromosome reported in this chapter. It is possible, however, that these banding patterns may be misinterpreted, as condensation in a spread of chromosomes can affect the visualization of bands (Brunst and Hauschka 1963).

More recently, linkage analyses and genetic association studies identified the sex-determining locus in the *Ambystoma* genome. Using a cross that was generated by backcrossing female *A. mexicanum*/*A. tigrinum* hybrids with male *A. mexicanum*, sex phenotypes were scored and genetic screens performed for sex-associated regions. A single marker (*E24C3*) was determined to be associated with segregation of the sex phenotype, which localized the sex-locus to the tip of *Ambystoma* LG9 [previously designated LG5, (Smith and Voss 2009)]. In addition, no evidence was found for different recombination frequencies between the sexes suggestive of recent evolution of sex chromosomes, however, these studies did not sample markers in close proximity to the sex locus. (Smith and Voss 2009).

Early theories proposed that homomorphic sex chromosomes must be new or recently arisen, but others have suggested alternate hypotheses that explain the lack of differentiation (Bachtrog et al. 2014). One hypothesis is a high turnover of sex chromosomes, which occurs when a critical sex-determining gene appears on an autosome and replaces the previous sex chromosomes before they differentiate (Schartl 2004). This can occur if a new sex-determining gene arises on a different autosome or if the existing sex-determining region moves to an autosome through transposition or translocation (Schartl 2004). As sex determination is a rapidly evolving trait in many lineages, sex-determining mechanisms may vary among divergent taxa as well as between closely related species or even populations of the same species (Bachtrog et al. 2014). In the order Anura, frog species Hylidae and Bufonidae have male heterogamety and female heterogamety, respectively, suggesting a turnover of sex chromosomes (Stock et al. 2011b; Guerrero et al. 2012). Presumably species with high turnover begin to undergo differentiation events similar to those hypothesized for the early

stages of mammalian and avian sex chromosome differentiation but are essentially reset as new Mendelian sex-determining factors arise.

Evidence supporting the high turnover of sex chromosomes hypothesis has been seen in multiple taxa with similar phenomena, including many fish (Mank and Avise 2009; Kitano and Peichel 2012; Yoshida et al. 2014), several other frogs (Miura 2007; Stock et al. 2011a), the platypus (Veyrunes et al. 2008) and some flies (Vicoso and Bachtrog 2013). Although it is known that many salamanders exhibit homomorphic sex chromosomes, and that both XY and ZW sex-determining mechanisms exist in the salamander, far less is known about the sex chromosome turnover. A recent cytogenetic study in proteid salamanders revealed two species, the Necturus and the Proteus salamanders had heteromorphic and homomorphic XY sex chromosomes, respectively (Sessions et al. 2016). Banding patterns suggest a translocation of what may be the sex-determining region to an autosome, forming homomorphic sex chromosomes in the proteus salamander, however, this study was based purely on cytogenetics and lacks genomic or functional experimentation (Sessions et al. 2016). This allowed the authors to speculate that a sex chromosome turnover occurred in the family, but it should be noted that the divergence time between the two species is quite large: ~100My (Hedges et al. 2015; Sessions et al. 2016).

Another hypothesis to explain “ever young” sex chromosomes assume that through sex reversal, the heterogametic sex chromosomes may recombine in the opposite sex (Perrin 2009; Stock et al. 2013). Deemed the fountain-of-youth hypothesis, these sex chromosomes keep their youthful appearance when an environmental stimulus, such as temperature, has some impact on the phenotypic development of sex, so that a male might be ZW, and a female might be XY. When this occurs, the sex chromosomes are expected to recombine, contributing to their lack of differentiation. Evidence of this hypothesis has been described in several species of tree frog (Stock et al. 2011b). It has been suggested that sex reversal might be an important evolutionary force within the amphibians, as temperature has been shown to affect gene expression and enzymatic activity in the sex-determining pathways of ectothermic vertebrates

(Stock et al. 2013). Temperature has been shown to alter sex ratios in some salamander species; however, in axolotl and other ambystomatid salamanders, these temperatures are extreme, falling outside the normal range of temperatures experienced by developing salamanders (Gilbert 1936; Wallace et al. 1999).

Relatively undifferentiated sex chromosomes have also been observed for species that are known to possess old sex chromosomes. For example, ratite birds are known to possess sex chromosomes that evolved before the divergence of the ancestral lineages that gave rise to ratites and neornithes ~140MYA. However, the sex chromosomes of ratite birds are characterized by a relatively small non-recombining region, making them nearly homomorphic. It has been hypothesized that ratite sex chromosomes that maintain their youthful appearance by virtue of sex-biased expression of genes, which is thought to select for the maintenance of genes on the W chromosome (Vicoso et al. 2013b). One study showed almost a 2 fold expression difference between sexes in a region that contains a putative male-determining gene *DMRT1* (Wang et al. 2014). Interestingly, the genes affected lie in the pseudoautosomal region, the region still recombining between sex chromosomes (Vicoso et al. 2013b). While dosage compensation, a mechanism by which the heterogametic sex equalizes gene expression of X-linked or Z-linked traits, has evolved in chickens, it is not present in ratites (Wang et al. 2014). Sexual antagonism may play a role in that fewer sexually antagonistic polymorphisms may not cause selection for reduced recombination (Rice 1987; Fry 2010).

To identify sex-linked (W-specific) regions in the undifferentiated sex chromosomes of axolotl, we generated sequence reads for 48 individuals backcrossed (*A. mexicanum/A. tigrinum* X *A. mexicanum*) salamanders of known sex and aligned these reads to the existing reference genome (Smith et al. 2005b; Keinath et al. 2015) (www.ambystoma.org). Analyses of read coverage identified 156 putative W-linked sequences, including two genes, an *ATRX* paralog and an ortholog of *MAP2K3*. These findings are useful for characterizing the sex-specific region of the axolotl and provide evolutionary perspective on the

homomorphic sex chromosomes of ambystomatid salamanders, indicating that the sex-determining gene in *A. mexicanum* may have arisen within the last 20MY. In addition, we anticipate that these sex-linked markers will be useful for identifying sex in juvenile axolotls, where sex is an important covariate for experimental studies, including studies of metamorphosis and regeneration.

Results

Identification of the sex-bearing chromosomes by FISH

Previous studies have demonstrated that sex is linked to the marker *E24C3*, at a distance of ~5.9cM distal to the terminal marker on LG9 (Smith and Voss 2009; Voss et al. 2013a). Consistent with linkage analyses, *E24C3* was detected near the tip of an average-sized chromosome. A second BAC corresponding to a marker from the opposite end of LG9 (*E12A6*) localized to the opposite tip of the same chromosome, indicating that this chromosome corresponds precisely to LG9 (Figure 5.1A). Notably, the BAC carrying *E12A6* also cross-hybridized with the centromere of all chromosomes, a feature that could potentially be useful for future experiments that require labeling of centromeres of comparison of distances to the centromere.

Comparative genomic hybridization was also performed to gain perspective on the degree to which otherwise-indistinguishable Z and W chromosomes might differ at the microscopic level. Competitive hybridization of differentially labeled male and female DNA revealed some enrichment for female-specific signals near the sex-specific region for axolotl and signal at the tips of two other chromosomes, which could be due to a polymorphism within the sampled animals or a sex-specific mutation (Figure 5.1B). Signals seem to be equally strong for both sexes and do not robustly identify a unique sex-specific region, likely due to the small size of sex-specific sequence.

Laser capture, sequencing and assembly of the Z chromosome

In an attempt to increase the number of markers that could be associated with the sex chromosome, we performed laser-capture sequencing on a chromosome corresponding to LG9. This library was generated from a single dyad that was collected in a larger series of studies on laser capture microscopy of axolotl chromosomes (Chapter 2). The sex chromosome library contained a total 143,156,920 reads between 40 and 100 bp after trimming and contained 995 reads that mapped to 23 distinct markers (transcripts) that had been previously placed on LG9, accounting for 40% of the markers that are known to exist on the linkage group (Figure 5.2). Given support for this, an additional lane of sequencing, yielded 935,736,694 additional reads (for a total of 1,078,893,614 reads). After trimming, 54,1884,866 reads remained. 9,272,583 paired end reads aligned concordantly to the human genome and 660,828 paired end reads aligned concordantly to bacterial genomes and were removed. Of the remaining reads (531,357,363), 68,844 of these reads aligned to 40 LG9 contigs, 70% of the LG (Figure 5.2). An error-corrected assembly of these data yielded a total of 1,232,131 scaffolds totaling 242.4Mb with a scaffold N50 length of 295, and contig N50 length of 126bp. (Table 5.1: results from other chromosomes are shown for comparison purposes). After aligning the paired end reads from the LG9 library to the whole genome assembly, 27,500 scaffolds were identified that could be reliably assigned to LG9. The size of these scaffolds totaled ~833Mb with an N50 of 4,515bp. This subset of assigned scaffolds could be used as a complementary assembly with higher contiguity.

Library evaluation and evolutionary conservation

Alignments between the sex chromosome assembly and *Ambystoma* reference transcripts were used to identify genes on the sex chromosome. These genes were aligned to the chicken genome assembly to confirm that homologs from the axolotl sex chromosome were heavily enriched on chicken chromosomes 7, 19 and 24, consistent with previous findings (Figure 5.3) (Voss et al. 2011).

Alignments from the sex chromosome assembly with chicken and newt

revealed that newt LG7 is homologous to axolotl LG9 (Figure 5.4). While ZW-type mechanism for sex determination has been inferred for the newt (National Research Council (US) Subcommittee on Amphibian Standards 1974), it is not known yet known which chromosome determines sex and no candidate genes currently exist.

Identification of female-specific regions

To identify sex-specific regions of the genome, we aligned low coverage sequence data from 26 males and 22 females to the first public draft assembly of the axolotl genome (Smith et al. 2005b; Keinath et al. 2015) (www.ambystoma.org). Notably, the draft assembly was generated using a female axolotl, which should contain genomic regions from both Z and W chromosomes. Males and females were drawn from a previously published mapping family, which was used in the initial mapping of the sex locus (Smith and Voss 2009). Each individual was sequenced at ~1X coverage with Illumina HiSeq short paired-end reads (125bp) resulting in 7.4 billion (7,426,348,268) total male reads and 6.4 (6,460,020,910) total female reads. The ratio of male to female coverage was calculated across 10,440,093 intervals covering ~19Gb of the draft assembly. Genome-wide coverage ratios generally fell within a tight distribution centered on equal coverage, after accounting for initial differences in average depth of coverage (Figure 5.5). Intervals were considered to be candidate sex-specific regions if enrichment scores [$\log_2$ (female coverage/adjusted male coverage)] exceeded two. In total, these analyses identified only 201 candidate female-specific intervals that were contained within 152 genomic scaffolds, with 30 genomic scaffolds having 2 or more intervals. A total of 47 intervals were represented by zero male reads, and average male coverage of male reads for other intervals ranged from 0.02 to 4.4.

PCR validation of candidate regions

PCR primers were designed for all candidate scaffolds and subject to initial PCR validation using a panel of six females and six males. In total, primers from 42 of

the 156 scaffolds yielded specific amplicons in all females, but no amplicons from males, and were considered sex-specific. The combined size of these scaffolds is approximately 300Kb or ~0.0094% of the genome. Aside from the validated female-specific scaffolds, 7 were present in a subset of the animals with no specific trend toward one sex or the other. Presumably these represent structural (insertion/deletion) variants that are segregating with in the lab population of *A. mexicanum*. Primers for another 82 scaffolds yielded amplification in both sexes with 20 showing brighter bands in females and 5 showing brighter bands in males. Primers for 27 other scaffolds yielded no amplification in either sex.

Homology

To search for evidence of sex-specific genes, all 42 validated sex-specific scaffolds were aligned (blastx) to the NCBI nonredundant protein database. In total, these searches yielded alignments to 17 protein-coding genes (Table 5.2.), several of which involved weak alignments to uncharacterized proteins (N = 4) or transposable elements (N = 5). However, two scaffolds yielded strong alignments to human protein coding genes. Specifically, Scaffold SuperContig_990642 aligned to transcriptional regulator *ATRX* (*ATRX*: 65% amino acid identity) and scaffold SuperContig_1084421 aligned to mitogen-activated protein kinase kinase kinase 2-like (MAP3K2: 97% amino acid identity). Notably, a conserved syntenic ortholog of MAP3K2 would be expected to occur on LG9, and thus it seems likely that MAP3K2 resided on the ancestral LG9 sex chromosome prior to the origin of the *A. mexicanum* sex-determining locus. However, a syntenic ortholog of *ATRX* would be expected to occur on LG2, which is syntenic a large region of the X chromosome that is conserved across all therian mammals (Smith et al. 2005a; Smith and Voss 2006; Smith and Voss 2007).

The identification of a sex-linked *ATRX* homolog is notable, as *ATRX* is known to play major roles in sex determination in mammals and other vertebrates (McElreavey and Fellous 1997; Neri and Opitz 1999; Pask et al. 2000; Huyhn et al. 2011). Alignments between scaffold SuperContig_990642 and the mapped ATRX homolog reveal that two distinct *ATRX* homologs exist in

axolotl (Figure 5.9). Henceforth, we refer to the syntenic homolog on LG2 as *ATRX* and the W-specific homolog as *ATRW*. A nucleotide alignment between the axolotl *ATRX* and *ATRW* genes shows that the genes share 90% identity across 1089 aligned nucleotides, and as such it appears that the two genes diverged relatively recently by transposition of a duplicate gene copy to the W chromosome. To further test this idea and better define the timing of this duplication, a tree was generated using *ATRX* homologs from several vertebrate taxa (Figure 5.6,5.7). Based on this tree, we infer that a duplication event gave rise to *ATRW* within *Ambystoma*, after divergence from its common ancestor with newt (the two lineages shared a common ancestor ~151 MYA) (Hedges et al. 2015). Considering the degree of sequence divergence and the relative length of shared vs. independent branches we estimate that the *ATRW* homolog may have arisen sometime in the last 20 MY (Figure 5.8), a timing that roughly coincides with a major adaptive radiation in the tiger salamander lineage (Shaffer 1984a; Shaffer 1984b).

To shed further light on the evolution of *ATRX* and *ATRW* within the *Ambystoma* lineage, we examined patters of derived substitutions in *ATRX* and *ATRW*. Across the 251bp alignment the number of nucleotide substitutions that can be attributed to *ATRW* since the divergence of axolotl is 9, which change 2 amino acids. By comparison, *ATRX* on LG2 shows only 1 nucleotide substitution since the duplication event (Figure 5.9). This suggests that *ATRW* may be evolving at a faster rate than *ATRX*, in which case 20 MY may represent a substantial overestimate for the origin of the duplication that gave rise to *ATRW*.

Discussion

Sex chromosome evolution in the axolotl

The results from this study show that the homomorphic sex chromosomes of the axolotl contain a small non-recombining region that is specific to the female W chromosome. The female-specific sequence is estimated to be about 300Kb, or roughly 1/100,000th of the enormous axolotl genome. Due to the physical size of

the genome and marker density in the recombinational map for the axolotl, it is not surprising that the differences in recombination were not initially evident (Smith and Voss 2009). With respect to the current fragmented assembly, it is still not possible to predict gene orders within this region or locate possible inversions, however the data are sufficient to identify robust markers for sex and genes that exist in the non-recombining region. Of the few protein-coding genes found within the validated sex-specific scaffolds two appear to represent non-repetitive coding sequences, including one that represents a relatively recent duplication of the transcriptional regulator *ATRX*.

The *ATRX* is a gene is located in the non-recombining region of the X chromosome in mammals. The gene is a chromatin remodeler that belongs to the SWI/SNF family. It is linked to the rare recessive disorder, alpha-thalassemia X-linked intellectual disability, which is characterized by severe intellectual disability, developmental delays, craniofacial abnormalities, and genital anomalies in humans (Stevenson 1993; Lee et al. 2015). In some cases, a mutation in the *ATRX* gene can lead to a female sex reversal due to early testicular failure (Ion et al. 1996). Gene expression studies performed in a marsupial and eutherian showed that *ATRX* expression was highly conserved between the two mammals and was necessary for the development of both male and female gonads (Huyhn et al. 2011). As one of the few protein-coding genes present in the region of W-specific sequence, and one that has been characterized in sex differentiation of mammals, we propose the *ATRW* as a candidate sex gene for axolotl, or alternately a strong candidate for an acquired sexually antagonistic gene.

Reanalysis of expression data from recent published tissue-specific transcriptomes showed expression of the *ATRX* gene (from LG2) in all major tissues and developing embryos, however, they showed no evidence of expression of the *ATRW* gene (Bryant et al. 2017). The tissues represented in the study included whole limb segments, blastemas from regenerating limbs, bone and cartilage, muscle, heart, blood vessel, gill, embryos, testis, and notably, ovaries. It is not clear at what stage the ovarian tissue was taken, however, the

author suggests multiple ovaries were sequenced from an adult, and multiple libraries exist for the tissue. It is possible that this sex-specific gene is simply not highly expressed at this specific stage (or in the adult stage, in general) and may only be expressed during early gonadogenesis. Similarly, the comprehensive annotation of W-linked genes in chicken was not known until it was revealed through RNAseq studies of gene expression prior to and during gonadogenesis (Ayers et al. 2013).

If *ATRW* is the primary sex-determining gene in axolotl, then the origin of this gene marks the origin of sex chromosomes in the tiger salamander lineage. A time-scaled gene tree based on sequence substitution rates of *ATRX* genes in multiple vertebrate placed the *ATRX* duplication event at ~20 MYA (Figure 5.8). This estimate places the *ATRX* duplication event within the *Ambystoma* clade but suggests all ambystomatids may not necessarily share the sex chromosome. Based on the *Ambystoma* species tree (Hedges et al. 2015), we expect the same sex chromosomes and sex locus to be present in tiger salamander complex but may not be in the more distantly related *A. jeffersonianum* complex or deeper ambystomatid lineages (Figure 5.10).

Given the relatively recent origin of *ATRW*, species within the tiger salamander complex are predicted to contain the same sex chromosomes. The tiger salamander specific complex consists of more than 30 named species that encompass a range of diversification dates (Shaffer 1984a; Shaffer 1984b; Shaffer and McKnight 1996). This complex should therefore facilitate future studies aimed at more precisely characterizing the timing of the *ATRX/W* duplication and the evolution of other W-specific sequences. Ongoing improvements to the *Ambystoma* assembly and development of assemblies for other salamander taxa should improve our ability to assess hypotheses related to the presence of homomorphic sex chromosomes (*e.g.* recent evolution, high-turnover, and fountain of youth). Additionally, recent efforts to develop genetic tools for the axolotl model should facilitate functional analyses that will be necessary to test whether *ATRW* is the primary sex-determining gene in axolotl or elucidate its role as a sexually antagonistic factor. Methods for achieving

targeted gene knockout and knockins have been developed in axolotl (Fei et al. 2014; Flowers et al. 2014; Woodcock et al. 2017) and could be adapted to better assess the functionality of *ATRW* in axolotls.

Utility of sex-linked markers in axolotl

Sex is an important biological variable in research, as it may contribute to differences in experimental data. Because axolotl is an important model for many areas of research and has shown sex-specific effects, such as tail regeneration, it is important for investigators to differentiate sex effects from other experimental variables (Voss et al. 2013a). Until now it was necessary to visualize the sex organs, utilize axolotls that had produced gametes, or perform experiments in hybrid crosses that segregate markers at the linked locus E24C3 in order to accurately determine sex in axolotls (Smith and Voss 2009). However, many experiments utilize juvenile animals that may not have completed gonadal differentiation or maturation. With several robust markers for W-specific sequences in hand, it is now possible to precisely differentiate sex of an axolotl with a simple PCR. These markers will also positively impact axolotl husbandry, as individuals may be housed and utilized in experiments accordingly (Chapter 6).

Methods

Metaphase chromosomes spreading for laser capture microdissection

Preparation of cells for metaphase spreads was done as previously (Chapter 2)(Keinath et al. 2015). Fixed cells were spread on UV-treated 1.0mm polyethylene naphthalate (PEN) membrane slides. Slides were inverted (membrane side down) over a steam bath of distilled water for 7 seconds. Immediately after steaming, 100µl of the fixed cells were dropped across the middle of the slide lengthwise. Each slide was subsequently placed in a steam chamber at ~35°C for 1 minute, then set on the hot plate for 5 minutes. After slides are dry, chromosomes were stained via immersion in freshly made Giemsa

stain for 2 minutes, rinsed in 95% ethanol, rinsed in distilled water, then allowed to dry in a desiccator until used.

Laser capture microdissection and amplification

The sex chromosome was captured using a Zeiss PALM Laser Microbeam Microscope at 40X magnification as done previously (Chapter 2, 3) (Keinath et al. 2015). The sex chromosome was dissected individually using a Zeiss PALM Laser Microbeam Microscope at 40X magnification and catapulted into Zeiss adhesive cap tubes (Zeiss 415190-9191-000). 10µl of a chromatin digestion buffer was pipetted into the cap (Keinath et al. 2015), and the tube was kept inverted overnight at 55°C. After incubation, the sample was centrifuged briefly and incubated at 75°C for 10 minutes and 95°C for 4 minutes to inactivate the Proteinase K. Along with 23 other samples, the sex chromosome sample was immediately carried through full amplification via Rubicon PicoPlex DNAseq Whole Genome Amplification (WGA) kit (R30050). Amplified material was outsourced for sequencing on an Illumina HiSeq 2000 (Hudson Alpha Institute for Biotechnology, Huntsville, Al.).

The standard manufacturer protocol was used with the exception of the cell extraction step, as chromatin digestion buffer was used prior to the second step. After amplification, an Agilent 2100 Bioanalyzer and accompanying DNA 12000 kit (Agilent DNA 12000 Kit 5067-1508), was used to approximate concentration and size distribution. The sex chromosome sample had a concentration >9ng/µl and sequenced on an Illumina HiSeq 2500 (Hudson Alpha Institute for Biotechnology, Huntsville, Al). After initial sequencing, the same sample was further sequenced paired-end 150bp length reads on a full lane of HiSeq 2500.

Sex chromosome sequence analyses and assembly

Because amplified sequences contain a non-complex leader sequence corresponding to the pseudorandom primers that are used for whole chromosome amplification, Trimmomatic was employed on the resulting reads to

remove leader sequence, remove other sequences derived from phiX, and perform quality trimming using the sliding window option to trim any window of 40 nucleotides with quality score lower than Q30 (Lohse et al. 2012). Reads were then aligned to the 945 model transcripts from the *Ambystoma* linkage map (Voss et al. 2011) using the Burrows Wheeler Aligner using the single-end mapping option and BWA-MEM algorithm (Li and Durbin 2009). They were also aligned to the human reference genome using the paired-end mapping option to identify exact matches for Bowtie 2 (Langmead et al. 2009). Paired reads mapping to the human genome concordantly were considered potential contaminants and removed. After trimming and removal of potential contaminants, the reads were corrected with Blue (Greenfield et al. 2014) using female *A. mexicanum* whole genome shotgun data (Keinath et al. 2015) and assembled with SOAPdenovo2 (Luo et al. 2012).

In order to assign whole genome scaffolds to LG9, the reads from the LG9 chromosome library were aligned paired-ended to the draft whole genome assembly using the Burrows Wheeler Aligner with the BWA-MEM algorithm (Li and Durbin 2009). They were filtered using samtools, and those scaffolds considered to reliably belong to LG9 contained at least 10 alignments, where each set of read pairs mapped concordantly with 100% identity, and the leftmost and rightmost mapping positions covered at least 50% of the scaffold.

FISH of sex-associated BAC *E24C3* and CGH

Fluorescent *in situ* hybridization of BACs to metaphase chromosome spreads were performed as previously described (Timoshevskiy et al. 2012; Timoshevskiy et al. 2017). A Qiagen Large Construct kit (Qiagen Science, 12462) was used to extract bacterial artificial chromosome (BAC) DNA for *E24C3* and *E12A6*, previously associated with sex (Smith and Voss 2009). Probes for *in situ* hybridization were labeled by nick-translation using direct fluorophores Cyanine 3-dUTP (Enzo Life Sciences, ENZ-42501) or Fluorescein-12-dUTP (Thermo Scientific, R0101) as described previously (Timoshevskiy et al. 2012).

Hybridization of BAC probes was performed as previously described for axolotl chromosomes (Woodcock et al. 2017).

To isolate repetitive DNA fractions from female salamander tissue, phenol-chloroform extraction in 1.2X SSC was used (Sambrook and Russell 2006). DNA was denatured for 5 minutes at 120°C, reassociated at 60°C for 1 hour to obtain $C_o t$ DNA. The DNA was placed on ice for 2 minutes then transferred to a bead bath at 42°C for 1 hour with 5X S1 nuclease buffer and S1 nuclease for a concentration of 100 units per 1mg DNA. DNA was precipitated with 0.1 volume of 3M sodium acetate and 1 volume isopropanol at room temperature, inverted several times and centrifuged at 14,000 rpm for 20 minutes at 4°C. DNA was washed with 70% ethanol, centrifuged at 14,000 rpm for 10 minutes at 4°C, air dried and solubilized in TE buffer.

Conservation and evolution of salamander chromosomes

To better evaluate the sex chromosome assembly, alignments between the sex chromosome assembly and reference transcripts (V4) were performed using megablast (Zhang et al. 2000) to identify genes that occurred on the sex chromosome. These genes were then aligned (tblastx) (Altschul et al. 1997) to the chicken genome assembly (Gallus_gallus-2.1, GCA_000002315.1) . Those with an alignment length of at least 50 amino acids and having at least 60% identity were considered potential homologs.

A similar approach was taken to identify the homologous newt linkage group to assess for potential sex candidate genes on the homologous group. *Ambystoma* reference transcripts from LG9 (V4) were aligned (tblastx) (Altschul et al. 1997) to the chicken genome assembly (Altschul et al. 1990). Using the same minimum thresholds as above, the potential homologs were then used to blast (tblastx) (Altschul et al. 1990) to the newt, *Notophthalmus viridescens*, reference transcripts (Keinath et al. 2017).

Scaffolds that were validated through PCR in a panel of 6 females and 6 males were aligned to the V4 and V5 *Ambystoma* transcriptome assemblies in order to identify the genes present from the transcriptome. If a transcript aligned

to the scaffold with a percent identity higher than 95%, that transcript was blasted (blastx) to the ncbi nonredundant protein database to search for homologous genes.

Identification of female-specific regions

We applied depth of coverage analysis to identify single-copy regions in the assembly that have approximately half of the modal coverage in females and underrepresented/absent coverage in males. An exact zero cutoff was not used in order to account for possible read mismapping. There are several analytical caveats in conducting the coverage analysis for the huge, complex salamander genome, such as the enormous number of scaffolds in the assembly, long gaps between contigs and the presence of repeats.

Reads generated on an Illumina HiSeq2000 (Hudson Alpha Institute for Biotechnology, Huntsville, Al.) from DNA that was isolated via phenol-chloroform extraction (Sambrook and Russell 2006) from 48 individuals that were drawn from a previously described backcross mapping panel (Smith et al. 2005a). The resulting reads were aligned to the axolotl draft genome assembly using BWA-MEM (using default parameters) followed by filtering of secondary alignments (samtools view –F2308) and alignments clipped on both sides of the read. Merging of female bam files and male bam files was performed using *Samtools merge* (Li and Durbin 2009; Li et al. 2009).

We used the algorithm DifCover to identify candidate female-specific regions. The DifCover method used works by computing the ratio of female: male average depth of coverage across continuous intervals containing approximately *V valid bases*. The valid bases are determined by lower and upper limits on depth of coverage for females (f) and males (m), defined respectively by min*f*, min*m*, max*f* and max*m.* If *Cf* and *Cm* are females and males coverage for a given valid base, then 1) *Cf* < max*f and Cm* < max*m;* and 2) *Cf*>min*f or Cm*>min*m.* Upper limits allow to determination and skipping of fragments that contain repeats, while lower limits serve to exclude underrepresented fragment - fragments with too small number of reads in both males and females. After some

testing, we defined V = 1000 and assigned lower limits to be equal to one third of modal coverage (min$f$ = 8, min$m$ = 9) and upper limits to 3X of modal coverage (max$f$ = 75, max$m$ = 87).

After testing, we chose V=1000 and assigned lower limits equal to one third of modal coverage, (8 for females and 9 for males) and upper limits 3X of modal coverage, (75 for females and 87 for males). The enrichment scores [log2(standardized sperm coverage/blood coverage)] were computed for each interval. If the average coverage in males for an interval had a zero value, we replaced it with non-zero positive value corresponding to alignment of half of one read. Some intervals are shorter than 1Kb and contain fewer than 1000 valid bases, and those that fell on the ends of scaffolds or if the interval were an entire scaffold, they were not used. Otherwise only intervals of at least 500 bases and with at least 200 valid bases were considered.

Primer design and PCR

Primers were designed within the sex candidate regions identified using Primer3 (Untergasser et al. 2012). Each primer was 25-28bp in length, with a target melting temperature of 60°C, 20-80% GC content and 150-400 bp product sizes depending on the size of the region and location of repeats (avoiding inclusion of repetitive sequence in primer and product). Fragments were amplified using standard PCR conditions (150ng DNA, 50ng of each primer, 200 mM each dATP, dCTP, dGTP, dTTP; thermal cycling at 94°C for 4 minutes; 34 cycles of 94°C for 45 seconds, 55°C for 45 seconds, 72°C for 30 seconds; and 72°C for 7 minutes). Reactions were tested on a panel of 6 males and 6 females to validate sex specificity. Gel electrophoresis was performed and presence/absence was recorded for each set of primers. The scaffolds from which primers were designed were considered female-specific if the primers yielded specific amplicons in all 6 females and in no males.

Phylogenetic reconstruction

Homologene was used to collect putative homology groups from the *ATRX* genes in a variety of eukaryotes (National Center for Biotechnology Information 2004). Sequence for axolotl *ATRX* was obtained from the *Ambystoma* reference transcripts, and the newt *ATRX* gene was obtained by aligning human *ATRX* to the newt reference transcriptome (Abdullayev et al. 2013). All sequences were aligned through MEGA7 (Kumar et al. 2016) via MUSCLE (Edgar 2004). Sequences were trimmed to compare only that sequence which was present in all species, a string of 251 codons. Divergence time ranges were added during the production of the timetree using estimations provided by TimeTree (Hedges et al. 2015).

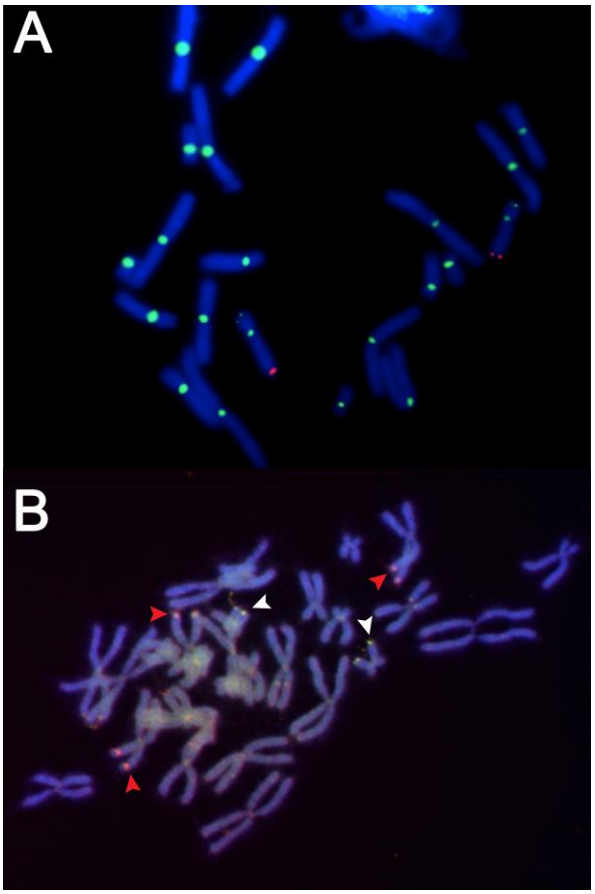

Figure 5.1. FISH of sex-linked BACs and CGH.

Cytogenetic methods improve visual identification of the sex-linked markers. **A)** FISH localizes two markers (*E24C3* and *E12A6*) associated with the sex locus, *ambysex*. DAPI stained metaphase spread of axolotl chromosomes. *E24C3* labeled with cy3 (red) and *E12A6* labeled with fluorescein isothiocyanate (green). **B)** CGH elucidates regions differing between male and female. Female DNA labeled red and male DNA labeled green. White arrows denote the location of green fluorescence on the sex chromosomes, consistent with the localization of *E24C3.* Red arrows denote locations where red fluorescence is more prevalent. While the leftmost red arrow seems to be located at the NOR region of chromosome 3, the other two red arrows point to what is seemingly fluorescence at the tip of another pair of chromosomes. These regions may be polymorphisms specific to the animal used for this preparation, or they may reflect chromosomal regions unique to sex.

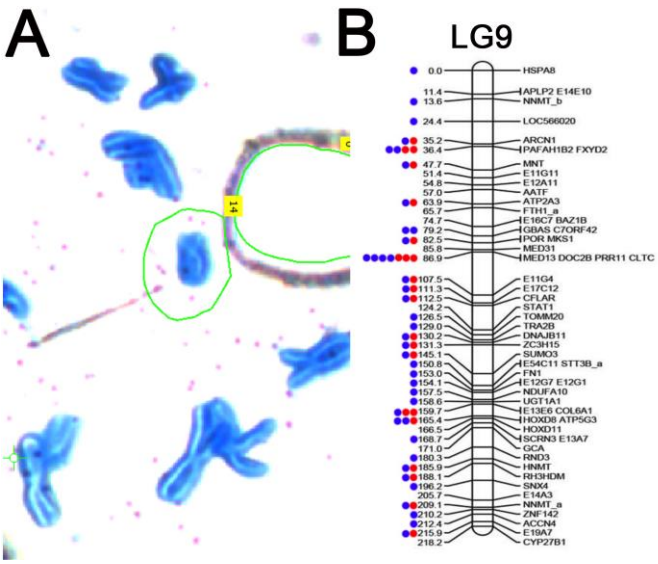

Figure 5.2. Individual sex chromosome dyad alignment results on LG9. Read mapping was used to assess the specificity of the laser capture, amplified library of the sex chromosome dyad. **A)** A partial metaphase spread of axolotl chromosomes stained with Giemsa on a membrane slide. The sex chromosome is circled in green. **B)** The distribution of markers sampled from the sex chromosome (LG9) via targeted sequencing of individual chromosomes. Dots represent markers with mapped reads from a single library. Red denotes the first sequencing attempt using the DNA-seq kit with 48 total barcoded samples on a single lane of an illumina HiSeq flowcell. Blue denotes re-sequencing of the same chromosome library on a single lane.

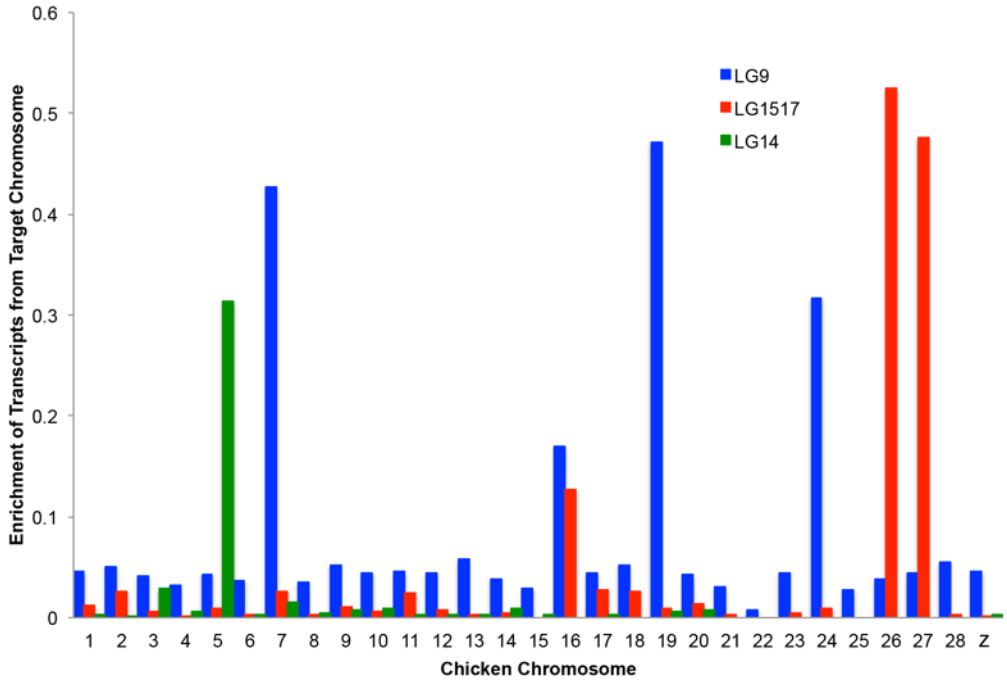

Figure 5.3. Conserved synteny between assembled A. *mexicanum* chromosomes and the chicken genome.

Tests for enrichment of AM13 denoted by red (LG15/17 targeted), AM14 denoted by green (LG14 targeted), sex chromosome denoted by blue (LG9 targeted) homologs across all assembled chicken chromosomes. Enrichment scores are calculated by dividing the observed number of homologs by the total number of genes annotated to the individual chicken chromosomes (Cunningham et al. 2015).

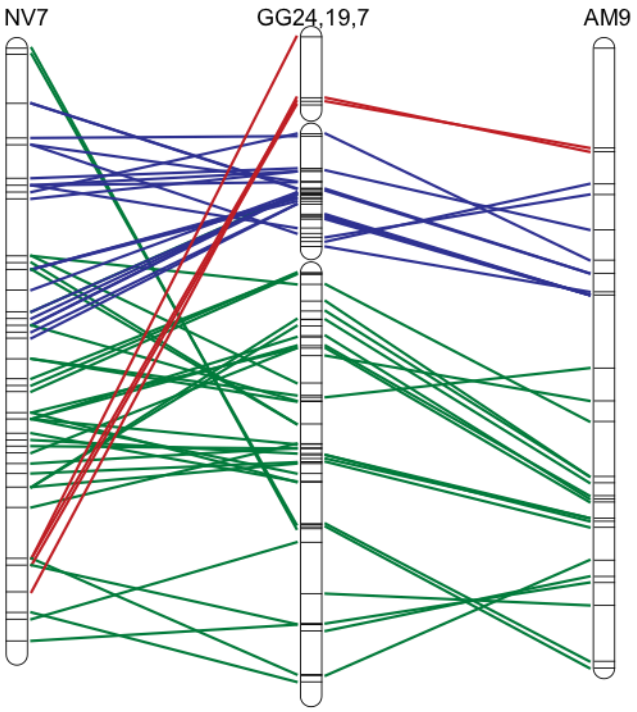

Figure 5.4. Conserved synteny between newt, chicken and axolotl. Conserved synteny studies show syntenic regions shared between newt linkage group 7 (LG7, left), chicken chromosomes 24 (top center), 19 (middle center) and 7 (bottom center), and axolotl LG9 (right). Each line corresponds to an alignment between a pair of presumptive chicken and salamander (newt or axolotl) orthologs. Alignments involving orthologs on chicken chromosome 7 are colored green, chromosome 19 are colored blue, and chromosome 24 are red. More alignments were found between newt and chicken, as the linkage map of the newt is more dense than that of the axolotl (Keinath et al. 2017).

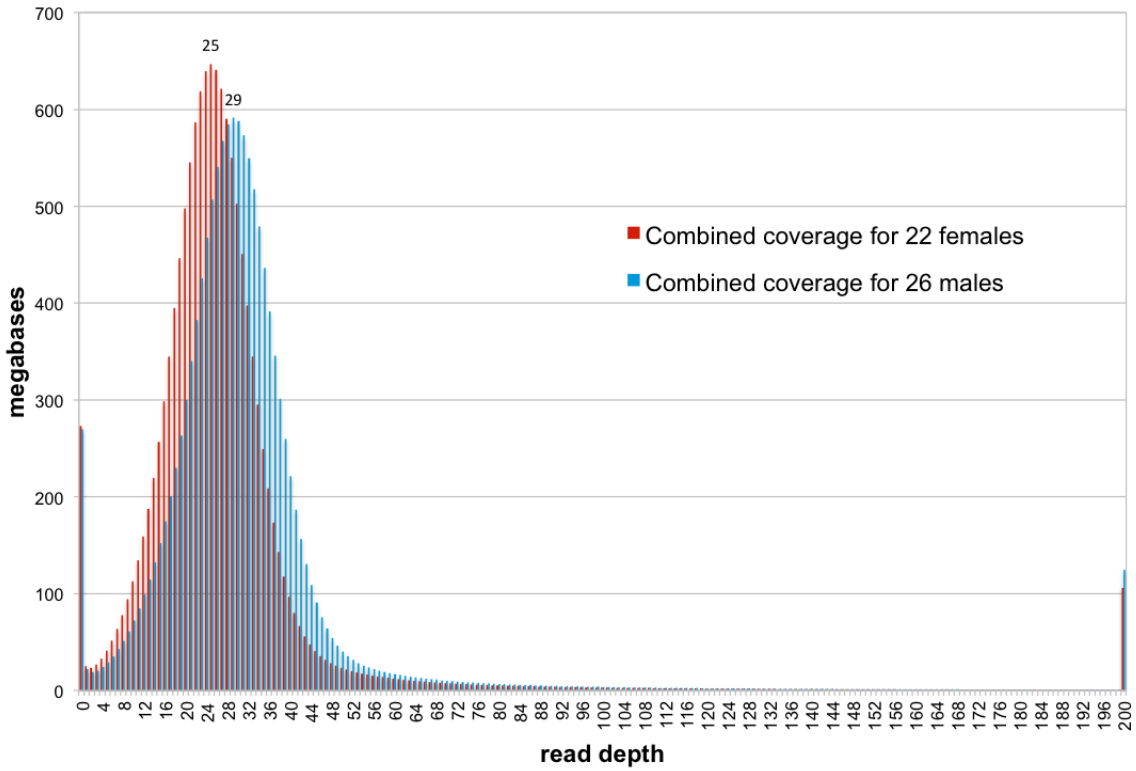

Figure 5.5. Distribution of read depth from combined female and males sequencing data.

Sequence reads from 48 individuals were mapped separately to the female whole genome assembly then alignment files were merged for across all individuals of a given sex (22 females and 26 males). Values represent the number of base pairs of the reference assembly that were sampled at a given depth of coverage. These distributions reveal that the modal coverage of reads from females was lower than the coverage of males, ~25X and ~29X, respectively, consistent with random sampling of sequence across individuals. There is no visible evidence that female sequences map to a larger proportion of the approximate single copy sequence within the female genome. The distribution of coverage ratios is tightly centered on equal coverage and only a small tail corresponds to intervals with higher sequence coverage in female relative to male.

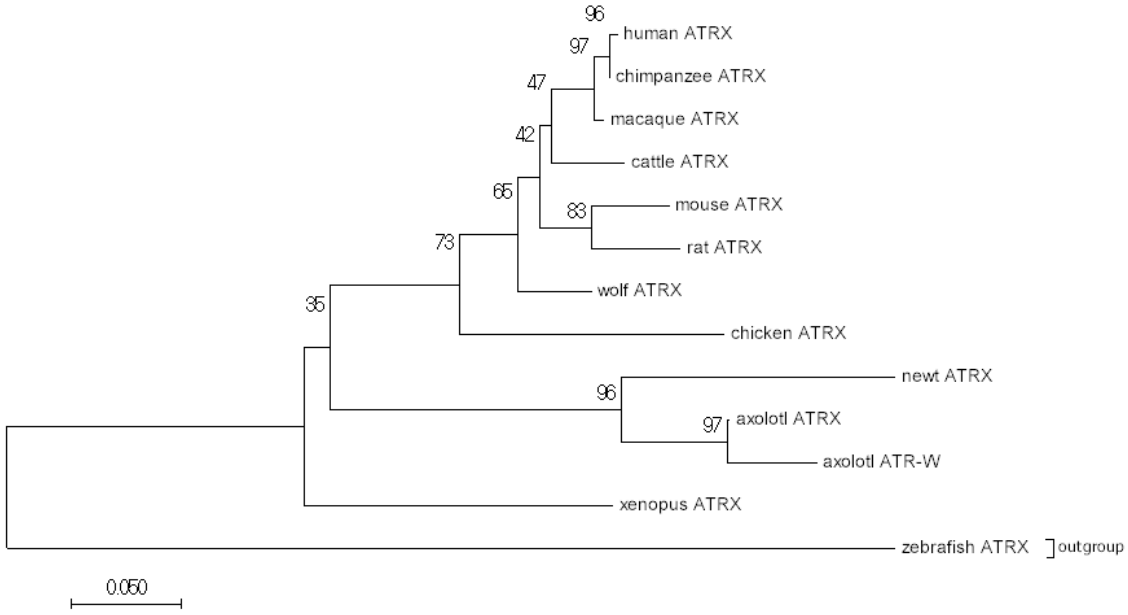

Figure 5.6. Neighbor-Joining tree for vertebrate *ATRX.*

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein 1985). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches (Felsenstein 1985). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004) and are in the units of the number of base substitutions per site. The analysis involved 13 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 251 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al. 2016).

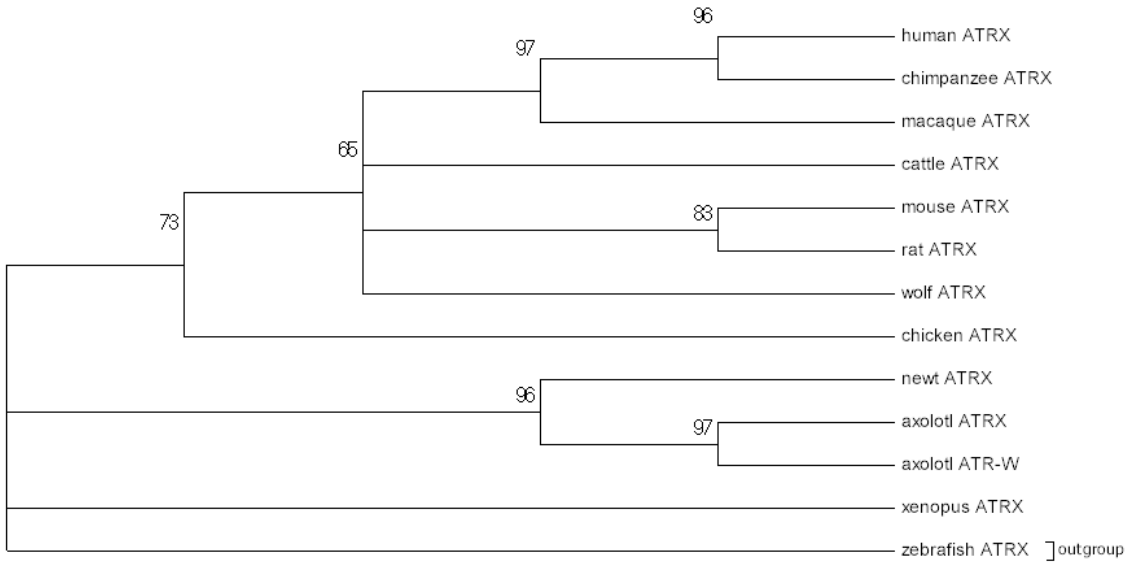

Figure 5.7. Neighbor-Joining tree with bootstraps for vertebrate *ATRX.*
The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches (Felsenstein 1985). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004) and are in the units of the number of base substitutions per site. The analysis involved 13 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 251 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al. 2016).

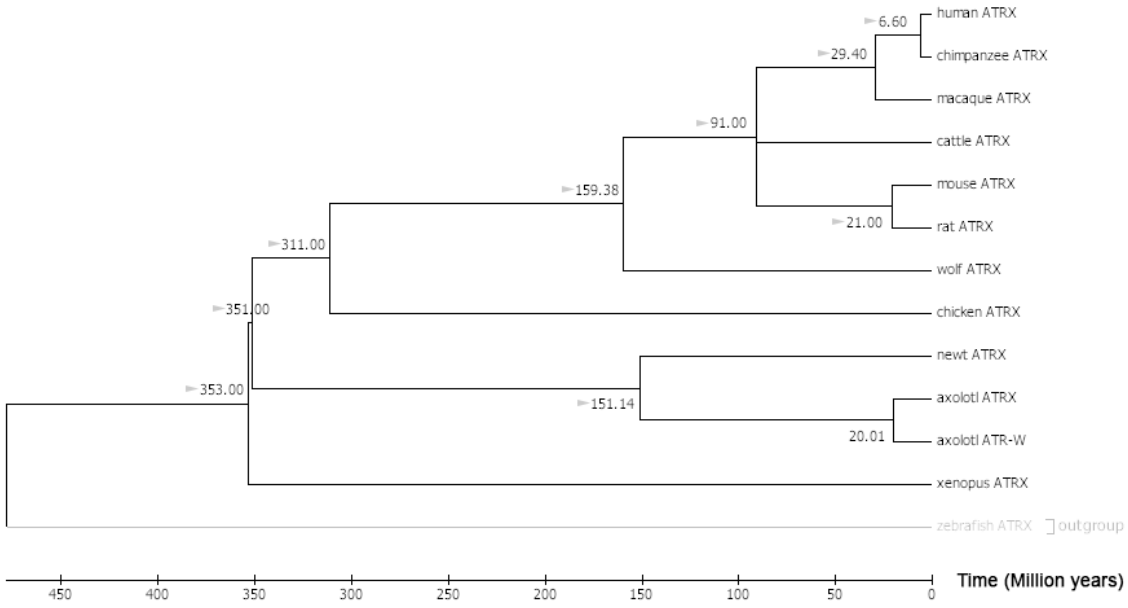

Figure 5.8. Neighbor-Joining vertebrate *ATRX* gene tree with divergence time estimations.

A time-scaled phylogenetic tree inferred using the Reltime method (Tamura et al. 2012) and estimates of branch lengths inferred using the Neighbor-Joining method (Saitou and Nei 1987). The tree was computed using 10 calibration constraints. The analysis involved 13 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 251 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al. 2016). Divergence time ranges estimated by Timetree were added manually and are marked with gray arrows (Hedges et al. 2015). This tree indicates that the duplication event giving rise to *ATRW* in axolotl may have occurred ~20MYA.

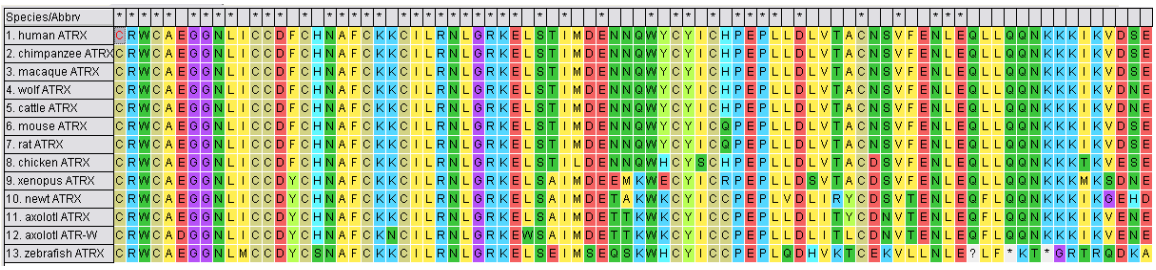

Figure 5.9. Alignment of translated nucleotides from *ATRX* in multiple vertebrate taxa

The alignment from MEGA7 (Kumar et al. 2016) of 84 amino acids of *ATRX* with conservation in 12 vertebrates, including *ATRX* from LG2 in axolotl and *ATRW* show the relative number of changes in codons specific to all amphibians, salamanders, axolotl and axolotl *ATRW.* A total of two out of nine nucleotide substitution events specific to the *ATRW* have altered the predicted codon. While the zebrafish sequence seems to share some conservation, in frame stop codons exist in this alignment and may point to errors in assembly.

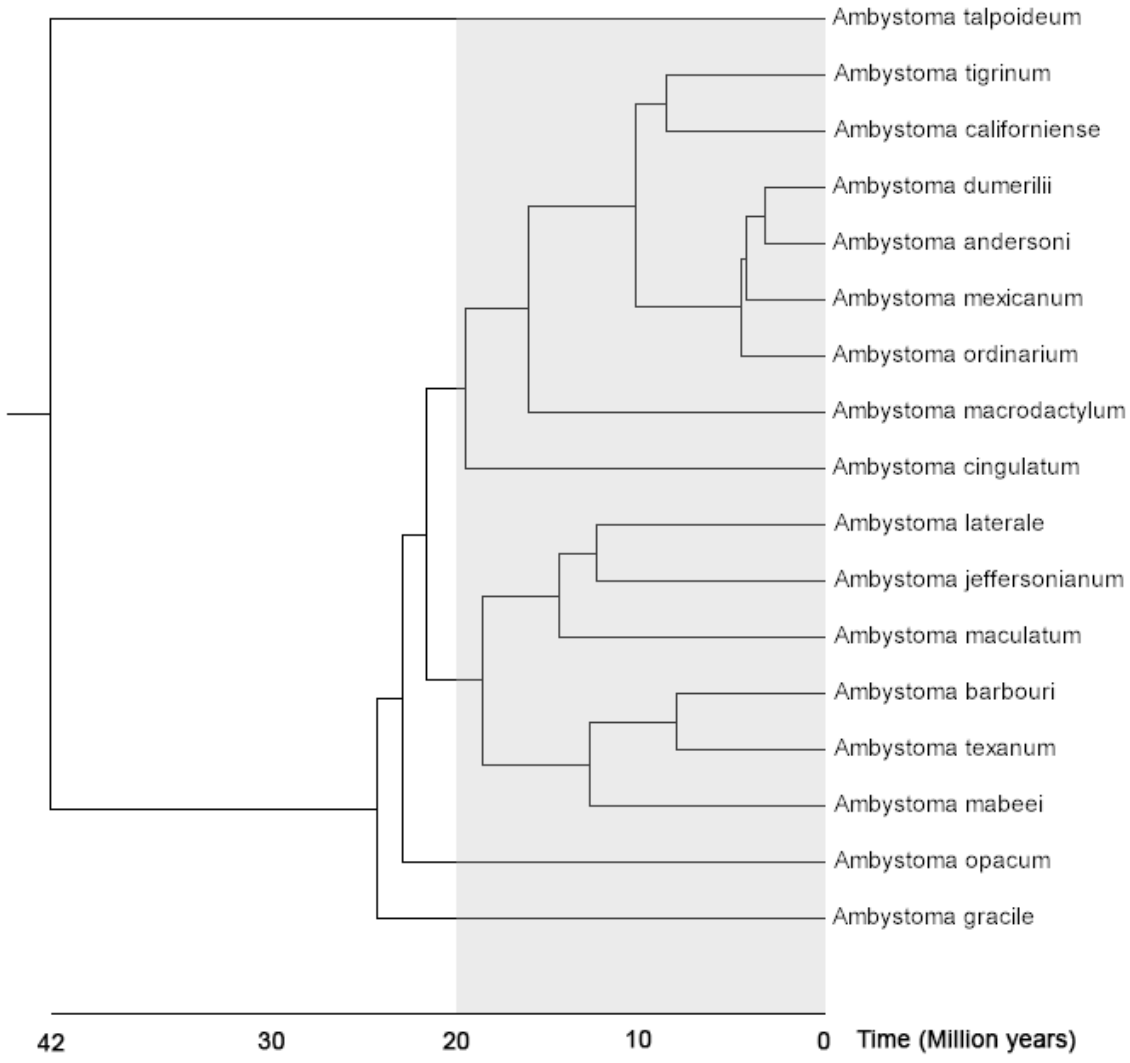

Figure 5.10. A species tree for the genus Ambystoma.

This tree was generated using Timetree (Hedges et al. 2015). The gray shaded region shows those species that may have *ATRW* as predicted by the timing of the duplication event. The tiger salamander complex spans this species list from *A. talpoideum* to *A. cingulatum*.

Table 5.1. Summary statistics for LG9, AM13 and AM14 chromosome assemblies

Summary statistics for de novo assembly of sequence data from the sex chromosome, which corresponds to linkage group 9 (LG9) as well as AM13 and AM14 for comparison as previously published (Keinath et al. 2015). Chromosomes correspond to *A. mexicanum* linkage groups 15/17 (LG15/17) and linkage group 14 (LG14). Statistics are presented for assemblies of raw sequence data (R) and assemblies post error correction (EC).

	Assembly			Contig		Scaffold	
	Length (Mb)	Number of Scaffolds	Number of Singletons	N50 Length Improvement	Proportion Scaffolded	N50 Length Improvement	Number >N50
LG9 (R)	189.7	1,054,224	760,174	118	0.352	256	285,628
LG9 (EC)	242.4	1,232,131	866,817	126 (6.8%)	0.429	295(15.2%)	335,062
LG15/17 (R)	302.5	604,617	243,354	231	0.598	705	136,682
LG15/17 (EC)	210.9	353,381	126,169	295 (28%)	0.643	830 (18%)	82,835
LG14 (R)	180.4	367,575	145,951	232	0.603	686	83,979
LG14 (EC)	143.0	258,214	93,931	290 (25%)	0.636	765 (12%)	62,022

Table 5.2. Blast results to nonredundant protein NCBI database

The table shows best match amino acid alignments for blast (blastx) hit results for all 42 sex-specific scaffolds. 17 scaffolds aligned to a protein-coding gene, and most shared <40% identity. The two highest identity hits to genes were to transcriptional regulator *ATRX* by SuperContig_990642 and mitogen-activated kinase kinase kinase 2 by SuperContig_108441.

Sex-specific Scaffold	Scaffold length (bp)	NCBI Best Hit	Query Cover	E value	% identity	Accession #
SuperContig 1084421	991	PREDICTED: mitogen-activated protein kinase kinase kinase 2-like [Phaethon lepturus]	18%	6.00E-33	98%	XP_010292439.1
SuperContig_990642	1488	PREDICTED: transcriptional regulator ATRX isoform X2 [Alligator sinensis]	17%	6.00E-13	64%	XP_006032758.2
SuperContig_1201750	725	PREDICTED: uncharacterized protein LOC101734340 [Xenopus tropicalis]	14%	0.13	50%	XP_017945915.1
SuperContig_1270996	631	hypothetical protein [Rhodopirellula baltica]	12%	9.7	50%	WP_011119337.1
SuperContig_1240926	668	PREDICTED: uncharacterized protein LOC106589496 [Salmo salar]	39%	2.00E-16	47%	XP_014035031.1
SuperContig_481414	11464	PREDICTED: dynein heavy chain 11, axonemal [Xenopus tropicalis] (reverse transcriptase)	5%	1.00E-32	43%	XP_017952780.1
SuperContig_1139773	843	aminotransferase class I and II [Streptomyces sp. CB00455]	17%	6.2	42%	WP_073917349.1
SuperContig_1398647	510	hypothetical protein [Massilia sp. BSC265]	36%	4.1	40%	WP_051933638.1
SuperContig_113461	850	flagellar autotomy protein [Micromonas pusilla CCMP1545] (reverse transcriptase)	12%	0.55	39%	XP_003062983.1
SuperContig_1105317	928	hypothetical protein A2Z37_15870 [Chloroflexi bacterium RBG_19FT_COMBO_62_14]	12%	6.3	37%	OGO67717.1
SuperContig_960617	1857	PREDICTED: uncharacterized protein LOC106605384 [Salmo salar]	20%	1.8	36%	XP_014056412.1
SuperContig_446459	12684	ORF2 protein [Salmo salar] (reverse transcriptase)	8%	1.00E-36	35%	AKP40998.1
SuperContig 556195	9021	PREDICTED: uncharacterized protein LOC108708171 [Xenopus laevis]	19%	2.00E-70	34%	XP_018102087.1
SuperContig_981147	1581	PREDICTED: LOW QUALITY PROTEIN: dynein heavy chain domain-containing protein 1 [Orcinus orca]	13%	8.4	32%	XP_004279330.1
SuperContig_1025868	1238	DNA primase [Pseudaminobacter manganicus]	23%	9.2	32%	WP_080921700.1
SuperContig_1035909	1185	hypothetical protein T12_433 [Trichinella patagoniensis]	16%	5.5	31%	KRY11477.1
SuperContig_1196200	734	DUF948 domain-containing protein [Lactobacillus buchneri]	41%	4.7	27%	WP_014939867.1

CHAPTER SIX

FUTURE DIRECTIONS

Abstract

Among living vertebrates, the amphibians are largely underrepresented in genomic studies despite their important phylogenetic location and rich biodiversity. In order to fill this gap in vertebrate evolutionary research, the content and structure of amphibian genomes and chromosomes must be investigated. The results of the studies in this dissertation lay the foundation for future studies using *Ambystoma mexicanum* as well as other amphibian species to shed light on vertebrate chromosome and genome evolution. Armed with the knowledge brought about by the studies in this dissertation, along with the new genomic and molecular resources in an ever-changing scientific environment, the research possibilities are endless. Several specific examples of future studies that further the research described in this dissertation are presented below, including improvements to methods for chromosome capture, comparative genomics, development of new amphibian resources, and analyses that may shed further light on the mechanisms and evolution of sex determination in *Ambystoma*.

Improvements to chromosome capture methods

While the methods for laser capture microscopy and subsequent amplification, sequencing and analyses for individual chromosomes were optimized in the laser capture chromosome study (Chapter 2), there were some caveats to using the data for chromosome assembly. Due to biases of the process, assembling chromosomes via modern de Bruijn graph-based genome assemblers, results in fragmentary assemblies (Chapter 3) (Keinath et al. 2015). If pooling chromosomes for a particular species were feasible via fluorescence or size, it may be possible to avoid the whole genome amplification step and associated bias, especially with the ever-decreasing quantity of DNA needed for some sequencing platforms. In the case of the axolotl, assignment of BACs to

individual chromosomes in the karyotype may help to effectively identify and pool chromosomes (Dolezel et al. 2012). Just as Rubicon Genomics worked with our group to alter the PicoPlex kit to accommodate sequencing, it is possible that changes to existing kits may change the way in which these dyads are amplified. With improved amplification, sequencing libraries may better represent the content of these chromosomes, making it a better avenue by which chromosome assemblies can be generated. Alternatively, whole genome sequencing may remain the most suitable option with assemblers and scaffolders paving the way to chromosome-scale assemblies.

Comparative genomics

As more resources become available for larger numbers of amphibian species, new comparative studies can elucidate key features of amphibian chromosomes and genomes. Even without full genome assemblies, linkage maps are providing great perspective on amphibian species. For example, new comparative linkage mapping studies in the common frog (*Rana temporaria*) identified variations in heterochiasmy and recombinational rates but also revealed conserved genome structure (Palomar et al. 2017). A recently published linkage map for the smooth newt (*Lissotriton vulgaris*) was used for speciation research to examine genomic architecture surrounding reproductive isolation in hybridizing newt genomes (Niedzicka et al. 2017). Just as genome wide comparison made among axolotl, newt, xenopus and chicken, Chapters 3 and 4 revealed a specific chromosomal fusion in Ambystoma, among other findings, new comparisons will provide more insight into the evolutionary history of vertebrate genomes and chromosomes.

The development of new amphibian resources is an important step in the improvement of comparative genomics. Mapping data could be generated for many species as the cost of sequencing continues to decrease. Chapter 4 shows that even in a species where only one parent is known, the development of a linkage map is possible. The same methods could be applied to other amphibians, of which, gametes or offspring can be collected. The development of amphibian transcriptomes, which are published regularly, may play a role in the

improvement of linkage mapping and construction of genomic resources for new amphibian species. In Chapter 4, I utilized a published newt transcriptome to identify and genotype polymorphisms (Looso et al. 2013; Keinath et al. 2017). Because this experiment only required obtaining a small number of embryos from a single wild-mated female, these same methods could be presumably applied to generate useful linkage maps for a wide diversity of amphibian species.

Sex chromosome evolution in the axolotl

Several experiments are needed to continue characterizing the evolution of sex chromosomes in the axolotl, in addition to this identification of other salamander and amphibian sex chromosomes. An obvious first step is to better understand the candidate sex-determining gene, *ATRW*, and its contribution toward gonadogenesis in the axolotl. As we only found one of the *ATR* exons in the sex-specific data, it will be important to amplify this the full-length transcript for this gene using a RACE (Rapid Amplification of cDNA Ends) kit or genome walking kit. Following this, RT-PCR may be performed in order to detect RNA expression of the *ATRW* gene during different stages of gonadogenesis. I expect that if *ATRW* plays a major role in sex differentiation, it will be expressed in developing female axolotls throughout gonadogenesis. Understanding when and where *ATRW* is expressed should permit further examination through RNAseq and aid in the development of useful transcriptional profiles for all genes involved in gonadogenesis.

Assessing the functionality of the *ATRW* gene will require more extensive work, including knockouts of the gene. If *ATRW* acts a critical female-determining gene, knockouts would be expected to yield all males. However, even with this result, it will be difficult to know if the *ATRW* gene is the determining gene or if it simply plays an important role in the development of the female phenotype. Studies in mice required transgenics to show *Sry* was sufficient for development of a male from a genetic female, and thus the critical male-determining gene for the species (Koopman et al. 1991). Similarly, by creating transgenic axolotls with the *ATRW* gene, the candidate determining *ATRW* gene can be tested. If the

only gene critical for determining sex were the *ATRW* gene, the expectation for a successful knock-in would be all female progeny.

In order to learn more about the sex-specific region in related ambystomatid salamanders, it would be important to screen a panel of known sex individuals of a variety of species using the primers developed for the axolotl. Because we expect the same sex locus to be working in closely related tiger salamander species that act as evolutionary replicates, verifying the extent to which these sex specific regions apply in the phylogeny may reveal when in evolutionary time these genes were acquired on the W.

New insights about the Z and W chromosomes in the axolotl may contribute to the overall theory of sex chromosome evolution. If the sex chromosomes of the axolotl are in fact young sex chromosomes in the early stages of sex chromosome evolution and headed for heteromorphy, then the characterization of these Z and W may shed light on the earlier stages. If instead, these sex chromosomes present evidence for another theory, such as the high turnover or the fountain of youth, the characterization of the sex chromosomes in a salamander species would be expected to provide new perspective on these theories.

With respect to the high turnover hypothesis, it would be expected that closely related species would provide evidence that the axolotl has undergone a duplication event not present in the other karyotypes, or perhaps the axolotl provides evidence of a fusion resulting in new sex chromosomes in another species. Because hybridization is possible among all tiger salamander species, this is not an expected outcome, however, it may be possible that other ambystomatids (outside of the tiger salamander complex) show that all tiger salamanders have a fusion event leading to new sex chromosomes not seen in the related species. In this case, it may be possible to test if evolutionary forces, such as sexual conflict, drive the turnover of sex chromosomes. For example, one possible avenue proposed for future sex chromosome studies in fish species suggests comparative studies of recently fused (new) sex chromosomes to the ancestral state in other species with the goal of identifying sexually antagonistic

genes that may have caused the fusion event (Yoshida et al. 2014). Because of the size of salamander genomes, this type of study at present may be too ambitious and expensive. With respect to the fountain of youth hypothesis, sex reversed animals, where the genotype is ZW and the phenotype is male) may offer insight into gene dosage or sex-specific gene expression. If the *ATRW* gene is the critical female determining sex for normal female sex differentiation, in the sex-reversed animals (i.e. ZW males), the gene should not be expressed. Regardless of the evolutionary story of these sex chromosomes, the genomic evaluation of sex-specific regions in the axolotl provides one of the first perspectives of sex chromosome evolution in a salamander and in the context of a large vertebrate genome.

Primers designed for sex-specific regions have already positively impacted salamander husbandry, and once publicly released, will continue aiding in experimental studies for which sex could be an important covariate. Because it is difficult to accurately sex axolotls based on appearance prior to the animal producing gametes in adulthood (>9 months), it is often necessary to dissect the animal in order to determine sex. In some cases, the external features that are associated with one sex (such as a swollen cloaca) may be mistaken for the other, which has occurred in the stock center many times. Now that a quick DNA extraction and subsequent PCR can determine sex within hours, and the animal does not need to be euthanized, as DNA can be acquired via blood, a tail clip, limb amputation or a scrape. This is the only sexing option for animals that have undifferentiated gonads.

With initiatives to build new genome assemblies for salamanders, it may be plausible in the future to apply similar depth of coverage analyses as presented in Chapter 5 in other salamander species. New studies are needed to elucidate sex-specific regions of other amphibian species in order to test the rapid turnover hypothesis, add to the evolutionary theory of sex chromosomes in vertebrates, and provide insight for salamander sex chromosome homomorphy.

References

.

Abdullayev I, Kirkham M, Bjorklund AK, Simon A, Sandberg R. 2013. A reference transcriptome and inferred proteome for the salamander Notophthalmus viridescens. *Exp Cell Res* **319**: 1187-1197.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *JMolBiol* **215**: 403-410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Amemiya CT, Alfoldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**: 311-316.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* **188**: 799-808.

Amphibian Survival Alliance, Amphibian Specialist Group. 2014. Genome Resources Working Group.

AmphibiaWeb. 2017. University of California, Berkeley, CA, USA.

Armstrong JB. 1984. Genetic mapping in the Mexican axolotl, Ambystoma mexicanum. *Canadian journal of genetics and cytology Journal canadien de genetique et de cytologie* **26**: 1-6.

Armstrong JB, Duhon ST, Malacinski GM. 1989. Raising the axolotl in captivity. In *Developmental Biology of the Axolotl*, (ed. JB Armstrong, GM Malacinski), pp. 220-227. Oxford University Press, New York.

Axelrod DI. 1980. Contributions to the neogene paleobotany of central California. *University of California Publications in Geological Sciences* **121**: 1-212.

Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grutzner F, Sinclair AH, Oshlack A, Smith CA. 2013. RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol* **14**: R26.

Bachtrog D. 2006. A dynamic view of sex chromosome evolution. *Curr Opin Genet Dev* **16**: 578-585.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, Hahn MW, Kitano J, Mayrose I, Ming R et al. 2014. Sex determination: why so many ways of doing it? *PLoS Biol* **12**: e1001899.

Baker M. 2012. De novo genome assembly: what every biologist should know. *Nature methods* **9**: 333-337.

Baldari CT, Amaldi F. 1976. DNA reassociation kinetics in relation to genome size in four amphibian species. *Chromosoma* **59**: 13-22.

Bentley DR Balasubramanian S Swerdlow HP Smith GP Milton J Brown CG Hall KP Evers DJ Barnes CL Bignell HR et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53-59.

Beukeboom LW, Perrin N. 2014. *The Evolution of Sex Determination*.

Blitz IL. 2012. Navigating the Xenopus tropicalis genome. *Methods Mol Biol* **917**: 43-65.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet* **48**: 427-437.

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R et al. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* **2**: 10.

Britten RJ, Kohne DE. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**: 529-540.

Brockes JP, Kumar A. 2005. Appendage regeneration in adult vertebrates and implications for regenerative medicine. *Science* **310**: 1919-1923.

Bruckskotten M, Looso M, Reinhardt R, Braun T, Borchardt T. 2012. Newt-omics: a comprehensive repository for omics data from the newt Notophthalmus viridescens. *Nucleic Acids Res* **40**: D895-900.

Brunst VV, Hauschka TS. 1963. *Length measurements of the diploid karyotype of the Mexican axolotl (Siredon mexicanum) with reference to a possible sex difference*. XVI International Congress of Zoology, Washington D.C.

Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, Lee TJ, Leigh ND, Kuo TH, Davis FG et al. 2017. A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep* **18**: 762-776.

Bull JJ. 1983. *Evolution of Sex Determining Mechanisms.* Benjamin/Cummings Publ. Co.

Callan HG. 1966. Chromosomes and nucleoli of the axolotl, Ambystoma mexicanum. *Journal of cell science* **1**: 85-108.

Capriglione T, Olmo E, Odierna B, Improta B, Morescalchi A. 1987. Cytoflurometric DNA base determination in vertebrate species with different genome sizes. . *Basic and Applied Histochemistry* **31**: 119-126.

Carlson BM. 1970. Relationship between the tissue and epimorphic regeneration of muscles. *Am Zool* **10**: 175-186.

Cavalier-Smith T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *JCell Sci* **34**: 247-278.

Chalopin D, Volff JN, Galiana D, Anderson JL, Schartl M. 2015. Transposable elements and early evolution of sex chromosomes in fish. *Chromosome Res* **23**: 545-560.

Charlesworth B. 1996. The evolution of chromosomal sex determination and dosage compensation. *CurrBiol* **6**: 149-162.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci* **355**: 1563-1572.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**: 118-128.
Chen M, Song P, Zou D, Hu X, Zhao S, Gao S, Ling F. 2014. Comparison of multiple displacement amplification (MDA) and multiple annealing and looping-based amplification cycles (MALBAC) in single-cell sequencing. *PLoS One* **9**: e114520.
Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC. 2013. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One* **8**: e62856.
Cohen BA, Mitra RD, Hughes JD, Church GM. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet* **26**: 183-186.
Collins JP, Mitton JB, Pierce BA. 1980. Ambystoma tigrinum: a multispecies conglomerate? *Copeia* **1980**: 938-941.
Connallon T, Clark AG. 2010. Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. *Evolution* **64**: 3417-3442.
Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grutzner F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**: 488-493.
Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S et al. 2015. Ensembl 2015. *Nucleic Acids Res* **43**: D662-669.
Cuny R, Malacinski GM. 1985. Banding differences between tiger salamander and axolotl chromosomes. *Canadian journal of genetics and cytology Journal canadien de genetique et de cytologie* **27**: 510-514.
da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, Zepeda-Mendoza ML, Campos PF, Heller R, Pereira RJ. 2016. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Mar Genomics* **30**: 3-13.
Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156-2158.
Davila Lopez M, Martinez Guerra JJ, Samuelsson T. 2010. Analysis of gene order conservation in eukaryotes identifies transcriptionally and functionally linked genes. *PLoS One* **5**: e10654.
de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* **7**: e1002384.
Dolezel J, Vrana J, Safar J, Bartos J, Kubalakova M, Simkova H. 2012. Chromosomes in the flow to simplify genome analysis. *Funct Integr Genomics* **12**: 397-416.
Duellman WE, Trueb L. 1986. *Biology of Amphibians.* McGraw-Hill, New York.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Edstrom JE, Kawiak J. 1961. Microchemical deoxyribonucleic acid determination in individual cells. *The Journal of biophysical and biochemical cytology* **9**: 619-626.
Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323**: 133-138.
Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends Ecol Evol* **29**: 51-63.
Ezaz T, Stiglec R, Veyrunes F, Marshall Graves JA. 2006. Relationships between vertebrate ZW and XY sex chromosome systems. *CurrBiol* **16**: R736-R743.
Fankhauser G, Humphrey RR. 1942. Induction of triploidy and haploidy in axolotl eggs by cold treatment. *Biological Bulletin Marine Biology Laboratory, Woods Hole* **83**: 367-374.
Fei JF, Schuez M, Tazaki A, Taniguchi Y, Roensch K, Tanaka EM. 2014. CRISPR-mediated genomic deletion of Sox2 in the axolotl shows a requirement in spinal cord neural stem cell amplification during tail regeneration. *Stem Cell Reports* **3**: 444-459.
Felsenstein J. 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **39**: 783-791.
Ferris DR, Satoh A, Mandefro B, Cummings GM, Gardiner DM, Rugg EL. 2010. Ex vivo generation of a functional and regenerative wound epithelium from axolotl (Ambystoma mexicanum) skin. *Development, growth & differentiation* **52**: 715-724.
Fishman L, Willis JH, Wu CA, Lee YW. 2014. Comparative linkage maps suggest that fission, not polyploidy, underlies near-doubling of chromosome number within monkeyflowers (Mimulus; Phrymaceae). *Heredity (Edinb)* **112**: 562-568.
Flowers GP, Timberlake AT, McLean KC, Monaghan JR, Crews CM. 2014. Highly efficient targeted mutagenesis in axolotl using Cas9 RNA-guided nuclease. *Development* **141**: 2165-2171.
Frankham R. 2007. Effective population size/adult population size ratios in wildlife: a review. *Genet Res* **89**: 491-503.
Fry JD. 2010. The genomic location of sexually antagonistic variation: some cautionary comments. *Evolution* **64**: 1510-1516.
Geng X, Li W, Shang H, Gou Q, Zhang F, Zang X, Zeng B, Li J, Wang Y, Ma J et al. 2017. A reference gene set construction using RNA-seq of multiple tissues of Chinese giant salamander, Andrias davidianus. *Gigascience* **6**: 1-7.
Gilbert WM. 1936. Amphisexuality and sex differentiation in Ambystoma. State University of Iowa, Unpublished Thesis.
Godwin JW, Pinto AR, Rosenthal NA. 2013. Macrophages are required for adult salamander limb regeneration. *Proc Natl Acad Sci U S A* **110**: 9415-9420.
Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**: 333-351.
Gould SJ. 1977. *Ontogeny and Phylogeny*. Belknap Press of Harvard University Press, Cambridge, MA.

Green DM, Sessions SK. 1991. Amphibian Cytogenetics and Evolution. *Journal of evolutionary biology* **6**: 300-302.
Greenfield P, Duesing K, Papanicolaou A, Bauer DC. 2014. Blue: correcting sequencing errors using consensus and context. *Bioinformatics* **30**: 2723-2732.
Gregory TR. 2002. Genome size and developmental complexity. *Genetica* **115**: 131-146.
Gregory TR. 2015. Animal Genome Size Database.
Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD. 2007. Eukaryotic genome size databases. *Nucleic Acids Res* **35**: D332-D338.
Grime JP, Mowforth MA. 1982. Variation in genome size—an ecological interpretation. *Nature* **299**: 151-153.
Guerrero RF, Kirkpatrick M. 2014. Local adaptation and the evolution of chromosome fusions. *Evolution* **68**: 2747-2756.
Guerrero RF, Kirkpatrick M, Perrin N. 2012. Cryptic recombination in the ever-young sex chromosomes of Hylid frogs. *J Evol Biol* **25**: 1947-1954.
Habermann B, Bebin AG, Herklotz S, Volkmer M, Eckelt K, Pehlke K, Epperlein HH, Schackert HK, Wiebe G, Tanaka EM. 2004. An Ambystoma mexicanum EST sequencing project: analysis of 17,352 expressed sequence tags from embryonic and regenerating blastema cDNA libraries. *Genome biology* **5**: R67.
Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of Life Reveals Clock-Like Speciation and Diversification. *Molecular Biology and Evolution* **32**: 835-845.
Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L et al. 2010. The genome of the Western clawed frog Xenopus tropicalis. *Science* **328**: 633-636.
Hillis DM, Green DM. 1990. Evolutionary changes of heterogametic sex in the phlogenetic history of amphibians. *Journal of evolutionary biology* **3**: 49-64.
Hormozdiari F, Hach F, Sahinalp SC, Eichler EE, Alkan C. 2011. Sensitive and fast mapping of di-base encoded reads. *Bioinformatics* **27**: 1915-1921.
Humphrey R. 1975. The Axolotl, Ambystoma mexicanum. In *Handbook of Genetics*, doi:10.1007/978-1-4613-4470-4_1 (ed. R King), pp. 3-17. Springer US.
Humphrey RR. 1945. Sex determination in the Ambystomatid salamanders: a study of the progeny of females experimentally converted into males. *Am J Anat* **76**: 33-66.
Humphrey RR. 1948. Reversal of sex in females of genotype WW in the axolotl (Siredon or Ambystoma mexicanum) and its bearing upon the role of the Z chromosomes in the development of the testis. *The Journal of experimental zoology* **109**: 171-185.
Humphrey RR, Frankhauser G. 1957. The origin of spontaneous and experimental haploids in the Mexican axolotl (Siredonor Ambystoma-mexicanum). *The Journal of experimental zoology* **134**: 427-447.
Hurst LD, Pal C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* **5**: 299-310.

Hutchison N, Pardue ML. 1975. The mitotic chromosomes of Notophthalmus (=Triturus) viridescens: localization of C banding regions and DNA sequences complementary to 18S, 28S and 5S ribosomal RNA. *Chromosoma* **53**: 51-69.
Huyhn K, Renfree MB, Graves JA, Pask AJ. 2011. ATRX has a critical and conserved role in mammalian sexual differentiation. *BMC Dev Biol* **11**: 39.
Ion A, Telvi L, Chaussain JL, Galacteros F, Valayer J, Fellous M, McElreavey K. 1996. A novel mutation in the putative DNA helicase XH2 is responsible for male-to-female sex reversal associated with an atypical form of the ATR-X syndrome. *Am J Hum Genet* **58**: 1185-1191.
Kai W, Nomura K, Fujiwara A, Nakamura Y, Yasuike M, Ojima N, Masaoka T, Ozaki A, Kazeto Y, Gen K et al. 2014. A ddRAD-based genetic map and its integration with the genome assembly of Japanese eel (Anguilla japonica) provides insights into genome evolution after the teleost-specific genome duplication. *BMC Genomics* **15**: 233.
Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, Fujita M, Suetake H, Suzuki S, Hosoya S et al. 2012. A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, Takifugu rubripes (fugu). *PLoS Genet* **8**: e1002798.
Keinath MC, Timoshevskiy VA, Timoshevskaya NY, Tsonis PA, Voss SR, Smith JJ. 2015. Initial characterization of the large genome of the salamander Ambystoma mexicanum using shotgun and laser capture chromosome sequencing. *Sci Rep* **5**: 16413.
Keinath MC, Voss SR, Panagiotis T, Smith JJ. 2017. A Linkage Map for the Newt Notophthalmus viridescens: Insights in Vertebrate Genome and Chromosome Evolution. *Dev Biol* **426**: 211-218.
Khattak S, Richter T, Tanaka EM. 2009. Generation of transgenic axolotls (Ambystoma mexicanum). *Cold Spring Harbor protocols* **2009**: pdb prot5264.
Khattak S, Schuez M, Richter T, Knapp D, Haigo SL, Sandoval-Guzman T, Hradlikova K, Duemmler A, Kerney R, Tanaka EM. 2013. Germline Transgenic Methods for Tracking Cells and Testing Gene Function during Regeneration in the Axolotl. *Stem Cell Reports* **1**: 90-103.
Kikuchi K, Hamaguchi S. 2013. Novel sex-determining genes in fish and sex chromosome evolution. *Dev Dyn* **242**: 339-353.
Kitano J, Peichel CL. 2012. Turnover of sex chromosomes and speciation in fishes. **94**: 549-558.
Koepfli KP, Paten B, Genome KCoS, O'Brien SJ. 2015. The Genome 10K Project: a way forward. *Annu Rev Anim Biosci* **3**: 57-111.
Köhler J, Vieites DR, Bonett RM, Garcia FH, Glaw F, Steinke D, Vences M. 2005. New Amphibians and Global Conservation: A Boost in Species Discoveries in a Highly Endangered Vertebrate Group. *BioScience* **55**: 693-696.
Kolnicki RL. 2000. Kinetochore reproduction in animal evolution: cell biological explanation of karyotypic fission theory. *Proc Natl Acad Sci U S A* **97**: 9493-9497.
Koonin EV. 2009. Evolution of genome architecture. *Int J Biochem Cell Biol* **41**: 298-306.

Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. 1991. Male development of chromosomally female mice transgenic for Sry. *Nature* **351**: 117-121.
Kragl M, Tanaka EM. 2009a. Axolotl (Ambystoma mexicanum) limb and tail amputation. *Cold Spring Harb Protoc* **2009**: pdb prot5267.
Kragl M, Tanaka EM. 2009b. Grafting axolotl (Ambystoma mexicanum) limb skin and cartilage from GFP+ donors to normal hosts. *Cold Spring Harbor protocols* **2009**: pdb prot5266.
Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639-1645.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**: 1870-1874.
Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: R25.
Laszlo AH, Derrington IM, Ross BC, Brinkerhoff H, Adey A, Nova IC, Craig JM, Langford KW, Samson JM, Daza R et al. 2014. Decoding long nanopore sequencing reads of natural DNA. *Nat Biotechnol* **32**: 829-833.
Lee H, Gurtowski J, Yoo S, Marcus S, McCombie WR, Schatz M. 2014. *Error correction and assembly complexity of single molecule sequencing reads*.
Lee JS, Lee S, Lim BC, Kim KJ, Hwang YS, Choi M, Chae JH. 2015. Alpha-thalassemia X-linked intellectual disability syndrome identified by whole exome sequencing in two boys with white matter changes and developmental retardation. *Gene* **569**: 318-322.
Lercher MJ, Urrutia AO, Hurst LD. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet* **31**: 180-183.
Levy SE, Myers RM. 2016. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet* **17**: 95-115.
Li H. 2012. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* **28**: 1838-1844.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
Li R Fan W Tian G Zhu H He L Cai J Huang Q Cai Q Li B Bai Y et al. 2010. The sequence and de novo assembly of the giant panda genome. *Nature* **463**: 311-317.
Licht LE, Lowcock LA. 1991. Genome Size and Metabolic-Rate in Salamanders. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **100**: 83-92.
Lindsley DL, Fankhauser G, Humphrey RR. 1956. Mapping Centromeres in the Axolotl. *Genetics* **41**: 58-64.
Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012**: 251364.

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic acids research* **40**: W622-627.

Looso M, Preussner J, Sousounis K, Bruckskotten M, Michel CS, Lignelli E, Reinhardt R, Hoffner S, Kruger M, Tsonis PA et al. 2013. A de novo assembly of the newt transcriptome combined with proteomic validation identifies new protein families expressed during tissue regeneration. *Genome Biol* **14**: R16.

Lu H, Giordano F, Ning Z. 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* **14**: 265-279.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**: 18.

Lynch M. 2007. *The Origins of Genome Architecture*. Sinauer Associates, Inc., Sunderland, MA.

Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* **302**: 1401-1404.

Macgregor HC. 1978. Some trends in the evolution of very large chromosomes. *Philos Trans R Soc Lond B Biol Sci* **283**: 309-318.

Maguire F, Henriquez FL, Leonard G, Dacks JB, Brown MW, Richards TA. 2014. Complex patterns of gene fission in the eukaryotic folate biosynthesis pathway. *Genome Biol Evol* **6**: 2709-2720.

Malcom JW, Kudra RS, Malone JH. 2014. The sex chromosomes of frogs: variability and tolerance offer clues to genome evolution and function. *J Genomics* **2**: 68-76.

Mank JE, Avise JC. 2009. Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sex Dev* **3**: 60-67.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764-770.

Martin AP, Naylor GJ, Palumbi SR. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature* **357**: 153-155.

McAllister BF. 2000. Fixation of Chromosomal Rearrangements. In *Comparative Genomics*, Vol 1 (ed. D Sankoff, JH Nadeau). Springer-Science+Business Media, B.V.

McCusker CD, Gardiner DM. 2014. Understanding positional cues in salamander limb regeneration: implications for optimizing cell-based regenerative therapies. *Dis Model Mech* **7**: 593-599.

McElreavey K, Fellous M. 1997. Sex-determining genes. *Trends EndocrinolMetab* **8**: 342-346.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297-1303.

McKnight ML, Shaffer HB. 1997. Large, rapidly evolving intergenic spacers in the mitochondrial DNA of the salamander family Ambystomatidae (Amphibia: Caudata). *Molecular biology and evolution* **14**: 1167-1176.

Miura I. 2007. An evolutionary witness: the frog rana rugosa underwent change of heterogametic sex from XY male to ZW female. *Sex Dev* **1**: 323-331.
Monaghan JR, Epp LG, Putta S, Page RB, Walker JA, Beachy CK, Zhu W, Pao GM, Verma IM, Hunter T et al. 2009. Microarray and cDNA sequence analysis of transcription during nerve-dependent limb regeneration. *BMC biology* **7**: 1.
Monaghan JR, Maden M. 2012. Visualization of retinoic acid signaling in transgenic axolotls during limb development and regeneration. *Developmental biology* **368**: 63-75.
Morescalchi A, Galgano M, Gargiulo G. 1973. Effects of cold and colcemid on the chromosomes of the fire-bellied toad, Bombina variegata pachypus Bonaparte. *Riv Biol* **66**: 183-214.
Morescalchi A, Olmo E, Serra V. 1974. Chromosomes and DNA of the ambystomatoid salamanders. *Experientia* **30**: 619-620.
Morescalchi A, Serra V. 1974. DNA renaturation kinetics in some paedogenetic Urodeles. *Experientia* **30**: 487-489.
Mueller RL, Gregory TR, Gregory SM, Hsieh A, Boore JL. 2008. Genome size, cell size, and the evolution of enucleated erythrocytes in attenuate salamanders. *Zoology (Jena)* **111**: 218-230.
Nachman MW, Searle JB. 1995. Why is the house mouse karyotype so variable? *Trends Ecol Evol* **10**: 397-402.
Nadeau JH, Taylor BA. 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Natl Acad Sci U S A* **81**: 814-818.
National Center for Biotechnology Information. 2004. Homologene [Internet]. Bethesda (MD) National Library of Medicine (US).
National Research Council (US) Subcommittee on Amphibian Standards. 1974. Amphibians: Guidelines for breeding, care, and management of laboratory animals. In *Classification and Description of Amphibians Commonly Used for Laboratory Research*, Vol II. National Academies Press (US), Washington (DC).
Neri G, Opitz J. 1999. Syndromal (and nonsyndromal) forms of male pseudohermaphroditism. *AmJMedGenet* **89**: 201-209.
Niedzicka M, Dudek K, Fijarczyk A, Zielinski P, Babik W. 2017. Linkage Map of Lissotriton Newts Provides Insight into the Genetic Basis of Reproductive Isolation. *G3 (Bethesda)* **7**: 2115-2124.
Nussbaum RA, Wilkinson M. 1989. On the Classification and Phylogeny of Caecilians (*Amphibia: Gymnophiona*), a critical review. *Herpetological Monographs* **3**: 1-42.
O'Neill EM, Schwartz R, Bullock CT, Williams JS, Shaffer HB, Aguilar-Miguel X, Parra-Olea G, Weisrock DW. 2013. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (Ambystoma tigrinum) species complex. *Mol Ecol* **22**: 111-129.
Ogata M, Ohtani H, Igarashi T, Hasegawa Y, Ichikawa Y, Miura I. 2003. Change of the heterogametic sex from male to female in the frog. *Genetics* **164**: 613-620.
Organ CL, Shedlock AM. 2009. Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biol Lett* **5**: 47-50.
Page RB, Boley MA, Kump DK, Voss SR. 2013. Genomics of a metamorphic timing QTL: met1 maps to a unique genomic position and regulates morph and

species-specific patterns of brain transcription. *Genome Biol Evol* **5**: 1716-1730.
Page RB, Voss SR, Samuels AK, Smith JJ, Putta S, Beachy CK. 2008. Effect of thyroid hormone concentration on the transcriptional response underlying induced metamorphosis in the Mexican axolotl (Ambystoma). *BMC genomics* **9**: 78.
Palaiokostas C, Bekaert M, Davie A, Cowan ME, Oral M, Taggart JB, Gharbi K, McAndrew BJ, Penman DJ, Migaud H. 2013. Mapping the sex determination locus in the Atlantic halibut (Hippoglossus hippoglossus) using RAD sequencing. *BMC Genomics* **14**: 566.
Palomar G, Ahmad F, Vasemagi A, Matsuba C, Nicieza AG, Cano JM. 2017. Comparative High-Density Linkage Mapping Reveals Conserved Genome Structure but Variation in Levels of Heterochiasmy and Location of Recombination Cold Spots in the Common Frog. *G3 (Bethesda)* **7**: 637-645.
Pardo-Manuel de Villena F, Sapienza C. 2001. Female meiosis drives karyotypic evolution in mammals. *Genetics* **159**: 1179-1189.
Pask A, Renfree MB, Marshall Graves JA. 2000. The human sex-reversing ATRX gene has a homologue on the marsupial Y chromosome, ATRY: implications for the evolution of mammalian sex determination. *ProcNatlAcadSciUSA* **97**: 13198-13202.
Pennell MW, Kirkpatrick M, Otto SP, Vamosi JC, Peichel CL, Valenzuela N, Kitano J. 2015. Y fuse? Sex chromosome fusions in fishes and reptiles. *PLoS Genet* **11**: e1005237.
Perrin N. 2009. Sex reversal: a fountain of youth for sex chromosomes? *Evolution* **63**: 3043-3049.
Pevzner P, Tesler G. 2003. Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. *Genome Res* **13**: 37-45.
Pevzner PA, Tang H, Waterman MS. 2001. An Eulerian path approach to DNA fragment assembly. *Proc Natl Acad Sci U S A* **98**: 9748-9753.
Phillippy AM. 2017. New advances in sequence assembly. *Genome Res* **27**: xi-xiii.
Putta S, Smith JJ, Walker JA, Rondet M, Weisrock DW, Monaghan J, Samuels AK, Kump K, King DC, Maness NJ et al. 2004. From biomedicine to natural history research: EST resources for ambystomatid salamanders. *BMCGenomics* **5**: 54.
Rand DM, Dorfsman M, Kann LM. 1994. Neutral and non-neutral evolution of Drosophila mitochondrial DNA. *Genetics* **138**: 741-756.
Rao N, Jhamb D, Milner DJ, Li B, Song F, Wang M, Voss SR, Palakal M, King MW, Saranjami B et al. 2009. Proteomic analysis of blastema formation in regenerating axolotl limbs. *BMC biology* **7**: 83.
Rastas P, Calboli FCF, Guo B, Shikano T, Merilä J. 2016. Construction of Ultradense Linkage Maps with Lep-MAP2: Stickleback F2 Recombinant Crosses as an Example. *Genome Biology and Evolution* **8**: 78-93.
Rice WR. 1984. Sex Chromosomes and the Evolution of Sexual Dimorphism. *Evolution* **38**: 735-742.
Rice WR. 1987. The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. *Evolution* **41**: 911-914.

Roberts RJ, Carneiro MO, Schatz MC. 2013. The advantages of SMRT sequencing. *Genome Biol* **14**: 405.
Rosbash M, Ford PJ, Bishop JO. 1974. Analysis of the C-value paradox by molecular hybridization. *Proceedings of the National Academy of Sciences of the United States of America* **71**: 3746-3750.
Roth G, Blanke J, Wake DB. 1994. Cell size predicts morphological complexity in the brains of frogs and salamanders. *Proc Natl Acad Sci U S A* **91**: 4796-4800.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.
Sambrook J, Russell DW. 2001. *Molecular Cloning: A laboratory Manual 3rd Eds*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Sambrook J, Russell DW. 2006. Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc* **2006**.
Sanger F, Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**: 441-448.
Sanger F, Coulson AR, Friedmann T, Air GM, Barrell BG, Brown NL, Fiddes JC, Hutchison CA, 3rd, Slocombe PM, Smith M. 1978. The nucleotide sequence of bacteriophage phiX174. *J Mol Biol* **125**: 225-246.
Schartl M. 2004. Sex chromosome evolution in non-mammalian vertebrates. *CurrOpinGenetDev* **14**: 634-641.
Schmid M, Nanda I, Steinlein C, Kausch K, Epplen JT, Haaf T. 1991. Sex determining mechanisms and sex chromosomes in amphibia. In *Amphibian Cytogenetics and Evolution*, (ed. DM Green, SK Sessions), pp. 393-430. Academic Press, New York.
Schmid M, Steinlein C. 2001. Sex chromosomes, sex-linked genes, and sex determination in the vertebrate class amphibia. *EXS*: 143-176.
Schreckenberg GM, Jacobson AG. 1975. Normal stages of development of the axolotl. Ambystoma mexicanum. *Developmental biology* **42**: 391-400.
Schubert I, Rieger R, Fuchs J. 1995. Alteration of basic chromosome number by fusion-fission cycles. *Genome* **38**: 1289-1292.
Seifertova E, Zimmerman LB, Gilchrist MJ, Macha J, Kubickova S, Cernohorska H, Zarsky V, Owens ND, Sesay AK, Tlapakova T et al. 2013. Efficient high-throughput sequencing of a laser microdissected chromosome arm. *BMC Genomics* **14**: 357.
Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, Fukui A, Hikosaka A, Suzuki A, Kondo M et al. 2016. Genome evolution in the allotetraploid frog Xenopus laevis. *Nature* **538**: 336-343.
Sessions SK. 1982. Cytogenetics of diploid and triploid salamanders of the *Ambystoma jeffersonianum* complex. *Chromosoma* **77**: 599-621.
Sessions SK. 2008. Evolutionary cytogenetics in salamanders. *Chromosome Res* **16**: 183-201.
Sessions SK, Bizjak Mali L, Green DM, Trifonov V, Ferguson-Smith M. 2016. Evidence for Sex Chromosome Turnover in Proteid Salamanders. *Cytogenet Genome Res* **148**: 305-313.

Sessions SK, Larson A. 1987. Developmental Correlates of Genome Size in Plethodontid Salamanders and Their Implications for Genome Evolution. *Evolution* **41**: 1239-1251.
Shaffer HB. 1984a. Evolution in a Paedomorphic Lineage. I. An Electrophoretic Analysis of the Mexican Ambystomatid Salamanders. *Evolution* **38**: 1194-1206.
Shaffer HB. 1984b. Evolution in a Paedomorphic Lineage. II. Allometry and Form in the Mexican Ambystomatid Salamanders. *Evolution* **38**: 1207-1218.
Shaffer HB, McKnight ML. 1996. The polytypic species revisited: genetic differentiation and molecular phylogenetics of the tiger salamander (Ambystoma tigrinum) (Amphibia: Caudata) complex. *Evolution* **50**: 417-433.
Smit AFA, Hubley R. 2015. RepeatModeler Open-1.0.
Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-3.0
Smith JJ, Keinath MC. 2015. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Res* **25**: 1081-1090.
Smith JJ, Kump DK, Walker JA, Parichy DM, Voss SR. 2005a. A comprehensive expressed sequence tag linkage map for tiger salamander and Mexican axolotl: enabling gene mapping and comparative genomics in Ambystoma. *Genetics* **171**: 1161-1171.
Smith JJ, Putta S, Walker JA, Kump DK, Samuels AK, Monaghan JR, Weisrock DW, Staben C, Voss SR. 2005b. Sal-Site: integrating new and existing ambystomatid salamander research and informational resources. *BMCGenomics* **6**: 181.
Smith JJ, Putta S, Zhu W, Pao GM, Verma IM, Hunter T, Bryant SV, Gardiner DM, Harkins TT, Voss SR. 2009. Genic regions of a large salamander genome contain long introns and novel genes. *BMCGenomics* **10**: 19.
Smith JJ, Voss SR. 2006. Gene order data from a model amphibian (Ambystoma): new perspectives on vertebrate genome structure and evolution. *BMCGenomics* **7**: 219.
Smith JJ, Voss SR. 2007. Bird and mammal sex-chromosome orthologs map to the same autosomal region in a salamander (ambystoma). *Genetics* **177**: 607-613.
Smith JJ, Voss SR. 2009. Amphibian sex determination: segregation and linkage analysis using members of the tiger salamander species complex (Ambystoma mexicanum and A. t. tigrinum). *Heredity* **102**: 542-548.
Sobkow L, Epperlein HH, Herklotz S, Straube WL, Tanaka EM. 2006. A germline GFP transgenic axolotl and its use to track cell fate: dual origin of the fin mesenchyme during development and the fate of blood cells during regeneration. *Developmental biology* **290**: 386-397.
Stam P. 1993. Construction of Integrated Genetic-Linkage Maps by Means of a New Computer Package - Joinmap. *Plant Journal* **3**: 739-744.
Stevenson RE. 1993. Alpha-Thalassemia X-Linked Intellectual Disability Syndrome. In *GeneReviews(R)*, (ed. RA Pagon, et al.), Seattle (WA).
Stock M, Croll D, Dumas Z, Biollay S, Wang J, Perrin N. 2011a. A cryptic heterogametic transition revealed by sex-linked DNA markers in Palearctic green toads. *J Evol Biol* **24**: 1064-1070.

Stock M, Horn A, Grossen C, Lindtke D, Sermier R, Betto-Colliard C, Dufresnes C, Bonjour E, Dumas Z, Luquet E et al. 2011b. Ever-young sex chromosomes in European tree frogs. *PLoS Biol* **9**: e1001062.

Stock M, Savary R, Betto-Colliard C, Biollay S, Jourdan-Pineau H, Perrin N. 2013. Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic green toads (Bufo viridis subgroup). *J Evol Biol* **26**: 674-682.

Straus NA. 1971. Comparative DNA renaturation kinetics in amphibians. *Proceedings of the National Academy of Sciences of the United States of America* **68**: 799-802.

Sun C, Mueller RL. 2014. Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. *Genome Biol Evol* **6**: 1818-1829.

Sun C, Shepard DB, Chong RA, Lopez Arriaza J, Hall K, Castoe TA, Feschotte C, Pollock DD, Mueller RL. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biol Evol* **4**: 168-183.

Sun YB, Xiong ZJ, Xiang XY, Liu SP, Zhou WW, Tu XL, Zhong L, Wang L, Wu DD, Zhang BL et al. 2015. Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. *Proc Natl Acad Sci U S A* **112**: E1257-1262.

Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. 2012. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A* **109**: 19333-19338.

Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* **101**: 11030-11035.

Tanaka K, Takehana Y, Naruse K, Hamaguchi S, Sakaizumi M. 2007. Evidence for different origins of sex chromosomes in closely related Oryzias fishes: substitution of the master sex-determining gene. *Genetics* **177**: 2075-2081.

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**: 2032-2034.

The World Conservation Union. 2014. Table 1: Numbers of threatened species by major groups of organisms (1996-2014). In *IUCN Red List of Threatened Species*, (ed. SSfGT Species).

Timoshevskiy VA, Lampman RT, Hess JE, Porter LL, Smith JJ. 2017. Deep ancestry of programmed genome rearrangement in lampreys. *Dev Biol* **429**: 31-34.

Timoshevskiy VA, Sharma A, Sharakhov IV, Sharakhova MV. 2012. Fluorescent in situ hybridization on mitotic chromosomes of mosquitoes. *J Vis Exp* doi:10.3791/4215: e4215.

Treangen TJ, Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* **13**: 36-46.

Tree of Sex Consortium. 2014. Tree of Sex: A database of sexual systems. *Scientific Data* **1**: 140015.

Uno Y, Nishida C, Takagi C, Ueno N, Matsuda Y. 2013. Homoeologous chromosomes of Xenopus laevis are highly conserved after whole-genome duplication. *Heredity (Edinb)* **111**: 430-436.

Unruh JR. 1991. The uplift of the Sierra Nevada and implications for late cenozoic epeirogeny in the western cordillera. *Geological Society of America Bulletin* **103**: 1395-1404.
Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**: e115.
Van Ooijen JW. 2011. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genetics research* **93**: 343-349.
Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M et al. 2014. Elephant shark genome provides unique insights into gnathostome evolution (vol 505, pg 174, 2014). *Nature* **513**.
Venter JC Adams MD Myers EW Li PW Mural RJ Sutton GG Smith HO Yandell M Evans CA Holt RA et al. 2001. The sequence of the human genome. *Science* **291**: 1304-1351.
Veyrunes F, Waters PD, Miethke P, Rens W, McMillan D, Alsop AE, Grutzner F, Deakin JE, Whittington CM, Schatzkamer K et al. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res* **18**: 965-973.
Vicoso B, Bachtrog D. 2013. Reversal of an ancient sex chromosome to an autosome in Drosophila. *Nature* **499**: 332-335.
Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013a. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol* **11**: e1001643.
Vicoso B, Kaiser VB, Bachtrog D. 2013b. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 6453-6458.
Villolobos M, Leon P, Sessions SK, Kezer J. 1988. Enucleated Erythrocytes in Plethodontid Salamanders. *Herpetologica* **44**: 243-250.
Vinogradov AE. 1998. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. *Cytometry* **31**: 100-109.
Voskoboynik A, Neff NF, Sahoo D, Newman AM, Pushkarev D, Koh W, Passarelli B, Fan HC, Mantalas GL, Palmeri KJ et al. 2013. The genome sequence of the colonial chordate, Botryllus schlosseri. *Elife* **2**: e00569.
Voss GJ, Kump DK, Walker JA, Voss SR. 2013a. Variation in salamander tail regeneration is associated with genetic factors that determine tail morphology. *PloS one* **8**: e67274.
Voss SR, Epperlein HH, Tanaka EM. 2009. Ambystoma mexicanum, the axolotl: a versatile amphibian model for regeneration, development, and evolution studies. *Cold Spring Harbor protocols* **2009**: pdb emo128.
Voss SR, Kump DK, Putta S, Pauly N, Reynolds A, Henry RJ, Basa S, Walker JA, Smith JJ. 2011. Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome research* doi:10.1101/gr.116491.110.
Voss SR, Kump DK, Walker JA, Shaffer HB, Voss GJ. 2012. Thyroid hormone responsive QTL and the evolution of paedomorphic salamanders. *Heredity (Edinb)* **109**: 293-298.

Voss SR, Palumbo A, Nagarajan R, Gardiner DM, Muneoka K, Stromberg AJ, Athippozhy AT. 2015. Gene expression during the first 28 days of axolotl limb regeneration. I: Experimental design and global analysis of gene expression. *Regeneration* **In Press**.

Voss SR, Putta S, Walker JA, Smith JJ, Maki N, Tsonis PA. 2013b. Salamander Hox clusters contain repetitive DNA and expanded non-coding regions: a typical Hox structure for non-mammalian tetrapod vertebrates? *Human genomics* **7**: 9.

Voss SR, Shaffer HB. 1996. What insights into the developmental traits of urodeles does the study of interspecific hybrids provide? *The International journal of developmental biology* **40**: 885-893.

Voss SR, Shaffer HB. 1997. Adaptive evolution via a major gene effect: paedomorphosis in the Mexican axolotl. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 14185-14189.

Voss SR, Smith JJ. 2005. Evolution of salamander life cycles: a major-effect quantitative trait locus contributes to discrete and continuous variation for metamorphic timing. *Genetics* **170**: 275-281.

Voss SR, Smith JJ, Gardiner DM, Parichy DM. 2001. Conserved vertebrate chromosome segments in the large salamander genome. *Genetics* **158**: 735-746.

Wallace H, Badawy GM, Wallace BM. 1999. Amphibian sex determination and sex reversal. *Cell MolLife Sci* **55**: 901-909.

Wang Z, Zhang J, Yang W, An N, Zhang P, Zhang G, Zhou Q. 2014. Temporal genomic evolution of bird sex chromosomes. *BMC Evol Biol* **14**: 250.

Waters PD, Wallis MC, Marshall Graves JA. 2007. Mammalian sex--Origin and evolution of the Y chromosome and SRY. *SeminCell DevBiol* **18**: 389-400.

White MJ. 1973. Animal cytology and evolution. *Cambridge*.

Wickbom T. 1950. THE CHROMOSOMES OF ASCAPHUS TRUEI AND THE EVOLUTION OF THE ANURAN KARYOTYPES. *Hereditas* **36**: 406-418.

Woodcock MR, Vaughn-Wolfe J, Elias A, Kump DK, Kendall KD, Timoshevskaya N, Timoshevskiy V, Perry DW, Smith JJ, Spiewak JE et al. 2017. Identification of Mutant Genes and Introgressed Tiger Salamander DNA in the Laboratory Axolotl, Ambystoma mexicanum. *Sci Rep* **7**: 6.

Yasue A, Mitsui SN, Watanabe T, Sakuma T, Oyadomari S, Yamamoto T, Noji S, Mito T, Tanaka E. 2014. Highly efficient targeted mutagenesis in one-cell mouse embryos mediated by the TALEN and CRISPR/Cas systems. *Sci Rep* **4**: 5705.

Yoshida K, Makino T, Yamaguchi K, Shigenobu S, Hasebe M, Kawata M, Kume M, Mori S, Peichel CL, Toyoda A et al. 2014. Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLoS Genet* **10**: e1004223.

You X, Shu L, Li S, Chen J, Luo J, Lu J, Mu Q, Bai J, Xia Q, Chen Q et al. 2013. Construction of high-density genetic linkage maps for orange-spotted grouper Epinephelus coioides using multiplexed shotgun genotyping. *BMC Genet* **14**: 113.

Zhang P, Wake DB. 2009. Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular phylogenetics and evolution* **53**: 492-508.
Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *JComputBiol* **7**: 203-214.
Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, Gilbert MT, Zhang G. 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* **346**: 1246338.

VITA

NAME: Melissa Keinath

EDUCATION: B.S., Biology, 2010
University of Kentucky
Lexington, KY, USA

PROFESSIONAL PUBLICATIONS:

Smith, J. J., **Keinath, M. C.** The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Research*. 2015 Aug;25(8):1081-90. Doi: 10.1101/gr.184135.114. Epub 2015 Jun 5.

Keinath M.C., Timoshevskiy V.A., Timoshevskaya N.Y., Tsonis P.A., Voss S.R., Smith, J.J. Initial characterization of the large genome of the salamander Ambystoma mexicanum using shotgun and laser capture chromosome sequencing. *Scientific Reports*. 2015;5:16413. doi: 10.1038/srep16413.

Keinath M.C., Voss S.R., Tsonis, P.A., Smith, J.J. A Linkage Map for the Newt *Notophthalmus viridescens*: Insights in Vertebrate Genome and Chromosome Evolution. *Developmental Biology*. 2017 Jun 15;426(2):211-218. doi: 10.1016/j.ydbio.2016.05.027. Epub 2016 Jun 2.

Timoshevskiy, V.A., Herdy, J.R., **Keinath, M.C.**, Smith, J.J. Cellular and Molecular Features of Developmentally Programmed Genome Rearrangement in a Vertebrate (Sea Lamprey: Petromyzon marinus). *PLOS Genetics*. 2016 Jun 24;12(6):e1006103. doi: 10.1371/journal.pgen.1006103. eCollection 2016 Jun.