Evolution of sex determination in Pipid frogs

Sex determination in frogs offers insights into how important things evolve

By Caroline Marie Sophie Cauret,

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University
Department of Biology
Doctor of Philosophy (2020)
Hamilton, Ontario

TITLE: Sex determination in frogs offers insights into how important things evolve
AUTHOR: Caroline Marie Sophie Cauret, MSc (Université Paul Sabatier)
SUPERVISOR: Dr. Ben J. Evans
COMMITTEE: Dr. Marie A. Elliot, Dr. G. Brian Golding
NUMBER OF PAGES: xiv, 186

Abstract

Developmental system drift refers to situations where, in different species, similar traits are controlled by diverged genetic systems. For example, sexual differentiation is crucial for reproduction and survival of species, but is controlled by rapidly evolving genes and genetic interactions. In my PhD thesis, I studied the evolution of triggers for sex determination in order to better understand how developmental systems drift occurs. The second chapter examines the diversity of sex determining systems in the frog family Pipidae (which includes the genus *Xenopus*). My work provides evidence for seven distinct sex determination systems in this group of frogs. This work also shows that a known trigger for sex determination (*dm-w*) has not always been efficient in driving female development of frogs that carry this gene. I propose that the inefficiency of a trigger for sex determination might play a major role in the fate of sex chromosomes. In the third chapter, I explore potential mechanisms by which *dm-w* and the supergene that contains *dm-w* evolved, and possible effects in its efficiency in driving female sexual differentiation. This work shows that the acquisition and evolution of the last exon of this gene might be associated with the onset of female-specificity. While the specific function of two other female-specific genes in *X. laevis* (*scan-w*, *ccdc69-w*) remains unclear, their origin seems to coincide with the empowerment of *dm-w*. We also found that an important domain in the function of *dmrt1*, another transcription factor that is related by gene duplication to *dm-w*, is often incapacitated by early stop codons. Taken together, this thesis documents dynamic functional evolution of individual components (*dm-w*, *dmrt1*, *scan-w*, *ccdc69-w*), entire genetic networks (sex chromosomes), and genomic contexts (recombination suppression) that govern an important phenotype (sex determination) in a diverse group of widely studied vertebrates (pipid frogs).

Acknowledgements

Thank you to my supervisor Ben Evans. You got me involved in many projects, including various international collaborations. I can't thank you enough for your understanding through all the ups and downs and for always encouraging me.

To my committee members Marie Elliot and Brian Golding. I could not have had better committee members than you. Thank you for all the support you both gave me, especially during Ben's sabbatical. Marie, I will also never forget how you made sure I had someone to spend holidays with my first year in Canada. Brian, thank you for complaining about the French language as much as I complained about English. Thank you for asking me to join your lab and checked on me when our lab was empty. I was lucky enough to have a supervisor and committee members as good as scientists as they are as people.

To the Horb lab. Thank you Marko for hosting us in your lab multiple years. Danielle, Nikko, Marcin thank you for the sassiness and the collaboration.

Thank you to my family. My parents, my brother and my nephews. Thank you for always believing in me and encouraging me in everything I ever wanted to do. My best friend and partner in all crimes, Andrew Tupper, I cannot imagine any milestone in my life without you by my side. All these years would have been so much harder without you bothering me all the way. Andy and Chris thank you for welcoming me in your family.

My friends from both sides of the Ocean. A special one to Marie. A crazy friendship started with horses that will never end. Seb thanks for being the happiest person when I told you about leaving for Canada. Sarah, my best enemy and roommate, for all the Canadian adventures. Renée, Zac and Sam for always making France closer.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Caroline Marie Sophie Cauret, declare that this thesis titled, "Sex determination in frogs offers insights into how important things evolve" and the work presented in it are my own. I confirm that: for Chapter 2 (published in *Molecular Biology and Evolution*) and Chapter 3 (in preparation for publication), I performed most of the analysis with input from Dr. Ben Evans. Preparation of the manuscripts was performed by myself and Ben Evans; co-authors contributing to the editing. Specific contributions of the co-authors are listed below. Marie-Theres T. Gansauge and Matthias Meyer performed the acquisition of the capture data for both studies. Ben Evans, Eli Greenbaum, Václav Gvoždík provided important frogs specimens. For Chapter 2, scripting was done by myself and Andrew S. Tupper; the mitochondrial phylogeny was estimated by Ben Evans. Knock out experiments for Chapter 3 were performed by myself, Ben Evans, Danielle Jordan and Marko Horb. Sanger Sequencing was done by myself, Ben Evans, Benjamin L. S. Furman, Martin Knytl, Xue Song, and Veronica Lopardo. Other analyses not itemized here were performed by myself with input from Ben Evans. Papers to which I have contributed but I am not the lead author are given in the Appendix. In those cases, I contributed to fieldwork, part of the analysis, Sanger Sequencing and editing of the manuscripts.

Chapter 1

Introduction

1.1 *Developmental system drift*: Different genetic pathways underpinning the same phenotypes in different species

Sometimes genetic networks that orchestrate the development of conserved phenotypes are shared among species, suggesting the existence of a "genetic toolkit". This similarity makes possible extrapolations across phylogenetically diverged species such as mice and humans. A striking example is the pathway involved in the eye development. The expression of a fly homolog of *Pax6*, which is involved in vertebrate eye development, in frog embryos can induce the formation of eye structures (Onuma et al. 2002). This highlights conservation of regulatory mechanisms underpinning eye development across vertebrates and invertebrates.

However, many conserved phenotypes are orchestrated by distinctive genetic networks – even in closely related species; this phenomenon is called *Developmental System Drift* (DSD) (True and Haag 2001). Many examples of DSD have been identified, such as bristles patterns of flies and limb development in vertebrates (reviewed by True and Haag 2001). A consequence of DSD and the evolution of genetic networks is that homologous genes in different species can have different functions. Between two closely related species of worms, for example, functional divergence has been found in 25% of conserved genes (Verster et al. 2014). Various mechanisms have been propsed to explain how DSD occurs, including alteration of expression patterns of pathway components, divergence of protein function (True and Haag 2001), deployment of alternative pathways, and addition of genes to existing pathways (Wang and Sommer 2011).

1.2 Sexual reproduction

Sexual reproduction, the fusion of gametes from two different individuals, is commonplace in plants and animals. Sexual reproduction allows for genetic recombination,

thereby increasing the diversity of allelic combinations across genes (Hickey and Golding 2018), which is considered to be the main benefit of sexual reproduction.

Different types of gametes can be produced either by the same (hermaphroditic) or different (gonochorism/diecious) individual(s). Producing separate sexes has evolved independently multiple times, which argues for benefits of having separate sexes over hermaphrodism in some circumstances. In angiosperms for example, it has been estimated that there were between 871 to 5000 independent origins of dioecy (Renner 2014). Reversion to hermaphrodism has also been observed, for example in fish (Pla et al. 2020). Having separate sexes is potentially costly because it requires finding a mate, which may expose an individual to predation or disease. A reduction in inbreeding depression has been suggested as an important benefit of separate sexes in plants (Dorken et al. 2002).

In species with separate sexes, generally we observe two sexes with two different types of gametes (anysogamy). Usually one type of gamete type (eggs) are larger, (almost) immobile and fewer number, whereas the other gamete type (sperm) are smaller, mobile, and in higher number. The large size of eggs allows resource storage, whereas increased mobility of sperm increases the probability of encountering an egg (Parker 1972). Co-evolution between nuclear and mitochondrial genomes, with especially the uniparental inheritance of mitochondria, has also been suggested to explain the evolution of separate sexes from hermaphroditic ancestors (Hadjivasiliou et al. 2012).

1.3 *Developmental system drift* in a crucial trait: Sex determination

Even though sexual differentiation is common, the genetic mechanisms underpinning sex determination are highly variable among species. In some organisms such as turtles (Bull and Vogt 1979), sex is determined by environmental cues. In others such as humans, sex is determined genetically. These two systems are not exclusive and in some cases temperature has been found to override the genetic sex determination (Radder et al. 2008). In the Australian dragons - in which sex determination has genetic and environmental components - Holleley et al. (2015) were able to obtain families of Australian dragons in which the sex relies solely on the temperature by using wild sex reversed individuals.

Within genetic sex determination, the heterogametic sex (which carries two different sex chromosomes) can be male (XY system with the male being XY and the female XX) or female (ZW system with the male being ZZ and the female ZW). Sex-determination can either be due to the expression of a sex-determining allele on the Y-chromosome (male heterogametic XY) or W-chromosome (female heterogametic ZW) or alternatively to differences in dosage of a gene on the X- or Z-chromosome compared to the Y- or W-chromosome, respectively. In the first scenario, a gene on the hemizygous chromosome has a distinctive function that is not present on genes on the homologous sex-chromosome. For example, in some fish species being a male results from the presence of *dm-y* (which is absent in females) (Matsuda et al. 2002). In the

second scenario, there may be a loss of function mutation that inactivated an allele on the W- or Y-chromosome with higher dosage on X- or Z-linked alleles in females or males respectively driving sexual differentiation. This may be the way that the Z-linked gene *dmrt1* influences sex-determination in birds (Smith et al. 2009). Additional levels of diversity in sex determination lie in the trigger itself and the chromosome carrying this trigger. For example, while the male sex determining gene *dm-y* is found in multiple closely related species of fish (Matsuda et al. 2002; Matsuda et al. 2003), it seems to have been lost in at least one species, suggesting the emergence of a new trigger (Tanaka et al. 2007). Sex chromosomes turnovers have been shown in various clades such as frogs (Furman and Evans 2016; Jeffries et al. 2018), fishes (Ross et al. 2009), and strawberries (Tennessen et al. 2018).

Although the diversity of sex determining systems is high, there is some redundancy in the recruitment of certain genetic triggers. Members of the *SOX* and *DM* families are often co-opted as important actors in the sex determining cascade. *Sox3* was identified as the trigger for sex determination some fish (Takehana et al. 2014) while *sry* - which derived from the X-linked *sox3* - is the main trigger for male determination in mammals (Foster and Graves 1994). Similarly, *dmrt1*, a DM-domain containing gene, appears to be sex-determining gene in birds (Smith et al. 2009) and is also involved in the sex-determination pathway in humans (Moniot et al. 2000). A mutation of *dmrt1* causes sex-reversal in humans (Moniot et al. 2000). *dm-y* (Matsuda et al. 2002) and *dm-w*, main trigger for sex-determination in respectively the fish medaka and the frog *X. laevis* (Yoshimoto et al. 2008), are duplicates of *dmrt1*

1.4 Sex-chromosome evolution

Following the acquisition of a trigger for sex-determination, sex-chromosomes continue to evolve and eventually may diverge in cytological appearance or remain cytologically similar. In many species, sex-chromosomes diverged due to one chromosome losing genes (i.e., becoming "degenerated") as a consequence of recombination suppression. Heteromorphic sex-chromosomes (cytogically different) initially evolved from an ancestral pair of recombining autosomes that are essentially identical (with the exception of population polymorphism). One explanation for this is that following the acquisition of a sex-determining gene, linkage between this locus and other sexually antagonistic loci (advantageous for one sex but detrimental for the other one) is favored (Fisher 1931), which leads to the extension of regions of recombination suppression (Rice 1987). The reduction of homology between the sex chromosomes can be rapid due to chromosomal rearrangements such as inversion or more gradual via accumulation of mutations and transposable elements. Sex-chromosomes may accumulate differences in gene content and chromosomal size. This is the case for the mammalian sex chromosomes for example: in humans the Y is much smaller than the X (∼60 Mb vs 165 Mb (Graves 2006)).

In contrast, some sex-chromosomes are not substantially distinct (such as those of many frogs and fish). Being a male or a female results in subtle differences in the genome sequence. For example, a small 150 kilobase region distinguishes the X and Y chromosomes of grapes and is involved in sex-determination (Picq et al. 2014). Previously, sex chromosome homomorphy was considered by many to be an early stage of sex-chromosome evolution. However, many homomorphic sex-chromosomes, such as those of ratite birds, are in fact quite old (Yazdi and Ellegren 2014). Several hypotheses have been proposed to explain how homomorphic sex-chromosomes persist over time. The "*Fountain of Youth*" hypothesis (Perrin 2009, Stöck et al. 2011) applies to species in which recombination is rare or absent in one phenotypic sex. In rare instances phenotypic sex reversal occurs, which allows periodic recombination between sex chromosomes. The maintenance of homomorphic sex-chromosomes via rare recombination events seems to occur in *Hyla* frogs (Stöck et al. 2011). Under the Fountain of Youth hypothesis, recombination is expected to be lower in the heterogametic sex. However in *Xenopus*, the opposite was found in (at least) two species (Furman and Evans 2018). The "*high-turnover*" hypothesis (Blaser et al. 2013) proposes that sex-chromosomes change over time, thereby preventing a high level of divergence. High turnover of triggers for sex-determination could occur, for example, if many deleterious mutations accumulate in a region of suppressed recombination around an ancestral trigger for sex-determination (Blaser et al. 2013). High-turnover of triggers for sex-determination has been identified in fishes (Kitano and Peichel 2012) and frogs (Dufresnes et al. 2015). Other hypotheses exist to explain the lack of degeneration of these sex-chromosomes that involves sexual antagonism. A mutation that is beneficial to one sex but detrimental to the other sex is expected to favor the cessation of recombination. However, if the involved genes instead evolve a sex-specific or sex-biased pattern of expression, the cessation of recombination between the chromosomes need not to be favored by selection (Bachtrog et al. 2014). The role of sexual antagonism in sex-chromosome homomorphy is thus potentially important as well.

1.5 Pipid frogs

Pipid frogs contain four genera: *Xenopus* (29 species), *Hymenochirus* (4), *Pseudhymenochirus* (1), and *Pipa* (7) (Frost 2016). In *Xenopus*, there is only one diploid species (*X. tropicalis*). Multiple independent polyploidization events occurred in *Xenopus* to generate tetraploid (e.g., *X. mellotropicalis*), octoploid (*e.g., X. andrei*), and dodecaploid species (*e.g., X. ruwenzoriensis*; see Evans et al. 2015 for a phylogeny of *Xenopus* genus). These polyploidization events result from sequential hybridization between different species (alloploidization) (Evans et al. 2015). It follows that autosomal loci are duplicated but not the female linked sex determining loci (review in Evans 2008). Most pipid species occur only in Africa, except *Pipa* which is found is South America (Cannatella 2015).

In pipid frogs, sex chromosomes are cytologically similar (homomorphic) (Tymowska 1991). In this group, changes in sex determination are particularly fast. Before the completion of this thesis, three different sex-determining systems were known in pipids: (1) *dm-w* (Yoshimoto et al. 2008), (2) the system in *X. borealis* (Furman and Evans 2016) and (3) a polymorphic system in *X. tropicalis* (Roco et al. 2015). The sex-determining region is located on chromosome 7 in *X. tropicalis*, chromosome 8L in *X. borealis* (Furman and Evans 2016) and chromosome 2L in *X. laevis* (Session et al. 2016). In *X. laevis*, a gene called *dm-w* is a major trigger for sex-determination. When present, the gonads become ovaries (Yoshimoto et al. 2010). This suggests that the hemizygous chromosome (the W-chromosome) has acquired a new function (different from the genes on the homologous sex-chromosome). A PCR assay found *dm-w* in various *Xenopus* species but it failed in others, suggesting an even higher diversity level of sex-determining systems (Bewick et al. 2011). As well, additional genes have been identified as W-specific in *X. laevis* (Mawaribuchi et al. 2017b). They are suspected to be involved in ovarian differentiation but their exact roles are unknown.

The genus *Xenopus* contains two species that are model organisms: *X. laevis* and *X. tropicalis*. Both are extensively used in research on human diseases (e.g. Schild et al. 1995; Boskovski et al. 2013). *X. laevis* was originally exported for pregnancy testing and is considered an invasive species that may be involved with the spread of chytrid, a fungus linked to the decline of amphibians worldwide (Weldon et al. 2004). Frogs provide advantages compared to other organisms including the production of a large number of eggs, and external development of the embryos starting at the fertilization of the eggs. Tools and resources have been developed for *Xenopus*, including whole chromosome-scale reference genomes for *X. laevis* and *X. tropicalis* (`http://www.xenbase.org/`), genome editing tools (Liu et al. 2014; Nakayama et al. 2014), and various resources centers: the National Xenopus Resource (NXR) and the European Xenopus Resource Centre (EXRC) (Pearl et al. 2012).

1.6 Aims and outlines of this thesis

1.6.1 Chapter 2: Developmental system drifts and the drivers of sex chromosomes evolution

The goals of Chapter 2 are to better understand the diversity of sex-determining systems in pipid frogs and to further investigate what factors can drive the evolution of sex chromosomes. Using genomic and genetics approaches (a combination of whole genome, reduced representation, targeted capture, and Sanger sequencing), we characterized the sex-chromosomes in three pipid species and studied the dynamics of the efficiency of *dm-w* in genus *Xenopus*. It was previously thought that if *dm-w* was present in a species, it would have the same role in female differentiation as it does in *X. laevis*. However we found that *dm-w* has actually variable effects on female differentiation in different species: in some species it was mainly found in females but not exclusively, and also sometimes

in males, and in one species it segregates as an autosomal (or pseudoautosomal) gene. Surprisingly we found that *dm-w* was female specific in only two species. Our results lead us to suggest the efficiency of genetic triggers for sex determination as a key determinant of the evolutionary paths taken by sex chromosomes. This is because in species with an inefficient system for sex determination - such as some *Xenopus* species - there need not be a genomic region that is strictly sex-specific. It follows, therefore, that the resolution of genomic conflict caused by sexual antagonism is unlikely to emerge by way of recombination suppression on the sex chromosomes that carry an inefficient trigger for sex determination.

1.6.2 Chapter 3: Evolution and function of a (small) sex determining supergene in a frog (*X. laevis*)

Chapter 3 investigates functional evolution of the supergene in *X. laevis* that carries *dm-w*. Using a combination of genomic and genome editing approaches, I examined the emergence of different exons of *dm-w* during the diversification of *Xenopus*, the linkage and function of two other female-specific genes (*scan-w*, *ccdc69-w*), and the evolution of *dmrt1*, which is an antagonistic duplicate of *dm-w*. We found that the fourth exon of *dm-w* arose in a subset of species that carry the first two coding exons, and that this corresponded with an empowerment of *dm-w* in its role as a trigger for female differentiation. Modification of the sequence of exon 4 led to the extension of the C-terminal of this gene, which is thought to have an important role in the antagonistic function of *dm-w*. The addition of *scan-w* and *ccdc69-w* in an ancestor of species in which *dm-w* is female-specific suggests a potential role in the empowerment of *dm-w*. However, our results from knock-out experiments imply that they are not crucial for female differentiation. We also demonstrate that pseudogenization occurred multiple times in *dmrt1* following duplication by allopolyploidization. Together these results provide context for the changed functionality of the small W-linked supergene of *X. laevis*.

Chapter 2

Developmental systems drift and the drivers of sex chromosome evolution

Cauret, C. M., Gansauge, M. T., Tupper, A. S., Furman, B. L., Knytl, M., Song, X. Y., ... & Evans, B. J. (2020). Developmental systems drift and the drivers of sex chromosome evolution. Molecular biology and evolution, 37(3), 799-810.

Abstract Phenotypic invariance – the outcome of purifying selection – is a hallmark of biological importance. However, invariant phenotypes might be controlled by diverged genetic systems in different species. Here we explore how an important and invariant phenotype – the development of sexually differentiated individuals – is controlled in over two dozen species in the frog family Pipidae. We uncovered evidence in different species for (i) an ancestral W chromosome that is not found in many females and is found in some males, (ii) independent losses and (iii) autosomal segregation of this W chromosome, (iv) changes in male versus female heterogamy, and (v) substantial variation among species in recombination suppression on sex chromosomes. We further provide evidence of, and evolutionary context for, the origins of at least seven distinct systems for regulating sex determination among four closely related genera. These systems are distinct in their genomic locations, evolutionary origins, and/or male versus female heterogamy. Our findings demonstrate that the developmental control of sexual differentiation changed via loss, sidelining, and empowerment of a mechanistically influential gene, and offer insights into novel factors that impinge on the diverse evolutionary fates of sex chromosomes.

Key words: Sex chromosomes, recombination suppression, developmental systems drift, sexual antagonism.

2.1 Introduction

2.1.1 Developmental systems drift and sex determination

An important discovery is that major developmental cascades are governed by conserved suites of genes, suggesting the existence of a "genetic toolkit" whose components orchestrate core developmental processes across diverse organisms (e.g., Pax6/eyeless in eye development Xu et al. 1997; distal-less in limb development Cohn and Tickle 1999). This conservation allows us to understand fundamental biological mechanisms by studying a handful of model organisms. However, conserved phenotypes may be controlled by genetic systems that differ among species. For example, there is variation among closely related species in transcription factor binding positions despite conservation of regulatory function (Villar et al. 2014). This divergence in genetic systems that underpin conserved phenotypes is called developmental systems drift (DSD) (Weiss and Fullerton 2000; True and Haag 2001). How DSD occurs remains an open question, and could involve pathway switching, convergence, and rapid evolution (Haag and Doty 2005), changed pleiotropic interactions (Pavlicev and Wagner 2012), and neutral variation could play an important role (Lynch and Hagner 2015).

Because it is fundamentally linked to reproduction and therefore fitness, it is surprising that the genetic control of two differentiated sex phenotypes (male and female) is so frequently subject to DSD. For example, differences in which sex is the heterogametic sex (male for XY systems and female for ZW systems) indicates DSD of the regulation of sexual differentiation in amphibians, fishes, and reptiles (e.g., Evans et al. 2012; Pennell et al. 2018). This is also true at finer phylogenetic scales, such as within gecko lizards (Gamble et al. 2015), stickleback fishes (Ross et al. 2009), and ranid frogs (Jeffries et al. 2018). Even in mammals, several species have independently lost the male determining gene *SRY* along with the rest of their Y chromosomes (Sutou et al. 2001; Matveevsky et al. 2017). More broadly, a diversity of environmental cues, which impinge on and are thus are coupled to variable genetic systems, can govern sexual differentiation (e.g., McCabe and Dunn 1997; Refsnider and Janzen 2015).

Diverse sex determination systems differ in the ways that genes act to trigger sexual differentiation (such as differences between the sexes in gene dosage, splice variants, gene content, and the ratio of X to autosomal chromosomes) and also in what genes actually trigger sexual differentiation (Graves 2008). Although triggers of sexual differentiation can evolve rapidly, several genes involved with the developmental process of sexual differentiation have conserved roles (Angelopoulou et al. 2012). An example of a highly conserved sex-related gene is *dmrt1*; this gene is a transcription factor that contains a DNA-binding DM-domain and operates in a downstream capacity in sexual differentiation of eutherian mammals (Matson et al. 2011), but is the sex determining locus in birds (Smith et al. 2009; Lambeth et al. 2014), and (as a paralog) in various fish and amphibians (Matsuda et al. 2002; Yoshimoto et al. 2008). Conservation of these sex-related genes suggests that DSD often involves one central system with distinctive inputs

in different species (i.e., variation on a theme), as opposed to wholesale replacement of major systems (Graves and Peichel 2010).

2.1.2 How does DSD influence sex chromosomes?

Sex chromosomes are distinguished from each other and from autosomal pairs by the presence of genetic variation that triggers (and thus is associated with) sexual differentiation. Some sex chromosome pairs are nearly identical in sequence and structure (homomorphic), whereas others evolve additional differences in gene content and repetitive elements, that causes them to be cytogenetically distinctive from each other (heteromorphic). Sex chromosome heteromorphy is caused by Hill-Robertson effects (Hill and Robertson 1966; Felsenstein 1974) that stem from recombination suppression over portions of the sex chromosomes for an extended period of evolutionary time (Charlesworth and Charlesworth 2000). In homomorphic sex chromosomes, recombination suppression is presumably restricted to a small region of the sex chromosomes and/or for a short period of time. Whether sex chromosomes evolve to be heteromorphic or stay homomorphic (the presumptive ancestral-state; Bachtrog et al. 2014) may depend on several factors. For example, *de novo* evolution or translocation of a sex determining gene could convert a pair of autosomal chromosomes into homomorphic sex chromosomes (Volff et al. 2007). The "*rapid turnover*" explanation for homomorphic sex chromosomes predicts that diverged species with homomorphic sex chromosomes should have non-homologous genomic locations of their triggers for sex determination. Rapid turnover could be favored or facilitated by temporal or geographic variation in the selective advantage of differing offspring sex ratios (West et al. 2000), neutral variation in genetic triggers for sex determination (Bull and Charnov 1977), or a sex-specific fitness advantage of a new sex determining system (Bull and Charnov 1977).

Another possibility is that sex chromosome homomorphy could exist if recombination is common enough between the sex chromosomes with sufficient frequency to prevent divergence between them (Perrin 2009), but not so common as to frequently disrupt the sex determining locus. The advantages of a high recombination rate might still be enjoyed in non-sex linked portions of the genome if the rate of recombination is sex biased (heterochiasmy or achiasmy) with a higher rate in the homogametic sex (females for XY systems, males for ZW systems). Under this mechanism, the "*fountain of youth*" explanation for homomorphic sex chromosomes, does not require variation among species in the genomic location of the trigger for sex determination, and predicts a lower rate of recombination rates in the heterogametic sex (males for XY systems, females for ZW systems).

Intra-locus sexual antagonism – when a mutation has opposite relative fitness effects on each sex – could also contribute to sex chromosome divergence in several ways. For instance, recombination suppression should be favored by natural selection when it links alleles with sexually antagonistic function to the phenotypic sex that these alleles benefit (e.g., Jordan and Charlesworth 2012). Thus, homomorphic sex chromosomes may

exist in species in which sexual antagonism is rare or resolved by mechanisms other than physical linkage to a sex-determining locus, such as by the evolution of sex-biased expression. Under this scenario, different species with homomorphic sex chromosomes may not vary in the genomic location of their trigger for sex determination, but if this is the case, sexual antagonism should either be rare, or resolved in other ways.

Conversely, it is also possible that the origin of a mutation with sexually antagonistic effects on an autosome could drive turnover of sex chromosomes (Doorn and Kirkpatrick 2010), in which case expectations would match those of the *rapid turnover* hypothesis. The "*hot potato*" model considers the distinctive responses of natural selection to mutations with sexually antagonistic effects, and to deleterious mutations that accumulate in non-recombining regions of a sex chromosome that contain the trigger for sex determination (Blaser et al. 2014). Natural selection on the former type of mutation is expected to favor the expansion of recombination suppression surrounding sex-linked regions, whereas selection on the latter type of mutation is expected to favor the establishment of new sex chromosomes with small non-recombining regions. In this way, the *hot potato* model predicts that transitions in sex chromosomes will occur continuously through evolution (Blaser et al. 2014). This cycle could be broken by "evolutionary traps" that disfavor sex chromosome transitions, such as when genes that are essential for development of one sex become sex-linked (Bachtrog et al. 2014). Divergence of sex chromosomes may also be limited if an inability to cope with sex-differences in gene dosage (i.e., hemizygosity in one sex) make the selective cost of sex chromosome degeneration too high (Adolfsson and Ellegren 2013).

2.1.3 Pipid frogs and DSD of sex determination and sex chromosomes

Species in the frog family Pipidae are a compelling group with which to explore possible links between genes that trigger sex determination and the evolutionary fate of the sex chromosomes on which they reside. Pipids have considerable variation in the genomic locations of genes that trigger sexual differentiation, and also have variation among and within species in male versus female heterogamy, and in the presence or absence of at least one trigger for sex determination (*dm-w*; Yoshimoto et al. 2008; Olmstead et al. 2010; Roco et al. 2015; Furman and Evans 2016; Bewick et al. 2011). There is also variation among pipid species in the extent of recombination suppression in genomic regions linked to different sex determining loci (Bewick et al. 2013; Furman and Evans 2018), even though morphologically diverged sex chromosomes have not been observed (Wickbom 1950; Tymowska 1991).

Using genomic approaches, we identify here new variation in the developmental systems that underpin sexual differentiation in pipid frogs by demonstrating that *dm-w* has variable effects on female differentiation in different species. We also find new variation in the degree to which recombination is suppressed in regions linked to sex determining genes. By studying sex chromosomes in this group, we elucidate three distinct ways in which DSD can happen, including loss, sidelining, or empowerment of a mechanistically

influential gene. Through interactions with genes with sexually antagonistic functions, we speculate that variation in the efficiency of genetic triggers for sex determination is a key determinant of the evolutionary paths taken by sex chromosomes.

2.2 Results

2.2.1 Loss, sidelining, and empowerment of an imperfect regulator of sexual differentiation

In the African clawed frog *X. laevis*, the gene *dm-w* exists as a single allele in females and triggers their sexual differentiation, and is a partial duplicate of the gene *dmrt1*, which is involved in the sex determining cascades of many species (Yoshimoto et al. 2008). *Dm-w* is also known from several species that are closely related to *X. laevis* (Bewick et al. 2011), and is distinguished from *dmrt1* by nucleotide divergence and several insertion/deletions (see Supplement for details). These features allowed us to use targeted next-generation sequencing of genomic DNA to identify *dm-w* in assembled reads from a panel of *Xenopus* species. Our probes targeted regions of *dm-w* that are homologous to *dmrt1* (i.e., *dm-w* exons 2 and 3) but not regions of *dm-w* that are not homologous to *dmrt1* (i.e., *dm-w* exons 1 and 4).

In these targeted next-generation sequences, we detected *dm-w* exons 2 and 3 in a paraphyletic assemblage of female individuals from 7 tetraploids, 2 octoploids, and 1 dodecaploid, and failed to detect *dm-w* in females of 1 diploid, 7 tetraploids, 3 octoploids, and 3 dodecaploids (Fig. 2.1). There was perfect concordance in the identification of independently assembled *dm-w* exons 2 and 3 (i.e., if *dm-w* was detected at all in an individual, both exons were independently detected in that individual). We did not detect missense or nonsense mutations in any of the *dm-w* sequences we recovered from exon 2 or 3.

As expected because *dm-w* originated in an ancestor of subgenus *Xenopus* after divergence of an ancestor of subgenus *Silurana* (Bewick et al. 2011), we did not identify *dm-w* in the targeted next-generation sequences from any of the species in subgenus *Silurana*. Included in our capture experiment were several male individuals of multiple species (Table A1.1). As expected, males from two species in subgenus *Xenopus* (*X. boumbaensis* and *X. lenduensis*) did not have either *dm-w* exon. Surprisingly, however, males of two other species (*X. itombwensis* and *X. pygmaeus*) carried both exons of *dm-w* that were targeted by our probes.

Our results are congruent with findings based on PCR assays by Bewick et al. (2011), and extend them by (i) ruling out a role of primer mismatch in the failure to detect *dm-w*, (ii) independently validating all detections with two exons, and (iii) surveying several species that were discovered recently, including *X. kobeli* – the only dodecaploid species with *dm-w*. Thus, *dm-w* is present in one or more tetraploid, octoploid, and dodecaploid

species. Bewick et al. (2011) detected *dm-w* exon 2 in a female *X. pygmaeus*, and here we detected exons 2 and 3 in a male individual of this species.

Several lines of evidence in addition to the perfect concordance with the PCR assays of Bewick et al. (2011) allow us to conclude that the targeted next generation sequence results were minimally influenced by sequencing depth, cross hybridization with non-target genomic regions, or rapid divergence of *dm-w* in some species. First, the sequencing depth of capture libraries of samples in which *dm-w* was detected was not substantially different from (and on average lower than) that of samples that where *dm-w* was not detected (3.4–5.6 million or 3.4–23.6 million sequence reads were captured from libraries with and without *dm-w*, respectively, Table A1.1). Second, we were able to detect both expected *dmrt1* ohnologs in all tetraploid species for both *dmrt1* exons, including allelic variation for several loci, and an additional paralog of *dmrt1* in *X. borealis* that arose after allotetraploidization and that was confirmed independently using Sanger sequencing (data not shown). This indicates that our approach efficiently captured sequences in each genome that were similar to our probes. We also identified most expected paralogs in all octoploid and dodecaploid species, although in most of these species, less than 4 or 6 paralogs (respectively) are expected due to gene loss (Bewick et al. 2011). Third, we compared the assembled captured sequences to other DM-domain containing genes from *X. tropicalis* (*dmrt1*, *dmrt2*, *dmrt3*, and *dmrt5*) and found them to be highly diverged in exonic and flanking intronic regions from the *dmrt1* and *dm-w* sequences that we assembled from the captured sequences (data not shown). Fourth, for essentially all *dm-w* and *dmrt1* sequences, we also captured and assembled flanking intronic sequences that were clearly homologous to the flanking intronic sequences from the genome sequences of *X. laevis* and *X. tropicalis* (Hellsten et al. 2010; Session et al. 2016), even though our capture probes did not target these intronic sequences. This indicates that, as expected, these intronic sequences are contiguous with exonic sequences of *dmrt1* and *dm-w* that matched our probes. And fifth, previous efforts provide unambiguous evidence that at least one species – *X. borealis* – does not use *dm-w* to trigger sex determination: whole genome sequences from one female and one male individual do not contain this gene, and the newly identified sex chromosomes of *X. borealis* (chromosome 8L) are not the same as the chromosome on which *dm-w* resides in *X. laevis* (chromosome 2L) (Furman and Evans 2016; Furman and Evans 2018). That at least one *Xenopus* species appears to have lost *dm-w* and evolved new sex chromosomes increases our prior expectations that this could happen in other species as well.

The region of *dm-w* that we sequenced using targeted next generation sequencing is relatively rapidly evolving. For example, uncorrected pairwise divergence between *dm-w* exon 2 from *X. laevis* and *X. clivii* is 8% at the nucleotide level (16 out of 209 ungapped bp) and 13.4% at the amino acid level (9 out of 67 ungapped amino acids). Uncorrected pairwise divergence of *dm-w* exon 3 from these two species is 11.0% at the nucleotide level (16 out of 145 ungapped bp) and 12.5% at the amino acid level (6 out of 48 ungapped amino acids). Thus, the failure to detect *dm-w* in some species (apart from *X. borealis*) using targeted next-generation sequencing could conceivably be false negatives due to rapid evolution of *dm-w*. However, if this were the case in some other species

(e.g., *X. longipes*), it would have to affect the entire DM domain (which is encompassed by our data from exons 2 and 3). To the extent that well supported relationships among mitochondrial genomes are congruent with evolutionary relationships among the primarily maternally inherited *dm-w*, false negatives due to rapid evolution of *dm-w* also would have to have occurred independently several times (i.e., separately in *X. parafraseri*, *X. allofraseri*, *X. wittei*, and more) and in some cases over rather short periods of time (e.g., *X. ruwenzoriensis*; Fig. 2.1). As a conservative measure that accommodates this possibility of false negatives, the itemization of new sex determination systems in the Discussion does not consider the sex-determining systems of species with undetected *dm-w* (apart from *X. borealis*) as necessarily distinctive (although they may be).

These results prompted us to turn to museum specimens of wild caught animals, and laboratory animals from six species, to test, using PCR and Sanger sequencing, whether *dm-w* was restricted to animals that were phenotypically female. We found that *dm-w* was female-specific in two species, including *X. laevis*, but not in four others, including *X. clivii*, which is the most divergent species from *X. laevis* that carries *dm-w* (Table 2.1, Fig. 2.1). For three species (*X. clivii* (Fig. A1.7), *X. pygmaeus* (Fig. A1.6), *X. victorianus*), *dm-w* was most commonly observed in phenotypic females, but it was also occasionally found in phenotypic males. For one of these species (*X. victorianus*), *dm-w* was found in all females (n = 20), but also found in one third of the males (5 out of 15 tested). For two of these three species (*X. clivii*, *X. pygmaeus*), *dm-w* was not detected in roughly one third of the females we surveyed (5 out of 16 female *X. clivii* and 3 out of 9 female *X. pygmaeus* did not have detectable *dm-w*) and *dm-w* was detected in several males (Table 2.1). In one species (*X. itombwensis*), *dm-w* was observed in all individuals of both phenotypic sexes (present in 24 individuals including 19 males, Fig. A1.3). Furthermore, in *X. itombwensis dm-w*, heterozygous positions were observed in a few individuals of both phenotypic sexes (Fig. A1.4). Heterozygous positions were also observed in *dm-w* carried by five male and one female *X. victorianus* (Table A1.4, Fig. A1.5).

A caveat to a lack of positive PCR amplification of *dm-w* in females of some species is that there could be polymorphism at primer sites that prevented amplification in some individuals. However, in females lacking *dm-w*, the negative amplifications were independently replicated with two separate amplicons that targeted different and non-overlapping portions of *dm-w* using different pairs of primers. We only considered samples for which a positive control (*i.e.* the amplification of a portion of the mitochondrial 16S ribosomal RNA gene) were successful. This caveat not withstanding, we observed positive independent amplifications of multiple independent regions of *dm-w* in several males from multiple species (Table 2.1). This provides unambiguous evidence that *dm-w* is not completely linked to female differentiation in some species, even though it is usually found in females in these same species. Ancestral-state reconstruction of these observations suggests that *dm-w* was not fixed only in females in the most recent common ancestor of species that carry this gene (Fig. 2.1). Further details about these methods and results are discussed in the Supplement.

Table 2.1: *Dm-w* is and is not female specific (Y or N respectively) in several *Xenopus* species. Targeted next generation sequencing, PCR, and Sanger sequencing were used to assess how many phenotypic females and males (# females and # males respectively) carry *dm-w* in parentheses for each phenotypic sex.

Species	# females (# with *dm-w*)	# males (# with *dm-w*)	Female-specificity of *dm-w*	Notes
X. itombwensis	5 (5)	20 (20)	N	[1]
X. pygmaeus	9 (6)	11 (2)	N	
X. clivii	16 (11)	29 (5)	N	[2]
X. victorianus	20 (20)	15 (5)	N	[3]
X. laevis	24 (24)	12 (0)	Y	
X. gilli	13 (13)	7 (0)	Y	

[1] *dm-w* is fixed or almost fixed; heterozygotes observed in both sexes suggesting autosomal segregation
[2] includes 24 wild samples analyzed in (Furman and Evans 2016)
[3] heterozygotes observed in 1 female and 5 males from eastern DRC

2.2.2 Additional variation in sex-linkage, recombination suppression, and heterogamy in other pipid species

Using reduced representation genome sequencing (RRGS) and shotgun whole genome sequencing (WGS) from captive-bred families, we extended our analysis to the sex chromosomes of three other pipid species whose ancestors never carried *dm-w*. This effort also uncovered variation in sex chromosomes, and definitively demonstrates that males are heterogametic in *Hymenochirus boettgeri* and *Pipa parva*, and suggests that females are heterogametic in *X. mellotropicalis*. The sex chromosomes of *H. boettgeri* and *P. parva* correspond to *X. tropicalis* chromosomes 4 and 6, respectively, but the sex chromosomes of *X. mellotropicalis* corresponds to *X. tropicalis* chromosome 7. Additionally, based on the genomic locations where scaffolds containing sex-linked SNPs in the RRGS data mapped to the *X. tropicalis* genome assembly, the extent of recombination suppression on the sex chromosomes varied substantially among these species, with a large region in *H. boettgeri* (∼79.8 Mb, i.e. ∼60% of the assembled chromosome, containing 11 sex-linked SNPs out of 29 SNPs total in the sex-linked region), a medium-sized region in *P. parva* (at least ∼8.6 Mb, containing 3 sex-linked SNPs out of 5 SNPs total in the sex-linked region), and a small region in *X. mellotropicalis* (only 1 scaffold <1 Mb contained one sex-linked SNP out of 5 SNPs total on this scaffold) (Fig. 2.2). Because these sex-linked SNPs in each family are based on RRGS data, they represent a small fraction of the total number of sex-linked SNPs in the genomes of each species. Embedded within these inferred regions of recombination suppression, we did observe several SNPs in the RRGS genotypes that did not have a completely sex-linked pattern of inheritance. We suspect that these are genotype errors stemming from low coverage because they are flanked by completely sex-linked SNPs that were confirmed by Sanger sequencing decribed below.

Our inferences concerning the extent of recombination suppression in each focal

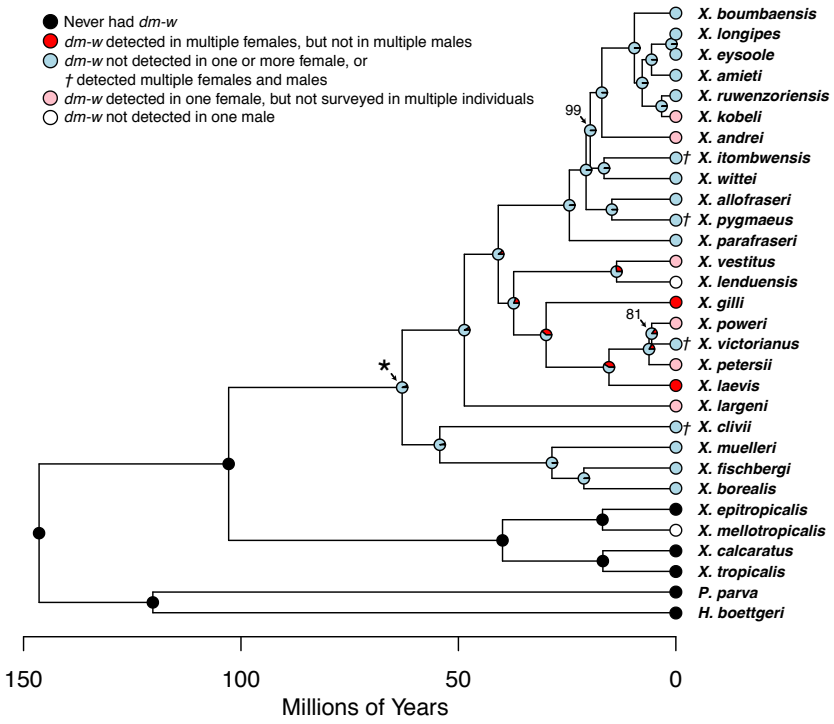


FIGURE 2.1: The sex determining gene *dm-w* was detected in at least one individual (red, pink, light blue with dagger) using targeted next generation sequencing, but was not detected in at least one individual in other species (light blue, black, white). In some species, *dm-w* was detected, but was not female-specific based on PCR assays (daggers). Data are plotted on a Bayesian consensus phylogeny estimated from complete mitochondrial genome sequences, as described in the Supplement. All nodes have 100% posterior probability except where labeled. The most recent common ancestor of all species in which *dm-w* was detected, indicated with an asterisk, has a 94% likelihood of having this gene not fixed only in females and a 5% likelihood of it being fixed only in females. The scale bar indicates the time in millions of years ago.

species are based on the locations of genomic regions in *X. tropicalis* that are homologous to dozens of sex-linked scaffolds on our three focal species (Supplementary Results, Fig. 2.2). For each focal species, we performed simulations to evaluate the probability of observing the number of sex-linked SNPs by chance. For *H. boettgeri* and *P. parva*, but not *X. mellotropicalis*, the simulations indicate that the number of identified sex-linked SNPs are highly unlikely to occur by chance (Supplemental A1.1).

Thus, even though the genome assemblies for each of our focal species were highly fragmented and the family sizes of each was relatively small, the generally high synteny among frog genomes (Sun et al. 2015; Session et al. 2016) and close phylogenetic affinities of pipid genomes allowed us to demarcate large regions of sex-linked recombination suppression using RRGS and WGS data in two pipid species. For *X. mellotropicalis*, our data indicate with high confidence that the sex-specific region of the sex chromosomes of this species is small; a consequence of this finding is that the support for female heterogamy is relatively weak in terms of the number of sex-linked SNPs in the RRGS data (one completely sex-linked SNP on an unplaced scaffold and one almost sex-linked SNP on chromosome 7; Supplementary Information, Table A1.2, Fig. A1.1).

To further validate these inferences based on RRGS data mapped to WGS assemblies, we Sanger sequenced genes from our laboratory families that we inferred to be within the sex-linked regions of each species based the RRGS and WGS data. These amplifications targeted regions that included sex-linked variation that was spanned by the RRGS data, and also regions that were not covered by the RRGS data but that were nonetheless putatively sex-linked in each focal species under the assumption of synteny with the *X. tropicalis* genome. The Sanger sequences verified complete sex-linkage for *H. boettgeri* (4 independent amplifications of portions of 3 genes) and *P. parva* (3 independent amplifications of portions of 3 genes). Sanger sequencing verification was unsuccessful for a sex-linked scaffold identified by the RRGS data for *X. mellotropicalis* due to repetitive sequences in the very small sex-linked region that we identified. However, a paralog of one gene (*or8h1*) was identified based on its genomic position in the *X. tropicalis* genome that had an almost female-specific sex-specific amplification. This suggests that *or8h1* is in close genomic proximity to the trigger for sex determination in *X. mellotropicalis*, but that recombination does occasionally occur between these two genes (Fig. A1.2; Supplementary Results). This locus and another sex-linked SNP from *X. mellotropicalis* map to the short arm of chromosome 7, which is also where the master sex determining gene of *X. tropicalis* is thought to reside (Fig. 2.1; Olmstead et al. 2010; Roco et al. 2015). The homologous genomic locations of the trigger for sex determination in the closely related species *X. tropicalis* and *X. mellotropicalis* provides further validation for our inference of the identity of the sex chromosomes of *X. mellotropicalis*, even though this region is very small in both species.

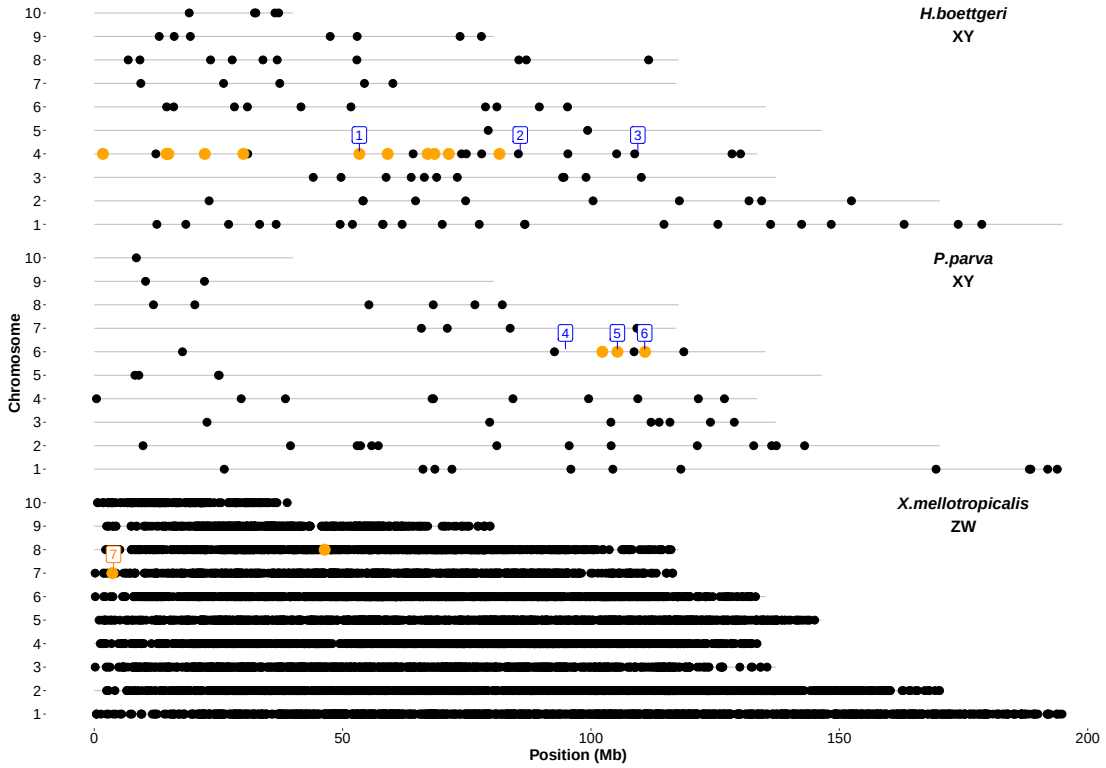


Figure 2.2: The sex chromosomes of *H. boettgeri* (Chr.04), *P. parva* (Chr.06) and *X. mellotropicalis* (Chr.07) are not homologous. For each species, orange or black dots represent the homologous genomic locations of sex-linked or not sex-linked SNPs, respectively, on the 10 chromosomes of *X. tropicalis*. The density of SNPs is highest for *X. mellotropicalis* because it is most closely related to *X. tropicalis* and because scaffolds from both subgenomes of this allotetraploid species map to only one region of the diploid *X. tropicalis* reference genome. Numbers refer to genomic regions that were validated by Sanger sequencing, and are in blue, or orange font based on whether the Sanger sequences had completely, or partially sex-linked SNPs, respectively, for the following loci: (1: *sall1*, 2: *dmrt5*, 3: *hmcn1*, 4: *kctd1*, 5: *ncoa2*, 6: *mmp16*, 7: *or8h1*). Putative false negatives for sex-linkage for *H. boettgeri* and *P. parva* (some black dots on chromosomes 4 and 6 respectively), and a false positive for sex-linkage for *X. mellotropicalis* (an orange dot on chromosome 8) are discussed in the Supplement.

2.3 Discussion

2.3.1 Developmental systems drift of sex determination and recombination suppression

Developmental processes are orchestrated by networks of genetic interactions. When compensatory changes occur in network components (i.e., genes) (Johnson and Porter 2007; Lynch and Hagner 2015), these networks can diverge in different species with minimal phenotypic consequence; this phenomenon is called developmental systems drift (DSD) (True and Haag 2001). Several lines of evidence – some newly reported here – demonstrate that DSD of sex determination occurred in pipid frogs. A gene called *dm-w* that reliably triggers female differentiation in *X. laevis* is not present in some species. In other species, this gene is genetically associated with female differentiation with imperfect efficiency (i.e., usually – but not always – because it is sometimes present in males). And in some species we failed to detect *dm-w* in some females, suggesting that this gene is not required for female differentiation in these species. Variation among pipid species in male versus female heterogamy also evidences DSD because it indicates variation in the dosages of sex-specific factors that trigger sexual differentiation.

Our findings concerning the presence, absence, and sex-specificity of a sex determination gene (*dm-w*) identify several mechanisms by which DSD can occur. In some species this gene is found in the majority of phenotypic females and also in a minority of phenotypic males. Interpretation of these results in a phylogenetic context (Fig. 2.1) argues that the ancestral capacity of *dm-w* to trigger female differentiation was strong but incomplete. By extension, this also argues that when *dm-w* initially began to contribute to the genetic control of sex determination, this gene was usually present as a single allele in females, because most of their fathers would have lacked *dm-w*. *Dm-w* is now strictly coupled to female differentiation in two closely related species (*X. laevis* and *X. gilli*), which indicates DSD occurred via an empowerment of the sex determination capacity of *dm-w* with respect to its ancestral condition.

Although *dm-w* originated in an ancestor of some extant *Xenopus*, in most (4 of 6) of the descendant species of this ancestor that we surveyed, this gene was not female-specific. In two of these species (*X. itombwensis* and *X. victorianus*), two alleles of *dm-w* were detected in both sexes (heterozygous sites are present in both sexes). Furthermore, the genomic location of *dm-w* in *X. itombwensis* may be on an autosome or pseudoautosomal region (because *dm-w* was present in all individuals of both sexes; Table 2.1, Table A1.4). That *X. itombwensis* is derived from an ancestor in which *dm-w* was usually segregating as a single allele in females, suggests that DSD also occurred via mechanistic sidelining of *dm-w* (i.e., this gene is still present and possibly functional, but not as in a sex-specific or sex-biased capacity as a trigger for sex determination). We did not find evidence in the coding regions of exons 2 and 3 that this sidelining rendered *dm-w* nonfunctional in *X. itombwensis*. Rather, its role in sexual differentiation may instead operate in a downstream capacity relative to some other, as yet unidentified, trigger

for sexual differentiation. In several species, we did not observe *dm-w* in the targeted next-generation sequencing of one female. Whether and how many times *dm-w* was lost depends on how many of these species also do not carry this allele in other individuals. However, several species in which one female lacks *dm-w* are closely related; this is consistent with the possibility that this gene was lost in their collective most recent common ancestor (Fig. 2.1) – a scenario which is consistent with DSD by gene deletion. It is possible some of these observations are a consequence of intraspecific regulatory variation in *dm-w* (e.g., *dm-w* alleles with high or low expression may tend to occur in females or males, respectively), and/or intraspecific variation in molecular variation in other portions of *dm-w* not surveyed here (e.g. coding regions of exon 4) or other genes (e.g., *scan-w* in *X. laevis* (Mawaribuchi et al. 2017b)).

Overall, this information combined with other reports cited below evidence at least *seven* distinct systems that trigger sex determination in pipid frogs, with each being distinguished by genomic location (expressed below in terms of orthology to chromosomes of *X. tropicalis*), independent origin, and heterogamy. These systems include:

- The male heterogametic system on chromosome 4 of *H. boettgeri* (this study)
- The male heterogametic system on chromosome 6 of *P. parva* (this study)
- The putative female heterogametic system on chromosome 7 in subgenus *Silurana* of *X. mellotropicalis* (this study), some populations of *X. tropicalis* (Roco et al. 2015), and possibly close relatives
- The male heterogametic system on chromosome 7 of *X. tropicalis* (Roco et al. 2015), which the principle of maximum parsimony suggests evolved from a female heterogametic ancestor (this study)
- The female heterogametic system on chromosome 2 in subgenus *Xenopus* of species where *dm-w* is female-specific (e.g., *X. laevis*; Yoshimoto et al. 2008; Session et al. 2016) or usually in females (e.g., *X. clivii*; this study)
- The female heterogametic system on chromosome 8 of *X. borealis*, and possibly close relatives whose ancestor lost *dm-w* (Furman and Evans 2016)
- The non-*dm-w* based systems on unknown chromosome(s) of other species in subgenus *Xenopus* that might have lost *dm-w* completely (e.g., *X. longipes*), or where it segregates autosomally (*X. itombwensis*) (this study; Table 2.1)

In species that do not vary in the heterogametic sex but do differ in the genomic location of the trigger for sex determination (e.g., *P. parva* and *H. boettgeri*), DSD may have occurred if there exists a novel trigger for sex determination, or alternatively translocation of a trigger may have occurred without DSD. We did not attempt to collect information about whether translocations account for the variation in the genomic locations of triggers for sex determination because, apart from *dm-w*, these triggers have not been identified.

Within pipid frogs, a roughly bimodal pattern is evident in which recombination suppression on the sex chromosomes is either restricted to a very small region (less than a few Mb) or alternatively spans a large region (greater than ∼10 Mb). Previous work on the allotetraploid species *X. laevis* and *X. borealis*, for example, illustrates that recombination suppression affects only a small region linked to *dm-w* on chromosome 2L in the former species, but a large region spanning ∼ 50 Mb that carries an unidentified sex determining factor on chromosome 8L in the latter (Furman and Evans 2018). Similar to *X. laevis*, *X. tropicalis* also has a very small region of suppressed recombination (Mitros et al. 2019; Bewick et al. 2013). Here we demonstrate that the pattern in the *X. mellotropicalis* resembles that of *X. laevis* and *X. tropicalis*, but the patterns in *H. boettgeri* and *P. parva* resemble *X. borealis*. In *P. parva*, we estimated the size of the region of suppressed recombination to minimally span multiple Mb (Fig. 2.2); there was a dearth of high confidence genotypes on one side, so the size of this region may be even larger. The sex chromosomes of pipid frogs thus demonstrate that homomorphic sex chromosomes might have large regions of suppressed recombination; a similar situation is found in ratite birds (Yazdi and Ellegren 2014; Vicoso et al. 2013).

One speculation that could account for these species-level differences is that there were (or are) more loci with sexually antagonistic effects on the chromosomes with more recombination suppression (chr 4, 6, 8) than those with less (chr 2, 7) (Rice 1987). However, another plausible explanation for these differences in the extent of recombination suppression instead relates to the efficacy of the trigger for sex determination in each species; we discuss this possibility next.

2.3.2 Inefficient sex determination as a mechanism for non-divergence of sex chromosomes

These findings of DSD of triggers for sex determination have the potential to affect genome evolution in many ways, including the evolution and recombination patterns of entire sex chromosomes. For example, there are several ways that the function and age of the trigger for sex determination could impinge on the genetic fate of entire sex chromosomes – including whether they are heteromorphic or homomorphic and whether a large or small region of recombination suppression is present. The *fountain of youth* hypothesis, for example, predicts that non-diverged sex chromosomes could be associated with phenotypic sex reversal if the heterogametic sex has low rates of recombination (Perrin 2009). However, in *X. laevis* and *X. borealis*, recombination occurs more frequently in oogenesis than spermatogenesis, including on the pseudoautosomal regions of the sex chromosomes, even though females are heterogametic in both of these species (Furman and Evans 2018). In *Xenopus*, recombination on the ends of chromosomes is lower in females than males (Furman and Evans 2018). Thus, because *dm-w* resides on the tip of chromosome 2L, recombination is less likely to disrupt this gene when it is found in females.

The *rapid turnover* hypothesis posits that new sex determination loci arise in different locations before recombination suppression results in sex chromosome heteromorphy (Volff et al. 2007). Our findings of at least seven distinct sex determination systems in pipid frogs are generally consistent with the expectations of the *rapid turnover* hypothesis, but suggest other mechanisms are at play as well. For example, the sex chromosomes of *X. clivii* and *X. laevis* – species whose ancestors diverged millions of years ago (Fig. 2.1; Evans et al. 2015) – are homomorphic (Tymowska 1991), even though they both have *dm-w* and presumably orthologous sex chromosomes. Another example is the sex chromosomes of *X. borealis*, which have a recently derived sex determining system, but widespread recombination suppression (Furman and Evans 2018), indicating that recent turnovers are not necessarily associated with small regions of recombination suppression.

Another possibility is that the efficiency of the trigger for sex determination could influence the evolutionary fate of entire sex chromosomes. The potential to resolve genomic conflict associated with sexual antagonism via linkage to an inefficient trigger for sex determination (i.e., that is not sex-specific) is presumably modest compared to linkage to an efficient trigger (that is sex-specific). If resolution of sexual antagonism favors recombination suppression in sex-linked regions, it follows that a small region of suppressed recombination would be expected near an inefficient trigger for sex determination, even in the absence of *rapid turnover*. Indeed, our findings suggest that the ancestral sex-determining locus of *X. laevis* used to be inefficient, and the region of recombination suppression in this species is known to be small (Furman and Evans 2018). Likewise, the sex determining gene in *X. mellotropicalis* is probably homologous to that of some strains of *X. tropicalis*, and the most recent common ancestor of these two species might have had a polymorphic and/or inefficient trigger of sexual differentiation as *X. tropicalis* does now (Roco et al. 2015; Mitros et al. 2019). This inefficient/polymorphic sex determining locus could explain why both of these species have small non-recombining regions on their sex chromosomes (Fig. 2.2; Bewick et al. 2013). Conversely, species with large regions of suppressed recombination, such as *H. boettgeri*, *P. parva*, and *X. borealis* (Furman and Evans 2018) might have triggers of sex determination that operate with high fidelity. In these species, expansion of recombination suppression in genomic regions that flank efficient sex determining loci may have been favored by natural selection because it resolved sexual conflict associated with mutations with sexually antagonistic fitness effects.

Other examples of inefficient triggers for sex determination have been reported. In the frog *Rana temporaria* – a species with homomorphic sex chromosomes – *dmrt1* is a candidate trigger for sex determination, but shows incomplete linkage to the sex phenotype (Rodrigues et al. 2017). This also could arise if this trigger were inefficient in some contexts (or populations), and/or if there is a polygenic basis of sex determination. Although the mechanism of sex determination differs from frogs (Hediger et al. 2004), a similar genomic situation is present in the housefly *Musca domestica*, where an inefficient sex determining locus on a Y chromosome is linked to male differentiation, with the extent of linkage depending on input from multiple sex-determining alleles on other chromosome pairs (Tomita and Wada 1989; Denholm et al. 1986). These examples of

inefficient triggers for sex determination, including *dm-w* in some *Xenopus*, allow for a novel explanation for why sex chromosomes may stay homomorphic; this explanation does not require or preclude rapid turnover of sex chromosomes, or sex differences in the rate of recombination.

2.3.3 What factors govern the efficiency of *dm-w*?

In *X. laevis*, *dm-w* is thought to drive female differentiation via competitive inhibition of the male-related gene *dmrt1* (Yoshimoto et al. 2008; Yoshimoto et al. 2010) and expression of one *dmrt1* paralog (*dmrt1-S*) is higher in male than female tadpole gonads (Mawaribuchi et al. 2017a). Related to this, the mechanism by which allopolyploidization is thought to occur in *Xenopus* does not involve duplication of the W-chromosome (or *dm-w*), but does involve duplication of autosomal genes such as *dmrt1* (Kobel and Du Pasquier 1986). Thus, a single dose of *dm-w* may compete with variable dosages of *dmrt1* in different individuals or species, depending on ploidy level and pseudogenization. Interestingly, pseudogenization of *dmrt1* homeologs occurred independently several times in *Xenopus* polyploids (Bewick et al. 2011), and it is conceivable that population-level variation in pseudogenization if *dmrt1* and/or other loci affects the efficacy of *dm-w* to determine sex.

2.3.4 Outlook

The dynamic mechanistic capacity of *dm-w* reported here demonstrates that DSD can arise from altered function or interactions of a core component of a developmental system. This could stem from changed pleiotropic interactions (Pavlicev and Wagner 2012), changed stoichiometric relationships as a result of gene duplication and pseudogenization, subfunctionalization (Force et al. 1999), and neofunctionalization (Lynch et al. 2001). Related to this, empirical and theoretical observations predict that some chromosomes or genomic regions may be more likely than others to be recruited as triggers for sex determination (Blaser et al. 2014; O'Meally et al. 2012; Brelsford et al. 2013; Furman and Evans 2016). The possibility that certain genomic regions are repeatedly co-opted as sex chromosomes is consistent with the mechanism observed here that DSD involves functional or interaction changes of components of conserved genetic pathways, as opposed to convergence in different species of pathways with distinctive genetic functions (Haag and Doty 2005). Indeed, the sex chromosomes we identified here harbour several sex-related genes (Table A1.7).

In various ranid and bufonid frogs, orthologs of *X. tropicalis* chromosomes 1, 2, 3, 5, and 8 are sex chromosomes (Jeffries et al. 2018; Miura 2007; Brelsford et al. 2013; Uno et al. 2015). The diversity of sex chromosomes in pipids – including chromosomes homologous to five of the ten *X. tropicalis* chromosomes (2, 4, 6, 7, or 8) – expands this list, leaving homologs of only *X. tropicalis* chromosomes 9 and 10 as not (yet) being identified as sex chromosomes in a frog species. However, several chromosome

pairs have been repeatedly co-opted as anuran sex chromosomes (e.g., homologs of *X. tropicalis* chromosome 5 in ranids (Jeffries et al. 2018), chromosome 1 in bufonids and ranids (Brelsford et al. 2013), chromosome 8 in ranids and pipids (Furman and Evans 2016; Miura 2007; Uno et al. 2015), and chromosome 7 in pipids and rhacophorids (Uno et al. 2015)). The newly identified sex chromosome 6 of *P. parva* is also homologous to the sex chromosomes of some snakes (Matsubara et al. 2006). Clearly, more information on the identities of genes that trigger sexual differentiation in species that lack *dm-w* would allow us to evaluate the degree to which DSD generally involves modulation of core developmental genes, or alternatively is largely sculpted by inputs from mechanistically divergent systems (e.g., Zhang et al. 2016; Cline et al. 2010).

Overall, this study recovers evidence for frequent DSD of sex determination in pipid frogs, identifies new sex chromosomes and novel variation in recombination suppression on sex chromosomes, and pinpoints three of mechanisms of DSD including loss, sidelining, and empowerment of a mechanistically influential gene. We speculate that the sex chromosomes of species with inefficient sex determination genes are less advantageous destinations for alleles with sexually antagonistic function. This points to a previously unsuspected mechanism for sex chromosome non-divergence (homomorphy) that will be fascinating to further explore in other groups with known triggers for sex determination.

2.4 Methods

2.4.1 Targeted next-generation sequencing and Sanger sequencing of *dm-w* in *Xenopus*

In order to survey *Xenopus* females for the presence of *dm-w*, we used oligonucleotide baits to enrich Illumina libraries for sequences similar to *dmrt1*, which is the gene from which exons 2 and 3 of *dm-w* originated (Yoshimoto et al. 2008). We captured sequences from libraries that were generated from one female from all described species in subgenus *Xenopus*, except *X. fraseri*, because we lacked a sample of high-quality genomic DNA from this species, and *X. lenduensis* and *X. mellotropicalis* where we generated libraries from male individuals. As a negative control, we captured sequences from libraries of one individual from each species in subgenus *Silurana* (this lineage diverged prior to the origin of *dm-w*). In order to evaluate female-specificity of *dm-w*, we used PCR and Sanger sequencing *dm-w* in several individuals from six species (Table 2.1) whose phenotypic sex was determined from external morphology, or surgically after euthanasia. Additional details of samples, PCR primers, library preparation, capture, sequencing, assembly, and analysis are provided in the Supplement. Capture sequences of *dm-w* and *dmrt1* are available on Genbank (accession MN030659-MN030916, Table A1.5) and an alignment of these data is available in the Supplementary Data (File S2 for exon 2, File S3 for exon 3).

2.4.2 The sex chromosomes of other pipids

The sex chromosomes of three pipid species that diverged before the origin of *dm-w* were also examined, including the dwarf clawed frog, *Hymenochirus boettgeri*, the Sabana Surinam toad, *Pipa parva*, and the Gabonese clawed frog, *Xenopus mellotropicalis*. The first two of these species are diploid (*H. boettgeri*: 2n = 24 and *P. parva*: 2n = 30; Cannatella and De Sa 1993) but the third (*X. mellotropicalis*) is allotetraploid (2n = 4x = 40; Evans et al. 2015). Half of the genome of *X. mellotropicalis* (i.e., one subgenome) is derived from a recent ancestor of the diploid species *X. tropicalis*, and its other subgenome is derived from another diploid species that is either extinct or undiscovered (Evans et al. 2015).

We generated one family from each of these three pipid species, performed WGS on one or both parents. Assembled WGS reads were used to anchor reduced representation genome sequencing (RRGS) and Sanger sequencing from both parents and their offspring to identify sex-linked genomic regions. The sex of postmetamorphic offspring was ascertained surgically after euthanasia. The *H. boettgeri* family included both parents, 11 daughters, and 21 sons. The *P. parva* family included both parents, five daughters, and seven sons. The *X. mellotropicalis* family included both parents, nine daughters, and nine sons. Additional details about these families and analysis are provided in the Supplement.

2.5 Acknowledgments

We are especially grateful to Svante Pääbo and Max Planck Institute for Evolutionary Anthropology for hosting BJE on a sabbatical in Leipzig, Germany. We thank Darcy Kelley for providing us with *X. mellotropicalis* and *X. clivii* individuals, Ben Vernot for assistance with capture probe design, Birgit Nickel for performing library capture, Brian Golding and the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca) and Compute/Calcul Canada for access to computational resources, and Brian Golding, Jonathan Dushoff, and Ian Dworkin for many enlightening discussions and comments on a draft of this manuscript. This work was supported by the Natural Science and Engineering Research Council of Canada (RGPIN/283102-2012 and RGPIN-2017-05770 to BJE), the Museum of Comparative Zoology (BJE), the Max Planck Institute for Evolutionary Anthropology (M-TG, SP, MM, BJE), a National Geographic Research and Exploration Grant (no. 8556-08 to EG), and the US National Science Foundation (DEB-1145459 to EG).

Chapter 3

Evolution and Function of a (Small) Sex Determining Supergene in a Frog (*X. laevis*)

This chapter aims at being submitted as a publication.

Authors: Cauret, C. M., Jordan, D. C., Gansauge, M. T., Lopardo, V., Greenbaum, E., Gvoždík, V., Horb, M. E. & Evans, B. J.

Abstract Supergenes are sets of tightly linked genes that influence a phenotype. To better understand how supergenes evolve and function, we examined a small supergene that drives female differentiation in the frog *X. laevis*. We found that function of this supergene as a trigger of female sexual differentiation evolved recently, depends on a locus called *dm-w*, and was coupled in time with the addition and extension of the coding region of the last exon of this gene and also with the addition of two other nonessential supergene components (*scan-w* and *ccdc69-w*). Further work is needed to determine the effects of these latter two genes on fertility. We also identified examples of species-specific and population-specific loss of supergene function. In addition, we investigated molecular evolution of an autosomal competitor of *dm-w* called *dmrt1*, and identified extensive pseudogenization after this locus was duplicated by multiple episodes of allopolyploidization. This dynamic genomic backdrop, incremental functionization, and nonfunctionalization of the *X. laevis* sex-determining supergene raises a 'chicken and egg' question of whether and how these variable genetic components influenced or were influenced by each other. Also unclear is what the functions of ancestral components were prior to assembly and the achievement of full functionality of this supergene.

3.1 Introduction

3.1.1 Where do new genes and new function come from?

Complex biological processes such as development, behavior, and metabolism are controlled by genetic networks of genes that interact with each other and with the environment. Genetic networks evolve via changes in pathway components (gain or loss of genes) and by altered interactions among these components. New genes arise from a variety of mechanisms, including horizontal gene transfer (Husnik and McCutcheon 2018), duplication (Tocchini-Valentini et al. 2005), exon shuffling (Wang et al. 2000), replication or modification by transposable elements (Velanis et al. 2020), gene fusion (Thomson et al. 2000) or fission (Wang et al. 2004), and *de novo* origin from previously non-coding genomic regions (Cai et al. 2008). Duplication may occur at a small (e.g., microsatellite expansion, gene duplication) or large scale (e.g., trisomy; whole genome duplication). Function of gene duplicates may diverge in several ways, such as subfunctionalization (partitioning of ancestral function among duplicates), neofunctionalization (the evolution of novel function), and pseudogenization (loss of function) (Lynch 2007). Sometimes duplicates evolve to have antagonistic roles wherein one copy represses the function of the other. For example, amphioxus have two steroid hormone receptors that are related by gene duplication; one is activated by estrogen while the other lost this ancestral function and instead acts as a repressor of the first (Bridgham et al. 2008). Antagonistic function also exists in paralogs involved in cortical development (Dennis et al. 2012) and sex determination (Yoshimoto et al. 2010). In addition, new genes may rapidly evolve essential functions. In fruit flies, for example, 30% of newly evolved genes appear to be essential since their inactivation is lethal (Chen et al. 2010). Evolution may also convergently exploit non-homologous or homologous genes to produce similar traits in different species. For example, multiple independent mutations of the same gene have been linked to the reduction of pigmentation in various species of fish living in caves (Espinasa et al. 2018), and non-homologous convergence is evidenced by lens optical proteins called crystallins that evolved several times from different ancestral proteins (Piatigorsky and Wistow 1989).

3.1.2 Supergenes and recombination suppression

Proteins with functional associations are sometimes encoded by genes that are genetically linked in the genome (Ogata et al. 2000; Thévenin et al. 2014) or in the same physical space in the nucleus (Thévenin et al. 2014), which may promote their co-regulation. Supergenes – groups of tightly linked genes that regulate a trait whose phenotype depends on genetic variation within the supergene – are an extreme case. Supergenes are thought to orchestrate several traits including aspects of behaviour (Wang et al. 2013), mimicry in butterflies (Joron et al. 2006), colour of fishes (Küpper et al. 2016), and heterostyly in plants (Labonne et al. 2010). In the latter case, the involvement of the different members of the supergene in the phenotype was confirmed by knockout (Labonne et al.

2010), whereas in butterflies, the role of the supergene components is inferred based on co-expression of supergene components in the developing wing discs, but has not yet been confirmed by genetic manipulation (Saenko et al. 2019).

Genetic associations between alleles of different loci can be generated by several processes such as heterogeneity of environmental conditions (if certain combinations of alleles are beneficial in some places but not others) or negative epistasis (wherein certain combinations of alleles are deleterious; Otto and Lenormand 2002). The formation of supergenes could be favoured by natural selection for reduced recombination in order to maintain advantageous combinations of alleles across multiple genes (Rice 1984; Rice 1987; Van Doorn and Kirkpatrick 2007; Fisher 1931; Charlesworth and Charlesworth 1980). Sex-linked portions of sex chromosomes can be thought of as supergenes because recombination suppression causes co-inheritance of genes linked to the sex determining locus (Charlesworth 2016). In some cases, supergenes carry loci with diverse functions. For example, the male-specific portion of the human Y-chromosome contains a gene that triggers sex determination (*SRY*) and also several genes that are necessary for male fertility long after primary sex determination has been achieved (Skaletsky et al. 2003). One factor that may favor recombination suppression in sex-linked supergenes is the resolution of genomic conflict associated with mutations with sexually antagonistic fitness effects, wherein a mutation benefits one sex but harms the other. Under this scenario (and depending on recessivity and in which sex a mutation is advantageous), it could be beneficial for the mutation to reside on the non-recombining portion of a sex-specific sex chromosome (a Y- or W-chromosome) as compared to an autosome (Rice 1987). Other factors that may lead to the evolution of recombination suppression include heterozygote advantage, meiotic drive, and genetic drift (Ponnikas et al. 2018). Of note, while recombination suppression may preserve adaptive combinations of alleles at different genes, it is also expected to reduce the efficacy of natural selection on variation in the non-recombining region, and eventually lead to more rapid accumulation of deleterious mutations as compared to genomic regions that recombine. Thus, one of the ironies of sexual reproduction is that it facilitates genetic recombination and associated benefits in the autosomes, but sexual reproduction is also frequently associated with recombination suppression on the sex chromosomes, which has fitness costs. In theory, increased mutational load of non-recombining regions can lead to sex-chromosome turnover, wherein new sex chromosomes arise from ancestral autosomes (Blaser et al. 2013; Blaser et al. 2014).

The effects of recombination suppression on sex chromosome turnover, coupled with other factors that contribute to turnover, such as sexual antagonism (Van Doorn and Kirkpatrick 2007), cause the systems for sex determination and their associated sex chromosomes (if present) in some groups to vary considerably (e.g. Bachtrog et al. 2014). Consequently, systems for sex-determination and the sex-chromosomes on which portions of these systems reside offer a compelling model with which to provide insights into how important traits evolve.

In principle, it is possible that different components of a supergene could have an

additive phenotypic effect; if this were the case, each component would be necessary but not sufficient to produce one of the phenotypes associated with a supergene (for example, for a sex determining supergene, one of the sex phenotypes – male or female). Another possibility is that supergene components individually have only a modest phenotypic effect, but collectively act synergistically. If this were the case, individual effects of each component would not be additive. Yet another possibility is that supergenes include components that are both necessary and sufficient for a particular phenotype, whereas the function of other components is peripheral to the main phenotype. These peripheral loci could be captured by expansion of the region of recombination suppression (e.g., by an inversion), or they could arrive by translocation after recombination suppression was achieved.

3.1.3 A (small) frog supergene: The female-specific region of the W-chromosome of *X. laevis*

To explore how supergenes arise and affect important traits, we focused on a simple supergene that is associated with a crucial phenotype: the female-determining region of the W-chromosome of the African clawed frog, *X. laevis*. There are only three female-specific genes on this W-chromosome (Mawaribuchi et al. 2017b): *dm-w*, *scan-w*, and *ccdc69-w*, and one of these – *dm-w*– is thought to be the main trigger for primary (gonadal) sexual differentiation of female *X. laevis* (Yoshimoto et al. 2008; Yoshimoto et al. 2010). In three of nine or three of seven transgenic (ZZ) males (depending on the construct used), insertion of *dm-w* by restriction enzyme-mediated integration resulted in the development of ovotestes, which contain both ovarian and testicular structures (Yoshimoto et al. 2008). In three of 11 transgenic (ZW) female tadpoles and 10 of 38 transgenic female adults, inactivation of *dm-w* by RNA interference produced abnormal gonads (Yoshimoto et al. 2008; Yoshimoto et al. 2010). The variable effects of *dm-w* transgenes and inactivation could indicate that dosages of other W-linked genes or Z-linked loci also influence sexual differentiation, or alternatively this could have methodological basis (e.g., positional effects of the *dm-w* transgene; incomplete inactivation of *dm-w* by RNA interference). *Dm-w* is expressed in the developing gonad during gonadal differentiation, and also in adult ovary and liver (Mawaribuchi et al. 2017b).

Little is known about the other two female-specific genes on the W-chromosome (*scan-w*, *ccdc69-w*) apart from their genomic location and expression domains. Similar to *dm-w*, both are expressed in female tadpole gonads during gonadal differentiation, suggesting a possible role in early ovarian differentiation, and also in adult ovary (Mawaribuchi et al. 2017b). As well, *scan-w* is expressed in adult brain and stomach, and *ccdc69-w* is expressed in adult brain (Mawaribuchi et al. 2017b). These genes originated by gene duplication of the autosomal loci (*scan.L* or *scan-like* and *ccdc69.L*, respectively Mawaribuchi et al. 2017b).

One gene – *capn5-z* – is known to be specific to the Z-chromosome of *X. laevis* (i.e., not present on the W-chromosome; Mawaribuchi et al. 2017b). Thus females, which

have one W and one Z-chromosome, have one *capn5-z* allele whereas males, which have two Z-chromosomes, have two. This gene is expressed in both sexes in the developing gonads, and also in adult gonads, brain, and spleen, and to a lesser extent in several other tissues (heart, liver, stomach, mesonephros; Mawaribuchi et al. 2017b).

Dm-w is a chimerical gene with its first coding exons (exons 2 and 3; exon 1 is non-coding) originating from partial duplication of *dmrt1*, which is an important transcription factor with male-related function in many metazoans (Zarkower 2001). The origin of the last exon of *dm-w* (exon 4) is unknown because it has no homology to *dmrt1* or any other known gene (Yoshimoto et al. 2008; Yoshimoto et al. 2010; Bewick et al. 2011; Mawaribuchi et al. 2017a). *Dm-w* is also present in other *Xenopus* species, but is not always female-specific (Bewick et al. 2011; Cauret et al. 2020). In some species, *dm-w* is also found in some males, and in at least one species (*X. itombwensis*), *dm-w* is present in all individuals and appears have autosomal segregation (Cauret et al. 2020). The variation in female-specificity indicates that the biological function of *dm-w* has been dynamic over time, and only evolved female-specific recently, and in only a subset of *Xenopus* species that carry this gene (Cauret et al. 2020).

Dmrt1 has a conserved role in male sexual differentiation in many species, including humans (Moniot et al. 2000; Masuyama et al. 2012; Yoshimoto et al. 2010). In birds, *dmrt1* is Z-linked and has two copies in males and one in females; differences in Z-dosage of this locus appear to determine whether an individual develops into a male or female (Smith et al. 2009). *Dmrt1* is a transcription factor with a DNA binding domain called a DM-domain that is crucial for its function (Zarkower 2001). It is this portion of *dmrt1* that was copied to create *dm-w* exons 2 and 3, which therefore also has a DM-domain. *Dm-w* is thought to be an antagonist of *dmrt1*, with *dm-w* potentially binding to similar regulatory regions that are bound by *dmrt1*, thereby preventing their activation by *dmrt1* (Yoshimoto et al. 2010). In the allotetraploid species *X. laevis*, two homeologous copies of *dmrt1* are present (*dmrt1.L*, *dmrt1.S*), and *dmrt1.L* has two splice variants (type 1 and 2; Mawaribuchi et al. 2017a). *Dmrt1.L* type 2 is expressed in the gonads of both sexes at multiple developmental stages but *dmrt1.L* type 1 and *dmrt1.S* appear to be primarily expressed in males (Mawaribuchi et al. 2017a).

3.1.4 Goals

The goal of this study is to understand the evolution and function of the W-linked supergene of *X. laevis*. Using targeted capture sequencing and PCR assays across the genus *Xenopus*, we investigated the evolutionary history and function of the three genes in this supergene (*dm-w*, *scan-w*, and *ccdc69-w*) and also its autosomal partner gene *dmrt1*. Then, using genome editing, we explored the phenotypic consequences of inactivation of *dm-w*, *scan-w*, and *ccdc69-w* by inactivating the coding region of these genes.

3.2 Results

3.2.1 Female-specificity of *dm-w* is coupled with the addition of exon 4 and extension of its coding region

Current understanding of female-specificity of *dm-w* is based on (i) capture data from exons 2 and 3 from one female from 28 of the 29 species of *Xenopus*, combined with (ii) surveys of multiple individuals of six species (Cauret et al. 2020). Additional information is therefore needed to fully evaluate *dm-w* efficiency as a sex determining gene – particularly those species that are closely related to *X. laevis* where *dm-w* appears to be completely linked to and required for female differentiation (Cauret et al. 2020). To address this gap in knowledge, we first collected capture data from the only other exon of *dm-w* with a coding region (exon 4) in the same panel of *Xenopus* species as Cauret et al. (2020) and then tested the female specificity with a PCR assay in five additional species beyond those considered by Cauret et al. (2020).

Targeted next-generation sequencing of one individual (usually a female) from all *Xenopus* species except *X. fraseri* detected *dm-w* exon 4 in the same species for which we previously identified exons 2 and 3 (Cauret et al. 2020) with the exception of *X. clivii*, a tetraploid, and *X. vestitus*, an octoploid, in which exon 4 was not detected (Fig. 3.1). Although other scenarios are possible, the phylogenetic position of *X. vestitus* relative to other species that carry exon 4 is consistent with this exon being lost after it became linked to *dm-w* exon 2 and 3 in the most recent common ancestor (MRCA) of *X. largeni* and *X. vestitus*, after divergence of this ancestor from *X. clivii*. Because *dm-w* exons 2 and 3 appear to have arisen in the MRCA of *X. clivii* and other *dm-w* containing species such as *X. largeni* and *X. vestitus* (Cauret et al. 2020), these data suggest that *dm-w* exon 4 arose after *dm-w* exons 2 and 3 and in an ancestor of a subset of species that carry *dm-w* exons 2 and 3. Another less parsimonious possibility that we cannot exclude is that exon 4 arose around the same time as exons 2 and 3, but was lost in *X. clivii*.

Nucleotide sequences of exon 4 indicates that an ancestral stop codon was present in the MRCA of all species that contain this exon that was lost in the MRCA of *X. laevis* and *X. gilli*. Loss of this ancestral stop codon had the effect of extending the coding region by 28 amino acids. The ancestral stop codon is still present in several species (*X. largeni*, *X. pygmaeus*, *X. kobeli*, *X. andrei*). The extension of the coding region of exon 4 was potentially coupled in time with the addition of *scan-w* and *ccdc69-w* to the ancestral W-linked supergene of *X. laevis* (see below), and with the onset of sex-specificity of *dm-w* (see below; Cauret et al. 2020).

In *X. petersii*, a −30 base pair (bp) deletion is present in the beginning of *dm-w* exon 4, although the stop codon is homologous to that in other closely related species (*X. laevis*, *X. victorianus*, *X. petersii*, *X. gilli*). This deletion in *X. petersii* removes 10 amino acids in the translated protein. In *X. poweri*, a 16 bp frameshift deletion near the end of the exon introduces an early stop codon that causes a loss of four amino acids at

the C-terminus of translated protein relative to other closely related species such as *X. laevis*.

Using PCR surveys for *dm-w* exons in four species for which the female specificity of *dm-w* was not tested using multiple individuals by Cauret et al. (2020), we found *dm-w* to be female-specific in *X. poweri* and *X. andrei* but not *X. petersii*, *X. kobeli*, or *X. largeni* (in *X. largeni* our small sample indicates *dm-w* is not present in all females but it is unclear whether it is sometimes present in males; Table 3.1). Within individual *X. kobeli* specimens, independent attempts to amplify *dm-w* exons 2, 3, and 4 were either all successful or all unsuccessful. This is consistent with these three exons being genetically linked and co-inherited in this species, which is also the case in *X. laevis* (Yoshimoto et al. 2008). Based on these results and also the consistent detection of all three exons in one female individual from several other species (Fig 3.1), we suspect these exons, when present, are genetically linked in other *Xenopus* species as well.

In *X. victorianus*, the geographic variation (Furman et al. 2015) seems to correspond with the efficiency of *dm-w*, in that *dm-w* is not female-specific in a population on the Lendu Plateau of the Democratic Republic of the Congo (DRC), but appears to be so elsewhere. The female used in the targeted sequencing was from Uganda; capture data were not obtained from *X. victorianus* from the Lendu Plateau.

In *X. clivii*, *dm-w* exon 2 and 3 are observed in both sexes (Cauret et al. 2020) but it is not clear whether this influences fertility. To test this, we attempted to cross female and male *X. clivii* that did not and did carry *dm-w* exons 2 and 3 with a partner frogs. In both crosses, we obtained viable offspring, which indicates that *dm-w* is not required for female fertility, and also does not disrupt male fertility.

Table 3.1: Variation in female-specificity of *dm-w* in *Xenopus* species that were not assayed in Cauret et al. 2020. Targeted next generation sequencing, PCR, and Sanger sequencing were used to assess how many phenotypic females and males (# females and # males respectively) carry *dm-w* in parentheses. *X. victorianus* is separated into two populations: P1 corresponding to the Lwiro and South Kivu region of the Democratic Republic of the Congo (DRC) and P2 to the Lendu region and Orientale Province, DRC. In *X. andrei* a question mark indicates uncertainty of the female specificity due to a small sample size.

Species	# females (# with *dm-w*)	# males (# with *dm-w*)	Female-specificity of *dm-w*	Notes
X. andrei	3 (3)	2 (0)	Y?	
X. kobeli	7 (3)	9 (2)	N	
X. largeni	4 (2)	1 (0)	N	
X. petersii	2 (2)	3 (3)	N	[1]
X. poweri	2 (2)	8 (0)	Y	
X. victorianus (P1)	13 (13)	8 (0)	Y	
X. victorianus (P2)	2 (2)	5 (5)	N	[1]

[1] heterozygous positions were observed in both sexes, which suggests autosomal segregation of *dm-w*

3.2.2 *Scan-w* and *ccdc69-w* arose before the diversification of the MRCA of *X. laevis* and *X. gilli*

Using the same approach detailed above and in Cauret et al. (2020), we performed targeted next generation sequencing of *scan-w* and *ccdc69-w* in all known *Xenopus* species except *X. fraseri*. *Scan-w* has six exons but exons 2-6 are small and contain repetitive sequences; we therefore focused our attention on exon 1, which has a protein coding region and also the largest non-repetitive region. *Ccdc69-w* has two exons and we analyzed both.

We were able to detect the first exon of *scan-w* in five species (*X. gilli*, *X. laevis*, *X. petersii*, *X. poweri* and *X. victorianus*; Fig. 3.1). For *ccdc69-w*, we recovered exon 1 and 2 only for species closely related to *X. laevis*, which are the same five species for which we identified *scan-w* exon 1 (Fig. 3.1). We also identified portions of *ccdc69-w* exon 1 from *X. largeni* and exon 2 from *X. andrei* but not other *ccdc69-w* exons from each of these species. Overall, these results point to an origin of *scan-w* and juxtaposition of both exons of *ccdc69-w* before the diversification of the MRCA of *X. gilli* from *X. laevis* (green diamond in Fig. 3.1). In the targeted next generation sequences, we were able to distinguish autosomal sequences from *scan.L* and *scan-like* although high similarity among these sequences prevented us from conclusively determining which of these autosomal loci was more closely related to *scan-w*.

3.2.3 *Dm-w* but not *scan-w* or *ccdc69-w* is required for female differentiation of *X. laevis*.

To further characterize their functional roles, we created knock out lines for *dm-w*, *scan-w* and *ccdc69-w* in *X. laevis* using CRISPR/Cas9. We targeted the beginning of the coding region of each gene in order to completely or maximally disrupt protein function by deleting the start codon and/or introducing frameshift mutations that caused early stop codons. For all three genes, sequence chromotograms of F0 individuals have overlapping sequences that begin at the targeted region; this indicates a mosaic genotype comprising wild-type and mutant sequences as expected because cutting occurs at a multicell stage of embryogenesis. These F0 females were then crossed with wild-type (J-strain) males to generate non-mosaic F1 knockout individuals (Fig. 3.2) which were confirmed by Sanger sequencing.

For *dm-w*, F1 individuals had a -10 bp deletion which created a frameshift mutation in the predicted coding region and an early stop codon (Fig. 3.2A). For *scan-w*, F1 individuals had a -20 bp deletion in the predicted coding region, which also created an early stop codon (Fig. 3.2B). For *ccdc69-w*, F1 each carried one of several different types of deletions: -22, -12, -6 bp in the predicted coding region. The -12 bp deletion in the coding region also has a large deletion upstream of the coding region of ∼201 bp, so overall this deletion is 213 bp long and included the start codon. The -22 bp also removed the predicted start codon and created an early stop codon (Fig. 3.2C).

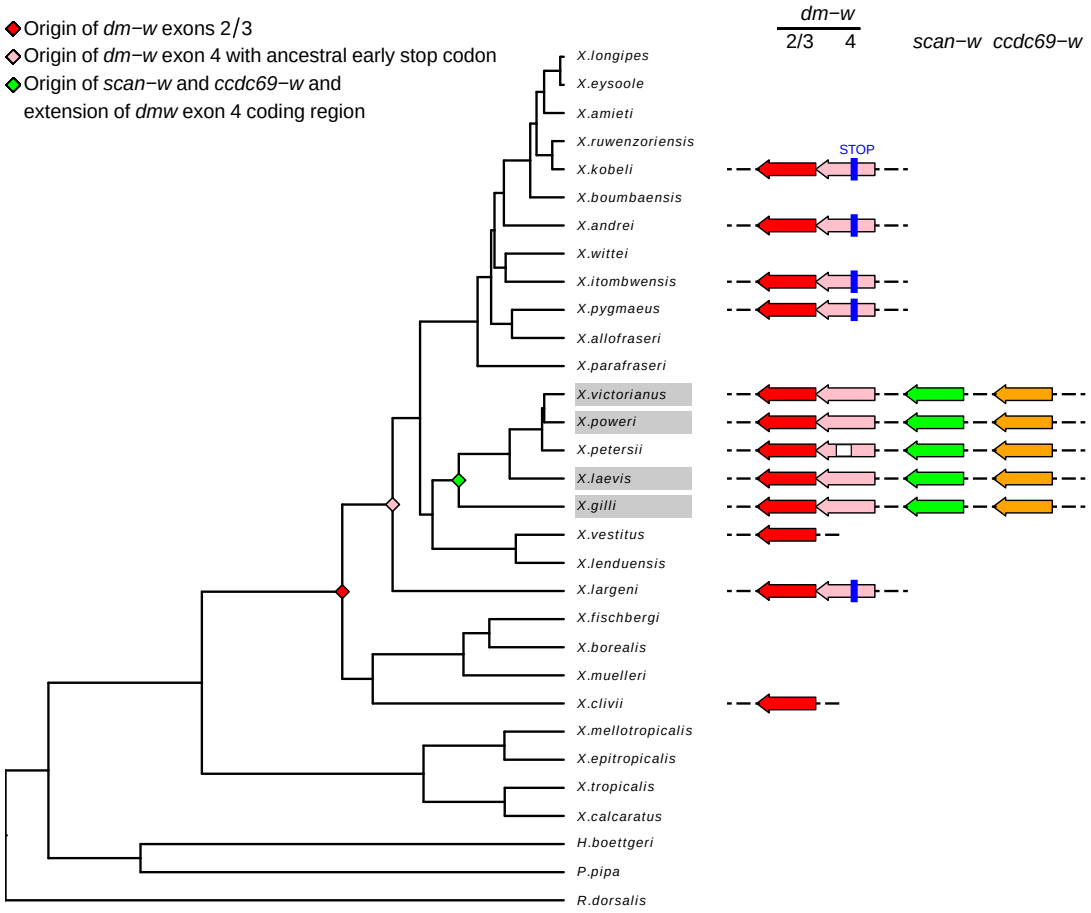


FIGURE 3.1: Evolutionary steps toward female-specificity of *dm-w* and linked loci on the W-chromosome of *X. laevis*. Grey rectangles highlight species names for which *dm-w* is female-specific. Arrows indicate W-linked orthologs of genes in *X. laevis*. For *X. victorianus*, the data represents the *dm-w* efficient clade and not the population from the Lendu Plateau (See Results). A white rectangle in *dm-w* exon 4 in *X. petersii* represents a deletion in the coding region. Diamonds represent major events: red, *dm-w* exons 2 and 3 are generated by partial duplication of *dmrt1*; pink, establishment of *dm-w* exon 4 in linkage following exons 2 and 3; green, origin of *scan-w* and *ccdc69-w*, respectively, by duplication of the autosomal loci *scan.L* or *scan-like* and *ccdc69.L*, and extension of the coding region of *dm-w* by mutation of the ancestral stop codon in exon 4. Data are plotted on the phylogeny from Cauret et al. 2020 which was estimated from complete mitochondrial genomes (Evans et al. 2019) and does not depict reticulating relationships among species that stem from allopolyploidation (Evans et al. 2015). The figure was produced using gggenes v.0.4.0 and ggtree v.2.1.2 in R version 3.4.4.

The -6bp is an in-frame mutation and is not expected to substantially disrupt protein function.

For the *dm-w* deletion, a F0 mutant developed into a phenotypic male. We were able to cross this *dm-w* mutant with a wild-type female and generate viable offspring, including a WW* individual that carried a wild-type W-chromosome and also a mutant W-chromosome (W*) that carried the mutant *dm-w* (Fig. 3.2A). We also obtained ZW* offspring carrying a Z-chromosome and a mutant W-chromosome. At this stage the sex of the F1 mutants are unknown. We are currently breeding more animals in order to determine whether all individuals with complete inactivation of *dm-w* develop into phenotypic males. However, based on a limited sample size, these observations demonstrate that *dm-w* is required for female differentiation, that two alleles of the Z-linked gene *capn5-Z* are not required for male differentiation or fertility, and that one allele of the W-linked genes *scan-w* and *ccdc69-w* does not prevent male development or fertility.

Both of our *scan-w* mutant individuals (2 F1 individuals with -20 bp) and all of our *ccdc69-w* mutant individuals (2 F1 with -22bp, 7 F1 individuals with -12bp, 2 F1 with -6bp) were female. These observations demonstrate that neither *scan-w* nor *ccdc69-w* is required for female differentiation. However, we were unable to obtain eggs from sexually mature (11 month old) F1 *scan-w* mutant, suggesting a possible impact on oogenesis. We are further investigating this possibility by attempting to generate a larger sample size of mutant individuals of each line.

3.2.4 Biased pseudogenization of *dmrt1* homeologs

As discussed above, *dm-w* is related to *dmrt1* by partial gene duplication and is thought to govern female differentiation via inhibition of transcription of male related genes that are otherwise activated in males by *dmrt1* (Yoshimoto et al. 2008; Yoshimoto et al. 2010). Although almost all *Xenopus* species are polyploid, *dm-w* (and other female-specific loci) are not duplicated in *Xenopus* owing to a particular mechanism by which allopolyploidization occurs (reviewed in Evans 2008). However, autosomal genes such as *dmrt1* are duplicated by allopolyploidization to generate up to two, four, or six homologous copies of this gene in tetraploid, octoploid, and dodecaploid *Xenopus* species, with each homolog having two alleles (Evans 2008; Bewick et al. 2011). This raises the question of whether and how variable dosages of *dmrt1*, which typically is associated with male phenotypes (Zarkower 2001), affect sex determination in species with differing ploidy levels. To begin to address this question and as a complement to our efforts to understand evolution and function of the *X. laevis* supergene, we broadened our previous efforts with *dmrt1* (Bewick et al. 2011) by using targeted capture sequencing of all six exons of this locus in the same panel of individuals that we discuss above for *dm-w*, *scan-w*, and *ccdc69-w* and in Cauret et al. (2020).

In the capture data, the maximum number of *dmrt1* homeologs that we detected in each species was generally consistent with expectations based on ploidy level. For

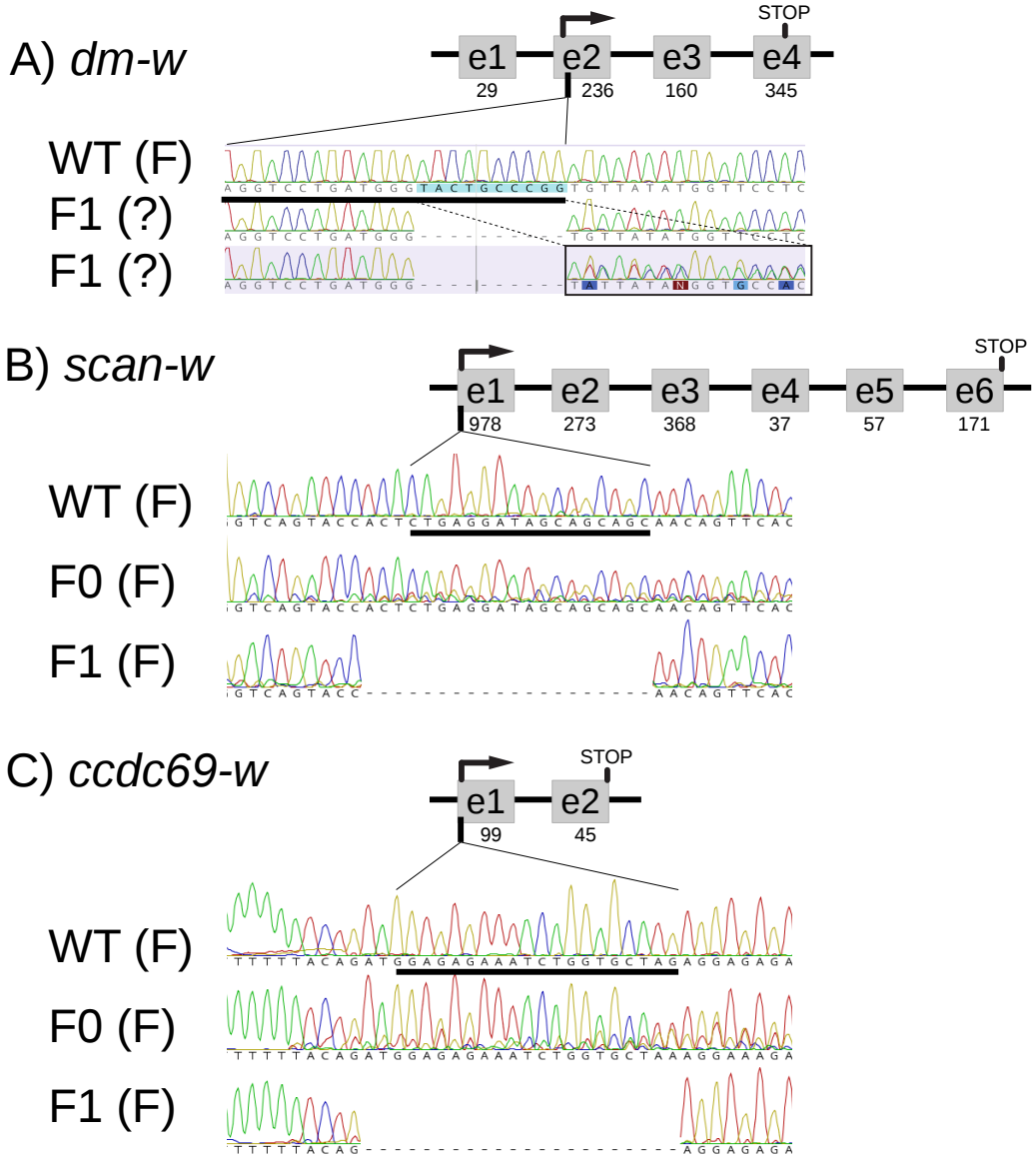


Figure 3.2: Inactivation of W-specific genes. (A) mutated region of exon 2 of *dm-w* which contains the start codon and the DM domain. Male F0 *dm-w* knock out was fertile and when crossed with a wild type J strain female produced individuals carrying one (ZW*) or two W chromosomes (WW*). The rectangle highlights overlapping sequences of two *dm-w* alleles, one with a deletion (W*) and one without (W). (B, C) mutated regions of the first coding exons of *ccdc69-w* and *scan-w*, respectively. F1 embryos were generated by mating F0 mosaic females with wild-type J strain males. Acronyms indicate wild-type (WT), and phenotypic female (F) or male (M). A question mark indicates the phenotypic sex is unknown. The thick line under each wild-type sequence represents the region targeted by the CRISPR guide RNA. Numbers under the schematic representation of exons/introns structure correspond to the size of the exon in base pairs.

example for the allotetraploid *X. laevis*, we identified two homeologs for each exon, and for the octoploid *X. lenduensis* we detected four (Table 3.2). We categorize these homeologs as either alpha or beta, depending on whether they are in the L- or S-subgenome respectively (based on their phylogenetic affinities). For *X. laevis* and *X. tropicalis*, sequences were confirmed by aligning the sequences obtained to various reference sequences (NCBI accession numbers: *X. laevis*: XM_018250680.1 (*dmrt1.L*) and NM_001085483.1 (*dmrt1.S*), and *X. tropicalis*: XM_031890717.1). For some species we were also able to identify sequences that were suggestive of allelic variation. In 11 instances, we detected what seemed to be additional duplicates that were species-specific.

Gene loss and pseudogenization after polyploidization was previously inferred based on a PCR survey of *dmrt1* exons 2 and 3 (Bewick et al. 2011). Results for the first two exons newly obtained here using targeted capture sequencing were generally consistent with the results from Bewick et al. (2011). Additional details about differences between Bewick et al. (2011) and this study are provided in the Supplement, along with potential caveats associated with our methodology.

We identified *dmrt1* pseudogenes in several species (Table 3.2) based on the presence of early stop codons and frameshift mutations relative to reference sequences from *X. tropicalis* and *X. laevis*. Most (7 out of 11) of the early stop codon were identified in exon 2 and almost all were unique to individual species. The only example of a shared early stop codon we identified was in *X. longipes* and *X. eysoole*, where an identical two bp deletion generates an early stop codon in exon 2 of one beta homeolog.

We also identified 6 large (> 20bp) deletions that could potentially impact protein function, and two of these were in exon 2. One of these, a -78 bp deletion in exon 2 was shared by *X. allofraseri*, *X. parafraseri*, *X. wittei*, *X. itombwensis*, and *X. kobeli*. Two of those deletions created a frameshift mutation that did not cause an early stop codon.

Overall, out of a total of 13 mutations that cause either a stop codon or frameshift mutation, 8 of them were in exon 2, even though this exon encodes only ∼22% of the coding region of *dmrt1*. Because these early stop codons, deletions and frameshift mutations interrupt and remove the early portion of the coding region that specifies the DM-domain, transcription and translation of these loci would presumably yield proteins with limited or no capacity to regulate genes with regulatory elements that are recognized by fully functional DM-domains. Also of interest and consistent with the results of Bewick et al. 2011, we found that more pseudogenes were beta (S subgenome) homeologs (n = 7) than alpha (L-subgenome) homeologs (n = 4). This is interesting because expression of the *dmrt1-S* homeolog appears to be more male-biased in gonads during and after sexual differentiation, including in adults, than either splice variant of *dmrt1-L* (Mawaribuchi et al. 2017a).

Table 3.2: *Dmrt1* homeologs identified in the capture data and categorized based on whether they are from the L (alpha) or S (beta) subgenome or, for *Silurana* from the subgenome more closely related to *X. tropicalis* (alpha) or not (beta). The number of homeologs identified for each of the first five coding exons of *dmrt1* is separated by a "/"; numbers in parenthesis correspond to the number of pseudogenes (carrying an early stop codon or frameshift mutation). The maximum number of homeologs expected for a diploid, tetraploid, octoploid, and dodecaploid species are respectively one, two, four, and six, respectively. * indicates that the homeolog category (alpha or beta) is uncertain.

Diploid	ID	Sex	alpha	beta
X. tropicalis	AMNH17274	F	1 / 1 / 1 / 1	
Tetraploid - *Silurana*				
X. calcaratus	VG09-364 or NMP6V 74630/1	F	1 / 1 / 1 / 1	1 / 0 / 1 / 1
X. epitropicalis	AMNH17278	F	1 / 1 / 1 / 1	1 / 0 / 1 / 1
X. mellotropicalis	AMNH17288	M	1 / 1 / 1 / 1	1 / 0 / 1 / 1
Tetraploid - *Xenopus*				
X. allofraseri	BJE3488	F	1 / 1 / 1 / 1	1 / 1 / 1 / 1
X. borealis	AMNH17315	F	1 / 2 / 2 / 2	2 / 1 / 1 / 2
X. clivii	AMNH17254	F	2 (1) / 2 / 2* / 2*	1 / 1 / 1 / 1
X. fischbergi	AMNH17297	F	1 / 1 / 3 (1) / 2	1 / 1 / 1 / 1
X. gilli	Xr(0-3)	F	1 / 1 / 1 / 1	1 / 1 / 1 / 1
X. laevis	kml(5)	F	1 / 1 / 1 / 1	1 / 1 / 1 / 1
X. largeni	AMNH17292	F	1 / 2 / 2 / 2	1 / 2 / 1 / 1
X. muelleri	Xen203	F	1 / 1 / 2 / 2	1 / 1 / 1 / 1
X. parafraseri	AMNH17282	F	1 / 1 / 1 / 1	2 / 1 / 1 / 1
X. petersii	AMNH17325	F	1 / 1 / 1 / 1	1 / 1 / 1 / 1
X. poweri	AMNH17260	F	1 / 1 / 1 / 1	1 / 1 / 1 / 1
X. pygmaeus	AMNH17320	M	1 / 1 / 1 / 1	1 / 1 / 1 / 1
X. victorianus	xen234	F	1 / 1 / 1 / 1	1 / 1 / 1 / 1
Octoploid				
X. amieti	AMNH17268	F	2 / 2 / 2 / 2	2 (1) / 1 / 1 / 1
X. andrei	AMNH17258	F	2 / 2 / 2 / 2	3 (1) / 2 / 2 / 2
X. boumbaensis	AMNH17284	F	2 / 1 / 4 (1) / 2	2 / 2 / 2 / 1
X. boumbaensis	AMNH17283	M	2 / 2 / 2 (1) / 1	3 / 1 / 3 (2) / 2
X. itombwensis	BJE00275	M	1 / 2 / 2 / 2	2 / 2 / 2 / 2
X. lenduensis	BJE2973	F	2 / 2 / 2 / 2	2 / 2 / 2 / 2
X. vestitus	RT2	F	1 / 2 / 2 / 2	2 (1) / 2 / 2 / 2
X. wittei	AMNH17307	F	2 / 2 / 2 / 2	3 (1) / 2 / 2 / 2
Dodecaploid				
X. eysoole	BJE3218	F	3 / 2 / 3 / 2	3 / 2 / 2 / 2
X. kobeli	BJE3075	F	3 / 2 / 3 / 2	3 / 2 / 2 / 2
X. longipes	BJE3190	F	2 / 2 / 3 / 3	3 (1) / 2 / 2 / 3
X. ruwenzoriensis	AMNH17317	F	2 / 2 / 2 / 3	4 (2) / 2 / 2 / 1

3.3 Discussion

Biological processes are orchestrated by interactions between biotic (e.g., protein) and abiotic (e.g., temperature) components that vary over time, through development, and within and among species. To explore how genetic components of important biological processes are assembled, we examined the sex determination system in a group of frogs (*Xenopus*), including three W-linked genes (*dm-w*, *scan-w*, *ccdc69-w*) which generally occur as a single copy that can be female-specific in some species, and an interacting partner gene in the autosomes (*dmrt1*), which was duplicated by allotetraploidization, allooctoploidization, and allododecaploidization.

3.3.1 Assembly of the (small) *X. laevis* W supergene

Our survey of *dm-w* exons 2 and 3 (Cauret et al. 2020) and exon 4 (this study) found that some species lost *dm-w* altogether after it arose, and that all species that carry exons 2 and 3 also carry exon 4, except *X. clivii* and *X. vestitus* (Fig. 3.1, Cauret et al. 2020). No species was identified that carries exon 4 but not exons 2 and 3. One interpretation of these data is that *dm-w* exons 2 and 3 appeared in the most recent common ancestor (MRCA) of all species that carry these two exons (i.e., the MRCA of *X. clivii* and *X. laevis*) and that *dm-w* exon 4 appeared in the MRCA of all species that carry this exon (i.e., the MRCA of *X. largeni* and *X. laevis*). It is also possible that *dm-w* exons 2, 3, and 4 arose around the same time in the same ancestral species, and that exon 4 was lost in *X. clivii*.

After the addition of exon 4, *dm-w* probably was still not female-specific, and it wasn't until the MRCA of *X. laevis* and *X. gilli* when the coding region of *dm-w* became longer due to deletion of an ancestral stop codon in exon 4, and when two other genes (*scan-w*, *ccdc69-w*) became W-linked, that this supergene became completely empowered to drive feminization. Moreover, in almost all species in which exon 4 either retains the ancestral stop codon, or where exon 4 is not present, the available data suggests that *dm-w* is not female-specific. An exception is *X. andrei* where *dm-w* may be female-specific based on a small sample size (3 females, 2 males; Table 3.1). Of note, is that we do not know if *dm-w* is female-specific in *X. vestitus*, although we did detect exons 2 and 3 in one female in this species. Our efforts also expand previous work (Cauret et al. 2020) by showing that *dm-w* is not completely female-linked in the dodecaploid *X. kobeli*, the tetraploids *X. largeni X. petersii*, or in some populations of the tetraploids *X. victorianus* (Table 3.1; discussed further below).

Previous research indicates that three W-linked loci in *X. laevis*, *dm-w*, *scan-w* and *ccdc69-w*, each became W-linked due to independent duplication/translocation because their closest paralogs in the autosomes are not tightly linked (Mawaribuchi et al. 2017b). Specifically, homeologs of *dmrt1* are on chr1L and chr1S at positions ∼127 and 104 Mb in *X. laevis* genome assembly 9.2, respectively; homeologs of *scan-w* are potentially also on chr1L and chr1S but far away from *dmrt1* (they are on scaffolds 77 and 8737 in the *X.*

laevis genome assembly 9.2 but an ortholog maps to 6.5 Mb on chr1 in the *X. tropicalis* genome assembly 10); homeologs of *ccdc69-w* are on chr3L and chr3S at positions ∼10 and 7.6 Mb, respectively, in the *X. laevis* genome assembly 9.2; Mawaribuchi et al. 2017b. Our results provide a temporal dimension to these independent origins, and suggest that *scan-w* and possibly *ccdc69-w* originated by duplication after the arrival to the W-chromosome of *dm-w* exons 2, 3, and 4. The most parsimonious scenario for the origin of *scan-w* is in an ancestor of *X. gilli* and *X. laevis*. *Ccdc69-w* may have also arose in this ancestor, or perhaps earlier, followed by degeneration of one or another exon in *X. largeni* and *X. andrei*.

Overall, these results thus suggest that components of the sex-determining pathway of *X. laevis*, such as *dm-w* exons 2+3, *dm-w* exon 4, *scan-w*, and *ccdc69-w* arrived independently, before or at the same time that this locus became the trigger for sex determination. Additionally, extension of the coding region of exon 4 probably occurred prior to or at the same time that supergene became female-specific.

In *X. laevis*, all three exons of *dm-w* reside on one end of chromosome 2L (Yoshimoto et al. 2008). At this time, we do not know the genomic location of exons 2 and 3 in species in which *dm-w* is not female-specific, including those that lack exon 4 such as *X. clivii*, and those that have exon 4, such as *X. kobeli*. In these other species, it is possible that *dm-w* resides in a genomic position that is orthologous to *X. laevis* chromosome 2L and that there are other sex determining genes located elsewhere in the genome as well.

A caveat to our conclusions from the targeted capture data regarding the timing of origin of components of the *X. laevis* W-linked supergene and gene loss of *dmrt1* is that it is possible that there are false negatives, where a gene was not detected in some species even though it was present. However, we suspect that the frequency of false negatives in our capture data is very low for several reasons. The congruence between the results from different capture data for *dm-w* exons 2 and 3 (Cauret et al. 2020) and also from a PCR survey for these exons (Bewick et al. 2011) is very high, with only two biologically plausible discrepancies (a failure to detect exon 4 in two species). Similarly, for *dmrt1*, we were able to detect a number of paralogs equal to or close to the expected number based on the ploidy levels for all exons, and with very high congruence with results from a PCR survey for *dmrt1* exons (Bewick et al. 2011). With only a few exceptions that could arise from allelic variation or partial duplication, the number of copies detected from capture of different *dmrt1* exons was highly consistent within each species (Table 3.2). Additionally, for *X. clivii*, the absence of *dm-w* exon 4 in a female carrying exons 2 and 3 is in agreement with the results of an extensive PCR survey in wild and lab cross individuals during which we were unable to amplify exon 4 (data not shown).

3.3.2 *dm-w* is required for female differentiation, but *scan-w* and *ccdc69-w* are not

The evolutionary correspondence between the onset of female specificity of *dm-w* and the arrival of *scan-w* and *ccdc69-w* to the female-specific portion of the W chromosome

opens the possibility that a combination of some or all of these loci are necessary for female differentiation. To explore this further, we generated knock-out animals of *scan-w* and *ccdc69-w* and found that they developed into phenotypic females. This allows us to reject the notion that all three of the female-specific loci (these two plus *dm-w*) on the *X. laevis* W-chromosome are essential components of a female-determining supergene, although their co-inheritance defines them as a supergene nonetheless. However, it is not yet clear whether these *scan-w* and *ccdc69-w* knockout animals are fertile, and future work is needed to assess whether *scan-w* and *ccdc69-w* function collaboratively with *dm-w* to orchestrate female fertility. Thus far, we have been unable to produce eggs from our knock out *scan-w* or *ccdc69-w* females. It is possible, for example, that these loci affect cellular structure of the gonads rather than viability or external morphology. Also possible is that removing one of these genes could cause female differentiation of a subset of the knockout individuals to fail. We are currently pursuing these possibilities by trying to generate a larger sample of knockout individuals for study.

Also possible is that *scan-w* and/or *ccdc69-w* could affect traits that have nothing to do with fertility but nonetheless could affect reproductive success. For example, in *Xenopus*, females metamorphose faster (Babošová et al. 2018) and grow into larger adults than males. It is thus conceivable that some aspects of secondary sexual differentiation have sexually antagonistic effects (e.g., if being large is generally costly, but beneficial in females for production and storage of eggs). Sexual antagonism has been suggested as a driving force for the extension of regions of recombination suppression (Rice 1987) because linkage to sex-specific genomic regions can resolve genomic conflict associated with sexual antagonism (Fisher 1931). In this way, natural selection may have favoured W-linkage of *scan-w* and/or *ccdc69-w* because females-specific inheritance permits benefits to be enjoyed by females but costs to not be suffered by males. However, contrary to this expectation, in some frogs, male reproductive success does not seem to be linked to the Y-chromosome differentiation (Veltsos et al. 2019), which suggests that genomic conflict may be rare or resolved via other mechanisms (e.g., sex-biased expression). Similarly in fungi, antagonistic selection between mating types does not seem to be the driver of the expansion of the non-recombining region (Bazzicalupo et al. 2019). Further scrutiny of behaviour and development of our *scan-w* and *ccdc69-w* knockout lines may help shed light on some of these possibilities.

3.3.3 Developmental systems drift in sex determination among populations of *X. victorianus* and *X. petersii*

Developmental system drift (DSD) refers to the evolution of diverse genetic underpinnings in different species of conserved traits (True and Haag 2001). In many species, developmental pathways linked to sexual differentiation are crucial for reproduction but are orchestrated by diverse genes and genetic interactions, and are thus a prime example of developmental systems drift (True and Haag 2001).

Findings discussed here and elsewhere (Cauret et al. 2020) evidence DSD of sex determination by demonstrating that *dm-w* is female-specific in some *Xenopus* species but not others. Another striking line of evidence of DSD of sex determination in *Xenopus* is found at the population level within *X. victorianus*, where *dm-w* is found in both sexes in animals from the Lendu Plateau, DRC, but only in females in populations elsewhere (e.g., Lwiro, DRC). *X. victorianus* from the Lendu Plateau are also genetically differentiated from other populations (Furman et al. 2015). The populations of *X. victorianus* in which *dm-w* is female-specific also have *scan-w* (minimally exon 1) and *ccdc69-w* (both predicted exons) with uninterrupted coding regions based on comparison to *X. laevis*. Interestingly, 310 bp of *scan-w* exon 3 was previously (and unknowingly) sequenced in *scan-w* in *X. gilli*, *X. laevis*, *X. poweri* and *X. victorianus* but not *X. petersii* before the annotation of *scan-w* was available (Furman et al. 2015). All the *X. victorianus* individuals sequenced by Furman et al. 2015 had a one bp deletion which creates a frameshift mutation and an early stop codon. Twelve out of 24 *X. victorianus* individuals also carried a -10 bp deletion after the first deletion but was not restricted to a specific geographical population: all the individuals from Lendu Plateau and Orientale Province carried this deletion as did all but one *X. victorianus* from South Kivu Province, DRC. Direct comparison of the *scan-w* data from Furman et al. 2015 and this study is not possible because the former study sequenced a portion of exon 3 and this study sequenced exon1.

Overall, our data suggest the existence of two sex determining systems within *X. victorianus*: one relying on the *dm-w* system and another with *dm-w* not being the main trigger for sex determination. It is possible that the Lwiro population is reproductively isolated from other *X. victorianus* species and therefore constitutes an undescribed species. Either way, these two groups are among the most closely related *Xenopus* (divergence < 4 million years ago; Evans et al. 2019) yet to be identified with distinctive systems for sex determination. The divergence between these *X. victorianus* populations is comparable to that between the dodecaploids *X. kobeli* and *X. ruwenzoriensis*, the former of which has a non-sex-linked *dm-w* whereas the latter does not (this study; Cauret et al. 2020), although it is not clear whether these two species share a common system for sex determination. These findings in *X. victorianus* thus underscore how rapidly DSD of sex determination occurs in *Xenopus*.

Multiple sex determining systems also have been identified in *X. tropicalis* (Roco et al. 2015), and several other frogs such as *Rana temporaria* (Rodrigues et al. 2016), *Rana rugosa* (Ogata et al. 2008), and *Leiopelma hochstetteri* (Green et al. 1993). Similar to *X. victorianus*, differences in sex determining systems in some other frogs are geographically structured (Ogata et al. 2008; Rodrigues et al. 2016). One exception appears to be *X. tropicalis*, where three sex chromosomes co-occur in the same population (Appendix E).

3.3.4 Duplication and pseudogenization of *dmrt1* in polyploid *Xenopus*

Our work on *dmrt1* extends previous efforts to quantify pseudogenization of *dmrt1* of all exons of all *Xenopus* species except *X. fraseri*. Results indicate that pseudogenization occurred independently several times in closely related homeologs of *dmrt1* – mostly from the S-subgenome – and that different species with the same ploidy level have different copy numbers of this important male-related gene (Table 3.2). We found that early stop codons and large deletions occur most often in exon 2 upstream or within the early portion of DM-domain which spans exons 2 and 3 of *dmrt1*. This implies that there are functional consequences of gene silencing of these homeologs and opens the posssibility that gene silencing was favoured by natural selection. Mutations near the DM-domain in *dmrt1* are correlated with switches between temperature and genetic sex determining systems in reptiles (Janes et al. 2014), although we do not have information with which to evaluate the effects of these mutations in *Xenopus*.

It remains unclear why *dm-w* appears to segregate as a single allele in *X. clivii*, *X. kobeli*, and other species (and possibly did so as well in their MRCA) as opposed to being a "regular" autosomal locus with two alleles in all individuals of both sexes (*X. itombwensis* being an exception to this; Cauret et al. 2020). A key unanswered question raised by these findings asks what was the ancestral function of *dm-w* before it became sex-specific, and whether *dm-w* operated in a polygenic sex determination system in this ancestor. It is possible that *dm-w* was (and in some species is) an "influencer" on female differentiation that tends to be found in females, depending on variation at other loci. Further insights into this issue could be gained with manipulative experiments that evaluate the effects on sexual differentiation of knocking out *dm-w* in species where this locus is not female-specific.

3.4 Methods

3.4.1 Targeted next-generation sequencing and Sanger sequencing of W-specific and autosomal loci

We used targeted next-generation sequencing to assess presence, absence, and sequence variation of *dm-w* exon 4, first exon of *scan-w*, and both exons of *ccdc69-w* in 28 of 29 *Xenopus* species using the same panel of individuals and genomic DNA libraries as detailed previously (Cauret et al. 2020). Briefly, we used oligonucleotide probes (GenScript) to enrich libraries for sequences similar to the loci of interest. Sequences obtained for each exon were independently assembled using Trinity 2.5.1 (Grabherr et al. 2011) and assembled sequences from each species were identified using blastn (Altschul et al. 1997). Due to repetitive regions in *scan-w*, a 300 bp cutoff on all blast hits was applied. Sequences from each exon were aligned using MAFFT version 7.271 (Katoh and Standley 2013), adjusted manually, and manually inspected for putatively chimerical sequences.

Our alignment included reference sequences from *X. tropicalis* genome assembly 10 and *X. laevis* genome assembly 9.2 for each exon plus 200 bp upstream and downstream. We then used PAUP v.4.0b10 (Swofford and Sullivan 2003) to produce a neighbor-joining tree to assist in identification and annotation of homeologs. We also used this approach to sequence *dmrt1* homeologs. When possible, annotation of homeologs follows Evans et al. 2005.

PCR assay and Sanger sequencing were also performed to evaluate the female-specificity of *dm-w* in additional species beyond those evaluated previously (Cauret et al. 2020): *X. kobeli*, *X. petersii*, *X. poweri*, *X. andrei*, and *X. largeni*. We also included additional *X. victorianus* individuals from two geographical areas to test for geographical differentiation in *dm-w* efficiency. The phenotypic sex of each specimen was determined based on external morphology, or surgically after euthanasia. PCR assays were also performed in order to test linkage of *dm-w* exon 4 to exons 2 and 3 in some species by attempting to amplify each of these exons in several individuals of some species: *X. kobeli*, *X. clivii* (this study), *X. pygmaeus* (Cauret et al. 2020), *X. gilli* (Bewick et al. 2011) , *X. laevis* (Bewick et al. 2011; Furman et al. 2015), *X. poweri*, *X. petersii* and *X. victorianus* (Furman et al. 2015).

3.4.2 Role of W-specific genes in *X. laevis*

Knock out of *dm-w*, *scan-w*, and *ccdc69-w*

Single guide RNAs (sgRNAs) were designed to target the beginning of the coding region for *dm-w*, *scan-w*, and *ccdc69-w* using CRISPRdirect (`https://crispr.dbcls.jp/`) with an aim of maximising disruption of protein function. The specificity of our guides was evaluated using the *X. laevis* genome assembly 9.1. Single stranded guide RNA (sgRNA) was generated from a DNA template that contained a promoter (T7 for *scan-w* and *ccdc69-w*; SP6 for *dm-w*) and a universal reverse primer for subsequent transcription. The DNA template was then used for sgRNA production using the Megascript T7 or SP6 kit (Invitrogen, Thermo Fisher Scientific).

SgRNAs were injected with the Cas9 protein into one cell embryos. Because cutting generally happens after several rounds of cell division, the resulting F0 embryos are mosaics of wild-type and mutant cells. F0 females (in the case of *scan-w* and *ccdc69-w*) or male (in the case of *dm-w*) were then back-crossed to wildtype (J strain) males or female respectively. Mutations were confirmed by sequencing and the genetic sex was verified by amplification of other W-specific genes. Phenotypic sex was inferred based on morphological characteristics (for females, larger size, larger cloaca, for males, smaller size, presence of nuptial pads). F1 individuals were also crossed to wild-type individuals to evaluate fertility, with ovulation (females) or clasping (males) facilitated by injection of human chorionic gonadotropin (Sigma).

3.4.3 Phylogenetic bias in *dmrt1* pseudogenization

To identify pseudogenes of *dmrt1*, assembled sequences were aligned using MAFFT and translated using Geneious v.6.1.8 (`http://www.geneious.com`, Kearse et al. 2012), using the reference genome of *X. tropicalis* (Genbank accession: XM_031890717.1) and *X. laevis* (Genbank accession: XM_018250680.1 (*dmrt1.L*) and NM_001085483.1 (*dmrt1.S*)) and associated reading frame annotations for each exon. Sequences were deemed to be a pseudogene if any exon had an earlier stop codon than the references. We also considered truncated loci that potentially could have partial function.

3.5 Acknowledgements

We thank the National Xenopus Resource for maintaining the knock out lines and hosting the Xenopus Gene Editing Workshop at the Marine Biological Laboratory, and Takuya Nakayama for assistance with sgRNA design. This work was supported by the Natural Science and Engineering Research Council of Canada (RGPIN-2017-05770), Resource Allocation Competition awards from Compute Canada, the Whitman Center Fellowship Program at the Marine Biological Laboratory, the Museum of Comparative Zoology at Harvard University, and National Institutes of Health grants R01-HD084409 and P40-OD010997.

Chapter 4

Conclusions and Outlook

This thesis explored genetic sex determination and sex chromosomes of closely related frogs species in the genus *Xenopus* with an aim of providing insights into how the genetic underpinnings of important traits evolve. We found that the function of a new trigger for female differentiation was variable over time and is variable among species, and provide evidence for empowerment, sidelining, and loss of function. This work considered the impact of changes in the structure of a gene, the linkage of new genes, and evolution of a competitor gene called *dmrt1* in the evolution of the efficiency of a trigger for sex determination.

Even though sexual reproduction is present in many species, the mechanisms of sex determination are diverse. Chapter 2 further characterized the diversity of sex determining systems in pipid frogs by identifying sex chromosomes for the first time in two genera – *Pipa* and *Hymenochirus*. These systems are distinct from each other and from those known in *Xenopus* in their genomic locations, evolutionary origins, and whether males or females are the heterogametic sex. We propose a new explanation for the maintenance of homomorphic sex chromosomes that relates to the efficiency of a trigger for sex determination. We posit that an inefficient sex determining locus is an unfavourable location for resolution of genomic conflict associated with mutations with sexually antagonistic fitness effects and these types of triggers are thus more likely to be associated with homomorphic sex chromosomes. This explanation differs from from the "*Fountain of youth*" hypothesis which presupposes a lower recombination rate in the heterogametic sex, whereas our explanation makes no such assertion. Moreover, in some *Xenopus* species (*X. borealis* and *X. laevis*), and perhaps all of them, the recombination rate is higher in females (the heterogametic sex) (Furman and Evans 2018).

It is surprising that *dm-w* is generally found as a single allele in species, even though it may not be female-specific (as in *X. clivii*, *X. pygmaeus*, and *X. kobeli*). No heterozygous sites were detected in these species, and in a *X. clivii* cross between a mother carrying *dm-w* and a male without this gene, some offspring did not inherit *dm-w*. If the mother carried two similar *dm-w* alleles, all the offspring should inherit a *dm-w* allele. This suggests that *dm-w* might still have a role in female differentiation and possibly in combination with other loci in a polygenic system as observed in some fish species (e.g. Liew et al. 2012). Alternatively, it could be that *dm-w* is close but incompletely linked to

another trigger for sex determination in these species where *dm-w* is not female-specific. Further narrowing down which genomic regions are sex-linked in *X. clivii* or *X. pygmaeus* (similarly to what was performed in Chapter 2) clearly could increase understanding of how sex determination is controlled in those species.

Chapter 3 explored function and evolution of the small female determining supergene in *X. laevis*. This work found that addition and extension of the coding region of the last exon of *dm-w* and also the addition of two other female-specific genes was associated with the onset of female specificity of this supergene. Knock-out lines of two female-specific genes (*scan-w* and *ccdc69-w*) were established and suggest that these loci are not necessary for female development. However knock-out animals of *dm-w* did cause phenotypic sex reversal, which is partially consistent with previous findings (Yoshimoto et al. 2008). Ongoing experiments aim to uncover whether *scan-w* and *ccdc69-w* are necessary for fertility and/or whether they influence the efficiency of *dm-w* in driving female development in *X. laevis*.

DSD in sex determination is sometimes thought to be due to the emergence of a novel trigger for sex determination while function of downstream genes remain largely conserved (reviewed in Bachtrog et al. 2014). Some of our results (e.g. loss of *dm-w* in multiple lineages) are in agreement with this idea. However, we also found that *dmrt1*- a gene with a conserved role in sex determination and autosomal in at least some *Xenopus* species - has different dosages in different species – even species with the same ploidy level. Though *dmrt1* is not the trigger for sex determination in species in which *dm-w* is female specific, *dmrt1* might still play an important role in the evolution of sex determination in this group. It is possible that *dmrt1* took over the role of "master" trigger for sex determination in some species where *dm-w* was lost, although we know that it is not the case in *X. borealis* (Furman and Evans 2016), *X. laevis* (Yoshimoto et al. 2008), and *X. tropicalis* (Roco et al. 2015). In the future, identification of sex chromosomes and other triggers for sex determination in species lacking *dm-w* or in which *dm-w* is not completely sex linked, as well as characterizing *dmrt1* pattern of expression in those species would improve our understanding of *developmental system drift* in sex determination. Ongoing and future efforts, especially using genome editing and genomic data promise to fill in some gaps in understanding of DSD.

Appendix A

Supplemental information: Developmental systems drift and the drivers of sex chromosome evolution

A1 Supplemental Methods

A1.1 Species identification

Species identification of individuals in the genus *Xenopus* (Table A1.1) was based on molecular and morphological variation (Evans et al. 2004; Evans et al. 2015). Species identification of the frogs in the genera *Hymenochirus* and *Pipa* was based on taxonomic annotations of submitted data in GenBank. This was assessed by (1) examining the best BLAST (Altschul et al. 1990) hit of a sequenced portion of the mitochondrial 16S ribosomal RNA gene, and by (2) evaluating the phylogenetic position of mitochondrial sequences from our genome assemblies with respect to several complete mitochondrial genome sequences in GenBank. The portion of the 16S genome was sequenced from both parents of the *Hymenochirus* and *Pipa* families using the 16sc-L and 16sd-H primers (Evans et al. 2003). The phylogenetic analysis included data from *H. curtipes*: KY080141.1, KY080143.1, *H. boettgeri*: AY341726.1, AY523756.1, *P. parva*: AY581622.1, KU495452.1, and *P. pipa*: AY581621.1, DQ283053.1. Both approaches supported our species designations of *H. boettgeri* and *P. parva* (i.e., respective best blast hits to each of these species, and strongly supported clades that contained mitochondrial DNA sequences from our genome assemblies and GenBank accessions of whole mitochondrial genome sequences from these species).

Library construction, sequencing, and bioinformatics of targeted next generation sequence data

Illumina libraries for targeted next generation sequencing were constructed using methods described in Meyer and Kircher (2010), and had sample-specific 7-base pair (bp) barcodes. Capture probes were prepared and used to hybridize to and pull out molecules in individual DNA libraries for sequencing, as described previously (Fu et al. 2013). A perl script was used to generate all possible unique overlapping 52-bp probes with 1 base pair tiling from a multifasta file containing *dmrt1* sequences from *X. tropicalis*, *X. laevis*, and *X. borealis* (Genbank accession numbers (XM018089347.1, XM018250680.1, NM001085483.1, MK907575, and MK907576). An 8-bp flanking sequence was attached to the 3' end of each of 8,337 unique probe sequences, and the resulting 60-bp probes were printed (with redundancy) on 30% the probes of a one million probe array (Agilent, USA) as described previously (Fu et al. 2013). Sequences were multiplexed and paired end sequencing was performed on a portion of one lane of an Illumina 2500 machine.

Reads were demultiplexed based on exact matches to species-specific barcodes. Overlapping paired-end reads were merged into a consensus read when possible. We then performed a *de novo* assembly using non-merged paired reads and merged consensus reads with the program TRINITY version 2.5.1 (Grabherr et al. 2011) using default parameters. The advantage of a *de novo* assembly over a reference-based assembly is that the former approach may more accurately reconstruct insertion/deletion polymorphisms and avoid possible genotyping biases from the reference sequence. Although Trinity is usually used for transcriptome assembly, this software performed better than SOAP denovo2 (a software designed for genome assembly) in assembling frog mitochondrial genomes from shotgun sequences (Yuan et al. 2016).

We then used BLAST with each of the exons from *X. tropicalis dmrt1* as a query against the assembled contigs from each species to identify contigs with homology to *dmrt1* and *dm-w*. These contigs were combined into a multifasta file, and aligned with MAFFT version 7.271 (Katoh and Standley 2013) with the 'adjustdirection' option. We then modified this alignment manually to identify non-homologous intronic sequences and to fine tune alignment of homologous sequences using Mesquite version 3.04 (Maddison 2008).

We included in these alignments portions of the *X. laevis* genome sequence, version 9.2, that surround and include exons 2 and 3 of *dm-w*, and the *X. laevis dm-w* mRNA sequence (GenBank accession $NM001114842.1$). These alignments allowed us to identify distinctive sequences in *X. laevis dm-w* that were and that were not homologous to *dmrt1*, and to conclusively identify *dm-w* sequences in some of the species. Manual inspection of this alignment also allowed us to identify and exclude portions of putatively chimerical sequences composed of different portions of multiple *dmrt1* homeologs, and also three sequences that were part *dmrt1* homeologs and part *dm-w* (the latter three chimerical sequences were verified by comparison to *dm-w* Sanger sequences from Furman et al. 2015).

To further test whether we correctly identified *dm-w* sequences, we performed phylogenetic analyses to test the hypothesis that the putative *dm-w* sequences that were identified by targeted next generation sequencing had the expected monophyletic relationship with respect to *dmrt1* sequences. Support for this relationship was assessed using maximum likelihood bootstrap values that were calculated using the ultrafast bootstrap algorithm of IQ-TREE version 1.6.8 (Nguyen et al. 2014), and a model of evolution that was selected using the Bayesian Information Criterion for each exon.

Table A1.1: Description of the samples used for the capture sequencing (Species, Field and Museum ID, Locality, Ploidy, Sex) and the number of reads (Reads) obtained from each library. Reads include sequences captured by *dmrt1* probes described here and also probes for two other autosomal genes (*rag1*, *rag2*) that were not included in this study.

Species	Field ID	Museum ID	Locality	Ploidy	Sex	Reads
Subgenus *Silurana*						
X. tropicalis	AMNH17274	Voucher lost (MNHG)	Sierra Leone - near Freetown	diploid	F	3411600
X. epitropicalis	AMNH17278	Voucher lost (MNHG)	DRC - Kinshasa	tetraploid	F	5130501
X. mellotropicalis	AMNH17288	Voucher lost (MNHG)	Cameroon - Nkoemvone	tetraploid	M	4172113
X. calcaratus	VG09-364	NMP6V 74630/1	Cameroon - Bakingili	tetraploid	F	3862925
Subgenus *Xenopus*						
X. clivii	AMNH17254	MHNG2645.069	Ethiopia - near Addis Ababa	tetraploid	F	4207310
X. muelleri	Xen203	No voucher	Tanzania - Ifakara	tetraploid	F	4013162
X. borealis	AMNH17315	Voucher lost (MNHG)	Kenya - Samburu Range	tetraploid	F	5643148
X. fischbergi	AMNH17297	Voucher lost (MNHG)	Nigeria - Jos	tetraploid	F	5460074
X. largeni	AMNH17292	MHNG2644.059	Ethiopia - near Kibre Mengist	tetraploid	F	3355198
X. pygmaeus	AMNH17320	Voucher lost (MNHG)	DRC - Boende	tetraploid	M	4151297
X. allofraseri	BJE3488	MCZ A-148164	Equatorial Guinea - Comedor	tetraploid	F	13181569
X. parafraseri	AMNH17282	MHNG2645.084	Cameroon - Yaounde	tetraploid	F	23571704
X. gilli	Xr(0-3)	No voucher	South Africa	tetraploid	F	5602674
X. petersii	AMNH17325	MHNG2645.094	Congo Rep - near town of Koulila	tetraploid	F	3858926
X. victorianus	xen234	No voucher	Uganda - Kitanga	tetraploid	F	4322852
X. poweri	AMNH17260	MHNG2645.073	Nigeria - Jos	tetraploid	F	5030963
X. laevis	kml(5)	No voucher	South Africa - Kleinmond	tetraploid	F	4680406
X. amieti	AMNH17268	MHNG2645.079	Cameroon - Galim	octoploid	F	4563323
X. wittei	AMNH17307	MHNG2645.092	Uganda - Chelima Forest	octoploid	F	3740421
X. boumbaensis	AMNH17284	MHNG2644.057	Cameroon - Moloundou	octoploid	F	4825896
X. andrei	AMNH17258	MHNG2645.072	Cameroon - Longyi	octoploid	F	4702872
X. itombwensis	BJE00275	MCZ A-138192	DRC - Miki	octoploid	M	5397761
X. lenduensis	BJE2973	MCZ A-140111	DRC - Emilio	octoploid	M	4399939
X. vestitus	RT2	No voucher	Uganda - near Kisoro	octoploid	F	4112075
X. boumbaensis	AMNH17283	MHNG2645.085	Cameroon - Moloundou	octoploid	M	3523883
X. ruwenzoriensis	AMNH17317	Voucher lost (MNHG)	Uganda - Semliki Forest	dodecaploid	F	3645069
X. kobeli	BJE3075	MCZ A-148037	Cameroon - Meganme	dodecaploid	F	5357943
X. eysoole	BJE3218	MCZ A-148095	Cameroon - Nkambe	dodecaploid	F	3948190
X. longipes	BJE3190	MCZ A-148072	Cameroon - Lake Oku	dodecaploid	F	5914696

A1.2 Tests for female-specificity of *dm-w*

Because we detected *dm-w* in male individuals of two species in the targeted next generation sequencing data (Table A1.1), we screened male and female individuals of six species for *dm-w* to further explore whether *dm-w* was female-specific in these species. Sample information for these screens is presented in Table A1.4. Primers are presented in Table A1.3 and (for *X. laevis*, *X. victorianus*, and *X. gilli*) in (Yoshimoto et al. 2008; Furman et al. 2015). Negative controls with no genomic DNA were included for each amplification, and for each of these DNA extractions a positive control (mitochondrial DNA or an autosomal gene) was performed.

A1.3 Ancestral state reconstruction of *dm-w* sex-specificity

We estimated phylogenetic relationships among complete mitochondrial DNA genome sequences from Evans et al. (2019) (except *X. fraseri*), and GenBank (*H. boettgeri* HM991331.1, *Rhinophrynus dorsalis* HM991334.1, and *P. pipa* CG244477.1 in lieu of *P. parva*) using the ultrafast bootstrap algorithm of IQ-TREE version 1.6.8 (Nguyen et al. 2014), with a model of evolution that was selected by the Bayesian Information Criterion (GTR+F+I+G4). The D-loop was excluded from this analysis because homology was difficult to ascertain in this region. The alignment for this analysis is provided in the Supplementary Data (File S1). The phylogeny was rooted with the mitochondrial genome of *R. dorsalis*, which was then pruned from the phylogeny for the ancestral state reconstruction. Using this phylogeny and our observations, we calculated a maximum likelihood estimation of ancestral sex-specificity of *dm-w* using the R package 'ape' (Paradis and Schliep 2018). Each species was categorized as either (0) never carrying *dm-w* (black in Fig. 2.1), (1) with *dm-w* detected in multiple females but not multiple males (red in Fig. 2.1), or (2) with *dm-w* not detected in one or more female (light blue in Fig. 2.1). Data were coded as missing for species in which only one male was used in the targeted next generation libraries (white in Fig. 2.1) and species in which one female was identified as carrying *dm-w* (pink in Fig. 2.1). This is because for these species, we lacked information about whether *dm-w* is fixed only in females. The transition matrix allowed for an equal rate of transition from state (0) to states (1) or (2) but no reversal from states (1) or (2) to state (0), and a separate reversible rate of transition between states (1) and (2).

We also performed the ancestral state reconstruction analysis on a time-calibrated Bayesian phylogeny that was estimated with BEAST version 2.5.2 (Bouckaert et al. 2014), and results (presented in Fig. 2.1) were essentially identical. In the Bayesian phylogenetic analysis, we used the same model that was selected above by IQTREE using the Bayesian Information Criterion, and we assumed a relaxed log-normal molecular clock. The prior probability for the age of the root was set assumed to be normally distributed with a mean of 159.4 million years and a standard deviation of 6.5 million years, based on the results of (Feng et al. 2017).

A1.4 The sex chromosomes of three other pipids

We characterized the sex chromosomes of three pipid species (*H. boettgeri*, *P. parva*, *X. mellotropicalis*) by studying how molecular variation across the genome was passed from parents to offspring using whole genome sequencing and reduced representation genome sequencing (RRGS). *H. boettgeri* parents were obtained from Xenopus Express (Brooksville, FL, USA) and a family was generated at McMaster University. A *P. parva* family was obtained from Standard Reptile (Wichita, KS, USA). *X. mellotropicalis* parents were provided by Darcy Kelley (Columbia University) and a family was generated at McMaster University. The geographic origins of the *H. boettgeri* and *P. parva* parents are unknown, but the *X. mellotropicalis* parents originated from Gabon.

As described below, we generated draft genome assemblies from one or both parents of each family. These assemblies were used as references to map and genotype whole genome sequences from each parent, map and genotype reduced representation genome sequencing (RRGS) data from each family, design primers to survey putative sex-linked SNPs in each family using Sanger sequencing, and to identify orthologous genomic regions in the *X. tropicalis* genome sequence.

A1.5 Draft genome assembly for *H. boettgeri* and *P. parva*

For *H. boettgeri*, the genome of each parent was sequenced on one half lane of an Illumina HiSeqX machine using paired-end sequencing, with an average insert size of 350 bp. For *P. parva*, the genome of each parent was sequenced on one third of a lane (2/3rds of a lane total). Adapters on the 5' ends of the sequences were removed with TRIMMOMATIC v-0.32 (Bolger et al. 2014), and retaining reads of at least 36 bp and with an average quality per base higher than 15 on a sliding window of 4 bp were retained for analysis. Adapters on the 3' ends of the sequences were removed with SCYTHE (https://github.com/vsbuffalo/scythe). A 19-mer (k-mers of size 19 bp) occurrence distribution was generated using JELLYFISH v-1.1.11 (Marçais and Kingsford 2011), and this was used as input for QUAKE v-0.3 (Kelley et al. 2010) in order to identify and correct reads with rare 19-mers, which were assumed to include errors. K-mer abundance thresholds of 4 and 2 for *H. boettgeri* and *P. parva*, respectively, were calculated and used by QUAKE. After these quality control steps, we obtained a total of 919–933 million reads for each *H. boettgeri* parent and 274-304 million for each *P. parva* parent.

We generated *de novo* genome assemblies for each of the parents, and also a combined assembly using data from both parents for each species. In principle, the combined assembly for each species should include sex-specific genomic regions, irrespective of which sex is heterogametic. Assemblies were performed using ABySS v.1.9.0 (Simpson et al. 2009), and SOAPdenovo v.2.04-r240 (Luo et al. 2012). For the ABySS assemblies we used a k-mer size of 64. For the SOAPdenovo genome assemblies, we explored k-mer values between 23 to 63 in increments of 10. These assemblies had N50 values ranging from 2.0 – 6.6 kilobasepairs (kbp) for *H. boettgeri* and 5.9 – 7.9 kbp for *P. parva*.

A1.6 Draft genome assembly of *X. mellotropicalis*

For *X. mellotropicalis*, a more intensive sequencing effort was conducted due to complexities associated with its allopolyploid genome. We suspected this species to have female heterogamy because all *Xenopus* species investigated so far have female heterogamy (Furman and Evans 2018) except *X. tropicalis*, which is polymorphic for male and female heterogamy (Roco et al. 2015). For this reason we selected a female individual for high coverage shotgun genome sequencing. We sequenced paired-end libraries with small insert sizes (180bp, 400bp, 1kb) and mate-pair libraries with longer insert sizes (6kb, 10kb) using an Illumina HiSeq 2500 machine. The read lengths were 101bp for all the libraries except the 6kb insert library that had a read length of 150bp. Adapters and low quality reads were filtered using TRIMMOMATIC v-0.32 (Bolger et al. 2014) and QUAKE v-0.3 (Kelley et al. 2010) (using a k-mer abundance threshold of 2) as described for *H. boettgeri* and *P. parva*.

We performed a de novo genome assembly with ALLPATHS-LG (Gnerre et al. 2011) and SOAPdenovo v.2.04-r240 (Luo et al. 2012). ALLPATHS-LG was used because it explicitly incorporates the variation in insert sizes of the sequence libraries that we generated. The ALLPATHS-LG assembly was obtained using default parameters and a ploidy level set to diploid ("ploidy=2") because we expected this allotetraploid genome to be disomic. For SOAPdenovo, we used a k-mer size of 33 without the gap closing option due to associated memory requirements. The ALLPATHS-LG assembly for *X. mellotropicalis* had an N50 of 2.0 kbp and the SOAPdenovo genome assembly (using a k-mer of 33) had an N50 of 11.6 kbp.

A1.7 Identification and masking of repetitive elements

Repetitive elements present challenges to our efforts to identify sex-linked variation in our focal species, and also to identify orthologous genomic regions in the genome sequence of *X. tropicalis*. For *H. boettgeri*, *P. parva* and *X. mellotropicalis*, no repeat library was available, and for each assembly we therefore identified and masked these regions using RepARK v. 1.3.0 (Koch et al. 2014) and REPEATMASKER version open-3.2.6. RepARK uses JELLYFISH (2.2.4) (Marçais and Kingsford 2011) to count 31-mers from NGS reads, identify abundant 31-mers, and then assemble them into contigs using VELVET (1.2.10) (Zerbino and Birney 2008). The k-mer abundance threshold calculated by RepARK was 92 and 139 for the *H. boettgeri* mother and father, respectively, 56 and 62 for the *P. parva* mother and father, respectively, and 50 for the *X. mellotropicalis* female. TEclass v. 2.1.3 (Abrusán et al. 2009) allowed us to identify different classes of repetitive elements (see Supplemental Results below). Species-specific repeat libraries were used as input libraries for REPEATMASKER version open-3.2.6 to mask repetitive elements in the respective draft genome assemblies.

A1.8 RRGS for three pipid families

We identified sex-linked genomic regions based on genotypes from RRGS data from each of the three pipid families that were mapped to the repeat-masked draft genome assemblies described above. For *P. parva* and *H. boettgeri*, library preparation of RADseq data was carried out by Floragenex, Inc. (Portland, OR, USA) using the SbfI enzyme, and sequencing was performed on an Illumina HiSeq 2500 machine. For *X. mellotropicalis*, library preparation for Genotype by Sequencing (GBS) (Elshire et al. 2011) was performed at Cornell University. Sequencing was performed using 100 bp reads on an Illumina HiSeq 2500 machine, as described previously (Furman and Evans 2018). Also as detailed previously (Furman and Evans 2018), the nature of the resulting data are similar for each family (i.e., high coverage reads from homologous genomic regions in parents and offspring) even though the protocol for library preparation differs for RADseq and GBS. For all three crosses, we intentionally gathered higher coverage data from both parents.

RRGS data were demultiplexed using the 'process_radtags' command of STACKS (Catchen et al. 2011). Reads were truncated to 70 and 75 bp respectively and over-represented sequences (identified with FASTQC v0.11.3) were removed using TRIMMOMATIC v-0.32 or 0.36 (Bolger et al. 2014). After filtering, we obtained 9–10, 4–6, and 17–23 million reads for each of the *H. boettgeri*, *P. parva* and *X. mellotropicalis* parents. We obtained 2-5 million reads each offspring of *H. boettgeri* and *P. parva*, and 2-10 millions reads for each offspring of *X. mellotropicalis*.

For each family, RRGS reads were mapped to a chimerical assembly obtained from both parents or, for *X. mellotropicalis*, the female-only draft assembly. Read mapping was accomplished using the MEM function of BWA v-0.7.8-r455. The reference genomes for *H. boettgeri* and *P. parva* were the respective SOAPdenovo chimerical assemblies (with k-mer sizes of 64 and 43 respectively), and for *X. mellotropicalis*, the ALLPATHS-LG assembly. These particular assemblies were selected based on the assembly statistics (e.g., N50 value and the proportion of missing data in scaffolds).

For *H. boettgeri* and *P. parva*, we then used the HaplotypeCaller function from the Genome Analysis Toolkit (GATK; DePristo et al. 2011, Auwera et al. 2013), to infer genotypes. The RealignerTargetCreator and IndelRealigner functions of GATK were used to realign insertion/deletion polymorphisms, and the BaseRecalibrator function was used to recalibrate quality scores. Because the RRGS data from *X. mellotropicalis* had lower coverage than that for *H. boettgeri* and *P. parva*, we instead used SAMTOOLS mpileup ($-d8000$) and BCFTOOLS ($bcftoolscall - mv - Oz$) for genotyping.

For *H. boettgeri* and *P. parva*, we retained high-quality genotypes with the following statistics: quality score normalized by allele depth (QD) > 2, Fisher strand bias (FS) < 60, map quality (MQ) > 40, strand odds ratio (SOR) < 3, rank sum test for mapping qualities (MQRankSum) > -12.5, and a test for bias in the position of a variable position with the reads (ReadPosRankSum) > -8. For the low coverage data from *X. mellotropicalis*, we required a minimum depth of 2X. For all three families, a final

quality control step was to consider only those sites with less than 20% missing data, and that were present in at least 80% of the offspring of each sex, or alternatively in at least 80% of the offspring of one sex. We excluded sites whose parent and offspring genotypes suggested errors or *de novo* mutation (e.g., a position with an A/T maternal genotype, G/C paternal genotype, and an A/A offspring genotype) in more then 20% of the individuals. Positions that were genotyped in only one sex (i.e. missing data in all individuals of the other sex), are also potentially in a sex-specific genomic region, but we also discarded these sites because they could also arise due to mutations at enzyme restriction sites which affect library construction (Gamble et al. 2015). Although coverage was low for the *X. mellotropicalis* RRGS data, these quality control steps contributed to the stringency of our inferences, and results were validated with Sanger sequencing (detailed next) and, for *X. mellotropicalis*, consistent with a closely related species (*X. tropicalis*, as described in the main text).

A1.9 Permutation test and validation of sex-linked genomic regions using Sanger Sequencing

For each family, we performed 1000 permutations in which the offspring sex was shuffled, and our pipeline for identifying sex-linked scaffolds re-run. The number of potentially sex-linked sites was recorded from each permutation and compared to the observed.

Using Perl and C++ scripts, we identified putatively sex-linked genotypes in RRGS data that was mapped to draft genome assemblies for each species. This allowed us to identify other putatively sex-linked genomic regions on the same scaffold that were not covered by the RRGS data, and to design primers that could be used to test for other sex-linked SNPs that were not covered by the RRGS data in each family using Sanger sequencing.

Primers were designed using PRIMER3 (Ye et al. 2012). For each amplification, Sanger sequencing was performed in forward and reverse directions for all individuals in the family for which the sex-linked scaffold was identified. Consensus sequences were obtained using GENEIOUS version 6.1.8 (http://www.geneious.com, Kearse et al. 2012). Primers and other information about Sanger sequencing verification are provided in Table A1.3. Sanger sequences were not obtained from a small number of individuals for some amplicons as a result of failed amplifications or failed sequence reactions.

A1.10 Candidate triggers for sex determination.

Although triggers of sexual differentiation may evolve rapidly, several genes involved with the developmental process of sexual differentiation have conserved roles (Angelopoulou et al. 2012; Bachtrog et al. 2014). One way that new triggers could arise is via changes in genetic pathways where downstream genes that are involved in sexual differentiation in an ancestor assume a new role at the top of the regulatory cascade governing sexual differentiation (True and Haag 2001). To provide descriptive information

that could be consistent with this possibility, we used *The Gene Ontology Resource* (http://geneontology.org/) to identify genes that had functional annotations related to "sex" in *X. tropicalis* and *X. laevis*. We were specifically interested in genes that were located in the region of the sex-chromosomes of *P. parva* and *H. boettgeri* that we identified as having a reduced recombination rate under the assumption of conserved synteny in pipid frogs. These genes are listed in Table A1.7. We also included in this list a few other genes that are involved in sex-determination in other species and located in the region of reduced recombination (based on our RADseq and PCR results) on the chromosomes of *P. parva* and *H. boettgeri*. No further analysis was performed on these loci.

A2 Supplemental Results

A2.1 Distinctive nucleotide sequences of *dm-w*

Genomic and transcribed sequences of *X. laevis* have homology between the 3' portions of exon 2 and all of exon 3 of *dmrt1* and *dm-w*, but with distinctions between genes in nucleotide sequence and insertion deletions (Yoshimoto et al. 2008). The 5' and 3' portions of intron 2 (immediately after exon 2 and immediately before exon 3) of *dmrt1* and *dm-w* are also homologous, but distinguished by nucleotide sequence and insertion/deletions. A 5' portion of exon 2 of *dm-w* that spans most of the untranslated region does not have discernible homology to *dmrt1*, and this is also the case for the 3' region of the first intron of *dm-w* that lies upstream of exon 2.

We were able to readily identify *dm-w* sequences in our assemblies based on distinctive nucleotide sequences. For example, there is a deletion in the coding region of *dm-w* exon 2 that was not found in any *dmrt1* homeolog from any species, and there were divergent nucleotides in putative *dm-w* exon 2 and 3 sequences that were shared with *X. laevis dm-w*, but not found in any *dmrt1* homeologs. In exon 3, a deletion was present in all putative *dm-w* sequences except the one from *X. clivii*. Additionally, the 5' UTR and upstream intronic sequences were clearly homologous to *X. laevis dm-w* in all putative *dm-w* sequences.

In both of the independent phylogenetic analyses, putative *dm-w* sequences from exons 2 and 3 formed a clade with respect to all *dmrt1* sequences in 99% of the bootstrap replicates (results not shown). This is expected under the hypothesis that *dm-w* arose only one time from partial gene duplication of *dmrt1*. The alignments for this analysis are provided in the Supplementary Data (File S2 for exon 2, File S3 for exon 3).

A2.2 *dm-w* is not female-specific in some species

For *X. victorianus*, *X. laevis* and *X. gilli*, we used the primers from Yoshimoto et al. 2008 to test for *dm-w* and concurrently amplify *dmrt1* as a positive control. We compared

these amplification and sequencing results from adult individuals to the phenotypic sex of each individual based on external anatomy, and confirmed phenotypic sex with surgical inspection of gonads after euthanasia for all males with *dm-w* and all females without *dm-w*. We additionally screened Sanger sequences from two earlier studies (Furman et al. 2015; Bewick et al. 2011) from each of these species for heterozygous positions. These Sanger sequences included 2–6 amplicons of different regions of *dm-w* that spanned coding regions of *dm-w* exons 2, 3, and 4, a portion of a non-transcribed region upstream of *dm-w*, and a portion of the coding region of *scan-w*, which is a female-specific gene upstream of *dm-w* in *X. laevis* (Mawaribuchi et al. 2017b) that has a shorter coding region in *X. victorianus* and *X. gilli* compared to *X. laevis*.

For *X. laevis* and *X. gilli*, *dm-w* was detected only in females (Table 2.1) and Sanger sequencing of exons and upstream regions did not identify any heterozygous positions in 67 and 14 individuals, respectively. However, in *X. victorianus*, *dm-w* was amplified and sequenced in some male individuals in addition to all female individuals (Table 2.1) and we identified 11 heterozygous positions in six of these 26 individuals (Fig. A1.5, Table A1.4). Each of these were unambiguous and supported by clean Sanger sequences in both directions and the heterozygous positions were confirmed in each of these individuals in independent DNA extractions. Unexpectedly, all but one of the six of the *X. victorianus* individuals with heterozygous sites in *dm-w* were male, (the phenotypic sex of these samples was not evaluated by Furman et al. (2015)). One of these heterozygous positions was a synonymous variant in *dm-w* exon 4 of three individuals, and the other ten heterozygous positions were in the untranscribed region between *dm-w* and *scan-w* in one or more individuals. These heterozygous positions imply that the father and the mother of several *X. victorianus* individuals both carried *dm-w*.

In *X. itombwensis* individuals, we succeeded in amplifying *dm-w* exons 2 and 3 in all the tested individuals (five out of five females and 20 out of 20 males) (Table 2.1, Fig. A1.3). Sanger sequences from five female and five male individuals (exon 3) and all individuals (exon 2) confirmed that these amplifications were *dm-w*. These Sanger sequences for exon 2 and 3 revealed several heterozygous positions in synonymous exonic positions and intronic positions in several individuals of each sex (Fig. A1.4, Table A1.4). No evidence of pseudogenization in the form of truncation mutations was detected in this 5' coding region of *dm-w*, which includes the main functional element of this gene, the DM-domain (Yoshimoto et al. 2008). These results suggest that *dm-w* is autosomal and fixed, or almost fixed, in *X. itombwensis*.

In *X. pygmaeus* individuals, exons 2 and 3 of *dm-w* both amplified in six out of 9 females and two out of 11 males, but neither exon amplified in any of the other samples (Table 2.1, Fig. A1.6). That each of these amplifications was *dm-w* was confirmed with Sanger sequencing. We did not rule out the possibility that two independent failed amplifications in some *X. pygmaeus* samples was due to linked polymorphisms in regions that were independently targeted by each of the primer pairs. This caveat notwithstanding, the independent failures of two *dm-w* amplifications in several individuals of each sex opens the possibility that *dm-w* is not required for female differentiation in *X.*

pygmaeus. The detection of *dm-w* in a male demonstrates it is not completely sex linked in this species. Using primers from Bewick et al. 2011, we also successfully obtained sequences of exon 4 for all the *X. pygmaeus* individuals for which we obtained exons 2 and 3, except for 1 female individual. In *dm-w* exon 4 of *X. pygmaeus*, a stop codon was present 81 bp earlier than the one in *X. laevis*, but the functional significance of this remains unclear.

For *X. clivii*, we used Sanger sequencing of *dm-w* in a family of *X. clivii* that we generated in our lab. *X. clivii* is the most distantly related species to *X. laevis* that carries *dm-w* (Fig. 2.1; Bewick et al. 2011; Furman and Evans 2016). Similar to the results from *X. pygmaeus*, *dm-w* was identified in both sexes but not fixed in either one (Table 2.1, Fig. A1.7). We were able to independently amplify and sequence exons 2 and 3 in two out of two females (including the mother) of our lab cross and four out of 12 sons. We were not able to amplify either *dm-w* exon in the father or any of the 10 other sons. We also tried to amplify these exons 2 and 3 in two adult females and succeeded for both exons in one of them, but neither exon in the other.

In the *X. clivii* family, we can rule out effects that polymorphism at primer sites might have because, barring *de novo* mutations, the parental alleles should match the offspring alleles. One possibility is that the father had a *dm-w* allele that could not be amplified, but (similar to the failed amplifications in *X. pygmaeus*) this would have to be a null amplification for two independent primer sets in the same individuals. A previous survey identified *dm-w* in eight of 12 wild-caught females but zero of 12 males (Furman and Evans 2016), suggesting that this gene generally occurs more frequently in females in nature, even though this was not the case in the one laboratory family we studied.

A2.3 A male heterogametic sex-linked region of *H. boettgeri* is homologous to *X. tropicalis* chromosome 4

In the *H. boettgeri* family, we identified 54 single nucleotide polymorphisms (SNPs) in the RRGS data that had a sex-linked inheritance pattern that were distributed among 51 scaffolds of our draft genome assembly, as compared to 1361 SNPs total that were detected throughout the genome. All 54 sex-linked SNPs matched expectations for male heterogamy. Of these 54 SNPs, 36 had a Y-linked pattern of inheritance in a male heterogametic (XY) system and were "almost" sex linked in that they were heterozygous in the father and at least 80% of the sons, but homozygous in the mother and all the daughters. A lack of heterozygosity in some sons could be either a genotyping error, or indicative of recombination between the X and Y chromosome during spermatogenesis. And 17 of the 54 SNPs had a X-linked pattern of inheritance in a male heterogametic system and were heterozygous in the father and at least 80% of the daughters, but homozygous in the mother and at least 80% of the sons (one SNP was completely sex-linked). Here, homozygous genotypes in some daughters could be due to genotyping error or reflect recombination between the X and Y chromosome during spermatogenesis. For 1 of the 54 SNPs, all the daughters were heterozygous (A/C) and at least 80% of the

sons were homozygous (A/A). However both parents were homozygous (mother: A/A, father: C/C). This pattern is most likely due to an undercalled heterozygous site in the father and compatible with a SNP on the X chromosome of the father. The number of XY sites that we found is significantly higher than the expectation we obtained from our permutation test, whereas the number of ZW sites fell within the null distribution of these permutations (Fig. A1.1).

Of 51 *H. boettgeri* scaffolds that carried putatively male heterogametic sex-linked SNPs, 11 had a positive BLAST hit to the *X. tropicalis* and *X. laevis* genome assemblies (defined as a BLAST expect value below 1e-5). All of these scaffolds matched chromosome 4 of *X. tropicalis* and 4L or 4S of the *X. laevis* genome assembly.

To test whether genotype error (e.g., "under-called" heterozygous positions that were incorrectly called as homozygotes) caused some of these sites to be "almost" as opposed to "completely" sex-linked, we attempted to Sanger sequence intronic regions of *hmcn1* (∼275 bp of intron 32) and *sall1* (∼260 bp of intron 1) in the *H. boettgeri* family. Both of these regions were initially identified in the RRGS data because they each contained one site with an almost sex-linked pattern of inheritance, which in both cases was heterozygous in the father and all but one son and homozygous in the mother and all daughters. Genotypes in Sanger sequences were in fact completely sex-linked, supporting our suspicion that genotyping error was present in the RRGS data. For *sall1* we obtained Sanger sequences for the entire family; for *hmcn1*, we obtained Sanger sequences for the entire family except one son. However, this son with missing Sanger data was heterozygous in the RRGS data, as expected for a completely sex-linked SNP.

To further assess the hypothesis that the sex-linked region in *H. boettgeri* is homologous to a large portion of chromosome 4 of *X. tropicalis*, we also Sanger sequenced genomic regions that were not covered by the RRGS data. We focused on two exons of the gene *dmrt5*, which is located on chromosome 4 in *X. tropicalis*.

In *dmrt5* exon3-intron3, four SNPs were identified that had a completely Y-linked inheritance pattern (the father and sons were heterozygous, the mother and daughters were homozygous; sequences included both parents, 20 sons, and 8 daughters). In *dmrt5* exon2-intron2, one SNP was identified with a completely Y-linked inheritance pattern (sequences included both parents, 15 sons, 5 daughters).

A2.4 A male heterogametic sex-linked region in *P. parva* is homologous to *X. tropicalis* chromosome 6

In the *P. parva* family, we identified 19 SNPs in the RRGS data that had a sex-linked inheritance pattern that were distributed among 18 scaffolds in our draft genome assembly, out of 564 SNPs total throughout the genome. Four of these 18 scaffolds matched a large scaffold of the *X. tropicalis* genome and all of these were to chromosome 6. All of these SNPs matched expectations for male heterogamy, and none matched expectations for female heterogamy. Nine of the 19 SNPs were completely sex-linked with a Y-linked

mode of inheritance (heterozygous in the father and all sons, homozygous in the mother and all daughters). 10 of the 19 SNPs were almost sex linked with a Y-linked mode of inheritance (heterozygous in the father and all but one son, homozygous in the mother and all daughters).

No sites were detected with a mode of inheritance from the paternal X (with the father and all daughters heterozygous and the mother and all sons homozygous). This could be a consequence of inbreeding in the ancestry of the captive-bred parents (because the father appears to carry an X chromosome that is very similar to at least one of those of the mother). Similar to *H. boettgeri*, the number of XY sites we found is significantly different from what we expected to obtain by chance (Fig. A1.1).

In the *P. parva* family, we amplified and Sanger sequenced portions of three genes on *X. tropicalis* chromosome 6, including: intronic and exonic region of *ncoa2* (exon 2, partial introns 1 and 2, ∼800bp), *mmp16* (partial exon 9 and UTR, 1kb, 460bp with seminested primer) and *kctd1* (exon4, partial introns 3 and 4, ∼700bp) for the whole family. Each of these regions contained at least one (two for *ncoa2*), SNP(s) with a completely Y-linked pattern of inheritance (heterozygous in the father and all sons, homozygous in the mother and all daughters).

A2.5 A female heterogametic sex-linked region of *X. mellotropicalis* is homologous to *X. tropicalis* chromosome 7

In the *X. mellotropicalis* family, we identified 3 SNPs in the RRGS data that had a sex-linked inheritance pattern out of 50,428 SNPs total that were detected throughout the genome. Each was on a different scaffold of the draft *X. mellotropicalis* genome sequence. Two SNPs matched expectations for W-linked female heterogamy (mother and all daughters heterozygous, father and >80% of sons homozygous). One of these three SNPs matched expectations for maternal Z-linkage (heterozygous in the mother and all the sons, homozygous in the father and all but one of the daughters). One of these *X. mellotropicalis* scaffolds matched the end of the short arm of chromosome 7 of *X. tropicalis*, another one matched chromosome 8, and the third had no close match.

To further explore the sex-linked region of *X. mellotropicalis*, we amplified and Sanger sequenced three genomic regions that are homologous to *X. tropicalis* chromosome 7: a portion of intron 1 of both *X. mellotropicalis* homeologs of *cyp17a1* (∼310-360bp), a portion of exon 1 of one *X. mellotropicalis* homeolog of *moxd2p* (∼315bp), and a portion of the coding region of both *X. mellotropicalis* homeologs of *or8h1* (∼400-450bp). No polymorphism was observed in the *cyp17a1* homeologs. 15 SNPs were identified in the *moxd2p* homeolog but none were sex-linked, suggesting that these sequences might have been from the chromosome that is homeologous to the sex chromosome of *X. mellotropicalis*. The *or8h1* homeolog had an almost sex-specific amplification wherein there was a strong amplification in all females and one male and a weak or absent amplification in the other males (Fig. A1.2). This pattern could arise if there were a partially sex-linked SNP in a genomic region spanned by one or both of the primers.

Because our assembly is highly fragmented, we were unable to distinguish contigs from each of the *X. mellotropicalis* subgenomes. However, we amplified both paralogs of *or8h1*, and one has an almost sex-specific amplification. Phylogenetic estimation of these homeologous data and homologous regions of *X. tropicalis* and *X. laevis* suggests that the sex-chromosome in *X. mellotropicalis* is not in the *X. mellotropicalis* subgenome that is most closely related to *X. tropicalis* (data not shown).

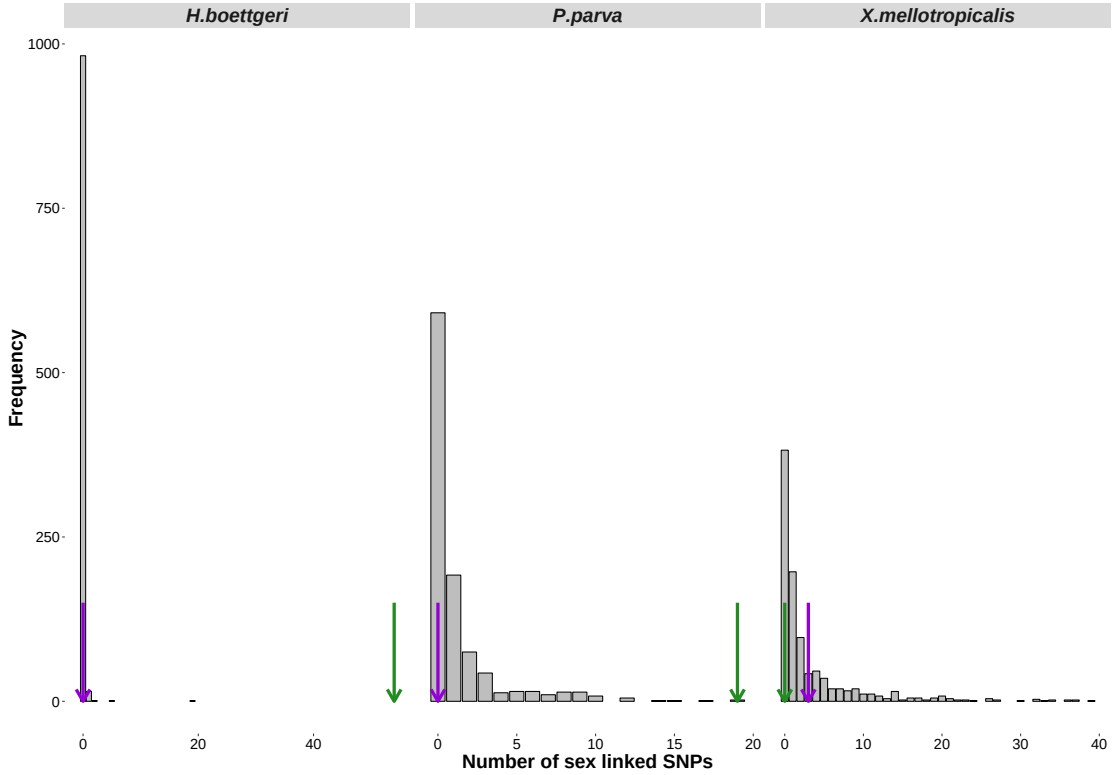


Figure A1.1: Results of the permutation test. The gray bars correspond to the distribution of sex-linked SNPs obtained when the sex of the offspring is permutated a thousand times. The green and purple arrows correspond respectively to the observed number of XY and ZW SNPs that were observed in the data of *H. boettgeri*, *P. parva*, and *X. mellotropicalis*.

A2.6 Variation among pipid species in the extent of sex linkage over one generation

Our analysis identified more sex-linked SNPs that map to a larger genomic region in *H. boettgeri* than in *P. parva* and *X. mellotropicalis* (Fig. 2.2). These results tend to suggest a larger sex-linked region in *H. boettgeri*. We found the fewest number of sex-linked SNPs for *X. mellotropicalis*. This suggests, similar to what has been found for the closely related diploid species *X. tropicalis* (Bewick et al. 2013), a large pseudoautosomal region in this species.

Some aspects of these results might be associated with methodological limitations. For *X. mellotropicalis*, our draft genome probably contains chimerical scaffolds from both subgenomes because, unlike the *X. laevis* genome, both subgenomes from *X. mellotropicalis* are not highly diverged. Co-mapping of RRGs to these chimerical contigs could mask sex-linked SNPs.

Differences in the RRGS coverage and offspring number could affect our inferences about the sizes of regions of recombination suppression in each of the three pipid families. A higher number of offspring increases the chance of detecting recombination events and thus should be associated with smaller regions of recombination suppression, all else being equal. The identification and extent of the sex-determining regions for each species were also influenced by how well our scaffolds mapped to the *X. tropicalis* reference genome (i.e., well for *X. mellotropicalis*, but less so for *H. boettgeri* or *P. parva*).

Table A1.2: Proportions of scaffolds containing sex-linked and not sex-linked SNPs by chromosomal location in our *H. boettgeri*, *P. parva* and *X. mellotropicalis* families. Scaffolds are from our draft genome assemblies that map respectively either to the identified sex-chromosome (numbered according to its homolog in *X. tropicalis*), to other chromosomes (which are autosomes), or to unplaced scaffolds in the *X. tropicalis* genome assembly (xenbase v. 9.1).

	# Sex-linked / # Not sex-linked scaffolds		
Species	Sex-chromosomes	Other Chromosomes	Scaffolds
P. parva	Chr.06: 3/4	0/61	0/5
H. boettgeri	Chr.04: 11/13	0/82	0/3
X. mellotropicalis	Chr.07: 1/764	1/8748	0/322

TABLE A1.3: Primer sequences used in this study. Forward and reverse sequences are given in 5' to 3' orientation.

Species	Gene	Primer Name	Forward/reverse Sequences
H. boettgeri	*dmrt5/dmrta2*	Hc_DMRT5_e3_F1	AGGACCTTTGTTCTGGTGGA
		Hc_DMRT5_e3_R1	CCCAATGCCACACACTTTGG
H. boettgeri	*dmrt5/dmrta2*	Hc_DMRT5_e2_F1	CTTCCACCTTCCCTACCCAAC
		Hc_DMRT5_e2_R1	CAGTTGCTACTTCGTGCAGC
H. boettgeri	*sall1*	Hc_Sall1_i1_F1	GAACTAAATGAAACCCGCCCC
		Hc_Sall1_i1_R1	CGGAGAAACACGGACAACAA
		Hc_Sall1_i1_R4	CCCCCATTACTGATCCYGTGT
H. boettgeri	*hmcn1*	Hc_hmcn1_i32_F1	CAACTAATGCTGCTGGGGAAG
		Hc_hmcn1_i32_R1	TGACCGAAAGGTGTAACCTGC
P. parva	*ncoa2*	Pp_NCOA2_e2_F1	TGTAAAATCGCAACCACGACA
		Pp_NCOA2_e2_R1	TGGCACTGATGAAAAAGCACTG
P. parva	*mmp16*	Pp_mmp16_e1_F1	AGTCGCTCCCACTGTCATTTT
		Pp_mmp16_e1_R1	TGGACAACACAGCAAGCACT
		Pp_mmp16_e1_R2	TTTTCCCATCAAACAAGGGGC
P. parva	*kctd1*	Pp_kctd1_e1_F1	ATGTAGTGGTTGCCTTGTGCC
		Pp_kctd1_e1_R1	GCTCCAGTCCCAAGTCATAGT
X. mellotropicalis	-	Xm_sc_159590_F1	GGACCAATAGGGTTTCGCTTC
		Xm_sc_159590_R1	GACATAGGGCAAAATCCTGCT
X. mellotropicalis	*or8h1*	Xm_or8h1_p1_F1	GTCATATTACAGATAATTCC
		Xm_or8h1_p1_R1	CAACAAAGCAATCATACGCCATG
		Xm_or8h1_p1_F3	GTATCTCTGGGTTCAGTTTATAAGC
		Xm_or8h1_p1_R2	CAATATGATAATTTAGATATATGTGC
X. mellotropicalis	*or8h1*	Xm_or8h1_p2_F1	GTCATATTACAGATAATTCC
		Xm_or8h1_p2_R1	CAACAAAGCAATCATACGCCATG
X. itombwensis/X. pygmaeus	*dm-w*	DMW_Xit_e3_F1	CACTGTATGTGGCAGCTGAAG
		DMW_Xit_e3_R1	GATAGGGTGGCTCTTGGTCA
X. itombwensis/X. pygmaeus	*dm-w*	DMW_Xit_e2_F1	CATTGTTTCCTGTTCTGAAAATGT
		DMW_Xit_e2_R2_l	TGCAGTATGAATGGTGCTGTT
		DMW_Xit_e2_R1_s	CTGCCCGGTGTTATATGGTT
X. clivii	*dm-w*	19Xc_DMW_ex2_f2b	GAACCATATAACACCGGGCAGT
		19Xc_DMW_ex2_r2	CCACGAAAAAATGCAGTAAATGCT
X. clivii	*dm-w*	2019cliv_e3_DMW_f1	AAGGGATCATAACTTTTTAAAACACAC
		2019cliv_e3_DMW_r1	CCTGAACTGGTTGTGGAACTG
		2019cliv_e3_DMW_f2	AGGGATCATAACTTTTTAAAACACAC
		2019cliv_e3_DMW_r2	TGAACTGGTTGTGGAACTGAA

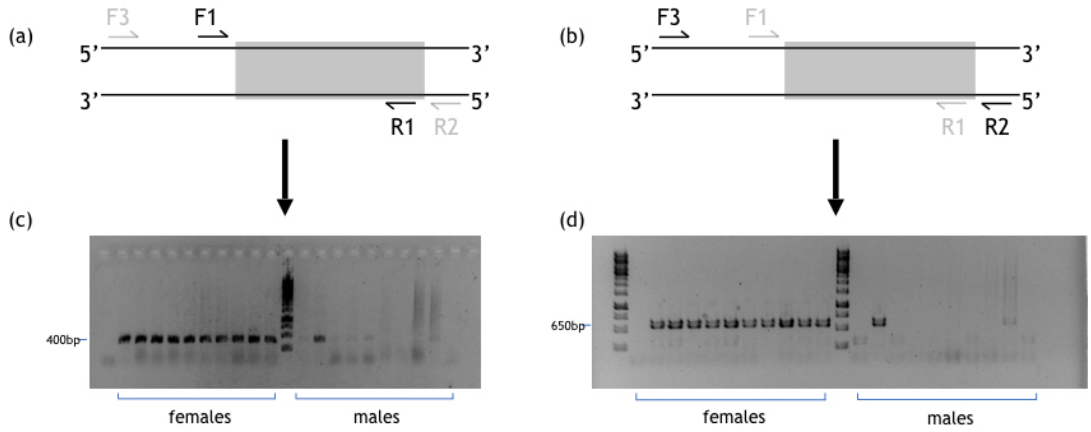


Figure A1.2: Amplification of a *or8h1* homeolog in the *X. mellotropicalis* family with nested primers Xm_or8h1_p1_F1/R1 (left) and Xm_or8h1_p1_F3/R2 (right). Dark font in (a) and (b) indicates locations of internal and external primer pairs respectively and the grey area represents the coding region. (c) and (d) show the resulting amplifications. The first well is a negative control and the first female and male individuals are the mother and the father of the cross, respectively.

A2.7 Amplification of dm-w

Table A1.4: Samples used in *dm-w* assay including Museum ID, Field ID, phenotypic sex (Sex), and whether *dm-w* was detected (Y) or not (N). Y* indicates that *dm-w* sequences had heterozygous sites. Only samples for which a positive control worked are reported.

Museum ID	Field ID	Sex	*dm-w*
X. gilli	13 females - 7 males		
no voucher	xm2-0	F	Y
no voucher	xr3-2	F	Y
no voucher	xr16	F	Y
no voucher	xr0-3	F	Y
no voucher	xr1-1	F	Y
no voucher	xr1-2	F	Y
no voucher	xr2-1	F	Y

Continued on next page

Continued from previous page

Museum ID	Field ID	Sex	DMW
no voucher	km1	F	Y
no voucher	km2	F	Y
no voucher	xr2-2	F	Y
no voucher	xm0-2	F	Y
no voucher	xr3-1	F	Y
no voucher	xm0-1	F	Y
no voucher	xr1-3	M	N
no voucher	xr3-3	M	N
no voucher	xr15	M	N
no voucher	xr2-12	M	N
no voucher	xr2-3	M	N
no voucher	xs91	M	N
no voucher	xs92	M	N
X. laevis	24 females - 12 males		
MCZ A-149234	BJE3525	F	Y
MCZ A-149235	BJE3526	F	Y
MCZ A-149236	BJE3527	F	Y
MCZ A-149243	BJE3534	F	Y
MCZ A-149244	BJE3535	F	Y
MCZ A-149245	BJE3536	F	Y
MCZ A-149247	BJE3538	F	Y
MCZ A-149254	BJE3545	F	Y
MCZ A-149256	BJE3547	F	Y
MCZ A-149257	BJE3548	F	Y
MCZ A-149258	BJE3549	F	Y
MCZ A-149259	BJE3572	F	Y
MCZ A-149260	BJE3573	F	Y
MCZ A-149261	BJE3574	F	Y
MCZ A-149262	BJE3575	F	Y
MCZ A-149264	BJE3609	F	Y
MCZ A-149265	BJE3639	F	Y
MCZ A-149266	BJE3640	F	Y
MCZ A-149272	BJE3646	F	Y
MCZ A-149274	BJE3648	F	Y
MCZ A-149275	BJE3649	F	Y
MCZ A-149276	BJE3650	F	Y
MCZ A-149269	BJE3643	F	Y
MCZ A-149246	BJE3537	F	Y
MCZ A-149237	BJE3528	M	N
MCZ A-149238	BJE3529	M	N
MCZ A-149239	BJE3530	M	N
MCZ A-149240	BJE3531	M	N
MCZ A-149241	BJE3532	M	N
MCZ A-149242	BJE3533	M	N
MCZ A-149253	BJE3544	M	N

Continued on next page

Continued from previous page

Museum ID	Field ID	Sex	DMW
MCZ A-149255	BJE3546	M	N
MCZ A-149263	BJE3608	M	N
MCZ A-149267	BJE3641	M	N
MCZ A-149268	BJE3642	M	N
MCZ A-149277	BJE3651	M	N
X. victorianus	20 females - 15 males		
no voucher	BJE1489	F	Y
MCZ A-138177	BJE260	F	Y
MCZ A-138179	BJE262	F	Y
MCZ A-138180	BJE263	F	Y
MCZ A-138183	BJE266	F	Y
MCZ A-138184	BJE267	F	Y
UTEP 21051	CFS1091	F	Y
no voucher	EBG1168	F	Y
UTEP 20166	EBG2148	F	Y
UTEP 20167	EBG2152	F	Y
UTEP 20168	EBG2170	F	Y
UTEP 20171	EBG2342	F	Y
UTEP 20174	EBG2463	F	Y*
UTEP 20175	EBG2464	F	Y
UTEP 21052	ELI1369	F	Y
UTEP 21053	ELI1370	F	Y
UTEP 21056	ELI1461	F	Y
UTEP 21057	ELI1462	F	Y
UTEP 21055	ELI503	F	Y
UTEP 21049	ELI527	F	Y
no voucher	BJE1488	M	N
MCZ A-138178	BJE261	M	N
MCZ A-138181	BJE264	M	N
MCZ A-138182	BJE265	M	N
MCZ A-140118	BJE2897	M	Y*
MCZ A-140119	BJE2898	M	Y*
MCZ A-140121	BJE2900	M	Y*
MCZ A-140123	BJE2902	M	Y*
UTEP 21050	CFS1090	M	N
UTEP 20165	EBG2147	M	N
UTEP 20169	EBG2171	M	N
UTEP 20170	EBG2329	M	Y*
UTEP 21035	ELI1142	M	N
UTEP 21054	ELI502	M	N
UTEP 21048	ELI526	M	N
X. itombwensis	5 females - 20 males		
no voucher	BJE3852	F	Y*
no voucher	Itom#3	F	Y*
no voucher	Itom#5	F	Y*

Continued on next page

Continued from previous page

Museum ID	Field ID	Sex	DMW
no voucher	Itom#8	F	Y*
no voucher	Itom#16	F	Y
no voucher	BJE3376	M	Y
no voucher	BJE4315	M	Y*
no voucher	BJE4316	M	Y*
no voucher	BJE4317	M	Y*
no voucher	BJE4318	M	Y*
no voucher	BJE4319	M	Y
no voucher	Itom#1	M	Y*
no voucher	Itom#2	M	Y
no voucher	itom#6	M	Y
no voucher	Itom#7	M	Y*
no voucher	Itom#9	M	Y*
no voucher	Itom#10	M	Y
no voucher	Itom#11	M	Y
no voucher	Itom#12	M	Y
no voucher	Itom#13	M	Y*
no voucher	Itom#14	M	Y
no voucher	Itom#15	M	Y
no voucher	Itom#17	M	Y
no voucher	Itom#18	M	Y*
MCZ A-138192	BJE00275	M	Y
X. pygmaeus	9 females -11 males		
UTEP 22019	ELI2150	F	N
UTEP 22021	ELI2074	F	N
UTEP 22022	ELI2081	F	N
UTEP 22023	ELI1726	F	Y
UTEP 22024	ELI3012	F	Y
UTEP 22026	ELI1682	F	Y
UTEP 22029	ELI2372	F	Y
UTEP 22030	ELI2880	F	Y
no voucher	Pygm#1	F	Y
UTEP 22028	ELI2371	M	N
UTEP 22031	ELI3243	M	Y
UTEP 22032	ELI2526	M	N
UTEP 22033	ELI2545	M	N
UTEP 22027	ELI2370	M	N
UTEP 22025	ELI1681	M	N
UTEP 22034	ELI2347	M	N
no voucher	Pygm#2	M	N
no voucher	Pygm#3	M	N
no voucher	Pygm#4	M	N
Voucher lost (MNHG)	AMNH17320	M	Y
X. clivii	16 females - 29 males		
MCZ-FS-A-30575/Cryogenic 380	BJE1552	F	Y

Continued on next page

Continued from previous page

Museum ID	Field ID	Sex	DMW
MCZ-FS-A-30576/Cryogenic 381	BJE1553	F	Y
MCZ-FS-A-30573/Cryogenic 378	BJE1550	F	N
MCZ FS Z-23323/Cryogenic 449	z-23323	F	N
MCZ-FS-A-30572/Cryogenic 377	BJE1549	F	N
MCZ-FS-A-30526/Cryogenic 331	BJE1503	F	Y
MCZ-FS-A-30527/Cryogenic 332	BJE1504	F	Y
MCZ-FS-A-30547/Cryogenic 352	BJE1524	F	Y
MCZ-FS-A-30548/Cryogenic 353	BJE1525	F	Y
MCZ-FS-A-30549/Cryogenic 354	BJE1526	F	Y
no voucher	BJE1510	F	N
no voucher	BJE1511	F	Y
no voucher	BJE4530	F	Y
no voucher	XCL02	F	Y
no voucher	XCL13	F	Y
no voucher	XCL14	F	N
MCZ-FS-A-30574/Cryogenic 379	BJE1551	M	N
MCZ-FS-A-30577/Cryogenic 382	BJE1554	M	N
MCZ FS Z-23319/Cryogenic 445	z-23319	M	N
MCZ FS Z-23320/Cryogenic 446	z-23320	M	N
MCZ FS Z-23321/Cryogenic 447	z-23321	M	N
MCZ FS Z-23322/Cryogenic 448	z-23322	M	N
MCZ-FS-A-30545/Cryogenic 350	BJE1522	M	N
MCZ-FS-A-30546/Cryogenic 351	BJE1523	M	N
no voucher	BJE1512	M	N
no voucher	BJE1513	M	N
no voucher	BJE1514	M	N
no voucher	BJE1515	M	N
no voucher	BJE1516	M	N
no voucher	BJE4531	M	N
no voucher	XCL01	M	N
no voucher	XCL03	M	Y
no voucher	XCL04	M	Y
no voucher	XCL05	M	N
no voucher	XCL06	M	N
no voucher	XCL07	M	N
no voucher	XCL08	M	N
no voucher	XCL09	M	N
no voucher	XCL10	M	N
no voucher	XCL11	M	N
no voucher	XCL12	M	N
no voucher	XCL15	M	Y
no voucher	XCL16	M	Y
no voucher	XCL17	M	N
no voucher	XCL18	M	N

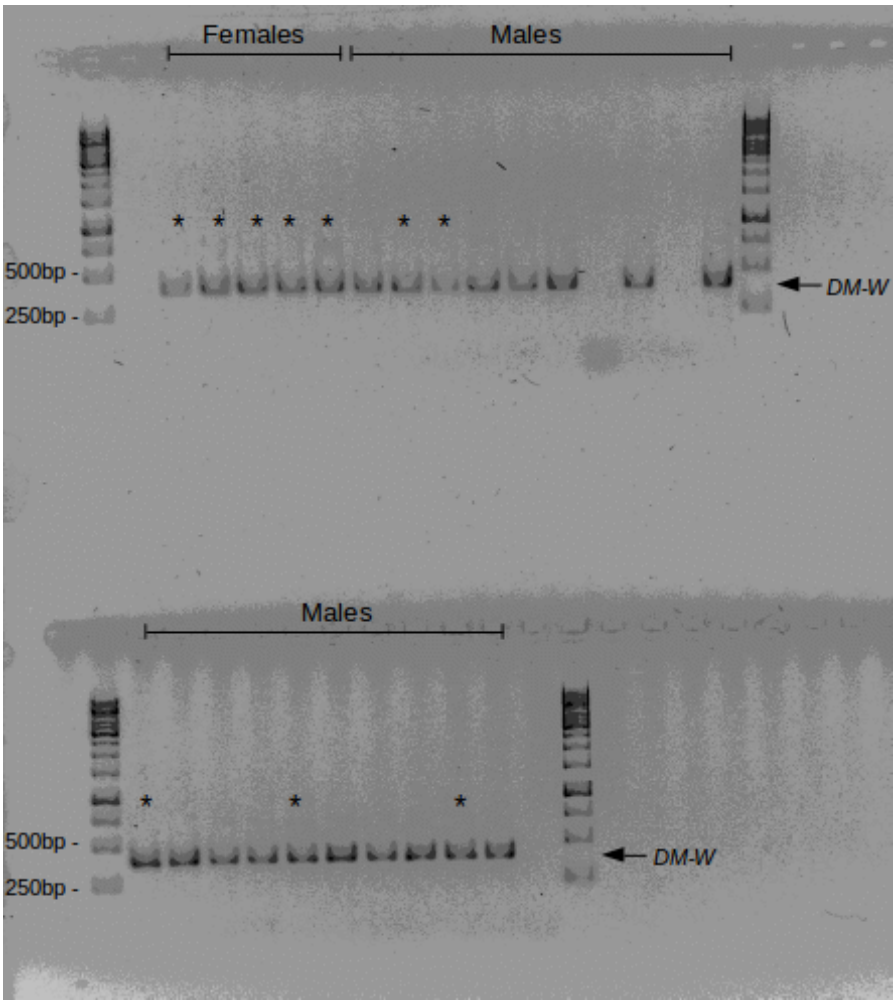
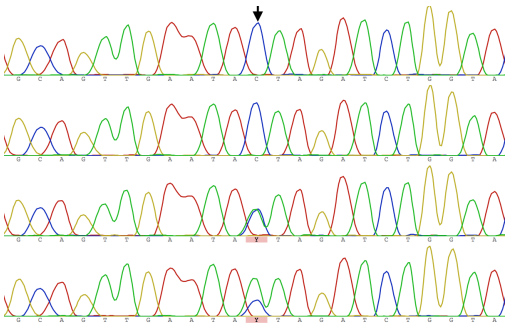
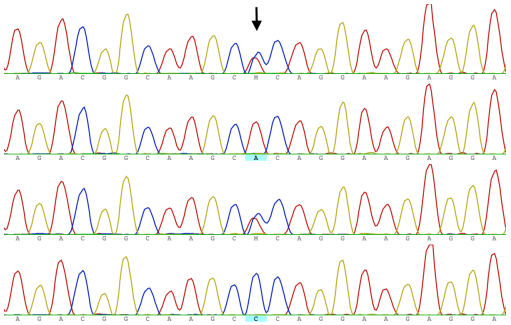


Figure A1.3: Amplifications of *dm-w* exon 3 and exon 2 (not shown) were successful in male and female *X. itombwensis*. Amplifications are flanked by a 1 Kb ladder; the first well before the female individuals and the last well after the male individuals are negative controls. Asterisks (*) represent amplifications that were sequenced. Two males did not amplify in this PCR reaction; one amplified in a separate temperature gradient PCR reaction for exons 2 and 3; these exons and the positive control did not amplify for the other individual, and it was thus not included in Table 2.1.



(A) From 37 bp to 59 bp



(B) From 114 bp to 136 bp

FIGURE A1.4: Sanger sequences from some *X. itombwensis dm-w* have heterozygous sites that are not sex-linked. The first sequence is from a female and the other sequences are from males; different regions of *dm-w* are depicted in (a) and (b). The number of base pairs is given relative to the first sequence (Genbank: MK907537). Black arrows highlight sites that are heterozygous in some individuals and homozygous in others. A forward read for each individual is shown here; the heterozygous sites are also clearly visible in the reverse reads (not shown).

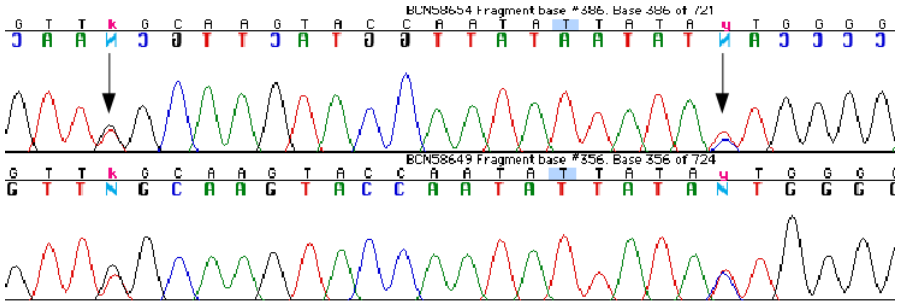


Figure A1.5: Sanger sequences from *X. victorianus* 5' flanking region of *dm-w* have heterozygous sites in some individuals. Sanger sequence of one male (BJE2902) in forward and reverse direction is shown. Black arrows highlight sites that are heterozygous in this individual.

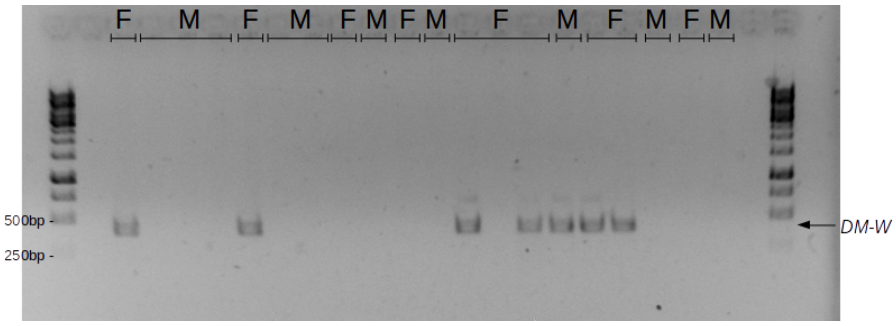


Figure A1.6: Amplifications of *dm-w* exon 3 and exon 2 (not shown) were successful in some male (M) and female (F) *X. pygmaeus* individuals. Amplifications are flanked by a 1 Kb ladder and the first and last wells are negative controls. The positive control (not shown) failed for one of the females and it was thus not included in Table 2.1. Successful amplifications were sequenced to confirm that they were *dm-w*.

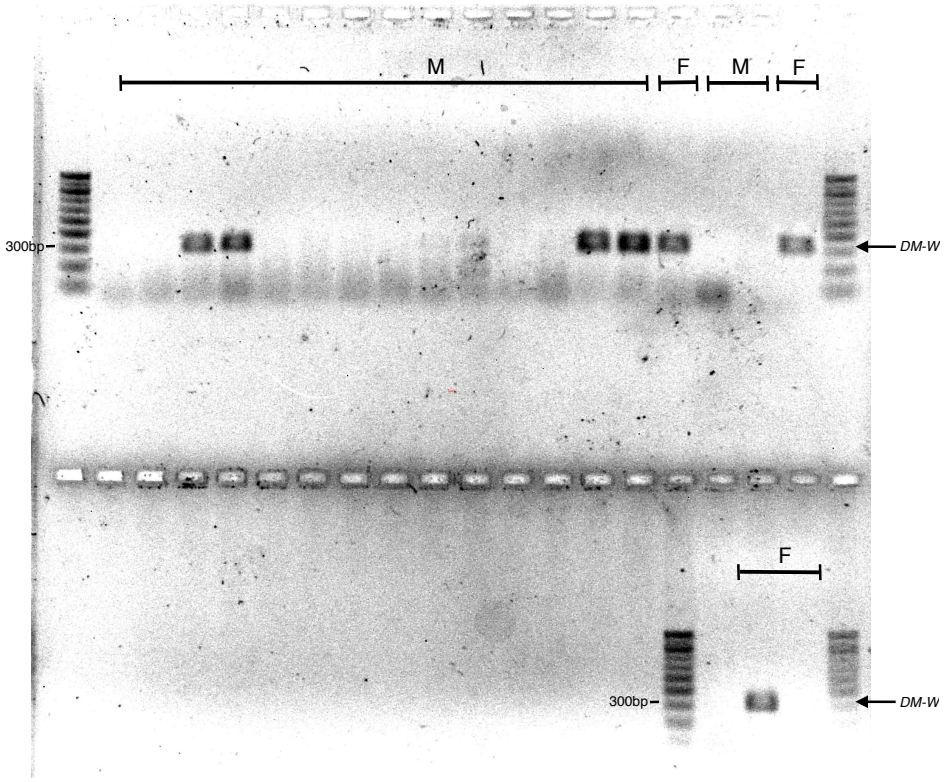


Figure A1.7: Amplifications of *dm-w* exon 2 and exon 3 (not shown) were successful in some male ("M") and female ("F") *X. clivii* individuals. Amplifications from a subset of individuals from Table 2.1 are flanked by a 100 bp ladder; the first well on the second row is a negative control. Successful amplifications were sequenced to confirm that they were *dm-w*.

A2.8 Data availability

Sequence data from this study are deposited in the Sequence Read Archive of the National Center for Biotechnology Information (NCBI), Bioproject PRJNA549161, including WGS and RRGS data. Assembled sequences from targeted next generation sequencing and Sanger sequencing are available in GenBank; accession numbers are in Table S5. Alignments for the mitochondrial, *dm-w*/*dmrt1* exons 2 and 3 analysis are available as Supplementary Files (respectively Files S1, S2 and S3).

A2.9 Repetitive elements in three pipid genomes

The composition of different types of repetitive elements was similar within species but different between species (Table A1.6). Overall, 25-29, 23, and 38% of the genomes of *H. boettgeri*, *P. parva*, and *X. mellotropicalis*, respectively, were repetitive elements. Thus, the *X. mellotropicalis* genome had a considerably higher proportion of repetitive elements than in the other two species. DNA transposons were the most abundant TE category (about 31.5% of the total identified

Table A1.5: Genbank accession numbers.

Loci	Species	Genbank Accession number	Method
dmrt1	*X. borealis*	MK907575-MK907576	RNAseq
dm-w exon 2	*X. itombwensis*	MK907543-MK907558	Sanger Sequencing (PCR)
dm-w exon 3	*X. itombwensis*	MK907533-MK907542	Sanger Sequencing (PCR)
dm-w exon 4	*X. pygmaeus*	MN365719	Sanger Sequencing (PCR)
dmrt1/dm-w exon 2	*Xenopus* sp.	MN030659-MN030798	Capture data
dmrt1/dm-w exon 3	*Xenopus* sp.	MN030799-MN030916	Capture data
16S ribosomal RNA	*P. parva* & *H. boettgeri*	MK907573-MK907574	Sanger Sequencing (PCR)
sall1	*H. boettgeri*	MK907563-MK907564	Sanger Sequencing (PCR)
dmrta2	*H. boettgeri*	MK907559-MK907560	Sanger Sequencing (PCR)
hmcn1	*H. boettgeri*	MK907561-MK907562	Sanger Sequencing (PCR)
ncoa2	*P. parva*	MK907565-MK907566	Sanger Sequencing (PCR)
kctd1	*P. parva*	MK907567-MK907568	Sanger Sequencing (PCR)
mmp16	*P. parva*	MK907569-MK907571	Sanger Sequencing (PCR)
or8h1	*X. mellotropicalis*	MK907572	Sanger Sequencing (PCR)

transposable elements for *X. mellotropicalis* and 40% for the other species). DNA transposons are also the most abundant TE class in *X. tropicalis* (Chalopin et al. 2015).

Table A1.6: Transposable element distributions. Repeat elements were identified using RepARK v. 1.3.0 (Koch et al. 2014) and classified using TEclass v. 2.1.3 (Abrusán et al. 2009).

Species	Sex	Repeat class	Count
H. boettgeri	Female (BJE3814)	DNA transposons	86784
		LTRs	46282
		LINEs	8506
		SINEs	15300
		Unclear	54341
		Total	211213
H. boettgeri	Male (BJE3815)	DNA transposons	56381
		LTRs	28799
		LINEs	5478
		SINEs	9257
		Unclear	33655
		Total	133570
P. parva	Female (BJE4294)	DNA transposons	39571
		LTRs	20028
		LINEs	3850
		SINEs	6493
		Unclear	23258
		Total	93200
P. parva	Male (BJE4295)	DNA transposons	39706
		LTRs	19981
		LINEs	3979
		SINEs	6563
		Unclear	23219
		Total	93448
X. mellotropicalis	Female (BJE3652)	DNA transposons	287594
		LTRs	222851
		LINEs	35790
		SINEs	98484
		Unclear	267515
		Total	912234

TABLE A1.7: Potential sex-related genes on the sex chromosomes of *H. boettgeri*, and *P. parva* are based on the Gene Ontology database (http://www.geneontology.org/, Consortium 2004) and on previous research on other organisms. *X. tropicalis* column corresponds to the location on *X. tropicalis* reference genome (v.9.1 on Xenbase).

Goterm	Gene abbreviation	Gene name	*X. tropicalis*	References
Sex	*sycp3*	synaptonemal complex protein 3	Chr04:2810834..2822710	
	LOC100491886/*astl2g*	astacin-like metallo-endopeptidase 2 gene g	Chr04:6113566..6123471	
	fshb	follicle stimulating hormone subunit beta	Chr04:10709930..10712426	
	lgr4	leucine-rich repeat containing G protein-coupled receptor 4	Chr04:11872155..11953020	
	irx5	iroquois homeobox 5	Chr04:44900594..44904952	
	cfap157	cilia and flagella associated protein 157	Chr04:67262199..67272088	
	brdt	bromodomain testis-specific	Chr04:94474641..94523774	
	asxl3	ASXL transcriptional regulator 3	Chr06:97487242..97558979	
	mybl1	MYB proto-oncogene like 1	Chr06:104080142..104115776	
No Goterm	*wt1*	Wilms tumor protein	Chr04:9611745..9651921	1
	dmrt5/dmrta2	DMRT like family A2	Chr04:85757614..85763009	2

[1] Miyamoto et al. 2008; Gammerdinger et al. 2016
[2] Shirak et al. 2006

Appendix B

Supplemental information: Evolution and Function of a (Small) Sex Determining Supergene in a Frog (*X. laevis*)

A1 Amplification of *dm-w*

Previously, females and males were shown to carry *dm-w* in *X. victorianus* (Cauret et al. 2020). Here, we re-explored *X. victorianus* data in a geographical context and expanded this survey to include additional *Xenopus* species (*X. petersii*, *X. poweri*, *X. kobeli*, *X. andrei*, and *X. largeni*). Samples used in these surveys are listed in Table A2.1.

Table A2.1: Samples used in *dm-w* assay including Museum ID, Field ID, phenotypic sex (Sex), and whether each of *dm-w* was detected (Y) or not (N). Asterisks indicate that primers for a non-coding region (Yoshimoto et al. 2008) instead of a coding region were used to infer the presence / absence of *dm-w*. Question marks indicate amplifications that either were not attempted or that produced a poor Sanger sequence. All samples reported also had a sucessful positive control.

Field ID	Museum ID	Sex	Exon2	Exon3	Exon4
X. clivii	12 females - 27 males				
AMNH17254	MHNG2645.069	F	Y	Y	N
BJE1552	MCZ-FS-A-30575/Cryogenic 380	F	Y	Y	N
BJE1550	MCZ-FS-A-30573/Cryogenic 378	F	N	N	N
z-23323	MCZ FS Z-23323/Cryogenic 449	F	N	N	N
BJE1549	MCZ-FS-A-30572/Cryogenic 377	F	N	N	N
BJE1503	MCZ-FS-A-30526/Cryogenic 331	F	Y	Y	N
BJE1525	MCZ-FS-A-30548/Cryogenic 353	F	Y	Y	N
BJE1526	MCZ-FS-A-30549/Cryogenic 354	F	Y	Y	N
BJE1510	no voucher	F	N	N	N
BJE1511	no voucher	F	Y	Y	N
BJE4530	no voucher	F	Y	Y	N
XCL02	no voucher	F	Y	Y	N
BJE1551	MCZ-FS-A-30574/Cryogenic 379	M	N	N	N
BJE1554	MCZ-FS-A-30577/Cryogenic 382	M	N	N	N

Continued on next page

Continued from previous page

Field ID	Museum ID	Sex	Exon 2	Exon 3	Exon 4
z-23319	MCZ FS Z-23319/Cryogenic 445	M	N	N	N
z-23321	MCZ FS Z-23321/Cryogenic 447	M	N	N	N
z-23322	MCZ FS Z-23322/Cryogenic 448	M	N	N	N
BJE1522	MCZ-FS-A-30545/Cryogenic 350	M	N	N	N
BJE1523	MCZ-FS-A-30546/Cryogenic 351	M	N	N	N
BJE1512	no voucher	M	N	N	N
BJE1513	no voucher	M	N	N	N
BJE1514	no voucher	M	N	N	N
BJE1515	no voucher	M	N	N	N
BJE1516	no voucher	M	N	N	N
BJE4531	no voucher	M	N	N	N
XCL01	no voucher	M	N	N	N
XCL03	no voucher	M	Y	N	N
XCL04	no voucher	M	Y	N	N
XCL05	no voucher	M	N	N	N
XCL07	no voucher	M	N	N	N
XCL08	no voucher	M	N	N	N
XCL09	no voucher	M	N	N	N
XCL10	no voucher	M	N	N	N
XCL11	no voucher	M	N	N	N
XCL12	no voucher	M	N	N	N
XCL15	no voucher	M	Y	N	N
XCL16	no voucher	M	Y	N	N
XCL17	no voucher	M	N	N	N
XCL18	no voucher	M	N	N	N
X. kobeli	7 females - 9 males				
BJE3075	MCZ A-148037	F	Y	Y	N
BJE3157	no voucher	F	N	N	N
BJE3074	no voucher	F	Y	Y	Y
BJE3077	no voucher	F	Y	Y	Y
BJE3155	no voucher	F	N	N	N
CAS253768	no voucher	F	N	N	N
CAS253767	no voucher	F	N	N	N
BJE3071	no voucher	M	Y	Y	?
BJE3072	no voucher	M	Y	Y	Y
BJE3150	no voucher	M	N	N	N
BJE3152	no voucher	M	N	N	N
BJE3154	no voucher	M	N	N	N
BJE3166	no voucher	M	N	N	N
BJE3167	no voucher	M	N	N	N
VG10-188	no voucher	M	N	N	N
CAS253769	no voucher	M	N	N	N
X. andrei	3 females - 2 males				
xen023	no voucher	F	?	Y	N
xen022	no voucher	F	?	Y	N
AMNH17258	MHNG2645.072	F	Y	Y	N
xen024	no voucher	M	?	N	N
AMNH17257	no voucher	M	?	N	N
X. victorianus - Lwiro (P1)	12 females - 7 males				
ELI503	UTEP 21055	F	?	Y	Y
ELI1369	UTEP 21052	F	Y	Y	Y
ELI1370	UTEP 21053	F	Y	Y	Y
ELI1461	UTEP 21056	F	Y	Y	Y
ELI1462	UTEP 21057	F	Y	Y	Y
ELI527	UTEP 21049	F	?	Y	Y
BJE260	MCZ A-138177	F	Y	Y	Y
BJE262	MCZ A-138179	F	Y	Y	Y
BJE263	MCZ A-138180	F	Y	Y	Y

Continued on next page

Continued from previous page

Field ID	Museum ID	Sex	Exon 2	Exon 3	Exon 4
BJE266	MCZ A-138183	F	Y	Y	Y
BJE267	MCZ A-138184	F	Y	Y	Y
CFS1091	UTEP 21051	F	Y	Y	Y
EBG2147 *	UTEP 20165	M	?	?	?
ELI526 *	UTEP 21048	M	?	?	?
ELI502 *	UTEP 21054	M	?	?	?
BJE261	no voucher	M	?	?	N
BJE264	no voucher	M	?	?	N
BJE265	no voucher	M	?	?	N
CFS1090 *	UTEP 21050	M	?	?	?
LWIRO 16-1	no voucher	U	Y	Y	Y
LWIRO 16-2	no voucher	U	Y	Y	Y
LWIRO 17-1	no voucher	U	Y	Y	Y
LWIRO 17-2	no voucher	U	Y	Y	Y
LWIRO 17-3	no voucher	U	Y	Y	Y
LWIRO 17-4	no voucher	U	Y	Y	Y
X. victorianus - Lendu (P2)	2 females - 5 males				
EBG2463	UTEP 20174	F	Y	Y	Y
EBG2464	UTEP 20175	F	Y	Y	Y
BJE2897	MCZ A-140118	M	Y	Y	Y
BJE2898	MCZ A-140119	M	Y	Y	Y
BJE2900	MCZ A-140121	M	Y	Y	Y
BJE2902	MCZ A-140123	M	Y	Y	Y
EBG2329	UTEP 20170	M	Y	Y	Y
X. victorianus - other					
xen234	no voucher	F	Y	Y	Y
X. poweri	2 females - 8 males				
AMNH17260	no voucher	F	Y	Y	Y
AMNH17262	no voucher	F	Y	Y	Y
BJE3252	no voucher	M	N	N	N
BJE3253	no voucher	M	N	N	N
BJE3254	no voucher	M	N	N	N
BJE3255	no voucher	M	N	N	N
VG09_100 *	no voucher	M	?	?	?
Xen058 *	no voucher	M	?	?	?
AMNH17259 *	no voucher	M	?	?	?
AMNH17263 *	no voucher	M	?	?	?
VG09_128 *	no voucher	U	?	?	?
private 33	no voucher	U	Y	Y	Y
ZFMK1 TN01	no voucher	U	Y	Y	Y
X. petersii	1 female - 3 males				
PM118	no voucher	F	Y	Y	Y
PM085	no voucher	M	Y	Y	Y
PM108	no voucher	M	Y	Y	Y
PM109	no voucher	M	Y	Y	Y
X. largeni	4 females - 1 male				
AMNH17292	MHNG2644.059	F	Y	Y	Y
BJE1505	no voucher	F	?	?	Y
BJE1508	no voucher	F	?	?	N
BJE1509	no voucher	F	?	?	N
BJE1506	no voucher	M	?	?	N
BJE1507	no voucher	U	?	?	N
BJE1580	no voucher	U	?	?	Y
BJE1581	no voucher	U	?	?	Y
BJE1582	no voucher	U	?	?	Y
BJE1583	no voucher	U	?	?	N
BJE1588	no voucher	U	?	?	N
BJE1589	no voucher	U	?	?	Y

Continued on next page

Continued from previous page

Field ID	Museum ID	Sex	Exon 2	Exon 3	Exon 4
BJE1590	no voucher	U	?	?	N
BJE1591	no voucher	U	?	?	Y
BJE1592	no voucher	U	?	?	Y
BJE1593	no voucher	U	?	?	N
BJE1594	no voucher	U	?	?	N
BJE1595	no voucher	U	?	?	N
BJE1596	no voucher	U	?	?	N
BJE1597	no voucher	U	?	?	N
z-23306	no voucher	U	?	?	N
Z-23307	no voucher	U	?	?	Y
Z-23308	no voucher	U	?	?	N
Z-23309	no voucher	U	?	?	Y

A2 *Dm-w* is inefficient in *X. petersii* but efficient in *X. poweri*

In *X. poweri*, we were able to amplify *dm-w* in all females but no males. None of the sequenced amplification had heterozygous sites, which is consistent with amplification of a single female-specific allele.

In *X. petersii*, we detected *dm-w* in all individuals of both sexes. Similar to previous findings from *X. itombwensis* (Cauret et al. 2020), heterozygous sites were observed in individuals from both sexes. This is consistent with the presence of two *dm-w* alleles segregating in an autosomal or pseudoautosomal region.

A3 Geographical variation in *dm-w* efficiency in *X. victorianus*

We were able to amplify *dm-w* in all *X. victorianus* from Lendu Plateau and Orientale Province, Democratic Republic of the Congo (DRC), which is near the northeast border with Uganda. We found heterozygous sites in exon 4 of *dm-w* in individuals from both sexes from this population. However, in *X. victorianus* from Lwiro, DRC, *dm-w* was successfully amplified only in females (12 females, 8 males were assayed, Table 1).

While it is unclear if the differentiated *X. victorianus* populations are actually different species, divergence time between these clades is relatively recent, <3 million years ago (Furman et al. 2015). We suspect that the *X. victorianus* population with female-specific *dm-w* efficient is more widespread (Kenya, DRC, Burundi, Uganda) than the *X. victorianus* population where *dm-w* is not female-specific. The uniqueness of the Lendu plateau is unsurprising because this region is part of the Albertine Rift which has been identified as a biodiversity hotspot with multiple endemic species, including two other

Xenopus species *X. lenduensis* (Evans et al. 2011) and *X. itombwensis* (Evans et al. 2008).

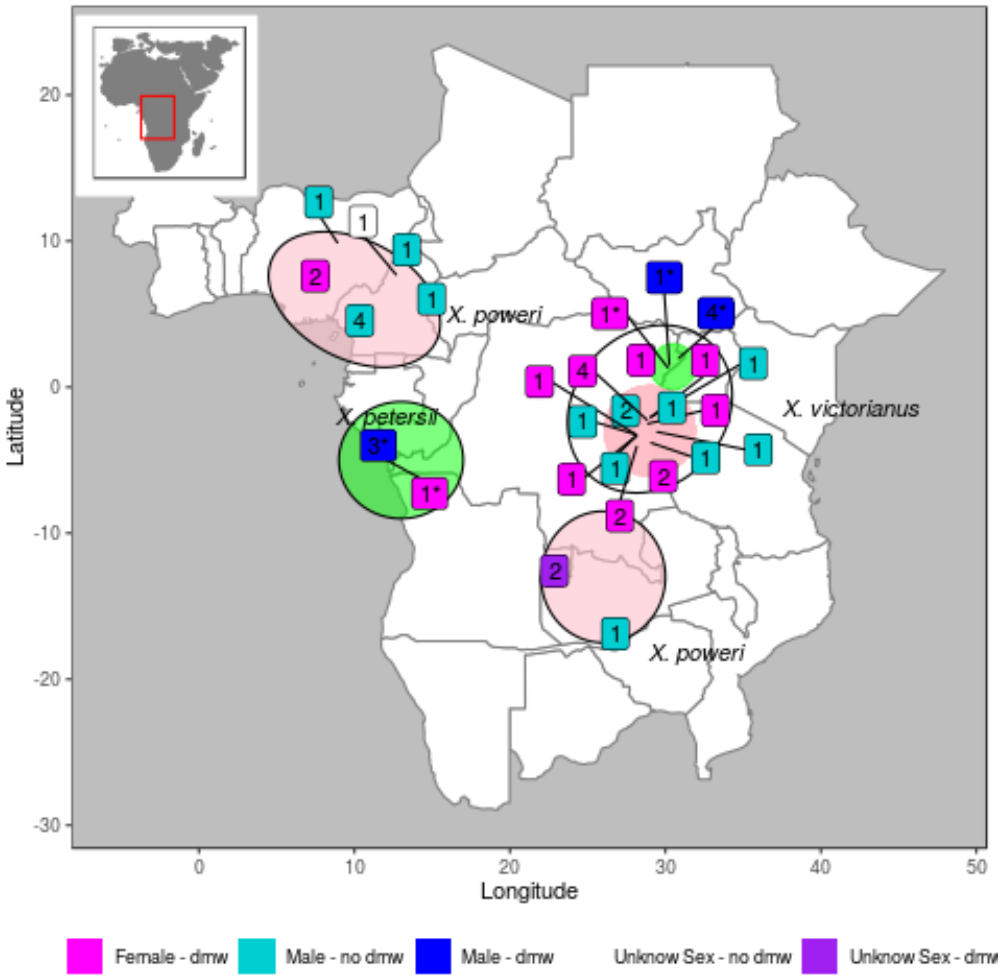


FIGURE A2.1: Geographical locations of *X. victorianus*, *X. petersii* and *X. poweri* samples. For each sex, the numbers of samples analyzed from each location are in boxes. Females carrying *dm-w* are represented in pink, males with and without *dm-w* are in blue and turquoise, respectively. Purple and white boxes indicate samples of unknown sex with and without *dm-w*. An asterisk next to a number indicates that heterozygous sites were identified in the samples. A green and a pink area represents respectively a population with an inefficient and efficient *dm-w*.

A4 Caveats to our interpretation of the evolution of the stop codon in *dm-w* exon 4

Using a phylogeny inferred from complete mitochondrial genome sequences, we infer that an early stop codon in exon 4 that is present in several extant species (*X. largeni*, *X. pygmaeus*, *X. kobeli*, *X. andrei*, *X. itombwensis*) was also present in the most recent common ancestor (MRCA) of these species and that a mutation before the divergence of the MRCA of *X. gilli* and *X. laevis* extended the C-terminal region of *dm-w*. However, there is some evidence that *X. largeni* might actually be more closely related to the clade containing *X. largeni*, *X. pygmaeus*, *X. kobeli*, *X. andrei*, and *X. itombwensis* (Furman and Evans 2018). If this were the case, it is possible that the earlier stop codon is derived rather than ancestral.

A5 Pseudogeneization of *dmrt1* paralogs and congruence with Bewick et al. 2011

We identified several differences between our capture data and the PCR/Sanger sequencing survey by Bewick et al. 2011 that highlight strengths and weaknesses of each approach.

For one species, *X. andrei*, we identified a new duplicates for exons 2 and 3 that was not identified by Bewick et al. 2011 which may have been missed because it was not successfully cloned or because too few clones were sequenced. It is also possible that there is variation among individuals in the presence of some duplicates because we did not use an identical panel of samples as Bewick et al. 2011.

For exon 2, we identified fewer *dmrt1* homeologs in the capture data from *X. largeni*, *X. itombwensis*, *X. vestitus* and *X. ruwenzoriensis* and *X. longipes* as compared to Bewick et al. 2011. For *X. largeni*, a duplicate identified by Bewick et al. 2011 contained a large (∼200bp) insertion (NCBI accession: HQ220777.1); this insertion may have reduced the efficacy of our capture probes. We suspect that one homeolog of *X. longipes* identified by Bewick et al. 2011 (NCBI accession: HQ220756.1) may be a chimerial sequence derived from alpha (L-subgenome) and beta (S-subgenome) homeologs. This could have a technical or biological origin, but either way it would have been manually removed in our analysis. For several other species the paralogs only had a few nucleotide differences which is suggestive of allelic variation and/or mutations introduced by the cloning process.

It is possible that sequence divergence could have caused us to miss some paralogs. However, we were previously able to obtain sequences from *dm-w* exons 2 and 3 using probe targeting *dmrt1* (Cauret et al. 2020). This region of *dm-w* is rapidly evolving which lead us to think that missing *dmrt1* paralogs due to sequence divergence is less likely.

Appendix C

Limited genomic consequences of hybridization between two African clawed frogs, *Xenopus gilli* and *X. laevis* (Anura: Pipidae).

Furman, B. L., Cauret, C. M., Colby, G. A., Measey, G. J., & Evans, B. J. (2017). Limited genomic consequences of hybridization between two African clawed frogs, *Xenopus gilli* and *X. laevis* (Anura: Pipidae). Scientific reports, 7(1), 1-11.

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Limited genomic consequences of hybridization between two African clawed frogs, *Xenopus gilli* and *X. laevis* (Anura: Pipidae)

Received: 13 October 2016
Accepted: 13 March 2017
Published online: 24 April 2017

Benjamin L. S. Furman[1], Caroline M. S. Cauret[1], Graham A. Colby[1], G. John Measey[2] & Ben J. Evans[1,2]

The Cape platanna, *Xenopus gilli*, an endangered frog, hybridizes with the African clawed frog, *X. laevis*, in South Africa. Estimates of the extent of gene flow between these species range from pervasive to rare. Efforts have been made in the last 30 years to minimize hybridization between these two species in the west population of *X. gilli*, but not the east populations. To further explore the impact of hybridization and the efforts to minimize it, we examined molecular variation in one mitochondrial and 13 nuclear genes in genetic samples collected recently (2013) and also over two decades ago (1994). Despite the presence of F_1 hybrids, none of the genomic regions we surveyed had evidence of gene flow between these species, indicating a lack of extensive introgression. Additionally we found no significant effect of sampling time on genetic diversity of populations of each species. Thus, we speculate that F_1 hybrids have low fitness and are not backcrossing with the parental species to an appreciable degree. Within *X. gilli*, evidence for gene flow was recovered between eastern and western populations, a finding that has implications for conservation management of this species and its threatened habitat.

Gene flow (introgression) between species may facilitate adaptive evolution through the exchange of beneficial genetic variation. This expedites the colonization of specialized ecological niches[1–3], and affects future adaptive potential by increasing genetic and phenotypic variation[2,4–7]. However, gene flow between species also poses risks by eroding species boundaries[8], disrupting adaptively evolved complexes of alleles[9,10], promoting the exchange of genetic variation associated with disease[11], influencing pathogen emergence[12], and facilitating species invasion[13,14]. As such, hybridization has important implications for biodiversity conservation.

Hybridization in African clawed frogs. Hybridization features prominently in the evolutionary history of African clawed frogs (genus *Xenopus*); 28 of 29 species are polyploid, and all of these are probably allopolyploid[15,16]. When backcrossed in the laboratory, there is variation among F_1 *X. gilli-laevis* hybrid females with respect to whether or not their progeny are polyploid[17]. Laboratory studies indicate that in some crosses (*X. gilli-X. laevis* and *X. laevis-X. muelleri*) F_1 hybrid males are sterile, but female F_1 hybrids are fertile[17,18]. F_1 *X. gilli-X. laevis* hybrid females are capable of backcrossing with either parental species, and both sexes of the F_2 backcross generation can be fertile[17]. Thus there exists the possibility that gene flow among *Xenopus* species could occur in nature. At least three *Xenopus* hybrid zones are thought to exist[19–21], and hybrids in each of these zones may have the same ploidy level as the parental species (pseudotetraploid; ref. 22).

The *X. gilli*/*X. laevis* hybrid zone. Classified by the IUCN as Endangered[23], *X. gilli*[24] occurs in south-western Western Cape Provence, South Africa[25–28]. *Xenopus gilli* is a found in seasonal ponds in lowland coastal fynbos habitat, a component of the Cape Floristic Region, which is a biodiversity hotspot[29] with an extreme level of plant endemism[30]. These ponds have high concentrations of humic compounds derived from the surrounding fynbos vegetation, and a characteristic dark color and low pH[25,31,32]. The range of *X. gilli* is disjunct and includes

[1]Biology Department, Life Sciences Building room 328, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada. [2]Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Private Bag X1, Matieland, 7602, Stellenbosch, South Africa. Correspondence and requests for materials should be addressed to B.J.E. (email: evansb@mcmaster.ca)

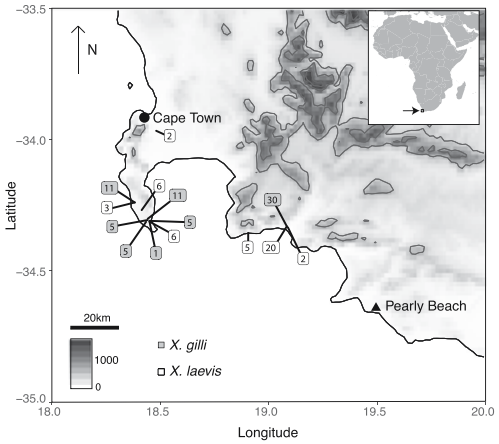


Figure 1. Sampling locations. For each species, numbers indicate the sum of number of individuals from each locality sampled in 1994 and 2013. An inset indicates the study area in southern Africa and altitude in meters is indicate on the scale. The map was made using the R package Marmap[81] using topographic data from the National Oceanic and Atmospheric Administration, USA.

the Cape of Good Hope section of Table Mountain National Park (CoGH), habitat near the town of Kleinmond, and habitat near the town of Pearly Beach (refs 25–28; Fig. 1). These three localities are interrupted by unsuitable, highly modified habitat that may impede contemporary gene-flow[26]. As with many amphibians[33], habitat degradation is a major threat to *X. gilli*[25,26].

In contrast, *X. laevis*[34], is found throughout southern Africa, in both natural and disturbed areas of South Africa and Malawi[35,36]. *Xenopus laevis* is syntopic throughout the range of *X. gilli*[25–28] and can tolerate a broad spectrum of environmental challenges including extremes of desiccation, salinity, anoxia, and temperature[37]. Picker *et al.*[32] proposed that there may be an ecological basis for speciation of *X. laevis* and *X. gilli* centered on higher tolerance of *X. gilli* embryos to low pH, allowing for habitat specialization.

Several aspects of external morphology readily distinguish these species, including smaller size of *X. gilli*, the presence of longitudinal dorsal mottling that does not connect over the midline in *X. gilli* only (for example, see Fig. 1 of ref. 19), and orange and black vermiculation on the venter of *X. gilli*. F_1 hybrids between *X. gilli* and *X. laevis* are readily identified based on individuals that are morphologically intermediate with respect to size and coloration, and this identification has been confirmed by molecular tests[19,28,38,39]. The reported abundance of F_1 hybrids varies from relatively common[26,38,39], to rare[27,28]. Morphological variation of some individuals has been previously interpreted as being derived from backcrosses of F_1 hybrids with each parental species[39].

The western extent of the *X. gilli* distribution occurs within the CoGH[25,26]. Following reports of hybrids and expansion of *X. laevis* populations, steps were taken in the mid-1980s to minimize co-occurence of these two species within the CoGH which included removal of *X. laevis* from *X. gilli* ponds, translocation of *X. gilli* to new sites[40], and construction of a wall around a known *X. gilli* pond[25,41]. The hope was to minimize hybridization and resource competition, for example, if larger *X. laevis* individuals are able to outcompete *X. gilli* for food[39,42]. With some interruptions, these efforts have continued for the last 30 years in the CoGH. Similar efforts have not been made for eastern populations of *X. gilli* which are located on private property, and in some ponds in these areas where *X. gilli* had been found in the past, now only *X. laevis* are found[26,41].

To further investigate the effect of hybridization on gene flow between *X. gilli* and *X. laevis*, we examined DNA sequence variation in these species from one mitochondrial DNA (mtDNA) marker and 13 nuclear DNA (nDNA) markers. Genetic samples were collected from within managed (west) and unmanaged (east) portions of the range of *X. gilli*. Samples were analyzed from both locations that were collected shortly after management began, and then in the same areas again 20 years later. We expected that if introgression was occurring in the populations during this time period, it would be more pronounced in the east population. If efforts to minimize hybridization in the west were successful, we expected more evidence of gene flow in the samples collected soon after management began as compared to more recently. However, in both localities and both sampling times, we found no evidence of shared mitochondrial haplotypes or nuclear alleles between these species, suggesting that the F_1 hybrids have low fitness and are not backcrossing with the parental species to an appreciable degree, despite potential fertility of F_1 females[17]. Within *X. gilli*, we recovered evidence of gene flow between east and west populations, and found genetic diversity to be higher in the unprotected eastern population. These findings have implications for management and conservation of this endangered habitat specialist.

Materials and Methods

Genetic samples analyzed in this study were collected either in 1994 or in 2013. Some of the samples from the earlier collection were also analyzed in two earlier studies[27,28]. The 2013 collection included *X. gilli* and *X. laevis* individuals from the same or geographically close (within 5 km) sites as the 1994 collection, and both sampling efforts used funnel traps. Animal sampling protocols approved by the Institutional Animal Care and Use Committee at Columbia University and work was performed in accordance with all relevant guidelines and regulations for animal experimentation, in accordance with laws for studying wildlife in South Africa and with appropriate collection permits from the Chief Directorate of Nature Conservation and Museums, and was approved by the Animal Ethics Committee at the University of Cape Town and the Stellenbosch University Research Ethics Committee: Animal Care and Use. Samples were obtained east and west of False Bay for both species and for both time periods (Fig. 1). We assigned individuals to species (*X. gilli* or *X. laevis*) based on dorsal and ventral patterning, shape of head, and overall size[43,44]. Because this study aimed to explore genetic effects of backcrossed hybrids, for both time points, we intentionally excluded individuals whose intermediate morphology (and genetic analysis in the case of the 1994 individual[28]) indicated that they were F_1 hybrids (1 individual from 1994 and 9 from 2013).

DNA was extracted from tissue samples using Qiagen DNEasy tissue extraction kits (Qiagen, Inc), following the manufacturer's protocol, or a phenol-chloroform protocol. A fragment of the mtDNA genome was amplified and sequenced for 36 and 33 *X. gilli* and *X. laevis* individuals, respectively, using primers from ref. 45 that target a portion of the 16S ribosomal RNA gene (*16S*). Exons of 13 nDNA genes ranging from 333–770 bp in length were sequenced for 20–41 *X. gilli* and 11–31 *X. laevis* individuals using paralog specific primers (primers are from ref. 46). These exons came from the genes B-Cell CLL/Lymphoma 9 (*BCL9*), BTB domain containing 6 (*BTBD6*), Chromosome 7 Open Reading Frame 25 (*C7orf25*), Fem-1 Homolog C (*FEM1C*), Microtubule Associated Serine/Threonine Kinase Like (*MASTL*), Mannosyl-oligosaccharide glucosidase (*MOGS-1*), Nuclear Factor, Interleukin 3 Regulated (*NFIL-3*), protocadherin 1 (*PCDH1*), phosphatidylinositol glycan anchor biosynthesis class O (*PIGO*), protein arginine methyltransferase 6 (*PRMT6*), Ras association domain family member 10 (*RASSF10*), SURP and G-patch domain containing 2 (*SUGP2*), and zinc finger BED-type containing 4 (*ZBED4*). A table of sample IDs and which loci were amplified for which samples is available in the Appendix. In the phylogenetic analysis of individual genes (discussed below), we used as an outgroup a sequence from *X. tropicalis* from the genome assembly version 9.0 on Xenbase[47]. When possible, we also included orthologous and homeologous sequences from *X. laevis* from the genome assembly version 9.1 on Xenbase[47], which was identified using BLAST[48]; this was not possible when a homeologous sequence was not identified, which could be due to gene loss or missing data in the genome sequence. Sequence data were aligned using MAFFT[49] and corrected by eye. Coding frame was estimated using the 'minimize stop codons' option in Mesquite v.3.04[50], and alignments were trimmed to begin at the first position and end at the third position of the reading frame.

We calculated the phase of nDNA alleles (i.e. haplotypes) using the 'best guess' option of PHASE[51,52] with default parameters. Each individual's allelic sequences for each locus were used in subsequent population genetic, clustering, and gene tree analyses. Thus, for each nuclear locus, an individual frog was represented by two sequences, each corresponding to one allele.

Gene trees. Gene trees were estimated for each phased nDNA exon and the mtDNA alignment using BEAST v1.8.3[53]. Substitution models were selected based on the Akaike Information Criterion using MRMODELTEST v.2[54], and xml files were prepared for BEAST using BEAUTI (part of the BEAST package). For each nDNA locus, we ran two Markov chain Monte Carlo runs for 25 million generations. For the mtDNA, the model selected by MRMODELTEST2 (GTR+Γ) failed to converge on stable parameter estimates, and we therefore instead used the simpler HKY+Γ model, and ran two chains for 50 million generations. For each analysis, convergence of parameter estimates on the posterior distribution was assessed using TRACER v.1.5[55] based on an effective sample size (ESS) value >200 and inspection of the trace of parameter estimates against the MCMC generation number. Based on this, for all phylogenetic analyses the first 25% of the posterior distribution was discarded as burn-in. Then, using TREEANNOTATOR, we produced consensus trees from the post-burn-in posterior distribution of trees.

Species tree. We also estimated a species tree (with the nuclear sequences used in the STRUCTURE analysis, see below) using the multi-species coalescent model of *BEAST[56]. We trimmed the dataset to include only nDNA genes with all populations sampled (see Genetic clusters section for details) and included only individuals sampled for all genes. All *X. laevis* individuals were considered to be the same species (17 individuals), and we separated the east and west *X. gilli* populations into separate species (10 and 11 individuals, respectively), and *X. tropicalis* was considered its own species. We set a simple HKY+Γ model joined for all data partitions (so that convergence of parameter estimates could be reached), assumed a strict molecular clock joined for all data partitions, and allowed the underlying gene tree structure to vary across data partitions. We ran the 8 chains for 170 million generations and removed 50 million generations as burn-in.

Genetic clusters. We used STRUCTURE v.2.3.4[57] to estimate individual assignment probabilities to genetic clusters using best-guess phased nDNA alleles on a subset of individuals. Three loci lacked data from the east *X. gilli* 1994 population (exons of the genes *MOGS-1*, *PCDH1* that also lacked data from this exon for *X. laevis* east 1994 samples, and *PIGO*), so we excluded them from STRUCTURE analysis. We also excluded individuals with >50% missing data for the remaining 10 loci. This resulted in a dataset of 13, 8, 11, and 6 *X. gilli* individuals from the following localities and years respectively: east 1994, east 2013, west 1994, and west 2013, where east and west refer to the sampling locations relative to False Bay. This analysis also included 9, 12, 6, 4 *X. laevis* individuals from east 1994, east 2013, west 1994, and west 2013 respectively. We used the admixture model of STRUCTURE and assumed no correlation between alleles at different loci. We ran the Markov chain Monte Carlo for 20 million generations, following a two million generation burn-in. We tested a number of clusters (K) ranging from 1–8,

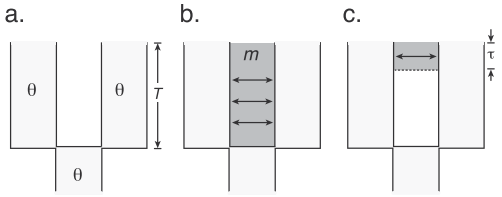


Figure 2. Evolutionary models considered for *X. gilli* sequence data from east and west populations included (**a**) population division without subsequent gene flow, (**b**) separation followed by ongoing gene flow, and (**c**) separation followed by secondary contact after a period of no gene flow. Model parameters include the population polymorphism parameter θ, which is assumed to be constant in the ancestral and both descendant populations, the time of speciation T, the amount of migration m, and the time of secondary contact τ.

with 5 replicate analyses for each setting of K. To correct for label switching and to average assignment probabilities across runs, we used CLUMPP v.1.1.2[58]. We first computed the D statistic, following recommendations in the CLUMPP manual[58], to decide on the particular algorithm to employ for maximizing similarity across runs. We then used the *ad hoc* method of ref. 57 and the method described by ref. 59 to evaluate the most likely number of genetic clusters (K).

Evolutionary models. As discussed below, our analyses did not detect mitochondrial DNA haplotypes or nuclear alleles that were shared between *X. laevis* and *X. gilli* but did detect shared haplotypes and alleles between populations of *X. gilli* and between populations of *X. laevis*. Because *X. gilli* is of conservation concern, we evaluated the fit of data from this species to three evolutionary models (Fig. 2). In the first (isolation) model, divergence of two *X. gilli* populations was followed by no migration between each population. Under the isolation model, all shared alleles between these populations would be due to incomplete lineage sorting (ILS). In the second (ongoing migration) model, *X. gilli* population divergence was followed by ongoing symmetrical migration between the populations. Under the ongoing migration model, shared alleles would be due to ILS or migration, and some of the alleles shared due to migration could have been exchanged millions of years ago. In the third (secondary contact) model, *X. gilli* population divergence was followed by a period of no migration and then by a period during which symmetrical migration occurred between east and west populations. Under the secondary contact model, shared alleles would again be due to ILS or migration, but alleles shared due to migration could only have been exchanged recently.

All models include a parameter T, which is the time of separation between the *X. gilli* populations and a parameter θ, which is the population polymorphism parameter of the ancestral and both descendant populations. The second and third models have an additional parameter m, which is the number of individuals in each population that are replaced per generation by individuals from the other population (east vs west), divided by the product of four times the effective population size of each population. The third model includes another parameter τ, which is the proportion of T going back in time from the present that secondary contact began. Thus, the ongoing migration and isolation models are special cases of the secondary contact model in which $\tau = 1$, or $\tau = 1$ and $m = 0$, respectively. We note that several assumptions of these models are undoubtedly violated (e.g., constancy of population size over time, equivalent population size of both descendant and the ancestral populations) but we made them nonetheless so we could complete the simulations (see below) within a reasonable amount of time, and because of the relatively small size of the dataset.

The approximate likelihood of combinations of values for these parameters was estimated using rejection sampling[60]. In this approach, the likelihood is approximated by the natural logarithm of the number of simulations for which the sum of four summary statistics from a simulation (discussed next) were within $\pm\varepsilon$ % of the sum of the observed four summary statistics from actual sequence data, divided by the number of simulations, where $\varepsilon = 25$. The value of ε determined how close the simulations must match the observed data in order to contribute to the likelihood, and was selected based on a compromise between the computational efficiency of the likelihood estimation and the accuracy of the estimate[60]. For the ongoing migration model and the secondary contact model, 40,000 simulations were performed for each combination of parameter values we considered. For the isolation model, no simulations had summary statistics within $\pm\varepsilon$ % of the observed; thus, 1,000,000 simulations were performed in order to achieve an upper bound for the likelihood estimate. The likelihood of the data over all combinations of the following parameter value intervals were estimated: T: every 1,000,000 generations in the interval of 0–20,000,000 generations; θ: every 0.001 units in the interval of 0.001–0.01 and every 0.01 units in the interval of 0.01–0.1; τ: every 10% in the interval of 10–100%; m every 0.1 units in the interval of 0–1 and every integer in the interval of 1–10.

We used the sum across loci of four summary statistics described by ref. 61 for these likelihood calculations, and simulations were performed using the program mimarsim[62]. These four summary statistics include the number of sites with a derived polymorphism (i) in the west population of *X. gilli* only, (ii) in the east population of *X. gilli* only, (iii) shared between the west and east populations of *X. gilli*, or (iv) fixed in either the west or in the east population of *X. gilli*. The simulations used a fixed value for the mutation rate equal to $2.69e^{-9}$ substitutions per site per generation, which was estimated based on the average synonymous divergence between a randomly selected *X. gilli* sequence and an orthologus sequence from *X. tropicalis*, and assuming a divergence time of 65

million years for the separation of these lineages[63], and a generation time of one year. Each locus had a mutation rate scalar based on synonymous divergence to *X. tropicalis* that accommodated variation among loci in the rate of evolution. To minimize the influence of natural selection on the polymorphism data, summary statistics and likelihood calculations were based only on variation at synonymous positions. Confidence intervals were estimated using the profile likelihood method (i.e., that the 95% confidence interval is defined by the two points that are 1.92 *log-likelihood (lnL)* units from the maximum).

Population dynamics over time and space. We performed various analyses to assess whether the genetic diversity varied among these species, over time, or among populations east and west of False Bay. Pairwise F_{ST} (with significance computed by a permutation test) was quantified for the same data used in the STRUCTURE analysis using ARLEQUIN v3.5.2.2[64]. Nucleotide diversity (π) of each locus was calculated using the pegas package in R[65,66]. We then calculated a mean value of π across loci for each of the eight populations, weighting the estimate by gene length for each locus. Confidence intervals were obtained by bootstrapping the weighted π values 5000 times.

Because allelic diversity is influenced by sample size, we using the program HP-RARE to calculate rarefied estimates of allelic diversity[67], which involves downsampling data to the smallest number of samples in each population across all nuclear loci for which there were data. This analysis was performed with the same data as used in the STRUCTURE analysis. For *X. laevis* populations there was one exception; the *prmt6* locus had only four sampled alleles for the *X. laevis* west 1994 population, thus we did one run with all of the data (using four as the smallest number of sampled alleles) and another run excluding *prmt6* (in which case, eight was the smallest number of sampled alleles). For all *X. gilli* populations, the smallest number of sampled alleles was eight. We generated confidence intervals by bootstrapping of the allelic diversity measurements 5000 times.

To statistically evaluate differences in genetic diversity over time, location and species, we constructed linear mixed models using the R package lme4[68]. We built models for the estimated values of nucleotide diversity (π) and allelic diversity independently with diversity values measured for each locus, using time (1994 or 2014), location (east or west) and species (*X. gilli* or *X. laevis*) as fixed effects (all additive, no interaction terms) and considering locus as a random effect. For each parameter of both models, we also used lme4 to compute confidence intervals with the confint function.

Results

Molecular polymorphism and gene trees. In the mitochondrial and 13 nuclear gene trees, alleles from *X. gilli* and *X. laevis* clustered in reciprocally monophyletic clades (Fig. 3, Fig. S1). No individuals were found to have introgressed loci, which would have been evidenced by an allele in one species having a closer relationship to the alleles of the other species (i.e. a paraphyletic relationship). Similar to previous studies[26,27], the mtDNA gene tree identified divergence between the east and west populations of *X. gilli* (Fig. 3). We identified one individual (Sample ID: XgUAE_08) from the west population of *X. gilli* that carried a mtDNA haplotype that was more closely related to haplotypes that were carried by individuals from the east population. This observation was also reported previously, from different samples[26,27]. The *BEAST analysis recovered the expected species tree of these three species with posterior probabilities of one (Fig. S2). This analysis estimated the divergence time of *X. gilli* and *X. laevis* at about 14.05 my and divergenc of the east and west *X. gilli* populations at about 1 my (0.51–1.36 my 95% HDP; when a calibration point of 65 my from *X. tropicalis* is assumed[63]).

Genetic clusters. STRUCTURE analyses assigned each individual to groups that corresponded with species assignment (Fig. 4a). All *X. laevis* individuals were assigned to a single genetic cluster at $K=2$–8, indicating a lack of allele frequency clustering, which is consistent with gene flow across the population range. The *X. gilli* samples were assigned to two clusters corresponding to sampling location (east and west) at $K=3$–8, indicating differences in allele frequencies, which is consistent with restricted gene flow between them (Fig. 4a). Assignment of individuals to clusters stabilized at $K=3$, with no new clusters being detected at higher values of K (Fig. 4a). The likelihood plot plateaus at $K=3$ (Fig. 4b); the Evanno method[59] supports $K=2$ and the *ad hoc* method of ref. 57 supports $K=3$.

Evolutionary models. Using simulations and summary statistics, we evaluated the fit of the *X. gilli* data to evolutionary models with no migration after speciation, with ongoing migration after speciation, or with secondary contact after speciation. The *lnL* of the secondary contact model was −8.032, the ongoing migration model was −8.987, and the isolation model was <−13.815. We were not able to more precisely estimate the likelihood of the isolation model because no simulations under this model resulted in data whose four summary statistics were within $\pm\varepsilon$ of the observed values (see Methods).

Nested models can be compared by assuming that twice the difference between the *lnL* of each model follows a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters in each model (denoted χ_1^2 for comparison between models that differ in one parameter). However, because comparison between these successively more complex models involves a boundary condition on one parameter ($\tau=1$ for the ongoing migration model, $m=0$ for the isolation model), this difference in model likelihoods follows a mixture of χ_0^2 and χ_1^2 distributions[69]. The secondary contact model is thus not supported over the ongoing migration model ($p=0.08$), but the ongoing migration model is supported over the isolation model ($p=0.009$). Overall then, these results support an inference of gene flow between *X. gilli* populations, but fail to discern substantial temporal heterogeneity in the level of gene flow.

The maximum likelihood parameter estimates and 95% confidence intervals for the ongoing migration model were θ: 0.002 (0.001–0.003) and m: 0.7 (0.1–2) individuals/generation. The maximum likelihood estimate for T was 8,500,000 generations; the 95% CI was unable to be estimated because it exceeded the boundaries we

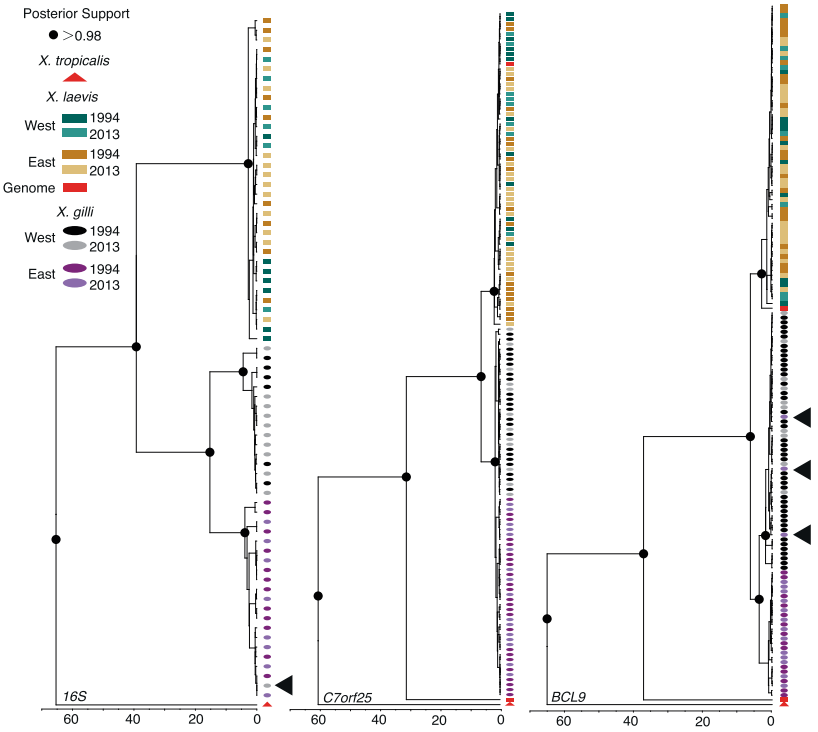


Figure 3. Representative gene trees that collectively provide no evidence of genetic exchange between *X. gilli* and *X. laevis*. The phylogeny on the left illustrates divergence between 16S rDNA mitochondrial sequences in the east and west populations of *X. gilli*, and with one shared sequence (indicated with an arrow) that occurred on both sides of False Bay. The nuclear phylogenies in the center and right provide examples of no shared alleles and shared alleles between the east and west *X. gilli* populations, respectively. Gene name acronyms are described in the Materials and Methods section. These and other phylogenies are depicted with sample labels in Fig. S1.

tested (1,000,000–20,000,000), suggesting low statistical power to estimate this parameter. Comparisons to similar parameters estimated for African clawed frogs by other studies using other methods[16,36,70] suggest that these estimates are biologically plausible. Our intuition that the shared identical alleles between east and west *X. gilli* populations are due to ongoing migration is thus supported, with caveats that several model assumptions, discussed below, are violated to some degree.

Population dynamics over time and space. In line with results from STRUCTURE analysis, a high F_{ST} was measured in all pairwise comparisons of the east and west populations of *X. gilli* (comparing within the same year 2013 east to 2013 west, and between years 1994 east to 2013 west and 2013 east to 1994 west; range: 0.55–0.60, $p < 0.05$). For *X. gilli*, between time points within each location (east or west), F_{ST} was not significantly different from zero (east 1994 to east 2013 and west 1994 to west 2013; $p > 0.05$, $F_{ST} < 0.02$). For *X. laevis*, pairwise comparisons of east and west populations, within the same year (1994 east to 1994 west or 2013 east to 2013 west) and between time points (1994 east to 2013 west and 2013 east to 1994 west), had intermediate F_{ST} values that departed significantly from zero ($p < 0.05$, $F_{ST} = 0.07–0.16$). But within locations comparing time points (1994 east to 2013 east and 1994 west to 2013 west), F_{ST} was not significantly different from zero ($p > 0.05$, $F_{ST} = 0.04$ for both comparisons).

Both nucleotide diversity and allelic diversity did not change drastically over time, but within species, both statistics were higher in the east population than the west (Fig. 5). In the linear mixed model analysis of π, the effect of species was significant with *X. laevis* higher than *X. gilli* by 0.00073 substitutions per site (95% CI: 0.00027–0.00119). The effect of location was significant with π lower in the west than the east population by 0.00091 (95% CI: 0.00045–0.00137). The effect of time of sampling was not significant, with the 2013 samples being lower by 0.00016 but the 95% CI of this difference spanning zero (−0.00062–0.00030). Similar results were recovered for allelic diversity, with *X. laevis* having higher allelic diversity than *X. gilli* (0.59, 95% CI: 0.21–0.97),

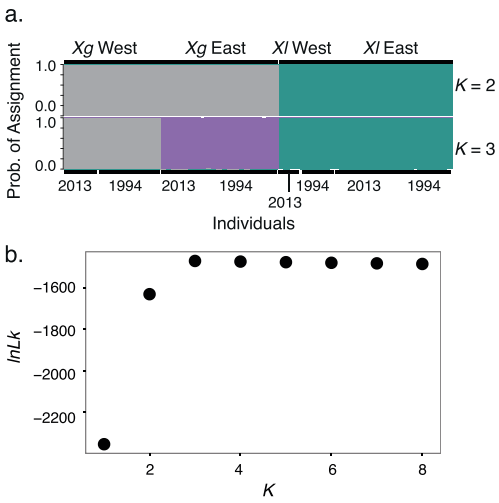


Figure 4. (**a**) Structure analyses for 10 loci, which had sequence data for all populations. (**b**) Likelihood for each value of *K*.

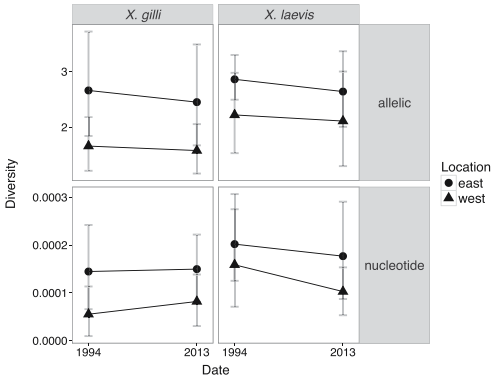


Figure 5. Genetic diversity statistics including rarefied estimates of allelic diversity (top panels) and nucleotide diversity (π) weighted by length of sequence; bottom panels). For allelic diversity, the analysis considered the same 10 loci as were analyzed by the STRUCTURE analysis (see Materials and Methods). Allelic diversity for *X. laevis* did not include *PRMT6* locus because this locus only had four alleles for the west 1994 population.

the west being less diverse than the east (−0.77, 95% CI: −1.14–−0.40), and no significant effect of sampling time (−0.15, 95% CI: −0.52–0.21).

Discussion

Gene Flow between *X. laevis* and *X. gilli*. Previous investigations of the genetic consequences of hybridization between *X. laevis* and *X. gilli* found no evidence of widespread genetic introgression[27,28], a result that seemed to be at odds with the incidence of morphologically and genetically identified hybrids in this and other studies[19,26,38,39,71]. In this study, we analyzed many of the samples from refs 27, 28 and also genetic samples that were collected more recently. Evidence of introgression between *X. laevis* and *X. gilli* was not detected in mitochondrial DNA or in any of 13 nuclear loci (Fig. 3; Fig. S1). Furthermore, each species formed separate genetic clusters with no evidence for similarities in allele frequencies (Fig. 4a). These findings were consistent in both sampling efforts examined here, which included targeting both populations of *X. gilli* and sampling time points separated by about two decades. Previous investigations into the extent of genetic introgression[28], used two

nuclear loci that were not used in this study. Combining that study with ours brings the total number of genomic regions studied to 15, and includes variation from 6 of the 18 chromosome pairs based on gene location in the *X. laevis* version 9 genome, on Xenbase. This expanded sampling is thus consistent with the interpretation by ref. 28 that genomic introgression is not extensive.

The lack of introgression is despite the continued identification (based on morphology) of a low frequency of putative F_1 hybrids in both localities. Though there could be an adaptive benefit for hybridization because *X. gilli* embryos can tolerate ponds with higher pH levels than *X. laevis*, which perhaps could allow for invasion of *X. gilli* habitat, we found no evidence that hybridization has led to gene flow of the genetic basis of this or other ecological adaptations that evolved after these two species diverged from their most recent common ancestor. Although not the focus of this study, the relatively low abundance of F_1 hybrids argues against the possibility that a new species of hybrid origin is evolving in this zone of sympatry between *X. laevis* and *X. gilli*. Reproductive isolation in amphibians has been shown to happen in a few million years for some lineages[72,73]. In *Xenopus* species, female individuals respond to species-specific calls evoked by males (phonotaxis)[74] and this presumably acts to some degree as a prezygotic barrier to hybridization. However, an observation is that at high densities, *Xenopus* individuals amplex indiscriminately (G. J. Measey, personal observation), potentially overriding some prezygotic barriers. In some ponds, *X. gilli* individuals can be outnumbered by *X. laevis* 3:1[41], and indiscriminate amplexus could mean *X. gilli* males (which are also smaller) are outcompeted for access to females. This may be why hybrids are occasionally seen, but the extended period of divergence between these species (~14 my) appears to have resulted in strong post-zygotic barriers preventing introgression.

Hybridization followed by back-crossing is expected to generate a mosaic of introgressed and non-introgressed genomic regions. Variation among genomic regions in the extent of introgression can be further augmented by natural selection favoring or disadvantaging genetic variants from one species in the genomic background of the other[5]. In California tiger salamanders (*Ambystoma californiense*), for example, some loci are fixed for foreign alleles from the introduced barred tiger salamander (*A. mavortium*), whereas other loci exhibit no sign of introgression[6]. That the barred tiger salamander was introduced only 60 years ago suggests that mosaicism of genomic introgression arose rapidly (in ~20 generations; ref. 6). In this study it is therefore possible that we failed to identify some introgressed regions of the genome due to the relatively sparse sampling of genomic regions. Future studies that survey variation across the entire genome, such as RAD-Seq[75], could more precisely quantify the extent of gene flow between these species, if it occurs.

Population structure in *X. gilli* and change over time. Analysis of mtDNA[26,27,45] and skin peptides secreted by these populations[76] support the existence of at least two distinct populations in *X. gilli* in the western and eastern portion of its range. Our mitochondrial analysis, STRUCTURE analysis, and some of the gene trees reported here (such as *MOGS-1* and *RASSF10*) also exhibit substantial geographic differences in *X. gilli* allele frequencies between these populations (Figs 3 and 4a, Fig. S1). In contrast, genetic diversity in *X. laevis* has minimal geographic structure, with most alleles occurring on both sides of False Bay, and STRUCTURE analyses assigning all *X. laevis* individuals to a single genetic cluster (Fig. S1, Fig. 4a). This is similar to findings reported by ref. 27.

When and why did population structure arise in *X. gilli*? Using mtDNA sequence data and a relaxed molecular clock[45], estimated that the divergence between *X. gilli* populations occurred 8.5 million years (my) (95% CI: 4.8–13.4), which is the same as the estimate obtained here using our coalescent modeling approach. This estimate is older than the 1 my divergence time estimated by the *BEAST analysis (Fig. S2), but this is not unexpected because *BEAST does not incorporate gene flow after divergence in its model. Using similar data and a coalescent modeling approach ref. 26 recovered a somewhat more recent divergence time of 4.63 my, but with confidence intervals that overlapped with the previous estimate (95% CI: 3.17–6.38). Evans *et al.*[27] proposed that the two populations split following inundation of the Cape Flats. Fogell *et al.*[26] pointed out that marine inundation probably occurred multiple times in the last few million years and that cycles of aridification also likely influenced the costal fynbos habitat, on which *X. gilli* relies. Our finding of gene flow after divergence supports the idea that these populations have been periodically reconnected, allowing exchange of migrants. Therefore, whatever the cause of divergence was, it was demonstrably not a permanent barrier.

Of note is that the evolutionary models we tested are almost certainly violated by the system we explored in many ways, including variation over time and among populations in population size, mutation rate, and migration rate. Although we do not anticipate that these violations are influential enough as to negate the rejection of the isolation model, a larger dataset might provide statistical power with which to better evaluate more complex scenarios, such as the secondary contact model.

The F_{ST} and linear mixed model analyses suggest that allele frequencies have not changed substantially in the last 20 years, though there is a trend of decreasing diversity (Fig. 5 and from values obtained in linear mixed models indicated a non- significant decline in diversity from 1994 to 2013). If generation time is about one year or less (which is based on laboratory studies and could be an underestimate, ref. 35), this represents 20 generations. Changes in allelic diversity may signal population declines earlier than nucleotide diversity, because loss of rare alleles (which happens during population declines) would have a greater impact on count based metrics, such as allelic diversity, than they would on frequency based metrics such as π[77]. Thus, though not significant, a declining trend seen for allelic diversity (Fig. 5) may be an early indication of population declines. Linear mixed models allowing for independent changes in diversity for each locus over time revealed declining genetic diversity (except for two loci in the π models; results not shown).

Management. Hybridization and introgression has the potential to threaten species survival[10]. In an attempt to reduce gene flow between species, three conservation actions were implemented in the mid-1980s. A wall was built around one impoundment in CoGH[25], populations of pure *X. gilli* were translocated to areas without *X. laevis*[40], and *X. laevis* were manually removed from CoGH[25,41,78]. Removal of *X. laevis* ceased in 2000, but

resumed in 2011[25,41,78]. The same management efforts have not been conducted for the population of *X. gilli* east of False Bay, most of which resides in non-protected areas[26,41].

Interestingly, the CoGH has greater juvenile recruitment of *X. gilli*[41] and fewer hybrids (2.5% vs 8–27% of individuals in ponds in the west and east respectively, ref. 26). Our results suggest that these hybrids are not producing successful offspring via backcrossing with either parental species frequently enough to produce large scale genomic impacts. These results suggest that the genomes of *X. gilli* and *X. laevis* are largely genetically distinct. Thus, the major benefits to *X. gilli* of removal of *X. laevis* from habitat shared with *X. gilli* probably stem from minimizing competition for ecological resources between these species[25,41,42].

For *X. gilli* and *X. laevis*, east populations harbor more genetic diversity than the west populations (Fig. 5). Allelic diversity and heterozygosity reflect a population's ability to respond to selection[79], and thus from a genetic perspective conservation of east populations of *X. gilli* is paramount.

This study suggests that patterns of gene flow within *X. gilli* included genetic exchange between populations in the east and west. The ancestral distribution of *X. gilli* was likely patchy to begin with and has contracted considerably in the last several decades, including in locations now occupied only by *X. laevis* or *X. laevis* and hybrids[25,26]. Ancestral patterns of gene flow are presumably imperiled by further habitat fragmented by human activity, including habitat altering effects of invasive species such as *Acacia saligna* (Port Jackson Willow), *Acacia mearnsii* (Black Wattle), and *Hakea sericea* (Silky Hakea)[80]. Continued efforts to conserve and restore coastal fynbos habitat both inside and outside of protected areas[43], such as removal of invasive vegetation, restoration of native vegetation, and removal of *X. laevis*, stands to benefit *X. gilli*. This is particularly important in the east population of *X. gilli* near Kleinmond where genetic diversity is highest and the population resides on private land.

References

1. Anderson, E. & Hubricht, L. Hybridization in tradescantia. iii. the evidence for introgressive hybridization. *Am. J. Bot.* 396–402 (1938).
2. Dowling, T. E. & Secor, C. L. The role of hybridization and introgression in the diversification of animals. *An. Rev. Ecol. Sys.* 593–619 (1997).
3. Rieseberg, L. H. *et al.* Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211–1216, doi:10.1126/science.1086949 (2003).
4. Stelkens, R., Brockhurst, M., Hurst, G., Miller, E. & Greig, D. The effect of hybrid transgression on environmental tolerance in experimental yeast crosses. *J. Evol. Biol.* **27**, 2507–2519 (2014).
5. Arnold, M. L. & Martin, N. H. Adaptation by introgression. *J. Biol.* **8**, 1 (2009).
6. Fitzpatrick, B. M. *et al.* Rapid fixation of non-native alleles revealed by genome-wide snp analysis of hybrid tiger salamanders. *BMC Evol. Biol.* **9**, 176 (2009).
7. Anderson, E. Introgressive hybridization. *Introgressive hybridization* (1949).
8. Arnold, M. L. *Evolution through genetic exchange*, vol. 3 (Oxford University Press Oxford, 2006).
9. Gilk, Sara E. *et al.* Outbreeding depression in hybrids between spatially separated pink salmon, *Oncorhynchus gorbuscha*, populations: marine survival, homing ability, and variability in family size. In *Genetics of Subpolar Fish and Invertebrates* 287–297 (Springer, Netherlands, 2004).
10. Rhymer, J. M. & Simberloff, D. Extinction by hybridization and introgression. *An. Rev. Ecol. Sys.* 83–109 (1996).
11. Simonti, C. N. *et al.* The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**, 737–741 (2016).
12. Stukenbrock, E. H. The role of hybridization in the evolution and emergence of new fungal plant pathogens. *Phytopathology* **106**, 104–112 (2016).
13. Figueroa, M. *et al.* Facilitated invasion by hybridization of *Sarcocornia* species in a salt-marsh succession. *J. Ecol.* **91**, 616–626 (2003).
14. Blair, A. C., Blumenthal, D. & Hufbauer, R. A. Hybridization and invasion: An experimental test with diffuse knapweed (*Centaurea diffusa* Lam). *Evol. Appl.* **5**, 17–28 (2012).
15. Evans, B. J. Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Front. Biosci.* **13**, 4687–4706 (2008).
16. Evans, B. J. *et al.* Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African clawed frog (*Xenopus*, Pipidae) from west and central Africa. *PLoS One* **10**, e0142823 (2015).
17. Kobel, H. R. Reproductive capacity of experimental *Xenopus gilli* x *X. l. laevis* hybrids. In *The Biology of Xenopus* (eds. Kobel, H. R. & Tinsley, R. R.) 73–80 (Oxford University Press, Oxford, 1996).
18. Malone, J. H., Chrzanowski, T. H. & Michalak, P. Sterility and gene expression in hybrid males of *Xenopus laevis* and *X. muelleri*. *PLoS One* **2**, e781 (2007).
19. Kobel, H. R., Pasquier, L. D. & Tinsley, R. C. Natural hybridization and gene introgression between *Xenopus gilli* and *Xenopus laevis laevis* (Anura: Pipidae). *J. Zool.* **194**, 317–322 (1981).
20. Fischer, W., Koch, W. & Elepfandt, A. Sympatry and hybridization between the clawed frogs *Xenopus laevis laevis* and *Xenopus muelleri* (Pipidae). *J. Zool.* **252**, 99–107 (2000).
21. Yager, D. D. Sound production and acoustic communication in *Xenopus borealis*. In *The Biology of Xenopus* (eds. Kobel, H. R. & Tinsley, R. R.) 121–140 (Oxford University Press, Oxford, 1996).
22. Tymowska J. Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. In *Amphibian Cytogenetics and Evolution* (eds. Green D. S. & Sessions S. K.) 259–297 (Academic Press, San Diego 1991).
23. South African Frog Re-assessment Group (SA-FRoG), IUCN SSC Amphibian Specialist Group. *Xenopus gilli*. The IUCN Red List of Threatened Species 2010: e.T23124A9417597. http://dx.doi.org/10.2305/IUCN.UK.2004.RLTS.T23124A9417597.en Accessed: 2016-09-29 (2010).
24. Rose, W. & Hewitt, J. Description of a new species of *Xenopus* from the Cape Peninsula. *Trans. Royal Soc. South Africa* **14**, 343–346 (1926).
25. Picker, M. D. & de Villiers, A. L. The distribution and conservation status of *Xenopus gilli* (Anura: Pipidae). *Biol. Conserv.* **49**, 169–183 (1989).
26. Fogell, D. J., Tolley, K. A. & Measey, G. J. Mind the gaps: Investigating the cause of the current range disjunction in the Cape Platanna, *Xenopus gilli* (Anura: Pipidae). *PeerJ* **1**, e166 (2013).
27. Evans, B. J., Morales, J. C., Picker, M. D., Kelley, D. B. & Melnick, D. J. Comparative molecular phylogeography of two *Xenopus* species, *X. gilli* and *X. laevis*, in the south-western Cape Province, South Africa. *Mol. Ecol.* **6**, 333–343 (1997).
28. Evans, B. J., Morales, J. C., Picker, M. D., Melnick, D. J. & Kelley, D. B. Absence of extensive introgression between *Xenopus gilli* and *Xenopus laevis laevis* (Anura: Pipidae) in Southwestern Cape province, South Africa. *Copeia* **1998**, 504–509 (1998).
29. Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
30. Kier, G. *et al.* A global assessment of endemism and species richness across island and mainland regions. *PNAS* **106**, 9322–9327 (2009).

31. Mitchell, D., Coley, P., Webb, S. & Allsopp, N. Litterfall and decomposition processes in the coastal fynbos vegetation, south-Western Cape, South Africa. *T. J. Ecol.* 977–993 (1986).
32. Picker, M. D., McKenzie, C. & Fielding, P. Embryonic tolerance of *Xenopus* (Anura) to acidic blackwater. *Copeia* 1072–1081 (1993).
33. Marsh, D. M. & Trenham, P. C. Metapopulation dynamics and amphibian conservation. *Conserv. Biol.* **15**, 40–49 (2001).
34. Daudin, F. M. *Histoire naturelle des rainettes, des grenouilles et des crapauds. Ouvrage orné de 38 planches représentant 54 espéces peintes d'après nature* (Levrault, 1802).
35. Kobel, H. R. & Tinsley, R. C. (eds.) *The Biology of Xenopus*. (Oxford University Press, 1996).
36. Furman, B. L. S. *et al*. Pan-African phylogeography of a model organism, the African clawed frog *Xenopus laevis*. *Mol. Ecol.* **24**, 909–925 (2015).
37. Measey, G. *et al*. Ongoing invasions of the African clawed frog, *Xenopus laevis*: a global review. *Biol. Invasions* **14**, 2255–2270 (2012).
38. Picker, M. D. Hybridization and habitat selection in *Xenopus gilli* and *Xenopus laevis* in the South-Western Cape Province. *Copeia* 574–580 (1985).
39. Picker, M. D., Harrison, J. & Wallace, D. Natural hybridization between *Xenopus laevis laevis* and *X. gilli* in the south-western Cape province, South Africa. In *The Biology of Xenopus* (eds. Kobel, H. R. & Tinsley, R. R.) 61–70 (Oxford University Press, Oxford, 1996).
40. Measey, G. J., de Villiers, A. L. & Soorae, P. Conservation introduction of the Cape Platanna within the Western Cape, South Africa. *Global Re-introduction Perspectives* 91–93 (2011).
41. de Villiers, F. A., de Kock, M. & Measey, G. J. Controlling the African clawed frog *Xenopus laevis* to conserve the Cape Platanna *Xenopus gilli* in South Africa. *Conserv. Evi.* **13**, 17 (2016).
42. Vogt, S., de Villiers, F.A., Ihlow, F., Rödder, D. & Measey, J. Competition and feeding ecology in two sympatric *Xenopus* species (Anura: Pipidae). *PeerJ*, **5**, e3130 (2017).
43. de Villiers, A. Species account: *Xenopus gilli* (Rose & Hewitt, 1927). In *Atlas and red data book of the frogs of South Africa, Lesotho and Swaziland* (eds. Minter, L., Burger, M., Harrison, J., Bishop, P. & Braack, H.) 260–263 (Smithsonian Institution Press, 2004).
44. Kobel, H. R., Loumont, C. & Tinsley, R. C. The extant species. In *The Biology of Xenopus* (eds. Kobel, H. R. & Tinsley, R. R.) 9–33 (Oxford University Press, Oxford, 1996).
45. Evans, B. J., Kelley, D. B., Tinsley, R. C., Melnick, D. J. & Cannatella, D. C. A mitochondrial DNA phylogeny of African clawed frogs: Phylogeography and implications for polyploid evolution. *Mol. Phylogenet. Evol.* **33**, 197–213 (2004).
46. Bewick, A. J., Anderson, D. W. & Evans, B. J. Evolution of the closely related, sex-related genes *DM-W* and *DMRT1* in African clawed frogs (*Xenopus*). *Evolution* **65**, 698–712 (2011).
47. Bowes, J. B. *et al*. Xenbase: gene expression and improved integration. *Nucleic Acids Res.* gkp953 (2009).
48. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
49. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–80 (2013).
50. Maddison, W. P. & Maddison, D. R. Mesquite: A modular system for evolutionary analysis. Version 3.04, http://mesquiteproject.org (2015).
51. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *T. Am. J. Hum. Genet.* **73**, 1162–1169 (2003).
52. Stephens, M. & Scheet, P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *T. Am. J. Hum. Genet.* **76**, 449–462 (2005).
53. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–73 (2012).
54. Nylander, J. MrModeltest v2 distributed by author. Evolutionary Biology Center, Uppsala University (2004).
55. Rambaut, A., Suchard, M. A., Xie, D. & Drummond, A. J. Tracer v1.6. http://beast.bio.ed.ac.uk/Tracer (2014).
56. Heled, J. & Drummond, A. J. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* **27**, 570–580 (2010).
57. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
58. Jakobsson, M. & Rosenberg, N. A. Clumpp: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801–1806 (2007).
59. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
60. Weiss, G. & von Haeseler, A. Inference of population history using a likelihood approach. *Genetics* **149**, 1539–1546 (1998).
61. Becquet, C. & Przeworski, M. A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* **17**, 1505–1519 (2007).
62. Becquet, C. & Przeworski, M. Learning about modes of speciation by computational approaches. *Evolution* **63**, 2547–2562 (2009).
63. Bewick, A. J., Chain, F. J. J., Heled, J. & Evans, B. J. The pipid root. *Syst. Biol.* **61**, 913–926 (2012).
64. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
65. Paradis, E. Pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).
66. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org (2015).
67. Kalinowski, S. T. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol. Ecol. Notes* **5**, 187–189 (2005).
68. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
69. Self, S. G. & Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* **82**, 605–610 (1987).
70. Evans, B. J., Bliss, S. M., Mendel, S. A. & Tinsley, R. C. The Rift Valley is a major barrier to dispersal of African clawed frogs (*Xenopus*) in Ethiopia. *Mol. Ecol.* **20**, 4216–4230 (2011).
71. Rau, R. E. The development of *Xenopus gilli* Rose & Hewitt (Anura, Pipidae). *T. Ann. South African Mus.* **76**, 247–263 (1978).
72. Dufresnes, C. *et al*. Timeframe of speciation inferred from secondary contact zones in the european tree frog radiation (*Hyla arborea* group). *BMC Evol. Biol.* **15**, 155 (2015).
73. Colliard, C. *et al*. Strong reproductive barriers in a narrow hybrid zone of West-Mediterranean green toads (*Bufo viridis* subgroup) with Plio-Pleistocene divergence. *BMC Evol. Biol.* **10**, 232 (2010).
74. Picker, M. D. *Xenopus laevis* (Anura: Pipidae) mating systems: a preliminary synthesis with some data on the female phonoresponse. *African Zool.* **15**, 150–158 (1980).
75. Davey, J. W. *et al*. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **12**, 499–510 (2011).
76. Conlon, J. M. *et al*. Evidence from peptidomic analysis of skin secretions that allopatric populations of *Xenopus gilli* (Anura: Pipidae) constitute distinct lineages. *Peptides* **63**, 118–125 (2015).
77. Greenbaum, G., Templeton, A. R., Zarmi, Y. & Bar-David, S. Allelic richness following population founding events – A stochastic modeling framework incorporating gene flow and genetic drift. *PLoS One* **9**, e115203 (2014).
78. Measey, J. *et al*. Invasive amphibians in southern Africa: a review of invasion pathways. *Bothalia-Applied Biodiv. Conserv.* **47**, a2117.
79. Caballero, A. & Garca-Dorado, A. Allelic diversity and its implications for the rate of adaptation. *Genetics* **195**, 1373–1384 (2013).

80. Wilson, J. R. *et al*. Biological invasions in the cape floristic region: history, current patterns, impacts, and management challenges. *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region* 273 (2014).
81. Pante, E. & Simon-Bouhet, B. Marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS One* **8**, e73051 (2013).

Acknowledgements
We thank Mike Picker for collaboration on early fieldwork, advice on *X. gilli* and for initiating a long-standing legacy of conservation of this species. We also thank Brian Golding for access to computational resources, Jonathan Dushoff for statistical advice, and two annonymous reviewers for helpful comments on an earlier version of this manuscript. We thank the staff at the CoGH for the efforts in preserving *X. gilli*, both genetically and its habitat. This work was supported by support to B.J.E. from the Natural Science and Engineering Research Council of Canada (RGPIN/283102-2012) and the Museum of Comparative Zoology, Harvard University, G.J.M. was funded by the National Research Foundation of South Africa (NRF Grant No. 87759) and the DST-NRF Centre of Excellence for Invasion Biology at Stellenbosch University. Research permission came from the Chief Directorate of Nature Conservation and Museums (2009/94), CapeNature (AAA007-01867), and SANParks.

Author Contributions
B.J.E. and G.J.M. collected samples. G.A.C. and B.L.S.F. generated sequence data. Analyses were performed by B.L.S.F., C.M.S.C., and B.J.E. B.L.S.F. and B.J.E. wrote the manuscript and all authors provided edits.

Additional Information
Supplementary information accompanies this paper at doi:10.1038/s41598-017-01104-9

Competing Interests: The authors declare that they have no competing interests.

Accession codes: Alignment for each locus, along with the corresponding BEAST XML files, resulting gene trees, and data for Structure analyses have been deposited in a Dryad repository (doi:10.5061/dryad.g6g2r). Genbank accession number for the sequences include the following: Rassf10: HQ221332–HQ221356, KY824194–KY824236 Sugp2: HQ221211–HQ221235, KY824150–KY824193 c7orf25: HQ220710–HQ220732, KY824433–KY824476 fem1c: HQ221309–HQ221331, KY824395–KY824432 mastl: HQ221046–HQ221070, KY824354–KY824394 mogs: HQ221071–HQ221092, KY824311–KY824353 nfil3: HQ221093–HQ221120, KP344016, KY852029–KY852072 BTBD6/p7e4: HQ220685–HQ220709, KY824477–KY824517 pigo: HQ221144–HQ221167, KY824237–KY824280 prmt6: HQ221168–HQ221178, HQ221180–HQ221190, zbed4: HQ221262–HQ221284, KY824107–KY824149, bcl9: KP345721, KP345621, KY851962–KY852028 pcdh1: HQ221129, KY824281–KY824310 16s: KP345307, KP345309–KP345313, KP345315–KP345318, KY852073–KY852133.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix D

Xenopus fraseri: Mr. Fraser, where did your frog come from?

Evans, B. J., Gansauge, M. T., Stanley, E. L., Furman, B. L., Cauret, C. M., Ofori-Boateng, C., ... & Meyer, M. (2019). *Xenopus fraseri*: Mr. Fraser, where did your frog come from?. PloS one, 14(9), e0220892.

PLOS ONE

RESEARCH ARTICLE

Xenopus fraseri: Mr. Fraser, where did your frog come from?

Ben J. Evans[1]*, Marie-Theres Gansauge[2], Edward L. Stanley[3], Benjamin L. S. Furman[1,4], Caroline M. S. Cauret[1], Caleb Ofori-Boateng[5], Václav Gvoždík[6,7], Jeffrey W. Streicher[8], Eli Greenbaum[9], Richard C. Tinsley[10], Matthias Meyer[2], David C. Blackburn[3]

1 Department of Biology, McMaster University, Hamilton, ON, Canada, **2** Max Planck Institute for Evolutionary Anthropology, Deutscher Platz, Leipzig, Germany, **3** Florida Museum of Natural History, University of Florida, Gainesville, FL, United States of America, **4** Department of Zoology, Beaty Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada, **5** Forestry Research Institute of Ghana, Kumasi, Ghana, **6** Institute of Vertebrate Biology of the Czech Academy of Sciences, Czech Republic, **7** Department of Zoology, National Museum, Prague, Czech Republic, **8** Department of Life Sciences, The Natural History Museum, London, United Kingdom, **9** Department of Biological Sciences, University of Texas at El Paso, El Paso, United States of America, **10** School of Biological Sciences, University of Bristol, Bristol, United Kingdom

* evansb@mcmaster.ca

OPEN ACCESS

Citation: Evans BJ, Gansauge M-T, Stanley EL, Furman BLS, Cauret CMS, Ofori-Boateng C, et al. (2019) *Xenopus fraseri*: Mr. Fraser, where did your frog come from? PLoS ONE 14(9): e0220892. https://doi.org/10.1371/journal.pone.0220892

Editor: Matthias Stöck, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, GERMANY

Received: July 4, 2019

Accepted: July 23, 2019

Published: September 11, 2019

Data Availability Statement: All new sequence data in this study are available on GenBank, (accessions MN259067–MN259098 for all new mitochondrial genomes except the partial sequence of CAS 146198, MN251070–MN251092 for all new Sanger sequences). An alignment of complete mitochondrial genome sequences and the Sanger sequence, including the partial sequence of CAS 146198, is provided in S1 Supplemental File. CT scans are available on MorphoSource (www.morphosource.org, doi: 10.17602/M2/M67528, 10.17602/M2/M48858, 10.

Abstract

A comprehensive, accurate, and revisable alpha taxonomy is crucial for biodiversity studies, but is challenging when data from reference specimens are difficult to collect or observe. However, recent technological advances can overcome some of these challenges. To illustrate this, we used modern approaches to tackle a centuries-old taxonomic enigma presented by Fraser's Clawed Frog, *Xenopus fraseri*, including whether *X. fraseri* is different from other species, and if so, where it is situated geographically and phylogenetically. To facilitate these inferences, we used high-resolution techniques to examine morphological variation, and we generated and analyzed complete mitochondrial genome sequences from all *Xenopus* species, including >150-year-old type specimens. Our results demonstrate that *X. fraseri* is indeed distinct from other species, firmly place this species within a phylogenetic context, and identify its minimal geographic distribution in northern Ghana and northern Cameroon. These data also permit novel phylogenetic resolution into this intensively studied and biomedically important group. *Xenopus fraseri* was formerly thought to be a rainforest endemic placed alongside species in the *amieti* species group; in fact this species occurs in arid habitat on the borderlands of the Sahel, and is the smallest member of the *muelleri* species group. This study illustrates that the taxonomic enigma of Fraser's frog was a combined consequence of sparse collection records, interspecies conservation and intraspecific polymorphism in external anatomy, and type specimens with unusual morphology.

Introduction

Alpha taxonomy–the discovery and naming of species–is the linchpin of our catalog of biodiversity. Accurate species identification and diagnosis allows us to study change in populations,

17602/M2/M49944, 10.17602/M2/M69818, 10.17602/M2/M69819, 10.17602/M2/M85216, 10.17602/M2/M85215).

Funding: Financial support for components of this work came from NSERC RGPIN-2017-05770 (BJE), NSF DEB1202609 (DCB), the Museum of Comparative Zoology, and the Max Planck Institute for Evolutionary Anthropology. EG acknowledges support from a National Geographic Research and Exploration Grant (no. 8556-08) and the US National Science Foundation (DEB-1145459). VG was supported by the Czech Science Foundation (GACR, project number 15-13415Y), Institute of Vertebrate Biology, Czech Academy of Sciences (RVO: 68081766), and Ministry of Culture of the Czech Republic (DKRVO 2019–2023/6.VII.a, National Museum, 00023272).

Competing interests: The authors have declared that no competing interests exist.

species, and communities, better understand how evolution occurs, and explore processes that drive diversification, extinction, and adaptation. Species names are generally assigned by the first publication to describe them, and rules exist for naming, revising, and synonymizing species names–for example as defined for animals by the International Code of Zoological Nomenclature (http://iczn.org/). One specimen, or a series of specimens, is designated to serve as a reference (a "type") that defines a species. In principle, comparison to type specimens allows one to attribute non-type specimens to a named species, or alternatively, to justify the recognition of a new species.

Taxonomic ambiguities may arise when type material is lost or destroyed, when intraspecific variation is high, when different species have undifferentiated morphology, and when new categories of data are collected, such as nucleotide sequences, which are difficult to assay in type material. An example of one such ambiguity is the case of Fraser's clawed frog, *Xenopus fraseri*. In 1852, two frog specimens collected by Louis Fraser (1819/20–1883; British zoologist and collector) were added to the catalog of the British Museum of Natural History (BMNH; now the Natural History Museum); in 1905, G. A. Boulenger designated these specimens to be syntypes of a new species, *Xenopus fraseri* [1]. The two type specimens of *X. fraseri* (probably males) are distinguished from all other *Xenopus* species by the combined presence of two morphological characters: (1) vomerine teeth, found only in these type specimens, *X. muelleri*, and *X. fischbergi*, and (2) prehallux claws, present in these type specimens but not in *X. muelleri* or *X. fischbergi*, and present in several other species that lack vomerine teeth (e.g., *X. allofraseri*, *X. parafraseri*) [2, 3]. The geographic origin of *X. fraseri* type specimens is listed in the BMNH catalog as "West Africa" and inferred to be Nigeria or "Fernando Po" (= Bioko Island, Equatorial Guinea) [1], or southern Benin or southwestern Nigeria [2]. The name *X. fraseri* was previously applied to populations now recognized as *X. allofraseri* and *X. parafraseri* (e.g., [3, 4]), but the combined presence of vomerine teeth and prehallux claws distinguish the syntypes of *X. fraseri* from these species [2]. A third formalin-preserved specimen that was collected in northwestern Ghana in 1975 (CAS 146198; male) was tentatively assigned to *X. fraseri* based on external morphology [2]. Thus, while distinctive molecular variation in the mitochondrial and nuclear genomes distinguishes all other species of African clawed frogs from one another (e.g., [2]), genetic data are lacking from *X. fraseri* because only two (or possibly three) specimens have been identified, and these are either old or formalin-preserved, and none has associated genetic samples. It thus remains unclear whether *X. fraseri* is in fact a distinctive species, or alternatively whether one of the more recently described species (e.g., [2]) is a synonym of *X. fraseri*. This latter possibility could arise, for example, if there is intraspecific polymorphism in the presence of either of the two morphological characters that distinguish *X. fraseri* types from all other species. Furthermore, if *X. fraseri* is distinct from other described species, both its phylogenetic position and geographic range remain unknown.

An increased understanding of *Xenopus* species diversity is important for several reasons. At the most basic level, an accurate inventory of species and their distributions allows us to monitor our planet's biosphere, including how diversification and extinction vary over time and space. There are also several implications for our understanding of evolution by allopolyploidization, and by extension, the genomic dynamics of duplicate genes. Other than the diploid *X. tropicalis*, all species of *Xenopus* are allopolyploid, including 16 allotetraploid, seven allooctoploid, and four allododecaploid species [2]. These diverse species of African clawed frogs evolved by "regular" speciation, where one ancestor diverges into two descendants, and also by "allopolyploid" speciation, where two ancestral species merge into one allopolyploid descendant species [5]. Interestingly, allopolyploidization happened independently several times in this group, and phylogenetic analyses point to the existence of three diploid and three tetraploid species that (1) existed in the past, (2) contributed their genomes to extant

tetraploid, octoploid, or dodecaploid species, and (3) do not have a known descendant with the same ploidy level as the pre-allopolyploidization ancestral species [2, 6–8]. These postulated species are the "lost ancestors" of *Xenopus* polyploids [2, 6–8], and any newly identified *Xenopus* species has the potential to be one of these lost ancestors. Discovery of one or more of these lost ancestral species could open up fascinating avenues of research that explore dynamics between each half (subgenome) of an allopolyploid genome, including mobility and suppression of transposable elements, Dobzhansky-Muller interactions, recombination, pseudogenization, subfunctionalization, and neofunctionalization (reviewed in [5]). Such studies would be catalyzed by currently available resources for *Xenopus* genomics, including two high quality genome sequences [9, 10] and powerful tools for gene editing (e. g., [11]).

Thus, to further understand the species status, phylogenetic placement, and geographic distribution of *X. fraseri*, we used sensitive techniques to capture and sequence almost complete mitochondrial genomes from both *X. fraseri* type specimens and the putative conspecific specimen. For comparative purposes, we also generated complete or almost complete mitochondrial DNA genomes from all other *Xenopus* species, and we analyzed external and internal morphology, including micro computed tomographic (μCT) scans of a *X. fraseri* type specimen, a putative conspecific, and populations of another *Xenopus* species (*X. fischbergi*, inferred to be closely related to *X. fraseri* by our phylogenetic analysis). We also followed up our phylogenetic inferences from complete mitochondrial DNA genomes, with Sanger sequencing of mitochondrial DNA from specimens we collected in the field. Our findings resolve the taxonomic quagmire presented by Mr. Fraser's frog by establishing *X. fraseri* as a distinct species that occurs in northern Ghana and northern Cameroon, and is the sister taxon to *X. fischbergi*.

Materials and methods

Targeted high-throughput sequencing of mitochondrial genomes

We used targeted high-throughput sequencing to obtain complete or almost complete mitochondrial genome sequences from all *Xenopus* species, including both of the ~170-year-old type specimens of *X. fraseri*. We additionally obtained a partial mitochondrial genome sequence from the putative conspecific specimen of *X. fraseri* (CAS 146198) and an almost complete mitochondrial genome sequence from an unusual specimen of *Xenopus* (MZUF 16294) that was purportedly collected in Eritrea. We also used Sanger sequencing to generate partial mitochondrial sequences from recently collected specimens of *X. fischbergi* individuals from Chad and the Democratic Republic of the Congo, and *X. fraseri* individuals from northern Ghana and northern Cameroon. Apart from the putative conspecific specimen of *X. fraseri* (CAS 146198), the partial mitochondrial DNA sequences spanned portions of the cytochrome c oxidase I gene, the 12S and 16S ribosomal RNA genes, and the intervening $tRNA^{val}$. Sample information is presented in S1 Table, and details of genetic data and capture probes, are provided in S1 Supporting Information.

Museum samples. Long-term storage of the four museum specimens studied here was in 60–70% ethanol or denatured alcohol, but the initial preservation treatments differed. Although we do not know what protocol was followed by Mr. Fraser when he collected the lectotype and the paralectotype specimens of *X. fraseri* (BMNH 1947.2.24.78 and BMNH 1947.2.24.79, formerly BMNH 1852.2.22.23 and BMNH 1852.2.22.24, respectively), it is likely that they were never exposed to formalin because the BMNH did not use formalin until the 1940s – 1950s (Colin McCarthy, personal communication). Liver and leg muscle were sampled from these specimens through small incisions made in the skin; the liver samples were used for targeted high-throughput sequencing. The putative conspecific specimen of *X. fraseri* (CAS 146198) was initially preserved in formalin; a sample of liver tissue from this specimen was

used. A fourth museum specimen (MZUF 16294) was probably also initially preserved with formalin; a sample of muscle tissue from this specimen was used. The fourth museum specimen, a female, was listed as originating from Tessenei, Eritrea, and collected by M. Levrini in 1956. This specimen was included because the morphology was unusual for this locality because the specimen has claws on three toes of each hind foot, but the only known species in Eritrea–*X. clivii*–has four (including one on the prehallux). This specimen also served as a technical replicate for the performance of our targeted high-throughout sequencing on formalin preserved tissues.

Assembly and inference of complete mitochondrial genomes

Reads were demultiplexed based on exact matches to sample-specific barcodes. Overlapping paired-end reads were then merged into consensus reads when possible. A *de novo* assembly of the complete mitochondrial genome was attempted for each sample with Trinity version 2.5.1 [12]. The advantage of a *de novo* assembly over a reference-based assembly is that the former approach may more accurately reconstruct insertion deletion polymorphisms, and presumably is less biased by a reference sequence.

The *de novo* assembly produced one ~17kb contig for the entire mitochondrial genome for almost all of the non-museum samples (28 out of 29), but none of the museum samples. For 5 of the 29 non-museum samples (*X. tropicalis*, *X. mellotropicalis*, *X. victorianus*, *X. laevis*, and *X. wittei*) the initial *de novo* assembly produced an assembly with two concatenated mitochondrial genomes. One sample (*X. poweri*) produced a fragmented assembly, and one of the two *X. boumbaensis* samples assembled only a small portion of the mitochondrial genome. For the five samples that initially produced concatenated mitochondrial genomes assembly, the *de novo* assembly was repeated using a k-mer size equal to either 21 or 32 instead of the initial setting of 29. For *X. victorianus* and *X. mellotropicalis*, a k-mer size equal to 21 provided a non-concatenated assembly, and the same for *X. laevis* and *X. wittei* with a k-mer size equal to 32. For *X. poweri*, the assembly was fragmented into 4 contigs that were independently aligned to the other *de novo* genomes with the others using Mafft [13] using the "—adjustdirection" option to align reverse-complemented contigs. The *X. poweri* sequence was then concatenated manually into a complete genome. For the *X. boumbaensis* sample that did not fully assemble *de novo*, we instead used the other *X. boumbaensis de novo* mitochondrial genome sequence as a reference, and generated a consensus sequence from reads that were mapped to this reference as described below for the museum samples.

Because *de novo* assembly of the museum samples failed to produce a complete mitochondrial genome, we instead generated a consensus sequence from unique and de-duplicated reads from each sample that were mapped to a reference mitochondrial genome. For each sample, we explored the effect of mapping to three different reference genomes that we generated *de novo* (*X. laevis*, *X. parafraseri*, *X. fischbergi*). We used as a reference the genome that yielded the most complete consensus, which was the *X. fischbergi* mitochondrial genome for the *X. fraseri* lectotype and paralectotype specimens, for the putative *X. fraseri* conspecific, and the *X. laevis* mitochondrial genome for the Eritrea specimen. Consensus calling on mapped reads requiring a coverage of at least 5x for each base, at least 80% agreement of the genotype, and a minimum map quality of 25.

Phylogenetic analysis

To evaluate the phylogenetic position of the mitochondrial genome sequences from *X. fraseri*, we performed phylogenetic analysis on these sequences along with previously published mitochondrial genome sequences of several species in Pipoidea (the clade that contains the sister

families Rhinophrynus and Pipidae) including *Rhinophrynus dorsalis* (HM991334.1), *Pipa carvalhoi* (HM991332.1) and *P. pipa* (GQ244477.1), *Hymenochirus boettgeri* (HM991331.1), and *Pseudhymenochirus merlini* (HM991333.1), *X.* cf. *tropicalis* (AP014695), *X. tropicalis* (AY789013), and *X. borealis* (JX155859). We excluded four regions from the alignment of these data where we deemed homology to be ambiguous based on visual inspection. These regions included the entire $tRNA^{Pro}$ and D-loop (positions 15,500–17,610 of AY789013.1), two portions of the 16S rDNA gene (positions 1099–1140 and 2214–2228 of AY789013.1), and a portion of the origin for light strand replication (positions 5231–5248 of AY789013.1). To explore the effects of missing data, we performed separate phylogenetic analysis on an alignment that included only the complete or almost complete mitochondrial DNA genomes, and on an alignment that included these data plus the partial mitochondrial DNA genome sequences (S1 Table, S1 Supplemental File).

IQ-TREE version 1.6.8 [14, 15] was used for maximum likelihood phylogenetic analysis and model selection (the GTR+F+I+G4 model was selected for both datasets–with and without the partial mitochondrial sequences–according to the Bayesian Information Criterion), and ultrafast bootstrap analysis (with 1,000 replicates) was used to assess support for this topology. We also analyzed the both datasets with BEAST version 2.52 [16], using the same model of evolution listed above, a Yule model for the tree prior, and assuming an uncorrelated lognormal relaxed molecular clock. Three calibration points were used, all with a normally distributed prior probability distribution, and all following estimates in [17]: (1) the age of genus *Xenopus*, which subtends the subgenera *Silurana* and *Xenopus* [2], was set to a mean of 45.3 million years (my), and a standard deviation of 5.7 my, (2) the age of family Pipidae was set to a mean of 117.6 and a standard deviation of 6.6 my, and (3) the age of Pipoidea was set to a mean of 159.4 my and a standard deviation of 6.0 my. For each BEAST analyses, 4 or 8 independent runs were performed respectively, each for 10 million generations, sampling every thousand generations, and starting from a random tree. Twenty-five percent of each run was discarded as burn-in. Tracer version 1.71 [18] was used to verify that the effective sample sizes of all parameters exceeded 200.

Morphology

We previously collected high-resolution X-ray μCT scans of several *Xenopus* specimens, including the paralectotype of *X. fraseri* and the holotype of *X. fischbergi* [2]. We used the same methodology to generate a scan of the putative *X. fraseri* specimen from Ghana (CAS 146198) and several more specimens of *X. fischbergi*, including one adult of each sex and two juveniles from Chad (specimens AMNH-H-A158343, AMNH-H-A158360, AMNH-H-A158350, and AMNH-H-A158356). These specimens were selected to capture variation between the sexes and during development. We lacked data from these post-metamorphic individuals from Chad, but inferred them to be *X. fischbergi* based on sequence from another individual (a tadpole) that was collected at the same time and localition (AMNH A-158377; S1 Table). We additionally examined external morphology and measured snout-vent lengths from several individuals for which we had sequence data to confirm species identification as *X. fraseri* or *X. fischbergi* (the sister taxon of *X. fraseri*, see below), and we also took measurements from the two adult *X. fischbergi* specimens from Chad.

Ethics statement

All procedures involving live animals have been approved by the Animal Use Committee at McMaster University (AUP 17-12-43).

Results

Xenopus fraseri is distinct from all other species, the sister species of *X. fischbergi*, and minimally distributed in northern Ghana and northern Cameroon

Using targeted high-throughput sequencing, we obtained complete or almost complete mtDNA sequences from at least one representative of all *Xenopus* species, and including three of the four museum samples, and including both of the type specimens of *X. fraseri* (average length of all but CAS 146198 was 17,192 base pairs (bp), the range was 14,204–17,833 bp). A partial (5,930 bp) mitochondrial genome sequence was obtained from the more recently collected putative *X. fraseri* specimen from Ghana (CAS 146198). Phylogenetic analysis of these genomes indicates that mitochondrial DNA from both type specimens of *X. fraseri* are closely related, substantially diverged from other *Xenopus* species, and sister to mitochondrial sequences of *X. fischbergi* (Fig 1, S1 Fig).

Analysis of complete or almost complete mitochondrial DNA genomes were inferred by Bayesian and maximum likelihood phylogenetic methods identified no strongly supported topological differences, and most nodes had strong posterior probability and bootstrap support (Fig 1, S1 Fig). The BEAST analysis indicates that the mitochondrial genomes of *X. fraseri* and *X. fischbergi* diverged ~4.7 million years ago (mya; 95% highest probability density interval 2.8–7.1 mya).

These findings in mind, we then examined other wild-caught individuals sampled throughout the putative ranges of *X. fischbergi* and *X. fraseri* in five countries in West and Central Africa: the Democratic Republic of the Congo, Cameroon, Nigeria, Chad, and Ghana. Phylogenetic analysis of partial mitochondrial DNA Sanger sequences from these individuals combined with the mitochondrial genome sequences described above demonstrates that *X. fraseri* is distributed in northern Ghana and northern Cameroon, including CAS 146198, whereas *X. fischbergi* occurs in the Democratic Republic of the Congo, Chad, and Nigeria (Fig 1, S1 Fig). In both of these analyses (complete mitochondrial genomes with or without the partial sequences), the phylogenetic position of *X. fraseri* places it unambiguously within the *muelleri* species group, which includes as well *X. muelleri*, *X. borealis*, and *X. fischbergi* [2]. The relatively recent divergence of *X. fraseri* within the *muelleri* species group indicates that it is not one of the lost ancestors, and suggests that *X. fraseri* could be an allotetraploid species with 36 chromosomes, although we did not generate a karyotype to assess this.

Divergence in mitochondrial DNA

Molecular divergence between the mtDNA sequences of the lectotype and paralectotype of *X. fraseri* was low: the uncorrected pairwise divergence (hereafter, divergence) was 0.11% out of 13,909 non-missing and non-gapped base pairs (hereafter, positions). Divergence between each of these sequences and the putative *X. fraseri* conspecific was also low: 0.22% out of 5,889 positions. But the complete genome of *X. fischbergi* from Nigeria was substantially diverged from almost complete mitochondrial genomes of the lectotype and paralectotype specimens of *X. fraseri* (4.37% or 4.53% out of 13,959 or 15,350 positions, respectively) and 2.7% diverged from the partial mitochondrial genome from the putative *X. fraseri* conspecific (out of 5889 positions). The smaller portions of the mitochondrial genome of *X. fischbergi* from Chad were also substantially diverged from the lectotype and paralectotype specimens of *X. fraseri* individuals (3.43% and 3.59% out of 2,796 or 2,842 positions, respectively) and from the partial mitochondrial genome from the putative *X. fraseri* conspecific (2.35% out of 2003 positions).

These divergences between *X. fraseri* and *X. fischbergi* exceed that between several other allotetraploid *Xenopus* sibling species pairs such as *Xenopus petersii* and *X. victorianus* (these

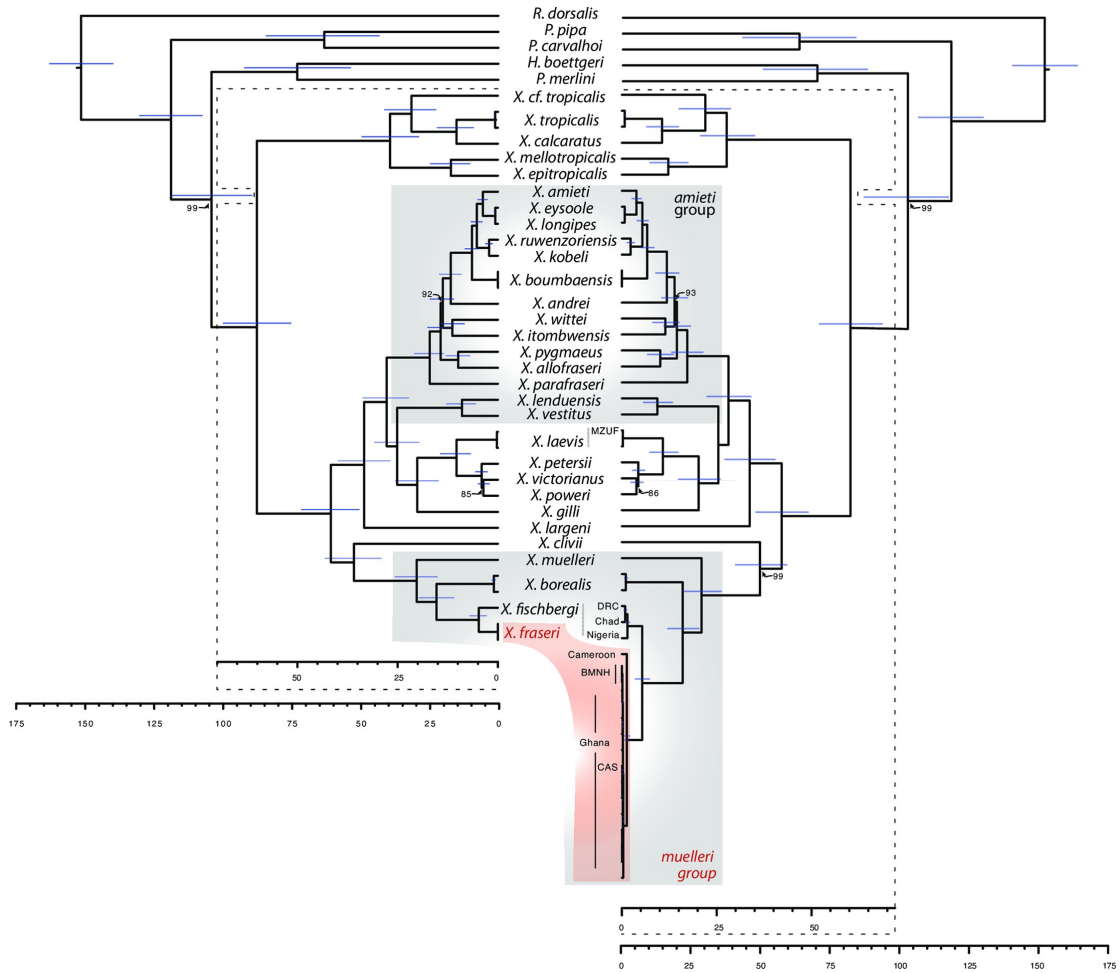


Fig 1. Baysian phylogenetic analysis of *Xenopus* mitochondrial DNA complete or almost complete genomes (left), and these data plus partial mitochondrial sequences (right). Two almost complete genomes from the lectotype and paralectotype of *X. fraseri* were included in both analyses; these genomes are labeled BMNH on the right and unlabeled on the left. All other *Xenopus* species have at least one complete or almost complete mitochondrial DNA genome. GenBank mitochondrial genomes from *X. borealis* and *X. tropicalis* were almost identical to ones generated, and the mitochondrial genome of specimen MZUF 16294 (MZUF), also unlabeled on the left, is almost identical to the mitochondrial genome we generated for *X. laevis*. The putative *X. fraseri* specimen CAS 146198 (CAS), type specimens of *X.* fraseri, and samples from Ghana are almost identical to one another. All nodes have 100% posterior probability except where indicated. Node bars indicate 95% highest posterior density for the node age. Scale bars are in millions of years.

https://doi.org/10.1371/journal.pone.0220892.g001

two species are 3.43% diverged over their whole mitochondrial genomes, 1.87% over the portion of the genome that was sequenced for the putative *X. fraseri* conspecific, and 1.36% over the portion of the mitochondria that was sequenced for the *X. fischbergi* individual from Chad). Divergences between various dodecaploid species pairs (e.g., *X. kobeli* and *X. ruwenzoriensis*) are even lower than that observed between *X. petersii* and *X. victorianus* (Fig 1, S1 Fig).

Divergence among mitochondrial DNA of *X. tropicalis* and *X.* cf. *tropicalis* is considerably larger than most other species owing to an unusual haplotype from Liberia [19, 20] that may constitute a distinct species [19] (with caveats based on one nuclear gene [2, 21]).

Role of reference genome in consensus sequences

Although most of the mitochondrial genomes reported here were assembled *de novo*, the mitochondrial DNA genomes for the museum samples were generated by mapping capture data to a closely related reference genome. A possible concern with this approach is that the reference-based consensus sequences for the museum samples might be affected by which reference genome was used. To explore this, we compared the consensus sequences from each *X. fraseri* type specimen that were generated when either *X. fischbergi* or *X. laevis* was used. This comparison demonstrated that the consensus sequences were minimally biased, and far more complete when a closely related reference genome was used compared to a more distantly related reference genome (Supplemental Results). Our analyses suggest that the reference genome imposes only a modest bias on consensus sequences, but to the extent that it does, our figures for divergence between the mitochondrial genomes of *X. fraseri* and *X. fischbergi* may be slight underestimates.

Comparative anatomy

Comparison of external and internal anatomy of the putative *X. fraseri* conspecific CAS 146198 to the *X. fraseri* paralectotype (BMNH 1947.2.24.79) and the holotype of *X. fischbergi* (CAS 255060) confirms the presence of vomerine teeth (insets in Fig 2). A pointed keratinized prehallux claw is absent in *X. fischbergi* and appears to be an intraspecific polymorphism in *X. fraseri* that distinguishes some specimens of *X. fraseri* from specimens of other species (Supplementary Results; S2 Fig). The small size (for the *muelleri* species group [2]) of the lectotype and paralectotype of *X. fraseri* and the *X. fraseri* from Ghana (S2 Table) could suggest that they are not fully mature. However, examination of 13 other adult *X. fraseri* individuals from Ghana indicates that this species is indeed substantially smaller than *X. fischbergi*. The average SVLs of seven adult female and six adult male *X. fraseri* were 46.8 mm and 38.5 mm, respectively (ranges 39.9–54.2 and 36.9–32.5 mm respectively; S2 Table). The average SVL of the holotype and two paratypes (all female) of *X. fischbergi*, and two other female specimen from the Democratic Republic of the Congo (UTEP21194), is 57.8 mm (S2 Table), which is 25% larger than the *X. fraseri* females we measured. An adult male *X. fischbergi* we measured was larger than all seven males of *X. fraseri* that we measured (S2 Table).

Subtle differences in skeletal morphology are suggested by μCT scans. In the ventral view, the optic foramina of the *X. fischbergi* holotype and other specimens have more rounded margins on the fronto-parietal bone, especially in the adult specimens, compared to the *X. fraseri* paralectotype and putative conspecific (Fig 2, S3 Fig). However, the small sample of μCT scans prevents us from assessing whether these differences are distinctive in each species.

Other phylogenetic insights

Our analysis of complete mitochondrial genome sequences from all *Xenopus* species provides strong statistical support for several novel or previously poorly supported relationships among mitochondrial DNA lineages in *Xenopus* (e.g., [2]), including strong support for (i) a sister relationship between *X. largeni* and a clade comprising mitochondrial lineages of species in the *laevis* and *amieti* species groups, and methodologically variable support (strong for Bayesian but modest for maximum likelihood) for (ii) a sister relationship between mitochondrial DNA from (*X. lenduensis* + *X. vestitus*) and species in the *laevis* species group, that renders

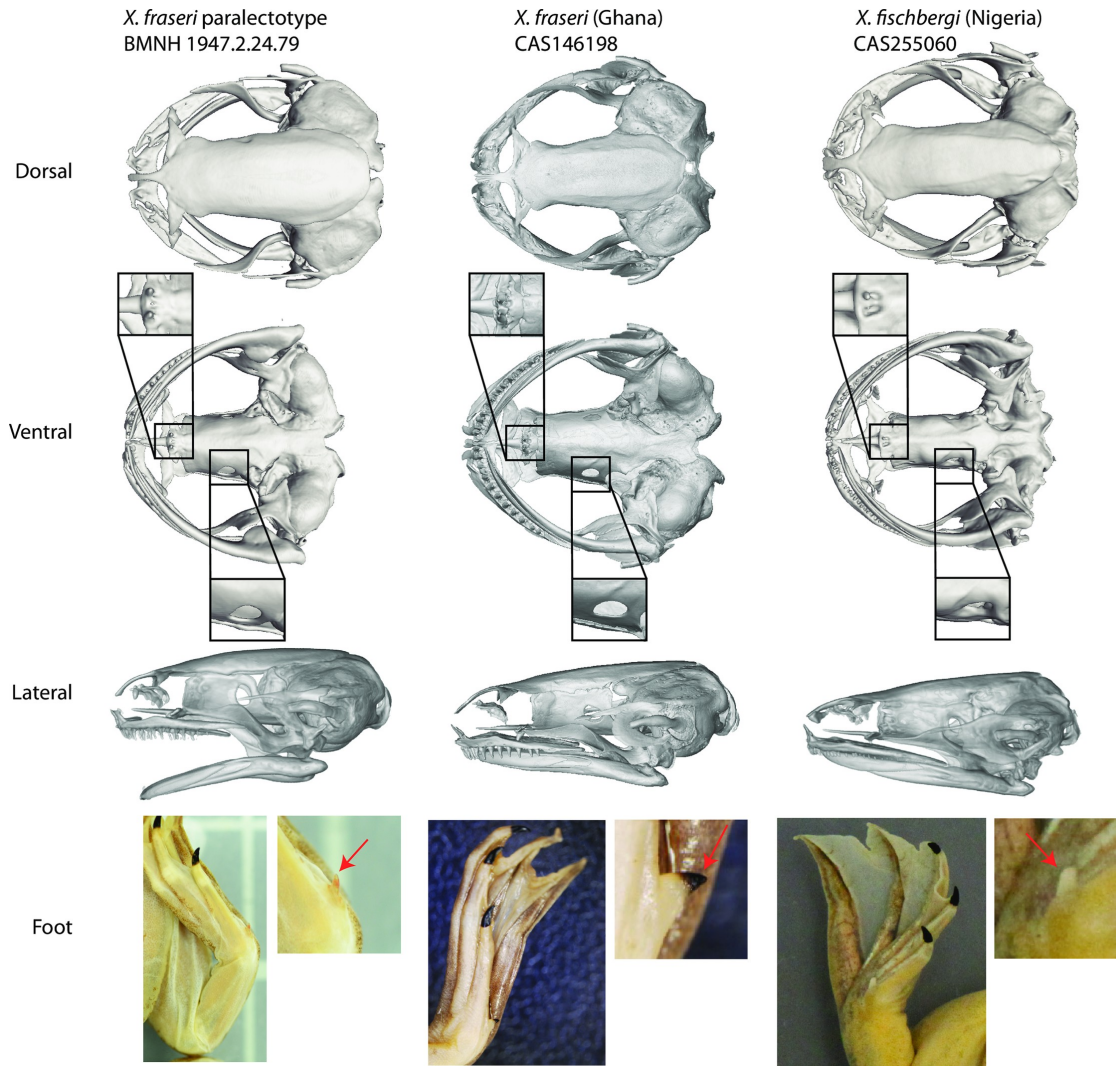


Fig 2. µCT scans and external morphology of the paralectotype of *X. fraseri* (BMNH 1947.2.24.79; probably male, SVL = 32.5 mm), a *X. fraseri* conspecific (CAS 146198; male, SVL = 34.1 mm), and the holotype of *X. fischbergi* (CAS 255060; female, SVL = 51.7 mm). On the ventral view, vomerine teeth and optic foramina are enlarged.

https://doi.org/10.1371/journal.pone.0220892.g002

mitochondrial DNA of the *amieti* group paraphyletic, and (iii) a sister relationship between *X. clivii* and species in the *muelleri* group (including *X. fraseri*). The complete mitochondrial genome from a museum specimen that was reportedly collected in Eritrea (MZUF 16294) was essentially identical to a complete mitochondrial genome from *X. laevis* from South Africa. This suggests that this specimen either was introduced to Eritrea from southern Africa where

X. laevis occurs, or (we suspect) was mislabeled, and actually originated from South Africa. This specimen indicates that formalin preserved material has the potential to yield complete mitochondrial DNA genomes, even though we did not obtain one from CAS 146198.

Discussion

Although we do not advocate using divergence of mitochondrial DNA as the sole justification for species distinctiveness, these data coupled with morphological characteristics discussed above, including the substantially smaller size of *X. fraseri* compared to other closely related species (S2 Table, [2]), and a pointed prehallux in some *X. fraseri* individuals, argues for the distinctiveness of *X. fraseri*. The specimens of *X. fraseri* were sampled from 4 different localities, and internal examination of one female indicated that she was gravid. This suggests the smaller size of these individuals compared to *X. fischbergi* is unlikely to be due to ontogeny. Our findings indicate that the sister taxon of *X. fraseri* is *X. fischbergi*, and that *X. fraseri* is a member of the muelleri species group along with *X. muelleri*, *X. fischbergi*, and *X. borealis* (Fig 1, S1 Fig). The minimal distribution of *X. fraseri*, established in this study, is in northern Ghana and Cameroon. *Xenopus fraseri* may be partially sympatric with *X. fischbergi*, the latter of which ranges at least from Nigeria to the northern Democratic Republic of the Congo (based on mitochondrial sequences), and possibly even farther west and east of this [2]. Given the similarity in external morphology of *X. fraseri* and *X. fischbergi*, our previous delineation of the range of *X. fischbergi* [2] may be an overestimate and, based on genetic information, we reassign several paratypes of *X. fischbergi* to *X. fraseri*, as detailed in the Supplemental Results.

Our recent systematic revision of West African *Xenopus* [2] included sequence data from two genetic samples that were identified here to be *X. fraseri*, but we failed to identify them then as being distinctive from *X. fischbergi* [2]. At that time, we did not realize that the shape and morphology of the prehallux was so variable within *X. fraseri*. The prehalluxes of the *X. fraseri* types and the formalin-preserved specimen from Ghana (CAS 146198) are the most keratinized specimens we have encountered (S2 Fig), so the type specimens of *X. fraseri* are in fact not particularly "typical" of this species in this regard. Additionally, because the mitochondrial DNA data in [2] was from a relatively slowly evolving portion of this genome, the divergence between the *X. fraseri* and *X. fischbergi* data in [2] was lower (~3%) than that between the almost complete mitochondrial genomes that were generated for this study (>4%; Supplemental Results).

In species with low dispersal rates, divergent and geographically separated intra-specific clades with narrow regions of overlap may arise without geographic barriers to gene flow [2]; this is seen, for example, in *X. laevis* from southwest South Africa [22]. In contrast, it is rare in species with low dispersal rates–including in *Xenopus* [23]–for diverged intra-specific clades of mitochondrial DNA to have broadly overlapping geographic distributions. This phylogeographic pattern is instead suggestive of cryptic species [2, 23], and the new genetic and morphological information from *X. fraseri* presented here allowed us to distinguish this cryptic species from another closely related, morphologically similar, and possibly co-distributed or parapatric species (*X. fischbergi*).

Other phylogenetic insights

The inference of a closer phylogenetic affinity of the mitochondrial genome of *X. clivii* to the mitochondrial genomes of the *muelleri* species group than to those of the *laevis* or *amieti* species groups is in line with inferences based on complete transcriptomes [24]. This observation is of interest for studies of sex chromosomes because *X. clivii* and several species in the *laevis* and *amieti* species group carry the female-specific sex determining gene *dm-w* [25], whereas

X. borealis does not carry *dm-w*, and has a newly evolved sex chromosome [26, 27]. This phylogenetic relationship of mitochondrial DNA, which is (also) maternally inherited, indicates that *dm-w* was lost at some point in the ancestry of *X. borealis* after the divergence of *X. clivii* [25].

Conclusions

Because he died over a century ago, Mr. Fraser is unavailable to tell us exactly where he collected his frog. Despite the uncertainty over the origin of Mr. Fraser's specimens, it has long been assumed that *X. fraseri* is distributed in lowland tropical forest [25]. This was reinforced by the many records of morphologically similar forest-dwelling specimens, including those now called *X. parafraseri* and *X. allofraseri* [28]. In fact, *X. fraseri* is distributed in relatively hot, arid savanna, including the borders of the Sahel. We argued previously based on historical records that the lectotype and paralectotype specimens of *X. fraseri* were probably collected in southern Benin or southwestern Nigeria [2] in areas that are in or near the Dahomey Gap, an area of savanna extending south from the Sahel that subdivides tropical rainforest habitat of West Africa, and that is poorly studied for *Xenopus*. Consistent with this proposal, the localities of this species identified here–northern Cameroon and northern Ghana–are in the same ecoregions as the Dahomey Gap (West Sudanian savanna and Guinean forest-savanna mosaic) [2]. Previous confusion about *X. fraseri* arose as a consequence of poor records of the provenance of the type specimens, the generally conserved interspecific morphology of several *Xenopus* species, high-intraspecific polymorphism of the prehallux of *X. fraseri*, and the unusual prehallux morphology of the *X. fraseri* type specimens. This study thus provides clarity to this taxonomic enigma presented by Fraser's frog by identifying distinctive molecular and morphological features of this species–as well as polymorphic aspects, and by delineating its phylogenetic position and geographic distribution.

Supporting information

S1 Supporting Information. Supplemental methods and results for this study.
(DOCX)

S1 Table. Species, sample and museum identification numbers, origin, and sex of genetic samples analyzed in this study.
(XLSX)

S2 Table. Snout-vent length of *X. fraseri* and *X. fischbergi* specimens (Voucher ID), including type and non-type specimens (Type Status) that are male or female (Sex).
(XLSX)

S1 Fig. Maximum likelihood bootstrap consensus tree of *Xenopus* mitochondrial DNA complete or almost complete genomes (left), and these data plus partial mitochondrial sequences (right). Bootstrap support is 100 percent except where indicated.
(TIF)

S2 Fig. Prehallux morphology of *X. fraseri*. Some *X. fraseri* individuals have a rounded prehallux (left two images), a slightly pointed prehallux (center), or a pointed and keratinized prehallux (right two images). Thus, a pointed and/or keratinized prehallux is a distinguishing, but not universal, characteristic of *X. fraseri*.
(TIF)

S3 Fig. μCT scans of four *X. fischbergi* specimens from Chad. On the ventral view, vomerine teeth and optic foramina are highlighted as in Fig 2. In the two juvenile specimens (right), one or both vomerine teeth are not X-ray opaque and are not visible on these scans, but they are

present.
(TIF)

S1 Supplemental File. Alignment of mitochondrial genome sequences analyzed in this study in nexus format.
(NEXUS)

Acknowledgments

We thank Lauren Scheinberg and Annamaria Nistri, the Museum of Natural History, and the California Academy of Sciences, and Museo di Storia Naturale dell'Università di Firenze sezione di Zoologia "La Specola" for allowing us to sample tissues from their specimens. We thank José Rosado for measurning *X. fraseri* specimens from Ghana. We thank Frank Burbrink, David Kizirian, and Lauren Vonnahme for a loan of *X. fischbergi* specimens from Chad, Adam Leache for a loan of a tissue sample of *X. fraseri* from Ghana, Ben Vernot for assistance with capture probe design, and A. Hánová, F. Snítilý and M. Dolinay for help in the field and/or laboratory, the Museum of Comparative Zoology at Harvard University for supporting a collection trip in Ghana, and Birgit Nickel for performing laboratory work associated with mitochondrial and nuclear DNA capture. We are grateful to Svante Pääbo and the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany for hosting BJE during a sabbatical, during which time this project was initiated and collection of molecular data performed.

Author Contributions

Conceptualization: Ben J. Evans, Matthias Meyer, David C. Blackburn.

Data curation: Ben J. Evans, Marie-Theres Gansauge, Edward L. Stanley.

Formal analysis: Ben J. Evans.

Funding acquisition: Ben J. Evans.

Investigation: Ben J. Evans, Benjamin L. S. Furman, Caroline M. S. Cauret, Caleb Ofori-Boateng, Václav Gvoždík, Jeffrey W. Streicher, Eli Greenbaum, Richard C. Tinsley.

Methodology: Ben J. Evans.

Project administration: Ben J. Evans.

Supervision: Ben J. Evans.

Visualization: Edward L. Stanley, David C. Blackburn.

Writing – original draft: Ben J. Evans.

Writing – review & editing: Ben J. Evans, Marie-Theres Gansauge, Edward L. Stanley, Benjamin L. S. Furman, Caroline M. S. Cauret, Caleb Ofori-Boateng, Václav Gvoždík, Jeffrey W. Streicher, Eli Greenbaum, Richard C. Tinsley, Matthias Meyer, David C. Blackburn.

References

1. Boulenger GA. On a collection of batrachians and reptiles made in South Africa by Mr. C. H. B. Grant, and presented to the British Museum by Mr. C. D. Rudd. Proceedings of the Zoological Society of London. 1905; 1905:248–55.
2. Evans BJ, Carter TF, Greenbaum E, Gvoždík V, Kelley DB, McLaughlin PJ, et al. Genetics, morphology, adverstisement calls, and historical records distinguish six new polyploid species of African clawed frog (*Xenopus*, Pipidae) from West and Central Africa. PLoS One. 2015; 10(12):e0142823. https://doi.org/10.1371/journal.pone.0142823 PMID: 26672747

3. Kobel HR, Loumont C, Tinsley RC. The extant species. In: Tinsley RC, Kobel HR, editors. The Biology of *Xenopus*. Oxford: Clarendon Press; 1996. p. 9–33.
4. Conlon JM, Mechkarska M, Kolodziejek J, Nowotny N, Coquet L, Leprince J, et al. Host-defense peptides from skin secretions of Fraser's clawed frog *Xenopus fraseri* (Pipidae): Further insight into the evolutionary history of the Xenopodinae. Comparative Biochemistry and Physiology, Part D Genomics Proteomics. 2014; 12:45–52. https://doi.org/10.1016/j.cbd.2014.10.001 PMID: 25463057
5. Evans BJ. Genome evolution and speciation genetics of allopolyploid clawed frogs (*Xenopus* and *Silurana*). Front Biosci. 2008; 13:4687–706. PMID: 18508539
6. Evans BJ, Carter TF, Hanner R, Tobias ML, Kelley DB, Hanner R, et al. A new species of clawed frog (genus *Xenopus*), from the Itombwe Plateau, Democratic Republic of the Congo: Implications for DNA barcodes and biodiversity conservation. Zootaxa. 2008; 1780:55–68.
7. Evans BJ, Greenbaum E, Kusamba C, Carter TF, Tobias ML, Mendel SA, et al. Description of a new octoploid frog species (Anura: Pipidae: *Xenopus*) from the Democratic Republic of the Congo, with a discussion of the biogeography of African clawed frogs in the Albertine Rift. J Zool, Lond. 2011; 283:276–90.
8. Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC. A mitochondrial DNA phylogeny of clawed frogs: Phylogeography on sub-Saharan Africa and implications for polyploid evolution. Mol Phylogenet Evol. 2004; 33:197–213. PMID: 15324848
9. Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, et al. Genome evolution in the allotetraploid frog Xenopus laevis. Nature. 2016; 538(7625):336–43. https://doi.org/10.1038/nature19840 PMID: 27762356; PubMed Central PMCID: PMC5313049.
10. Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kaptonov V, et al. The genome of the western clawed frog *Xenopus tropicalis*. Science. 2010; 328:633–6. https://doi.org/10.1126/science.1183670 PMID: 20431018
11. Tandon P, Conlon F, Furlow JD, Horb ME. Expanding the genetic toolkit in *Xenopus*: Approaches and opportunities for human disease modeling. Dev Biol. 2016; 426(2):325–35. https://doi.org/10.1016/j.ydbio.2016.04.009 PMID: 27109192.
12. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011; 29(7):644–52. https://doi.org/10.1038/nbt.1883 PMID: 21572440; PubMed Central PMCID: PMC3571712.
13. Katoh K, Standley SM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. MBE. 2013; 30:772–80.
14. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013; 30(5):1188–95. https://doi.org/10.1093/molbev/mst024 PMID: 23418397; PubMed Central PMCID: PMC3670741.
15. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32(1):268–74. https://doi.org/10.1093/molbev/msu300 PMID: 25371430; PubMed Central PMCID: PMC4271533.
16. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. MBE. 2012; 29:1969–73.
17. Feng YJ, Blackburn DC, Liang D, Hillis DM, Wake DB, Cannatella DC, et al. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. PNAS. 2017; 114(29):E5864–E70. https://doi.org/10.1073/pnas.1704632114 PMID: 28673970; PubMed Central PMCID: PMC5530686.
18. Rambaut A, Drummond AJ. Tracer v1.5, Available from http://beast.bio.ed.ac.uk/Tracer. 2007.
19. Haramoto Y, Oshima T, Takahashi S, Asashima M, Ito Y, Kurabayashi A. Complete mitochondrial genome of "*Xenopus tropicalis*" Asashima line (Anura: Pipidae), a possible undescribed species. Mitochondrial DNA. 2016; 6(27):3341–3.
20. Evans BJ, Cannatella DC, Melnick DJ. Understanding the origins of areas of endemism in phylogeographic analyses: a reply to Bridle *et al*. Evolution. 2004; 58(6):1397–400.
21. Evans BJ, Kelley DB, Melnick DJ, Cannatella DC. Evolution of RAG-1 in polyploid clawed frogs. MBE. 2005; 22(5):1193–207.
22. Irwin DE. Phylogeographic breaks without geographic barriers to gene flow. Evolution. 2002; 56 (12):2383–94. PMID: 12583579
23. Furman BL, Bewick AJ, Harrison TL, Greenbaum E, Gvozdik V, Kusamba C, et al. Pan-African phylogeography of a model organism, the African clawed frog 'Xenopus laevis'. Mol Ecol. 2015; 24(4):909–25. https://doi.org/10.1111/mec.13076 PMID: 25583226.
24. Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winkley K, et al. Cryptic species as a window on diversity and conservation. TREE. 2006; 22(3):148–55. PMID: 17129636

25. Furman BLS, Evans BJ. Sequential turnovers of sex chromosomes in African clawed frogs (*Xenopus*) suggest some genomic regions are good at sex determination. G3. 2016; 6:3625–33. https://doi.org/10.1534/g3.116.033423 PMID: 27605520

26. Yoshimoto S, Okada E, Umemoto H, Tamura K, Uno Y, Nishida-Umehara C, et al. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. PNAS. 2008; 105 (7):2469–74. https://doi.org/10.1073/pnas.0712244105 PMID: 18268317

27. Bewick AJ, Anderson DW, Evans BJ. Evolution of the closely related, sex-related genes DM-W and DMRT1 in African clawed frogs (*Xenopus*). Evolution. 2011; 65(3):698–712. https://doi.org/10.1111/j.1558-5646.2010.01163.x PMID: 21044062

28. Tinsley RC, Loumont C, Kobel HR. Geographical distribution and ecology. In: Tinsley RC, Kobel HR, editors. The Biology of *Xenopus*. Oxford: Clarendon Press; 1996. p. 35–59.

Appendix E

A frog with three sex chromosomes that co-mingle together in nature: *Xenopus tropicalis* has a degenerate W- and a Y- that evolved from a Z-

This manuscript has been submitted and is under review.

A frog with three sex chromosomes that co-mingle together in nature: *Xenopus tropicalis* has a degenerate W- and a Y- that evolved from a Z-

Benjamin L. S. Furman[1,2], Caroline M. S. Cauret[1], Martin Knytl[1,3], Xue-Ying Song[1], Tharindu Premachandra[1], Caleb Ofori-Boateng[3], Danielle C. Jordan[4], Marko E. Horb[4], and Ben J. Evans[1,*]

[1]Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada

[2]Department of Zoology, University of British Columbia, 6270 University Blvd Vancouver, British Columbia, V6T 1Z4 Canada

[3]Department of Cell Biology, Charles University, 7 Vinicna Street, Prague, 12843, Czech Republic

[4]CSIR-Forestry Research Institute of Ghana, Kumasi, Ghana

[5]Eugene Bell Center for Regenerative Biology and Tissue Engineering and National *Xenopus* Resource, Marine Biological Laboratory, 7 MBL St, Woods Hole, MA 02543 USA

[*]Correspondence: evansb@mcmaster.ca

Abstract

In many species, sexual differentiation is a vital prelude to reproduction, and disruption of this process can have severe fitness effects, including sterility. It is thus interesting that genetic systems governing sexual differentiation vary among – and even within – species. To understand these systems more, we investigated a rare example of a frog with three sex chromosomes: the Western clawed frog, *Xenopus tropicalis*. We demonstrate that natural populations from the western and eastern edges of Ghana have a young Y-chromosome, and that a male-determining factor on this Y-chromosome is in a very similar genomic location as a previously known a female-determining factor on the W-chromosome. Compared to

the rest of the genome, a small sex-associated portion of the sex chromosomes exhibits higher nucleotide differentiation and has a 50-fold enrichment of transcripts with male-biased expression during early gonadal differentiation. Nucleotide polymorphism of expressed transcripts suggests regulatory degeneration on the W-chromosome, emergence of a new Y-chromosome from an ancestral Z-chromosome, and natural co-mingling of the W-, Z-, and Y-chromosomes in the same population. Additionally, *Xenopus tropicalis* has sex-differences in the rates and genomic locations of recombination events during gametogenesis that are similar to at least two other *Xenopus* species, which suggests that sex differences in recombination are genus-wide. These findings are consistent with theoretical expectations associated with recombination suppression on sex chromosomes, demonstrate that several characteristics of old and established sex chromosomes (e.g., nucleotide divergence, sex biased expression) can arise well before a sex chromosomes become cytogenetically distinguished, and show how these characteristics can have lingering consequences that are carried forward through sex chromosome turnovers.

Author Summary

Sex chromosomes often come in pairs (an X and a Y, or a Z and a W) and variation among species evidences widespread rapid evolutionary change of sex chromosomes. To understand why, we examined a rare example of a frog (*Xenopus tropicalis*) with three sex chromosomes. We discovered a small sex-linked sliver of the genome that has a high proportion of genes with higher expression in males than females during gonadal differentiation. Molecular variation in expressed transcripts from this genomic region suggests that this pattern stems from decreased or lost expression of alleles on the W-chromosome combined with a recent origin of the Y-chromosome from an ancestral Z-chromosome. These findings are consistent with theoretical expectations associated with reduced genetic recombination, and demonstrate that features of ancestral chromosomes have persistent genomic effects that bleed through sex chromosome transitions.

Author Contributions

Field collection was performed by BLSF, CMS, CO-B, and BJE. Laboratory and computational analyses were performed by BLSF, CMS, MK, X-YS, TP, DCJ, and BJE. Writing was performed by BLSF and BJE. All authors contributed to editing.

Introduction

During eukaryotic evolution, genetic control of sexual differentiation changed many times [1]. In some instances, the establishment of a new master regulator for sexual differentiation is associated with cessation of recombination, and extensive divergence in nucleotides, gene content, and gene expression between non-recombining regions of each sex chromosome [2–7]. In other species, extensive recombination between sex chromosomes may occur, and gene content, function (in terms of gene expression), and cytological appearances of each sex chromosome may be almost identical (e.g., [8]). Between these extremes, there exists an astonishing range of variation in the extent of recombination suppression and the degree of sex chromosome divergence [6]. For those sex chromosome pairs that do diverge, it is unclear how fast and in what order differences between them arise. The ability to cope with differences between the sexes in the dosage of gene products stemming from degeneration of sex-linked alleles on the W- or the Y- chromosome [9], periodic recombination [10], and genomic conflict associated with mutations with sexually antagonistic fitness effects via the origin of sex-biased expression patterns [11] all may influence whether or not – and how much – sex chromosomes diverge from each other.

Another phenomenon that may influence recombination and divergence between sex chromosomes is turnover, wherein the genomic location, genetic function (i.e., whether female or male determining) or gene that triggers sexual differentiation changes [12, 13]. A sex chromosome turnover is considered "homologous" when a new variant that assumes the role of sex determination arises on an ancestral sex chromosome [14–17]. A turnover is considered "non-homologous" if it involves establishment of a novel chromosome pair as the sex chromosomes. Non-homologous and homologous turnovers may involve a new variant taking over with the opposite effects of an ancestral sex determining locus; this changes which sex is heterogametic (females for WZ systems, males for XY systems). For example, in medaka fishes, a new trigger for female development replaced an ancestral trigger for male development, creating a turnover of XY to WZ sex chromosomes [18]. Turnover can also occur via translocation of a sex determination allele, which is the case in strawberries [19] and some salmonids [20]. Non-homologous XY to XY turnovers may be favoured by natural selection if the ancestral Y-chromosome has a high load of deleterious mutations [21, 22]. However, Y-linked deleterious mutations may disfavour an XY to WZ transition if this transition involves autosomal segregation of the ancestral Y-chromosome [16]. So far, only a handful of master sex determining genes are known [23]. Understanding why sex determination systems and their associated sex chromosomes change is a challenging prospect (reviewed in [17]), but catching them in the act – during evolutionary windows where multiple sex determination systems co-exist in one species – may help us understand why and how this occurs.

Specifically, these transitions periods may offer insights into whether and how characteristics of ancestral sex chromosomes (e.g., nucleotide divergence, sex-biased expression, degeneracy) bleed through to effect new sex chromosome systems, and how these ancestral characteristics influence what type of sex chromosome system comes next.

In amphibians, many changes between male and female heterogamety have been inferred [15, 24–26], making this group a compelling focus for studies of new sex determining systems and early evolutionary events of young sex chromosomes. Within-species variation in the heterogametic sex has been identified in a handful of amphibians such as the Japanese wrinkled frog, (*Glandirana rugosa*; [14]), Hochstetter's frog (*Leiopelma hochstetteri*; [27]), and the Western clawed frog (*Xenopus tropicalis*; [28]), studied here. In *X. tropicalis*, a W-, Z-, and Y-chromosome have been identified [28–30], but no cytological divergence among sex chromosomes of this or any other *Xenopus* species has been detected [31]. Most of the sex chromosomes of *X. tropicalis* are pseudoautosomal regions where genetic recombination occurs [32]. Current understanding is that the W- is dominant for female differentiation over the Z-, and the Y- is dominant for male differentiation over the W- [28]. Although it is technically no longer a Z-chromosome after the Y-chromosome appeared, we use this term as a placeholder to refer to the extant non-male-specific sex-chromosome that descended from the ancestral Z-chromosome, following [28]. WW and WZ individuals develop into females and ZZ, ZY, and WY individuals develop into males [28]. Thus females carry at least one W-chromosome but not all males carry a Y-chromosome. In principle, YY offspring could be generated if a genetic male (WY or ZY) were sex reversed and developed into a phenotypic female and then crossed with another genetic male. To our knowledge, natural sex reversal has not been reported in *X. tropicalis*, and we assume here that this is rare.

The genomic location of the female-associated region of the W-chromosome was narrowed down in a laboratory strain to a 95% Bayes credible interval positions 0–3.9 megabases (Mb) on chromosome 7 in genome assembly 9.1 (v9) [30]. However, this region was not completely linked to the female phenotype in this study, and it was proposed that this lack of complete linkage could stem from ancestral admixture with an individual carrying a Y-chromosome [30]. The male determining factor of the Y-chromosome of *X. tropicalis* is thought to be in a similar location as the female-determining factor [28], but its precise location has not been determined. Within the genus *Xenopus*, the most recent common ancestor of subgenus *Silurana*, which includes *X. tropicalis*, probably had heterogametic females [26]. This implies that the Y-chromosome of *X. tropicalis* is younger than the W-chromosome and Z-chromosome and thus derived from an ancestor of one of these chromosomes. Mitochondrial genomes of species in subgenus *Silurana* diverged about 25 million years ago [33], implying that the Y-chromosome of *X. tropicalis* is younger than that. This variation raises the possibility that *X. tropicalis* is currently in the midst of a homologous sex chromosome

turnover.

X. tropicalis is a model organism for study of developmental biology and human disease [34–36]. Y-chromosomes have been detected in laboratory strains of *X. tropicalis* that are thought to originate in Sierra Leone, Ivory Coast, Nigeria, and Cameroon [28], although this has not been confirmed directly in specimens sampled in nature. Also unknown is whether populations with male and female heterogamy geographically overlap and interact genetically in nature, whether the variation that defines these chromosomes occurs in the same gene or genomic region, or exactly when cytogenetically undifferentiated sex chromosomes, such as those of *X. tropicalis*, acquire characteristics that are often associated with old sex chromosomes (nucleotide divergence, sex-biased gene expression). We are also interested in evaluating whether and how recombination differs between the sexes of *X. tropicalis* – including in the sex-linked region and across the genome. To address these gaps in knowledge, the goals of this study are to (i) test whether there is male or female heterogamy in natural populations of *X. tropicalis*, (ii) narrow down the region of sex linkage in this species, (iii) evaluate genome-wide patterns of sex-biased expression and nucleotide differentiation, and (iv) characterize patterns of recombination across the genome and between the sexes of wild-caught individuals of this species.

Results

We collected *X. tropicalis* individuals in their natural habitat in Ghana in 2015, and generated three families at McMaster University from frogs that were imported from Ghana and their offspring (Fig. 1). To examine sex-linkage, nucleotide divergence, gene expression, and recombination, two families were used for reduced representation genome sequencing (RRGS), and offspring from the third family were used for transcriptome sequencing (RNAseq). In the RRGS data, we also included genetic samples from *X. tropicalis* individuals derived from Sierra Leone and Nigeria. A focused survey of sex linkage was also performed using Sanger sequencing of amplicons from different portions of the sex-linked region. The Sanger sequencing assays additionally included individuals from a laboratory colony at the National *Xenopus* Resource (NXR), Woods Hole, USA, that are derived from Ivory Coast (RRID:NXR_1009) and Nigeria (RRID:NXR_1018).

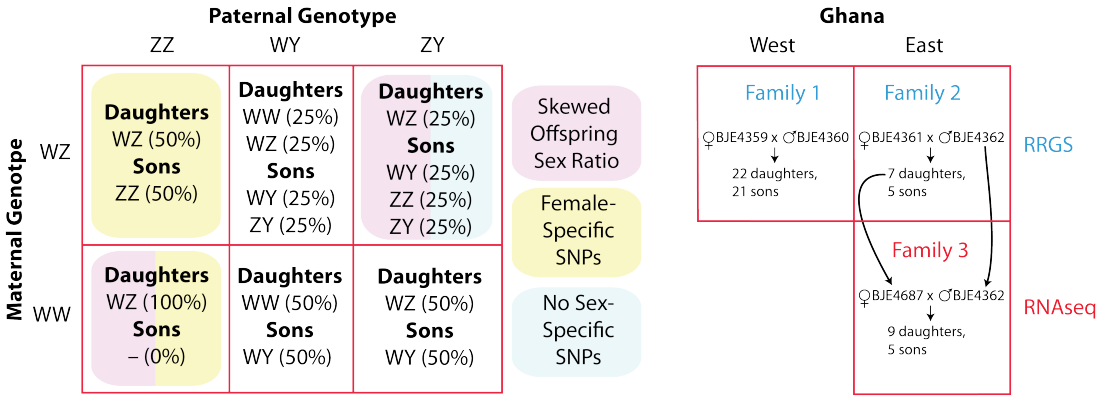


Figure 1: The three sex chromosomes of *X. tropicalis* can be crossed in six ways to produce offspring with different types of sex-linkage and/or skewed offspring sex ratios (left). Crosses on the left that are not shaded are expected to have male-specific SNPs passed from father to all sons in the male-specific portion of the Y-chromosome. We generated three laboratory families from west and east Ghana for RRGS and RNAseq analyses (right). For the RNAseq analysis (Family 3), offspring were analyzed from a cross between the father and a daughter from Family 2 (indicated with arrows).

Males have a Y-chromosome in Two *X. tropicalis* Families From Ghana

There are two possible sex chromosome genotypes in females (WZ, WW) and three in males (ZZ, ZY, WY), which can be combined in six ways for reproduction (Fig. 1). One combination (WZ x ZY) produces a 1:3 female:male offspring sex ratio (WZ daughters and ZZ, ZY, or WY sons); this type of family does not have completely sex-linked genetic variation passed from either parent to all of the same-sex offspring (because the W- and Z-chromosomes are both inherited by sons and daughters, and the Y-chromosome is not inherited by some sons). Another parental combination (WW x ZZ) produces only WZ daughters. This parental combination, and one other with no offspring sex-bias (WZ x ZZ), are expected to have completely female-linked genetic variation passed from mother to daughters on the W-chromosome. The three other parental combinations (unshaded in Fig. 1) are all expected to have no sex-bias in offspring numbers, and also completely male-linked genetic variation passed from father to sons on the Y-chromosome: WZ x WY, WW x ZY, WW x WY.

In the RRGS data from two families (Family 1 and 2) that were generated from individuals from the western and eastern edges of Ghana, genome-wide inheritance of SNPs provides unambiguous evidence for a sex determining system where males carry a Y-chromosome. There were no maternal heterozygous sites that were sex linked prior to or after FDR correction in either family, but several paternal SNPs on the distal end of chromosome 7 were significantly associated with sex in each cross (Fig. 2). We intentionally sampled an approximately equal number of offspring of each sex in Family 1 and 2 (22 and 21 daughters and sons

for Family 1, seven and five daughters and sons for Family 2) and the presence of sons allows us to rule out the possibility that either of these two crosses was between a WW mother and a ZZ father. Additionally, completely (100%) male-linked SNPs were detected similar genomic locations in both families - on one end of chromosome 7. Together these observations demonstrate that the father of each of these families carried a Y-chromosome, and that at least one of the parents in both crosses did not carry a Z-chromosome, because any combination with both parents carrying one or more Z-chromosomes would not have any completely male-linked SNPs (Fig. 1). Thus the sex-chromosome genotypes of Families 1 and 2 both could be any one of the three unshaded crosses in Fig. 1. Additional details about polymorphism and sex-linkage in Family 1 and 2 are provided in the Supplement. There are differences in the v9 genome assembly used by [30] and the v10 genome assembly used in this study in the sex-linked region of chromosome 7 (Suppl. Fig. S3). To facilitate comparison between these studies, we report genomic coordinates of both assemblies for the sex-linkage and F_{ST} results. Findings discussed below from RNAseq are reported using v10 coordinates and the genome-wide recombination analyses were performed using v9.

None of the sex-linked SNPs from Families 1 and 2 were sequenced in the other family, and this is presumably a consequence of variable coverage of the RRGS data between each family, and variation in the presence of *SbfI* restriction sites. There were non-sex-linked SNPs in Ghana west Family 1 that occurred in the region that was sex linked in the Ghana east Family 2, and vice-versa (Fig. 2). One explanation for this discrepancy is that all of the sex-linked positions in both families are near to, but not within, the sex-determining region. These offspring thus provide a sample of recombination events that occur near the sex determining locus and also elsewhere in the genome. In Ghana East Family 2, we lacked markers between ∼6-11Mb on chromosome 7, and consequently were unable to further pin down the extent of sex-linkage in this family in this region using the RRGS data.

On chromosome 7 <6Mb, we had variable markers that overlapped in Family 1 and 2, and these markers were male-linked in Family 2 but not Family 1 (Fig. 2). Using analyses discussed below, we were able to conclude that the sex chromosome genotype of the father of Family 2 (BJE4362) was WY. Although we were not able to discern the sex chromosome genotype of the father of Family 1 (BJE4360), one possibility is that his sex chromosome genotype was ZY, and that recombination occurs between Z- and Y-chromosomes, but not between W- and Y-chromosomes (and possibly not also between W- and Z-chromosomes though we do not attempt to address this possibility here). This scenario would explain why there were sex-linked sites on the end of chromosome 7 in Family 2 but not Family 1. This scenario is also consistent with evidence presented below for a lack of recombination in the sex-linked region during spermatogenesis of the father of Family 3 (BJE4362, also the father of Family 2), an origin of the Y-chromosome from an ancestral Z-

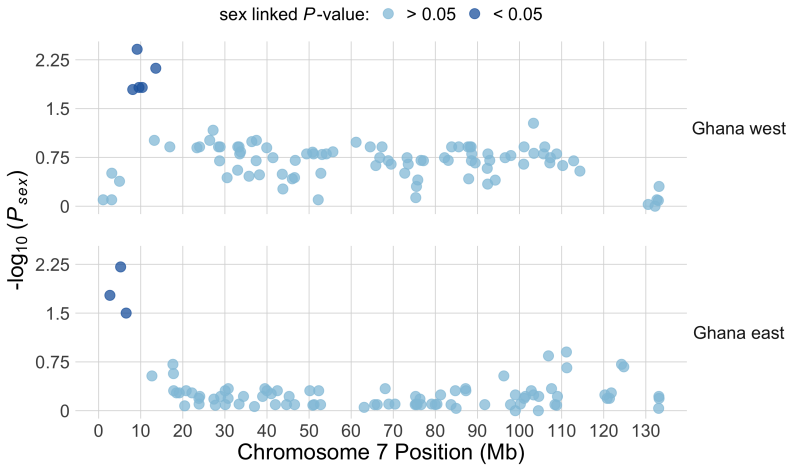


Figure 2: Manhattan plot of sex linkage for chromosome 7 in the Family 1 from Ghana west (top) and Family 2 from Ghana east (bottom) for paternal heterozygous sites. Dark and light dots indicate variants with a significant or not significant association with sex, respectively, after FDR correction in (top), and before FDR correction in (bottom).

chromosome, and also with degeneration of the sex-linked portion of the W-chromosome as a mechanism for a high density of transcripts with male-biased expression.

To explore sex-linkage further, we used Sanger sequencing to survey variation in each family from Ghana, in wild caught individuals, and in various laboratory strains; results are summarized in Table 1. Using Sanger data, we were able to identify one 100% sex-linked SNP and two almost completely sex linked SNPs in Family 2. With a very small sample size we identified three 100% sex-linked SNPs in the Ghana east wild population, but none in Family 1 or the Ghana west wild population. The sample sizes of sequenced amplicons was small for several strains due to a low number of individuals collected in the field or propagated in the laboratory, and unclear Sanger sequences in due to insertion/deletion polymorphisms. Although other explanations are possible, the lack of sex-specific SNPs (even with this small sample size) in Family 1 is consistent with the hypothesis proposed above that the sex chromosome genotype of the father may have been ZY. Under this scenario, recombination between the Z- and Y-chromosomes during spermatogenesis could cause some of the sons to not carry Y-linked SNPs at LOC100488897 (Table 1), which is located at ~8Mb, even though they may have inherited the Y-linked male determining factor that is located ~10–11Mb on chromosome 7. In Family 2, the lack of complete sex-linkage at two loci <11Mb on chromosome 7 was due to homozygous positions in one son that were heterozygous in four other sons. Based on conclusions discussed below, two different sex chromosome genotypes are expected in sons of Family 2 (see top center of left side of Fig. 1). One possibility is that the one unusual son was ZY and the other sons were WY, and that their

heterozygous positions were due to divergence between the W and Y chromosomes that was not present in the ZY son. Overall, while the extent of the Sanger sequencing data were limited by difficulties with obtaining clean sequences from our amplicons, the findings from the available data are generally congruent with the results from the RRGS data in the sense that some portion <11Mb on chromosome 7 does not appear to be sex-linked in Family 1, but we did identify a male-linked SNP in this same region in Family 2).

One amplicon (LOC100495284) had a male-specific SNP in two strains at the NXR from Nigeria (Nigerian and Superman) and a small sample of Ghana east wild individuals, but not in other strains we surveyed. In the NXR strains from Nigeria, males were A/G and females G/G at position 1,928,777 or 10,257,211 in v9 or v10, respectively. A male-specific SNPs is not definitive evidence of a Y-chromosome because a segregating polymorphism on a Z-chromosome or an autosome could by chance be present only in males. However, without invoking a Y-chromosome, the chance of observing a heterozygous genotype in 17 of 17 males and zero of 18 females is very low ($\leq 7 \times 10^{-6}$). The sex-linked SNPs in the Nigeria strain was different from a nearby sex-linked SNP in three wild Ghana east males.

The *X. tropicalis* Sex Chromosomes Have a Small Region of Differentiation

We quantified F_{ST} between females and males over a moving average of 50 single nucleotide polymorphisms (SNPs) in wild individuals from Ghana, and georeferenced lab individuals from Sierra Leone and Nigeria. Some of these individuals have a Y-chromosome (demonstrated above) and all of the females (presumably) have a W-chromosome because this chromosome is required for female differentiation; some individuals of either sex may additionally have a Z-chromosome. In this analysis, differences in nucleotide polymorphism at various localities coupled with sex differences in sample sizes are expected to have a genome-wide effect on F_{ST} between females and males due to population subdivision (see below). In non-recombining portions of the sex chromosomes, F_{ST} additionally should be influenced by nucleotide divergence between the Y-, Z-, and W-chromosomes, nucleotide polymorphism on these chromosomes, and the different frequencies of these chromosomes in females and males. Thus, F_{ST} between the sexes across populations has the potential to reveal patterns of differentiation between the sex chromosomes.

Consistent with expectations associated with divergence between the sex chromosomes due to recombination suppression, we observed a generally higher F_{ST} in the sex-linked region(<11Mb) of chromosome 7 that was identified in the RRGS data from Family 2. There was a genome-wide F_{ST} peak equal to 0.13 at position 9,940,000 in the sex-linked region of chromosome 7 in v10; F_{ST} was >0.09 from positions 9,775,600

Table 1: Results of the Sanger sequencing survey of 18 amplified regions (Locus) for seven groups of *X. tropicalis*: a lab cross and wild individuals from west Ghana (Family1, GWwild) or east Ghana (Family2, GEwild) that were used for the RRGS data (but not the RNAseq data), and captive strains at the National *Xenopus* Resource from Ivory Coast (IC) and Nigeria (Nigerian and Superman). The genomic position of each locus in the *X. tropicalis* v9 and v10 are indicated, with the chromosome or scaffold followed by a range of genomic coordinates. For each group, the number of males and females sequenced are separated by slashes, followed by whether a sex-specific SNP was detected (Y) or not (N); "NV" indicates no variation in the sequences. A dash indicates that the amplification was not attempted or that Sanger sequences were not clean. For two loci from the Family 2, an asterisk indicates that 4/5 males had a male-specific SNP and in both of these amplicons, the same male individual did not have this SNP; thus variants at these loci were almost but not completely sex-linked.

Locus	Xenopus tropicalis 9.1	Xenopus tropicalis 10.0	Family1	GWwild	Family2	GEwild	IC	Nigerian	Superman	Notes
-	scaffold_486: 109006-109688	Chr7: 590989-591647	–	3/2/N	–	2/1/N	–	–	–	
-	scaffold_1093: 25288-26168	Chr7: 2401099-2401979	–	3/2/N	–	4/1/N	–	–	–	
vwa2	scaffold_132: 209950-210334	Chr7: 3988998-3989381	–	1/1/Y	–	2/0/NV	–	–	–	a
LOC108644822	scaffold_83: 134640-135403	Chr7: 6350671-6351443	–	–	–	2/1/N	–	–	–	
LOC108644867	scaffold_130: 572284-572814	Chr7: 7445611-7446140	–	3/2/NV	–	–	–	–	–	
phc1	scaffold_130:760304-761055	Chr7: 7992897-7992184	–	2/2/N	–	1/1/NV	–	–	–	
LOC100488897	scaffold_130: 643554-643884	Chr7: 8109808-8110138	2/2/N	5/5/N	5/7/Y	5/1/N	–	–	–	
aicda	Chr7: 665453-665920	Chr7: 9026686-9027153	–	4/3/N	–	3/1/N	–	–	–	
SNP1	Chr7: 901880-902194	Chr7: 9256905-9257219	2/2/NV	–	5/7/N*	1/1/Y	–	–	–	
trpg1l-like	Chr7: 1364981-1365454	Chr7: 9677066-9677539	–	4/3/N	–	3/1/N	–	–	–	
grp162	Chr7: 1386997-1387487	Chr7: 9698865-9699356	–	4/3/N	–	3/1/N	–	–	–	
LOC100495284	Chr7:1928340-1928761	Chr7: 10256773-10257194	2/2/NV	5/4/N	5/7/N*	3/1/Y	3/5/N	5/8/Y	12/10/Y	b
LOC100493019	Chr7: 2045411-2045931	Chr7: 10389186-10389706	–	4/3/N	–	3/1/N	–	–	–	
SNP4	Chr7: 5207535-5208256	Chr7: 13469130-13469851	–	–	5/7/N	1/1/Y	–	–	–	

a: This is possibly an allele-specific amplification in some populations; no amplification occurred in several male and female individuals, and in the Ghana west population, one male sequence differed from one female sequence by 1 homozygous nucleotide; sequences from two Ghana east males were invariant and identical to the Ghana west male.

b: The same SNP was present in Superman and Nigerian males, but a different SNP was present in Ghana east males. The male-linked allele in Superman and Nigerian males amplified weakly, but consistently.

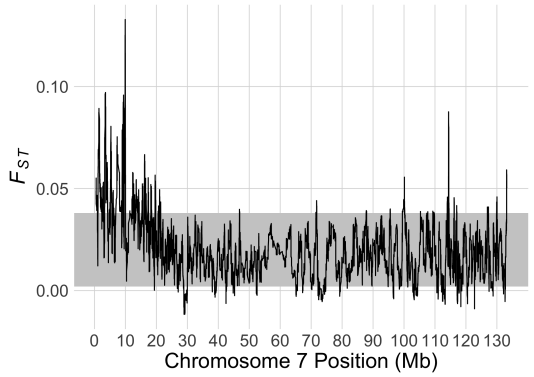


Figure 3: F_{ST} between females and males for *X. tropicalis* chromosome 7 of wild samples from Ghana, and georeferenced lab strains from Nigeria and Sierra Leone. The grey band represents the whole genome bootstrap CI for the mean F_{ST} that generated by resampling F_{ST} measured on the autosomes.

– 9,999,600 (Figs. 3, S2). The locations of this peak and range are 1,615,479 and 1,454,645 – 1,664,477 in v9. This F_{ST} peak and the male-specific SNPs on the *X. tropicalis* Y-chromosome discussed above (Table 1) overlap with a small genomic window of strongly female-linked variation on the *X. tropicalis* W-chromosome found previously by [30] (Suppl.Fig. S3). Specifically, the margins of the F_{ST} peak overlap with the most strongly female-linked linkage group (super_547:0; positions 1,365,917 – 1,693,249 in v9; LOD score: 13.13296453 [30]). Additionally, a male-linked SNP from the Sanger sequencing (locus LOC100495284) is within an adjacent – and almost equivalently strongly female-linked – linkage group (super_547:1; positions 1,731,784 – 2,045,519 bp in v9; LOD score: 13.13296449 [30]). The 100% sex-linked SNP in the Ghana west family was also very close to this peak (at position 9,149,465 in v10, 795,572 in v9; see above). Over the entire sex-linked region <10.4Mb, the mean F_{ST} value is 0.049 (standard deviation = 0.023) which is above the genome-wide 95% CI of 0.002 – 0.038.

The Sex-Linked Portion of the *X. tropicalis* Sex Chromosomes Has a Very High Density of Genes with Male-Biased Expression

We generated and analyzed whole transcriptome data from the gonad/mesonephros complex during an early stage of sexual differentiation. Although this study does not explore this issue, the initial motivation for selecting this tissue and developmental stage was that it corresponds with the timing of gonadal differentiation in *X. laevis* [37] and the sex determining gene of *X. tropicalis* could also be expressed in this tissue type and developmental stage. These tissues were dissected from tadpoles of Family 3 which was generated from a cross between the wild-caught father of Ghana east Family 2 that was used for the RRGS analysis, and a daughter

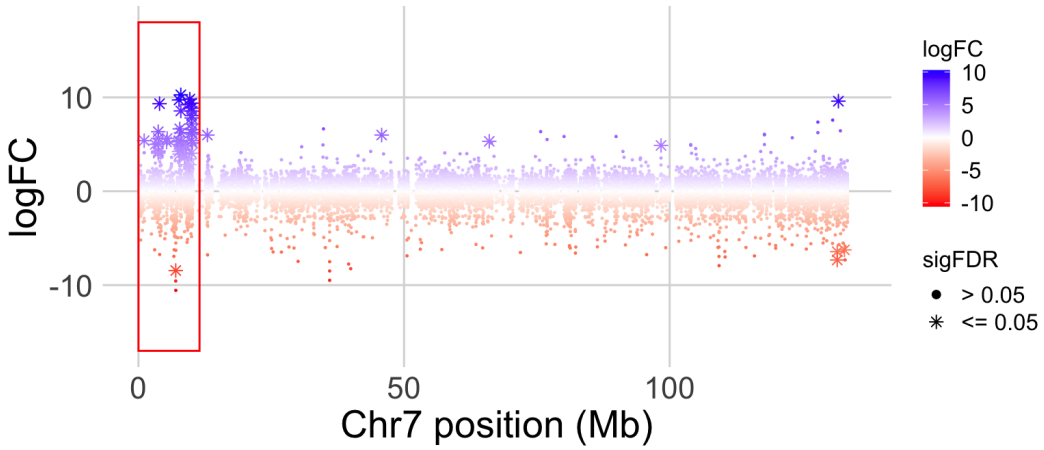


Figure 4: Log_2 transformed male/female transcript expression ratio (logFC) along *X. tropicalis* chromosome 7 in offspring of Family 3. The x-axis indicates the genomic coordinates in millions of base pairs (Mb). Small dots represent individual transcripts and * represent transcripts that are significantly differentially expressed after FDR correction (sigFDR). Positive values reflect male biased expression, negative values are female biased. A red box highlights a cluster of genes on the sex-linked portion of chromosome 7 with mostly male-biased expression.

of this cross (Fig. 1). The sex of each tadpole was assessed using Sanger sequencing (Methods). Although we did not assess sex-linkage in Family 3, evidence discussed above indicates that the father carried a Y-chromosome. There was not a male-sex-bias in offspring numbers (nine daughters, five sons) which, following the same reasoning above, indicates that at least one of these parents did not carry a Z-chromosome, and that the sex-chromosome genotypes of the parents of Family 3 could be any one of the same three unshaded crosses in Fig. 1 that were possible for Families 1 and 2.

We used these expression data to evaluate how genes with sex-biased expression are distributed over the genome, including in sex-linked and non-sex-linked portions of the sex chromosomes. A total of 259,197 transcripts were assembled that mapped to one of the 10 chromosomes in v10; 296 transcripts were detected that mapped to scaffolds in this assembly, and 2816 did not map to these assemblies. Of these, 151, 1, and 5, respectively, had significant sex-biased expression after FDR correction.

For neutral variants in an autosomal locus, genetic drift of the transcriptome is not expected to produce a skew in sex-biased expression. In the non-sex-linked portion of the genome (including the pseudoautosomal region of chromosome 7), for example, the numbers of transcripts with significantly male- or female-biased expression were relatively similar (n = 63 and 44, respectively). However, there were two genomic regions

with a high density of genes with sex-biased gene expression (Figs. 4, S5). The first is the sex-linked portion of chromosome 7 (<3.3 Mb in v9 or <10.4 Mb in v10), which has a very high density of genes with male-biased expression (45 transcripts from 30 genes in v10) but not female-biased expression (1 transcript; Table S2, Fig. 4). Twenty-seven of these transcripts from 20 genes were male-specific (expressed only in males); 18 transcripts from 13 genes were male-biased (expressed in both sexes but significantly higher in males), and only one transcript was female-specific. The proportions of differentially expressed transcripts from this region with male-biased or male-specific expression are significantly higher than expected based on the proportion from the rest of the genome ($P < 0.00001$, binomial tests). As a consequence of the high density of these genes on the sex-linked region, the number of transcripts with significantly male-biased and male-specific expression was far higher on chromosome 7 than any of the other chromosomes (Table S4), even though the proportion of this chromosome that is sex-linked is small (<10% of chromosome 7; Fig. 2; [32]), and even though chromosome 7 is intermediate in size.

The second region with a high density of sex-biased transcripts is on chromosome 3 between 114 – 128 Mb in v10; this area encodes a high density of female-biased transcripts (21 transcripts from 14 genes) but not male-biased (one transcript; Table S3). However, the density of sex-biased transcripts on this region of chromosome 3 (1.5 transcripts/Mb) is substantially lower than the density of sex-biased transcripts on the sex-linked portion of chromosome 7 (4.4 transcripts/Mb) (Fig. S5). We do not know why this region has a high density of female-biased transcripts; hereafter we focus on the sex-linked region.

Why Do Genes in the Sex-linked Region Encode So Many Transcripts with Male-biased Expression?

There are several possible explanations for the strong skew towards male-biased expression of transcripts encoded by genes in the sex-linked region of these frogs. One possibility is that this particular region had a high density of male-biased transcripts in an ancestor when this region was not sex-linked (that is, prior to the origin of a sex-determining locus on chromosome 7 in *X. tropicalis*). To gain perspective into this possibility, we turned to expression data that we collected for another study from *X. borealis*, a closely related allotetraploid species, from the same tissue (gonad/mesonephros) and a similar developmental stage (tadpole stage 48). We determined genomic locations of assembled transcripts in the *X. laevis* genome assembly version 9.2 using the same methods as described here for *X. tropicalis*, and as described in more detail elsewhere [38]. Because *X. borealis* is allotetraploid and because it has different sex chromosomes than *X. tropicalis* (on chromosome 8L [39]), this species has two autosomal chromosomes (chromosomes 7L

and 7S), that are orthologous to the sex chromosomes of *X. tropicalis*. Inspection of these regions identified only one significantly male-biased transcript on *X. borealis* chromosome 7L, one on chromosome 7S, and no significantly female-biased transcripts on either of these chromosomes (Suppl. Fig. S4). This comparison does not rule out the possibility that a strongly male-biased expression skew was present ancestrally but lost during evolution of *X. borealis*, but it does suggest that there is no reason to expect that transcripts in this genomic region was somehow predisposed to have male-biased expression. Inspection of non-sex-linked regions in *X. tropicalis* (Suppl. Fig. S5) leads to a similar conclusion: no other region in *X. tropicalis* has such high densities of male- or female-biased expression as the sex-linked region. Thus, the extreme concentrations of transcripts with sex-biased expression that we observe on the sex-linked region of *X. tropicalis* are rare on autosomes. Taken together, these comparisons favor the interpretation that the evolution of male-biased expression occurred in concert with the origin of sex-linkage <11Mb on chromosome 7 in *X. tropicalis*.

How might sex-linkage lead to a skew towards male-biased expression of transcripts encoded by sex-linked genes? One possible explanation is that expression of some alleles in the sex-linked region of the W-chromosome decreased or was lost due to recombination suppression, and that the Y-chromosome evolved from an ancestral non-degenerate Z-chromosome, and thus retains functional regulation of these alleles that are now Y-linked. To explore this possibility, we calculated pairwise nucleotide diversity in sex-linked expressed transcripts of each individual offspring of Family 3 using the RNAseq data. Estimates of polymorphism in expressed transcripts are expected to underestimate divergence between the sex chromosomes due to unexpressed and differentially expressed alleles, but are a useful indicator of sex chromosome regulatory degeneration because transcripts encoded by only one sex chromosome should have no heterozygous positions, and those co-expressed by gametologous alleles could have heterozygous genotypes due to divergence between sex chromosomes as a result of recombination suppression [17].

For two of the three possible sex chromosome genotype combinations (WW mother x WY father; WW mother x ZY father), if recombination is suppressed in the sex-linked portion of the W-chromosome, we expected divergent sites in expressed transcripts of sex-linked genes to be similar within each offspring sex. This is because there is only one sex chromosome genotype for each sex for each of these parental sex chromosome genotypes (bottom middle and bottom right of the left side of Fig. 1). However, for offspring from a WZ mother and a WY father, daughter and sons each have two possible sex chromosome genotypes (WW or WZ for daughters, WY or ZY for sons (top middle of of the left side of Fig. 1). If recombination is suppressed in the sex-linked portion of the W-chromosome, we therefore expected that this type of cross could have two distinctive levels of within female polymorphism in sex-linked expressed transcripts (in WW and WZ females), and also two distinctive levels of within male polymorphism in sex-linked expressed transcripts

(in ZY and WY males). After filtering (Methods), we retained for analysis an average of 782 (range: 653–904) expressed transcripts from the sex-linked region (<11Mb on chromosome 7) per individual, and an average of 50 bp (range: 43-57) per transcript within each individual.

Consistent with our predictions associated with a cross between a WZ mother and a WY father, we observed two distinct levels of polymorphism in expressed transcripts of sex-linked genes within daughters, and also two within sons (Fig. 5). In addition to resolving the sex chromosome genotypes of the parents of Family 3 (the mother was WZ and the father was WY), these results also indicate that the genotype of the mother of Family 2 (BJE4361, Fig. 1) was also WZ because her daughter, who was the mother of Family 3 (BJE4687) carried a Z-chromosome, and her father did not. These findings also indicate that the Y-chromosome is derived from the Z-chromosome and not from the W-chromosome because divergence between the Z- and Y-chromosomes is lower than divergence between the Z- and W- or between the Y- and W-chromosomes. Additionally, and perhaps most surprisingly, these results demonstrate that W-, Z-, and Y-chromosomes all co-occur in nature in individuals from the same small seep (<6 feet wide) in east Ghana.

We initially expected WW offspring to have no polymorphism in sex-linked transcripts because both of these W-chromosomes descend from the BJE4362, who was the father and the grandfather of offspring of Family 3 (Fig. 1), but this expectation was not borne out (Fig. 5). We suspect that this is a consequence of mapping error, for example due to repetitive regions in untranslated regions of these transcripts. Another non-exclusive possibility is that recombination occurred in the distal portion of the sex-linked region of the W- and Z-chromosomes during oogenesis in the mother or in distal portion of the sex-linked region of the W- and Y-chromosomes during spermatogenesis in the father/grandfather. To test these possibilities, we separately quantified polymorphism in transcripts encoded by genes <6Mb and between 6Mb and 11Mb of the sex-linked region (Suppl. Fig. S6). If recombination occurs between the W- and Y- sex chromosomes (in the father of Family 3) or between the W- and Z chromosomes (in the mother of Family 3), we expected diversity of the first half would be similar to levels seen in other recombining regions of the genome (i.e. the pseudoautosomal and autosomal regions). However, within three of the four sex chromosome genotypes, the first portion and second portions of the sex-linked region had similar diversities and these diversity levels were generally not similar to the pseudoautosomal and autosomal regions (see below). In the putative ZY individuals, the diversity in the 5' portion of the sex-linked region was similar to that in the autosomal region. One interpretation of this small sample is that recombination between the Z- and Y-chromosomes causes polymorphism to be similar in the first half of the sex-linked region. We therefore favor a technical explanation for the inference of polymorphism in these putative WW individuals over the biological explanation associated with recombination between W-chromosome and the Y- or Z- chromosomes (Fig. 5).

We had no expectation about the relative level of pairwise nucleotide polymorphism in expressed transcripts in recombining compared to non-recombining regions because we do not know the proportion of sex-linked transcripts that are only expressed on one sex chromosome (and thus completely homozygous, which decreases polymorphism compared to expressed non-sex linked transcripts) or the level of divergence between expressed transcripts on different sex chromosomes (which generates heterozygous genotypes and increases polymorphism compared to expressed non-sex linked transcripts). Moreover, the level of divergence between non-recombining portions of sex chromosomes need not be constant throughout the non-recombining region [2]. The average polymorphism of expressed non-sex-linked transcripts was 0.013 (range: 0.012-0.015). That this level of polymorphism is lower than that of expressed sex-linked transcripts in putative WZ females and WY males argues for the presence of diverged sites between the sex-linked regions of the W- and Z-chromosomes and the W- and Y-chromosomes.

If these inferred sex chromosome genotypes are correct, we predicted that patterns of male-biased expression might be distinctive if we reanalyzed the data using a subset of individuals (subset 1) that included only putative WW females (XT3, XT9, XT11, XT20) and only putative WY males (XT7, XT8) as compared to a subset (subset 2) that included only putative WZ females (XT2, XT6, XT10, XT16, XT17) and only putative ZY males (XT1, XT13, XT19) because females do not have a Z-chromosome in the first subset but they do in the second subset. As expected, more significantly male-biased transcripts were identified in the analysis of subset 1 (Fig. S7) than subset 2 (Fig S8), even though subset 1 included fewer individuals of both sexes. Another interesting finding that emerges from these analyses is that the portion of chromosome 7 with the high density of genes with sex-biased expression extends slightly beyond the sex-linked region that we identified when all individuals were included for subset 1, but not for subset 2. This suggests that the degeneracy of the W-chromosome may extend beyond 11Mb, but the male-biased expression associated with this degeneracy on the W-chromosome in this region is not as apparent in WZ females due to the presence of the Z-chromosome.

When we performed an analysis of differential expression between putative WW females and putative ZY males (subset 3), the results were similar to the subset 1 analysis in terms of number and identities of differentially expressed transcripts in and near the sex-linked region (Suppl. Fig. S9). We also performed an analysis of only putative WZ females and putative WY males (subset 4). If male-biased expression were solely due to degeneration on the W-chromosome, we expected this comparison to have even more modest sex-differences than the other subsets. However, the pattern of male biased expression was relatively similar to the analysis of subset 2 (which is putatively WZ females versus ZY males; Fig. S10). This suggests that, in addition to degeneration of the W-chromosome, there may be regulatory differences of alleles on

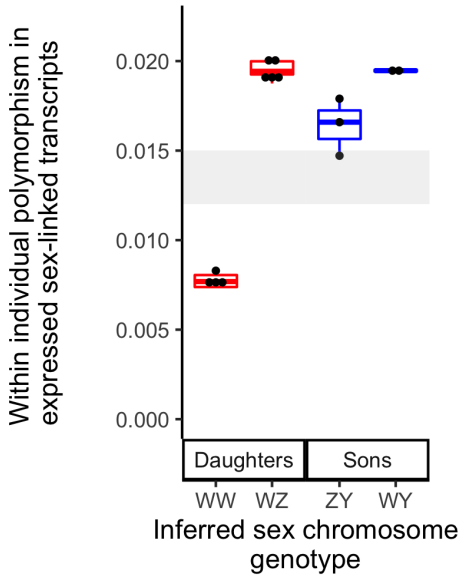


Figure 5: In daughters and sons of Family 3, two distinct levels of within individual polymorphism in expressed sex-linked transcripts imply that there are two distinct sex chromosome genotypes in offspring of each sex. Inferred sex chromosome genotypes (x-axis) are based on within individual polymorphism of expressed sex-linked transcripts (y-axis). The range of pairwise nucleotide diversity for non-sex-linked transcripts in the 14 individuals for which RNAseq was performed is depicted in gray.

the sex-linked portions of the Z- and Y-chromosomes that contribute to male-biased expression in the main analysis (Fig. 4). Alternatively or in addition, regulation of several genes in the sex-linked portion of the Z-chromosome may vary depending on whether it this chromosome is in a female or male individual.

Rates and Locations of Recombination are Sex-specific in *X. tropicalis*

We used RRGS data from Families 1 and 2 to compare genome-wide rates and locations of crossover events in females and males. For both families, the total length of the female linkage map greatly exceeded that of the male map, even though the female and male maps had a similar number of markers and spanned similar proportions of the genome. For the Ghana west family, the female map length was 920 cM (including 1504 SNPs), and the male map was 367 cM (including 1645 SNPs). For the Ghana east family, the female map length was 1495 cM (2061 SNPs), and the male map length was 630 cM (1857 SNPs). This indicates that recombination is far more common during oogenesis than during spermatogenesis.

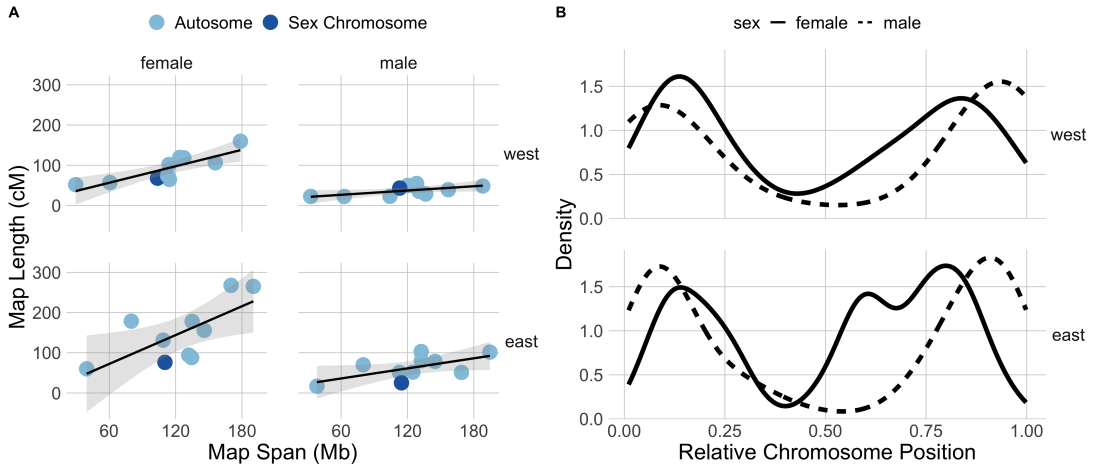


Figure 6: A) Linkage map length in centimorgans (cM) is positively correlated with the length of the genomic region in millions of base pairs (Mb) in females (left) but not in males (right) from Family 1 and Family 2 (top and bottom rows, respectively). B) Crossover density for is more strongly biased towards chromosome tips in males than females in Family 1 and Family 2.

There was a positive relationship between linkage map length and the size in base pairs of the genomic region to which the linkage map corresponds in females maps, but this was not evident in males maps. The slope of this relationship in females from Family 1 is 28.7 (CI: 6.7 – 50.7) and in females from Family 2 is 50.9 (95% confidence interval (CI): 28.5 – 71.8), whereas this slope for males from Family 1 is 7.5 (CI: -13.7 – 28.6) and males from Family 2 is 17.6 (CI: -3.9 – 39.0; Fig. 6A). That male recombination rates were unrelated to the size in base pairs of the genomic region to which the linkage map corresponds, argues that their crossover events in males occur in more concentrated genomic regions, as compared to females. Consistent with this, crossover events were more biased towards chromosome tips in during male recombination compared to female recombination, although both sexes had a lower density of crossover events near the centers of chromosomes as compared to the first and third quartiles (Fig. 6B). These differences are unlikely to be related to sex-differences in coverage of the RRGS data because markers included in the linkage maps spanned similar proportions of the chromosomes in both sexes, and in both populations (Family 1: an average of 95.1% of the chromosome lengths were covered for females and 93.2% for males; Family 2: 99.0% for females and 98.8% for males). We did not detect a substantial disparity in the number of crossovers on the sex chromosome (chromosome 7) in male maps of either family (Fig. 6 A), which is consistent with most of this chromosome being pseudoautosomal [32].

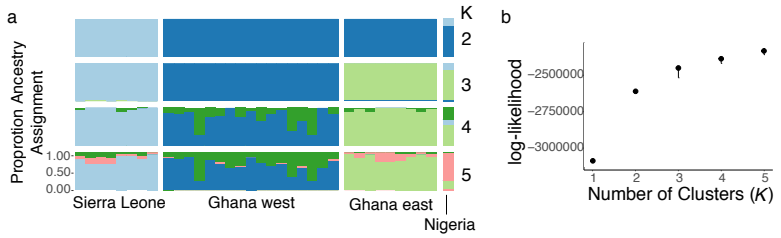


Figure 7: Genetic cluster analysis of RRGS data illustrates geographic structure of wild *X. tropicalis*. (a) Ancestry assignments of individual samples for 2–5 populations (K). (b) log-likelihood of values of K from 1–5.

Population Structure in *X. tropicalis*

We tested whether there was a genome-wide signature of population differentiation in *X. tropicalis* samples derived from nature and georeferenced laboratory animals. This analysis identifies population differentiation between samples from Sierra Leone and Ghana + Nigeria with two partitions, and between Sierra Leone, Ghana west, and Ghana east + Nigeria with three (Fig. 7). With more than three partitions, additional subdivision is within individuals from each geographic locality.

Discussion

The Geographical Context of *X. tropicalis* Sex Chromosomes

Xenopus tropicalis is distributed in tropical habitats in West Africa, ranging from Sierra Leone to western Cameroon [40]. Rainforest habitat in West Africa is interceded by savanna in a region called the Dahomey Gap, which lies roughly in the center of the distribution of *X. tropicalis*, including southeastern Ghana, the southern portions of the countries of Togo and Benin, and southwestern Nigeria [41, 42]. Few records of *Xenopus* are available from the Dahomey Gap, and it is possible that there is a discontinuity in the range of *X. tropicalis* in this region. Limited data from mitochondrial DNA sequences suggests that there may be population subdivision within *X. tropicalis* that is associated with the Dahomey Gap [32]. Genome-wide data analyzed here point more strongly to subdivision between *X. tropicalis* populations from Sierra Leone and Ghana + Nigeria, and suggest populations from east Ghana, which were sampled in a forested patch within the Dahomey Gap, are less differentiated from populations from Nigeria (east of the Dahomey Gap) than from populations from west Ghana (west of the Dahomey Gap; Fig. 7).

Our results identify, for the first time, a Y-chromosome in *X. tropicalis* individuals sampled directly from nature (two localities Ghana) and suggest a Y-chromosome is present in a laboratory strain from Nigeria. In a recent study of sex-linkage in *X. tropicalis* that included the Nigeria laboratory strain that was used for the genome sequencing, the nature of the crosses did not permit assessment of whether a Y-chromosome was present [30]. However, the authors concluded that if there was a Y-chromosome, it would have originated from the Nigeria strain [30], which is consistent with our findings. Our limited survey of variation in a laboratory strain from Ivory Coast did not identify sex-linked SNPs (Table 1). This does not rule out the possibility that the Ivory Coast strain also carries a Y-chromosome because some of the males may have two Z-chromosomes. The findings from the transcriptomic analysis also provide geo-referenced evidence of a Z-chromosome in east Ghana, and suggest as well that a Z chromosome was present with W- and Y-chromosomes in the wild caught individuals in west Ghana. Overall, these results demonstrate that the W-, Z-, and Y-chromosomes co-mingle in the natural range of *X. tropicalis* in Ghana, and perhaps elsewhere. A key question motivated by these co-mingling sex chromosomes asks how and why they co-exist when they are associated with substantial offspring sex-ratio skew [28], which is expected to often be disadvantageous [43–46]. It may be the case that the Z-chromosome has a low frequency in the populations we sampled in Ghana, and that the W-chromosome segregates in these populations, more or less, like an X-chromosome. Eventual extinction of the Z-chromosome would transition the W-chromosome into a new X-chromosome, which is one way to prevent offspring sex-ratio skew with the new sex determining system associated with the newly emerged Y-chromosome. Alternatively, if the Y-chromosome were rare in a population, most crosses also would have a balanced sex ratio governed by the W- and Z-chromosomes. Extinction of the Y-chromosome thus could also prevent offspring sex-ratio skew via reversion to the ancestral (WZ/ZZ) system for sex determination. Although either of these scenarios are possible, results presented here provide direct or indirect evidence that all three sex chromosomes were present in Family 2+3 and 1, respectively, which suggests that none of these chromosomes are on the brink of extinction. Further efforts to genotype sex chromosomes of *X. tropicalis* sampled in nature could evaluate these possibilities.

Are the Three Sex Chromosomes of *X. tropicalis* Defined by Different Variants at One Gene?

The three sex chromosomes of *X. tropicalis* could be part of an evolutionary transition, with a new system for sex determination on its way to fixation. Alternatively, this variation may be stable over evolutionary time, with different frequencies of each sex chromosome favored by unique factors in distinct environments

[47]. In two laboratory bred families from Ghana and a laboratory strain from Nigeria, we identified male-specific SNPs that fall within the tightly female-linked region on the W-chromosome of another *X. tropicalis* strain [30] and also a closely related species, *X. mellotropicalis* (Supplemental Results; [26]). That the male-specific region of the Y-chromosome and the female-specific region of the W-chromosome occur in gametologous locations (Fig. 2) along with a peak of nucleotide differentiation (Figs. 3, S3) points to the possibility that the female-specific (W-linked) and male-specific (Y-linked) variation that drives male and female differentiation, respectively, could be in the same locus or possibly segmental duplicates that are closely situated.

Mechanistically, genetic sex determination can be realized in many ways [1] which could include, for example, a dominant sex-specific allele (such as *Sry* in eutherian mammals), or dosage mechanisms that involve differences in copy number of a shared allele (such as *Dmrt1* in birds). At this point we can only speculate about mechanisms in *X. tropicalis*. One possibility is that there is a combination of these mechanisms, such as a loss of function allele for male differentiation on the W-chromosome (which causes WZ individuals to be phenotypically female and ZZ individuals to be male). However, this is inconsistent with previous findings that WZZ triploids develop into females [28]. Additionally, WWY triploids develop into males, which argues against sensitivity of the Y-chromosome to dosage of the W- [28]. A more plausible alternative is that the W-chromosome carries female-determining allele whose function is not present on the Z-chromosome, whereas the Y-chromosome carries a dominant negative regulator of the female-determining allele on the W. A dominant negative regulatory role has been proposed for *dm-w*, which is a W-linked trigger for female differentiation in *X. laevis* over the male-related *dmrt1* gene (which is autosomal) and closely related to *dmw* by partial gene duplication [37, 48]. Future efforts aimed at identifying the variants that trigger female and male variation in *X. tropicalis* is crucial to unravel their fascinating evolutionary histories and genetic interactions.

Signs of Age in Cytologically Indistinguishable Sex Chromosomes

During gonadal differentiation, a total of 151 transcripts in the gonad/mesonephros transcriptome were identified with significant sex bias and a known genomic location; one third of these transcripts (n = 50) were in the $\sim$ 11Mb male-linked region of chromosome 7 in in v10 (Fig. 4, S5), which comprises less than one hundred and fiftieth of the 1.7 Gb genome of *X. tropicalis* [49]. Put another way, we detected 50 times more significantly sex-biased transcripts in sex-linked region of chromosome 7 than expected if these transcripts were encoded by genes that were uniformly distributed throughout the genome. Of the transcripts in this

genomic region with significantly sex-biased expression (n = 46), almost all were male-specific (n = 27) or male-biased (n = 18), none were female-biased, and only one was female-specific (Table S2). An excess of sex-linked genes with male-biased expression was also observed in adult tissues, although that excess was not significant [50], which is a possible consequence of the lower quality genome assembly that was available at that time for determining the genomic locations of transcripts. In the sex-linked portion of the *X. tropicalis* sex chromosomes there exists more substantial nucleotide divergence compared to the rest of the genome (Figs. 3, S2) and divergence between the sex-linked regions of the W-chromosome and the Z- or Y-chromosomes is higher than that between the Z- and Y-chromosomes (Fig. 5). This information, combined with the observation that the closely related tetraploid species *X. mellotropicalis* has a female-linked genomic region in a homologous location to *X. tropicalis* [26], suggests that the Y-chromosome evolved from the Z-chromosome rather than from the W-chromosome. If the Y-chromosome eventually fixes in *X. tropicalis* (and the Z- goes extinct), the mechanism of turnover would appear to follow the scenario depicted in Table 1D of [51] but with an ancestor with female heterogamy and a descendant with male heterogamy. In the absence of dosage compensation, sex chromosome turnover may be favored due to the accumulation of deleterious mutations and associated lowered or lost expression of alleles on the non-recombining sex chromosome [21,22]. However, in *X. tropicalis* this scenario does not appear to apply since the degenerate W-chromosome is staged to survive a transition to male heterogamy if the Y-chromosome fixes in the future.

Transcriptome evolution is influenced by genetic drift and natural selection in regulatory regions of genes. Due to differences in effective population size (N_e), ploidy, recombination, and mutation, the effects of these phenomena are expected to be distinctive between the sex chromosomes and autosomes [52, 53]. Several factors have the potential to influence regulatory evolution on sex chromosomes, such as faster-X or faster-Z effects [54, 55]). The faster-X effect may be heightened in species with dosage compensation [54], although there is no strong evidence of dosage compensation in amphibians [56]. Evidence presented here is most consistent with degeneration of the W-chromosome, presumably prior to the origin of the Y-chromosome from the Z-chromosome, as a mechanism for male-biased expression of transcripts encoded by sex-linked loci. For example, a comparison between putative WW females and WY males identifies more substantial sex-biased expression than a comparison between WZ females and ZY males (Suppl. Fig. S7, S8). We also detected higher polymorphism in expressed transcripts encoded by genes in the sex-linked region of putative WZ and WY individuals than in transcripts encoded by non-sex-linked (autosomal and pseudoautosomal) transcripts. This is also suggestive of divergence due to recombination suppression on the sex-linked portion of the W-chromosome. One prediction that is associated with the mechanism behind male-biased expression of sex-linked transcripts in *X. tropicalis* is that *X. mellotropicalis* should also have a degenerate W-chromosome

and also exhibit male-biased expression in the sex-linked portion of its sex chromosomes. This is another interesting direction for further exploration.

Sex Differences in Recombination

In *X. tropicalis* from Ghana, the rate of recombination is higher during oogenesis than spermatogenesis, and the crossover densities vary during these meiotic events as well, with proportionately more crossovers occurring in more central region of chromosomes during oogenesis compared to spermatogenesis (Fig. 6 B). This pattern was evident in a relatively small sample of crossover events that were observed in two biological replicates, each one generation in length, but are congruent with results recovered from the other two *Xenopus* species examined so far – *X. laevis* and *X. borealis* [57]. A lower density of crossover events in the center of chromosomes was also detected in another study of *X. tropicalis* [30], although sex-differences in these densities were not evaluated in that study. Overall, this suggests that these sex-biases in recombination rate and location are widespread in *Xenopus*, including across ploidy levels (*X. laevis* and *X. borealis* are both allotetraploid), and probably as well the most recent common ancestor of extant *Xenopus*. In several other species, including other frogs, the recombination rate is also higher in females compared to males, though the opposite pattern has also been observed [58–64]. Paternal crossovers are more concentrated at the ends of chromosomes than maternal crossovers in other vertebrates as well, including humans [58–60,63,64]. Why sex differences in the locations of recombination exist is not entirely clear, but is mechanistically achieved by sex-differences in the rate that double strand breads occur and in the rate that they are resolved into crossover or gene conversion events [64], which are influenced by the unique ways that meiosis occurs in females and males [65].

These sex-differences in recombination rate and location have interesting ramifications for the genomic positions of male-specific and female-specific variation on sex-chromosomes. In females, triggers for female-determination should frequently be located on the ends of a W-chromosome because there they should be disrupted less frequently by recombination as compared to alleles that are not near chromosome ends [64]. This prediction is supported by the W-chromosome of *X. tropicalis* and in *X. laevis* where another female-determining gene – *dm-w* – is also positioned on the end of a chromosome (2L) [37], where the rate of recombination in females is relatively low [57]. The position of the male-determining factor on the end of the Y-chromosome is not expected because recombination is higher in this region. However, it appears that suppressed recombination between the W and Z-chromosomes <11 Mb was already in place prior to the origin of the Y-, and this would presumably prevent disruption of the trigger for male differentiation on the

Y-.

Outlook

We report here the co-occurrence of W-, Z-, and Y-chromosomes in a natural population of *X. tropicalis* from Ghana. We identified a high density of transcripts with a strong skew towards male-biased expression that originate from a small, differentiated, sex-linked genomic region in this frog. The findings of this study are consistent with the expectation that recombination suppression can lead to degeneration of sex-chromosomes [66]. These results also evidence W-chromosome degeneracy in a species with cytologically undifferentiated sex chromosomes, show a small male-linked region on the Y-chromosome overlaps with a female-linked region of the W-chromosome [30], and demonstrate that these three sex chromosome co-occur in the same population in nature. These findings open the possibility that variation at a single locus or a set of tightly linked loci define the three sex chromosomes of *X. tropicalis*, with alternative pairings of these variants governing whether an individual develops into a female or male. Exactly what genetic variation governs sex determination in *X. tropicalis* and how this variation is distributed across the natural range of this species remain uncharacterized, and are a promising direction for future efforts. Together, these features illustrate that several characteristics that are frequently attributed to old sex chromosomes (regulatory degeneration, nucleotide divergence) can in fact be present before divergence is detectable at the cytogenetic level, and persist through the evolutionary windows during which new sex chromosomes arise and replace ancestral sex chromosomes.

Methods

Genetic Samples; Reduced Representation Genome Sequencing

To study sex-linkage, recombination, and population structure in *X. tropicalis*, we performed reduced representation genome sequencing (RRGS, [67]) on laboratory generated and wild caught individuals. The RRGS samples included 22 and 21 female and male offspring, respectively, from Family 1 whose parents were both from west Ghana (mother: BJE4359; father: BJE4360), seven and five female and male offspring, respectively, from Family 2 whose parents originated from east Ghana (mother: BJE4361; father: BJE4362), both parents from both of these families, 18 and seven additional wild caught samples from Ghana west and Ghana east, eight samples from individuals derived from Sierra Leone, and one from an individual derived

from Nigeria. Parents of the lab crosses were performed at higher (∼four times) coverage than the offspring in order to increase the genotype quality in these individuals. Libraries were constructed with the *Sbf1* restriction enzyme by Floragenex (Portland, Oregon, USA), and multiplexed on one lane of an Illumina 2500 machine.

The wild *X. tropicalis* samples were collected from two locations near the western and eastern borders of Ghana: Ankasa Nature Reserve (GPS: 5.24424 -2.64044, altitude: 48 m; Ghana west), and near the town of Admedzofe (GPS: 6.83165 0.43642, altitude: 738 m; Ghana east). Families from each population were generated by injecting parents with Human Chorionic Gonadrotropin (Biovendor, Asheville, USA) to induce ovulation and clasping, and offspring were reared until post-metamorphic maturation. The Sierra Leone individuals (four females, four males) and a Nigeria individual (a female) were derived from georeferenced populations that were maintained at the Station de Zoologie Expérimentale at the University of Geneva [68]. The phenotypic sex of lab offspring were determined by surgical examination of gonads after euthanasia via transdermal overdose of MS222 (Sigma-Aldrich, St. Louis, MO, USA). The sexes of individuals from Ghana, Nigeria, and Sierra Leone were determined based on external morphology (females with larger size and larger cloacal lobes; males with nuptial pads on the forearms and smaller cloacal lobes). DNA was extracted from webbing, liver, muscle, or blood using the DNeasy blood and tissue extraction kit (Qiagen, Toronto, Canada).

RRGS reads from each individual were de-multiplexed using using Radtools [69], and trimmed with Trimmomatic version 0.39 [70], enforcing a minimum length of 36 base pairs (bp), removing 3 bp from the leading and trailing ends, and requiring less than four ambiguous base pairs in a sliding window of 15 base pairs. This resulted in an average of 5,000,000 reads per individual (range ∼700,000–20,000,000). We aligned these data to the *X. tropicalis* genomes v9.1 and v10.0 using BWA [71], and used samtools/bcftools [72, 73] to call genotypes. We set individual genotypes to missing that did not have a minimum depth of 15, had a genotype quality below 20, or for the offspring. Additionally, all individual genotypes were discarded from a genomic position if >20% of laboratory offspring had missing genotypes, Hardy-Weinberg equilibrium in the lab offspring was violated, or >10% of lab offspring had a genotype that was not possible given the parental genotypes. This last category of sites are often a consequence of genotyping errors where heterozygous positions are called as homozygous [57, 74, 75]. For Family 1, we also filtered any individuals that had greater than 20% missing data, leaving 36 offspring; we did not apply the same quality filter to the east population, because of the substantially smaller family size. Finally, we filtered the data to one randomly selected SNP per restriction-site associated region (RADTag) in each family.

Analysis of Sex-linkage

With the filtered SNP datasets for the two families aligned to v9 and v10, we calculated allelic association with sex following [76]; results were essentially the same for both genomes and v10 is presented here. This analysis was performed within each of our two families (Family 1 and 2) for bi-allelic sites that were heterozygous in the father or mother. Genotyping errors can reduce power to detect sex linkage, so we developed an approach to detect putative genotyping errors that resembled double recombination events in a small genomic window, and set them to missing data, thereby reducing their impact (additional details are provided the Supplement). This substantially reduced the frequency of false positive signals of genotype association with sex (comparing Fig. S1 to Fig. S11).

In an attempt to narrow down the sex-linked region in populations of *X. tropicalis* with male heterogamy beyond the signal that was present in the RRGS data, we used Sanger sequencing to survey for sex-linked variants. We analyzed both of our laboratory crosses, our wild caught samples from both of these localities in Ghana, and male and female individuals from a colony at the National *Xenopus* Resource, Woods Hole, MA, USA, that are thought to be derived from Ivory Coast and Nigeria. We focused these efforts on genomic regions in the vicinity of the sex-linked regions that were identified by our RRGS analysis, and also that were identified by another study of a strain of *X. tropicalis* which has female heterogamy [30].

Sex Chromosome Differentiation and Population Subdivision

To test for differentiation of the sex chromosomes, we quantified F_{ST} between females and males for each SNP following the bi-allelic F_{ST} approach of [77]. Because we do not know the sex chromosome genotype of almost all individuals for which we performed RRGS (the parents of Family 2 are an exception, see Results), we were unable to evaluate F_{ST} between cohorts of females and males that each had the same sex genotype. Instead, we evaluated F_{ST} between males and females across all samples for which we collected RRGS data, except the offspring of the two laboratory crosses. This included in wild individuals from Ghana east (1 female, 8 males), Ghana west (6 females, 14 males), and georeferenced lab individuals from Sierra Leone (4 females, 4 males) and Nigeria (1 female).

For the samples originating from Sierra Leone, Nigeria, and Ghana – including the parents of each laboratory cross but not including the offspring of these crosses – we assessed admixture proportions by analyzing the RRGS data using NGSadmix version 32 [78]. We removed reads with a map quality <20 from the bam files, set the `SNP_pval` (likelihood of their being a SNP) parameter of NGSadmix to 1e-6, and used

a minimum minor allele frequency (`minMAF`) of 0.05. We estimated genetic ancestry for partitions of 1–5 clusters (K), and ran 20 replicates for each value of K. We used CLUMPP [79] to combine the replicates while averaging the population assignments and correcting for label switching.

Transcriptome Analysis

We dissected gonad/mesonephros tissue from 14 tadpoles at developmental stage 50 [80] that were offspring of Family 3, which had a wild caught father from Ghana east that was used in the RRGS (BJE4362) and one of his daughters from Family 2 (BJE4687). Tadpole stage 50 was chosen for analysis because this is the stage where expression of the sex determining gene *dm-w* has been detected in *X. laevis* [37]. The tadpole gonad/mesonephros tissue was preserved in RNAlater, and RNA was extracted individually from each sample using the RNeasy micro kit (Qiagen, Toronto, Canada). For each tadpole, we also preserved tail tissue in ethanol, and genomic DNA was extracted using the DNeasy kit.

Based on our results from RRGS and Sanger sequencing (see Results), the sex of each tadpole from the second Ghana east cross was determined based on the presence (males) or absence (females) of heterozygous genotypes in two completely or almost completely sex-linked amplicons based on results from the Ghana east RRGS data and family (LOC100488897, primers: Scaf2_f1 + Scaf2_r2 and SNP1, primers: trop_east_SNP1_F1 + trop_east_SNP1_R1; Table S1). In the Ghana east family that was used for the RRGS data, the first of these amplicons had a sex-specific heterozygous SNP in five of five sons and none of seven daughters, and the second of these amplicons had an almost sex-specific SNP in four of five sons and none of seven daughters (Table 1). For the tadpoles that were used for RNAseq, heterozygosity at both of these amplicons was concordant for all individuals in the sense that heterozygosity was observed either at both amplicons or at neither amplicon (results from these tadpoles are not presented in Table 1 because we were not able to infer sex from adult individuals). This effort indicated that nine of the 14 tadpoles used in the RNAseq analysis were female and five were male. The accuracy of this indirect approach to sexing these tadpoles is evidenced by the very strong signature of expression divergence in the sex-linked region (Fig. 4).

Library preparation and transcriptome sequencing was performed the Centre for Applied Genomics (Toronto, Canada), multiplexing all 14 samples on one lane of an Illumina 2500 machine and 150 base pair reads. Reads were trimmed using Trimmomatic version 0.36, removing the first and last three bases, retaining reads with a minimum length of 36 base pairs, and a 'maxinfo' setting of 30 and 0.7 (which determines the nature of an adaptive quality trim that aims to balance the benefits of preserving longer reads

against the costs of retaining sequences that have errors). A *de novo* transcriptome assembled using Trinity version 2.8.2 with a minimum k-mer coverage of two.

Transcript counts were quantified for each sample using Kallisto v.0.43.0 following the methods of [81] with default parameters for indexing (using a kmer size of 31) and quantification (using quant parameter settings: -b 0 -t 1). We discarded genes with an average of less than one raw read per sample. Read counts from Kallisto were then used for differential expression between males and females with the EdgeR package version 3.4, using the vanilla pipeline (i.e., calcNormFactors, estimateCommonDisp, estimateTagwiseDisp, exactTest for comparison between males and females), following the edgeR vignette [82]. EdgeR was used to calculate the $\log_2$-transformed male/female expression ratio (logFC), wherein values above or below zero indicate genes that are more highly expressed in males or females, respectively. To avoid ratios equal to zero or undefined, a default prior count of 0.125 was added to all samples using the exactTest function of edgeR. Using estimated read counts from Kallisto, we also performed an independent differential expression analysis with the DESeq2 package [83], following the DESeq2 vignette. Shrinkage was used with adaptive t prior shrinkage estimator from the package "apeglm" [84]. This option reduces the mean squared error of expression levels of each gene relative to the classical estimator, especially for genes with low expression levels [83,84]. Because the results of the EdgeR and DeSeq2 analyses were similar (Fig. S12), we report only the EdgeR results here.

We defined significantly sex-biased genes based on a false discovery rate (FDR) with Benjamini-Hochberg correction cutoff of 0.05 from the EdgeR output, and requiring the absolute value of logFC to be > 2. Genomic locations of individual transcripts were then ascertained based on the best match of each transcript against the *X. tropicalis* v10 (NCBI BioProject AAMC00000000.4; GenBank Assembly submission GCA_000004195.4) using a splice-aware aligner GMAP [85]. Median expression values for each sex were quantified after transcripts per million normalization (TPM) [86]. Confidence intervals for these medians were obtained using the DescTools package [87].

To quantify variation in expressed transcripts encoded by sex-linked genes, for each offspring from Family 3 we mapped the RNAseq data to the transcriptome assembly and called genotypes using bwa and bcftools. We filtered genotypes with less than 4X coverage, genotype quality below 20, and map quality below 20. Pairwise nucleotide diversity was then calculated for each individual using a perl script and collated with transcript location as assessed above using R.

Linkage Mapping

In order to evaluate whether and how the rates and genomic locations of recombination differ between the sexes, we used the RRGS data to build and compare sex-specific linkage maps for each chromosome of the *X. tropicalis* families using Onemap v1.0 [88] and v9.1 [30]. Using the same approach as [57], we first identified the largest linkage group per chromosome using all genotypes (maternal-specific, paternal-specific, or both parents heterozygous), setting the minimum logarithm of the odds (LOD) score to five and the maximum recombination fraction to 0.4. We then separated heterozygous markers that were maternal-specific or paternal-specific for each chromosomal linkage map, and reconstructed sex-specific linkage maps for each chromosome, using a minimum LOD score of three. In this way we were able to reconstruct sex-specific rates and locations of recombination during oogenesis and spermatogenesis, respectively. Ordering of markers used in the linkage map was based on their mapping positions to the v9.1 genome.

After an initial build, we inspected individuals and set as missing data any single markers or sets of markers within a 10 Mb window that indicated a double recombination event, under the assumption that these genotypes are most likely due to genotyping errors because two recombination events are usually rare in very small genomic windows. We then reconstructed a sex- and chromosome-specific linkage maps with these filtered sets of markers. To determine how chromosome lengths (as covered by markers used in the linkage map) related to inferred map lengths for both families, we used a linear model with fixed effects of sex in which recombination occurs, family used for the linkage map, and interaction between those fixed effects, and a three-way interaction between sex, family, and amount of base-pairs covered by the extreme markers used in the linkage map (i.e., $map\ length \sim sex * family + sex : family : bp\_covered$). Residuals were evaluated for non-normality to ensure proper model fit. Analyses were performed in R using the `lm` function and confidence intervals were generated with `confint` [89].

Acknowledgements

This work was supported by the Natural Science and Engineering Research Council of Canada (RGPIN-2017-05770), Resource Allocation Competition awards from Compute Canada, the Whitman Center Fellowship Program at the Marine Biological Laboratory, the Museum of Comparative Zoology at Harvard University, and National Institutes of Health grants R01-HD084409 and P40-OD010997. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank

Brian Golding for access to computational resources, and Jessen Bredeson, Sofia Medina Ruiz, Dan Rohksar and their colleagues, and Xenbase [90], for making the v10 *X. tropicalis* genome assembly available.

Ethics Statement

This work was approved by the Animal Care Committee at McMaster University (AUP# 17-12-43).

Data Accessibility

The RRGS and RNAseq data from *X. tropicalis* have been deposited in the Short Read Archive of NCBI (Bioproject PRJNA627066) as has the RNAseq data from *X. borealis* (accession number SRR11844024-5, SRR11844031, SRR11844033-42). Representative Sanger sequences have been deposited in GenBank (accession numbers XXX). Colonies of offspring of wild individuals from Ghana east are available upon request from the National Xenopus Resource (USA) or McMaster University, and are also being raised at the Institute for Amphibian Biology in Hiroshima, Japan. Colonies of offspring from Ghana west are available upon request from McMaster University.

References

1. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, et al. Sex determination: Why so many ways of doing it? PLoS Biology. 2014;12(7):e1001899.

2. Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. Science. 1999;286(5441):964–967.

3. Zhou Q, Zhang J, Bachtrog D, An N, Huang Q, Jarvis ED, et al. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science. 2014;346(6215):1246338.

4. Matsubara K, Tarui H, Toriba M, Yamada K, Nishida-Umehara C, Agata K, et al. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. Proceedings of the National Academy of Sciences. 2006;103(48):18190–18195.

5. Bachtrog D. Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. Nature Reviews Genetics. 2013;14(2):113–124.

6. Wright AE, Dean R, Zimmer F, Mank JE. How to make a sex chromosome. Nature Communications. 2016;7(1):1–8.

7. Deakin J. Chromosome evolution in marsupials. Genes. 2018;9(2):72.

8. Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, et al. A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). PLoS Genetics. 2012;8(7):e1002798.

9. Adolfsson S, Ellegren H. Lack of dosage compensation accompanies the arrested stage of sex chromosome evolution in ostriches. Molecular Biology and Evolution. 2013;30(4):806–810.

10. Perrin N. Sex reversal: A fountain of youth for sex chromosomes? Evolution: International Journal of Organic Evolution. 2009;63(12):3043–3049.

11. Vicoso B, Kaiser VB, Bachtrog D. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proceedings of the National Academy of Sciences. 2013;110(16):6453–6458.

12. Bull JJ. Evolution of sex determining mechanisms. The Benjamin/Cummings Publishing Company, Inc.; 1983.

13. Volff JN, Nanda I, Schmid M, Schartl M. Governing sex determination in fish: Regulatory putsches and ephemeral dictators. Sexual Development. 2007;1(2):85–99.

14. Miura I. An evolutionary witness: The frog *Rana rugosa* underwent change of heterogametic sex from XY male to ZW female. Sexual Development. 2007;1(6):323–331.

15. Jeffries DL, Lavanchy G, Sermier R, Sredl MJ, Miura I, Borzée A, et al. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nature Communications. 2018;9(1):4088.

16. Saunders PA, Neuenschwander S, Perrin N. Impact of deleterious mutations, sexually antagonistic selection, and mode of recombination suppression on transitions between male and female heterogamety. Heredity. 2019;123(3):419–428.

785 17. Vicoso B. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nature Ecology
786 and Evolution. 2019; p. 1–10.

787 18. Myosho T, Takehana Y, Hamaguchi S, Sakaizumi M. Turnover of sex chromosomes in celebensis group
788 medaka fishes. G3: Genes, Genomes, Genetics. 2015;5(12):2685–2691.

789 19. Tennessen JA, Wei N, Straub SC, Govindarajulu R, Liston A, Ashman TL. Repeated translocation of
790 a gene cassette drives sex-chromosome turnover in strawberries. PLoS Biology. 2018;16(8):e2006062.

791 20. Yano A, Nicol B, Jouanno E, Quillet E, Fostier A, Guyomard R, et al. The sexually dimorphic on the
792 Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids.
793 Evolutionary applications. 2013;6(3):486–496.

794 21. Blaser O, Grossen C, Neuenschwander S, Perrin N. Sex-chromosome turnovers induced by deleterious
795 mutation load. Evolution: International Journal of Organic Evolution. 2013;67(3):635–645.

796 22. Blaser O, Neuenschwander S, Perrin N. Sex-chromosome turnovers: The hot-potato model. The
797 American Naturalist. 2014;183(1):140–146.

798 23. Herpin A, Schartl M. Plasticity of gene-regulatory networks controlling sex determination: Of masters,
799 slaves, usual suspects, newcomers, and usurpators. EMBO Reports. 2015;16(10):1260–1274.

800 24. Evans BJ, Pyron RA, Wiens JJ. Polyploidization and sex chromosome evolution in amphibians. In:
801 Polyploidy and Genome Evolution. Springer; 2012. p. 385–410.

802 25. Pennell MW, Mank JE, Peichel CL. Transitions in sex determination and sex chromosomes across
803 vertebrate species. Molecular Ecology. 2018;27(19):3950–3963.

804 26. Cauret CM, Gansauge MT, Tupper AS, Furman BL, Knytl M, Song XY, et al. Developmental systems
805 drift and the drivers of sex chromosome evolution. Molecular Biology and Evolution. 2020;37(3):799–
806 810.

807 27. Green DM, Zeyl CW, Sharbel TF. The evolution of hypervariable sex and supernumerary (B) chro-
808 mosomes in the relict New Zealand frog, *Leiopelma hochstetteri*. Journal of Evolutionary Biology.
809 1993;6(3):417–441.

28. Roco ÁS, Olmstead AW, Degitz SJ, Amano T, Zimmerman LB, Bullejos M. Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis*. Proceedings of the National Academy of Sciences. 2015;112(34):E4752–E4761.

29. Olmstead AW, Lindberg-Livingston A, Degitz SJ. Genotyping sex in the amphibian, *Xenopus* (*Silurana*) *tropicalis*, for endocrine disruptor bioassays. Aquatic Toxicology. 2010;98(1):60–66.

30. Mitros T, Lyons J, Session A, Jenkins J, Shu S, Kwon T, et al. A chromosome-scale genome assembly and dense genetic map for *Xenopus tropicalis*. Developmental Biology. 2019;.

31. Tymowska J. Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. Amphibian cytogenetics and evolution. 1991;259:297.

32. Bewick AJ, Chain FJ, Zimmerman LB, Sesay A, Gilchrist MJ, Owens ND, et al. A large pseudoautosomal region on the sex chromosomes of the frog *Silurana tropicalis*. Genome Biology and Evolution. 2013;5(6):1087–1098.

33. Evans BJ, Gansauge MT, Stanley EL, Furman BL, Cauret CM, Ofori-Boateng C, et al. *Xenopus fraseri*: Mr. Fraser, where did your frog come from? PloS One. 2019;14(9).

34. Grainger RM. *Xenopus tropicalis* as a model organism for genetics and genomics: Past, present, and future. In: *Xenopus* Protocols. Springer; 2012. p. 3–15.

35. Blum M, Ott T. *Xenopus*: an undervalued model organism to study and model human genetic disease. Cells Tissues Organs. 2018;205(5-6):303–313.

36. Showell C, Conlon FL. The Western clawed frog (*Xenopus tropicalis*): An emerging vertebrate model for developmental genetics and environmental toxicology. Cold Spring Harbor Protocols. 2009;2009(9):pdb–emo131.

37. Yoshimoto S, Okada E, Umemoto H, Tamura K, Uno Y, Nishida-Umehara C, et al. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. Proceedings of the National Academy of Sciences. 2008;105(7):2469–2474.

38. Song XY, Furman BL, Premachandra T, Knytl M, Cauret CMS, Wasonga DV, et al. Sex-biased expression of sex-linked transcripts in African clawed frogs (*Xenopus*). Philosophical Transactions of the Royal Society B;submitted.

837 39. Furman BLS, Evans BJ. Sequential turnovers of sex chromosomes in African clawed frogs (*Xeno-*
838 *pus*) suggest some genomic regions are good at sex determination. G3: Genes|Genomes|Genetics.
839 2016;6(11):3625–3633.

840 40. Tinsley R, Loumont C, Kobel H. Geographical distribution and ecology. In: Tinsley R, Kobel H,
841 editors. The Biology of *Xenopus*. Oxford: Clarendon Press; 1996. p. 35–41.

842 41. Miller CS, Gosling WD. Quaternary forest associations in lowland tropical West Africa. Quaternary
843 Science Reviews. 2014;84:7–25.

844 42. Demenou BB, Doucet JL, Hardy OJ. History of the fragmentation of the African rain forest in the Da-
845 homey Gap: Insight from the demographic history of *Terminalia superba*. Heredity. 2018;120(6):547–
846 561.

847 43. Düsing K. The regulation of the gender ratio in the multiplication of the people, animals and plants.
848 Fischer; 1884.

849 44. Fisher RA. The genetical theory of natural selection. ; 1958.

850 45. Hamilton WD. Extraordinary sex ratios. Science. 1967;156(3774):477–488.

851 46. Vuilleumier S, Lande R, Van Alphen J, Seehausen O. Invasion and fixation of sex-reversal genes.
852 Journal of Evolutionary Biology. 2007;20(3):913–920.

853 47. Bateman A, Anholt B. Maintenance of polygenic sex determination in a fluctuating environment: an
854 individual-based model. Journal of Evolutionary Biology. 2017;30(5):915–925.

855 48. Bewick AJ, Anderson DW, Evans BJ. Evolution of the closely related, sex-related genes DM-W and
856 DMRT1 in African clawed frogs (*Xenopus*). Evolution: International Journal of Organic Evolution.
857 2011;65(3):698–712.

858 49. Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, et al. The genome of the
859 Western Clawed Frog *Xenopus tropicalis*. Science. 2010;328(5978):633–636.

860 50. Chain FJ. Sex-biased expression of young genes in *Silurana (Xenopus) tropicalis*. Cytogenetic and
861 Genome Research. 2015;145(3-4):265–277.

862 51. Bull JJ, Charnov EL. Changes in the heterogametic mechanism of sex determination. Heredity.
863 1977;39(1):1.

52. Connallon T, Clark AG. Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. Evolution: International Journal of Organic Evolution. 2010;64(12):3417–3442.

53. Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics. 2009;10(3):195–205.

54. Charlesworth B, Coyne JA, Barton NH. The relative rates of evolution of sex chromosomes and autosomes. The American Naturalist. 1987;130(1):113–146.

55. Vicoso B, Charlesworth B. Effective population size and the faster-X effect: an extended model. Evolution: International Journal of Organic Evolution. 2009;63(9):2413–2426.

56. Malcom JW, Kudra RS, Malone JH. The sex chromosomes of frogs: Variability and tolerance offer clues to genome evolution and function. Journal of Genomics. 2014;2:68.

57. Furman BL, Dang UJ, Evans BJ, Golding GB. Divergent subgenome evolution after allopolyploidization in African clawed frogs (*Xenopus*). Journal of Evolutionary Biology. 2018;31(12):1945–1958.

58. Ottolini CS, Newnham LJ, Capalbo A, Natesan SA, Joshi HA, Cimadomo D, et al. Genome-wide maps of recombination and chromosome segregation in human oocytes and embryos show selection for maternal recombination rates. Nature Genetics. 2015;47(7):727.

59. Sardell JM, Cheng C, Dagilis AJ, Ishikawa A, Kitano J, Peichel CL, et al. Sex differences in recombination in sticklebacks. G3: Genes, Genomes, Genetics. 2018;8(6):1971–1983.

60. Brelsford A, Dufresnes C, Perrin N. High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. Heredity. 2016;116(2):177.

61. Berset-Brändli L, Jaquiéry J, Broquet T, Ulrich Y, Perrin N. Extreme heterochiasmy and nascent sex chromosomes in European tree frogs. Proceedings of the Royal Society B: Biological Sciences. 2008;275(1642):1577–1585.

62. Theodosiou L, McMillan W, Puebla O. Recombination in the eggs and sperm in a simultaneously hermaphroditic vertebrate. Proceedings of the Royal Society B: Biological Sciences. 2016;283(1844):20161821.

63. Sutherland BJ, Rico C, Audet C, Bernatchez L. Sex chromosome evolution, heterochiasmy, and physiological QTL in the salmonid brook charr *Salvelinus fontinalis*. G3: Genes, Genomes, Genetics. 2017;7(8):2749–2762.

64. Sardell JM, Kirkpatrick M. Sex differences in the recombination landscape. The American Naturalist. 2020;195(2):361–379.

65. Brandvain Y, Coop G. Scrambling eggs: meiotic drive and the evolution of female recombination rates. Genetics. 2012;190(2):709–723.

66. Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2000;355(1403):1563–1572.

67. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3(10):e3376.

68. Rungger D. *Xenopus helveticus*, an endangered species? International Journal of Developmental Biology. 2002;46(1):49–63.

69. Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, et al. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. PloS One. 2011;6(4):e19315.

70. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014; p. btu170.

71. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754–1760.

72. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079.

73. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–2993.

74. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics. 2016;17(2):81.

75. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PloS One. 2014;9(2):e90346.

76. Goudet J, Raymond M, de Meeüs T, Rousset F. Testing differentiation in diploid populations. Genetics. 1996;144(4):1933–1940.

77. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: The impact of rare variants. Genome Research. 2013;23(9):1514–1521.

78. Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. Genetics. 2013;195(3):693–702.

79. Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23(14):1801–1806.

80. Nieuwkoop PD. Normal table of *Xenopus laevis* (Daudin). Normal table of *Xenopus laevis* (Daudin). 1956; p. 162–203.

81. Bray NL, Pimentel P Haroldand Melsted, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology. 2016;34(5):525–527. doi:10.1038/nbt.3519.

82. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–140. doi:10.1093/bioinformatics/btp616.

83. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550. doi:10.1186/s13059-014-0550-8.

84. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019;35(12):2084–2092. doi:10.1093/bioinformatics/bty895.

85. Wu TD, Watanabe CK. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 2005;21(9):1859–1875. doi:10.1093/bioinformatics/bti310.

86. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in biosciences. 2012;131(4):281–285.

87. Signorell A, et al. DescTools: Tools for Descriptive Statistics; 2020. Available from: `https://cran.r-project.org/package=DescTools`.

88. Margarido GRA, de Souza AP, Garcia AAF. OneMap: Software for genetic mapping in outcrossing species. Hereditas. 2007;144:78–79.

89. R Core Team. R: A Language and Environment for Statistical Computing; 2016. Available from: `https://www.R-project.org/`.

90. Karimi K, Fortriede JD, Lotay VS, Burns KA, Wang DZ, Fisher ME, et al. Xenbase: a genomic, epigenomic and transcriptomic model organism database. Nucleic Acids Research. 2018;46(D1):D861–D868.

91. Wei KH, Bachtrog D. Ancestral male recombination in *Drosophila albomicans* produced geographically restricted neo-Y chromosome haplotypes varying in age and onset of decay. PLoS genetics. 2019;15(11).

Appendix F

Sex chromosome degeneration, turnover, and sex-biased expression of sex-linked transcripts in African clawed frogs (Xenopus)

This manuscript has been submitted and is under review.

Sex chromosome degeneration, turnover, and sex-biased expression of sex-linked transcripts in African clawed frogs (*Xenopus*)

Xue-Ying Song[1], Benjamin L. S. Furman[1,2,*], Tharindu Premachandra[1], Martin Knytl[1,3], Caroline M. S. Cauret[1], Domnick Victor Wasonga[4], John Measey[5], Ian Dworkin[1], and Ben J. Evans[1,*]

[1]Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada

[2]Department of Zoology, University of British Columbia, 6270 University Blvd Vancouver, British Columbia, V6T 1Z4 Canada

[3]Department of Cell Biology, Charles University, 7 Vinicna Street, Prague, 12843, Czech Republic

[4]National Museums of Kenya, P.O. Box 40658 - GPO 00100, Nairobi, Kenya

[5]Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Private Bag X1, Matieland, 7602, Stellenbosch, South Africa

[*]Correspondence: evansb@mcmaster.ca

July 10, 2020

Abstract

The tempo of sex chromosome evolution – how quickly, in what order, why, and how their particular characteristics emerge during evolution – remain poorly understood. To understand this further, we studied three closely related species of African clawed frog (genus *Xenopus*), that each have independently evolved sex chromosomes. We identified population polymorphism in the extent of sex chromosome differentiation in wild-caught *X. borealis* that corresponds to a large, previously identified region of recombination suppression. This large sex-linked region of *X. borealis* has an extreme concentration of genes that encode transcripts with sex-biased expression, and we recovered similar findings

in the smaller sex-linked regions of *X. laevis* and *X. tropicalis*. In two of these species, strong skews in expression (mostly female-biased in *X. borealis*, mostly male-biased in *X. tropicalis*) are consistent with expectations associated with recombination suppression, and in *X. borealis*, we hypothesize that a degenerate ancestral Y-chromosome transitioned into its contemporary Z-chromosome. These findings demonstrate that Xenopus appear to be tolerant of differences between the sexes in dosage of the products of multiple genes, and offer insights into how evolutionary transformations of ancestral sex chromosome carry forward to affect function of new sex chromosomes.

Keywords: Comparative transcriptomics, recombination suppression, heterogamy, dosage tolerance

1. Introduction

Many vertebrates have separate male and female individuals rather than hermaphrodites, with sex-specific traits in each sex being regulated by transcripts with sex-biased expression. In species with genetic sex determination, genetic variation that triggers sexual differentiation is present on sex chromosomes. This is important because sex-specific genetic variation establishes a genomic compartment – which may expand – that has unique features as compared to autosomes (copy number, mode of inheritance, rates of mutation and recombination) that strongly influence evolution in important ways [1, 2]. For example, compared to autosomes, non-recombining portions of sex chromosomes tend to have a higher abundance of repetitive elements and more extensive pseudogenization [3], and distinctive evolutionary dynamics of non-neutral mutations that depend on the degree of dominance [4, 5]. Sex chromosomes evolve from autosomes and, in theory, genetic recombination near the genomic region that contains the master regulator of sex determination can decline or become absent due to several non-exclusive factors. For example, it is conceivable that neutral processes may drive recombination suppression [4]. Inheritance (or non-inheritance) of a non-recombined genomic region may be advantageous in order to avoid sterile progeny [6], genetic linkage to a sex determining locus may resolve genomic conflict associated with mutations with opposite fitness effects in each sex [5, 7], or recombination suppression could be favoured to maintain heterozygosity, especially in inbred populations [8]. Mechanistically, recombination suppression may be achieved by inversion or translocation [9, 10], or potentially without genomic rearrangement [11] due to genetic mutations that alter recombination (recombination modifiers; [12]).

There are costs of recombination suppression in sex-linked regions that are associated with Hill-Robertson effects, which reduce the efficacy of natural selection [13]. Some of these costs may be curtailed by the origin of and gene conversion among duplicates [14]. Benefits of recombination suppression also exist that are

related to sex-specific or sex-biased modes of inheritance. This can help resolve genomic conflict, and these regions can also foster the elaboration of sex-specific genetic functions, such as the many genes that are crucial for spermatogenesis on the male-specific portion of the human Y-chromosome [14]. Gene loss on the sex-specific sex chromosome (the Y- or W-) creates hemizygous genotypes, allowing natural selection to respond more efficiently to recessive alleles in the heterogametic sex (XY males, or WZ females); the associated rapid evolution is known as the faster-X (or faster-Z) effect [4, 15]. One of the consequences of their unique evolution is that sex chromosomes play an oversized role in the establishment of reproductive incompatibilities among species [4, 15, 16].

a. Signs of rapid sex chromosome evolution

In species that have separate sexes, sexual differentiation is usually a necessary prelude to reproduction, and disruption of this developmental process can have dire fitness consequences. Consequently, one might expect purifying selection on sex determination pathways – including the trigger for sex determination – to be strong. Contrary to this expectation, components of the developmental cascade that govern sexual differentiation actually vary considerably among species, including and especially the triggers for sex determination and the chromosomes on which they reside. This variation is sometimes evidenced by a known trigger for sex determination that is present in one lineage but not another. An example of this is the *Sry* locus, which triggers male differentiation in most placental and marsupial mammals, but not in monotremes [17]. Another line of evidence for rapid evolution comes from studies that document variation among species or populations in whether females or males are heterogametic. In Diptera, for instance, the heterogametic sex in some species is male but female in others, and sex-linked regions in several species are non-homologous [18]. Another example of rapid evolution involves dramatic differences in the extent of sex chromosome heteromorphy – microscopic differences in the size of the sex chromosomes or genomic differences in the magnitude and genomic extent of differentiation between sex chromosomes – with some species having non-diverged (homomorphic) sex chromosomes and others with heteromorphic sex chromosomes. These differences are due to variation in the genomic extent of recombination suppression surrounding a trigger for sex determination, and an example of this variation is found in livebearing teleost fish [19]. Some aspects of rapid evolution eventually act as negative feedbacks for further change. For example, sex chromosome hemizygosity associated with loss of alleles on the non-recombining portion of the sex-specific sex chromosome creates imbalances in the stoichiometries of interacting sex-linked and autosomal genes. This has the potential to arrest sex chromosome divergence [20], favor the origin of new sex chromosomes [21], or spur the evolution of systems for dosage compensation [22]. Over time, the origin of systems for dosage compensation and sex-

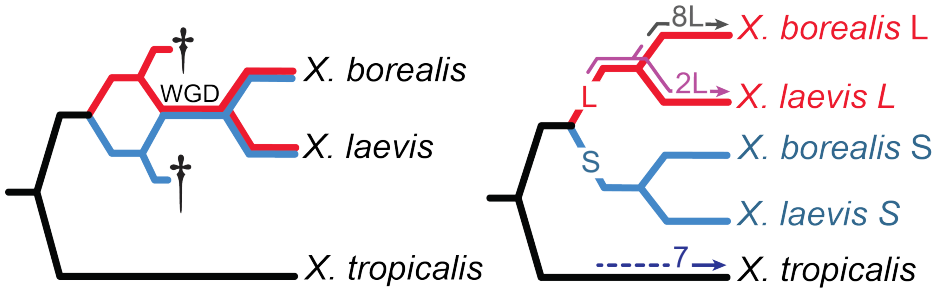


Figure 1: Evolutionary relationships among focal species (left) and their orthologs and homeologs (right) in this study. Allopolyploidization involved whole genome duplication (WGD) via the merger of two extinct ancestral species (daggers) that generated two subgenomes (L and S, red and blue) in the tetraploid species *X. borealis* and *X. laevis*. Sex chromosomes in each species are indicated on the right above branches in gray, pink or blue. The pink line corresponds with the origin of the female-determining gene *dm-w* on chr2L. A dashed blue line indicates the unknown timing of the origin of the *X. tropicalis* sex chromosomes; branches not to scale.

specific viability alleles may act to dampen the evolutionary meanderings of sex chromosomes [23].

b. Young sex chromosomes of *Xenopus* frogs

How can we make sense of the diverse and sometimes correlated or counterposed drivers that choreograph the evolutionary fates of sex chromosomes? One approach is to examine young sex chromosomes with an eye on identifying early steps in evolution, before they become blurred or confounded by subsequent change. To this end, the sex chromosome of African clawed frogs (genus *Xenopus*) – and particularly the Marsabit clawed frog, *Xenopus borealis* – provide compelling subjects. Within *Xenopus*, all of the examples of rapid sex chromosome evolution discussed above are present (Fig. 1). Variation in heterogamy exists within *X. tropicalis*, with variation on chromosome 7 (hereafter chr7) defining W-, Y-, and Z- sex chromosomes [24]. In *X. laevis* and several other closely related species, a recently evolved genetic trigger for female sex determination called *dmw* established chromosome 2L (chr2L) as the sex chromosomes [25]. *Xenopus tropicalis* diverged before *dmw* arose [26], whereas *X. borealis* lost *dmw* and evolved new sex chromosomes on chromosome 8L (chr8L; [27]). Though all *Xenopus* species have homomorphic sex chromosomes [28], there exists considerable variation in recombination suppression around the sex-linked portions of sex chromosomes in different *Xenopus*. *Xenopus laevis* and *X. tropicalis* have a very small region of recombination suppression, whereas this region in *X. borealis* comprises almost half of the sex chromosomes; the other half of chr8L undergoes recombination in both sexes, and therefore is a pseudoautosomal region [29, 30].

Xenopus borealis is an allotetraploid species, meaning that it was formed via the fusion of two diploid ancestral species (Fig 1; reviewed in [31]). For this reason, its genome contains autosomal chromosomes that

are homeologous to the sex chromosomes (chromosome 8S; chr8S). The closely related species *X. laevis* is also allotetraploid, shares a most recent common allotetraploid ancestor with *X. borealis* ~18-34 Ma [32] or older [33], and has autosomal chromosomes that are orthologous (chr8L) and homeologous (chr8S) to the sex chromosomes of *X. borealis*. Likewise, the sex chromosomes of *X. laevis* are orthologous and homeologous to autosomes of *X. borealis* (chr2L and chr2S, respectively). *Xenopus tropicalis*, a diploid, has autosomal chromosomes that are orthologous to the sex chromosomes of *X. borealis* and of *X. laevis* (chr7 and chr2, respectively), and the sex chromosomes of *X. tropicalis* are orthologous to pairs of homeologous autosomes in *X. borealis* and also in *X. laevis* (chr7L and chr7S – both of these homeologous chromosomes are in both of these allotetraploid species). *Xenopus tropicalis*, *X. laevis*, and *X. borealis* share a most recent common ancestor ~60 Ma [33]. Together, these three sets of alternatively sex-linked, orthologous, and/or homeologous chromosomes in these frogs provide informative contrasts with which to consider evolution of sex-linked regions (Fig. 1).

With an aim of better understanding sex chromosome dynamics in *Xenopus*, we used genome-wide approaches to test for population structure and variation in sex chromosome differentiation in natural populations of *X. borealis* sampled in their native range in Kenya. We then evaluated whether and how expression of genes in the sex-linked region of the sex chromosomes differs from the rest of the genome, and when these changes emerged during *Xenopus* diversification. We considered whole transcriptome data from *X. borealis* adult liver and gonad/mesonephros tissues from two tadpole stages during the onset of gonadal differentiation, and we also contextualize findings in *X. borealis* with similar comparisons with adult liver transcriptomes of *X. laevis* and *X. tropicalis*.

2. Materials and methods

a. Samples, analyses of population structure in *X. borealis*

Xenopus borealis naturally occurs in Kenya and northern Tanzania. We analyzed a total of 93 wild-caught *X. borealis* specimens that were sampled through the central portion of its range in Kenya in 2018. Information on the sampling localities, voucher specimens, and sex of these samples is provided in Table S1 and Fig. 2; specimens from this collection, including genetic samples, are accessioned at the Museum of Comparative Zoology at Harvard University, USA, and the National Museums of Kenya in Nairobi, Kenya. As described in the Supplement, we performed phylogenetic analysis on a portion of mitochondrial DNA from a subset of these samples (84 specimens) and also two laboratory specimens. We also collected and analyzed reduced representation genome sequences (RRGS, [34]) from a subset of these samples (51 specimens) combined with

whole genome sequencing data from the two laboratory specimens. These analyses allowed us to discern population structure and sex-chromosome differentiation within *X. borealis*.

b. Comparative transcriptomics of *X. borealis*, *X. laevis*, and *X. tropicalis*

We sequenced and analyzed transcriptomes from gonads at two juvenile developmental stages and adult liver from our laboratory strain of *X. borealis*. In order to gain an evolutionary perspective on these transcriptomes, we also analyzed transcriptomes from adult liver from *X. laevis* and *X. tropicalis*. *Xenopus borealis* and *X. laevis* specimens were obtained directly or were offspring of animals that were purchased from XenopusExpress (Brooksville, Florida, USA) and the *X. tropicalis* adult liver data were obtained from the NCBI-SRA (from two males: SRR5412273 and SRR5412274, and two females: SRR5412275 and SRR5412276). Details of RNA extraction and sequencing, and the transcriptome assembly and analysis of sex-biased expression are provided in the Supplement.

Genomic locations of assembled transcripts from each tetraploid species (*X. borealis* and *X. laevis*) were assessed in the *X. laevis* genome assembly version 9.2 [32], which was obtained from Xenbase [35], using the splice-aware aligner GMAP [36] with parameters: -A -B 5 -t 20 -f samse. The *X. tropicalis* transcripts were mapped to the version 10 *X. tropicalis* genome assembly (NCBI BioProject AAMC00000000.4; GenBank Assembly submission GCA_000004195.4) using the same approach. The parameter –cross-species was used for the *X. borealis* transcripts to accommodate divergence from the *X. laevis* reference genome.

In our laboratory strain of *X. borealis*, the sex-linked region of the sex chromosomes corresponds to positions 4,605,306 to 56,690,925 base pairs (bp) on chr8L of the *X. laevis* genome assembly version 9.2 [29]. For this study we assume that the sex-linked region starts from the beginning of chr8L because no RRGS data from [29] mapped to <4,605,306 bp on chr8L. In order to establish approximate margins of regions of orthology and homeology for sex-linked regions in each of our three focal species, we generated dot plots from chromosomes in the genome assemblies of *X. laevis* and *X. tropicalis* using Gopard version 1.40 [37] and re-DOTable version 1.0 [38]. Based on these plots, we concluded that the sex-linked region of *X. borealis* is homeologous to 16–45 and 51–68 Mb of *X. laevis* chr8S, and orthologous to 0–12 and 20–73 Mb of the *X. tropicalis* chromosome 8 (e.g., Supplemental Fig. S1). The sex-linked region of the *X. laevis* sex chromosomes are >178 Mb on chr2L [29] and, based on dotplots, this region is homeologous to 154-158 Mb on *X. laevis* chr2S and orthologous to 172-177 Mb on chr2 of *X. tropicalis*. The sex-linked region of the *X. tropicalis* sex chromosome is <11Mb on chr7 [30, 39]. This region is orthologous to <9Mb and <6Mb on *X.laevis* chr7L and chr7S, respectively.

For each tetraploid transcriptome, we used a generalised linear model to test whether a binary response variable – whether or not a transcript was significantly differentially expressed (1) or not (0) according to FDR corrected p-values from edgeR ($significance$) – was affected by the interaction between two predictor variables: whether a transcript was from the L (1) or S (0) subgenome ($subgenome$), and whether a transcript was encoded by a gene in the sex-linked region or the region that is homeologous to the sex-linked region (1), or by a gene from elsewhere in the genome (0) ($SL\_homeologous$):

$$significance \sim SL\_homeologous * subgenome$$

We had a one-sided expectation that there would be a positive interaction between *subgenome* and $SL\_homeologous$, which would indicate that odds of sex-biased expression are higher for transcripts encoded by genes in the sex-linked portion of chr8L compared to the rest of genome, after controlling for the subgenome effects. It is important to consider subgenome effects because, in several *Xenopus* allotetraploids, subgenome S has more rearrangements, shorter coding regions, and more pseudogenization than subgenome L, and transcripts from the S subgenome tend to be more lowly expressed than homeologous transcripts from the L [32, 40]. For the diploid transcriptome of *X. tropicalis* there was no *subgenome* parameter, and the model simplifies to a logistic regression with one predictor variable (sex_linked) – that indicates whether or not a transcript was from a sex-linked (1) or non-sex-linked gene (0).

For juvenile tissues from *X. borealis* we were only able to make intraspecific comparisons, for example between the sex-linked portion of chr8L and the homeologous portion of chr8S (the latter of which is autosomal). For adult liver tissues, we had comparable data from all three focal species. In order to compare evolutionary trends across these species for this tissue and developmental stage, we performed three sets of analyses with sex-linked, and homeologous or orthologous to sex-linked regions across each species defined to match each one of our three focal species. For example, when adult liver of *X. borealis* was the focal species, $SL\_homeologous$ and sex_linked were set to a value of one for transcripts encoded by genes in the sex-linked, homeologous, and orthologous regions of chr8L, chr8S, and chr8 for models in each of the three focal species, and transcripts encoded by genes in the rest of each genome were set to zero. The same was done for models in each species for chr2L, chr2S, and chr2, or chr7L, chr7S, and chr7, respectively, when *X. laevis* or *X. tropicalis*, respectively, were the focal species. These tests allowed us to test whether sex-linked changes in one species arose concurrently with sex-linkage, or whether they were present ancestrally.

Analysis of individual transcripts allowed us to evaluate sex-biased expression of splice variants or possibly

individual alleles. We also analyzed differential expression at the gene level based on summed normalised expression levels of all transcripts from individual genes. For this gene-level analysis, expression levels were summed for transcripts that mapped within or overlapped by at least 200 bp with annotated genes in version 9.2 of the *X. laevis* (for *X. borealis* and *X. laevis*) and version 10 of the *X. tropicalis* genome (for *X. tropicalis*). We also included in this analysis the summed expression levels of transcripts of unannotated loci that overlapped by at least 200 bp in unannotated portions of these genomes.

3. Results

a. Population Variation in Sex Chromosome Differentiation in *X. borealis*

Partial mitochondrial DNA sequences from wild caught *X. borealis* cluster into four closely related, geographically structured clades. Some of these clades have only modest bootstrap support, and relationships among them are not well resolved (Fig. 2). At two localities in central Kenya (Njoro and Nairobi), two of these four clades were present; at all other localities only one clade was present in our sample. In Kenya, a geographic boundary between the most diverged intraspecific mitochondrial lineages of *X. borealis* may roughly correspond to the eastern arm of the Great Rift Valley. The two sequences from our lab strain clustered in different clades: the male (BJE3896) clustered in the black clade in Fig 2, whereas the female (BJE3897) clustered in the blue clade, with no divergent sites compared to four other sequences, all of which were from Nairobi specimens. To our knowledge, sequences from the western (red) clade and the Kitobo (purple) clade have not been previously reported. The black and blue clades (Fig. 2) have been previously reported and are estimated to have diverged from one another about 2 million years ago [33, 41].

Although samples from Kitobo, which has a unique mitochondrial lineage, were not included in the RRGS data, available information from other samples is consistent with the findings from mitochondrial DNA in terms of evidencing geographic structure, and also admixture between Nairobi and Njoro (Fig 3). With two partitions, a gradient of ancestry is evident, with specimens from the central Kenya localities of Nairobi and Njoro having a mixture of two ancestry components, specimens from Wundanyi (in the east) assigned to one ancestry component, and other specimens (all of whom are from west Kenya) assigned to another. With three partitions, the Wundanyi population is further distinguished from the others, and a gradient of ancestry components is still evident in the Nairobi and Njoro populations.

Using RRGS data, we calculated F_{ST} between females and males for the wild samples of *X. borealis*. For the autosomes and pseudoautosomal regions, this statistic is expected to be influenced by polymorphism and

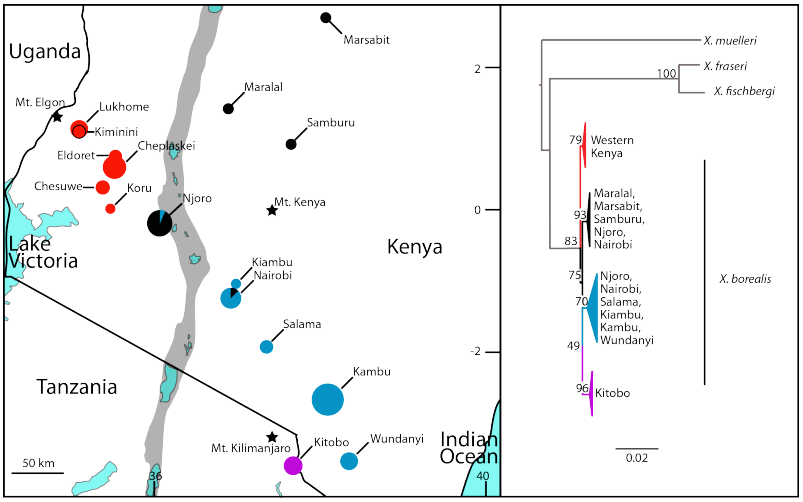


Figure 2: Sampling locations for *X. borealis* (left) and estimated relationships among four mitochondrial clades in *X. borealis* with respect to other species in the *muelleri* species group (right); bootstrap values are above branches and a scale bar indicates substitutions per site. The size of the circles corresponds to the number of mitochondrial sequences analyzed from each locality; pie charts from two localities indicate the proportions of each clade sampled with colors matching the phylogeny. Stars indicate locations of major mountains and gray shading indicates location of the eastern arm of the Great Rift Valley.

differences in the number of males and females in different populations (because an unequal sample of each sex in different subdivided populations should increase F_{ST} between males and females across the sample). For the sex chromosomes, F_{ST} should be additionally influenced by divergence between the W- and Z- sex chromosomes. We detected substantial differentiation in the region of recombination suppression on the sex chromosomes (chr8L, [27]) of the *X. borealis* specimens from central (Nairobi) and eastern (Wundanyi) Kenya, but not elsewhere on this chromosome (Fig. 4). In populations from western Kenya, there was no strong evidence of differentiation between the sexes in any substantially sized genomic region, including the sex-linked region of chr8L (Fig. 4). In both of these populations, or when all samples were considered, there were no other substantially sized regions where F_{ST} was substantially elevated beyond the 95% confidence interval for the non-sex-linked portion of the genome (data not shown). In the sex-linked region of *X. borealis* specimens from central and eastern Kenya, there was a dip in F_{ST} around 20 Mb (Fig. 4); we suspect this could be because RRGS data from *X. borealis* chr8S inappropriately mapped to this part of chr8L in the *X. laevis* reference genome.

b. The genomic landscape of genes with sex-biased expression in *X. borealis*

The transcriptome assemblies had 523,313, 562,550, 391,563, and 207,184 transcripts, respectively, for *X. borealis* tadpoles (a combined assembly for gonad/mesonephros from stages 46 and 48), *X. borealis* adult

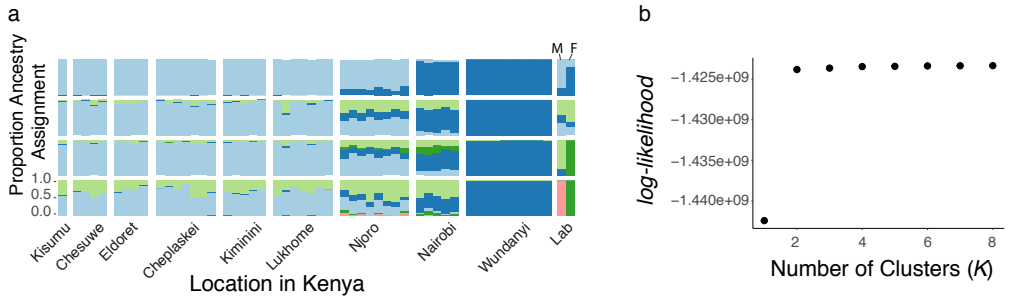


Figure 3: Genetic cluster analysis of RRGS data and WGS data. (a) Ancestry components of wild *X. borealis* from Kenya exhibit geographic structure, and genetic affinities of a male (M) and female (F) laboratory animal (Lab, right side) suggest origins from central Kenya (Njoro, Nairobi). (b) log-likelihood of values of K from 1–8.

liver, and independent assemblies for *X. laevis* adult liver and *X. tropicalis* adult liver. These transcripts include biological complexity stemming from splice variants and allelic variation, and technical variation due to incompletely assembled transcripts and assembly errors. We performed analyses of individual transcripts and also of groups of transcripts that were binned by genomic location into putative genes.

In the analysis of *X. borealis* transcripts from tadpole stages 46 and 48 gonad/mesonephros, and adult liver, a total of 453, 763, and 614 transcripts, respectively, were significantly differentially expressed (Table 1). In all three of the *X. borealis* transcriptomes that we examined (tadpole stages 46 and 48 gonad/mesonephros, and adult liver), there were significant and substantially higher densities of genes in the sex-linked region of the *X. borealis* sex chromosome that encoded sex-biased transcripts compared to (i) the homeologous region of chromosome 8S in *X. borealis* and (ii) the rest of the genome of *X. borealis* ($p \leq 0.001$ for all comparisons, binomial tests, Table 1, Fig. 5). More specifically, between 31–50% of all of the sex-biased transcripts that mapped to chromosomes in these three *X. borealis* transcriptomes were encoded by genes in the sex-linked region of chr8L, which comprises only ~2% of the genome. The interaction terms from the generalized linear models were significant for each *X. borealis* transcriptome ($p \ll 0.001$ for all three tests). The interaction coefficient of these models provides an estimate of the ratio of odds ratios of the probability a transcript is significantly differently expressed in the sex-linked region of chr8L compared to the rest of the L subgenome, divided by the probability that a transcript is significantly differently expressed in the region of chr8S that is homologous to the sex-linked region of chr8L compared to the rest of the S subgenome. Exponentiating the 95% confidence intervals for these interaction coefficients from each model indicates that the probability of significance in the sex-linked region of chr8L is 2.9–8.6, 3.9–8.7, and 2.4–5.6 times higher than the rest of the genome after controlling for subgenome effects (for the gonad/mesonephros tadpole stages 46 and 48, and adult liver, respectively). In the *X. borealis* transcriptomes, the effect of *subgenome* was not significant

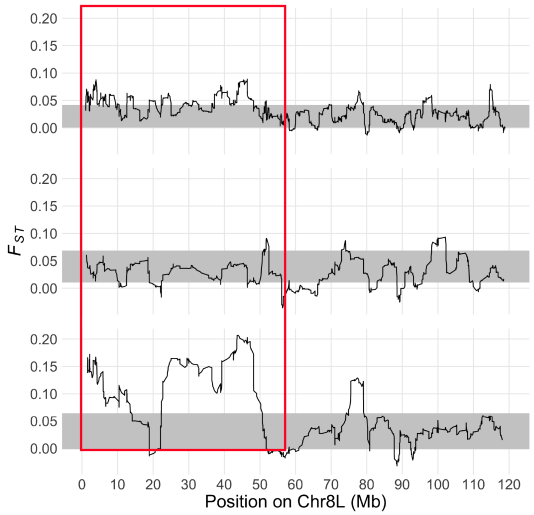


Figure 4: F_{ST} between males and females is substantial on the sex-linked portion (red box) of the sex chromosomes (Chr8L) for the samples from central and east Kenya (bottom; 9 females, 6 males) but not west Kenya (middle; 17 females, 19 males), and is slightly elevated when all samples are considered (top). Gray boxes indicate the 95% confidence interval for F_{ST} values in the non-sex-linked portion of the genome.

in any of these models (p = 0.90, 0.12, and 0.85, respectively), and the results were essentially the same when the *subgenome* predictor and interaction term were not included in each model (data not shown).

In the non-sex-linked portion of the *X. borealis* genome, the proportions of the sex-biased transcripts that mapped to chromosomes that were female-biased were similar in all three transcriptomes (53%, 59%, and 66%, respectively, for tadpole stages 46 and 48 gonad/mesonephros, and adult liver), as was the density of sex-biased genes (0.07, 0.10, and 0.12 sex-biased transcripts/Mb, respectively, assuming that the non-sex-linked portion of the genome is ~3.0Gb). However, in the sex-linked region, the proportions of sex-biased transcripts with female-biased expression were uniformly higher than the rest of the genome (89%, 87%, and 84% in tadpole stages 46 and 48 and adult liver, respectively) and the densities of sex-biased transcripts in the sex-linked region was 23 – 52 fold higher than the rest of the genome (2.89, 5.32, and 2.84 sex-biased transcripts per Mb, respectively, assuming the sex-linked region of the *X. borealis* sex chromosomes is ~57Mb [29]). This large disparity is qualitatively obvious when expression divergence of transcripts is visualized relative to genomic position (red asterisks denote significantly negative log(male/female) fold change – indicating female-biased expression – in Fig. 5). It is also apparent when densities are calculated relative to the number of annotated genes in each region instead of Mb (Supplemental Results). Although the proportions of male-biased transcripts was lower in the sex-linked region of *X. borealis* compared to other genomic regions (due to so many female-biased transcripts that are encoded by genes in this region), there

was still a 10–16 times higher density of male-biased transcripts relative to the non-sex-linked part of the genome. Thus, the sex-linked region is enriched in female-biased as well as male-biased transcripts, but more so for the female-biased transcripts.

For adult liver, additional comparisons to orthologous transcriptomes of *X. laevis* and *X. tropicalis* were possible. These comparisons illustrate that the proportion of genes with differentially expressed transcripts in the sex-linked portion of the *X. borealis* sex chromosomes is higher than (i) the orthologous region of chr8L in *X. laevis*, (ii) the homeologous regions chr8S in *X. laevis*, and (iii) the orthologous region of chr8 in *X. tropicalis*. In contrast to the findings from *X. borealis*, the interaction term from the generalized linear model was not significant for the *X. laevis* adult liver transcriptome when the sex-linked and homeologous regions were defined to match *X. borealis* ($p = 0.82$). Likewise, in *X. tropicalis*, a linear model also found no significant increase in the odds of differential expression for transcripts in the genomic region that is orthologous to the sex-linked region of *X. borealis* as compared to other genomic regions ($p = 0.80$). Previous work has demonstrated that the sex chromosomes of *X. borealis* evolved after divergence from the most recent common ancestor with *X. laevis* (Fig. 1, [27]). Thus, these results from *X. laevis* and *X. tropicalis* together with intraspecific comparisons within the *X. borealis* demonstrate that rapid regulatory evolution of the sex-linked region of *X. borealis* corresponded in time with this region becoming sex-linked (i.e., after divergence from the most recent common ancestor with *X. laevis*), as opposed to this region of chr8L encoding a high number of differentially expressed transcripts in an ancestor prior to the origin of the new sex chromosomes of *X. borealis*. These transcriptome results demonstrate that the *X. borealis* transcriptome evolved to have sex-specific expression due to rapid evolution in *cis*- of the regulatory regions of sex-linked genes of the new sex chromosomes of *X. borealis*. This effect was pronounced (Fig. 5) and significant, but affects only the small proportion of the transcriptome that is encoded by sex-linked genes.

On the homeologous chr8S of *X. borealis*, we observed a lower density of sex-biased genes than the sex chromosome (chr8L) for all three transcriptomes, although this was still higher than the rest of the genome (Supplemental Figs S2–S4). A probable explanation is that some of these sex-biased transcripts erroneously mapped to chr8S of the *X. laevis* reference genome, even though they actually are encoded by genes on chr8L in *X. borealis*. Additionally, in some of the *X. borealis* transcriptomes, various non-sex-linked genomic regions had a high density of female-biased transcripts (e.g., <15Mb on chr 3S in the gonad/mesonephros tissue of tadpole stage 48 transcriptome, and 130–155 Mb on chromosome 5L in the adult liver transcriptome, Supplemental Figs S3, S4). However, these patches were much smaller and had far lower densities of sex-biased transcripts than the sex-linked region of the new sex chromosomes of *X. borealis*.

Table 1: The number of significantly sex-biased transcripts by chromosome in *X. borealis* tadpole stages 46 and 48 gonad/mesonephros (St46, St48), and adult liver (Adult), including female-biased (FB), female-specific (FS), male-biased (MB), and male-specific (MS) transcripts, with biased transcripts being expressed in both sexes but significantly higher in one, and specific transcripts being expressed exclusively in one sex. For the sex chromosome (Chr8L), numbers in parentheses refer to the number of differentially expressed transcripts that are from genes within the sex-linked region. For the chromosome that is homeologous to the sex chromosome (Chr8S), numbers in parentheses indicate differentially expressed transcripts that are from genes within the portion of this chromosome that is homeologous to the sex-linked region of the sex chromosome. Some sex-biased transcripts mapped to scaffolds whose chromosomal locations are unknown (Scaffolds), or did not map to the *X. laevis* reference genome (Unmapped).

	St46				St48				Adult			
Chromosome	FB	FS	MB	MS	FB	FS	MB	MS	FB	FS	MB	MS
Chr1L	4	6	8	16	3	6	9	3	7	9	5	4
Chr1S	2	3	6	1	0	4	0	1	2	1	6	5
Chr2L	1	3	1	1	0	5	6	3	7	10	3	1
Chr2S	0	4	0	3	2	6	5	0	2	12	3	3
Chr3L	4	8	4	7	6	6	9	2	4	9	1	1
Chr3S	1	2	0	2	8	13	7	3	5	1	3	0
Chr4L	2	5	2	2	3	3	9	3	11	6	2	3
Chr4S	2	2	1	1	4	4	2	2	6	2	3	1
Chr5L	3	1	2	4	4	7	9	2	12	22	10	8
Chr5S	1	3	3	1	5	3	1	2	3	3	4	6
Chr6L	3	3	2	1	3	3	1	4	5	4	3	1
Chr6S	0	3	1	0	4	3	7	0	2	4	4	7
Chr7L	1	1	2	0	1	5	5	1	7	2	1	5
Chr7S	0	2	1	2	2	0	3	1	2	3	2	0
Chr8L	53(52)	99(95)	10(9)	12(9)	74(69)	206(195)	31(28)	14(11)	41(34)	123(102)	12(12)	16(14)
Chr8S	4(3)	18(13)	2(2)	1(1)	12(7)	29(25)	6(3)	0(0)	8(6)	28(19)	2(1)	3(2)
Chr9_10L	1	0	1	1	2	3	5	7	0	5	11	4
Chr9_10S	2	4	5	4	3	3	2	2	4	8	3	4
Scaffolds	12	34	4	6	0	48	14	8	12	21	6	15
Unmapped	11	19	2	4	15	40	13	13	6	22	1	1

c. The genomic landscapes of genes with sex-biased expression in *X. laevis* and *X. tropicalis*

Another question we were able to investigate is whether the sex-linked regions of *X. laevis* and *X. tropicalis* sex chromosomes also are enriched with transcripts with sex-biased expression. Similar to the findings in *X. borealis*, each of the independently evolved sex-linked regions of *X. laevis* and *X. tropicalis* also were significantly enriched with sex-biased transcripts compared to other homeologous and/or orthologous genomic regions. In *X. laevis*, the density of sex-biased transcripts in the sex-linked region was 27 sex-biased transcripts/Mb), which is over 40 times the density elsewhere (0.57 sex-biased transcripts/Mb; Table 2, Fig. S7). When sex-linked and homeologous regions were defined to match the sex chromosomes of *X. laevis* (i.e., portions of chr2L and 2S respectively), the interaction term of the linear model for *X. laevis* was significant ($p \ll 0.001$) and the 95% confidence interval of the odds that a transcript encoded by genes in the sex-linked region of *X. laevis* would have sex-biased expression was 1.9–5.7 times higher than the odds for transcripts encoded by genes in the rest of the genome after controlling for subgenome effects. When the linear models were run on *X. borealis* and *X. tropicalis* transcriptomes but defining orthologous and homeologous to the sex-linked region to match *X. laevis*, there was no significant correlation (for *X. borealis* for the interaction

term ($p = 0.953$) or for *X. tropicalis* for the *sex_linked* predictor variable ($p = 0.951$)). These contrasts are consistent with the proposal that sex-biased expression of the portion of the *X. laevis* that is encoded by the sex-linked portion of chr2L arose after divergence from the most recent common ancestor with *X. tropicalis*, and coincided with this region becoming sex-linked in *X. laevis* and then were lost in *X. borealis* after that species acquired a new sex chromosome (on chr8L).

There was also strong evidence for male-biased expression of transcripts encoded by the independently evolved sex-linked portion of chr7 in the genome of *X. tropicalis*. In this region, the density of sex-biased transcripts in the sex-linked portion of chr7 was 1.9 sex-biased transcripts/Mb, which is ten times the density of transcripts that map to other portions of chromosomes (0.18 sex-biased transcripts/Mb; Table 3, Fig. S7). This enrichment is reflected by a significant correlation in a linear model ($p \ll 0.001$) that indicates that the odds of sex-biased expression are 3.5–8.5 higher in the sex linked region of chr7 than in other genomic regions. Although comparison to *X. laevis* and *X. borealis* is complicated by genome duplication, we ran the linear models anyway, but defining the *SL_homeologous* predictor variable to match the sex-linked portions of *X. tropicalis* chr7. In these analyses, there was no significant correlation the *SL_homeologous* predictor variable or the interaction with *subgenome* for either species ($p > 0.05$ for both terms for both species). These contrasts are consistent with the proposal that the sex-biased expression of the sex-linked portion of the *X. tropicalis* transcriptome evolved in association with sex-linkage in this species as opposed to being an ancestral condition, and that this enrichment was either lost when new sex chromosomes emerged in *X. laevis* and *X. borealis*, or that it was never present if the most recent common ancestor of these three species had a sex chromosome elsewhere in the genome.

In contrast to *X. borealis*, the sex-linked region of *X. tropicalis* has a higher proportion of male-biased transcripts (n = 20 out of 21 sex-biased transcripts; 95%) than the rest of the genome, where 38% of sex-biased transcripts are male-biased (Table 3). In *X. laevis* there is a more similar proportion of female- and male-biased transcripts in the sex-linked region (50% female-biased) compared to the rest of the genome (60% female-biased; Table 2), than in *X. borealis*.

One unexpected finding was that the *X. laevis* adult liver transcriptome was more strongly sex-biased compared to *X. borealis* and *X. tropicalis* (Supplemental Figs S5 compared to Supplemental Figs S4 and S6). In the *X. laevis* adult liver transcriptome, we detected 4.7 times as many sex-biased expressed transcripts in the non-sex-linked portion of the genome (n = 1718) compared to *X. borealis* (n = 368), although this disparity is magnified slightly by the larger size of the non-sex-linked region of the *X. laevis genome*. Interestingly, even after excluding the sex-linked region, there were significantly more sex-biased expressed

Table 2: The number of differentially expressed transcripts by chromosome in *X. laevis* adult liver, with labeling following Table 1. For the sex chromosome (Chr2L) and the ortholog to the *X. borealis* sex chromosome (Chr8L), numbers in parentheses refer to the number of differentially expressed transcripts that are from genes within the genomic region that is sex-linked or that is orthologous to the *X. borealis* sex-linked region, respectively. For the chromosome that is homeologous to the sex chromosome (Chr8S), numbers in parentheses indicate differentially expressed transcripts that are from genes within the portion of this chromosome that is homeologous to the sex-linked region of the sex chromosome.

Chromosome	FB	FS	MB	MS
Chr1L	45	97	28	33
Chr1S	18	23	16	24
Chr2L	21(18)	28(22)	8(7)	42(33)
Chr2S	11	55	4	30
Chr3L	34	63	29	87
Chr3S	11	17	5	12
Chr4L	63	76	13	47
Chr4S	20	27	15	58
Chr5L	13	12	4	15
Chr5S	2	23	5	5
Chr6L	23	27	7	13
Chr6S	3	8	6	11
Chr7L	44	121	18	41
Chr7S	7	18	6	21
Chr8L	44(19)	55(12)	15(6)	45(16)
Chr8S	8(4)	10(5)	3(2)	10(3)
Chr9_10L	20	13	11	6
Chr9_10S	13	20	4	8
Scaffolds	25	35	23	35
Unmapped	1	3	0	4

transcripts encoded by genes from the L (n=1181) than the S subgenome in *X. laevis* (n = 537, $P \ll 0.0001$, binomial test). The *subgenome* predictor variable was significant in the linear model for *X. laevis* when this species was the focal species (i.e., with *SL_homeologous* set to 1 for the ends of chr2L and chr2S, see Methods; $p \ll 0.001$). Exponentiating the correlation coefficient indicates that the 95% confidence interval for the odds of sex-biased expression of transcripts encoded by genes in the L subgenome is 1.63–2.03 higher than for those encoded by genes in the S subgenome. Thus, not only are genes more highly expressed in the L subgenome [32], but it appears that they are more likely to encode transcripts with sex-biased expression, even after controlling for the effects of sex-linkage on chr2L. One important caveat is that *subgenome* and $significance$ are conflated because differential expression is statistically easier to detect for highly expressed transcripts (on the L), which violates assumptions of the generalized linear model. Other caveats of this and other transcriptome comparisons are discussed in the Supplement.

Another unexpected observation relates to the gonad/mesonephros transcriptomes from tadpole stage 48 and adult liver, where the high density of sex-biased transcripts extends slightly beyond the sex-linked region identified by [29]. This was not the case in the corresponding transcriptome from tadpole stage 46 (Fig. 5). At this time we do not have an explanation for this observation.

Table 3: The number of differentially expressed transcripts by chromosome in *X. tropicalis* adult liver, with labeling following Table 1. For the sex chromosome (Chr7) and the ortholog to the *X. borealis* sex chromosome (Chr8), numbers in parentheses refer to the number of differentially expressed transcripts that are from genes within the genomic region that is sex-linked or that is orthologous to the *X. borealis* sex-linked region, respectively.

Chromosome	FB	FS	MB	MS
Chr1	13	30	5	7
Chr2	3	4	7	1
Chr3	6	10	1	12
Chr4	53	35	13	2
Chr5	6	2	6	4
Chr6	5	8	3	12
Chr7	2(1)	1(0)	7(3)	24(17)
Chr8	16(14)	12(10)	8(6)	7(5)
Chr9	4	1	3	2
Chr10	4	4	4	1
Scaffolds	0	0	0	0
Unmapped	1	0	0	0

d. Sex-biased expression is less pronounced in gene-level than transcript-level analyses

When expression levels of transcripts that mapped to the same genomic region were summed into putative genes, analysis of the resulting genic expression recovered a less substantial or non-significant enrichment of sex-biased expression in the sex-linked region of these three frog species. With *X. borealis* as the focal species (i.e., when the *SL_homeologous* parameter defined to match the sex-linked and homeologous regions of *X. borealis*), the interaction term of the linear models was significant for *X. borealis* gonad/mesonephros tadpole stage 48 ($p = 0.01$) and adult liver ($p = 0.003$), but not gonad/mesonephros tadpole stage 46 ($p = 0.16$). Thus, after controlling for subgenome effects, the odds of sex-biased expression at the gene level were 1.6–20.4 and 1.5–8.1 times higher in the sex-linked region as compared to elsewhere in *X. borealis* gonad/mesonephros tadpole stage 48 and adult liver, respectively. When summed transcripts from *X. laevis* and *X. tropicalis* were analyzed with each of these species as the focal species, the interaction term for the *X. laevis* data was not significant, and the correlation between sex-linkage and sex-bias in *X. tropicalis* also was not significant ($p > 0.05$ for all coefficients). These results are consistent with plots of gene-level expression that illustrate that the summed (gene-level) sex-biased expression in the sex-linked regions is less distinguished from the rest of the respective genomes of these three focal species (Supplemental Figs. S8 – S12) as compared to the analysis of individual transcripts (Supplemental Figs. S2–S4). Thus it appears that much of the regulatory evolution that occurred on the sex-linked region influenced expression at the level of splice variants and alleles, as opposed to concurrently governing expression of all transcripts encoded by sex-linked genes.

4. Discussion

a. Among population variation of the sex chromosomes of *X. borealis*

To help disentangle concurrent and confounding drivers of sex chromosome evolution, we scrutinized the independently evolved sex chromosomes of three closely-related frog species: the Marsabit clawed frog *X. borealis*, the African clawed frog *X. laevis*, and the Western clawed frog *X. tropicalis*. Using field collected animals and genomic and transcriptomic approaches, we demonstrate that (i) *X. borealis* has modest but significant population structure that is relatively recently evolved, (ii) there exists variation among *X. borealis* populations in the extent of sex chromosome differentiation that corresponds to a previously identified region of recombination suppression [27], (iii) the sex-linked portion of the *X. borealis* sex chromosomes (chr8L) carry genes that encode a high number of transcripts with sex-biased expression, and especially female-biased expression, and (iv) sex-biased regulatory evolution evolved in *X. borealis* independently after divergence from *X. laevis*, and in association with the origin of sex-linkage on the new sex chromosomes of this species. In the independently evolved sex-linked regions of *X. laevis* (chr2L) and *X. tropicalis* (chr7), we also identified a high density of sex-biased transcripts encoded by sex-linked transcripts, and we showed that sex-biased regulation evolved independently in association with sex-linkage in each of these species as well. These findings are thus concordant across these three frog species, each with independently evolved sex chromosomes, and each with distinctive sex-linked genomic regions. These results point to independent evolution of *cis*- regulation in the sex-linked regions of each species that occurred contemporaneously with the onset of sex-linkage, as opposed to having been present ancestrally in an autosomal precursor of each sex chromosome. When analysis was performed at the gene level on the transcriptome data from all three species, the signal of sex-biased expression by sex-linked genes was less substantial, which suggests that this effect is largely driven by allele-specific and splice-variant-specific expression. Technical artifacts undoubtedly also affect our quantification of sex-biased expression, but we have no reason to suspect that this would result in an enrichment of sex-biased expression of transcripts encoded by genes in three different sex-linked regions of three species.

In the laboratory strain of *X. borealis* previously studied by Furman and Evans (2016, 2018), sex chromosome differentiation was detected that coincided with recombination suppression [29]. Our study suggests that this strain originates from central Kenya (Figs. 2, 3), shows that the extent of sex chromosome differentiation varies among populations of *X. borealis* (Fig. 4), and links recombination suppression and sex chromosome differentiation to sex-biased expression of transcripts encoded by sex-linked genes in *X. borealis*

from central and eastern Kenya. A lack of substantial differentiation between male and female *X. borealis* from western Kenya (Fig. 4) suggests that the sex chromosomes in these specimens are mostly pseudoautosomal regions that (by definition) recombine in both sexes. We were unable to identify sex-linked variation on chromosome 8L in the western population of *X. borealis*, so it is conceivable that this population has a system for sex determination that is distinctive from *X. borealis* from central and eastern Kenya in terms of where in the genome the trigger for sex determination resides, and/or whether males or females are the heterogametic sex.

At this time, we can only speculate about the mechanisms behind recombination suppression in the sex-linked portions of the sex chromosomes of *X. borealis* from central and eastern Kenya (Fig 4). One possibility is that specimens from central and eastern Kenya (including our laboratory strain) have an inversion on the W-chromosome but not on the Z-chromosome, and that suppresses recombination between this portion of these sex chromosomes. This inversion could be absent on both sex chromosomes of *X. borealis* from western Kenya – either because it was lost or because it was never present. If recombination suppression in the central and eastern populations is newly evolved compared to sex chromosomes in the western populations, this would suggest that sex-biased expression of sex-linked transcripts evolved quickly because divergence between these populations is relatively recent. Alternatively and more plausibly, *X. borealis* from west Kenya may have recently lost recombination suppression on chr8L, for example, if a new sex determining locus evolved on another chromosome in an ancestor of specimens from that part of Kenya. This scenario would be consistent with a more protracted period for sex-biased expression of sex-linked transcripts to arise in the central and eastern population, but it would still be more recently than the age of the most recent common ancestor of *X. laevis* and *X. borealis* (~40 Ma [33]) because the sex chromosomes of *X. borealis* arose after this ancestor diversified [27]. Further study of sex-linkage in *X. borealis* from west Kenya, and in other closely related species in the *muelleri* species group (Fig. 2) may help distinguish between these alternatives.

b. Effects of sex chromosome evolution on the transcriptome

In three frog species, we observed a higher density of sex-biased transcripts encoded by genes on the sex-linked portion of the sex chromosomes as compared to those encoded by genes located in the rest of the genome. In *X. borealis*, which has female heterogamy, this higher density of sex-biased transcripts was significantly skewed towards female-biased transcripts compared to the rest of the genome (Table 1; Fig. 5), but in *X. tropicalis*, a higher proportion of transcripts with male-biased expression are encoded in the sex-linked region. In two *X. tropicalis* families from Ghana – the fathers of which were each demonstrated to carry a

Y-chromosome – the sex-linked region of the sex chromosomes also had a significant enrichment of sex-linked transcripts with male-biased expression in tadpole gonads/mesonephros during sexual differentiation [30]. In *X. laevis*, there are heterogametic females and a high density of sex-biased transcripts in the sex-linked region, but no substantial skew towards these transcripts being more female- or male-biased as compared to sex-biased transcripts encoded by genes elsewhere in the genome.

One factor that could contribute to the strong skew of sex-biased expression in *X. tropicalis*, and *X. borealis* is degeneration of the sex-specific sex chromosome (the W- or Y-). In *X. tropicalis*, there is polymorphism in sex chromosomes, with W-, Z-, and Y-chromosomes being present in different laboratory strains [24], and the Y-chromosome evolved from an ancestral Z-chromosomes [30]. Patterns of nucleotide polymorphism in *X. tropicalis* are consistent with degeneration of the W-chromosome as a primary mechanism driving this male-biased expression of sex-linked genes in tadpole gonad/mesonephros [30]. Expression results from adult liver presented here are consistent with these observations in *X. tropicalis* tadpoles.

In *X. borealis*, however, this explanation does not immediately explain the pattern of sex-biased expression that we observed because a strong skew towards female-biased expression is not predicted from W-chromosome degeneration. One hypothetical scenario that could explain these observations is that this biased expression is a consequence of degeneration of an ancestral Y-chromosome that eventually evolved into a Z-chromosome, thereby causing an ancestral X-chromosome to become a W-chromosome (Table 1C in Bull and Charnov, 1977). A transition from an XY to a ZW system is theoretically disfavoured by natural selection if it involves fixation of a degenerate Y-chromosome [21], but is still possible if the deleterious load of the ancestral Y-chromosome is modest. Amphibians are not known to have dosage compensation [43] and tolerance of dosage differences between the sexes is evidenced here by the sex-biased expression of transcripts encoded by genes in the sex-linked region (Fig 5). The observation of population variation in sex chromosomes of *X. borealis* (Fig 4) opens the possibility that a sex chromosome turnover may have taken place in *X. borealis* from western Kenya. An interesting question to address in the future asks what the sex determination system of the western population is, where it is located in the genome, and whether this population still has the Z-chromosome that is present in the central/eastern population and that potentially could be derived from a degenerate Y-chromosome.

Another factor that potentially could contribute to higher expression in the heterogametic sex is adaptive evolution on the sex-specific sex chromosome, such as the W-chromosome of *X. borealis*. While we do not have evidence for or against this possibility, adaptive evolution is generally expected to be faster on the X- or Z-chromosome than the Y-chromosome or W-chromosome [44, 45], and we therefore view degeneration

of the sex-specific sex chromosome as more likely to have a larger contribution to the high enrichment and skewed sex-biased expression of sex-linked transcripts in *Xenopus* than adaptive evolution of sex-specific sex chromosomes. It would be fascinating for future efforts to evaluate the significance of sex-biased expression patterns by comparing populations with distinctive sex chromosomes (e.g., the western versus eastern populations of *X. borealis*, including zones of intergradation if they exist.

5. Concluding remarks

In the Marsabit clawed frog, *X. borealis*, there exists population variation in the extent of differentiation between the W- and Z-chromosomes on a large sex-linked region of these sex chromosomes. This sex-linked region has a far higher higher density of transcripts with sex-biased expression – especially female-biased expression – as compared to orthologous and homeologous genomic regions, and to other recombining pseudoautosomal and autosomal genomic regions. This demonstrates that sex-biased expression of sex-linked genes occurred in concert with the origin of the new sex chromosomes of *X. borealis*. Similar findings were also recovered from transcriptomes of *X. laevis* and *X. tropicalis*, each of which have independently evolved sex chromosomes. Additionally, we found that transcripts encoded by sex-linked genes in *X. borealis* and *X. tropicalis* have a strong sex-bias (female- or male-biased, respectively). A plausible explanation for the intense, sex-biased of transcripts that emerge from the sex-linked region is that it is due to degeneration of a sex-specific sex chromosome (the W- in *X. tropicalis* and an ancestral Y- in *X. borealis*). For *X. borealis* this scenario would involve a recent homologous sex chromosome turnover, wherein an ancestral sex determination system on chromosome 8L with male (XY) heterogamy evolved into the current system on the same chromosome with female (ZW) heterogamy, with the ancestral X-chromosome becoming the contemporary W-chromosome and the ancestral Y-chromosome becoming the contemporary Z-chromosome. This hypothesis that can be tested with comparative data from other populations and species. Still unknown are the functional consequences of extensive sex-biased expression of transcripts that are tethered to various sex determining loci in different *Xenopus* species. This raises the question of whether and how this variation might clash or synergize in individuals in nature.

Ethics Statement

This work was approved by the Animal Care Committee at McMaster University (AUP# 17-12-43).

Data Accessibility

The RRGS and RNAseq data from this study have been deposited in the Short Read Archive of NCBI (accession numbers XXX, SRR11844024–SRR11844044, and SRR11836336–SRR11836343). All new Sanger sequences were submitted to GenBank (accession numbers XXX – XXX).

Authors' contributions

BLSF, VW, and BJE performed the fieldwork. X-YS, BLSF, TP, MK, CMSC, ID, and BJE collected and analyzed data. X-YS and BJE wrote the paper, and all authors edited the paper.

Funding

This work was supported by the Natural Science and Engineering Research Council of Canada (RGPIN-2017-05770, BJE), Resource Allocation Competition awards from Compute Canada (BJE), the Museum of Comparative Zoology at Harvard University (BJE), and the National Research Foundation (NRF) of South Africa (NRF Grant No. 87759, JM).

Acknowledgements

We thank Brian Golding for access to computational resources, Jonathan Dushoff and Ben Bolker for statistical advice, and our funding sources for supporting this research.

References

[1] B. Charlesworth. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* 10(3) (2009), 195–205.

[2] B. Vicoso and B. Charlesworth. Evolution on the X chromosome: Unusual patterns and processes. *Nature Reviews Genetics* 7(8) (2006), 645–653.

[3] B. Charlesworth and D. Charlesworth. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 355(1403) (2000), 1563–1572.

[4] B. Charlesworth, J. A. Coyne, and N. H. Barton. The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist* 130(1) (1987), 113–146.

[5] W. R. Rice. *The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes.* Vol. 41. 4. 1987.

[6] R. Spigler, K. Lewers, D. Main, and T. Ashman. Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity* 101(6) (2008), 507–517.

[7] B. Charlesworth. The evolution of sex chromosomes. *Science* 251(4997) (1991), 1030–1033.

[8] B. Charlesworth and J. Wall. Inbreeding, heterozygote advantage and the evolution of neo–X and neo–Y sex chromosomes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266(1414) (1999), 51–56.

[9] B. T. Lahn and D. C. Page. Four evolutionary strata on the human X chromosome. *Science* 286(5441) (1999), 964–967.

[10] D. Charlesworth, B. Charlesworth, and G. Marais. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95(2) (Aug. 2005), 118–128. ISSN: 0018-067X.

[11] Y. Sun, J. Svedberg, M. Hiltunen, P. Corcoran, and H. Johannesson. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nature Communications* 8(1) (2017), 1140.

[12] L. Brooks. The evolution of recombination rates. *The evolution of sex* (1988), 87–105.

[13] W. G. Hill and A. Robertson. The effect of linkage on limits to artificial selection. *Genetics Research* 8(3) (1966), 269–294.

[14] H. Skaletsky, T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L. G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423(6942) (2003), 825–837.

[15] B. Vicoso and B. Charlesworth. Effective population size and the faster-X effect: An extended model. *Evolution: International Journal of Organic Evolution* 63(9) (2009), 2413–2426.

[16] J. P. Masly and D. C. Presgraves. High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. *PLoS Biology* 5(9) (2007).

[17] M. Wallis, P. Waters, M. Delbridge, P. J. Kirby, A. J. Pask, F. Grützner, W. Rens, M. A. Ferguson-Smith, and J. A. Graves. Sex determination in platypus and echidna: autosomal location of SOX3 confirms the absence of SRY from monotremes. *Chromosome Research* 15(8) (2007), 949.

[18] B. Vicoso and D. Bachtrog. Numerous transitions of sex chromosomes in Diptera. *PLoS Biology* 13(4) (2015).

[19] I. Darolti, A. E. Wright, B. A. Sandkam, J. Morris, N. I. Bloch, M. Farré, R. C. Fuller, G. R. Bourne, D. M. Larkin, F. Breden, et al. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proceedings of the National Academy of Sciences* 116(38) (2019), 19031–19036.

[20] S. Adolfsson and H. Ellegren. Lack of dosage compensation accompanies the arrested stage of sex chromosome evolution in ostriches. *Molecular Biology and Evolution* 30(4) (2013), 806–810.

[21] P. A. Saunders, S. Neuenschwander, and N. Perrin. Impact of deleterious mutations, sexually antagonistic selection, and mode of recombination suppression on transitions between male and female heterogamety. *Heredity* 123(3) (2019), 419–428.

[22] L. Gu and J. R. Walters. Evolution of sex chromosome dosage compensation in animals: a beautiful theory, undermined by facts and bedeviled by details. *Genome biology and evolution* 9(9) (2017), 2461–2476.

[23] D. Bachtrog, J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T.-L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose, R. Ming, et al. Sex determination: Why so many ways of doing it? *PLoS Biology* 12(7) (2014), e1001899.

[24] Á. S. Roco, A. W. Olmstead, S. J. Degitz, T. Amano, L. B. Zimmerman, and M. Bullejos. Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis*. *Proceedings of the National Academy of Sciences* 112(34) (2015), E4752–E4761.

[25] S. Yoshimoto, E. Okada, H. Umemoto, K. Tamura, Y. Uno, C. Nishida-Umehara, Y. Matsuda, N. Takamatsu, T. Shiba, and M. Ito. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proceedings of the National Academy of Sciences* 105(7) (2008), 2469–2474.

[26] A. J. Bewick, D. W. Anderson, and B. J. Evans. Evolution of the closely related, sex-related genes DM-W and DMRT1 in African clawed frogs (*Xenopus*). *Evolution: International Journal of Organic Evolution* 65(3) (2011), 698–712.

[27] B. L. S. Furman and B. J. Evans. Sequential turnovers of sex chromosomes in African clawed frogs (*Xenopus*) suggest some genomic regions are good at sex determination. *Genes|Genomes|Genetics* (2016).

[28] J. Tymowska. Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. In: *Amphibian Cytogenetics and Evolution*. Ed. by D. M. Green and S. K. Sessions. Academic Press San Diego, 1991, 259–297.

[29] B. L. S. Furman and B. J. Evans. Divergent evolutionary trajectories of two young, homomorphic, and closely related sex chromosome systems. *Genome Biology and Evolution* 10(3) (2018), 742–755.

[30] B. L. S. Furman, C. M. S. Cauret, M. Knytl, X.-Y. Song, T. Premachandra, C. Ofori-Boateng, J. D. C., M. E. Horb, and B. J. Evans. A frog with three sex chromosomes that co-mingle together in nature: *Xenopus tropicalis* has a degenerate W- and a Y- that evolved from a Z-. *Plos Genetics* Submitted (2020).

[31] B. J. Evans. Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Front Biosci* 13 (2008), 4687–4706.

[32] A. M. Session, Y. Uno, T. Kwon, J. A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, et al. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* 538(7625) (2016), 336–343.

[33] B. J. Evans, M.-T. Gansauge, E. L. Stanley, B. L. Furman, C. M. Cauret, C. Ofori-Boateng, V. Gvoždık, J. W. Streicher, E. Greenbaum, R. C. Tinsley, et al. *Xenopus fraseri*: Mr. Fraser, where did your frog come from? *PloS One* 14(9) (2019).

[34] N. A. Baird, P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3(10) (2008), e3376.

[35] K. Karimi, J. D. Fortriede, V. S. Lotay, K. A. Burns, D. Z. Wang, M. E. Fisher, T. J. Pells, C. James-Zorn, Y. Wang, V. G. Ponferrada, et al. Xenbase: A genomic, epigenomic and transcriptomic model organism database. *Nucleic Acids Research* 46(D1) (2018), D861–D868.

[36] T. D. Wu and C. K. Watanabe. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21(9) (May 2005), 1859–1875. ISSN: 1367-4803.

[37] J. Krumsiek, R. Arnold, and T. Rattei. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23(8) (2007), 1026–1028.

[38] S. Andrews. *re-DOT-able: version 1.1*. `https://github.com/s-andrews/redotable`. 2019.

[39] T. Mitros, J. Lyons, A. Session, J. Jenkins, S. Shu, T. Kwon, M. Lane, C. Ng, T. Grammer, M. Khokha, et al. A chromosome-scale genome assembly and dense genetic map for *Xenopus tropicalis*. *Developmental Biology* (2019).

[40] B. L. Furman, U. J. Dang, B. J. Evans, and G. B. Golding. Divergent subgenome evolution after allopolyploidization in African clawed frogs (*Xenopus*). *Journal of Evolutionary Biology* 31(12) (2018), 1945–1958.

[41] B. J. Evans, D. B. Kelley, R. C. Tinsley, D. J. Melnick, and D. C. Cannatella. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Molecular phylogenetics and evolution* 33(1) (2004), 197–213.

[42] J. J. Bull and E. L. Charnov. Changes in the heterogametic mechanism of sex determination. *Heredity* 39(1) (1977), 1.

[43] J. W. Malcom, R. S. Kudra, and J. H. Malone. The sex chromosomes of frogs: Variability and tolerance offer clues to genome evolution and function. *Journal of Genomics* 2 (2014), 68.

[44] W. R. Rice. Evolution of the Y sex chromosome in animals. *Bioscience* 46(5) (1996), 331–343.

[45] H. A. Orr and Y. Kim. An adaptive hypothesis for the evolution of the Y chromosome. *Genetics* 150(4) (1998), 1693–1698.

Bibliography

Abrusán, G., Grundmann, N., DeMester, L., and Makalowski, W. (2009). TEclass-a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 25(10), 1329–1330.

Adolfsson, S. and Ellegren, H. (2013). Lack of dosage compensation accompanies the arrested stage of sex chromosome evolution in ostriches. *Molecular Biology and Evolution* 30(4), 806–810.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology* 215(3), 403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25(17), 3389–3402.

Angelopoulou, R., Lavranos, G., and Manolakou, P. (2012). Sex determination strategies in 2012: towards a common regulatory model? *Reproductive Biology and Endocrinology* 10(1), 13.

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Angel, G. del, Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics*, 11–10.

Babošová, M., Vašeková, P., Porhajašová, J., and Noskovič, J. (2018). Influence of temperature on reproduction and length of metamorphosis in Xenopus laevis (Amphibia: Anura). *The European Zoological Journal* 85(1), 150–157.

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J. C., and Sex Consortium, T. T. of (2014). Sex Determination: Why So Many Ways of Doing It? *PLOS Biology* 12(7), 1–13.

Bazzicalupo, A. L., Carpentier, F., Otto, S. P., and Giraud, T. (2019). Little evidence of antagonistic selection in the evolutionary strata of fungal mating-type chromosomes (Microbotryum lychnidis-dioicae). *G3: Genes, Genomes, Genetics* 9(6), 1987–1998.

Bewick, A. J., Anderson, D. W., and Evans, B. J. (2011). Evolution of the closely related, sex-linked genes *DM-W* and *DMRT1* in African clawed frogs (*Xenopus*). *Evolution* 65(3), 698–712. ISSN: 1558-5646.

Bewick, A. J., Chain, F. J., Zimmerman, L. B., Sesay, A., Gilchrist, M. J., Owens, N. D., Seifertova, E., Krylov, V., Macha, J., Tlapakova, T., et al. (2013). A large pseudoautosomal region on the sex chromosomes of the frog *Silurana tropicalis*. *Genome Biology and Evolution* 5(6), 1087–1098.

Blaser, O., Grossen, C., Neuenschwander, S., and Perrin, N. (2013). Sex-chromosome turnovers induced by deleterious mutation load. *Evolution: International Journal of Organic Evolution* 67(3), 635–645.

Blaser, O., Neuenschwander, S., and Perrin, N. (2014). Sex-chromosome turnovers: the hot-potato model. *The American Naturalist* 183(1), 140–146.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

Boskovski, M. T., Yuan, S., Pedersen, N. B., Goth, C. K., Makova, S., Clausen, H., Brueckner, M., and Khokha, M. K. (2013). The heterotaxy gene GALNT11 glycosylates Notch to orchestrate cilia type and laterality. *Nature* 504(7480), 456–459.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., and Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10(4), e1003537.

Brelsford, A., Stöck, M., Betto-Colliard, C., Dubey, S., Dufresnes, C., Jourdan-Pineau, H., Rodrigues, N., Savary, R., Sermier, R., and Perrin, N. (2013). Homologous sex chromosomes in three deeply divergent anuran species. *Evolution* 67(8), 2434–2440.

Bridgham, J. T., Brown, J. E., Rodríguez-Marí, A., Catchen, J. M., and Thornton, J. W. (2008). Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. *PLoS Genet* 4(9), e1000191.

Bull, J. J. and Charnov, E. L. (1977). Changes in the heterogametic mechanism of sex determination. *Heredity* 39(1), 1.

Bull, J. and Vogt, R. C. (1979). Temperature-dependent sex determination in turtles. *Science* 206(4423), 1186–1188.

Cai, J., Zhao, R., Jiang, H., and Wang, W. (2008). De novo origination of a new protein-coding gene in Saccharomyces cerevisiae. *Genetics* 179(1), 487–496.

Cannatella, D. (2015). Xenopus in space and time: fossils, node calibrations, tip-dating, and paleobiogeography. *Cytogenetic and genome research* 145(3-4), 283–301.

Cannatella, D. C. and De Sa, R. O. (1993). Xenopus laevis as a model organism. *Systematic Biology* 42(4), 476–507.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics* 1(3), 171–182.

Cauret, C. M., Gansauge, M.-T., Tupper, A. S., Furman, B. L., Knytl, M., Song, X.-Y., Greenbaum, E., Meyer, M., and Evans, B. J. (2020). Developmental systems drift and the drivers of sex chromosome evolution. *Molecular biology and evolution* 37(3), 799–810.

Chalopin, D., Naville, M., Plard, F., Galiana, D., and Volff, J.-N. (2015). Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biology and Evolution* 7(2), 567–580.

Charlesworth, B. and Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 355(1403), 1563–1572.

Charlesworth, D. and Charlesworth, B. (1980). Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genetics Research* 35(2), 205–214.

Charlesworth, D. (2016). The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evolutionary Applications* 9(1), 74–90.

Chen, S., Zhang, Y. E., and Long, M. (2010). New genes in Drosophila quickly become essential. *science* 330(6011), 1682–1685.

Cline, T. W., Dorsett, M., Sun, S., Harrison, M. M., Dines, J., Sefton, L., and Megna, L. (2010). Evolution of the *Drosophila* feminizing switch gene Sex-lethal. *Genetics* 186(4), 1321–1336.

Cohn, M. J. and Tickle, C. (1999). Developmental basis of limblessness and axial patterning in snakes. *Nature* 399(6735), 474.

Consortium, G. O. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* 32(suppl1), D258–D261.

Denholm, I., Franco, M., Rubini, P., and Vecchi, M. (1986). Geographical variation in house-fly (*Musca domestica L.*) sex determinants within the British Isles. *Genetics Research* 47(1), 19–27.

Dennis, M. Y., Nuttle, X., Sudmant, P. H., Antonacci, F., Graves, T. A., Nefedov, M., Rosenfeld, J. A., Sajjadian, S., Malig, M., Kotkiewicz, H., et al. (2012). Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell* 149(4), 912–922.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* 43(5), 491–498.

Doorn, G. S. van and Kirkpatrick, M. (2010). Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics*.

Dorken, M. E., Friedman, J., and Barrett, S. C. (2002). The evolution and maintenance of monoecy and dioecy in Sagittaria latifolia (Alismataceae). *Evolution* 56(1), 31–41.

Dufresnes, C., Borzé, A., Horn, A., Stöck, M., Ostini, M., Sermier, R., Wassef, J., Litvinchuck, S. N., Kosch, T. A., Waldman, B., Jang, Y., Brelsford, A., and Perrin, N. (2015). Sex−Chromosome Homomorphy in Palearctic Tree Frogs Results from Both Turnovers and X−Y Recombination. *Molecular Biology and Evolution* 32(9), 2328–2337.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5), e19379.

Espinasa, L., Robinson, J., and Espinasa, M. (2018). Mc1r gene in Astroblepus pholeter and Astyanax mexicanus: Convergent regressive evolution of pigmentation across cavefish species. *Developmental biology* 441(2), 305–310.

Evans, B. J. (2008). Genome evolution and speciation genetics of clawed frogs (Xenopus and Silurana). *Front Biosci* 13, 4687–4706.

Evans, B. J., Brown, R. M., McGuire, J. A., Supriatna, J., Andayani, N., Diesmos, A., Iskandar, D., Melnick, D. J., and Cannatella, D. C. (2003). Phylogenetics of fanged frogs: testing biogeographical hypotheses at the interface of the Asian and Australian faunal zones. *Systematic Biology* 52(6), 794–819.

Evans, B. J., Carter, T. F., Tobias, M. L., Kelley, D. B., Hanner, R., and Tinsley, R. C. (2008). A new species of clawed frog (genus Xenopus) from the Itombwe Massif, Democratic Republic of the Congo: implications for DNA barcodes and biodiversity conservation. *Zootaxa* 1780(1), 55–68.

Evans, B. J., Greenbaum, E., Kusamba, C., Carter, T. F., Tobias, M. L., Mendel, S. A., and Kelley, D. B. (2011). Description of a new octoploid frog species (Anura: Pipidae: Xenopus) from the Democratic Republic of the Congo, with a discussion of the biogeography of African clawed frogs in the Albertine Rift. *Journal of Zoology* 283(4), 276–290.

Evans, B. J., Kelley, D. B., Melnick, D. J., and Cannatella, D. C. (2005). Evolution of RAG-1 in polyploid clawed frogs. *Molecular Biology and Evolution* 22(5), 1193–1207.

Evans, B. J., Kelley, D. B., Tinsley, R. C., Melnick, D. J., and Cannatella, D. C. (2004). A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Molecular phylogenetics and evolution* 33(1), 197–213.

Evans, B. J., Pyron, R. A., and Wiens, J. J. (2012). Polyploidization and sex chromosome evolution in amphibians. In: *Polyploidy and genome evolution*. Springer, 385–410.

Evans, B. J., Carter, T. F., Greenbaum, E., Gvoždík, V., Kelley, D. B., McLaughlin, P. J., Pauwels, O. S., Portik, D. M., Stanley, E. L., Tinsley, R. C., et al. (2015). Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African clawed frog (Xenopus, Pipidae) from West and Central Africa. *PLoS One* 10(12), e0142823.

Evans, B. J., Gansauge, M.-T., Stanley, E. L., Furman, B. L., Cauret, C. M., Ofori-Boateng, C., Gvoždík, V., Streicher, J. W., Greenbaum, E., Tinsley, R. C., et al. (2019). Xenopus fraseri: Mr. Fraser, where did your frog come from? *PloS one* 14(9), e0220892.

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* 78(2), 737–756.

Feng, Y.-J., Blackburn, D. C., Liang, D., Hillis, D. M., Wake, D. B., Cannatella, D. C., and Zhang, P. (2017). Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary. *Proceedings of the National Academy of Sciences USA*, 201704632.

Fisher, R. (1931). The evolution of dominance. *Biol. Rev. Biol. Proc. Camb. Philos. Soc.* 6, 345–368.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y.-l., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151(4), 1531–1545.

Foster, J. W. and Graves, J. (1994). An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. *Proceedings of the National Academy of Sciences* 91(5), 1927–1931.

Frost, D. R. (2016). *Amphibian Species of the World: an Online Reference. Version 6.0 (Date of access:7/12/2016). Electronic Database.* American Museum of Natural History, New York, USA. URL: `http://research.amnh.org/herpetology/amphibia/index.html`.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences USA*, 201221359.

Furman, B. L., Bewick, A. J., Harrison, T. L., Greenbaum, E., Gvoždík, V., Kusamba, C., and Evans, B. J. (2015). Pan-African phylogeography of a model organism, the African clawed frog 'Xenopus laevis'. *Molecular Ecology* 24(4), 909–925.

Furman, B. L. and Evans, B. J. (2016). Sequential turnovers of sex chromosomes in African clawed frogs (*Xenopus*) suggest some genomic regions are good at sex determination. *G3: Genes| Genomes| Genetics* 6(11), 3625–3633.

Furman, B. L. and Evans, B. J. (2018). Divergent evolutionary trajectories of two young, homomorphic, and closely related sex chromosome systems. *Genome Biology and Evolution* 10(3), 742–755.

Gamble, T., Coryell, J., Ezaz, T., Lynch, J., Scantlebury, D. P., and Zarkower, D. (2015). Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. *Molecular Biology and Evolution* 32(5), 1296–1309.

Gammerdinger, W. J., Conte, M. A., Baroiller, J.-F., D'Cotta, H., and Kocher, T. D. (2016). Comparative analysis of a sex chromosome from the blackchin tilapia, *Sarotherodon melanotheron*. *BMC genomics* 17(1), 808.

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., Sharpe, T., Hall, G., Shea, T. P., Sykes, S., et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences USA* 108(4), 1513–1518.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* 29(7), 644.

Graves, J. A. M. (2006). Sex Chromosome Specialization and Degeneration in Mammals. *Cell* 124(5), 901–914. ISSN: 0092-8674.

Graves, J. A. M. (2008). Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annual Review of Genetics* 42, 565–586.

Graves, J. A. M. and Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biology* 11(4), 205. ISSN: 1474-760X.

Green, D. M., Zeyl, C. W., and Sharbel, T. F. (1993). The evolution of hypervariable sex and supernumerary (B) chromosomes in the relict New Zealand frog, Leiopelma hochstetteri. *Journal of evolutionary biology* 6(3), 417–441.

Haag, E. S. and Doty, A. V. (2005). Sex determination across evolution: Connecting the dots. *PloS Biology* 3(1), e21.

Hadjivasiliou, Z., Pomiankowski, A., Seymour, R. M., and Lane, N. (2012). Selection for mitonuclear co-adaptation could favour the evolution of two sexes. *Proceedings of the Royal Society B: Biological Sciences* 279(1734), 1865–1872.

Hediger, M., Burghardt, G., Siegenthaler, C., Buser, N., Hilfiker-Kleiner, D., Dübendorfer, A., and Bopp, D. (2004). Sex determination in *Drosophila melanogaster* and *Musca domestica* converges at the level of the terminal regulator doublesex. *Development Genes and Evolution* 214(1), 29–42.

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L., et al. (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328(5978), 633–636.

Hickey, D. A. and Golding, G. B. (2018). The advantage of recombination when selection is acting at many genetic loci. *Journal of theoretical biology* 442, 123–128.

Hill, W. G. and Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research* 8(3), 269–294.

Holleley, C. E., O'Meally, D., Sarre, S. D., Graves, J. A. M., Ezaz, T., Matsubara, K., Azad, B., Zhang, X., and Georges, A. (2015). Sex reversal triggers the rapid transition from genetic to temperature-dependent sex. *Nature* 523(7558), 79–82.

Husnik, F. and McCutcheon, J. P. (2018). Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology* 16(2), 67.

Janes, D. E., Organ, C. L., Stiglec, R., O'Meally, D., Sarre, S. D., Georges, A., Graves, J. A., Valenzuela, N., Literman, R. A., Rutherford, K., et al. (2014). Molecular evolution of Dmrt1 accompanies change of sex-determining mechanisms in reptilia. *Biology letters* 10(12), 20140809.

Jeffries, D. L., Lavanchy, G., Sermier, R., Sredl, M. J., Miura, I., Borzée, A., Barrow, L. N., Canestrelli, D., Crochet, P.-A., Dufresnes, C., et al. (2018). A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nature Communications* 9(1), 4088.

Johnson, N. A. and Porter, A. H. (2007). Evolution of branched regulatory genetic pathways: directional selection on pleiotropic loci accelerates developmental system drift. *Genetica* 129(1), 57–70.

Jordan, C. Y. and Charlesworth, D. (2012). The potential for sexually antagonistic polymorphism in different genome regions. *Evolution: International Journal of Organic Evolution* 66(2), 505–516.

Joron, M., Papa, R., Beltrán, M., Chamberlain, N., Mavárez, J., Baxter, S., Abanto, M., Bermingham, E., Humphray, S. J., Rogers, J., et al. (2006). A conserved supergene locus controls colour pattern diversity in Heliconius butterflies. *PLoS Biol* 4(10), e303.

Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30(4), 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12), 1647–1649.

Kelley, D. R., Schatz, M. C., and Salzberg, S. L. (2010). Quake: quality-aware detection and correction of sequencing errors. *Genome biology* 11(11), 1.

Kitano, J. and Peichel, C. L. (2012). Turnover of sex chromosomes and speciation in fishes. *Environmental biology of fishes* 94(3), 549–558.

Kobel, H. R. and Du Pasquier, L. (1986). Genetics of polyploid *Xenopus*. *Trends in Genetics* 2, 310–315.

Koch, P., Platzer, M., and Downie, B. R. (2014). RepARK-de novo creation of repeat libraries from whole-genome NGS reads. *Nucleic acids research* 42(9), e80–e80.

Küpper, C., Stocks, M., Risse, J. E., Dos Remedios, N., Farrell, L. L., McRae, S. B., Morgan, T. C., Karlionova, N., Pinchuk, P., Verkuil, Y. I., et al. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature genetics* 48(1), 79–83.

Labonne, J., Tamari, F., and Shore, J. (2010). Characterization of X-ray-generated floral mutants carrying deletions at the S-locus of distylous Turnera subulata. *Heredity* 105(2), 235–243.

Lambeth, L. S., Raymond, C. S., Roeszler, K. N., Kuroiwa, A., Nakata, T., Zarkower, D., and Smith, C. A. (2014). Over-expression of *DMRT1* induces the male pathway in embryonic chicken gonads. *Developmental Biology* 389(2), 160–172.

Liew, W. C., Bartfai, R., Lim, Z., Sreenivasan, R., Siegfried, K. R., and Orban, L. (2012). Polygenic sex determination system in zebrafish. *PloS one* 7(4), e34397.

Liu, Y., Luo, D., Lei, Y., Hu, W., Zhao, H., and Cheng, C. H. (2014). A highly effective TALEN-mediated approach for targeted gene disruption in Xenopus tropicalis and zebrafish. *Methods* 69(1), 58–66.

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(1), 1.

Lynch, M. (2007). *The origins of genome architecture*. Vol. 98. Sinauer Associates Sunderland, MA.

Lynch, M. and Hagner, K. (2015). Evolutionary meandering of intermolecular interactions along the drift barrier. *Proceedings of the National Academy of Sciences USA* 112(1), E30–E38.

Lynch, M., O'Hely, M., Walsh, B., and Force, A. (2001). The probability of preservation of a newly arisen gene duplicate. *Genetics* 159(4), 1789–1804.

Maddison, W. P. (2008). Mesquite: A modular system for evolutionary analysis. *Evolution* 62, 1103–1118.

Marçais, G. and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6), 764–770.

Masuyama, H., Yamada, M., Kamei, Y., Fujiwara-Ishikawa, T., Todo, T., Nagahama, Y., and Matsuda, M. (2012). Dmrt1 mutation causes a male-to-female sex reversal after the sex determination by Dmy in the medaka. *Chromosome research* 20(1), 163–176.

Matson, C. K., Murphy, M. W., Sarver, A. L., Griswold, M. D., Bardwell, V. J., and Zarkower, D. (2011). DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature* 476(7358), 101.

Matsubara, K., Tarui, H., Toriba, M., Yamada, K., Nishida-Umehara, C., Agata, K., and Matsuda, Y. (2006). Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proceedings of the National Academy of Sciences USA* 103(48), 18190–18195.

Matsuda, M., Sato, T., Toyazaki, Y., Nagahama, Y., Hamaguchi, S., and Sakaizumi, M. (2003). Oryzias curvinotus has DMY, a gene that is required for male development in the medaka, O. latipes. *Zoological science* 20(2), 159–161.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C. E., Shibata, N., Asakawa, S., Shimizu, N., et al. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417(6888), 559–563.

Matveevsky, S., Kolomiets, O., Bogdanov, A., Hakhverdyan, M., and Bakloushinskaya, I. (2017). Chromosomal evolution in mole voles Ellobius (Cricetidae, Rodentia): Bizarre sex chromosomes, variable autosomes and meiosis. *Genes* 8(11), 306.

Mawaribuchi, S., Musashijima, M., Wada, M., Izutsu, Y., Kurakata, E., Park, M. K., Takamatsu, N., and Ito, M. (2017a). Molecular evolution of two distinct *DMRT1* promoters for germ and somatic cells in vertebrate gonads. *Molecular Biology and Evolution* 34(3), 724–733.

Mawaribuchi, S., Takahashi, S., Wada, M., Uno, Y., Matsuda, Y., Kondo, M., Fukui, A., Takamatsu, N., Taira, M., and Ito, M. (2017b). Sex chromosome differentiation and the W-and Z-specific loci in Xenopus laevis. *Developmental biology* 426(2), 393–400.

McCabe, J. and Dunn, A. (1997). Adaptive significance of environmental sex determination in an amphipod. *Journal of Evolutionary Biology* 10(4), 515–527.

Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* 2010(6), pdb–prot5448.

Mitros, T., Lyons, J. B., Session, A. M., Jenkins, J., Shu, S., Kwon, T., Lane, M., Ng, C., Grammer, T. C., Khokha, M. K., et al. (2019). A chromosome-scale genome assembly and dense genetic map for Xenopus tropicalis. *Developmental Biology*.

Miura, I. (2007). An evolutionary witness: the frog Rana rugosa underwent change of heterogametic sex from XY male to ZW female. *Sexual Development* 1(6), 323–331.

Miyamoto, Y., Taniguchi, H., Hamel, F., Silversides, D. W., and Viger, R. S. (2008). A GATA4/WT1 cooperation regulates transcription of genes required for mammalian sex determination and differentiation. *BMC Molecular Biology* 9(1), 44.

Moniot, B., Berta, P., Scherer, G., Südbeck, P., and Poulat, F. (2000). Male specific expression suggests role of *DMRT1* in human sex determination. *Mechanisms of development* 91(1), 323–325.

Nakayama, T., Blitz, I. L., Fish, M. B., Odeleye, A. O., Manohar, S., Cho, K. W., and Grainger, R. M. (2014). Cas9-based genome editing in Xenopus tropicalis. In: *Methods in enzymology*. Vol. 546. Elsevier, 355–375.

Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von, and Minh, B. Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32(1), 268–274.

Ogata, H., Fujibuchi, W., Goto, S., and Kanehisa, M. (2000). A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucleic acids research* 28(20), 4021–4028.

Ogata, M., Hasegawa, Y., Ohtani, H., Mineyama, M., and Miura, I. (2008). The ZZ/ZW sex-determining mechanism originated twice and independently during evolution of the frog, Rana rugosa. *Heredity* 100(1), 92–99.

Olmstead, A. W., Lindberg-Livingston, A., and Degitz, S. J. (2010). Genotyping sex in the amphibian, *Xenopus (Silurana) tropicalis*, for endocrine disruptor bioassays. *Aquatic Toxicology* 98(1), 60–66.

O'Meally, D., Ezaz, T., Georges, A., Sarre, S. D., and Graves, J. A. M. (2012). Are some chromosomes particularly good at sex? Insights from Amniotes. *Chromosome Research* 20(1), 7–19.

Onuma, Y., Takahashi, S., Asashima, M., Kurata, S., and Gehring, W. J. (2002). Conservation of Pax 6 function and upstream activation by Notch signaling in eye development of frogs and flies. *Proceedings of the National Academy of Sciences* 99(4), 2020–2025.

Otto, S. P. and Lenormand, T. (2002). Resolving the paradox of sex and recombination. *Nature Reviews Genetics* 3(4), 252–261.

Paradis, E. and Schliep, K. (2018). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35(3), 526–528.

Parker, G. A. (1972). The origin and evolution of gamete dimorphism and the male-female phenomenon. J Theor Biol. *J. theor. Biol* 36, 553.

Pavlicev, M. and Wagner, G. P. (2012). A model of developmental evolution: selection, pleiotropy and compensation. *Trends in Ecology and Evolution* 27(6), 316–322.

Pearl, E. J., Grainger, R. M., Guille, M., and Horb, M. E. (2012). Development of Xenopus resource centers: the national Xenopus resource and the European Xenopus resource center. *genesis* 50(3), 155–163.

Pennell, M. W., Mank, J. E., and Peichel, C. L. (2018). Transitions in sex determination and sex chromosomes across vertebrate species. *Molecular Ecology*.

Perrin, N. (2009). Sex reversal: a fountain of youth for sex chromosomes? *Evolution* 63(12), 3043–3049.

Piatigorsky, J. and Wistow, G. J. (1989). Enzyme/crystallins: gene sharing as an evolutionary strategy. *Cell* 57(2), 197–199.

Picq, S., Santoni, S., Lacombe, T., Latreille, M., Weber, A., Ardisson, M., Ivorra, S., Maghradze, D., Arroyo-Garcia, R., Chatelet, P., This, P., Terral, J.-F., and Bacilieri, R. (2014). A small XY chromosomal region explains sex determination in wild dioecious *V. vinifera* and the reversal to hermaphroditism in domesticated grapevines. *BMC Plant Biology* 14(1), 229. ISSN: 1471-2229.

Pla, S., Benvenuto, C., Capellini, I., and Piferrer, F. (2020). A phylogenetic comparative analysis on the evolution of sequential hermaphroditism in seabreams (Teleostei: Sparidae). *Scientific reports* 10(1), 1–12.

Ponnikas, S., Sigeman, H., Abbott, J. K., and Hansson, B. (2018). Why do sex chromosomes stop recombining? *Trends in Genetics*.

Radder, R. S., Quinn, A. E., Georges, A., Sarre, S. D., and Shine, R. (2008). Genetic evidence for co-occurrence of chromosomal and thermal sex-determining systems in a lizard. *Biology Letters* 4(2), 176–178.

Refsnider, J. M. and Janzen, F. J. (2015). Temperature-Dependent Sex Determination under Rapid Anthropogenic Environmental Change: Evolution at a Turtle's Pace? *Journal of Heredity* 107(1), 61–70.

Renner, S. S. (2014). The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of botany* 101(10), 1588–1596.

Rice, W. R. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, 735–742.

Rice, W. R. (1987). The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* 41(4), 911–914.

Roco, Á. S., Olmstead, A. W., Degitz, S. J., Amano, T., Zimmerman, L. B., and Bullejos, M. (2015). Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis*. *Proceedings of the National Academy of Sciences* 112(34), E4752–E4761.

Rodrigues, N., Vuille, Y., Brelsford, A., Merilä, J., and Perrin, N. (2016). The genetic contribution to sex determination and number of sex chromosomes vary among populations of common frogs (Rana temporaria). *Heredity* 117(1), 25–32.

Rodrigues, N., Studer, T., Dufresnes, C., Ma, W.-J., Veltsos, P., and Perrin, N. (2017). *Dmrt1* polymorphism and sex-chromosome differentiation in Rana temporaria. *Molecular Ecology* 26(19), 4897–4905.

Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D., and Peichel, C. L. (2009). Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLoS Genet* 5(2), e1000391.

Saenko, S. V., Chouteau, M., Piron-Prunier, F., Blugeon, C., Joron, M., and Llaurens, V. (2019). Unravelling the genes forming the wing pattern supergene in the polymorphic butterfly Heliconius numata. *EvoDevo* 10(1), 16.

Schild, L., Canessa, C. M., Shimkets, R. A., Gautschi, I., Lifton, R. P., and Rossier, B. C. (1995). A mutation in the epithelial sodium channel causing Liddle disease increases channel activity in the Xenopus laevis oocyte expression system. *Proceedings of the National Academy of Sciences* 92(12), 5699–5703.

Session, A. M., Uno, Y., Kwon, T., Chapman, J. A., 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(7625), 336–343.

Shirak, A., Seroussi, E., Cnaani, A., Howe, A. E., Domokhovsky, R., Zilberman, N., Kocher, T. D., Hulata, G., and Ron, M. (2006). Amh and Dmrta2 genes map to tilapia (*Oreochromis spp.*) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics* 174(3), 1573–1581.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome research* 19(6), 1117–1123.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423(6942), 825–837.

Smith, C. A., Roeszler, K. N., Ohnesorg, T., Cummins, D. M., Farlie, P. G., Doran, T. J., and Sinclair, A. H. (2009). The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. *Nature* 461(7261), 267.

Stöck, M., Horn, A., Grossen, C., Lindtke, D., Sermier, R., Betto-Colliard, C., Dufresnes, C., Bonjour, E., Dumas, Z., Luquet, E., Maddalena, T., Sousa, H. C., Martinez-Solano, I., and Perrin, N. (2011). Ever-Young Sex Chromosomes in European Tree Frogs. *PLOS Biology* 9(5), 1–9.

Sun, Y.-B., Xiong, Z.-J., Xiang, X.-Y., Liu, S.-P., Zhou, W.-W., Tu, X.-L., Zhong, L., Wang, L., Wu, D.-D., Zhang, B.-L., et al. (2015). Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proceedings of the National Academy of Sciences USA* 112(11), E1257–E1262.

Sutou, S., Mitsui, Y., and Tsuchiya, K. (2001). Sex determination without the Y chromosome in two Japanese rodents *Tokudaia osimensis osimensis* and *Tokudaia osimensis* spp. *Mammalian Genome* 12(1), 17–21.

Swofford, D. L. and Sullivan, J. (2003). Phylogeny inference based on parsimony and other methods using PAUP*. *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny, cáp* 7, 160–206.

Takehana, Y., Matsuda, M., Myosho, T., Suster, M. L., Kawakami, K., Shin, T., Kohara, Y., Kuroki, Y., Toyoda, A., Fujiyama, A., et al. (2014). Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish Oryzias dancena. *Nature communications* 5(1), 1–10.

Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S., and Sakaizumi, M. (2007). Evidence for different origins of sex chromosomes in closely related Oryzias fishes: substitution of the master sex-determining gene. *Genetics* 177(4), 2075–2081.

Tennessen, J. A., Wei, N., Straub, S. C., Govindarajulu, R., Liston, A., and Ashman, T.-L. (2018). Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS biology* 16(8), e2006062.

Thévenin, A., Ein-Dor, L., Ozery-Flato, M., and Shamir, R. (2014). Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. *Nucleic acids research* 42(15), 9854–9861.

Thomson, T. M., Lozano, J. J., Loukili, N., Carrió, R., Serras, F., Cormand, B., Valeri, M., Díaz, V. M., Abril, J., Burset, M., et al. (2000). Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene. *Genome research* 10(11), 1743–1756.

Tocchini-Valentini, G. D., Fruscoloni, P., and Tocchini-Valentini, G. P. (2005). Structure, function, and evolution of the tRNA endonucleases of Archaea: an example of subfunctionalization. *Proceedings of the National Academy of Sciences* 102(25), 8933–8938.

Tomita, T. and Wada, Y. (1989). Multifactorial sex determination in natural populations of the housefly (*Musca domestica*) in Japan. *The Japanese Journal of Genetics* 64(5), 373–382.

True, J. R. and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evolution and Development* 3(2), 109–119.

Tymowska, J. (1991). Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. *Amphibian Cytogenetics and Evolution* 259, 297.

Uno, Y., Nishida, C., Takagi, C., Igawa, T., Ueno, N., Sumida, M., and Matsuda, Y. (2015). Extraordinary diversity in the origins of sex chromosomes in anurans inferred from comparative gene mapping. *Cytogenetic and Genome Research* 145(3-4), 218–229.

Van Doorn, G. and Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual conflict. *Nature* 449(7164), 909–912.

Velanis, C. N., Perera, P., Thomson, B., Leau, E. de, Liang, S. C., Hartwig, B., Förderer, A., Thornton, H., Arede, P., Chen, J., et al. (2020). The domesticated transposase ALP2 mediates formation of a novel Polycomb protein complex by direct interaction with MSI1, a core subunit of Polycomb Repressive Complex 2 (PRC2). *PLoS Genetics* 16(5), e1008681.

Veltsos, P., Rodrigues, N., Studer, T., Ma, W.-J., Sermier, R., Leuenberger, J., and Perrin, N. (2019). No evidence that Y-chromosome differentiation affects male fitness in a Swiss population of common frogs. *Journal of evolutionary biology*.

Verster, A. J., Ramani, A. K., McKay, S. J., and Fraser, A. G. (2014). Comparative RNAi screens in C. elegans and C. briggsae reveal the impact of developmental system drift on gene function. *PLoS Genet* 10(2), e1004077.

Vicoso, B., Kaiser, V. B., and Bachtrog, D. (2013). Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proceedings of the National Academy of Sciences USA* 110(16), 6453–6458.

Villar, D., Flicek, P., and Odom, D. T. (2014). Evolution of transcription factor binding in metazoans —- mechanisms and functional implications. *Nature Reviews Genetics* 15(4), 221.

Volff, J.-N., Nanda, I., Schmid, M., and Schartl, M. (2007). Governing sex determination in fish: regulatory putsches and ephemeral dictators. *Sexual Development* 1(2), 85–99.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D., and Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* 493(7434), 664–668.

Wang, W., Yu, H., and Long, M. (2004). Duplication-degeneration as a mechanism of gene fission and the origin of new genes in Drosophila species. *Nature genetics* 36(5), 523–527.

Wang, W., Zhang, J., Alvarez, C., Llopart, A., and Long, M. (2000). The origin of the Jingwei gene and the complex modular structure of its parental gene, yellow emperor, in Drosophila melanogaster. *Molecular Biology and Evolution* 17(9), 1294–1301.

Wang, X. and Sommer, R. J. (2011). Antagonism of LIN-17/Frizzled and LIN-18/Ryk in nematode vulva induction reveals evolutionary alterations in core developmental pathways. *PLoS Biol* 9(7), e1001110.

Weiss, K. M. and Fullerton, S. M. (2000). Phenogenetic drift and the evolution of genotype–phenotype relationships. *Theoretical Population Biology* 57(3), 187–195.

Weldon, C., Du Preez, L. H., Hyatt, A. D., Muller, R., and Speare, R. (2004). Origin of the amphibian chytrid fungus. *Emerging infectious diseases* 10(12), 2100.

West, S. A., Herre, E. A., and Sheldon, B. C. (2000). The benefits of allocating sex. *Science* 290(5490), 288–290.

Wickbom, T. (1950). The chromosomes of *Pipa pipa*. *Hereditas* 36, 363–366.

Xu, P.-X., Woo, I., Her, H., Beier, D. R., and Maas, R. L. (1997). Mouse Eya homologues of the *Drosophila* eyes absent gene require *Pax6* for expression in lens and nasal placode. *Development* 124(1), 219–231.

Yazdi, H. P. and Ellegren, H. (2014). Old but Not (So) Degenerated−Slow Evolution of Largely Homomorphic Sex Chromosomes in Ratites. *Molecular biology and evolution* 31(6), 1444–1453.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics* 13(1), 134.

Yoshimoto, S., Ikeda, N., Izutsu, Y., Shiba, T., Takamatsu, N., and Ito, M. (2010). Opposite roles of DMRT1 and its W-linked paralogue, DM-W, in sexual dimorphism of Xenopus laevis: implications of a ZZ/ZW-type sex-determining system. *Development* 137(15), 2519–2526.

Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., Matsuda, Y., Takamatsu, N., Shiba, T., and Ito, M. (2008). A W-linked DM-domain gene, *DM-W*, participates in primary ovary development in Xenopus laevis. *Proceedings of the National Academy of Sciences USA* 105(7), 2469–2474.

Yuan, S., Xia, Y., Zheng, Y., and Zeng, X. (2016). Next-generation sequencing of mixed genomic DNA allows efficient assembly of rearranged mitochondrial genomes in *Amolops chunganensis* and *Quasipaa boulengeri*. *PeerJ* 4, e2786.

Zarkower, D. (2001). Establishing sexual dimorphism: conservation amidst diversity? *Nature Reviews Genetics* 2(3), 175–185.

Zerbino, D. and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, gr–074492.

Zhang, X., Guan, G., Li, M., Zhu, F., Liu, Q., Naruse, K., Herpin, A., Nagahama, Y., Li, J., and Hong, Y. (2016). Autosomal gsdf acts as a male sex initiator in the fish medaka. *Scientific Reports* 6, 19738.